Utah State University

DigitalCommons@USU

5-2011

Mathematical Justification of Introductory Hypothesis Tests and Development of Reference Materials

Jennifer L. Loveland
Utah State University

Recommended Citation

UtahStateUniversity
MERRILL-CAZIER LIBRARY

MATHEMATICAL JUSTIFICATION OF INTRODUCTORY HYPOTHESIS TESTS AND
DEVELOPMENT OF REFERENCE MATERIALS

by

Jennifer L. Loveland

A project submitted in partial fulfillment of the requirements for the degree

of

MASTER OF SCIENCE

in

Mathematics

Approved:

_____   _____
Kady Schneiter   Chris Corcoran
Major Professor   Committee Member

_____
Brynja Kohler
Committee Member

UTAH STATE UNIVERSITY

2011

ABSTRACT

Mathematical Justification of Introductory Hypothesis Tests and

Development of Reference Materials

by

Jennifer L. Loveland, Master of Science

Utah State University, 2011

Major Professor: Dr. Kady Schneiter
Department: Mathematics

This purpose of this project was to create a set of reference material for an introductory level statistics course. The focus was on the Neyman-Pearson approach to hypothesis testing. A brief historical development of the Neyman-Pearson approach is followed by mathematical proofs of each of the hypothesis tests covered in the reference material. The reference material includes the basic hypothesis tests taught in an introductory statistics course, the accompanying distributions, and prerequisite information.

(132 pages)

ACKNOWLEDGMENTS

Contents

CHAPTER 1

Introduction

This project consisted of three sections. The bulk of this project is a set of reference materials that I developed for introductory level hypothesis testing. There is also a section with the mathematical proofs for every test included in the reference materials. The last section is a historical development of the Neyman-Pearson approach to hypothesis testing that is commonly taught in introductory classes.

The historical development sheds light on the methods used in the Neyman-Pearson approach. There are various conventions taught in introductory level classes and this section discusses their origins.

The mathematical proofs section contains a proof of every hypothesis test. It also contains any preliminary proofs that were needed. Properties of expected value and moment generating functions are included because I used them in the proofs of the hypothesis tests. Some of the proofs are accessible to introductory level students and I expect those students to look at the proofs, but many of the proofs are at the math 5720 level. As a result, and because there are so many preliminary proofs needed for the proof of each hypothesis test, rather than put some proofs with the reference material and some in a different section, I placed all the proofs in one section to achieve a linear ordering of the proofs.

Teachers at the introductory level can pick and choose the proofs that they want to share with their students. For example, walking a student through the proof that

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

will help a student remember to use a t test when the variance is unknown and that there are $n-1$ degrees of freedom. A student should also be able to understand the proofs for the z test for independent samples. This should help the student remember the test statistic or at the very least not view the test statistic as something pulled out of the teacher's magic hat. Some of the tests' proofs have more complicated algebra than others, but at least part of the proof for each hypothesis test could be shared with introductory level students. The arguments for the degrees of freedom for the chi square tests for multinomial data could easily be shared and would make the concept of degrees of freedom less mysterious to student.

The section of mathematical proofs would also be very valuable to a student in a upper level statistics class as he could see the proofs for every test in one place as well as all the background proofs that are needed.

While I want to eventually create a set of reference materials for an entire statistics course, I focused on hypothesis testing for this project. The basic tests are each covered as well as the distributions needed for the tests. My goal was to create a set of reference material that was not a textbook, but a supplement to a statistics class or book. Each concept is covered in 1-2 pages to facilitate easy checking of facts and to help with basic procedural knowledge.

While tutoring, I realized that many students view each hypothesis test as a different concept and test statistics are random complicated formulas given by professors. It is difficult for the students to remember which test is applicable in a given situation and they don't realize that each test is the same process. It is my hope that with the layout of the reference material that they can easily flip through the tests and compare assumptions to realize when each test is appropriate. The layout should also help students realize that each test follows the same procedure.

The reference material can be used with any introductory statistics course, but I designed it with Stat 2000 in mind because that course uses mathematical formulas and operations but doesn't require calculus. I used several textbooks as reference, but the bulk of the material comes from Moore and McCabe's *Introduction to the Practice of Statistics.*

Although the reference material is the main project, I designed the layout in a way that does not agree with the required format of a project report and so it is included in the appendix.

I have placed this project on my website, http://mathnstats.com. I hope to soon have reference material for an entire statistics course along with relevant proofs available online.

CHAPTER 2

A Partial Development of the Neyman-Pearson Approach to Hypothesis Testing

Historical Development of Neyman-Pearson Approach

Karl Pearson and the Chi-Square Test for Goodness of Fit. In 1900, Karl Pearson published a paper, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". In this paper he proposed his P, χ^2 test, now commonly referred to as the chi-square test for goodness of fit. The χ^2 test can be used to measure how well a proposed model fits a set of data.

In 1892, Pearson explains that mathematical models can describe, but not explain natural phenomenon. Mathematical models could be used to predict and infer without actually understanding the "biological mechanism". Models are good if they are a good fit to the data, but the model has no truth value. [**17**, 8]

It is clear from one of his papers that Karl based his methods on using large samples from a population.

R.A. Fisher and Tests of Significance. R.A. Fisher coined the phrase "tests of significance". A test of significance consisted of proposing one (and only one) probability model that would fit a data set. There is no alternative hypothesis. The data is reduced to a test statistic. Then the distribution of the test statistic is used to determine how likely the data is to occur under the model. Fisher was the first to refer to the "p-value" as the probability that determines significance. In one example in his book, *Statistical Methods for Research Workers,* he found that the p-value was less than 0.01 and interprets it as, "Only one value in a hundred will exceed [the calculated test statistic] by chance, so that the difference between the results is clearly significant." [**8**, 122]

From examples in Fisher's book and papers, it appears that he had three possible conclusions for tests of significance.

(1) If the p-value is small, generally less than 0.01, an effect has been shown.

(2) If the p-value is large, generally more than 0.20, any effect is so small, if it exists at all, that it cannot be detected without a larger sample size.

(3) If the p-value is between 0.01 and 0.20, Fisher discusses how to design the next experiment to obtain an improved understanding of the effect. [**20**, 100]

In 1929, Fisher stated:

> An observation is judged significant, if it would rarely have been produced, in the absences of a real cause of the kind we are seeking. It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once in twenty trials. This is an arbitrary, but convenient, level of significance for the practical investigator, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation. [**9**, 191]

According to Fisher:

(1) To be significant, an observation would have to be unlikely to occur if the effect of interest is absent. (The hypothesis is true.)

(2) A common significance level is 0.05.

(3) All experiments without significant results should be ignored. This implies that there is no "acceptance" of the hypothesis.

(4) To claim there is an effect, an experimenter should be able to design an experiment so that it will give significant results the majority of the time.

It is interesting to note that while hypothesis testing is often thought to be done on a sample from a population, Fisher believed that in scientific applications, there is no true underlying population. "...the only populations that can be referred to in a test of significance have no objective reality, being exclusively the product of the statistician's imagination through the hypotheses he has decided to test". [**17**, 8] Contrary to Pearson's approach with small samples, Fisher developed methods to use hypothesis testing with smaller samples.

Pearson Versus Fisher. In 1935, Hugo John Buchanan-Wollaston wrote a letter that was printed in *Nature* in 1935 about a problem he saw in Karl Pearson's P, χ^2 test. Buchanan-Wollaston argued that there are many models that could give an acceptable value of the test statistic and would therefore fit the data according to a χ^2 test. He believed that one of the possible models should not be chosen as correct just because a χ^2 hypothesis test failed to reject the model. Therefore, according to Buchanan-Wollaston, Pearson's χ^2 test is only valid to test the lack of fit of a model. It could not be used to assert that a model does fit the data. He thought that this logic applied to all other statistical tests. The replies of Fisher and Pearson give insight into their methods and reasoning.

Fisher published a letter in reply in 1935.

> Mr. Buchanan-Wollaston's point that the χ^2 test, like the other tests of significance, is cogent for the rejection of hypotheses, but, in the opposite case, by no means cogent for their acceptance, deserves to be widely appreciated. For the logical fallacy of believing that a hypothesis has been proved to be true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistical than in other kinds of scientific reasoning. Yet it does so only too frequently. Indeed, the "error of accepting an hypothesis when it is false" has been specially named by some writers "errors of the second kind". It would, therefore, add greatly to the clarity with which the tests of significance are regarded if it were generally understood that tests of significance, when used accurately, are capable of rejecting or invalidating hypotheses, in so far as these are contradicted by the data; but that they are never capable of establishing them as certainly true. In fact that "errors of the second kind" are committed only by those who misunderstand the nature and application of tests of significance.
> [**17**, 5]

Fisher agreed with Buchanan-Wollaston's view that the χ^2 test could not be used to argue that a model does fit the data or that the hypothesis should be accepted. He even argues that a type II error, accepting the hypothesis when it is false, shouldn't be an issue, because the hypothesis is never accepted. A test of significance can only be used to reject a hypothesis when it is "contradicted by the data". Later in 1973, he stated "a test of significance contains no criterion for accepting a hypothesis". [**17**, 8]

Karl Pearson replied with two letters published in *Nature* in 1935.

> The difference between Prof. Fisher and myself lies in the use (and abuse) of the acceptance and rejection of 'hypotheses'. There is only one case in which an hypothesis can be definitely rejected, namely when its probability is zero...Now Prof. Fisher refers to rejecting hypotheses as a function of the P, χ^2 method, and

of accepting them as a logical fallacy. I have in my letter of August 24 stated that the tests are there to ascertain whether a reasonable *graduation* has been reached; not to assert whether one or another hypothesis is true or false... It is not for statisticians to say whether an hypothesis is false except when $P = 0$. All that they can legitimately say is that it gives a poor *graduation*. In particular, it is very unwise in *my* opinion to form tables which provide only the values of $P = 0.01$ and $P = 0.05$, and consider 'hypotheses' which give a value of $P < 0.01$ as 'false', and those with a value between 0.01 and 0.05 as 'doubtful', and for the rest of the scale of P have no descriptive category, for you must not say that such values prove hypotheses to be true. Hence I repeat my assertion, in the face of all the authority of Prof. Fisher and his followers, that all the P, χ^2 test asertains is goodness of graduation, and I hold that 'goodness' of graduation is relative to the nature of the material investigated, our experience of similar material and the purpose to which we intend to put our graduation. The value of P at which we consider goodness or badness of graduation starts cannot be fixed without regard to the special problem under consideration.

There seems somewhere a logical fallacy in the position of both Prof. Fisher and Mr. Buchanan-Wollaston. They both apparently assert that the P, χ^2 test enables one to say that an hypothesis is false, yet never to say that an hypothesis is true, but if an hypothesis be *false*, its reverse must be true....As a matter of fact, the P, χ^2 has only measured its 'goodness of fit' by a probability coefficient, and it is as idle to say as a result of it, that the hypothesis is 'false', as that the reverse of it is 'true'.

The 'laws of Nature' are only constructs of our minds; none of them can be asserted to be true or false, they are good in so far as they give good fits to our observations of Nature, and are liable at any time to be replaced by a better 'fit', that is, by a construct giving a better graduation. [**17**, 6]

Pearson makes several interesting points in this reply:

(1) He agrees with Fisher that a hypothesis is not accepted, but only because he believes that hypotheses should neither be accepted nor rejected.

(2) Pearson intends the p-value to simply be a measure of how well a model fits the data, and not a criterion of significance.

(3) There is no one "correct model" that determines natural phenomenon, so a hypothesis can not be true or false. Instead, all a person can hope to do is find the model that best describes the data.

(4) It is apparent that there is no alternative hypothesis.

Henry Inman summarizes with, "According to Pearson, scientists are not in the business of searching for truth; rather they seek to construct verbal or mathematical summaries of relevant perceptual data." [**17**, 8]

Neyman & Pearson. In 1933, Egon Pearson, son of Karl Pearson, and Jerzy Neyman published a paper, "On the Problem of the most Efficient Tests of Statistical Hypotheses". Their methods are the basis of the Neyman-Pearson approach to hypothesis testing that is commonly taught to introductory students.

> Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviors with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. Here, for example, would be such a "rule of behaviors": to decide whether a hypothesis, H, of a given type be rejected or not, calculate a specified character, x, of the observed facts; if $x > x_0$ reject H, if $x \leq x_0$ accept H. Such a rule tells us nothing as to whether in a particular case H is true when $x \leq x_0$ or false when $x > x_0$. But it may often be proved that if we behave according to such a rule, then in the long run we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false....observed facts are described as "samples," and the hypotheses concern the "populations" from which the samples have been drawn... [**2**, 291]

Neyman and Pearson contributed the ideas:

(1) The null hypothesis is the hypothesis being tested. It is either rejected or accepted.

(2) The idea of an alternative hypothesis.

(3) The alternative hypothesis does not have to be the negation of the null hypothesis.

(4) If there are too many alternative hypotheses, the test will not be powerful.

(5) The formalization of type I and type II errors.

(6) A p-value less than the significance level, does not mean that H_0 is false.

(7) The significance level can be thought of as the number of times in the long run that H_0 will be rejected when it is true (a type I error).

(8) Methods of finding the most efficient test that will maximize the power or the probability of rejecting H_0 when it is false.

(9) The observed data is the sample which comes from a population. The hypotheses concern the population.

The Current Neyman-Pearson Approach to Hypothesis Testing

The approach to hypothesis testing generally taught in introductory level statistic courses is referred to as the Neyman-Pearson approach. However, it encompasses ideas from Karl Pearson, R.A. Fisher, Jerzy Neyman, Egon Pearson, and other researchers.

Ideas from Karl Pearson.

(1) Students are encouraged to report the p-value as well as the conclusion of rejecting or failing to reject H_0 so that readers can see how strong the evidence was for or against H_0. This follows Karl Pearson's idea that the probability is a measure of how well the model fits the data.

(2) The hesitancy to accept a hypothesis has translated into the convention of saying, "we fail to reject H_0", even though Pearson believed that H_0 should be neither rejected nor accepted.

(3) The data comes from a sample chosen from a larger population upon which inferences are made.

Ideas from R.A. Fisher.

(1) A test statistic that summarizes the data is found.

(2) The p-value is the probability of seeing the test statistic or something more extreme if the null hypothesis is true.

(3) Students are taught that the level of significance is the amount of evidence needed to reject a null hypothesis. If $\alpha = 0.05$, the probability of the observation occurring, if the null hypothesis is true, would have to be less than 5% to be judged significant.

(4) The hesitancy to accept a hypothesis has translated into the convention of saying, "we fail to reject H_0", even though Fisher believed that H_0 should be never be accepted and the results of any experiments with non significant p-values should be ignored and the experiment redesigned.

Ideas from Neyman and Pearson.

(1) There is both a null and alternative hypothesis. The alternative hypothesis is assumed to be the negation of the null hypothesis. (This goes against the advice of Neymand and Pearson to not have too many alternative hypotheses.)

(2) The null hypothesis can be accepted, but it is preferable to say "we fail to reject H_0" rather than "we accept H_0". This slight difference in terminology is convention based on Karl Pearson and Fisher's arguments against accepting a hypothesis.

(3) There are type I and type II errors. The power is the probability of rejecting H_0 when it is false.

(4) Tests can be chosen on the criterion of power. This is largely ignored in introductory classes as students are given basic tests to choose from.

(5) The data comes from a sample chosen from a larger population upon which inferences are made.

CHAPTER 3

Mathematical Proofs

Expected value and variance

DEFINITION 1. If X is a random variable with pdf $f(x)$ or pmf $p(x)$, and a function $g(X)$ is measurable, then the expected value of $g(X)$ is defined as

$$E\left(g(X)\right) = \begin{cases} \int\limits_{-\infty}^{\infty} g(x)f(x)dx, & \text{if } X \text{ is continuous} \\ \sum\limits_{-\infty}^{\infty} g(x)p(x)dx, & \text{if } X \text{ is discrete} \end{cases}$$

if $E\left(|g(X)|\right) < \infty$, otherwise $E\left(g(X)\right)$ does not exist.

THEOREM 1. *If a and b are finite real numbers, and X is a random variable with expected value $E(X)$, then $E(aX + b) = aE(X) + b$.*

PROOF. Note that $\int\limits_{X} f(x)dx = 1$ and $\sum\limits_{X} p(x) = 1$ by definition of pdf and pmf.

Existence for Continuous Case

$$
\begin{aligned}
E(|aX + b|) &= \int\limits_{X} \left(|ax + b|\right) f(x)dx \\
&\leq \int\limits_{X} \left(|a||x| + |b|\right) f(x)dx \\
&\qquad \text{by Triangle Inequality} \\
&= \int\limits_{X} |a||x| f(x)dx + \int\limits_{X} |b| f(x)dx \\
&= |a| \int\limits_{X} |x| f(x)dx + |b| \int\limits_{X} f(x)dx \\
&= |a|E(|X|) + |b|(1) \\
&= |a|E(|X|) + |b| \\
&< \infty
\end{aligned}
$$

because $E(X)$ exists, so $E(|X|) < \infty$, and a and b are finite. Therefore, $E(aX + b)$ exists.

Calculation for Continuous Case

$$
\begin{aligned}
E(aX + b) &= \int_X (ax + b)\, f(x)dx \\
&= \int_X ax\, f(x)dx + \int_X b\, f(x)dx \\
&= a \int_X x\, f(x)dx + b \int_X f(x)dx \\
&= aE(X) + b(1) \\
&= aE(X) + b
\end{aligned}
$$

Existence for Discrete Case

$$
\begin{aligned}
E(|aX + b|) &= \sum_X (|ax + b|)\, p(x) \\
&\leq \sum_X (|a||x| + |b|)\, p(x) \\
&\quad \text{by Triangle Inequality} \\
&= \sum_X |a||x|\, p(x) + \sum_X |b|\, p(x) \\
&= |a| \sum_X |x|\, p(x) + |b| \sum_X p(x) \\
&= |a|E(|X|) + |b|(1) \\
&= |a|E(|X|) + |b| \\
&< \infty
\end{aligned}
$$

because $E(X)$ exists, so $E(|X|) < \infty$, and a and b are finite. Therefore, $E(aX + b)$ exists. Therefore, $E(aX + b)$ exists.

Calculation for Discrete Case

$$
\begin{aligned}
E(aX + b) &= \sum_X (ax + b)\, p(x) \\
&= \sum_X ax\, p(x) + \sum_X b\, p(x) \\
&= a \sum_X x\, p(x) + b \sum_X p(x) \\
&= aE(X) + b(1) \\
&= aE(X) + b
\end{aligned}
$$

$\square$

THEOREM 2. *If X and Y are jointly distributed random variables with expected values $E(X)$ and $E(Y)$, then $E(X + Y) = E(X) + E(Y)$.*

PROOF. Let $f(x, y)$ be the joint pdf, and $f_X(x)$ and $f_Y(y)$ the marginal pdfs.

Continuous Case Existence

$$
\begin{aligned}
E(|X + Y|) &= \int\int (|x + y|)\, f(x, y) dx dy \\
&\leq \int\int (|x| + |y|)\, f(x, y) dx dy \\
&\quad \text{by Triangle Inequality} \\
&= \int\int |x|\, f(x, y) dy dx + \int\int |y|\, f(x, y) dx dy \\
&= \int |x| \left[\int f(x, y) dy \right] dx + \int |y| \left[\int f(x, y) dx \right] dy \\
&= \int |x| f_X(x) dx + \int |y| f_Y(y) dy \\
&= E(|X|) + E(|Y|) \\
&< \infty
\end{aligned}
$$

because $E(X)$ and $E(Y)$ both exist. Therefore, $E(X + Y)$ exists.

Continuous Case Calculation

$$
\begin{aligned}
E(X+Y) &= \int\int (x+y)\ f(x,y)dxdy \\
&= \int\int (x+y)\, f(x,y)dxdy \\
&\quad \text{by Triangle Inequality} \\
&= \int\int x\, f(x,y)dydx + \int\int y\, f(x,y)dxdy \\
&= \int x \left[\int f(x,y)dy \right] dx + \int y \left[\int f(x,y)dx \right] dy \\
&= \int x\, f_X(x)dx + \int y\, f_Y(y)dy \\
&= E(X)+E(Y)
\end{aligned}
$$

Therefore, $E(X+Y) = E(X) + E(Y)$.

The proof for the discrete case is similar.

This argument is valid whether X and Y are independent or dependent.

$\square$

COROLLARY 1. *When X and Y are jointly distributed, $E(X)$ and $E(Y)$ exist, and a, b, and c are finite real numbers, by Theorems 1 and 2,*

$$
E(a+bX+cY) = a + bE(X) + cE(Y)
$$

DEFINITION 2. Random variables X and Y are independent if and only if

$$
f_{X,Y}(x,y) = f_X(x)f_Y(y)
$$

THEOREM 3. *If X and Y are independent, and $E[g(X)]$ and $E[h(Y)]$ both exist, then*

$$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$$

PROOF.

$$
\begin{aligned}
E[g(X) \cdot h(Y)] &= \int \int g(x) \cdot h(y) f(x,y) dx dy \\
&= \int \int g(x) \cdot h(y) f_X(x) \cdot f_Y(y) dx dy \\
&\quad \text{(by assumption of independence)} \\
&= \int g(x) f_X(x) dx \cdot \int h(y) f_Y(y) dy \\
&= E[g(X)] \cdot E[h(Y)]
\end{aligned}
$$

$\square$

DEFINITION 3. The variance of X is

$$Var(X) = E\left[(X - E(X))^2\right]$$

if the expected value exists.

THEOREM 4. $Var(X) = E(X^2) - (E(X))^2$

PROOF.

$$
\begin{aligned}
Var(X) &= E\left[(X - E(X))^2\right] \\
&= E\left[X^2 - 2X \cdot E(X) + (E(X))^2\right] \\
&= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 \\
&= E(X^2) - 2(E(X))^2 + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

Note that $Var(X)$ exists if both $E(X^2)$ and $E(X)$ exist.

$\square$

THEOREM 5. *If a and b are finite real numbers, and X is a random variable with finite $E(X)$ and $Var(X)$, then $Var(aX + b) = a^2 Var(X)$.*

PROOF. Continuous case only.

Existence

To check existence of $Var(aX+b)$ it is only necessary to check that $E((aX+b)^2)$ exists, because $E(aX + b)$ exists by Theorem 1.

$$
\begin{aligned}
E\left(\left|(aX+b)^2\right|\right) &= E\left((aX+b)^2\right) \\
&= E(a^2 X^2 + 2abX + b^2) \\
&= E(a^2 X^2) + E(2abX) + E(b^2) \\
&= a^2 E(X^2) + 2ab E(X) + b^2 \\
&< \infty
\end{aligned}
$$

because a and b are finite constants and $E(X^2)$ and $E(X)$ are both assumed to exist, therefore $Var(aX + b)$ exists.

