Conditional Probability

Probability of event B, given event $A = P(B|A)$.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Independent Events

Two events A and B are said to be independent **if and only if**

$$P(B|A) = P(B) \qquad \text{or} \qquad P(A|B) = P(A)$$

Bayes' Rule

Total Probability / Rule of Elimination

If the events $B_1 \ldots B_k$ are a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1 \ldots k$, then for any event A in S:

$$P(A) = \sum_{i=1}^{k} = P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

We find $P(A)$ through finding the intersections between events A and $B_1 \ldots B_k$ and summing them together.

Example: Machines B_1, B_2, and B_3 make 30% 45%, and 25% of the products. Respectively, 2%, 3%, and 2% of the products made by each machine are defective.

Given event A that the product is defective, find the probability $P(A)$ that it is defective.

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$

Bayes' Rule

Finds $P(B_i|A)$ (Product was randomly selected and defective, what is the probability that it was made by Machine B_i).

If the events $B_1 \ldots B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1 \ldots k$, then for any event A in S such that $P(A) \neq 0$

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

Some Discrete Probability Distributions

Binomial Distribution

A Bernoulli trial can result in a success with probability p and failure with probability $q = 1 - p$. The probability distribution o the number of successes in n independent trials–X–is

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

Summations of Binomial Distributions

For $P(a \leq X \leq b)$,

$$B(x; n, p) = \sum_{x=a}^{b} b(x; n, p)$$

For $b(x; n, p)$, the **mean** is $\mu = np$ and the **variance** is $\sigma^2 = npq$.

Multinomial Experiments and Distributions

If a given trial can result in any one of k possible outcomes $E_1, \ldots, E_k$ with probabilities $p_1, \ldots, p_k$, then the **multinomial distribution** will give the probability that E_1 occurs x_1 times and E_k occurs x_k times in n independent trials where

$$x_1 + x_2 + \ldots + x_k = n \text{and} p_1 + p_2 + \ldots + p_k = 1$$

The joint probability distribution is denoted by

$$f(x_1, \ldots, x_k; p_1, \ldots, p_k, n) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

Hypergeometric Distribution

Similar to binomial, but does not require independence between trials.

Used in testing where the item tested cannot be replaced.

Find probability of x successes from the k items labeled successes and $n - x$ failures from the $N - k$ items labeled failures when a random sample size n is selected from N items.

A **hypergeometric experiment** possesses the following two properties:

1. A random sample of size n is selected without replacement from N items
2. Of the N items, k may be classified as successes and $N - k$ may be classified as failures.

The number X of successes of a hypergeometric experiment is called a **hypergeometric random variable**.

The **hypergeometric probability distribution** is

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

with a range of

$$max\{0, n - (N - k)\} \leq x \leq min\{n, k\}$$

When both k (# of successes) and $N - k$ (# of failures) are larger than sample size n, the range of a hypergeometric random variable will be $x = 0, 1, \ldots, n$

The mean of a hypergeometric distribution $h(x; N, n, k)$ is $\mu = \frac{nk}{N}$.

The variance of a hypergeometric distribution is

$$\sigma^2 = \frac{N - n}{N - 1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

Multivariate Hypergeometric Distribution

If N items can be partitioned into the k cells $A_1, A_2, \ldots, A_k$ with $a_1, a_2, \ldots, a_k$ elements, respectively, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of elements selected from $A_1, A_2, \ldots, A_k$, is a random sample of size n is

$$f(x_1, \ldots, x_k; a_1, \ldots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2} \cdots \binom{a_k}{x_k}}{\binom{N}{n}}$$

Example:

- $N = 10$ individuals used for a study
- $A_1 \ldots A_3$ = Blood types O, A, and B
- $a_1 \ldots A_3 = 3$ with O, 4 with A, 3 with B
- n = random sample of 5 individuals
- $x_1 \ldots x_3 = 1$ person with O, 2 with A, and 2 with B

$$f(1, 2, 2; 3, 4, 3, 10, 5) = \frac{\binom{3}{1}\binom{4}{2}\binom{3}{2}}{\binom{1}{0}5} = \frac{3}{14}$$

Negative Binomial and Geometric Distributions

In a binomial distribution, we are interested in the probability of x successes in a fixed n trials.

In a **negative binomial** distribution, the experiment has the same properties, but we are interested in the probability that the kth success occurs on the xth trial.

If repeated independent trials can result in a success with probability p and failure with a probability $q = 1 - p$, then the probability distribution of the random variable X (the number of the trial on which the kth success occurs) is

$$b(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \qquad x = k, k+1, k+2, \ldots$$

The probability distribution of the random variable X, the number of the trial on which the **first success** occurs, is the **geometric distribution**

$$g(x; p) = pq^{x-1}, \qquad x = 1, 2, 3, \ldots$$

The mean of the geometric distribution is

$$\mu = \frac{1}{p}$$

and the variance is

$$\sigma^2 = \frac{1-p}{p^2}$$

Applications of Negative Binomial and Geometric Distributions

Geometric distributions deal with situations involving the number of trials prior to a success (repeatedly calling a busy phone line).

Applications of the negative binomial distribution are similar, where attempts are costly and *are occurring in sequence.*

Poisson Distribution and the Poisson Process

Poisson experiments – Experiments yielding numerical values of a random variable X, which is the number of outcomes occurring during a given time interval or in a specified region.

Examples of Poisson distributions:

- Number of phone calls received per hour in an office
- Number of days school is closed due to snow
- Number of games postponed due to rain in a baseball season

Properties of the Poisson process

1. No memory. # of outcomes in one interval/region is independent of the number of outcomes in a disjoint interval/region.
2. Probability that a 1 outcome will occur in a small interval/region is proportional to interval/region size and **not** related to the number of outcomes occurring outside the interval/region.
3. Probability that > 1 outcome will occur in a small interval/region is negligible.

The PDF of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

where λ is the average number of outcomes per unit time, distance, area, etc. . .

Poisson probability sum

$$P(r, \lambda t) = \sum_{x=0}^{r} p(x; \lambda t)$$

Mean number of outcomes and variance are both $\mu = \sigma^2 = \lambda t$.

Approximation of a Binomial Distribution by a Poisson Distribution

If a binomial distribution has a large n and a small p, the conditions simulate the *continuing space or time* of the Poisson process.

If n is large and p is close to 0, the Poisson distribution can be used with the mean $\mu = np$.

If p is close to 1, we flip success and failure to get a new p close to 0 and continue with the approximation.

Let X be a binomial random variable with probability distribution $b(x; n, p)$. When $n \to \infty$, $p \to 0$, and $np \overset{n\to\infty}{\to} \mu$ remains constant,

$$b(x; n, p) \overset{n\to\infty}{\to} = p(x; \mu)$$

Continuous Probabiltiy Distributions

Continuous Uniform Distribution

Density function that is flat; probability function is uniform on a closed interval $[A, B]$

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & \text{elsewhere} \end{cases}$$

This density function forms a rectqangle with base $B - A$ and **constant height** $\frac{1}{B-A}$

The mean is $\mu = \frac{A+B}{2}$ and the variance is $\sigma^2 = \frac{(B-A)^2}{12}$.

Normal Distribution / Gaussian Distribution

A continuous random variable x with a normal distribution is called a **normal random variable**.

The density of the normal random variable X with mean μ and variance σ^2 is

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

Once μ and σ are specified, the normal curve is completely determined.

A larger variance means a flatter normal curve, because the area under the curve must equal 1. The mean defines where the peak and mode of the curve is.

The normal curve is symmetric on $x = \mu$ and has inflection points at $x = \mu \pm \sigma$. It is concave downward if $\mu - \sigma < X < \mu + \sigma$ and concave upward otherwise.

The normal curve approaches the horizontal axis asymptotically.

Areas Under The Normal Curve

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mathrm{d}x$$

It is really hard to solve that above integral, we use the following simplification.

Transform X into random variable Z with mean 0 and variance 1

$$Z = \frac{X - \mu}{\sigma}$$

This turns the integral into

$$P(z_1 < Z < z_2) = \int_{z_1}^{z_2} n(z; 0, 1)\mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} \mathrm{d}z$$

This is called the **standard normal distribution**.

Using the normal curve in reverse

To begin with a known area or probability, find the z value, and use that to determine the x value, we have $x = \sigma z + \mu$

Applications of the Normal Distribution

When turning a probability in X, like $P(X \ge 1)$, into a probability in Z, you have to subtract the mean and divide by the standard deviation on both sides of the inequality.

Normal Approximation to the Binomial

For large n, the normal distribution can approximate the binomial distribution.

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$ then the normal distribution variable is

$$Z = \frac{X + 0.5 - np}{\sqrt{npq}}$$

The **continuity correction** is **+0.5**.

$$\begin{aligned} P(X \le x) &= \sum_{k=0}^{x} b(k; n, p) \\ &\approx \text{area under normal curve to left of } x + 0.5 \\ &= P\left(Z \le \frac{x + 0.5 - np}{\sqrt{npq}}\right) \end{aligned}$$

Gamma and Exponential Distributions

The **gamma function** is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$$

The function has the following properties:

- $(\alpha-1)\Gamma(\alpha-1)$
- For a positive integer n, $\Gamma(n) = (n-1)!$ and $\Gamma(1) = 1$
- $\Gamma(1) = 1$
- $\Gamma(1/2) = \sqrt{\pi}$