Calculation

Then by Theorem 1

$$
\begin{aligned}
Var(aX+b) &= E\left((aX+b)^2\right) - [E(aX+b)]^2 \\
&= \left[a^2 E(X^2) + 2ab E(X) + b^2\right] - [aE(X) + b]^2 \\
&= a^2 E(X^2) + 2ab E(X) + b^2 - a^2 \left(E(X)\right)^2 - 2ab E(X) - b^2 \\
&= a^2 \left[E(X^2) - (E(X))^2\right] \\
&= a^2 Var(X)
\end{aligned}
$$

$\square$

THEOREM 6. *If X and Y are independent random variables with finite $E(X)$ and $Var(X)$ and $E(Y)$ and $Var(Y)$, then $Var(X + Y) = Var(X) + Var(Y)$.*

PROOF. Continuous case only.

Existence

To check existence of $Var(X+Y)$ it is only necessary to check that $E((X+Y)^2)$ exists, because $E(X + Y)$ exists by Theorem 2.

$$
\begin{aligned}
E\left(\left|(X+Y)^2\right|\right) &= E\left((X+Y)^2\right) \\
&= E(X^2 + 2XY + Y^2) \\
&= E(X^2) + E(2XY) + E(Y^2) \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) \\
&\quad (\text{because } X \text{ and } Y \text{ are independent}) \\
&< \infty
\end{aligned}
$$

because $E(X^2)$, $E(X)$, $E(Y)$, and $E(Y^2)$ all exist by assumption. Therefore, $Var(X + Y)$ exists.

Calculation

By Theorems 1 and 2

$$
\begin{aligned}
Var(aX + b) &= E\left((X+Y)^2\right) - [E(X+Y)]^2 \\
&= \left[E(X^2) + 2E(X)E(Y) + E(Y^2)\right] - [E(X) + E(Y)]^2 \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) - \left[(E(X))^2 + 2E(X)E(Y) + (E(Y))^2\right] \\
&= \left[E(X^2) - (E(X))^2\right] + \left[E(Y^2) - (E(Y))^2\right] \\
&= Var(X) + Var(Y)
\end{aligned}
$$

$\square$

COROLLARY 2. *When X and Y are independent random variables, $Var(X)$ and $Var(Y)$ exist, and a, b, and c are finite real numbers, by Theorems 5 and 6,*

$$
Var(a + bX + cY) = b^2 Var(X) + c^2 Var(Y)
$$

Moment generating functions

DEFINITION 4. A random variable X has moment generating function (mgf)

$$M(t) = E\left(e^{tX}\right)$$

if the expected value exists. Specifically,

$$M(t) = \sum_X e^{tx} p(x)$$

if X is discrete, and

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

if X is continuous.

THEOREM 7. *If the random variable X has a moment generating function that exists in an open interval around zero, then*

$$M^{(r)}(0) = E(X^r)$$

PROOF. For the continuous case, $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.

For $r = 1$

$$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

under certain conditions, integration and differentiation can be interchanged, so

$$
\begin{aligned}
M'(t) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial t} e^{tx} f(x) dx \\
&= \int_{-\infty}^{\infty} x e^{tx} f(x) dx
\end{aligned}
$$

and evaluating at zero gives

$$
\begin{aligned}
M'(0) &= \int_{-\infty}^{\infty} x e^{(0)x} f(x) dx \\
&= \int_{-\infty}^{\infty} x f(x) dx \\
&= E(X)
\end{aligned}
$$

General Case

If we take the rth derivative of $M(t)$, we get

$$
\begin{aligned}
M^{(r)}(t) &= \frac{d^r}{dt^r} \int_{-\infty}^{\infty} e^{tx} f(x) dx \\
&= \int_{-\infty}^{\infty} \frac{\partial^r}{\partial t^r} e^{tx} f(x) dx \\
&= \int_{-\infty}^{\infty} x^r e^{tx} f(x) dx
\end{aligned}
$$

and evaluating at zero gives

$$
\begin{aligned}
M^{(r)}(0) &= \int_{-\infty}^{\infty} x^r e^{(0)x} f(x) dx \\
&= \int_{-\infty}^{\infty} x^r f(x) dx \\
&= E(X^r)
\end{aligned}
$$

$\square$

REMARK 1. The rth derivative of the moment generating function evaluated at zero is the expected value of X^r. This gives

$$
M'(0) = E(X)
$$

$$
M''(0) = E(X^2)
$$

PROPOSITION 1. *The moment generating function uniquely determines the probability distribution of a random variable, if the mgf exists in an open interval around zero.*

REMARK 2. If we can find the moment generating function of a random variable, we can find the probability distribution associated with it.

THEOREM 8. *If $M_X(t)$ is the moment generating function of X and $Y = a + bX$, then the moment generating function of Y is*

$$M_Y(t) = e^{at} M_X(bt)$$

PROOF.

$$
\begin{aligned}
M_Y(t) &= E\left(e^{tY}\right) \\
&= E\left(e^{t[a+bX]}\right) \\
&= E\left(e^{at+btX}\right) \\
&= E\left(e^{at}e^{btX}\right) \\
&= e^{at}E\left(e^{btX}\right) \\
&= e^{at}M_X(bt)
\end{aligned}
$$

$\square$

THEOREM 9. *If $M_X(t)$ and $M_Y(t)$ are the moment generating functions of X and Y, where X and Y are independent, then the mgf of $Z = X + Y$ is*

$$M_Z(t) = M_X(t)M_Y(t)$$

If $M_X(t)$ exists on the interval A, and $M_Y(t)$ exists on the interval B, then $M_Z(t)$ exists on the interval $A \cap B$.

PROOF.

$$
\begin{aligned}
M_Z(t) &= E\left(e^{tZ}\right) \\
&= E\left(e^{t[X+Y]}\right) \\
&= E\left(e^{tX+tY}\right) \\
&= E\left(e^{tX}e^{tY}\right)
\end{aligned}
$$

Because X and Y are independent, the expected value can be factored into the product of expected values by Theorem 3, so

$$
\begin{aligned}
M_Z(t) &= E\left(e^{tX}e^{tY}\right) \\
&= E\left(e^{tX}\right)E\left(e^{tY}\right) \\
&= M_X(t)M_Y(t)
\end{aligned}
$$

$\square$

Normal distribution

DEFINITION 5. A normal distribution with mean μ and variance σ^2 is a continuous distribution with probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

DEFINITION 6. The standard normal distribution is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.

THEOREM 10. *If $X \sim N(\mu, \sigma^2)$, and a and b are finite constants, then*

$$Y = aX + b \sim N(a\mu + b, \, a^2\sigma^2)$$

PROOF. If $a > 0$, the cumulative density function of Y is

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(aX + b \leq y) \\
&= P\left(X \leq \frac{y-b}{a}\right) \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

If we differentiate the cumulative density function, we get the probability density function. Therefore

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\
&= \frac{1}{a} f_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Since $X \sim N(\mu, \sigma^2)$,

$$
\begin{aligned}
f_Y(y) &= \frac{1}{a} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(\left(\frac{y-b}{a}\right)-\mu\right)^2} \right] \\[2em]
&= \frac{1}{(a\sigma)\sqrt{2\pi}} e^{-\frac{1}{2(a\sigma)^2}(y-[b+a\mu])^2}
\end{aligned}
$$

This is the pdf of a normally distributed random variable with mean $a\mu + b$ and variance $a^2\sigma^2$.

If $a < 0$, the cumulative density function of Y is

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\[1em]
&= P(aX + b \le y) \\[1em]
&= P\left(X \ge \frac{y-b}{a}\right) \\[1em]
&= 1 - F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

If we differentiate the cumulative density function, we get the probability density function. Therefore

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}\left[1 - F_X\left(\frac{y-b}{a}\right)\right] \\[1em]
&= \frac{1}{a} f_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Since $X \sim N(\mu, \sigma^2)$,

$$
\begin{aligned}
f_Y(y) &= \frac{1}{a} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(\left(\frac{y-b}{a}\right)-\mu\right)^2} \right] \\[2em]
&= \frac{1}{(a\sigma)\sqrt{2\pi}} e^{-\frac{1}{2(a\sigma)^2}(y-[b+a\mu])^2}
\end{aligned}
$$

and this is the pdf of a normally distributed random variable with mean $a\mu + b$ and variance $a^2\sigma^2$.

If $a = 0$, then $Y = 0 \cdot X + b = b$ which is a Dirac distribution. It can also be thought of as being normally distributed with mean b and variance 0. So $Y \sim N(0 \cdot \mu + b, 0^2\sigma^2)$ or $Y \sim N(b, 0)$.

$\square$

COROLLARY 3. *If $X \sim N(\mu, \sigma^2)$, X can be expressed as $X = \sigma Z + \mu$ where $Z \sim N(0, 1)$.*

PROOF. If $Z \sim N(0, 1)$, then $X = \sigma Z + \mu$ would be distributed $N(0 \cdot \sigma + \mu, \sigma^2 \cdot 1)$ or $N(\mu, \sigma^2)$ by Theorem 10.

$\square$

THEOREM 11. *A standard normal distribution, see Definition 6, has mgf*

$$M(t) = e^{t^2/2}$$

PROOF. By Definitions 5 and 6, the density function of a standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

By Definition 4, the mgf is

$$
\begin{aligned}
M(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x^2 - 2tx + (-t)^2\right) + \frac{t^2}{2}} dx \\
&= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x - \frac{t}{4}\right)^2} dx \\
&= e^{t^2/2} \cdot 1
\end{aligned}
$$

because

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x - \frac{t}{4}\right)^2}$$

is the density function for a random variable distributed $N(t/4, 1)$ and density functions integrate to 1.

$\square$

COROLLARY 4. *If $X \sim N(\mu, \sigma^2)$, then the moment generating function is*

$$M(t) = e^{\mu t} e^{\sigma^2 t^2 / 2}$$

PROOF. If $X \sim N(\mu, \sigma^2)$, then $X = \sigma Z + \mu$ where $Z \sim N(0, 1)$ by Corollary 3. Then Theorem 8 can be used.

$$
\begin{aligned}
M_X(t) &= M_{\sigma Z + \mu}(t) \\
&= e^{\mu t} M_Z(\sigma t) \\
&= e^{\mu t} e^{(\sigma t)^2 / 2} \\
&= e^{\mu t} e^{\sigma^2 t^2 / 2}
\end{aligned}
$$

$\square$

THEOREM 12. *If $X \sim N(\mu, \sigma^2)$ and $Y \sim N\left(\nu, \tau^2\right)$ and X and Y are independent, then*

$$X + Y \sim N\left(\mu + \nu, \sigma^2 + \tau^2\right)$$

PROOF. By Theorem 9,

$$
\begin{aligned}
M_{X+Y}(t) &= M_X(t) M_Y(t) \\
&= \left[e^{\mu t} e^{\sigma^2 t^2 / 2}\right] \left[e^{\nu t} e^{\tau^2 t^2 / 2}\right] \\
&= e^{\mu t + \nu t} e^{\sigma^2 t^2 / 2 + \tau^2 t^2 / 2} \\
&= e^{(\mu + \nu) t} e^{(\sigma^2 + \tau^2) t^2 / 2}
\end{aligned}
$$

We see that $M_{X+Y}(t)$ is the moment generating function of a normal distribution with mean $\mu + \nu$ and variance $\sigma^2 + \tau^2$. Therefore by the uniqueness of moment generating functions, Proposition 1,

$$X + Y \sim N\left(\mu + \nu, \sigma^2 + \tau^2\right)$$

$\square$

THEOREM 13. *If X is a normally distributed random variable with mean μ and standard deviation σ, then $Z = \dfrac{X - \mu}{\sigma}$ has a standard normal distribution.*

PROOF. By Corollary 4,

$$M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}$$

Then

$$
\begin{aligned}
M_Y(t) &= M_{\frac{X - \mu}{\sigma}}(t) \\
&= M_{(1/\sigma)X - (\mu/\sigma)}(t) \\
&= e^{(-\mu/\sigma)t} M_X\left(\frac{1}{\sigma}t\right)
\end{aligned}
$$

by Theorem 8. Then

$$
\begin{aligned}
M_Y(t) &= e^{(-\mu/\sigma)t}\left[e^{\mu\left(\frac{1}{\sigma}t\right)} e^{\sigma^2\left(\frac{1}{\sigma}t\right)^2/2} \right] \\
&= e^{(-\mu/\sigma)t}\left[e^{(\mu/\sigma)t} e^{\sigma^2 \cdot \frac{1}{\sigma^2} \cdot t^2/2} \right] \\
&= e^{(-\mu/\sigma + \mu/\sigma)t} e^{1 \cdot t^2/2} \\
&= e^{0t} e^{t^2/2} \\
&= e^{t^2/2}
\end{aligned}
$$

which is the mgf of a standard normal distribution. Therefore by uniqueness of moment generating functions, Proposition 1, $Z \sim N(0, 1)$.

$\square$

Distributions derived from the normal distribution

DEFINITION 7. If $Z \sim N(0,1)$, then Z^2 has a chi-square distribution with 1 degree of freedom, denoted χ_1^2.

DEFINITION 8. If $V_1, V_2, \ldots, V_n$ are independent random variables each with a chi-square distribution with 1 degree of freedom, then $\sum_{i=1}^{n} V_i$ has a chi-square distribution with n degrees of freedom, denoted χ_n^2.

DEFINITION 9. If Z and U are independent random variables with $Z \sim N(0,1)$ and $U \sim \chi_n^2$, then

$$\frac{Z}{\sqrt{\frac{U}{n}}}$$

has a t distribution with n degrees of freedom, denoted t_n.

DEFINITION 10. If U and V are independent random variables, $U \sim \chi_n^2$ and $V = \chi_m^2$, then

$$\frac{U/n}{V/m}$$

has a F distribution with n and m degrees of freedom, denoted $F_{n,m}$.

FACT 1. *The pdf of a chi-square random variable with k degrees of freedom is*

$$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} \cdot I_{[0,\infty)}(x)$$

FACT 2. *A random variable with a gamma distribution with parameters α and λ has pdf*

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

for $x \geq 0$, $\alpha > 0$ and $\lambda > 0$.

LEMMA 1. *The moment generating function of a chi-square random variable with k degrees of freedom is $(1 - 2t)^{-k/2}$ for $t < 1/2$.*

PROOF. The pdf of a chi-square random variable with k degrees of freedom is

$$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} \cdot I_{[0,\infty)}(x)$$

So by Definition 4, the moment generating function is

$$
\begin{aligned}
M(t) &= \int_0^\infty e^{tx} \cdot \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} dx \\
&= \int_0^\infty \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{tx-x/2} dx \\
&= \frac{2^{k/2}(1-2t)^{-k/2}}{2^{k/2}} \int_0^\infty \frac{\left(\dfrac{1-2t}{2}\right)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x\frac{(1-2t)}{2}} dx \\
&= (1-2t)^{-k/2}
\end{aligned}
$$

because

$$\frac{\left(\dfrac{1-2t}{2}\right)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x\frac{(1-2t)}{2}}$$

is the pdf of a gamma distribution with $\alpha = k/2$ and $\lambda = \dfrac{(1-2t)}{2}$ and probability density functions integrate to one. By the definition of the pdf of a gamma distribution, $\lambda > 0$ or $\dfrac{(1-2t)}{2} > 0$. Therefore the moment generating function for a chi-square random variable is only defined if $t < 1/2$.

$\square$

LEMMA 2. *If $X \sim \chi_k^2$, then*

$$E(X) = k$$

$$Var(X) = 2k$$

PROOF. By Theorem 7,

$$M^{(r)}(0) = E(X^r)$$

so

$$
\begin{aligned}
E(X) &= \frac{d}{dt}(1-2t)^{-k/2}|_{t=0} \\
&= -\frac{k}{2}(1-2t)^{-k/2-1} \cdot -2|_{t=0} \\
&= -\frac{k}{2}(1-2\cdot 0)^{-k/2-1} \cdot -2 \\
&= k
\end{aligned}
$$

and

$$
\begin{aligned}
E(X^2) &= \frac{d^2}{dt^2}(1-2t)^{-k/2}|_{t=0} \\
&= k\left(-\frac{k}{2}-1\right)(1-2t)^{-k/2-2} \cdot -2|_{t=0} \\
&= k(k+2)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
&= k(k+2) - k^2 \\
&= k^2 + 2k - k^2 \\
&= 2k
\end{aligned}
$$

$\square$

One sample z-test for the population mean

DEFINITION 11. If $X_1, X_2, \ldots, X_n$ are random variables, let

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

and

$$S = \frac{\sum_{i=1}^{n}(X_i - \overline{X})}{n-1}$$

REMARK 3. The mean of a sample from a population is not a constant. It varies depending on which sample from the population is chosen. Therefore, the sample mean is a random variable with its own distribution, called the sampling distribution.

LEMMA 3. *If $\overline{X}$ is the random variable of the sample means of all the simple random samples of size n from a population with expected value $E(X)$ and variance $Var(X)$, then the expected value and variance of $\overline{X}$ are*

$$E(\overline{X}) = E(X)$$
$$Var(\overline{X}) = \frac{Var(X)}{n}$$

PROOF. Consider all of the possible samples of size n from a population with expected value $E(X)$ and variance $Var(X)$. If a sample $X_1, X_2, \ldots, X_n$ is chosen, each X_i comes from the same population so each X_i has the same expected value, $E(X)$ and variance, $Var(X)$. By Corollary 1,

$$\begin{aligned}
E(\overline{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i)\right) \\
&= \frac{1}{n}\left(n \cdot E(X)\right) \\
&= E(X)
\end{aligned}$$

and by Corollary 2

$$
\begin{aligned}
Var(\overline{X}) \;&=\; Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \\
&=\; \frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i\right) \\
&=\; \frac{1}{n^2}\left(\sum_{i=1}^{n}Var(X_i)\right) \\
&=\; \frac{1}{n^2}\left(n\cdot Var(X)\right) \\
&=\; \frac{Var(X)}{n}
\end{aligned}
$$

$\square$

REMARK 4. As the sample size n increases, the variance of $\overline{X}$ decreases.

THEOREM 14. *If $X_1, X_2, \ldots, X_n$ are normally distributed random variables with mean μ and variance σ^2, then*

$$
\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)
$$

PROOF. By Lemma 3 and Theorems 10 and 12,

$$
\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)
$$

So by Theorem 13,

$$
\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)
$$

$\square$

COROLLARY 5. *If $X_1, X_2, \ldots, X_n$ are normally distributed random variables, then the statistic $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ has a standard normal distribution where μ_0 is the hypothesized population mean under the null distribution.*

PROOF. This corollary is a direct result of Theorem 14.

$\square$

REMARK 5. Therefore, the standard normal distribution can be used to find p-values for a one sample z-test for a population mean.

Paired sample z test for the population mean of paired differences

THEOREM 15.   *If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are normally distributed random variables and X_i and Y_i are independent, then*

$$Z = \frac{\overline{D} - \delta}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0, 1)$$

where $D_i = X_i - Y_i$ is a pairwise difference, $\overline{D} = \dfrac{\sum_{i=1}^{n} D_i}{n}$, δ is the mean of the pairwise differences under the null hypothesis, and σ_D is the population standard deviation of the pairwise differences.

PROOF.   Note that X_i and Y_i are normal random variables, so $D_i = X_i - Y_i$ is also a normal random variable by Theorem 12. Then $D_1, D_2, \ldots, D_n$ can be thought of as the sequence of random variables $\{X_i\}_{i=1}^{n}$ described in Corollary 5. Therefore, $Z \sim N(0, 1)$.

$\square$

REMARK 6.   Therefore, the standard normal distribution can be used to find p-values for the paired sample z-test for the population mean of pairwise differences.

Two sample z-test for the difference of population means

LEMMA 4. *If* $X_1, X_2, \ldots, X_n$ *and* $Y_1, Y_2, \ldots, Y_m$ *are independent with* $X_i \sim N(\mu_X, \sigma_X^2)$ *and* $Y_i \sim N(\mu_Y, \sigma_Y^2)$ *then*

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

PROOF. By Lemma 3, $\overline{X} \sim N\left(\mu_X, \dfrac{\sigma_X^2}{n}\right)$ and $\overline{Y} \sim N\left(\mu_Y, \dfrac{\sigma_Y^2}{m}\right)$. Then by Theorems 10 and 12,

$$
\begin{aligned}
\overline{X} - \overline{Y} &= \overline{X} + (-1)\overline{Y} \\
&\sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + (-1)^2\frac{\sigma_Y^2}{m}\right)
\end{aligned}
$$

$\square$

THEOREM 16. *If* $X_1, X_2, \ldots, X_n$ *and* $Y_1, Y_2, \ldots, Y_m$ *are independent with* $X_i \sim N(\mu_X, \sigma_X^2)$ *and* $Y_i \sim N(\mu_Y, \sigma_Y^2)$ *then*

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \sim N(0,1)$$

PROOF. By Lemma 4, $\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}\right)$, so the mean of $\overline{X} - \overline{Y}$ is $\mu_X - \mu_Y$ and the variance is $\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}$. Then $\overline{X} - \overline{Y}$ can be standardized using Theorem 13, giving

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \sim N(0,1)$$

$\square$

REMARK 7. Therefore, the p-value for the two sample z-test for the difference of population means (independent samples) can be found by with the standard normal distribution.

One sample z-test for the population proportion

PROPOSITION 2. ***Lindeberg-Levy Central Limit Theorem***

If $\{X_n\}_{i=1}^{\infty}$ is a sequence of random variables that are independent and identically distributed, each with mean $E(X_i) = \mu$ and positive finite variance $Var(X_i) = \sigma^2$, with $\overline{X}_n = \dfrac{1}{n} \sum_{i=1}^{n} X_i$, then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z$$

*where $Z \sim N(0,1)$, see [**9**, **30**].*

THEOREM 17. *Therefore, if $X \sim Bin(n,p)$, and $\hat{p} = \dfrac{X}{n}$, then for large n,*

$$\hat{p} \text{ is approximately distributed } N\left(p, \frac{p(1-p)}{n}\right)$$

PROOF. If $X \sim Bin(n,p)$, then $X = \sum_{i=1}^{n} X_i$ where each X_i is a Bernoulli random variable with probability of success p. Therefore $\hat{p} = \dfrac{X}{n} = \dfrac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$. $E(X_i) = p$ and $Var(X_i) = p(1-p) < \infty$ $\forall i$. Then it holds for $\hat{p}_n = \overline{X}_n$ that

$$\frac{\hat{p}_n - p}{\sqrt{\dfrac{p(1-p)}{n}}} \xrightarrow{d} Z$$

where $Z \sim N(0,1)$ by the Lindeberg-Levy Central Limit Theorem.

$\square$

COROLLARY 6. *The test statistic*

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

is approximately normally distributed under the null hypothesis of $p = p_0$ for large n.

PROOF. Since $\hat{p}$ is approximately normally distributed with mean p_0 and variance $\dfrac{p_0(1-p_0)}{n}$, by Theorem 13, Z is approximately normally distributed.

$\square$

REMARK 8. Therefore, the standard normal distribution can be used to find p-values for the one sample z-test for a population proportion.