The continuous random variable X has a **gamma distribution**, with parameters α and β, with the density function

$$f(x;\alpha,\beta) = \left\{ \begin{array}{l|c} \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & x>0 \\ 0 & \text{elsewhere} \end{array} \right.$$

where $\alpha > 0$ and $\beta > 0$

The CDF of the **gamma distribution** is

$$P(X \le x) = F(z;\alpha)\int_0^z \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\mathrm{d}y$$

where $z = x/\beta$

The **exponential distribution** is where $\alpha = 1$

$$f(x;\beta) = \left\{ \begin{array}{l|c} \frac{1}{\beta}e^{-x/\beta} & x>0 \\ 0 & \text{elsewhere} \end{array} \right.$$

where $\beta > 0$

The mean and variance of the gamma / exponential distributions are

$$\mu = \alpha\beta$$
$$\sigma^2 = \alpha\beta^2$$

Relationship to the Poisson Process

The Poisson distribution is used to compute the probability of a specific number of "events" occurring during a particular *period of time or span of space.*

The Poisson distribution had the parameter λ, which was the mean number of events per unit "time".

When applying the exponential distribution to the Poisson process, we have the density function

$$f(x) == \frac{1}{\beta}e^{-x/\beta}$$

where β is the **mean time between events**.

This application of the exponential distribution is used to calculate mean time between equipment failures.

The Memory-less Property and Its Effect on The Exponential Distribution

If a component makes it to t_0 hours, the probability of it lasting an additional t is the same as the probability of reaching t_0.

With the exponential distribution, there is no "punishment" for wear and tear on the component, and is only appropriate when the memory-less property is justified.

Gamma Distribution and the Poisson Process

Exponential distribution is the gamma distribution with $\alpha = 1$

Exponential distribution is the mean time until the occurrence of Poisson events or mean time between Poisson events.

The *time or space* between a *specific number* of Poisson events is described by the Gamma distribution where α is the number of events.

The Gamma distribution works well even in situations without a clear Poisson structure, such as **survival time** problems in engineering and medical applications.

Chi-Squared Distributions

Another special case of the gamma distribution

$\alpha = v/2$ and $\beta = 2$ where v is a positive integer.

v is called the **degrees of freedom**.

$$f(x;v) = \left\{ \begin{array}{l|c} \frac{1}{2^{v/2}\Gamma(v/2)}x^{v/2-1}e^{-x/2} & x>0 \\ 0 & \text{elsewhere} \end{array} \right.$$

Mean and variance are

$$\mu = v$$
$$\sigma^2 = 2v$$

The Chi-Squared distribution plays a vital role in statistical inference, such as statistical hypothesis testing and estimation.

Beta Distribution

The **beta function** is defined by

$$\beta(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

for $\alpha, \beta > 0$

The **beta distribution** density function is

$$f(x) = \left\{ \begin{array}{l|l} \frac{1}{\beta(\alpha,\beta)}x^{n-1}(1-x)^{\beta-1} & 0<x<1 \\ 0 & \text{elsewhere} \end{array} \right.$$

The uniform distribution on $(0,1)$ is a beta distribution with parameters $\alpha = 1$ and $\beta = 1$

Mean and variance are

$$\mu = \frac{\alpha}{\alpha+\beta}$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Lognormal Distribution

Applies in cases where a natural log transformation results in a normal distribution

The continuous random variable X hasa **lognormal distribution** if the random variable $Y = ln(X)$ has a normal distribution with mean μ and standard deviation σ. The resulting density function of X is

$$f(x;\mu,\sigma) = \left\{ \begin{array}{l|l} \frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2\sigma^2}[ln(x)-\mu]^2} & x\ge 0 \\ 0 & x<0 \end{array} \right.$$

The mean and variance are

$$\mu = e^{\mu+\sigma^2/2}$$
$$\sigma^2 = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$$

Functions of Random Variables

Introduction

Transformations of Variables

Discrete Case

Suppose that X is a **discrete** random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a **one-to-one** transformation between the values of X and Y.

Then the probability distribution of Y is

$$g(y) = f[u^{-1}(y)]$$

Suppose that X_1 and X_2 are **discrete** random variables with joint probability distribution $f(x_1,x_2)$. Let $Y_1 = u_1(X_1,X_2)$ and $Y_2 = u_2(X_1,X_2)$ define a one-to-one transformation between the points (x_1,x_2) and (y_1,y_2) so that the equations

$$y_1 = u_1(x_1,x_2) \quad \text{and} \quad y_2 = u_2(x_1,x_2)$$

may be uniquely solved liked

$$x_1 = w_1(y_1, y_2) \quad \text{and} \quad x_2 = w_2(y_1, y_2)$$

Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]$$

We can use the above rule to find the probability distribution of the variable $Y_1 = u_1(X_1, X_2)$ where X_1 and X_2 are discrete random variables with joint probability distribution $f(x_1, x_2)$.

We find the distribution of Y_1 by defining a $Y_2 = u_2(x_1, x_2)$ to maintain one-to-one correspondence between points and get probability distribution $g(y_1, y_2)$. The marginal distribution $h(y_1)$, which is the summation of $g(y_1, y_2)$ over y_2, is the probability distribution for Y_1.