Two sample z-test for population proportions

THEOREM 18. *If $X \sim Bin(p_X, n)$ and $Y \sim Bin(p_Y, m)$, then $\hat{p}_X = \dfrac{X}{n}$ is approximately distributed*

$$N\left(p_X, \frac{p_X(1 - p_X)}{n}\right)$$

and $\hat{p}_Y = \dfrac{Y}{m}$ is approximately distributed

$$N\left(p_Y, \frac{p_Y(1 - p_Y)}{m}\right)$$

by Theorem 17. Then $\hat{p}_X - \hat{p}_Y$ is approximately distributed

$$N\left(p_X - p_Y, \frac{p_X(1 - p_X)}{n} + \frac{p_Y(1 - p_Y)}{m}\right)$$

by Theorems 10 and 12. Therefore

$$\frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\dfrac{p_X(1 - p_X)}{n} + \dfrac{p_Y(1 - p_Y)}{m}}}$$

has an approximate standard normal distribution.

PROOF. See theorem 13 with $E\left(\hat{p}_X - \hat{p}_Y\right) = p_X - p_Y$ and

$$Var\left(\hat{p}_X - \hat{p}_Y\right) = \frac{p_X(1 - p_X)}{n} + \frac{p_Y(1 - p_Y)}{n}$$

$\square$

COROLLARY 7. *Under the null hypothesis that $p_X = p_Y = p$,*

$$\frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{p(1-p)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

has an approximate standard normal distribution.

PROOF.

$$\frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\dfrac{p_X(1-p_X)}{n} + \dfrac{p_Y(1-p_Y)}{m}}} = \frac{(\hat{p}_X - \hat{p}_Y) - (p - p)}{\sqrt{\dfrac{p(1-p)}{n} + \dfrac{p(1-p)}{m}}}$$

$$= \frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{p(1-p)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

$\square$

REMARK 9. When conducting a two sample z-test for population proportions to see if $p_X = p_Y = p$, it is assumed that p is unknown. So $\hat{p} = \dfrac{X + Y}{n + m}$, which is the overall proportion of success for both populations, is used as an estimate for p.

REMARK 10. As a result of Corollary 7, the test statistic

$$Z = \frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

has an approximate normal distribution.

REMARK 11. Therefore, the standard normal distribution can be used to find p-values for the two sample z-test for population proportions.

One sample t-test for the population mean

LEMMA 5. $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$

PROOF.

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X} \\
&= \sum_{i=1}^{n} X_i - n\overline{X} \\
&= \sum_{i=1}^{n} X_i - n\frac{\sum_{i=1}^{n} X_i}{n} \\
&= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i \\
&= 0
\end{aligned}
$$

$\square$

THEOREM 19. $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $n-1$ degrees of freedom.

PROOF.

$$
S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}
$$

so

$$
\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)}{\sigma^2}\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2
$$

Now

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \\
&= \sum_{i=1}^{n}(Z_i)^2 \\
&= \sum_{i=1}^{n} V_i
\end{aligned}
$$

where $Z_i \sim N(0,1)$ by Theorem 13 and $V_i \sim \chi_1^2$ by Definition 7. Then by Definition 8,

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi_n^2$$

We also have

$$\begin{aligned}
\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left([X_i - \overline{X}] + [\overline{X} - \mu] \right)^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[(X_i - \overline{X})^2 + 2(X_i - \overline{X})(\overline{X} - \mu) + (\overline{X} - \mu)^2 \right] \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 + \frac{2}{\sigma^2}(\overline{X} - \mu) \sum_{i=1}^{n}(X_i - \overline{X}) + \frac{1}{\sigma^2} \sum_{i=1}^{n}(\overline{X} - \mu)^2
\end{aligned}$$

Note that $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ by Lemma 5 and $(\overline{X} - \mu)^2$ is a constant. Therefore

$$\begin{aligned}
\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 &= \frac{1}{\sigma^2} \sum_{i=1}^{n}(X_i - \overline{X})^2 + 0 + \frac{n}{\sigma^2}(\overline{X} - \mu)^2 \\
&= \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2
\end{aligned}$$

Let $W = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$, $U = \frac{(n-1)S^2}{\sigma^2}$, and $V = \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2$. Then $W = U + V$.

We know $W \sim \chi_n^2$ and

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

by Theorem 14, so $V \sim \chi_1^2$. The moment generating function of a random variable with n degrees of freedom is $M(t) = (1 - 2t)^{-n/2}$ by Lemma 10. By Theorem 9, $M_W(t) = M_U(t)M_V(t)$. Then

$$\begin{aligned}
M_U(t) &= \frac{M_W(t)}{M_V(t)} \\
&= \frac{(1 - 2t)^{-n/2}}{(1 - 2t)^{-1/2}} \\
&= (1 - 2t)^{-(n-1)/2}
\end{aligned}$$

which is the mgf of random variable with a chi-square distribution with $n - 1$ degrees of freedom. By uniqueness of moment generating functions, Proposition 1, $U \sim \chi_{n-1}^2$.

$\square$

THEOREM 20. *If $\{X_i\}_{i=1}^n$ is a sequence of independent random variables with $X_i \sim N(\mu, \sigma^2)$, then*

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

PROOF.

$$\begin{aligned}
\frac{\overline{X} - \mu}{S/\sqrt{n}} &= \frac{1}{S} \cdot \frac{\overline{X} - \mu}{1/\sqrt{n}} \cdot \sqrt{\frac{\sigma^2}{\sigma^2}} \\
&= \sqrt{\frac{\sigma^2}{S^2}} \cdot \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \\
&= \sqrt{\frac{(n-1)\sigma^2}{(n-1)S^2}} \cdot \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \\
&= \frac{\dfrac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{\dfrac{\left(\dfrac{(n-1)S^2}{\sigma^2}\right)}{(n-1)}}} \\
&= \frac{Z}{\sqrt{\dfrac{U}{n-1}}}
\end{aligned}$$

where $Z = \dfrac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ by Theorem 14 and $U = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ by Theorem 19. Then by Definition 9, $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$.

$\square$

COROLLARY 8. *The test statistic $t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$ has a t distribution with $n-1$ degrees of freedom under the null hypothesis.*

PROOF. This corollary is a direct result of Theorem 20.

$\square$

REMARK 12. Therefore, the t-distribution with $n-1$ degrees of freedom can be used to find p-values for the one sample t-test for the population mean.

Paired sample t-test for the population mean of paired samples

THEOREM 21. *If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are normally distributed random variables, X_i and Y_i are independent, then*

$$Z = \frac{\overline{D} - \delta}{\dfrac{S_D}{\sqrt{n}}} \sim N(0, 1)$$

where $D_i = X_i - Y_i$ is a pairwise difference, $\overline{D} = \dfrac{\sum_{i=1}^{n} D_i}{n}$, δ is the mean of the pairwise differences under the null hypothesis, and S_D is the sample standard deviation of the pairwise differences.

PROOF. Note that X_i and Y_i are normal random variables, so $D_i = X_i - Y_i$ is also a normal random variable by Theorem 12. Then $D_1, D_2, \ldots, D_n$ can be thought of as the sequence of random variables $\{X_i\}_{i=1}^{n}$ described in Theorem 20. Therefore $t \sim t_{n-1}$.

$\square$

REMARK 13. Therefore, the t distribution with $n - 1$ degrees of freedom can be used to find p-values for the paired sample t-test for the population mean of pairwise differences.

Two sample t-test for population means (equal variances)

LEMMA 6. *If* $X_1, X_2, \ldots, X_n$ *and* $Y_1, Y_2, \ldots, Y_m$ *are independent with* $X_i \sim N(\mu_X, \sigma^2)$ *and* $Y_i \sim N(\mu_Y, \sigma^2)$ *then*

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

PROOF. By Lemma 3, $\overline{X} \sim N\left(\mu_X, \frac{\sigma^2}{n}\right)$ and $\overline{Y} \sim N\left(\mu_Y, \frac{\sigma^2}{m}\right)$. Then by Theorems 10 and 12,

$$\begin{aligned} \overline{X} - \overline{Y} \;\; &= \;\; \overline{X} + (-1)\overline{Y} \\ &\sim \;\; N\left(\mu_X - \mu_Y, \frac{\sigma^2}{n} + (-1)^2\frac{\sigma^2}{m}\right) \end{aligned}$$

$\square$

LEMMA 7. *If* $X_1, X_2, \ldots, X_n$ *and* $Y_1, Y_2, \ldots, Y_m$ *are independent with* $X_i \sim N(\mu_X, \sigma^2)$ *and* $Y_i \sim N(\mu_Y, \sigma^2)$, *then*

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim N(0, 1)$$

PROOF. By Lemma 6, $\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$, so the mean $\overline{X} - \overline{Y}$ is $\mu_X - \mu_Y$ and the variance is $\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$.

$\overline{X} - \overline{Y}$ can be standardized by Theorem 13, giving

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim N(0, 1)$$

.

$\square$

THEOREM 22. *If* $X_1, X_2, \ldots, X_n$ *and* $Y_1, Y_2, \ldots, Y_m$ *are independent with* $X_i \sim N(\mu_X, \sigma_X^2)$ *and* $Y_i \sim N(\mu_Y, \sigma_Y^2)$, *then the statistic*

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$$

where

$$S_p = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}$$

has a t distribution with $n + m - 2$ degrees of freedom.

PROOF.

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}} \cdot \frac{\sigma}{\sigma}$$

$$= \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \div \sqrt{\left[\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2}\right] \cdot \frac{1}{n+m-2}}$$

Let $Z = \dfrac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}$. By Lemma 7, $Z \sim N(0, 1)$.

Let $U = \left[\dfrac{(n-1)S_X^2}{\sigma^2} + \dfrac{(m-1)S_Y^2}{\sigma^2}\right]$. Then $U \sim \chi_{n+m-2}^2$ by Definition 8 because

$\dfrac{(n-1)S_X^2}{\sigma^2} \sim \chi_{n-1}^2$ and $\dfrac{(m-1)S_Y^2}{\sigma^2} \sim \chi_{m-1}^2$ by Theorem 19. Therefore,

$$T = \frac{Z}{\sqrt{\dfrac{U}{n+m-2}}}$$

which has a t distribution with $n + m - 2$ degrees of freedom by Definition 9.

$\square$

REMARK 14. Therefore, the t distribution with $n + m - 2$ degrees of freedom can be used to find p-values for the two sample t-test for the difference of population means under the assumption of equal population variances.

Two sample t-test for population means (unequal variances)

LEMMA 8. *If $X_1, X_2, \ldots, X_n$ is a sequence of random variables, $Var(S^2) = \dfrac{2\sigma^4}{n-1}$.*

PROOF. We know from Theorem 19, that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

and we know that the variance of a chi-square random variable with k degrees of freedom is $2k$, by Lemma 2. So

$$Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

By Corollary 2,

$$Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{(n-1)^2 Var(S^2)}{\sigma^4}$$

Therefore

$$\frac{(n-1)^2 Var(S^2)}{\sigma^4} = 2(n-1)$$

or

$$Var(S^2) = \frac{2\sigma^4}{(n-1)}$$

$\square$

THEOREM 23.

$$\hat{\beta}\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)^{-1}\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right) \overset{approx}{\sim} \chi^2_{\hat{\beta}}$$

where

$$\hat{\beta} = \frac{\left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)^2}{\dfrac{S_X^4}{n^2(n-1)} + \dfrac{S_Y^4}{m^2(m-1)}}$$

PROOF. Satterthwaite suggested the approximation

$$\beta\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)^{-1}\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right) \sim \chi^2_{\beta}$$

see [21] or [22, 111].

$\square$

If $V = \beta \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right)$ and $V \sim \chi_\beta^2$, then we would expect

$$Var(V) = Var\left[\beta \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right) \right]$$

and we know that the variance of a chi-square random variable with k degrees of freedom is $2k$, see Lemma 2. So $Var(V) = 2\beta$. By Corollary 2, where β, n, m, σ_X^2, and σ_Y^2 are constants,

$$Var\left[\beta \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right) \right] = \frac{\beta^2 \left(\frac{Var(S_X^2)}{n^2} + \frac{Var(S_Y^2)}{m^2} \right)}{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^2}$$

but by Lemma 8, $Var(S_X^2) = \frac{2\sigma_X^4}{n-1}$ and $Var(S_Y^2) = \frac{2\sigma_Y^4}{m-1}$, so

$$Var\left[\beta \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right) \right] = \frac{\beta^2 \left(\frac{2\sigma_X^4}{(n-1)\,n^2} + \frac{2\sigma_Y^4}{(m-1)m^2} \right)}{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^2}$$

but $Var(V) = 2\beta$, implying

$$\frac{\beta^2 \left(\frac{2\sigma_X^4}{(n-1)\,n^2} + \frac{2\sigma_Y^4}{(m-1)m^2} \right)}{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^2} = 2\beta$$

Solving this equation gives

$$\beta = \frac{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^2}{\frac{\sigma_X^4}{(n-1)\,n^2} + \frac{\sigma_Y^4}{(m-1)m^2}}$$

or $\beta = 0$, which can be ignored because the degrees of freedom must be greater than zero.

Of course, σ_X^2 and σ_Y^2 are unknown for a t hypothesis test. So an estimate is

$$\hat{\beta} = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}$$

and

$$\hat{\beta} \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right) \overset{\text{approx}}{\sim} \chi_{\hat{\beta}}^2$$

THEOREM 24. *If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are independent with $X_i \sim N(\mu_X, \sigma_X^2)$ and $Y_i \sim N(\mu_Y, \sigma_Y^2)$, then the statistic*

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$$

has an approximate t distribution with v degrees of freedom where

$$v = \frac{\left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)^2}{\dfrac{S_X^4}{n^2(n-1)} + \dfrac{S_Y^4}{m^2(m-1)}}$$

PROOF.

$$
\begin{aligned}
T &= \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}} \cdot \frac{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \\[2em]
&= \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \div \sqrt{\dfrac{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}
\end{aligned}
$$

Let $Z = \dfrac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}}$, by Theorem 4, $Z \sim N(0, 1)$.

$\square$

Now

$$\sqrt{\frac{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} = \sqrt{\frac{\hat{\beta} \left(\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}\right)^{-1} \left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)}{\hat{\beta}}}$$

but by Theorem 23,

$$V = \hat{\beta} \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)^{-1} \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right) \overset{\text{approx}}{\sim} \chi_{\hat{\beta}}^2$$

Therefore,

$$T = \frac{Z}{\sqrt{\dfrac{V}{\hat{\beta}}}}$$

where $Z \sim N(0,1)$, and V is approximately distributed $\chi^2_{\hat{\beta}}$. Then by Definition 9, T is approximately distributed with $\hat{\beta}$ degrees of freedom, where

$$\hat{\beta} = \frac{\left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)^2}{\dfrac{S_X^4}{n^2(n-1)} + \dfrac{S_Y^4}{m^2(m-1)}}$$

REMARK 15. Therefore, the t distribution with v degrees of freedom as stated above can be used to find p-values for the two sample t-test for the difference of population means under the assumption of unequal population variances.

Chi-square test for population variance

COROLLARY 9. *Under the null hypothesis that $\sigma^2 = \sigma_0^2$, the test statistic*

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

has a chi-square distribution with $n-1$ degrees of freedom.

PROOF. This corollary is a direct consequence of Theorem 19.

$\square$

REMARK 16. Therefore, the chi-square distribution with $n-1$ degrees of freedom can be used to find p-values for the chi-square test for population variance.

Two sample F test for population variances

THEOREM 25. *Under the null hypothesis that $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, the test statistic*

$$F = \frac{S_X^2}{S_Y^2}$$

has a F distribution with $n-1$ and $m-1$ degrees of freedom.

PROOF.

$$
\begin{aligned}
F &= \frac{S_X^2}{S_Y^2} \\[2mm]
&= \frac{S_X^2}{S_Y^2} \cdot \frac{\dfrac{1}{\sigma^2}}{\dfrac{1}{\sigma^2}} \cdot \frac{\dfrac{(n-1)}{(n-1)}}{\dfrac{(m-1)}{(m-1)}} \\[2mm]
&= \frac{\left(\dfrac{(n-1)S_X^2}{\sigma^2}\right) \div (n-1)}{\left(\dfrac{(m-1)S_Y^2}{\sigma^2}\right) \div (m-1)} \\[2mm]
&= \frac{U/(n-1)}{V/(m-1)}
\end{aligned}
$$

where $U = \dfrac{(n-1)S_X^2}{\sigma^2} \sim \chi_{n-1}^2$ and $V = \dfrac{(m-1)S_Y^2}{\sigma^2} \sim \chi_{m-1}^2$ by Theorem 19. Then by Definition 10, $F \sim F_{n-1,m-1}$.

$\square$

REMARK 17. Therefore, the F distribution with $n-1$ and $m-1$ degrees of freedom can be used to find the p-values for the two sample F test for population variances.

Chi-square test for goodness of fit

PROPOSITION 3. *The test statistic*

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

where X_i is the observed frequency or count in each group, p_i is the theoretical probability for each group, and n is the sample size, has an approximate chi-square distribution with

- $r - m$ degrees of freedom where r is the number of independent counts and m is the number of parameters estimated, see [**15**, 260], or
- if the experiment is set up in a way that $r = k - 1$, there are $k - m - 1$ degrees of freedom where k is the number of groups or cells and m is the number of parameters estimated, see [**4**, 579].

REMARK 18. Therefore, the p-values for the chi-square test for goodness of fit can be found by using the chi-square distribution with $k - m - 1$.

Chi-square test for independence

COROLLARY 10. *For the chi-square test for independence,*

the test statistic

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

has an approximate $\chi^2_{(I-1)(J-1)}$ distribution where I is the number of rows and J is the number of columns.

PROOF. Under the null hypothesis that the two variables are independent, the probability of a cell is the product of the probability of the column and the probability of the row. So $p_l = \frac{c_j}{n} \cdot \frac{r_i}{n} = \frac{c_j r_i}{n^2}$ where c_j is the column total and r_j is the row total and n is the grand total or sample size. Then the expected value of a cell is $np_l = \frac{c_j r_i}{n} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$. Therefore

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{l=1}^{k} \frac{(X_i - np_l)^2}{np_l}$$

Then by Proposition 3, X^2 has an approximate χ^2_{k-m-1} distribution.

The number of cells, k can be found by the product of the number of columns and number of rows, $k = IJ$. The expected probabilities are found by estimating the row probabilities and the column probabilities. Since the row probabilities must sum to 1 and the column probabilities must sum to one, there are $(I-1)$ and $(J-1)$ independent parameters that are being estimated so $m = (I-1) + (J-1)$. Therefore the degrees of freedom are

$$\begin{aligned}
k - m - 1 &= IJ - [(I-1) + (J-1)] - 1 \\
&= IJ - I + 1 - J + 1 - 1 \\
&= IJ - I - J + 1 \\
&= (I-1)(J-1)
\end{aligned}$$

$\square$

REMARK 19. Therefore, the chi-square distribution with $(I-1)(J-1)$ degrees of freedom can be used to find p-values for the chi-square test for independence.

Chi-square test for homogeneity

COROLLARY 11. *For the chi-square test for homogeneity,*

the test statistic

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

has an approximate $\chi^2_{(I-1)(J-1)}$ distribution where I is the number of rows and J is the number of columns.

PROOF. Under the null hypothesis that the J distributions are the same, the probability of the ith category is the same for each jth distribution. Call that probability p_i. If the probability of each ith category is the same, then an estimate for $p_i = \dfrac{r_i}{n}$ or the category total divided by the grand total. Then the expected value in the ith category in the jth distribution could be found by the product of the probability of the ith category and the sample size of the jth distribution, $E_{ij} = \dfrac{r_i c_j}{n}$. Therefore

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{l=1}^{k} \frac{(O_l - E_l)^2}{E_l}$$

has an approximate χ^2_{r-m} distribution, by Proposition 3. Each distribution j has $I-1$ independent counts because the category total is fixed. So there are $r = J(I-1)$ independent counts. Since $\sum_{i=1}^{I} p_i = 1$, it is only necessary to estimate $m = I - 1$ parameters. Therefore, the degrees of freedom are

$$\begin{aligned} r - m &= J(I-1) - (I-1) \\ &= IJ - J - I + 1 \\ &= (I-1)(J-1) \end{aligned}$$

$\square$

REMARK 20. Therefore, the chi-square distribution with $(I-1)(J-1)$ degrees of freedom can be used to find p-values for the chi-square test for homogeneity.

Bibliography

[1] Anderson, David Ray, Dennis J. Sweeney, and Thomas Arthur Williams. *Statistics for Business and Economics.* Mason, OH: Thomson/South-Western, 2005.

[2] Bowerman, Bruce L. *Essentials of Business Statistics.* Boston: McGraw-Hill/Irwin, 2008.

[3] Casella, George, and Roger L. Berger. *Statistical Inference.* Pacific Grove, CA: Duxbury Thomson Learning, 2002.

[4] Chernoff, Herman, and E. L. Lehmann. "The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit." *Annals of Mathematical Statistics* 25.3 (1954): 579-86.

[5] Christensen, Ronald. "Testing Fisher, Neyman, Pearson, and Bayes." American Statistician 59.2 (2005): 121-26.

[6] DiSantostefano, Rachael L., and Keith E. Muller. "A Comparison of Power Approximations for Satterthwaite's Test." *Communications in Statistics* 24.3 (1995): 583-93.

[7] Everitt, Brian. *The Cambridge Dictionary of Statistics.* Cambridge, UK: Cambridge UP, 1998.

[8] Fisher, Ronald Aylmer. *Statistical Methods, Experimental Design and Scientific Inference: a Re-issue of Statistical Methods for Research Workers, the Design of Experiments, and Statistical Methods and Scientific Inference.* Ed. J. Henry. Bennett. Oxford: Oxford UP, 1990.

[9] Fisher, R. A. "The Statistical Method in Psychical Research." *Proceedings of the Society for Psychical Research* 39 (1929): 185-92.

[10] Hayter, Anthony J. *Probability and Statistics for Engineers and Scientists.* Pacific Grove, CA: Duxbury, 2002.

[11] Johnson, Robert Russell, and Patricia Kuby. *Elementary Statistics.* Belmont, CA: Thomson/Brooks/Cole, 2004.

[12] Maity, Arnab, and Michael Sherman. "The Two-Sample T Test with One Variance Unknown." *American Statistician* 60.2 (2006): 163-66.

[13] Moore, David S., George P. McCabe, and Bruce A. Craig. *Introduction to the Practice of Statistics.* New York: W.H. Freeman, 2009.

[14] Neyman, J., and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." Philosophical Transactions of the Royal Society of London A 231 (1933): 289-337.

[15] Neyman, J. "Contribution to the theory of the χ^2-test." *Proceedings of the Berkely Symposium on Mathematical Statistics and Probability.* CA: University of California Press, 1949.

[16] Pearson, Karl. "Historical Note on the Distribution of the Standard Deviations of Samples of any Size Drawn from an Indefinitely Large Normal Parent Population." *Biometrika* 23.3/4 (1931): 416-418.

[17] Pearson, Karl, R. A. Fisher, and Henry F. Inman. "Karl Pearson and R.A. Fisher on Statistical Tests: A 1935 Exchange from Nature." *American Statistician* 48.1 (1994): 2-11.

[18] Plackett, R. L. "Karl Pearson and the Chi-Squared Test." *International Statistical Review* 51.1 (1983): 59-72.

[19] Rice, John A. *Mathematical Statistics and Data Analysis.* Belmont, CA: Thomson/Brooks/Cole, 2007.

[20] Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century.* New York: Henry Holt, 2002.

[21] Satterthwaite, Franklin E. "Synthesis of Variance." *Psychometrika* 6.5 (1941): 309-16.

[22] Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2.6 (1946): 110-14.

[23] Symanzik, Jürgen. *Mathematical Statistics Notes.*

[24] Tabak, John. *Probability and Statistics: The Science of Uncertainty.* New York, NY: Checkmark Books, 2005.

[25] Weiss, Neil A. *Introductory Statistics.* San Francisco: Pearson Addison-Wesley, 2008.

[26] Welch, B. L. "The Significance of the Difference Between Two Means When the Population Variances Are Unequal." *Biometrika* 29.3/4 (1938): 350-62.

APPENDIX A

Reference Material

Basic Terms

Population: The set of all people, items, events, objects, etc. that are of interest.
Examples: everyone in the US, the results of 10000 coin tosses, every car built in 2010, etc.

Sample: A subset (smaller set) of the population. This is the part of the population for which data is gathered.
Examples: every thousandth person in the US, the results of 100 coin tosses, etc.

Population vs Sample: The population refers to all the individuals or events of interest. However, it is usually too difficult, time consuming, or expensive to collect data on an entire population such as everyone in the United States or every car built in 2010. So a smaller subset of the population called the sample is chosen on which to gather data. It is assumed that the sample is representative of the population. One way to do insure this is by choosing a simple random sample.

Sampling with Replacement: After choosing an element from a population, the element is placed back into the population and can be drawn again.

Sampling without Replacement: After choosing an element from a population, it is withheld from the population and cannot be drawn again.

Simple Random Sample: A simple random sample means that each element in the population has the same chance of being chosen.
Note that random does not mean chaotic or without order, but that each element has the same chance of being drawn.
The elements in the sample need to be chosen independently. That means that drawing an element has no effect on which element is chosen next.

Element: One of the individuals, objects, etc. in the population or sample.

Variable: A characteristic of interest of the population or sample. This is what is measured.
For example, the height of everyone in the US, the number of heads in 100 coin tosses, etc.

Discrete Variable: A quantitative variable that can assume a countable number of values. There is a distance between any two values.
An example is how many states a person can live in during his life. He could have lived in 1, 2, 3, 4, 5, 6, ..., 49, or 50 states. He couldn't have lived in 3.68797869 states. Notice that there is a difference of at least 1 between any two values.
Figure skating judges give fractional scores of 9.9, 7.5, etc. Yet the scores are still discrete because any two scores differ by 0.1 increments. No one will get a score such as 9.8473.

Continuous Variable: A quantitative variable that can assume an uncountable number of values. A continuous variable can assume any value in an interval.
An example is how much gas can be used to fill a car. It could take 13, 13.423425, 13.746456 gallons, or any fractional amount of gas to fill the tank.

Discrete vs Continuous: A variable can be either discrete or continuous depending how the variable is measured.

When a person is asked his age, it is typical to give an answer in terms of years. If a person turned 25 two months ago, he would simply say he is 25 years old. With this reasoning, age would be a discrete variable.

However, he isn't really 25 years old. He is 25 years and 2 months old. Of course, he could also measure his age by weeks, days, minutes, seconds, or nano seconds. In this context, age would be a continuous variable.

Data: The set of values collected for the variables.

Observation: The measurement for a specific element or one data value.

Parameter: A numerical value summarizing all the data of the **population.**

Examples: If the population is every person in the US, a parameter could be the mean of the height of everyone in the US or the standard deviation of the height of everyone in the US.

Symbols for parameters are usually Greek letters such as $\mu, \sigma,$ or τ.

Statistic: A numerical value summarizing all the data of the **sample.**

For example, if the sample is 1000 randomly chosen people in the US, a statistic could be the mean of the height of the 1000 randomly chosen people.

Symbols for statistics are usually letters from the English alphabet such as $\overline{x}, s,$ or T.

Random Variable: A variable that assumes a numerical value to each of the outcomes in an experiment. It is called a random variable because it represents the unknown outcome of an experiment. The outcomes are uncertain until the experiment is actually carried out.
Examples:
A random variable could assign a 1 to every head result in a coin toss and a 0 to every tail.
A random variable could assign the number 1 to every blond, 2 to every brunette, etc.
A random variable could assign the number 23 to a 23 year old, 50 to a 50 year old, etc.

Describing Data

Measures of Central Tendency. One of the easiest and most common ways to describe data is by the average value. Average has been, and often is, used to refer to any of the measures of central tendency. To avoid misconceptions, the most common measures of central tendency are defined below.

Mean: The mean is commonly referred to as the "average". To find the mean, add all the values and divide by the number of values.

Population Mean:

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum\limits_{i=1}^{N} x_i}{N}$$

where x_i is an observation, and N is the number of observations in the population.

Sample Mean:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

where x_i is an observation, and n is the number of observations in the sample.

The mean is "sensitive" to outliers. If there is a large outlier in the data, the mean will be much larger than if the outlier was removed. Similarly, if there is a small outlier in the data, the mean will be much smaller than if the outlier was removed. So the mean is not a good representation of the "average" value of the data when outliers are present. We say the mean is not a **resistant measure** of central tendency.

Trimmed Mean: To lessen the effect of outliers on how well the mean represents the data, a small percentage (such as 5%) of both the smallest and largest values can be deleted. This removes the effect of the outliers. Then the mean is computed for the rest of the data.

Median: The median is the "middle value". Half of the observations in the data are less than the median and half are larger than the median. There is no formula for the median. Instead, write all the observations in order from least to greatest. If the number of values, n, is odd, the middle value is the median. This would be the $\frac{(n+1)}{2}$ observation. If n is even, there will be two middle numbers. The median is the average or mean of these two numbers. If n is even, $\frac{(n+1)}{2}$ will be a fraction. This serves as a reminder that the median is in between two observations.

The median is not sensitive to outliers. So the median is a **resistant measure** of central tendency.

Mode: The value that occurs with greatest frequency. To find the mode, count how many times each value shows up. The mode is the value that shows up the most.

Measures of Spread or Variability. Variability describes how far apart the data values are. Two sets of data might have the same mean, but the data might look very different.

For example: $X = \{4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6\}$ and $Y = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ both have a mean of 5, but Y has much more variability, because the data is spread further apart.

Range: This is the largest value minus the smallest value. $X_{(n)} - X_{(1)}$

Variance: This measure of variability is based on the difference between each value and the mean. There are different formulas for variance for populations and samples.

Population Variance:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

where x_i is an observation, N is the number of observations in the population, and μ is the population mean.

Sample Variance :

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

where x_i is an observation, n is the number of observations in the sample, and $\overline{x}$ is the sample mean.

Standard Deviation: This is the square root of the variance. Variance is changed to standard deviation to get correct units. Variance will be the original units squared, but standard deviation will have the same units as the original data.

For example, if the data is about money, variance would have units dollars squared, which has no practical application or interpretation. However, the standard deviation would be in units of dollars.

Population Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N}}$$

where x_i is an observation, N is the number of observations in the population, and μ is the population mean.

Sample Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

where x_i is an observation, n is the number of observations in the sample, and $\overline{x}$ is the sample mean.
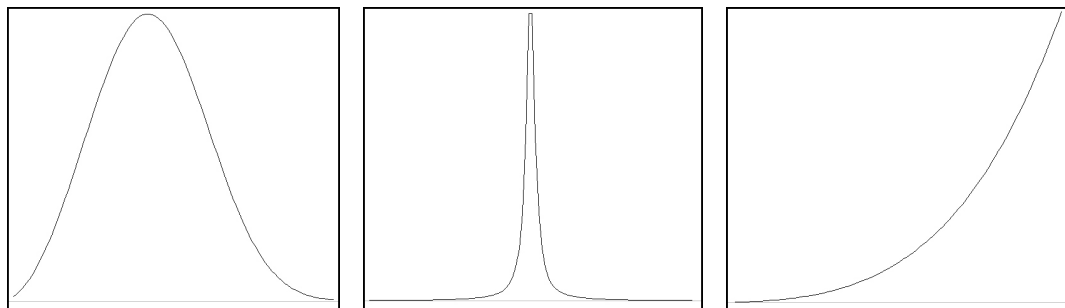
Distributions

A probability distribution describes the probability of each possible outcome, set of outcomes, or the probability that an outcome is in a particular interval. Distributions can be expressed with a table, equation, or graph.

Discrete or Continuous

Discrete Distribution: A discrete distribution describes the probability of discrete values, see page 58. The probability of a single value or sets of values can be found, such as $P(X = 3)$ or $P(X \geq 3)$. An example of discrete values is the number rolled on a die.
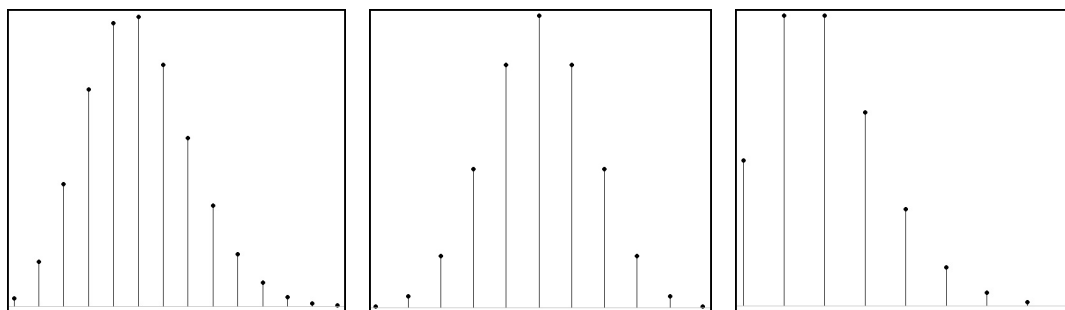
Continuous Distribution: A continuous distribution describes the probability of continuous values, see page 58. Note: The probability of a single value is zero. So only the probability of intervals is of interest. For example, $P(X = 3) = 0$, but $P(X \geq 3)$ can be found. An example of continuous values is a person's salary.


Probability of a point is zero: If a distribution is continuous, then the probability of a single point is zero, or $P(X = x) = 0$. A basic explanation is that a continuous distribution is defined on uncountably infinite values. So if there are infinitely many points to choose from, how likely is it to choose a specific value x which is only one out of infinite points? Of course the probability is zero.

Shape of Distribution

The shape of a distribution can be described several ways including symmetry and the number of modes.

Symmetry. A distribution can be left skewed, right skewed or symmetric.

Left Skewed: In a left skewed distribution, the bulk of the distribution is in the higher values on the right side, but there are extreme low values in the left. Since the bulk of the distribution is in the higher values, a person might expect to see a high mean. But the extreme low values in the left have to be taken into account as well and as a result the mean is lower than expected. If a distribution is left skewed, the mean is less than the median. (Picture it as a slide. If it is left skewed a person would slide down the left side.)
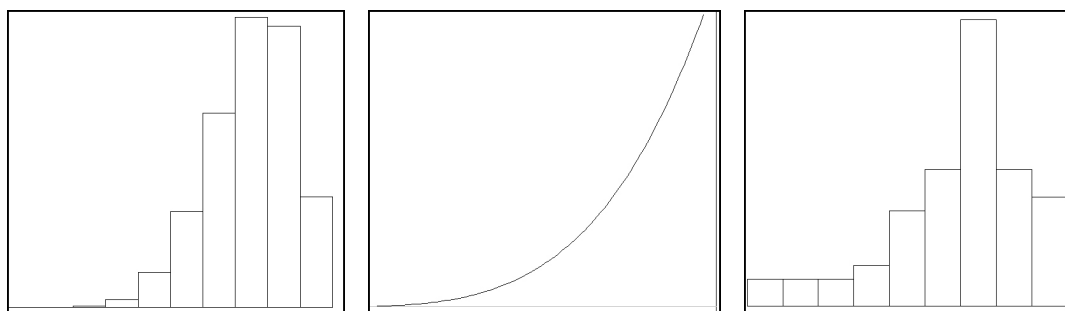
Right Skewed: In a right skewed distribution, the bulk of the distribution is in the lower values on the left side, but there are extreme high values on the right that when accounted for make the mean higher than would have been expected after seeing all the low values. The mean is greater than the median. (If it was a slide a person would slide down the right side.)

Skewness can be extreme (picture below on left) or very slight (picture below on right).
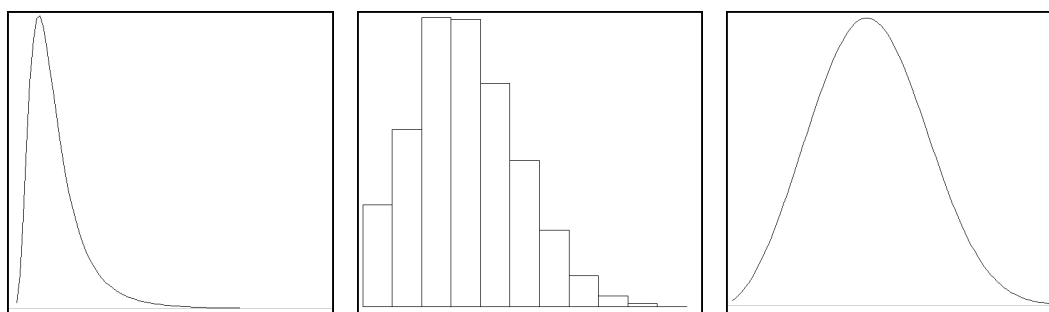
Example: Most people have an small or moderate income, but if a few rich people are included in the distribution, the mean salary will suddenly be much higher. This is a right skewed distribution.

Note: If a distribution is skewed, the mean is not a good representation of the "average" value of the data. This is another reason that the mean is not a **resistant measure** of central tendency (see page 60).

Symmetric: A symmetric distribution looks symmetric, i.e., it looks the same for values less than the mean and values greater than the mean. A symmetric distribution is not skewed to the left or right.
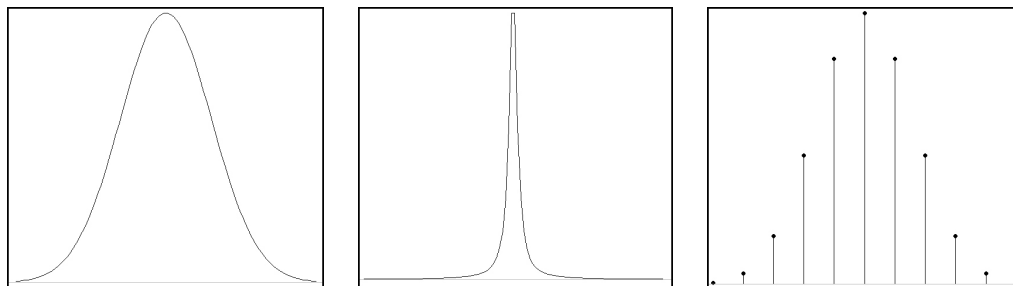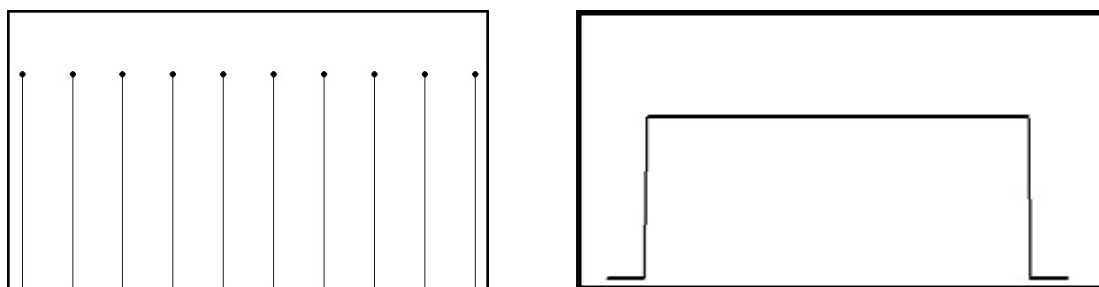
Bell Shaped Curves. Some distributions have what is called a bell shape. They are symmetric with a peak at the mean.
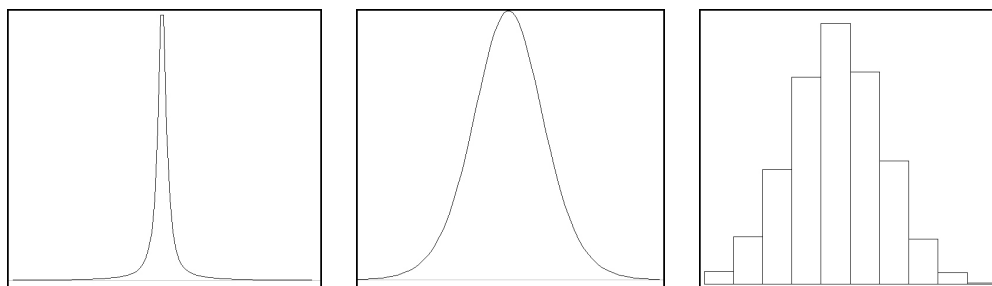
The most famous bell shape is the Gaussian curve or normal curve, see page 71.
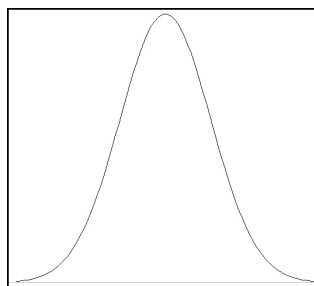
Rectangular Curves. A rectangular curve is the uniform distribution, see page 70. Each value has the same probability or frequency.

Modes. A distribution can be described by the number of modes. The mode is the number that occurs the most, or the value at which the distribution has a maximum value. When looking at a graph, the mode will be the peak. There can be more than one mode.

Unimodal: A distribution that has only one mode or "peak".
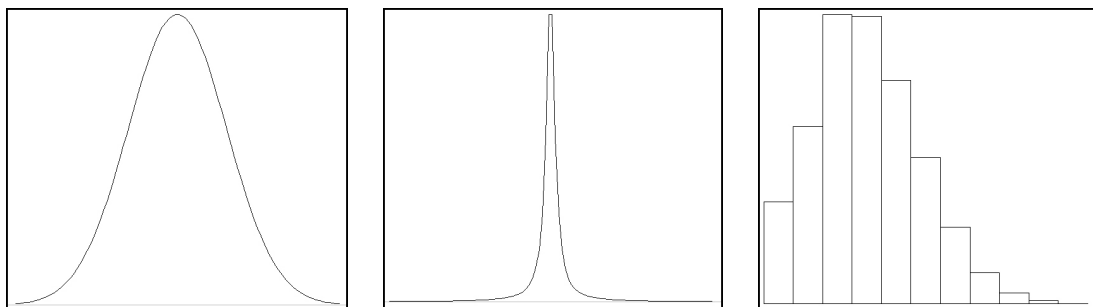
Bimodal: A distribution that has two modes or "peaks".
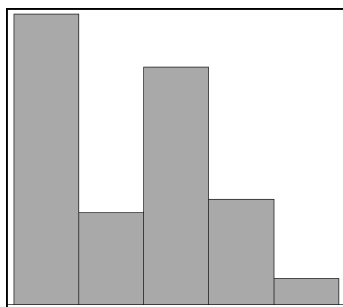
Multimodal: These distributions have two or more modes or "peaks".
This means that bimodal distributions are also multimodal. The rule of thumb is to be as specific as possible and call a distribution bimodal if it has two peaks.

Basic Inference

The population refers to all the individuals or events of interest. However, it is usually too difficult, time consuming, or expensive to collect data on an entire population. So a smaller subset of the population called t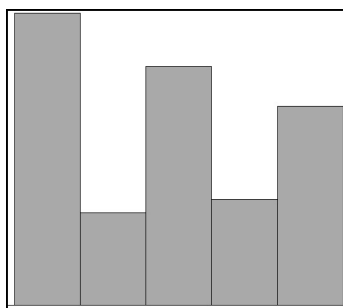he sample is chosen on which to gather data. It is assumed that the sample is representative of the population. If the sample is representative of the population, it can be used to make inferences about the population.

A parameter is a number that describes a characteristic of the population. We wish to know the value of the parameter, but that is impossible without having data on the entire population.

So a statistic which describes a characteristic of the sample is used to make inferences about the parameter.

Example. Suppose a bread factory wished to know if the mean weight of all the loaves produced is 2 pounds. The parameter of interest would be the mean weight of the bread.

It would be difficult and time consuming to weigh every loaf of bread. So a sample of bread could be chosen to weigh. Perhaps one in every hundred loaves could be weighed. The mean of this sample would be the statistic.

The sample mean serves as an estimate of the population mean.

Sampling Distribution

All Possible Samples. From any population, there are many different samples of size n that can be chosen.

Example. Let a population be $\{0, 1, 2, 3, 4\}$. The following are samples of size 2 that could be chosen without replacement:

$\{0,1\}$	$\{0,2\}$	$\{0,3\}$	$\{0,4\}$	$\{1,2\}$
$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$

Notice that there are 10 different samples that could have been chosen.

Sampling Distribution. Since there are different samples that can be chosen, any sample statistic, (i.e. sample mean, sample mode, sample standard deviation, sample range, sample maximum value), will be different based on which sample is chosen. Each statistic will be a random variable because its value changes from sample to sample. That means that each statistic also has its own distribution, called the sampling distribution.

Example. Calculate the sample mean of each of the samples from the previous example.

Sample	$\overline{x}$	Sample	$\overline{x}$
$\{0,1\}$	0.5	$\{1,3\}$	2
$\{0,2\}$	1	$\{1,4\}$	2.5
$\{0,3\}$	1.5	$\{2,3\}$	2.5
$\{0,4\}$	2	$\{2,4\}$	3
$\{1,2\}$	1.5	$\{3,4\}$	3.5
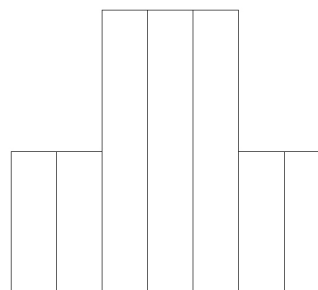
Notice that there are 10 different values for the sample mean depending on which sample is chosen. That means that the sample mean has its own distribution. Also shown is the histogram for the sampling distribution of the sample mean.

Binomial Distribution

Binomial Distribution. The binomial distribution is discrete. The binomial distribution describes the probability of getting x successes out of n trials and can be used when:

- The number of trials, n, is fixed.
- The trials are independent. (The outcome of one trial does not affect the outcome of the next trial.)
- There are only two possible outcomes: success and failure.
- The probability of the success, or failure, is the same for each trial.

REMARK. Trials are called identical when the possible outcomes are the same for each trial and each possible outcome has the same probability. So the binomial distribution can be used when there are n identical and independent trials.