Continuous Case

Suppose that X is a **continuous** random variable with pdf $f(x)$. Let $Y = u(X)$ define a one-to-one correspondence between X and Y. Also, $w(y) = u^{-1}(y)$.

Then the pdf of Y is

$$g(y) = f[w(y)]|J|$$

where $J = w'(y)$ and is called the **Jacobian** of the transformation.

Continuous Joint Probability Case

X_1 and X_2 are **continous** with distribution $f(x_1, x_2)$. Let the functions define the one-to-one transformation:

$$Y_1 = u_1(X_1, X_2)$$
$$Y_2 = u_2(X_1, X_2)$$

so they can be solved for

$$x_1 = w_1(y_1, y_2)$$
$$x_2 = w_1(y_1, y_2)$$

Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]|J|$$

where the Jacobian is the 2×2 determinant

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

and $\frac{\partial x_1}{\partial y_1}$ is the derivative of $x_1 = w_1(y_1, y_2)$ with respect to y_1 with y_2 held constant.

Fundamental Sampling Distributions and Data Descriptions

Populations, Samples, and Random Sampling

Population — Totality of observations which we are concerned, whether they are finite or infinite.

Sample — A subset of a population.

Sometimes we cannot study the entire population, so we take samples.

Any sampling procedure that consistently overestimate or underestimate some characteristic of the population is said to be **biased**.

To elimiate bias, choose a **random sample**.

In selecting a random sample of size n from a population $f(x)$, define $X_i, i = 1, 2, \ldots, n$ to be the ith measurement or sample value that we observe.

$X_1, \ldots, X_n$ is the **random sample** of size n from the population $f(x)$, each with the probability distribution $f(x_1), \ldots, f(x_n)$.

Because the random elements are selected under identical conditions, the random variables $X_1, \ldots, X_n$ **should be independent** and that each has the **same probability distribution** $f(x)$.

The joint probability distribution
$f(x_1, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$

Important Statistics

We select random samples to find information about the unknown poulation parameters.

We take a large random sample, find the parameter $\hat{p}$ that applies to the sample, and then use it to determine the parameter p that applies to the entire population.

$\hat{p}$ is a function of the observed values in the random sample, and it varies from sample to sample. Thus $\hat{p}$ is the value of the random variable P.

Statistic — Any function of the random variables constituting a random sample

The sample median shows the middle value of a sample.

The sample mode is the value of the sample that occurs the most often.

Sample Variance, Standard Deviation, and Range

Sample variance

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

The sample variance is defined to be the average of the squares of the differences of the observations from their means.

Sample standard deviation $= s = \sqrt{s^2}$

Sample range $= X_{max} - X_{min}$

Sampling Distributions

Sampling Distribution — The probability distribution of a statistic

Statistic — Any function of the random variables constituting a random sample

The sampling distribution of a statistic depends on:

- Population distribution
- Sample size
- Method of chooing samples

Sampling distribution of the mean — The probability distribution of $\bar{X}$

The sampling distributions of $\bar{X}$ and S^2 allow us to make inferences on μ and σ^2.

Sampling Distribution of Means and the Central Limit Theorem

Suppose a random sample of n observations is taken from a normal population with a mean μ and variance σ^2.

Each observation $X_i = 1, 2, \ldots, n$ of the random sample will have the same normal distribution as the population being sampled.

The random variable $\bar{X}$ has a normal distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$

If we are sampling from a population with unknown distribution, then the sampling distribution of $\bar{X}$ will be approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ because of the **Central Limit Theorem**.

Central Limit Theorem — If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and variance σ^2, then the limiting form of the dsitribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $n \to \infty$ is the standard normal distribution $n(z; 0, 1)$.

This approximation will generally be good if $n \geq 30$, provided that the population distribution is not terribly skewed.

Sampling Distribution of the Difference between Two Means

To find the difference $\bar{X}_1 - \bar{X}_2$:

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

IF both $n_1 \geq 30$ and $n_2 \geq 30$, then the normal approximation is good when the underlying distribution is not too far from normal.

Even when the above condition does not apply, the approximation is still pretty good unless the distribution is decidedly non-normal.

Normal Approximation to the Binomial Distribution

The binomial random variable is the X number of successes in n trials. Both X and X/n are approximately normal with a sufficiently large n, given that $np \geq 5$ and $nq \geq 5$.

Sampling Distribution of S^2

The sampling distribution of S^2 is used to infer the total population variance σ^2.

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

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

has a chi-squared distribution with $v = n - 1$ degrees of freedom.