Probability Mass Function. The Binomial distribution has the probability mass function

$$p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x \left(1 - p \right)^{n-x}$$

where

$$
\begin{aligned}
p\left(x\right) &= \text{the probability of } x \text{ successes in } n \text{ trials} \\
n &= \text{the number of trials} \\
x &= \text{the number of successes} \\
p &= \text{the probability of a success on any one trial} \\
1 - p &= \text{the probability of a failure on any one trial} \\
\binom{n}{x} &= \frac{n!}{x!\,(n-x)!}
\end{aligned}
$$

The binomial distribution is only defined for integer values 0 or greater, i.e. $0, 1, 2, 3, \ldots$, because the binomial distribution describes the number of successes.

The graphs of the probability mass functions for values of $n = 30$ and $p = 0.3, 0.5,$ and 0.8 are shown below.
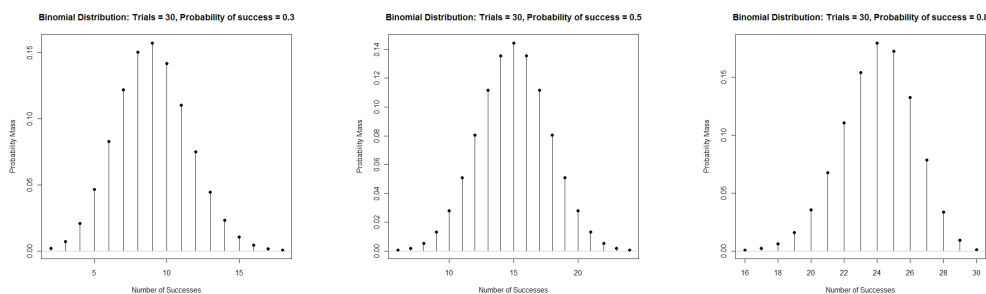
How the probability mass function is derived. If two events are independent, the probability of both events happening is the product of the probability of each event.

The binomial distribution describes how many successes happen in n trials. When finding the probability of getting x successes, it doesn't matter which of the n trials were successes, just that there were x total successes. There are $\binom{n}{x}$ or "n choose x" ways to choose x successes from n total trials.

We need x successes. The probability of each success is p. So the probability of getting x successes is p^x (p multiplied x times).

Of course, if there are only x successes, then the rest of the trials must be failures. There are $n - x$ trials left and the probability of each failure is $1 - p$. So the probability of $n - x$ failures is $(1 - p)^{n-x}$.

Therefore the probability of getting x successes out of n trials is $\binom{n}{x} p^x (1 - p)^{n-x}$.

Expected Value and Variance.

$$E(X) = \mu = np$$

$$Var(X) = \sigma^2 = np(1 - p)$$

Normal Approximation. The normal distribution (page 71) can be used to approximate the binomial distribution when the sample size is large and the probability of a success isn't too close to 0 or 1. The rule of thumb is that the normal distribution can be used when $np \geq 10$ **and** $n(1 - p) \geq 10$.

The number of successes, X, is approximately normally distributed with $\mu = np$ and $\sigma^2 = np(1 - p)$.

$$X \overset{approx}{\sim} N(np, np(1 - p))$$

The sample proportion, $\hat{p}$, is approximately normally distributed with $\mu = p$ and $\sigma^2 = \dfrac{p(1 - p)}{n}$.

$$\hat{p} = \frac{X}{n}$$

$$\hat{p} \overset{approx}{\sim} N\left(p, \frac{p(1 - p)}{n}\right)$$

Continuity Correction. The binomial distribution is discrete and the normal distribution is continuous, so a correction is needed to handle the differences between the two distributions. Using the continuity correction makes the normal approximation a better approximation. Notice in the picture below the difference between the discrete and continuous functions.
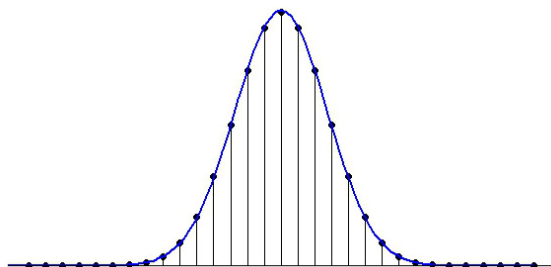

For example, $P(X = x)$ can be found for a binomial distribution, but $P(X = x) = 0$ for any continuous distribution. So to approximate $P(X = x)$ with the normal distribution, find $P(x - 0.5 \leq X \leq x + 0.5)$. The continuity correction can be made by adding or subtracting 0.5 from x to be on the "safe side".

Binomial	Normal Approximation with Continuity Correction
$P(X = x)$	$P(x - 0.5 \leq X \leq x + 0.5)$.
$P(X \geq x)$	$P(X \geq x - 0.5)$
$P(X \leq x)$	$P(X \leq x + 0.5)$

For a continuous function, there is no difference in $P(X \leq x)$ and $P(X < x)$. However, there is a big difference between $P(X \leq x)$ and $P(X < x)$ if the distribution is discrete. When talking about the binomial distribution, saying $X \leq 5$ means X can be $\{0, 1, 2, 3, 4, 5\}$, but saying $X < 5$ means X is $\{0, 1, 2, 3, 4\}$. This means that $X < 5$ is equivalent to $X \leq 4$ when working with integers.

Binomial	What is Meant	Normal Approximation with Continuity Correction
$P(X > x)$	$P(X \geq x + 1)$	$P(X \geq (x + 1) - 0.5)$ or $P(X \geq x + 0.5)$
$P(X < x)$	$P(X \leq x - 1)$	$P(X \leq (x - 1) + 0.5)$ or $P(X \leq x - 0.5)$
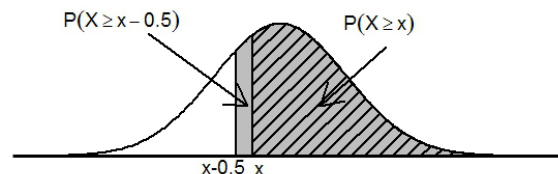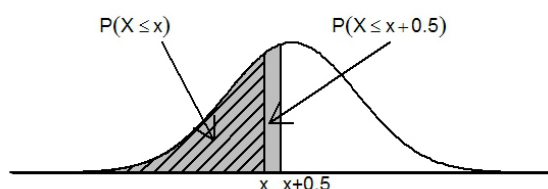
Discrete Uniform Distribution

The uniform distribution is used when every outcome has the same chance of occuring.

Probability Mass Function. The discrete uniform distribution has the probability mass function

$$p(x) = \frac{1}{n}$$

where n is the number of possible outcomes.
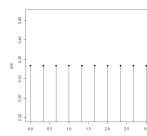
Integers are the numbers $\{..., -3, -2, -1, 0, 1, 2, 3, ...\}$. If the possible outcomes are integers, call the smallest possible outcome a and the greatest possible outcome b. Then,

$$n = b - a + 1$$

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a+1)^2 - 1}{12} = \frac{n^2 - 1}{12}$$

Continuous Uniform Distribution

The continuous uniform distribution is defined on an interval, $a \leq X \leq b$, with constant probability on the interval.

Probability Density Function. A continuous uniform random variable has the probability density function

$$f(x) = \frac{1}{b-a}$$

where x can be any value between a and b. The distribution can be abbreviated as $U(a,b)$.

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution

The normal distribution is a continuous probability distribution. The density function is bell shaped, centered at the mean, and is determined by the mean μ and the standard deviation σ.

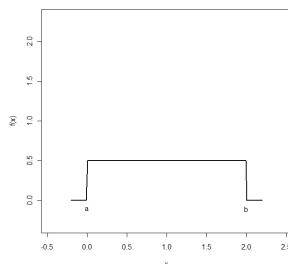The density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

The notation $X \sim N\left(\mu, \sigma^2\right)$ means that the random variable X is normally distributed with mean μ and variance σ^2. A few books use the notation $X \sim (\mu, \sigma)$ meaning that the standard deviation is σ.

The cumulative density function is represented with the notation Φ;

$$\Phi(x) = P\left(X \leq x\right)$$

Standard Normal Distribution. There are infinitely many normal distributions with different means and variances, far too many to make a table of probabilities for each. Instead only one table is used, the table for the standard normal distribution.

The standard normal distribution has mean $\mu = 0$ and standard deviation $\sigma = 1$. If the data or values fit any normal distribution, the data can be "standardized" by the formula,

$$z = \frac{x - \mu}{\sigma}$$

This is called the z-score. The distance between x and μ in terms of standard deviations is given by z. For example, if $z = 3$, then x is 3 standard deviations from the mean.

After the data is standardized, it will have a standard normal distribution. The standard normal distribution is represented by Z.

$$Z \sim N\left(0, 1\right)$$

Given an x value, find probabilities. If x is a value from a normal distribution, to find any associated probabilities, such as $P(X \geq x)$, $P(X < x)$, etc., first convert the x value to the z-score and then use a standard normal table or calculator. Drawing pictures is often helpful.

Example 1. If $X \sim N\left(3, 2.25\right)$, find the probability that a number drawn from the distribution is greater than 2. This means find $P\left(X > 2\right)$. Since the normal distribution is continuous, this is the same as finding $P\left(X \geq 2\right)$. See page A. The first step is to convert the x value to a z-score. It is given that

$$x = 2,\ \mu = 3,\ \sigma^2 = 2.25\,\text{or}\,\sigma = 1.5$$

So the z-score is

$$z = \frac{2 - 3}{1.5} = -0.66$$

Once the x value is standardized, $P(X \geq 2)$ is the same as $P\left(Z \geq -0.66\right)$. So find $P\left(Z \geq -0.66\right)$ by either looking at a table or using a calculator.

area=0.745
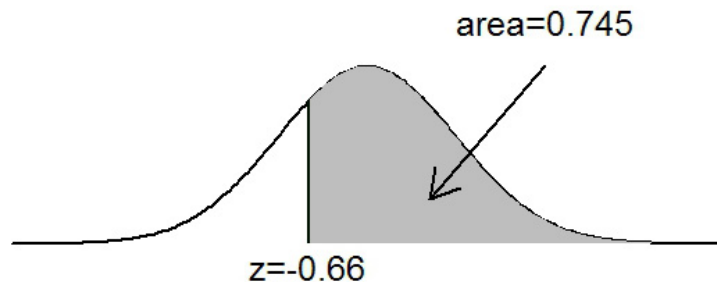
z=-0.66

Therefore,

$$
\begin{aligned}
P(X > 2) & = & P(X \geq 2) \\
& = & P(Z \geq -0.66) \\
& = & 0.745
\end{aligned}
$$

Example 2. If $X \sim N(20, 4)$, find $P(X \leq 17 \text{ or } X \geq 22)$. First find the z-scores for each x value.

$$
z = \frac{17 - 20}{4} = -0.75 \text{ and } z = \frac{22 - 20}{4} = 0.5
$$

So find $P(Z \leq -0.75 \text{ or } Z \geq 0.5) = 0.226 + 0.308 = 0.534$.
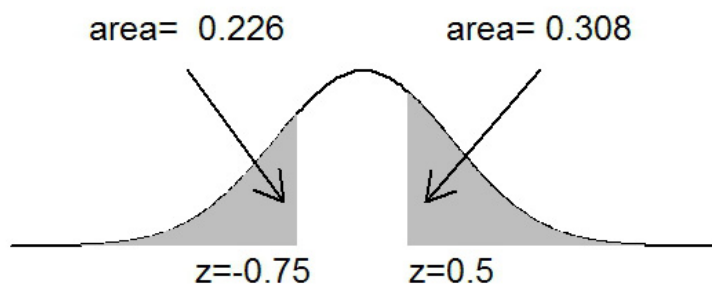
The 68-95-99.7 Rule of Thumb. For the normal distribution with mean μ and standard deviation σ:

- Approximately 68% of the observations are within one standard deviation of the mean.
- Approximately 95% of the observations are within two standard deviations of the mean.
- Approximately 99.7% of the observations are within three standard deviations of the mean.
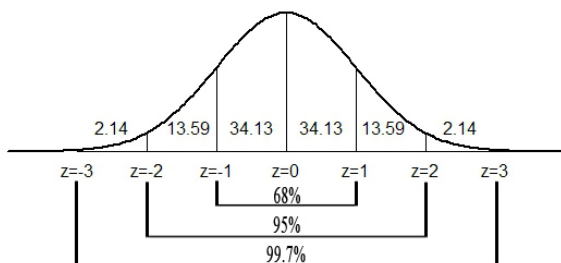
What is it used for? The normal distribution can be used to describe many natural phenomenon. For example, the heights of people in a city would follow the normal distribution. The normal distribution is used as an approximation for other distributions including the binomial distribution (page 68), and t-distribution (page 74).

The sampling distribution of

$$
\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}
$$

is a standard normal distribution. The standard normal distribution is useful for inferences about the population mean when σ^2 is known.

Chi-Square Distribution

This continuous distribution is right skewed. The shape of the distribution depends on the degrees of freedom or df. The symbol for this distribution is χ^2_{df}.

$df = 1$	$df = 5$	$df = 10$
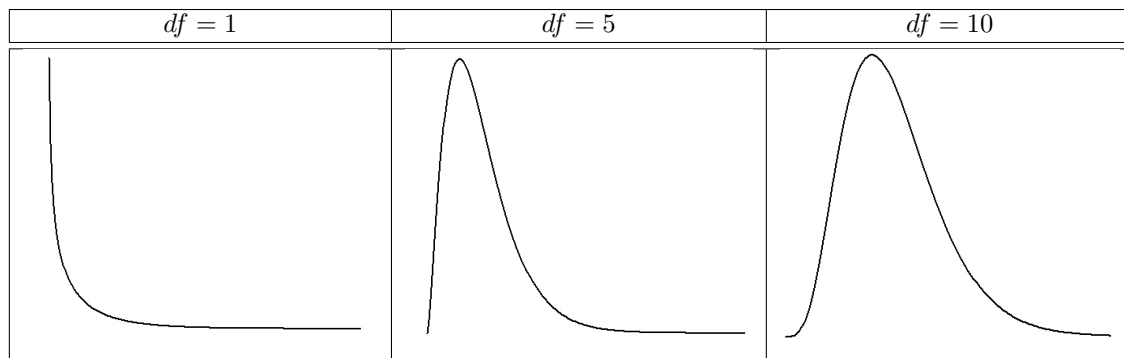
Epected Value and Variance. For a chi square distribution with v degrees of freedom, the mean and variance are

$$\mu = E(X) = v$$
$$\sigma^2 = Var(X) = 2v$$

What is it used for? The sampling distribution of

$$\frac{(n-1)\,S^2}{\sigma^2}$$

is a chi square distribution with $n-1$ degrees of freedom. It is useful for inferences about the population variance. It is also used to analyze multinomial or contigency tables for goodness of fit, independence, and homogeneity.

Where does it come from? The chi-square distribtuion can be found by squaring standard normal random variables. If Z is a random variable with a standard normal distribution, then $U = Z^2$ has a chi square distribution with 1 degree of freedom. If V is the sum of n independent chi square random variables that each have 1 degree of freedom, then V has a chi square distribution with n degrees of freedom or $V \sim \chi^2_n$.

t Distribution

This continuous distribution is symmetric about its mean which is zero. It is bell shaped and looks similar to the normal distribution, but the shape of the distribution depends on the degrees of freedom or df. The symbol is t_{df}.

The t distribution can be more spread out, or have greater variance, than the normal distribution. As the sample size becomes large, the t-distribution approaches the normal distribution. The rule of thumb is that if $n \geq 30$, the normal curve can be used to approximate the t distribution because the two distributions are so close.

The picture below shows the t-distributions with degrees of freedom of 1, 5, and 30. The dashed line is the standard normal distribution. Notice that as the degrees of freedom increases, the graph of the t-distribution approaches the graph of the standard normal distribution.
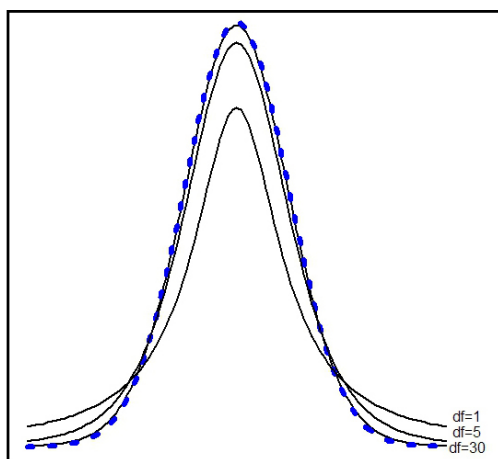
Expected Value and Variance. If the t distribution has v degrees of freedom, then

$$\mu = E(X) = 0$$
$$\sigma^2 = Var(X) = \frac{v}{v-2}$$

The variance only exists if the degrees of freedom is greater than 2. Otherwise the variance is undefined.

What is it used for? The t-distribution is the sampling distribution of

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

It is useful for inferences about the population mean when the population variance is unknown.

Where does it come from? The t-distribution is the ratio of a standard normal random variable and a variable with a chi-square distribution with v degrees of freedom. If $Z \sim N(0,1)$ and $V \sim \chi_v^2$ then

$$\frac{Z}{\sqrt{\frac{V}{v}}}$$

has a t-distribution with v degrees of freedom.

Degrees of Freedom. What does degrees of freedom mean? For one sample, if $\overline{x}$ is known, there are only $n-1$ independent sample values. This is because if $n-1$ values and the sample mean are known, then it is possible to find the nth value. So the nth value can only be one value that it can be.

Example: If a sample has size $n=4$, and it is known that the sample mean is $\overline{x}=4$, and three of the sample values are known to be 2, 3, and 5, find the 4th sample value.

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\
&= \frac{x_1 + x_2 + x_3 + x_4}{4} \\
&= \frac{2 + 3 + 5 + x_4}{4}
\end{aligned}$$

Since $\overline{x}=4$,

$$4 = \frac{2 + 3 + 5 + x_4}{4}$$

This equation can be solved giving $x_4 = 6$. So the 4th sample value had to be 6.

In a general sense, the degrees of freedom is how many values are independent, or free to be any value, when a statistic is calculated. The degrees of freedom of a sample statistic can also be thought of as the number of independent or free data values (usually the sample size) minus the number of parameters that were estimated in the process of finding the statistic. For example, when trying to estimate the population mean by finding the sample mean, there is no need to estimate any other parameters, so the degrees of freedom is the sample size. But when trying to estimate the variance, it is necessary to know the sample mean which is an estimate of the population mean. Since one parameter is estimated in the process of finding the variance, the degrees of freedom would be $n-1$.

F Distribution

This is a continuous distribution, that is not symmetrical. It is skewed to the right. The shape of the curve depends on two different degrees of freedom, the numerator degrees of freedom, df_n and the denominator degrees of freedom df_d.
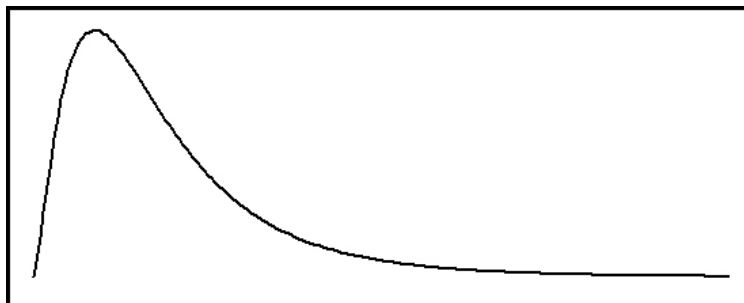
What is it used for? The F distribution is the sampling distribution of

$$\frac{S_X^2}{S_Y^2}$$

It is useful to make inferences when comparing two populations' variances.

Where does it come from? The F distribution is the ratio of two chi-square random variables and their degrees of freedom. If U and V are independent chi-square random variables with degrees of freedom u and v, then

$$\frac{\frac{U}{u}}{\frac{V}{v}} \sim F_{u,v}$$

Central Limit Theorem

The central limit theorem says that for a population with **any** distribution with mean μ and variance σ^2, if the simple random sample size n is large enough, then the mean will be approximately normally distributed. That means that the sampling distribution of the mean will have an approximate normal distribution.

If the original population had mean μ and variance σ^2, then

$$\overline{X} \overset{approx}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Caution! This result can only be used if n is large and the population variance is finite (any number but $\pm\infty$).

Alternate Central Limit Theorem. Sometimes the central limit theorem talks about the sum of all the values in the sample instead of the mean. Call the sum of the values in the sample S. Then if a simple random sample is drawn from a population with any distribution with mean μ and variance σ^2, then the sampling distribution of S will be approximately normal.

$$S \overset{approx}{\sim} N\left(n\mu, n\sigma^2\right)$$

This result holds when n is large and the population variance is finite (any number but $\pm\infty$).

Sampling Distribution of Means and Proportions

Sampling Distribution of the Sample Mean. When a sample is drawn from a population, there are many different samples that could have been chosen. That means that the mean of the sample, $\overline{x}$ will vary depending on which sample is picked. Thus $\overline{X}$ is a random variable because the sample mean varies from sample to sample. Every sample has a mean and variance. Since $\overline{X}$ depends on the sample picked, theoretically all the possible sample means could be found and then the true mean and variance of the sample means could be calculated.

If $\overline{x}$ is the mean of a sample of size n from any population with known mean μ and variance σ^2, then the mean and standard deviation of all the $\overline{x}'s$ (the random variable $\overline{X}$) can be found.

$$\mu_{\overline{X}} = \mu$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

Note. The symbol $\overline{x}$ is used to refer to the mean of a single sample. $\overline{X}$ is the random variable that describes the different sample means from all possible samples.

Note. The larger the sample size, the less variability there is and $\overline{x}$ becomes a better estimate of μ. See the example below.

Note. As n becomes large, the distribution of $\overline{X}$ approaches the normal distribution and

$$\overline{X} \overset{approx}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

See the Central Limit Theorem on the preceding page.

Example 1. 500 samples were drawn from a normal population with mean $\mu = 100$ and standard deviation $\sigma = 10$. The mean of each sample was calculated. The histograms of the sample means are shown below along with the normal curve with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$ for sample sizes of 5, 50, and 500. Notice that as the sample size increases, the distribution of $\overline{X}$ becomes closer to the normal distribution. Also as the sample size increases, there is less variability as the values of $\overline{X}$ become closer to the mean $\mu = 100$.

Sample Size 5   Sample Size 50   Sample Size 500
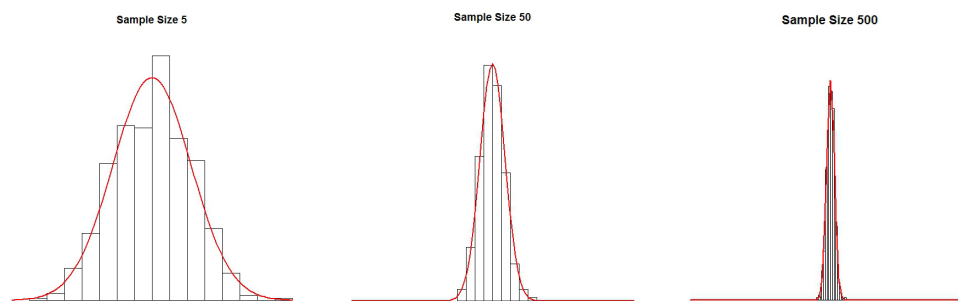
Of course, it is easy to believe that the sample means from a normal distribution will be normally distributed. However, the sample means from any distribution will be normally distributed.

Example 2. Below is shown a chi-square distribution with 3 degrees of freedom.
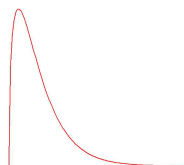
Notice that the distribution does not look normal. It is in fact very right skewed. 500 samples were drawn from this distribution and the sample mean calculated for each sample. This was done for sample sizes of 5, 50, and 500. Then the histograms of the sample means were overlaid with the normal curve with $\mu_{\overline{X}} = 3$ and $\sigma^2_{\overline{X}} = \dfrac{6}{n}$ (See page 73 to find μ and σ^2 for a chi-square distribution).

Sample Size 5   **Sample Size 50**   **Sample Size 500**

Notice that even with a sample size of 5, the histogram of $\overline{X}$ looks more symmetric and more like a normal distribution than the original distribution of X.

It takes a much larger sample size for $\overline{X}$ to be close to the normal distribution when the original distribution of X is chi square than when it is normal.

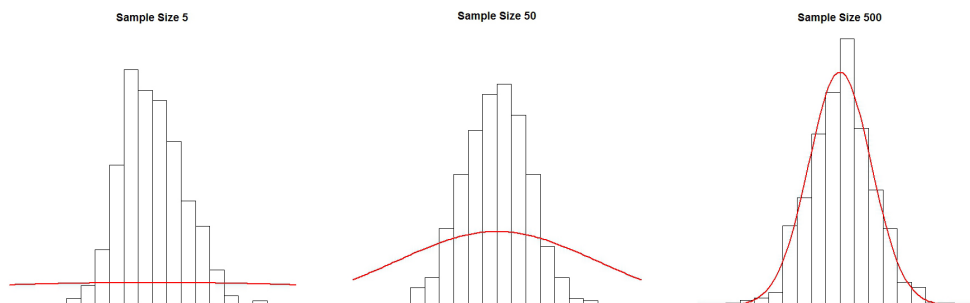**Sampling Distribution of a Sample Proportion.** There are many samples that can be drawn from a population. That means that the proportion of successes of the sample, $\hat{p}$ will vary depending on which sample is picked. This means that $\hat{P}$ is a random variable because it varies from sample to sample. Since $\hat{P}$ varies depending on the sample, the sample proportions of success could be found for every possible sample and then the true mean and variance of the sample proportions could be calculated.

If $\hat{p}$ is the proportion of success of a sample of size n from any population with proportion p, then the mean and standard deviation of all $\hat{p}'s$ can be found.

$$\mu_{\hat{P}} = p$$

$$\sigma^2_{\hat{P}} = \frac{p\,(1-p)}{n}$$

Note: As the sample size increases, the variability decrease and $\hat{p}$ becomes a better estimate of p.

Hypothesis Testing

What is Hypothesis Testing? Hypothesis testing is used to obtain information about a population parameter. A hypothesis is created about the population parameter, and then a sample from the population is collected and analyzed. The data found will either support or not support the hypothesis.

Steps for Hypothesis Testing.

(1) Choose the null and alternative hypotheses.
(2) Choose a significance level.
(3) Calculate the test statistic.
(4) Calculate the p-value.
(5) State the conclusion.

Null and Alternative Hypotheses. The null hypothesis, denoted H_0, is the statement that is being tested. Usually the null hypothesis is the "status quo" or "no change" hypothesis. The hypothesis test looks for evidence against the null hypothesis.

The alternative hypothesis, denoted H_A or H_1, is the statement that we are hoping is true or what we wish to prove. It is the "opposite" of the null hypothesis. Since we wish to prove the alternative hypothesis, we usually write the alternative hypothesis first and then the null hypothesis.

A food company has just implemented a new bread recipe. The company wishes to know if the mean weight per loaf has changed. Since the company wishes to know if the mean weight has changed, that is the alternative hypothesis, while the null hypothesis, or status quo, is that the mean weight of the bread hasn't changed. The hypotheses could be:

H_0: The mean weight per loaf of bread hasn't changed.
H_A: The mean weight per loaf of bread has changed.

This can be written symbolically. Let μ_0 be the old mean weight of the bread. It is the hypothesized value. That is the weight that the bread would be if it doesn't change. The true mean weight of the bread in the population is called μ.

$H_0 : \mu = \mu_0$
$H_A : \mu \neq \mu_0$

One-sided and Two-sided Alternative Hypotheses. The alternative hypothesis can be two-sided or one sided.

In the bread example above, the company wanted to know if the mean weight had changed. The alternative hypothesis only says that the mean weight changed. The alternative hypothesis doesn't give any information about whether the mean weight has increased or decreased. If the management suspects that the mean weight had increased the hypotheses could be:

$H_0 : \mu = \mu_0$ (The mean weight per loaf of bread hasn't changed.)
$H_A : \mu > \mu_0$ (The mean weight per loaf of bread has increased.)

This is an example of a one-sided alternative hypothesis. Another one-sided hypothesis test would be if the management thinks that the mean weight had decreased. Then the hypotheses could be:

$H_0 : \mu = \mu_0$ (The mean weight per loaf of bread hasn't changed.)
$H_A : \mu < \mu_0$ (The mean weight per loaf of bread has decreased.)

Caution. It might be tempting to look at the data and then write the alternative hypothesis or pick the significance level. DON'T! Pick the hypotheses and significance level before computing anything for the hypothesis test.

Level of Significance.

Most hypothesis tests fall in the category of significance tests. Before the test is started (before the sample is chosen and anything is computed), a significance level, α is chosen. The most commonly used significance levels are $\alpha = 0.10$, 0.05, or 0.01. If a significance level isn't specified, $\alpha = 0.05$ is the common choice. The significance level is how much evidence is needed to reject the null hypothesis. For example, if $\alpha = 0.05$ is chosen, the evidence is considered strong enough to reject the null hypothesis if the data in the sample would only happen 5% of the time, or less, **when the null hypothesis is true**. That means that the null hypothesis will only be rejected when the data in the sample isn't very likely if the null hypothesis is true.

Test Statistic. A statistic is any value that is computed from the data in the sample. A test statistic is a statistic that can be used to find evidence in a hypothesis test.

If a hypothesis test is conducted to find information about the population mean, the sample mean would be a logical choice of a statistic that would be useful. Similarly, if a someone wishes to know about the population variance, the sample variance is a possible test statistic.

P-values. The p-value is the probability of observing an outcome as extreme or more extreme as the observed sample outcome if the null hypothesis is true.

Example. Consider the set of hypotheses that the mean weight of a loaf of bread hasn't changed versus the mean weight of a loaf of bread has increased. If a sample of bread loaves is chosen, and the mean weight is found to be $\overline{x} = 2.14$ pounds, then the p-value is $P(\overline{X} \geq 2.14)$.

Rejecting or Failing to Reject the Null Hypothesis

If the p-value is less than the significance level, reject the null hypothesis. For example, if $\alpha = 0.05$ and the p-value is 0.03, reject the null hypothesis because we expect to see the observed outcome only 3% of the time if the null hypothesis is true. So the observed outcome isn't very likely. More specifically the probability of the observed outcome happening was less than 5% if the null hypothesis is true. So we reject the null hypothesis in favor of the alternative hypothesis and say, "there is sufficient evidence to reject the null hypothesis". To summarize with non technical language, if something is not very likely, reject it.

If the p-value is greater than the significance level, we fail to reject the null hypothesis. For example, if $\alpha = 0.05$ and the p-value is 0.15, we fail to reject the null hypothesis. The observed outcome is expected 15% of the time if the null hypothesis is true. This may not seem very likely, but it is more likely than 5% so the conclusion is to fail to reject the null hypothesis, and we say, "there is not sufficient evidence to reject the null hypothesis."

Why shouldn't the conclusion be, "there is sufficient evidence to accept the null hypothesis"? It is a convention based on the fact that in mathematics, statements are not proved with examples. A claim can be disproven with one example, but even one million examples in favor of the claim can't prove it. To borrow a common phrase, "Absence of evidence is not evidence of absence". However, hypothesis tests don't actually prove anything anyways. They are just a method of judging the evidence for or against a hypothesis. Yet, the tradition is strong enough, that a conclusion should never be, "there is sufficient evidence to accept the null hypothesis".

Analogy. Until the 17th century Europeans thought every swan was white because for centuries, every swan they saw was white. Then, a black swan was discovered in Australia, instantly disproving the hypothesis that all swans are white.

Analogy. Suppose a person thinks that there might have been a skunk in his yard the previous night. A null hypothesis is that there was no skunk in the yard (status quo). The alternative hypothesis would then be that there was a skunk in the yard.

H_0: There was no skunk in the yard.

H_A: There was a skunk in the yard.

He could go outside the next day and look for evidence that there was a skunk. If he finds skunk fur or smells a skunk, then he would have evidence to reject the null hypothesis in favor of the alternative hypothesis (that there was a skunk).

On the other hand, if he doesn't find evidence that a skunk was there, that does not mean that the null hypothesis is true. A skunk could have been there without leaving evidence. That is why he shouldn't say he accepts the null hypothesis. He doesn't know for sure that there wasn't a skunk. He just doesn't have evidence to support the claim that there was a skunk. So he says there is not sufficient evidence to reject the null hypothesis or that he fails to reject the null hypothesis.

If he rejects the null hypothesis, he could technically say that he accepts the alternative hypothesis. However, tradition dictates that conclusions are always stated as rejecting or failing to reject hypothesis rather than accepting hypothesis.

If he finds skunk fur in the yard, he would reject the null hypothesis. Yet he still hasn't proved that there was a skunk. The dog could have brought the fur into the yard. Because he hasn't proved the alternative hypothesis, (he might have strong evidence that there was a skunk, but he hasn't proven it) he shouldn't say that he accepts the alternative hypothesis.

Hypothesis Testing Errors

Type I Error: A type I error happens when a true null hypothesis is rejected. The probability of a type I error is denoted α.

Type II Error: A type II error occurs when a false null hypothesis is not rejected. The probability of a type II error is denoted β.

Decision	H_0 is true	H_0 is false
Reject H_0	Type I Error	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error

The goal is to make both α and β as small as possible. Unfortunately for a fixed sample size, decreasing α will increase β and vice versa. It is necessary to choose which error is more important to decrease based on the scenario. In most cases, β is difficult to calculate so α is set between 0.01 and 0.10.

The probability of a type I error, α, is also the level of significance. If $\alpha = 0.05$ is chosen, then a null hypothesis is rejected if the sample would only happen 5% of the time if the null hypothesis is true. That means that there is a probability of 0.05 that the null hypothesis will be rejected when it is true.

Analogy. Trials are like hypothesis tests. Since a person is innocent until proven guilty, innocence is the status quo.

$H_0 =$ the plaintiff is innocent

$H_A =$ the plaintiff is guilty

If a jury convicts an innocent man, the jury has made a type I error. If the jury comes to the conclusion that a man is innocent, but he was actually guilty, they have made a type II error.

In this case, a decision has to be made about whether it is better to minimize the probability of a type I error or a type II error. Is it better to send an innocent man to jail or to release a guilty man? In criminal trials the precedent is that a man is only convicted if the evidence is beyond a reasonable doubt. They We don't want to convict an innocent man, so the courts try to minimize the probability of a type I error. Of course this means that more guilty people are not convicted or the probability of a type II error is higher.

Power of a Test

The power of a test is the probability of rejecting the false null hypothesis. This is also the same as the probability of not making a type II error. Therefore,

$$Power = 1 - \beta$$

Analogy. In the jury example, the power of the test would be the probability of concluding a guilty man is guilty.

In the bread example, the power of the test is the probability of concluding that the mean weight of the bread has changed when the weight of the bread really did change.

One Sample z-Test for the Population Mean

When can this test be used?

- The population is normally distributed or the sample size is large enough that the mean is normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 30$.)
- The population standard deviation, σ, is **known**.

How is this test used?

(1) State the hypotheses:
$H_0 : \mu = \mu_0$
$H_A : \mu \neq \mu_0$ or $H_A : \mu > \mu_0$ or $H_A : \mu < \mu_0$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\text{the sample mean} - \text{the hypothesized true population mean}}{\text{the population standard deviation} \big/ \sqrt{\text{the sample size}}}$$

(4) Find the p-value:
The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the standard normal curve.

If $H_A : \mu > \mu_0$, p-value $= P\left(Z \geq z\right)$

If $H_A : \mu < \mu_0$, p-value $= P\left(Z \leq z\right)$

If $H_A : \mu \neq \mu_0$, p-value $= P\left(Z \leq -|z| \text{ or } Z \geq |z|\right)$ or $2P\left(Z \geq |z|\right)$
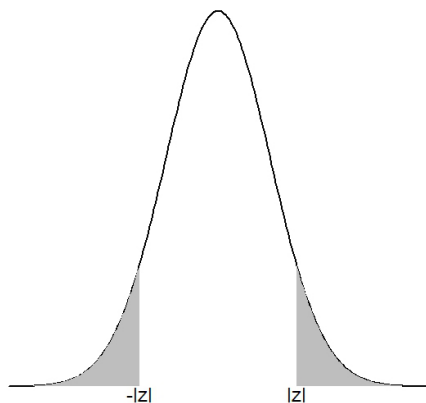
(5) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

How does the one sample z-test work?

The z-test is used to make inferences about the mean of a population.

(1) State the hypotheses which could be:

$H_0 : \mu = \mu_0$

$H_A : \mu \neq \mu_0$ or $H_A : \mu > \mu_0$ or $H_A : \mu < \mu_0$

(2) Pick a significance level, α.

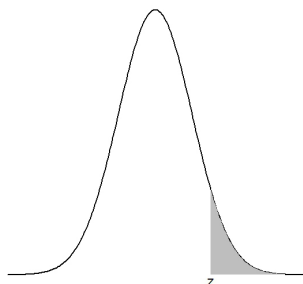This is the cutoff for how likely the data should be to fail to reject or reject the null hypothesis.

(3) Why is the test statistic $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$?

The population needs to be normally distributed so that the sample mean will have a normal sampling distribution. Specifically,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

If the population isn't normally distributed, but the sample size is large, (30 is sufficient), then the sample mean has an approximately normal sampling distribution by the central limit theorem.

Because $\overline{X}$ has a normal distribution, it is necessary to standardize any value of $\overline{X}$ before finding probabilities (see page 71). The mean of $\overline{X}$ is μ and the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n}$, but under the null hypothesis, we assume that $\mu = \mu_0$. So the standardized value of $\overline{X}$ which is used for the test statistic is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

(4) How are the p-values found?

Case 1: $H_A : \mu > \mu_0$

If μ was larger than μ_0, then it would be logical to expect to see larger sample means than if $\mu = \mu_0$. The larger the population mean, the larger the sample means should be. Remember that there are many different samples that can be chosen and each of those samples will have different sample means. For this one sided alternative hypothesis, it is necessary to find the probability that any sample would have a mean greater than or equal to the mean of the chosen sample, $P(\overline{X} \geq \overline{x})$. Since the sample mean has a normal sampling distribution, it is necessary to standardize (see page 71) the value before finding the probability, which is why the test statistic is $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ and not $\overline{X}$. So the p-value is $P(Z \geq z)$.
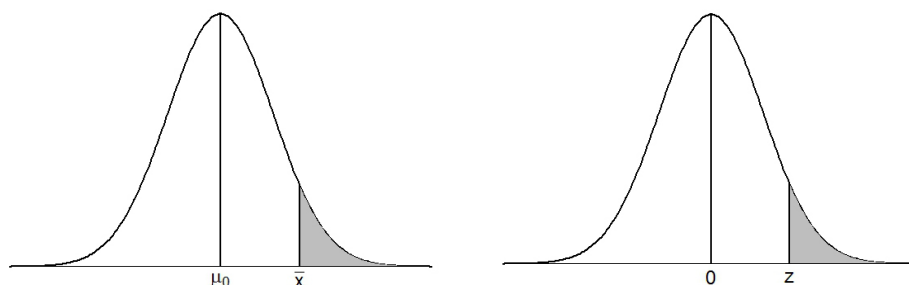
Case 2: $H_A : \mu < \mu_0$

If μ was smaller than μ_0, smaller sample means would be expected than if $\mu = \mu_0$. The p-value is the probability of a sample mean being less than or equal to the mean of the chosen sample, $P(\overline{X} \leq \overline{x})$. So the p-value is $P(Z \leq z)$.
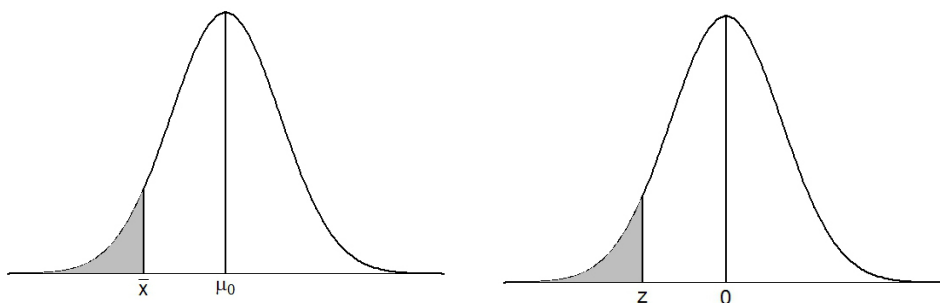
Case 3: $H_A : \mu \neq \mu_0$

If μ is not equal to μ_0, it is either larger than μ_0 or smaller than μ_0. So look for the probability that any sample mean could be further away from μ_0 than $\overline{x}$. (This is what is meant by find the probability of getting a sample mean more extreme than $\overline{x}$.) This is $P(|\overline{X} - \mu_0| \leq |\overline{x} - \mu_0|)$. Now $\overline{x}$ could be less than or greater than μ_0 which means that z could be positive or negative. To simplify calculations, assume that z is positive or find $P(Z \geq |z|) + P(Z \leq -|z|)$. Because the normal curve is symmetrical, this is the same as finding $2P(Z \geq |z|)$.
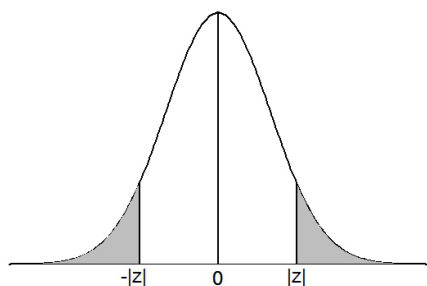
(5) How is the conclusion determined?

If the population was normally distributed, then the p-values are exact. If the population wasn't normally distributed, but the sample size was large enough, then the p-values are approximations.

The p-values describe the probability getting the test statistic (or a value more extreme than the test statistic) if the null hypothesis is true (i.e. $P(Z \geq z)$).

If the p-value is large, the data in the sample is likely to occur if the null hypothesis is true. There is not enough evidence to reject the null hypothesis.

If the p-value is small, then the data values aren't very likely to occur if the null hypothesis is true. This provides evidence that the null hypothesis isn't true. (It doesn't prove that the null hypothesis is false, it just presents evidence against the null hypothesis.) The null hypothesis is rejected.

Remember that α is the cutoff for how likely the data should be to fail to reject or reject the null hypothesis. For example, if $\alpha = 0.05$ is chosen, the null hypothesis will only be rejected if the probability to see the data in the sample is less than 5%. In summary,

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Notes. Hypothesis testing deals with probabilities. Even if an event has a really small probability it can still occur. Therefore, the p-values should be reported as well as the conclusion so that the readers can see how likely it was that the null hypothesis was true.

One Sample t-Test for the Population Mean

When can this test be used?

- The population is normally distributed or the sample size is large enough that the mean is normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 15$.)
- Do not use this test if there are outliers or the population is very skewed.
  (Skewness can be ignored if $n \geq 40$)
- The population standard deviation, σ, is **unknown**. The sample standard deviation s will be used in calculations instead.

How is this test used?

(1) State the hypotheses:
   $H_0 : \mu = \mu_0$
   $H_A : \mu \neq \mu_0$ or $H_A : \mu > \mu_0$ or $H_A : \mu < \mu_0$

(2) Compute the test statistic:

   $$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} = \frac{\text{the sample mean} - \text{the hypothesized true population mean}}{\text{the sample standard deviation} / \sqrt{\text{the sample size}}}$$

(3) Find the degrees of freedom:

   $$df = n - 1$$

   The degrees of freedom is the sample size minus 1.

(4) Find the p-value:
   The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the t distribution with $n - 1$ degrees of freedom.

   If $H_A : \mu > \mu_0$, p-value $= P\left(T \geq t\right)$
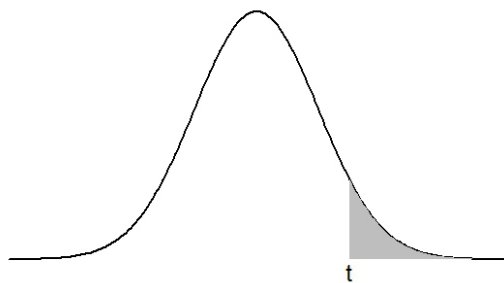
If $H_A : \mu < \mu_0$, p-value $= P\left(T \leq t\right)$
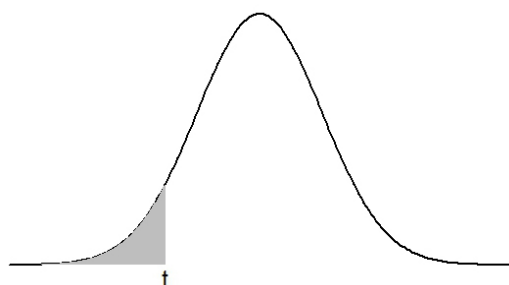
If $H_A : \mu \neq \mu_0$, p-value $= P\left(T \leq -|t| \text{ or } T \geq |t|\right) \text{ or } 2P\left(T \geq |t|\right)$

(5) State the conclusion:
Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Notes.
- If the sample size is at least 30, the t distribution can approximate be approximated with the normal distribution. So the z test could be used instead. (The standard normal table is easier to use than the t-tables.)
- In practice, σ is rarely known. To calculate σ which is the population variance, the data for the whole population would have to be known. Then the true population mean, μ, could be calculated too and there would be no need for a hypothesis test. Therefore, the t-test is almost always used.

Paired Sample z-Test for Population Mean of Paired Differences

When can this test be used?

- There are two samples that are the same size.
- The two samples are **dependent**. (This can be when the two similar subjects are matched or paired, or when subjects come from the same source, or when 2 observations are made on the same subject such as pre and post tests.)
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 30$.)
- The standard deviation of the population of pairwise differences is **known**.

Notation.

x_i	Data from sample 1	
y_i	Data from sample 2	
d_i	The pairwise difference	$d_i = x_i - y_i$
$\overline{d}$	The mean of the sample of pairwise differences	$\overline{d} = \dfrac{\sum_{i=1}^{n} d_i}{n}$
σ_d	The standard deviation of the population of pairwise differences	
n	The sample size	
μ_d	The true mean of the population of pairwise differences	
D	The **hypothesized** mean of the pairwise differences	

How is this test used?

(1) State the hypotheses:
   $H_0 : \mu_d = D$
   $H_A : \mu_d \neq D$ or $H_A : \mu_d > D$ or $H_A : \mu_d < D$

   The hypothesized difference in the means is D.