A given $\mathcal{X}^2$ value will have an area of α. We represent that as $\mathcal{X}^2_\alpha$

Exactly 95% of a ch-squared distribution lies between $\mathcal{X}^2_{0.975}$ and $\mathcal{X}^2_{0.025}$. An $\mathcal{X}^2$ value outside of this interval is not likely to occur.

t-Distribution

Use of the Central Limit Theoremand the normal distribution *assumes the population standard deviation is known.*

The t-distribution has the random variable T with $v = n - 1$ degrees of freedom

$$T = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

for independent *normal* random variables $X_1, \ldots, X_n$ with mean μ.

$\bar{X}$ is the sample mean and S is the sample standard deviation (square root of sample variance).

The variance of T depends on sample size n and is always greater than 1.

A t-value with v degrees of freedom and leaves an area of α to the right is represented by t_α.

$t_{1-\alpha} = -t_\alpha$

Exactly 95% of a t-distribution with $v = n - 1$ degrees of freedom lie between $-t_{0.025}$ and $t_{0.025}$. Other values also have 95% of the area, but these are the simplest.

A t value that falls below $-t_{0.025}$ or above $t_{0.025}$ is either a rare occurance or our assumption about μ is wrong.

F-Distribution

The F-distribution is used to compare sample variances, and is used in problems with two or more samples.

F is defined to be the ratio of two independent chi-squared random varaibles, each divided by its number of degrees of freedom.

$$F = \frac{U/v_1}{V/v_2}$$

f_α is the f-value that gives an area of α.

For v_1 and v_2 degrees of freedom,

$$f_{1-\alpha} = \frac{1}{f_\alpha(v_1, v_2)}$$

F-distribution with two sample variances

If S_1^2 and S_2^2 are variances of independent random samples of size n_1 and n_2, taken from *normal populations* with variances σ_1^2 and σ_2^2, then

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

has an F-disribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$.

One- and Two-Sample Estimation Problems

Introduction

Formal outline of the procedure of statistical inference.

Estimation of Population Parameters

Statistical Inference

Statistical Inference— Methods that one uses to make inferences or generalizations about a population

Two methods of statistical inference:

- **Classical method** — Inferences are strictly based on information from a random sample obtained from the population.
- **Bayesian method** — Inferences are based on prior subjective data about the unknown parameters' distributions *and* sample data information.

Two major areas of statistical inference:

- **Estimation** — Use a sample to estimate a property of the population. Ex: election polling
- **Test of hypothesis** — Use sample data to prove or disprove a hypothesis. Ex: prove that one brand of paper towel is more absorbant than another.

Classical Methods of Estimation

A **point estimate** of some population parameter θ is a single value $\hat{\theta}$.

Ex: $\bar{x}$ of the statistic $\bar{X}$, computed from sample size n, is a point estimate of μ.

Ex 2: $\hat{p} = x/n$ of a binomial distribution is a point estimate of the true proportion p.

All point estimates have some error from their true value.

To estimate μ, sometimes the sample median $\tilde{X}$ is a better estimate than the sample mean $\bar{X}$.

Unbiased Estimator

Let $\hat{\Theta}$ be an estimator whose value $\bar{\theta}$ is a point estimate of some unknown population parameter θ.

$\hat{\Theta}$ is an **unbiased estimator** of θ if

$$\mu_{\bar{\Theta}} = E(\bar{\Theta}) = \theta$$

Variance of a Point Estimator

If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of θ, we want to choose the estimator with the *smaller variance*, as it is the **more efficient estimator** of θ.

For *normal populations*, both $\bar{X}$ and $\tilde{X}$ are unbiased estimators of μ, but the variance of $\bar{X}$ is smaller.

Interval Estimation

Interval Estimation — An interval within which we would expect to find the value of the parameter. It is of the form

$$\hat{\theta_L} < \theta < \hat{\theta_U}$$

where teh bounds depend on the value of $\hat{\Theta}$ for a particular sample and the sampling distribution of $\hat{\Theta}$.

As the sample size increases, $\sigma^2_{\bar{X}} = \sigma^2/n$ decreases and our estimate is likely to be closer to μ.

Interpretation of Interval Estimates

$\hat{\theta_L}$ and $\hat{\theta_U}$ are values of the corresponding random variables $\hat{\Theta_L}$ and $\hat{\Theta_U}$.

If we find $\hat{\Theta_L}$ and $\hat{\Theta_U}$ such that

$$P(\hat{\Theta_L} < \theta < \hat{\Theta_U}) = 1 - \alpha$$

for $0 < \alpha < 1$, then we have a probability of $1 - \alpha$ of selecting a random sample that will produce an interval containing θ.

$\hat{\theta_L} < \theta < \hat{\theta_U}$ is called a $100(1 - \alpha)\%$ **confidence interval**.

The fraction $1 - \alpha$ is called the **confidence coefficient** or **degree of confidence**.

The endpoints of the interval are called the **confidence limits**.

When $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we have a **wider** 99% confidence interval.

The wider the confidence interval, the more confident we are that the interval contains the unknown parameter.