   For instance, if Tire Company A thinks that the mean distance of its tires is at least 20,000 miles more than the mean distance of Tire Company B's tires, then $D = 20,000$. If both companies put one tire on each car, then the samples would be matched or dependent. The hypotheses could be:
   $H_0 : \mu_d = 20,000$
   $H_A : \mu_d > 20,000$

   Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
   $H_0 : \mu_d = 0$
   $H_A : \mu_d \neq 0$ or $H_A : \mu_d > 0$ or $H_A : \mu_d < 0$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$z = \frac{\overline{d} - D}{\frac{\sigma_d}{\sqrt{n}}} = \frac{\text{sample mean of pairwise differences} - \text{hypothesized mean of pairwise differences}}{\text{standard deviation of population of pairwise differences}/\sqrt{\text{sample size}}}$$

(4) Find the p-value:
The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the standard normal curve.

If $H_A : \mu_d > D$, p-value $= P\left(Z \geq z\right)$

If $H_A : \mu_d < D$, p-value $= P\left(Z \leq z\right)$

If $H_A : \mu_d \neq D$, p-value $= P\left(Z \leq -|z| \text{ or } Z \geq |z|\right)$ or $2P\left(Z \geq |z|\right)$

(5) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note. Once all the pairwise differences, d_i, have been found, they compose a sample. Then the one sample z-test for the population mean can be used on the sample of the pairwise differences. It is equivalent to the test above, but without the new notation.

Paired Sample t-Test for Population Mean of Pairwise Differences

When can this test be used?
- There are two samples that are the same size.
- The two samples are **dependent**. (This can be when the two similar subjects are matched or paired, or when subjects come from the same source, or when 2 observations are made on the same subject such as pre and post tests.)
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 15$.)
- The standard deviation of the population's difference is **unknown**.

Notation.

x_i	Data from sample 1	
y_i	Data from sample 2	
d_i	The pairwise difference	$d_i = x_i - y_i$
$\overline{d}$	The mean of the sample of pairwise differences	$\overline{d} = \dfrac{\sum_{i=1}^{n} d_i}{n}$
s_d	The standard deviation of the sample pairwise differences	$s_d = \sqrt{\dfrac{\sum_{i=1}^{n} \left(d_i - \overline{d}\right)^2}{n-1}}$
n	The sample size	
μ_d	The true mean of the population of pairwise differences	
D	The **hypothesized** mean of the pairwise differences	

How is this test used?

(1) State the hypotheses:
$H_0 : \mu_d = D$
$H_A : \mu_d \neq D$ or $H_A : \mu_d > D$ or $H_A : \mu_d < D$

The hypothesized difference in the means is D.

For instance, if Tire Company A thinks that the mean distance of its tires is at least 20,000 miles more than the mean distance of Tire Company B's tires, then $D = 20,000$. If both companies put one tire on each car, then the samples would be matched or dependent. The hypotheses could be:
$H_0 : \mu_d = 20,000$
$H_A : \mu_d > 20,000$

Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
$H_0 : \mu_d = 0$
$H_A : \mu_d \neq 0$ or $H_A : \mu_d > 0$ or $H_A : \mu_d < 0$

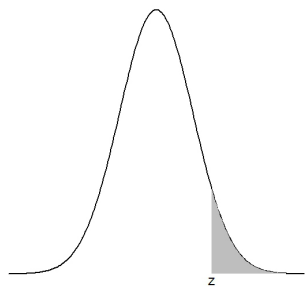(2) Pick a significance level, α.

(3) Compute the test statistic:

$$t = \dfrac{\overline{d} - D}{\dfrac{s_d}{\sqrt{n}}} = \dfrac{\text{sample mean of pairwise differences} - \text{hypothesized mean of pairwise differences}}{\text{sample standard deviation of pairwise differences}/\sqrt{\text{sample size}}}$$
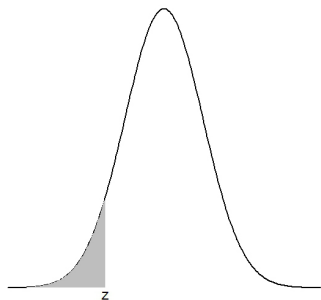
(4) Find the degrees of freedom:

$$df = n - 1$$

The degrees of freedom is the sample size minus 1.

(5) Find the p-value:
The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the t distribution with $n - 1$ degrees of freedom.

If $H_A : \mu_d > D$, p-value $= P(T \geq t)$

If $H_A : \mu_d < D$, p-value $= P(T \leq t)$

If $H_A : \mu_d \neq D$, p-value $= P(T \leq -|t| \text{ or } T \geq |t|) \text{ or } 2P(T \geq |t|)$

(6) State the conclusion:
Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note. Once all the pairwise differences, d_i, have been found, they compose a sample. Then the one sample t-test for the population mean can be used on the sample of the pairwise differences. It is equivalent to the test above, but without the new notation.

Two Sample z-Test for Difference of Population Means

When can this test be used?
- There are two samples from two populations. (The samples can be different sizes.)
- The two samples are **independent**.
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
  (A rule of thumb is each sample size is $n \geq 30$.)
- Both population standard deviations, σ_x and σ_y, are **known**.

Notation.

Population	Data	Mean	Standard Deviation	Sample Size	Sample Mean
1	x_i	μ_x	σ_x	n	$\overline{x}$
2	y_i	μ_y	σ_y	m	$\overline{y}$

How is this test used?
(1) State the hypotheses:
$H_0 : \mu_x - \mu_y = D$
$H_A : \mu_x - \mu_y \neq D$ or $H_A : \mu_x - \mu_y > D$ or $H_A : \mu_x - \mu_y < D$
The hypothesized difference in the means is D.

For instance, if Tire Company A thinks that the mean distance of its tires is at least 20,000 miles more than the mean distance of Tire Company B's tires, then $D = 20,000$. If Company A and Company B put their tires on different randomly chosen cars, the samples would be independent. The hypotheses could be:
$H_0 : \mu_x - \mu_y = 20,000$
$H_A : \mu_x - \mu_y > 20,000$

Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
$H_0 : \mu_x = \mu_y$
$H_A : \mu_x \neq \mu_y$ or $H_A : \mu_x > \mu_y$ or $H_A : \mu_x < \mu_y$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$z = \frac{(\overline{x} - \overline{y}) - D}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} = \frac{\text{difference in sample means} - \text{hypothesized difference in population means}}{\sqrt{\dfrac{\text{population 1 variance}}{\text{sample 1 size}} + \dfrac{\text{population 2 variance}}{\text{sample 2 size}}}}$$
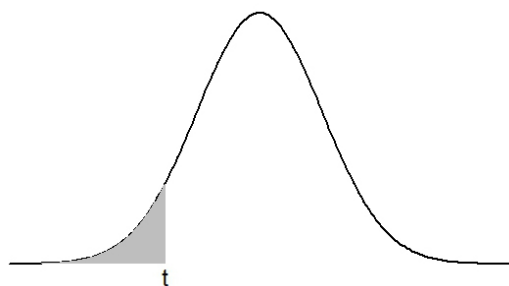
(4) Find the p-value:

The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the standard normal curve.

If $H_A : \mu_x - \mu_y > D$, p-value $= P(Z \geq z)$

If $H_A : \mu_x - \mu_y < D$, p-value $= P(Z \leq z)$

If $H_A : \mu_x - \mu_y \neq D$, p-value $= P(Z \leq -|z| \text{ or } Z \geq |z|)$ or $2P(Z \geq |z|)$

(5) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Two Sample t-Test for Difference of the Population Means (Equal Variances)

When can this test be used?
- There are two samples from two populations. (The samples can be different sizes.)
- The two samples are **independent**.
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 15$.)
- Both population standard deviations, σ_x and σ_y, are **unknown**, but are assumed to be equal.

Notation.

Population	Data	Mean	Standard Deviation	Sample Size	Sample Mean	Sample Standard Deviation
1	x_i	μ_x	σ_x	n	$\overline{x}$	s_x
2	y_i	μ_y	σ_y	m	$\overline{y}$	s_y

How is this test used?

(1) State the hypotheses:
$H_0 : \mu_x - \mu_y = D$
$H_A : \mu_x - \mu_y \neq D$ or $H_A : \mu_x - \mu_y > D$ or $H_A : \mu_x - \mu_y < D$
The hypothesized difference in the means is D.

Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
$H_0 : \mu_x = \mu_y$
$H_A : \mu_x \neq \mu_y$ or $H_A : \mu_x > \mu_y$ or $H_A : \mu_x < \mu_y$

(2) Pick a significance level, α.

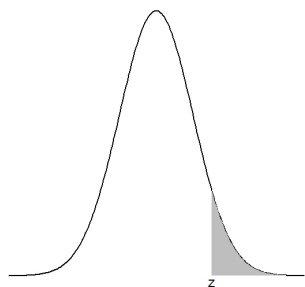(3) Compute the test statistic:

$$t = \frac{(\overline{x} - \overline{y}) - D}{\sqrt{s_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} = \frac{\text{difference in sample means} - \text{hypothesized difference in population means}}{\sqrt{(\text{pooled variance estimate}) \left(\frac{1}{\text{sample 1 size}} + \frac{1}{\text{sample 2 size}} \right)}}$$

where
$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2}$$
.

(4) Find the degrees of freedom:

$$df = n + m - 2$$

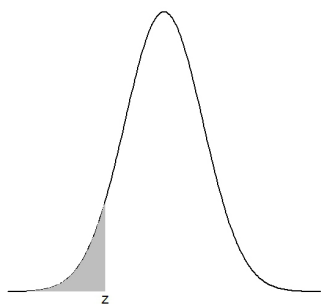The degrees of freedom is the sum of the sample sizes minus 2.

(5) Find the p-value:

The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the t distribution with $n+m-2$ degrees of freedom.
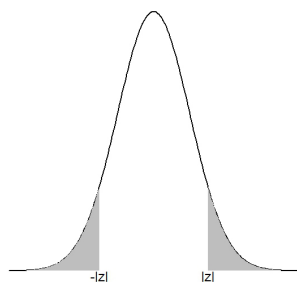
If $H_A : \mu_x - \mu_y > D$, p-value $= P\left(T \geq t\right)$

If $H_A : \mu_x - \mu_y < D$, p-value $= P\left(T \leq t\right)$

If $H_A : \mu_x - \mu_y \neq D$, p-value $= P\left(T \leq -|t| \text{ or } T \geq |t|\right)$ or $2P\left(T \geq |t|\right)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p-value < \alpha$	reject H_0
$p-value > \alpha$	fail to reject H_0

Two Sample t-Test for Difference of the Population Means (Unequal Variances)

When can this test be used?
- There are two samples from two populations. (The samples can be different sizes.)
- The two samples are **independent**.
- Both populations are normally distributed or both sample sizes are large enough that the means are normally distributed.
  (A rule of thumb is that the sample size is large enough if $n \geq 15$.)
- Both population standard deviations, σ_x and σ_y, are **unknown**, but are assumed to be not equal.

Notation.

Population	Data	Mean	Standard Deviation	Sample Size	Sample Mean	Sample Standard Deviation
1	x_i	μ_x	σ_x	n	$\overline{x}$	s_x
2	y_i	μ_y	σ_y	m	$\overline{y}$	s_y

How is this test used?

(1) State the hypotheses:
$H_0 : \mu_x - \mu_y = D$
$H_A : \mu_x - \mu_y \neq D$ or $H_A : \mu_x - \mu_y > D$ or $H_A : \mu_x - \mu_y < D$
The hypothesized difference in the means is D.

Usually, the hypothesized difference is $D = 0$. In this case the hypotheses simplify to:
$H_0 : \mu_x = \mu_y$
$H_A : \mu_x \neq \mu_y$ or $H_A : \mu_x > \mu_y$ or $H_A : \mu_x < \mu_y$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$t = \frac{(\overline{x} - \overline{y}) - D}{\sqrt{\dfrac{s_x^2}{n} + \dfrac{s_y^2}{m}}} = \frac{\text{difference in sample means} - \text{hypothesized difference in population means}}{\sqrt{\left(\dfrac{\text{sample 1 variance}}{\text{sample 1 size}} + \dfrac{\text{sample 2 variance}}{\text{sample 2 size}}\right)}}$$

(4) Find the degrees of freedom:

$$df = \frac{\left(\dfrac{s_x^2}{n} + \dfrac{s_y^2}{m}\right)^2}{\dfrac{s_x^4}{n^2(n-1)} + \dfrac{s_y^4}{m^2(m-1)}}$$

This is very unlikely to be a whole number, so it can be rounded down to the nearest integer.

If the test is done without statistical software, the degrees of freedom can be very tedious to calculate. So the smaller of the values $n - 1$ and $m - 1$ can be used instead. This will not be as accurate, but it is a conservative choice.
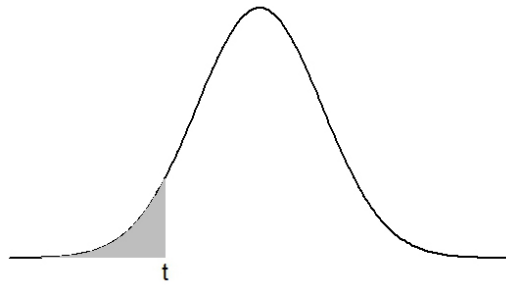
(5) Find the p-value:

The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the t distribution with the degrees of freedom described above.
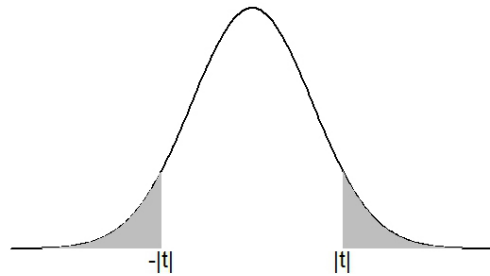
If $H_A : \mu_x - \mu_y > D$, p-value $= P(T \geq t)$

If $H_A : \mu_x - \mu_y < D$, p-value $= P(T \leq t)$

If $H_A : \mu_x - \mu_y \neq D$, p-value $= P(T \leq -|t| \text{ or } T \geq |t|)$ or $2P(T \geq |t|)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note: The unequal variances test can be used even if the variances are equal. If the variances are equal, it is not as powerful as the pooled variance test, but it is the safe option.

One Sample z-Test for the Population Proportion

When can this test be used?

- Data comes from a binomial experiment.

- The sample size is large.
  (A rule of thumb is that the sample size is large enough if the number of expected successes $np_0 \geq 10$ and the number of expected failures $n(1 - p_0) \geq 10$.)

- The population is at least 20 times the size of the sample.

Notation.

X	The number of successes in the sample	
n	The sample size	
p	The population proportion	
$\hat{p}$	The sample proportion	$\hat{p} = \dfrac{X}{n}$
p_0	The hypothesized population proportion	

How is this test used?

(1) State the hypotheses:
$H_0 : p = p_0$
$H_A : p \neq p_0$ or $H_A : p > p_0$ or $H_A : p < p_0$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{\text{the sample proportion} - \text{the hypothesized true population proportion}}{\sqrt{\dfrac{\text{hypothesized proportion} \, (1\text{-hypothesized proportion})}{\text{the sample size}}}}$$

(4) Find the p-value:

The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the standard normal curve.

If $H_A : p > p_0$, p-value $= P(Z \geq z)$

If $H_A : p < p_0$, p-value $= P(Z \leq z)$

If $H_A : p \neq p_0$, p-value $= P(Z \leq -|z| \text{ or } Z \geq |z|) \text{ or } 2P(Z \geq |z|)$
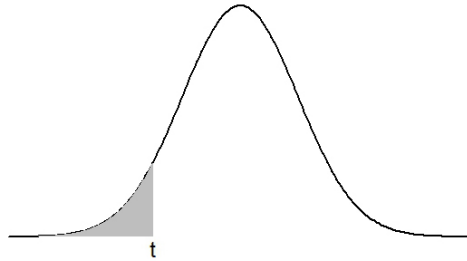
(5) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Two Sample z-Test for Population Proportions

When can this test be used?

- The data comes from a binomial experiment.

- Both sample sizes are large.
  (A rule of thumb is that the sample sizes are large enough the number of success and the number of failures is at least 5 for both samples)

Notation.

Population	Count of Successes	Sample Size	Population Proportion	Sample Proportion
1	X	n	p_x	$\hat{p}_x = \dfrac{X}{n}$
2	Y	m	p_y	$\hat{p}_y = \dfrac{Y}{m}$

How is this test used?

(1) State the hypotheses:
$H_0 : p = p_0$
$H_A : p \neq p_0$ or $H_A : p > p_0$ or $H_A : p < p_0$

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}(1 - \hat{p})\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

where $\hat{p} = \dfrac{X + Y}{n + m}$ is the overall proportion of successes.

(4) Find the p-value:
The p-value depends on which alternative hypothesis is being used. The p-value is the probability or the area in the tail(s) of the standard normal curve.

If $H_A : p_x > p_y$, p-value $= P(Z \geq z)$
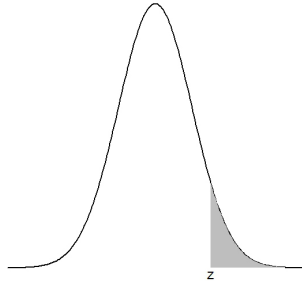
If $H_A : p_x < p_y$, p-value $= P(Z \leq z)$
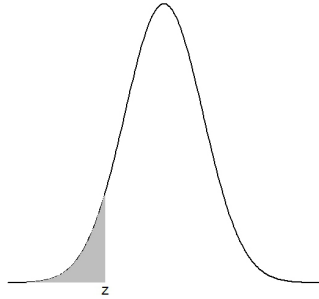
If $H_A : p_x \neq p_y$, p-value $= P(Z \leq -|z| \text{ or } Z \geq |z|)$ or $2P(Z \geq |z|)$

(5) State the conclusion:
Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Chi-Square Test for the Population Variance

When can this test be used?

- The population is normally distributed.
- The population variance, σ^2 is **unknown**.

How is this test used?

(1) Pick a significance level, α.

(2) State the hypotheses:
$H_0 : \sigma^2 = \sigma_0^2$
$H_A : \sigma^2 \neq \sigma_0^2$ or $H_A : \sigma^2 > \sigma_0^2$ or $H_A : \sigma^2 < \sigma_0^2$

(3) Compute the test statistic:

$$X^2 = \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{(\text{sample size -1})(\text{the sample variance})}{\text{the hypothesized population variance}}$$

(4) Find your degrees of freedom.

$$df = n - 1$$

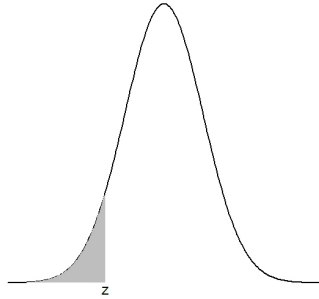The degrees of freedom is the sample size minus 1.

(5) Find the p-value:
The p-value depends on which alternative hypothesis is being used. The p-value is found by the probability or the area in the tail(s) of the χ^2 distribution with $n - 1$ degrees of freedom.
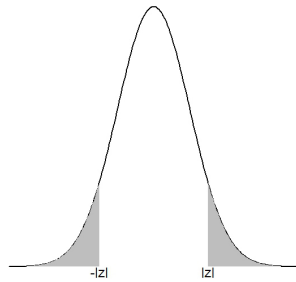
Warning! The χ^2 distribution is not symmetric. The probability of each tail must be calculated separately.

If $H_A : \sigma^2 > \sigma_0^2$, p-value $= P\left(\chi^2 \geq X^2\right)$
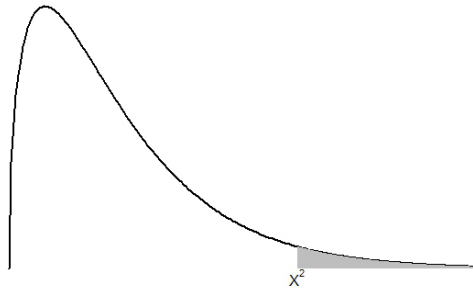
If $H_A : \sigma^2 < \sigma_0^2$, p-value $= P\left(\chi^2 \leq X^2\right)$
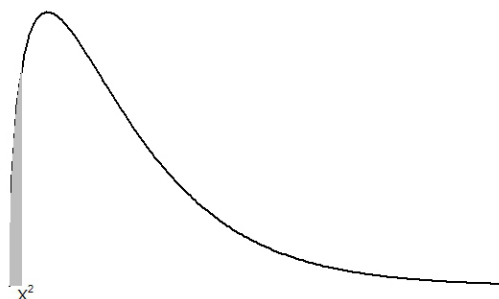
If $H_A : \sigma^2 \neq \sigma_0^2$, p-value $= 2\left(\chi^2 \leq X^2\right)$ if the test statistic X^2 is less than the median or $2P\left(\chi^2 \geq X^2\right)$ if X^2 is greater than the median.

or
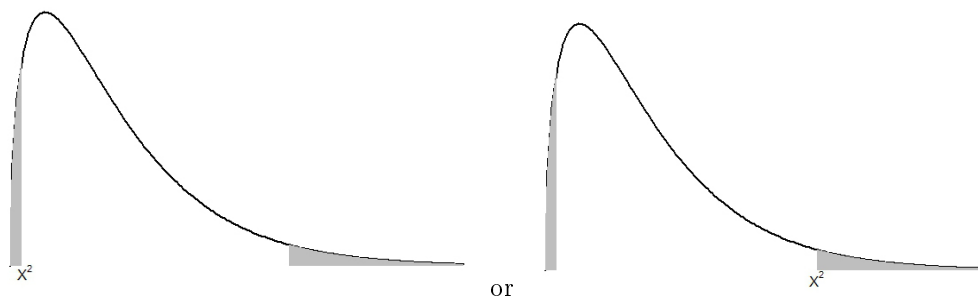
Hint: The median is the number, a, such that $P(\chi^2 \geq a) = .5$. If the p-value is greater than one, it is a sign that X^2 is on the other side of the median.

(6) State the conclusion:
Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Two Sample F Test for Population Variances

When can this test be used?

- There are two samples from two populations. (The samples can be different sizes.)
- The two samples are **independent**.
- Both populations are normally distributed.
  Note: The populations must be normally distributed even if the sample sizes are large!
- Both population variances, σ_x^2 and σ_y^2, are **unknown**.

Notation.

Population	Data	Population Variance	Sample Size	Sample Variance
1	x_i	σ_x^2	n	s_x^2
2	y_i	σ_y^2	m	s_y^2

How do I use it?

(1) State the hypotheses:
$H_0 : \sigma_x = \sigma_y$
$H_A : \sigma_x \neq \sigma_y$ or $H_A : \sigma_x > \sigma_y$ or $H_A : \sigma_y > \sigma_x$

This tests the hypothesis that the two populations have the same standard deviation.

(2) Pick a significance level, α.

(3) Compute the test statistic:

$$F = \frac{s_x^2}{s_y^2} = \frac{\text{the sample variance of population 1}}{\text{the sample variance of population 2}}$$

To simplify calculations, let the sample that has the larger variance be sample 1. Then $F \geq 1$ which will always be in the right tail of the distribution. Then

$$F = \frac{\text{larger } s^2}{\text{smaller } s^2}$$

(4) Find the 2 degrees of freedom.

$df = n - 1$ and $m - 1$.

That is right. The F distribution has 2 degrees of freedom called the numerator and denominator degrees of freedom.

df_n is the numerator degrees of freedom. It is associated with the sample whose sample variance is in the numerator of the F statistic. (If the sample variance of population 1 is the numerator of the F statistic, then $n - 1$ is the numerator degrees of freedom.)

df_d is the denominator degrees of freedom. It is associated with the sample whose sample variance is in the denominator.

The F distribution is written as F_{df_n, df_d}. Notice that the numerator degrees of freedom comes first. Don't switch the order of the degrees of freedom!

(5) Find the p-value:

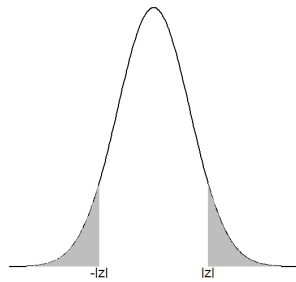The p-value depends on which alternative hypothesis is being used. The p-value is found by the probability or the area in the right hand tail of the F distribution with df_n, df_d degrees of freedom.

If $H_A : \sigma_x > \sigma_y$ or $H_A : \sigma_y > \sigma_x$, p-value= $P(\mathcal{F} \geq F)$ (This works if the sample with the bigger variance is the numerator.)
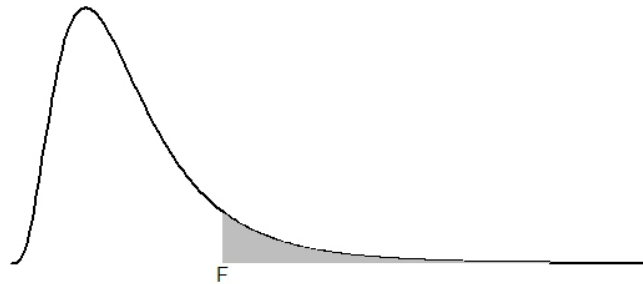
If $H_A : \sigma_x \neq \sigma_y$, p-value= $2P(\mathcal{F} \geq F)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note. This test is **very** sensitive to non normal populations. If the populations aren't normal the results won't be valid. Therefore this test isn't used very often.

Chi-Square Test for Goodness of Fit

When can this test be used?

- The data is from a multinomial experiment.
- All expected values are at least 5. (Another rule of thumb is that there are more than 4 groups, the average of the expected values is at least 5, and the smallest expected value is at least 1.)

This test is used to see if the data 'fits' a distribution. The data will be divided into k groups. (Asking if the data fits the distribution, is the same as asking if we can predict how many items will be in each group.)

How is this test used?

(1) State the hypotheses:

H_0 : the data fits the proposed distribution
H_A : the data does not fit the proposed distribution

or

H_0 : the probability of each group is the expected probability.
H_A : the probability of at least one group does not match the expected probability.

(2) Pick a significance level, α.

(3) Compute the test statistic: (This is actually Pearson's Approximation)

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \sum_{\text{all cells}} \frac{(Observed - Expected)^2}{Expected}$$

n is the sample size.

O_i is the observed frequency in each group.

p_i is the theoretical or expected probability for each group.

E_i is the expected frequency in each group. Find the expected frequency by multiplying the sample size by the theoretical or expected probability for each group.

$E_i = np_i$

(4) Find the degrees of freedom.

$df = k - 1 - m$.

m is the number of parameters being estimated and k is the number of groups or cells in the contingency table. If the parameters are all known, or constant probabilities have been guessed for each cell, there is no need to estimate any parameters.

For example, if the proposed distribution is the normal distribution, it is necessary to use $\overline{x}$ and s as estimates for μ and σ so $m = 2$. However, if it is proposed that for 4 groups, the probability of the groups are 0.25, 0.5, 0.25, and 0.25 respectively, then all the probabilities have been stated and there is no need to estimate any parameters.

(5) Find the p-value:

The p-value is the probability or the area in the right tail of the χ^2_{k-1-m} distribution.

Warning! The χ^2 distribution is not symmetric.

p-value $= P\left(\chi^2 \geq X^2\right)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note. If the observed and expected values are not close, then X^2 will be large resulting in a small p-value. This is an indication that the model does not fit the data.

Chi-Square Test for Independence

When is this test used?
- The data is multinomial data in a contingency table or two way cross-classification table.
- All expected values are at least 5. (Another rule of thumb is that there are more than 4 cells, the average of the expected values is at least 5, and the smallest expected value is at least 1.)
- The cells have counts or frequencies. It doesn't work if the data is percentages or relative frequencies!
- The data comes from **one** population.

This test is used to see if two variables are independent.

Each cell will have a frequency or count. It is necessary to find row, column, and grand totals. One variable will be the rows and one will be the columns.

How is this test used?

(1) State the hypotheses:

H_0 : the two variables are independent
H_A : the two variables are dependent

or

H_0 : the probability of a cell is (row probability)(column probability)
H_A : the probability of at least one cell is not (row probability)(column probability).

(2) Pick a significance level, α.

(3) Compute the test statistic: (This is actually Pearson's Approximation)

$$X^2 = \sum_{\text{all cells}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{\text{all cells}} \frac{(Observed - Expected)^2}{Expected}$$

n is the sample size or grand total.

O_{ij} is the observed frequency or count in the cell that is in row i and column j.

r_i is the row total for row i.

c_j is the column total for row j.

E_{ij} is the expected frequency in each cell. Find the expected frequency by multiplying the row total by the column total and dividing by the grand total (the sample size).

$$E_{ij} = \frac{r_i c_j}{n} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

(So the expected count in each cell can be found by the total of that row multiplied by the total of that column divided by the sample size.)

(4) Find the degrees of freedom.

$df = (r-1)(c-1)$ or $df = (\text{number of rows} - 1)(\text{number of columns} - 1)$.

(5) Find the p-value:

The p-value is the probability or the area in the right tail of the $\chi^2_{(r-1)(c-1)}$ distribution.

Warning! The χ^2 distribution is not symmetric.

p-value $= P\left(\chi^2 \geq X^2\right)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p - value < \alpha$	reject H_0
$p - value > \alpha$	fail to reject H_0

Note. If the observed and expected values are not close, then X^2 will be large resulting in a small p-value. This is an indication that the two variables or categories are not independent.

Chi-Square Test for Homogeneity

When is this test used?
- The data is multinomial data in a contingency table or two way cross-classification table.
- All expected values are at least 5. (Another rule of thumb is that there are more than 4 cells, the average of the expected values is at least 5, and the smallest expected value is at least 1.)
- The cells have counts or frequencies. It doesn't work if the data is percentages or relative frequencies!
- Either the row totals or column totals are fixed.
- The data comes from multiple samples which are independent.

This test is used to see if the different samples come from populations with the same distribution.

Each cell will have a frequency or count. It is necessary to find row, column, and grand totals. There will be i categories and j samples or distributions. The notation below assumes the categories are the rows and the samples are the columns. If they are switched, all the calculations and results will be the same.

How is this test used?

(1) State the hypotheses:

H_0 : the J distributions are the same

H_A : the J distributions are not the same

or

H_0 : the probability of the ith category is the same for each jth distribution. (This means that if a category is chosen to look at, all the samples should have the same relative frequencies.)

H_A : at least one distribution or sample does not have the same probability as the other distributions for one category.

(2) Pick a significance level, α.

(3) Compute the test statistic: (This is actually Pearson's Approximation)

$$X^2 = \sum_{\text{all cells}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{\text{all cells}} \frac{(Observed - Expected)^2}{Expected}$$

n is the sample size or grand total.

O_{ij} is the observed frequency or count in the cell that is in row i and column j.

r_i is the row total for row i.

c_j is the column total for row j.

E_{ij} is the expected frequency in each cell. Find the expected frequency by multiplying the row total by the column total and dividing by the grand total (the sample size).

$$E_{ij} = \frac{r_i c_j}{n} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

(So the expected count in each cell can be found by the total of that row multiplied by the total of that column divided by the sample size.)
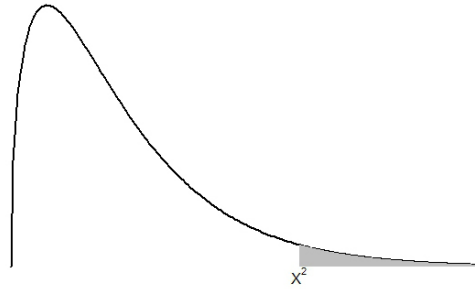
(4) Find the degrees of freedom.

$df = (r - 1)(c - 1)$ or $df = (\text{number of rows} - 1)(\text{number of columns} - 1)$.

(5) Find the p-value:

The p-value is the probability or the area in the right tail of the $\chi^2_{(r-1)(c-1)}$ distribution.

Warning! The χ^2 distribution is not symmetric.

p-value $= P\left(\chi^2 \geq X^2\right)$

(6) State the conclusion:

Once the p-value is known, compare it to α, the significance level.

If the p-value is smaller than α, the observed outcome wasn't very likely given that the null hypothesis is true. So reject the null hypothesis in favor of the alternative hypothesis.

If the p-value is greater than α then the observed outcome was likely enough that it is reasonable to assume that the null hypothesis is true so fail to reject the null hypothesis.

$p-value < \alpha$	reject H_0
$p-value > \alpha$	fail to reject H_0

Note. If the observed and expected values are not close, then X^2 will be large resulting in a small p-value. This is an indication that the J distributions are not the same.

Table of Hypothesis Tests

Test	When to use this test	Calculated Test Statistic	Distribution of test statistic given H_0	P-value
One sample z test for the population mean	• Test is for μ • Population is normally distributed or $n \geq 30$ • σ is known	$z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$	$Z \sim N(0,1)$	$H_A : \mu > \mu_0$, the p-value is $P(Z \geq z)$ $H_A : \mu < \mu_0$, the p-value is $P(Z \leq z)$ $H_A : \mu \neq \mu_0$, the p-value is $2P(Z \geq |z|)$
One sample t test for the population mean	• Test is for μ • Population is normally distributed or $n \geq 15$ • σ is not known	$t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$	$T \sim t_{n-1}$	$H_A : \mu > \mu_0$, the p-value is $P(T \geq t)$ $H_A : \mu < \mu_0$, the p-value is $P(T \leq t)$ $H_A : \mu \neq \mu_0$, the p-value is $2P(T \geq |t|)$
Paired sample z test for population mean of pairwise differences	• The two samples are dependent • The two sample sizes are equal • Both populations are normally distributed or $n \geq 30$ • σ_d is known	$z = \dfrac{\overline{d} - D}{\sigma_d/\sqrt{n}}$	$Z \sim N(0,1)$	$H_A : \mu_d > D$, p-value is $P(Z \geq z)$ $H_A : \mu_d < D$, p-value is $P(Z \leq z)$ $H_A : \mu_d \neq D$, p-value is $2P(Z \geq |z|)$
Paired sample t test for population mean of pairwise differences	• The two samples are dependent • The two sample sizes are equal • Both populations are normally distributed or $n \geq 15$ • σ_d is not known	$t = \dfrac{\overline{d} - D}{s_d/\sqrt{n}}$	$T \sim t_{n-1}$	$H_A : \mu_d > D$, p-value is $P(T \geq t)$ $H_A : \mu_d < D$, p-value is $P(T \leq t)$ $H_A : \mu_d \neq D$, p-value is $2P(T \geq |t|)$

Test	When to use this test	Calculated Test Statistic	Distribution of test statistic given H_0	P-value
Two sample z test for difference of population means	• The two samples are independent • Both populations are normally distributed or $n \geq 30$ and $m \geq 30$. • σ_x and σ_y are known	$z = \dfrac{(\overline{x} - \overline{y}) - D}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}}$	$Z \sim N(0,1)$	$H_A : \mu_x - \mu_y > D$, p-value is $P(Z \geq z)$ $H_A : \mu_x - \mu_y < D$, p-value is $P(Z \leq z)$ $H_A : \mu_x - \mu_y \neq D$, p-value is $2P(Z \geq |z|)$
Two sample t test for difference of population means (equal variances)	• The two samples are independent • Both populations are normally distributed or $n \geq 15$ and $m \geq 15$. • σ_x and σ_y are not known but are assumed to be equal	$t = \dfrac{(\overline{x} - \overline{y}) - D}{\sqrt{s_p^2 \left(\dfrac{1}{n} + \dfrac{1}{m} \right)}}$ $s_p^2 = \dfrac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$	$T \sim t_{n+m-2}$	$H_A : \mu_x - \mu_y > D$, p-value is $P(T \geq t)$ $H_A : \mu_x - \mu_y < D$, p-value is $P(T \leq t)$ $H_A : \mu_x - \mu_y \neq D$, p-value is $2P(T \geq |t|)$
Two sample t test for difference of population means (unequal variances)	• The two samples are independent • Both populations are normally distributed or $n \geq 15$ and $m \geq 15$. • σ_x and σ_y are not known but are assumed to be not equal	$t = \dfrac{(\overline{x} - \overline{y}) - D}{\sqrt{\dfrac{s_x^2}{n} + \dfrac{s_y^2}{m}}}$	$T \sim t_{df}$ $df = \dfrac{\left(\dfrac{s_x^2}{n} + \dfrac{s_y^2}{m} \right)^2}{\dfrac{s_x^4}{n^2(n-1)} + \dfrac{s_y^4}{m^2(m-1)}}$ (or the smaller of $n-1$ and $m-1$)	$H_A : \mu_x - \mu_y > D$, p-value is $P(T \geq t)$ $H_A : \mu_x - \mu_y < D$, p-value is $P(T \leq t)$ $H_A : \mu_x - \mu_y \neq D$, p-value is $2P(T \geq |t|)$

Test	When to use this test	Calculated Test Statistic	Distribution of test statistic given H_0	P-value
One sample z test for a population proportion	• $np_0 \geq 10$ and $n(1-p_0) \geq 10$ • The population is at least 20 times the size of the sample	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ $\hat{p} = \dfrac{X}{n}$	$Z \sim N(0,1)$	$H_A : p > p_0$, p-value is $P(Z \geq z)$ $H_A : p < p_0$, p-value is $P(Z \leq z)$ $H_A : p \neq p_0$, p-value is $2P(Z \geq |z|)$
Two sample z test for population proportions	• There are at least 5 success and 5 failures for both samples.	$z = \dfrac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$ $\hat{p} = \dfrac{X+Y}{n+m}$	$Z \sim N(0,1)$	$H_A : p_x > p_y$, p-value is $P(Z \geq z)$ $H_A : p_x < p_y$, p-value is $P(Z \leq z)$ $H_A : p_x \neq p_y$, p-value is $2P(Z \geq |z|)$
Chi square test for population variance	• The population is normally distributed • σ^2 is unknown	$X^2 = \dfrac{(n-1)\,s^2}{\sigma_0^2}$	$\chi^2 \sim \chi_{n-1}^2$	$H_A : \sigma^2 > \sigma_0^2$, p-value is $P(\chi^2 \geq X^2)$ $H_A : \sigma^2 < \sigma_0^2$, p-value is $P(\chi^2 \leq X^2)$ $H_A : \sigma^2 \neq \sigma_0^2$, p-value is $2P(\chi^2 \leq X^2)$ if X^2 is less than the median or $2P(\chi^2 \geq X^2)$ if X^2 is greater than the median
Two sample F test for population variances	• There are two independent samples • Both populations are normally distributed • σ_x^2 and σ_y^2 are unknown	$F = \dfrac{\text{larger } s^2}{\text{smaller } s^2}$	$\mathcal{F} \sim F_{n-1,m-1}$ where n is the sample size of the sample with larger variance and m is the other sample size	$H_A : \sigma_x > \sigma_y$, p-value is $P(\mathcal{F} \geq F)$ $H_A : \sigma_x \neq \sigma_y$, p-value is $2P(\mathcal{F} \geq F)$ (σ_x is the standard deviation from the population that had a larger sample variance.)

Test	When to use this test	Calculated Test Statistic	Distribution of test statistic given H_0	P-value
Chi square test for goodness of fit	• The data is from a multinomial experiment • All expected values are at least 5	$X^2 = \sum_{i=1}^{n} \dfrac{(O_i - E_i)^2}{E_i}$ $E_i = np_i$	$\chi^2 \sim \chi^2_{k-1-m}$ where k is the number of groups and m is the number of parameters estimated	p-value is $P\left(\chi^2 \geq X^2\right)$
Chi square test for independence	• The data is from a multinomial experiment • All expected values are at least 5 • The data comes from one population	$X^2 = \sum_{\text{all cells}} \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$ $E_{ij} = \dfrac{r_i c_j}{n}$ $r_i = $ row total $c_j = $ column total	$\chi^2 \sim \chi^2_{(r-1)(c-1)}$ where r is the number of rows and c is the number of columns	p-value is $P\left(\chi^2 \geq X^2\right)$
Chi square test for homogeneity	• The data is from a multinomial experiment • All expected values are at least 5 • The data comes from multiple independent samples	$X^2 = \sum_{\text{all cells}} \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$ $E_{ij} = \dfrac{r_i c_j}{n}$ $r_i = $ row total $c_j = $ column total	$\chi^2 \sim \chi^2_{(r-1)(c-1)}$ where r is the number of rows and c is the number of columns	p-value is $P\left(\chi^2 \geq X^2\right)$

Hypothesis Testing Flow Chart

```
                              ┌──────────────┐
                              │  One Sample  │
                              └──────────────┘
```

One Sample

Mean of Population

Population Proportion

Population Variance

Population is normally distributed or the sample size is large

The sample size is large

The population is normally distributed

The population standard deviation is known

The population standard deviation is unknown

Use the one sample z-test for a population proportion

The population variance is unknown.

Use the one sample z-test for a population mean

Use the one sample t-test for a population mean

Use the chi-square test for population variance

```
                      ┌─────────────────────┐
                      │   Paired Samples    │
                      └─────────────────────┘
                                 │
                   ┌─────────────────────────────┐
                   │   Mean of Population of      │
                   │    Pairwise Differences      │
                   └─────────────────────────────┘
                                 │
                   ┌─────────────────────────────┐
                   │      Population is           │
                   │  normally distributed or     │
                   │   the sample size is large   │
                   └─────────────────────────────┘
                          /              \
          ┌─────────────────────┐   ┌─────────────────────┐
          │  The population of  │   │  The population of  │
          │ pairwise differences'│  │ pairwise differences'│
          │ standard deviation is│  │ standard deviation is│
          │        known        │   │       unknown       │
          └─────────────────────┘   └─────────────────────┘
                    │                          │
          ┌─────────────────────┐   ┌─────────────────────┐
          │   Use the paired    │   │   Use the paired    │
          │   sample z-test     │   │   sample t-test     │
          └─────────────────────┘   └─────────────────────┘
```

```
                        ┌─────────────────────┐
                        │ Independent Samples │
                        └─────────────────────┘
          ┌───────────────────┼───────────────────┐
┌──────────────────┐  ┌──────────────────┐  ┌──────────────────┐
│   Difference of  │  │    Population    │  │    Population    │
│ Population Means │  │    Proportion    │  │     Variance     │
└──────────────────┘  └──────────────────┘  └──────────────────┘
```

Difference of Population Means

Both populations are normally distributed or both sample sizes are large

- Both population variances are known
  - Use the two sample z-test for difference of population means

- Both population variances are unknown, but assumed to be equal
  - Use the two sample t-test for difference of population means (equal variances)

- Both population variances are unknown, but assumed to be not equal
  - Use the two sample t-test for difference of population means (unequal variances)

Population Proportion

Sample sizes are large

Use the two sample z-test for population proportions

Population Variance

Both populations are normally distributed

Both population variances are unknown

Use the two sample F test for population variances

Symbols

Symbol	Meaning	Formula
π	The number π	$\pi = 3.141592...$
e	The number e	$e = 2.718281...$
$x!$	Factorial	$x! = x\,(x-1)\,(x-2)\cdots(3)\,(2)\,(1)$
N	The population size	
n	The sample size	
x_i	The value of a single piece of data in the sample or population	
μ	The population mean	$\mu = \dfrac{\sum\limits_{i=1}^{N} x_i}{N}$
$\overline{x}$	The sample mean	$\overline{x} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$
σ	The population standard deviation	$\sigma = \sqrt{\dfrac{\sum\limits_{i=1}^{N} (x_i - \mu)^2}{N}}$
s	The sample standard deviation	$s = \sqrt{\dfrac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$
σ^2	The population variance	$\sigma^2 = \dfrac{\sum\limits_{i=1}^{N} (x_i - \mu)^2}{N}$
s^2	The sample variance	$s^2 = \dfrac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$
$X_{(n)}$	The largest data value	
$X_{(1)}$	The smallest data value	
p	The population proportion	$P = \dfrac{X}{N}$
$\hat{p}$	The sample proportion	$\hat{p} = \dfrac{X}{n}$ $\hat{p} = \dfrac{\text{count of successes in the sample}}{\text{sample size}}$
d_i	A paired difference	For two samples, $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$, $d_i = x_i - y_i$
μ_d	The mean of the population of paired differences	$\mu_d = \dfrac{\sum\limits_{i=1}^{N} d_i}{N}$

Symbol	Meaning	Formula
$\bar{d}$	The mean of a sample of paired differences	$\bar{d} = \dfrac{\sum\limits_{i=1}^{n} d_i}{n}$
s_d	The standard deviation of a sample of paired differences	$s_d = \sqrt{\dfrac{\sum\limits_{i=1}^{n} \left(d_i - \bar{d}\right)^2}{n-1}}$
$E\left(X\right)$	The expected value of X	$E\left(x\right) = \mu = \sum x f\left(x\right)$
$Var\left(X\right)$	The variance of X	$Var\left(x\right) = \sigma^2 = \sum \left(x - \mu\right)^2 f\left(x\right)$ $Var\left(x\right) = E\left(x^2\right) - \left(E\left(x\right)\right)^2$
$\dbinom{n}{k}$	The binomial coefficient	$\dbinom{n}{k} = \dfrac{n!}{k!\left(n-k\right)!}$
$\sim$	is distributed	
$\overset{approx}{\sim}$	is distributed approximately	
$P\left(X \leq x\right)$	The probability that the random variable X is less than or equal to the value x	
$\Phi(x)$	The cumulative standard normal distribution	$P\left(X \leq x\right)$ for $X \sim N(0,1)$
α	The probability of a type I error	
β	The probability of a type II error	
$1 - \beta$	Power of a statistical test	

Greek Alphabet

Greek Letter	Upper Case	Lower Case
alpha	A	α
beta	B	β
gamma	Γ	γ
delta	Δ	δ
epsilon	E	ϵ
zeta	Z	ζ
eta	H	η
theta	Θ	θ
iota	I	ι
kappa	K	κ
lambda	Λ	λ
mu	M	μ
nu	N	ν
xi	Ξ	ξ
omicron	O	o
pi	Π	π
rho	P	ρ
sigma	Σ	σ
tau	T	τ
upsilon	Υ	υ
phi	Φ	ϕ
chi	X	χ
psi	Ψ	ψ
omega	Ω	ω