It is better to be 95% confident that the variable is between 6 and 7 than to be 99% confident that it is between 3 and 10.

Single Sample: Estimating the Mean

If $\bar{x}$ is the mean of a random sample of size n from a population with a known variance σ^2, a $100(1 - \alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the z-value leaving an $\alpha/2$ area to the right.

We can be $100(1 - \alpha)\%$ confident that the error (size of difference between μ and $\bar{x}$) will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

If $\bar{x}$ is used to estimate μ, we can be $100(1 - \alpha)\%$ confident that the error will not exceed a specific amount e when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

One-Sided Confidence Bounds

If $\bar{X}$ is the mean of the random sample of size n from a population with a variance σ^2, the one-sided $100(1 - \alpha)\%$ confidence bounds for μ are given by

Upper one-sided bound:	$\bar{x} + z_\alpha\sigma/\sqrt{n}$
Lower one-sided bound:	$\bar{x} - z_\alpha\sigma/\sqrt{n}$

The case of σ Unknown

If we have a random sample from a *normal distribution*, then the random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a Student t-distribution wtih $n - 1$ degrees of freedom and a sample standard deviation S.

The procedure with the bounds is the same, but σ is replaced by S and the standard normal distribution is replaced by the t-distribution.

If $\bar{x}$ and s are the mean and standard deviation of a randomsample from a normal population with unknown variance σ^2, a $100(1 - \alpha)\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Upper one-sided bound:	$\bar{x} + t_\alpha s/\sqrt{n}$
Lower one-sided bound:	$\bar{x} - t_\alpha s/\sqrt{n}$

For σ known, we use the Central Limit Theorem.

For σ unknown, we use the t-distribution, which is based on the premise that the sampling is from a normal (approximately bell-shaped) distribution.

Concept of a Large-Sample Confidence Interval

Even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

This is the **Large-Sample Confidence Interval**, based on the Central Limit Theorem. This is only an approximation, though.

Standard Error of a Point Estimate

A point estimate supplies a single number extracted from experimental data.

A confidence interval estimate provides a reasonable estimate **given the experimental data**.

The **standard error (s.e)** of $\bar{X}$ is $\sigma/\sqrt{n}$.

The **estimated standard error (e.s.e)** of $\bar{X}$ is $s/\sqrt{n}$.

The confidence interval is no better (in terms of width) than the qualit yof the point estimate.

The confidence limits are multiplied by the s.e. or the e.s.e.

Prediction Intervals

Prediction intervals are for:

- Value of a future observation, based on observed data
- Uncertainty of a single observation

To predict a new observation, we need to account for the variation of estimating the mean *and* the variation of a future observation.

Known Population Variance σ^2

For a normal distribution of measurements with unknown mean μ and known variance σ^2, a $100(1 - \alpha)\%$ **prediction interval** of a future observation x_0 is

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1 + 1/n}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Unknown Population Variance σ^2

For a normal distirbution of measurements with unknown mean μ and variance σ^2, a $100(1 - \alpha)\%$ **prediction interval** of a future observation x_0 is

$$\bar{x} - t_{\alpha/2}\sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2}\sqrt{1 + 1/n}$$

where $t_{\alpha/2}$ with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Use of Prediction Limits for Outliar Detection

An outliar can be seen as an observation that comes from a population with a different mean.

An observation is an outliar if it falls outside the prediction ionterval computed without including the questionable observation in the sample.

Two Samples: Estimating the Difference Between Two Means

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random variales of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_1^2, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 <$$
$$(\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

The Experimental Conditions and the Experimental Unit

For confidence interval estimation of $\mu_1 - \mu_2$, the experimental conditions must keep μ_1 and μ_2 completely independent.

Experimental Unit — The part of the experiment that produces experimental error and is responsible for σ^2. Ex: the patient in a drug study, the plot of ground in an agricultural study, the raw materials in a chemical experiment

To avoid biasing the experimental units, make sure the two populations are randomly assigned to the experimental units.

Variances Unknown but Equal

A point estimate of the unknown common variance σ^2 can be obtained by the pooled estimator S_p^2

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 2)S_2^2}{n_1 + n_2 - 2}$$

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2 from approximately normal populations with equal unknown variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 <$$
$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where s_p is the pooled estimator of the population standard deviation and $t_{\alpha/2}$ is the t-value with $v = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Interpretation of the Confidence Interval

For a single parameter, the confidence interval just provides error bounds–and contains reasonable values (given the experimental data) inside it.

If we have a high confidence that $\mu_1 - \mu_2$ is positive, then we can infer that $\mu_1 > \mu_2$.

Equal Sample Sizes

It is okay if the variances are not completely equal or the populations are not completely nomral.

If the population variances are considerably different, we still get reasoanble results when the populations are normal and the sample sizes $n_1 = n_2$.

When planning an experiment, one should make every effort to equalize the size of the samples.

Unknown and Unequal Variances

If $\bar{x}_1$ and s_1^2, $\bar{x}_2$ and s_2^2 are the means and variances of independent random samples of sizes n_1 and n_2 from *approximately normal populations with unknown unequal variances*, an an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 <$$
$$(\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $t_{\alpha/2}$ is the t-value with

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}$$

degrees of freedom (rounded down), leaving an area of $\alpha/2$ to the right.

All of these confidence intervals can be written as

$$\text{point estimate} \pm t_{\alpha/2} \wedge \text{ s.e.(point estimate)}$$

Paired Observations

Paired Observations – Each experimental unit has a pair of observations, one for each population.

The observations in a pair have something in common.

Measuring differences $d_1 \dots d_n$, values of random sample $D_1 \dots D_n$ from a population of differences that we assume is normaly distributed with $\mu_D = \mu_1 - \mu_2$ and σ_D^2.

We estimate σ_D^2 with s_D^2. The point estimator of μ_D is $\bar{D}$.

When Should Pairing Be Done?

The i-th pair difference is $D_i = X_{1i} - X_{2i}$.

The two observations are taken on the same sample experimental unit and *are not independent*.

$$\text{Var}(D_i) = Var(X_{1i} - X_{2i}) = \sigma_1^2 + \sigma_2^2 - 2\text{Cov}(X_{1i}, X_{2i})$$

One expects a positive covariance if the unit is homogenous.

The intent of pairing is to reduce σ_D

The gain in quality of the confidence interval (compared to not pairing) will be greatest when there is homogenity within units and large differences between units.

The performance of the confidence interval will depend on the standard error of $\bar{D} = \sigma_D/\sqrt{n}$, where n is the number of pairs.

Tradeoff Between Reduding Variance and Losing Degrees of Freedom

Although pairing should reduce variance and standard error of the point estimate, the degrees of freedom are reduced by changing the problem to one-sample.

Changing the degrees of freedom changes $t_{\alpha/2}$ and may be counterproductive if the reduction in variance is only moderate.

Another illustration of pairing involves choosing n pairs of subjects, with each pair having a similar characteristic such as IQ, age, or breed, and then selecting one member of each pair at random to yield a value for X_1 and have the other member yield X_2.

X_1 and X_2 may represent the grades obtained by two kids with equal IQ, when the two are randomly assigned to two different teaching methods.

Confidence Interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observations

If $\bar{d}$ and s_d are the mean and standard deviation of the nomrally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Single Sample: Estimating a Proportion

A point estimator of the proportion p in a binomial experiment is given by the statistic $\hat{P} = X/n$, where X represents the number of successes in n trials.

The sample proportion $\hat{p} = x/n$ will be used as the point estimate of the parameter p.

If $\hat{p}$ is the proportion of succeses in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate $100(1-\alpha)\%$ confidence interval for the binomial parameter $\hat{p}$ is given by

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

This method is unreliable when n is small and the unknown p is believed to be close to 0 or 12.

Both $n\hat{p}$ and $n\hat{q}$ should be ≥ 5.

If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confient that the error will not exceed $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$

If $\hat{p}$ is used asn estimate of p, we can be **at least** $100(1-\alpha)\%$ confident that the error will be less than e when the sample size is about

$$n = \frac{z^2_{\alpha/2}}{4e^2}$$

Two Samples: Estimating the Difference between Two Proportions

If $\hat{p}_1$ and $\hat{p}_2$ are the proportions of successes in random samples of size n_1 and n_2, an approximate $100(1-\alpha)\%$ confidence interval for the difference of two binomial parameters $p_1 - p_2$ is given by

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 <$$
$$(\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Single Sample: Estimating the Variance

If s^2 is the variance of a random sample of size n from a normal population, a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\mathcal{X}^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\mathcal{X}^2_{1-\alpha/2}}$$

where $\mathcal{X}^2_{\alpha/2}$ and $\mathcal{X}^2_{1-\alpha/2}$ are Chi-squared values with $v = n-1$ degrees of freedom.

Two Samples: Estimating the Ratio of Two Variances

If s_1^2 and s_2^2 are the variances of independent samples of sizes n_1 and n_2 from normal populations, then a $100(1-\alpha)\%$ confidence interval for σ_1^2/σ_2^2 is

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2, v_1)$$

where $f_{\alpha/2}(v_1, v_2)$ is an f-value with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

One- and Two-Sample Tests of Hypotheses

Statistical Hypotheses: General Concepts

Sometimes the problem is not estimating the mean/variance, but making a data-based decision.

Statistical Hypothesis — Assertion or conjecture concerning one or more populations.

Random sampling is used to find data that supports or refutes the hypothesis.

The Role of Probability in Hypothesis Testing

The decision procedure must include an awareness of *the probability of a wrong conclusion.*

Rejection of a hypothesis means there is a small probabilty of obtaining the sample when the hypothesis is true.

Failure to reject does not rule out other possibilities. Rejection is the *firm conclusion.*

The hypothesis should be the opposite of the conclusion you are looking for, so you can reject it if you're right.

The Null and Alternative Hypothesis

Null Hypothesis (H_0) — Any hypothesis we wish to test. Defendent is innocent.

Alternative Hypothesis (H_1) — Accepted hypothesis based on rejection of H_0. Represents the question to be answered or the theory to be tested. Defendent is guilty.

The null hypothesis nullifies or opposes the alternative, and is frequently the logical compliment.

One of two concusions:

- **Reject H_o:** In favor of H_1 because of sufficient evidence in the data
- **Fail to reject H_o:** Because of insufficient evidence

Statement of H_0 represents the "status quo" in response to the new idea.

Testing a Statistical Hypothesis

The Test Statistic

Test Statistic — Random variable X

Critical value — If the observed x for X is greater than the critical value, H_0 is rejected. If $x \leq$ critical value, H_0 fails to be rejected.

Critical Region – All $x >$ critical value.

Teh Probabilty of a Type I Error

Type I Error — Rejection of H_0 when it is true. Happens if sample randomly has data that suggest H_0 is false.

Type II Error — Nonrejection of H_0 when it is false. Sample randomly has data that fails to reject H_0.

	H_0 is true	**H_0 is false**
Do not reject H_0	Correct	Type II
Reject H_0	Type I	Type I

Level of significance, size of test, (α) — Probability of committing a Type I error.

$$\alpha = P(\text{type I error}) = P\left(X > c\text{when}H_0 = \text{true}\right)$$

where c is the *critical value.*

The Probability of a Type II Error

$\beta = P(\text{type II error}) = P(X \leq c\text{when}H_1 = \text{true})$ where c is the critical value

The Role of α, β, and Sample Size

Increasing the size of the critical region will decrease β and increase α.

Decreasing the size of the critical region will increase β and decrease α.

Both β and α can be decreased by increasing n.

Important Properties of a Test of Hypothesis

- Type I error and Type II error are related. A decrease in the probability of one will increase the probability of the other.

- The size of the critical region, and therefore the probability of committing a type I error, can always be reduced by adjusting the critical value(s).
- An increase in n will reduce α and β simultaneously.
- If H_0 is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The greater the distance between the true value and the hypothesized value, the smaller β will be.

The **power** of a test is the probability of rejecting H_0 given that a specific alternative is true. Power can be computed as $1 - \beta$.

One- and Two-Tailed Tests

One-tailed test:

$$\begin{aligned} &H_0 : \theta = \theta_0 \\ &H_1 : \theta > \theta_0 \\ &\text{or} \\ &H_1 : \theta < \theta_0 \end{aligned}$$

The inequality symbol points in the direction of the critical region.

Two-tailed test:

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

The critical region is split into two parts, often having equal probabilities.

The Use of P-Values for Decision-Making in Testing Hypotheses

A **P-value** is the lowest level (of significance) at which the observed value of the test statistic is significant.

Classical Steps

1. State the null and alternative hypotheses
2. Choose α
3. Choose an appropriate test statistic and establish critical region based on α
4. Reject H_0 if the computed test statistic is in the critical region. Otherwise, do not reject.

P-Value Significance Testing

1. State null and alternative hypotheses
2. Choose an appropriate test statistic
3. Compute the P-Value based on the computed value of the test statistic
4. Use judgement based on the P-value and the knowledge of the scientific system

Single Sample: Tests Concerning a Single Mean

Test Procedure for a Single Mean (Variance Known)

For the hypothesis

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \\ z &= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \end{aligned}$$

If $-z_{\alpha/2} < z < z_{\alpha/2}$, do not reject H_0.

If $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$, reject H_0 and accept H_1.

There will be probability α of rejecting H_0 (falling into the critical region) when $\mu = \mu_0$.

One-sided hypotheses involve the same statistic, but only one tail is involved.

Tests on a Single Sample (Variance Unknown)

For the hypothesis

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

exceeds $t_{\alpha/2,n-1}$ or is less than $-t_{\alpha/2,n-1}$.

When $n \geq 30$ with a bell shaped population, we can use the Central Limit Theorem to do the calculation with

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and still get an *estimate* of the probability. The confidence interval may be wider, though.

Two Samples: Tests on Two Means

Unknown But Equal Variances

For the hypothesis:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$$

exceeds $t_{\alpha/2,n_1+n_2-2}$ or is less than $-t_{\alpha/2,n_1+n_2-2}$ where

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 + 2}$$

Unknown But Unequal Variances

If one cannot assume that $\sigma_1 = \sigma_2$, then if the populations are *normal*

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

has an approximate t-distribution with approximate degrees of freedom

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

Do **not** reject H_0 when

$$-t\alpha/2, v < t' < t_{\alpha/2,v}$$