Statistical Questions Besides Hypothesis Testing

Estimation. We also use statistics to calculate estimates of parameters. Examples might be:

- Average lifespan.
- Infant mortality rate.
- Mean body weight (could be as a function of sex, age, etc.).
- Can also use it to measure model parameters.

\[ N(t) = N_0 e^{rt} \]
Random Variables

What’s a random variable?

• A variable that takes different values with different probabilities is called a random variable.

• Typically, random variables are numerical.

• Can be continuous or discrete.
Types of Random Variables

• Continuous (real numbers)
  – Weight, height ($0$ to $\infty$)
  – Age ($0$ to $\infty$)
  – Position ($-\infty$ to $\infty$)

• Discrete (integers)
  – Number of leaves on a tree
  – Individuals (e.g. people, animals) in a population).

• Ordinal – ranks
  – Class grades
  – Bond ratings (Standard & Poor AAA, AA, A, BBB, BB, B, CCC, CC, C, D)

• Nominal – attributes that have no natural order
  – Species
  – Flavors
  – Colors
Types of Random Variables

- **Continuous (real numbers)**
  - Weight, height (0 to \( \infty \))
  - Age (0 to \( \infty \))
  - Position (-\( \infty \) to \( \infty \))

- **Discrete (integers)**
  - Number of leaves on a tree
  - Individuals (e.g. people, animals) in a population.

- **Ordinal – ranks**
  - Class grades
  - Bond ratings (Standard & Poor AAA, AA, A, BBB, BB, B, CCC, CC, C, D)

- **Nominal – attributes that have no natural order**
  - Species
  - Flavors
  - Colors
Probability
Outcomes

Suppose a phenomenon can occur in $k$ different ways. Let’s call these outcomes. Examples:

- Coin toss: 2 outcomes – head or tail.
- Card drawn from deck: 52 outcomes.
Events

Denote a specific outcome as an event. Examples:

• Coin toss: Tossing a head.
• Dice: Rolling a 1.
• Cards: Drawing an Ace.
Probability

The probability of an event is defined as:

\[
Pr(\text{event}) = \frac{\text{total number of ways the focal event can occur}}{\text{total number of outcomes}}
\]

Pr(head) = \frac{1}{2} = Pr(tail)

Pr(rolling a 3) = \frac{1}{6}
Probability of Independent Events: Coin Tosses

HHHHH = \( \left( \frac{1}{2} \right)^5 \)

TTTTT = \( \left( \frac{1}{2} \right)^5 \)
Any arrangement has the same probability

\[ \text{HTHTHT} = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^5 \]

\[ \text{THTHT} = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^5 \]

\[ \text{TTTTH} = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^5 \]

\[ \text{TTTTTH} = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^5 \]
Suppose we have a long series of heads:

HHHHHHHHHHHHH...

what’s the probability that the next coin toss is also a head?
Probability

The probability of an event is defined as:

\[
Pr(\text{event}) = \frac{\text{total number of ways the event can occur}}{\text{total number of outcomes}}
\]

\[
Pr(\text{snake eyes - 2 ones}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}
\]

\[
Pr(\text{boxcars - 2 sixes}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}
\]
Factorials

\[ n! = n(n-1)(n-2) \ldots (3)(2)(1) \]

\[ 0! = 1 \quad \text{by mathematical convention} \]

\[ 3! = 3 \times 2 \times 1 = 6 \]
\[ 4! = 4 \times 3 \times 2 \times 1 = 24 \]
\[ 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \]
\[ 50! = 30,414,093,201,713,378,043,612,608,166,064,768,844,377,641,568,960,512,000,000,000,000,000 \]

Factorials increase very rapidly!
Permutations

Permutations are the number of ways we can arrange a set of $n$ things. There are $n!$ ways to arrange a set of $n$ objects.

Think of three balls of different colors:

$$3 \times 2 \times 1 = 3! = 6$$
Permutations

We can also talk about the number of ways to arrange subsets of size \( k \) drawn from a set of \( n \) things.

\[
n^\text{P}_k = \frac{n!}{(n-k)!}
\]
Permutations

For example how many ways can we arrange two balls drawn from our set of three?

\[
\frac{3!}{(3-2)!} = \frac{6}{1} = 6
\]

Note that once we’ve chosen the first two balls, the third is determined. So \(3P_2=3!/1!=3!\), which is the same as number of permutations of the three balls taking three at a time \(3P_2=3!/0!=3!\).
Combinations

Combinations are the number of distinct subsets we can draw from a larger set, irrespective of the order of the elements.

\[ nC_k = \frac{n!}{k!(n-k)!} \]

Usually, however this is written as:

\[ \binom{n}{k} \]

Read as ‘\(n \text{ choose } k\)’.
Combinations

Combinations are the number of distinct subsets we can draw from a larger set:

\[
\frac{3!}{1!(3-1)!} = \frac{6}{2} = 3
\]
Combinations

Note that:

\[
\binom{n}{k} = \frac{n \cdot P_k}{k!}
\]

So, to get the number of combinations of \( n \) things taken \( k \) at a time, \textit{we divide} \( nP_k \) \textit{by the number of ways we can arrange the} \( k \) \textit{selected objects}.
Combinations

For example how many combinations of two balls can we draw from a set of three?

\[
\frac{3!}{2!(3-2)!} = \frac{6}{2 \times 1} = 3
\]
Combinations – Example

How many possible hands are there in 5 card stud poker?

\[
\frac{52!}{5!(52 - 5)!} = 2,598,960
\]
Probability Distribution

- Let $x$ be a random variable.
- Probability distribution assigns probabilities to the values that $x$ takes.
- Sum of probabilities for all possible values of $x$ is 1.
Discrete Distributions
Binomial Distribution

\[ \Pr(k) = \binom{n}{k} p^k (1 - p)^{n-k} \]

Probability of \( k \) “successes” in \( n \) trials.
Poisson Distribution

If we let $p \to 0$ and $n \to \infty$ but let the product $np \to \lambda$ we get the distribution:

$$\Pr(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$
Continuous Distributions
Normal Distribution

• Also called the Gaussian distribution after Karl Freidrich Gauss, AKA “the prince of mathematicians” who discussed it in 1809.

• Others including de Moivre (1733) and the Marquis de Laplace (1774) had worked on it earlier, however.

• Colloquially, it is often called the ‘bell curve” because it looks bell shaped.
More formally, \( f(x) \) is called the Normal Probability Density Function (PDF)

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

- \( x \) is the random variable.
- \( \mu \) is the mean
- \( \sigma \) is the standard deviation
Normal Distribution
Normal Distribution

• Formula for the normal distribution (PDF) is:

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

Where \( \mu \) is the mean of the distribution and \( \sigma^2 \) is the variance.

• We use the shorthand \( x \sim N(\mu , \sigma^2) \) and say “\( x \) is distributed normal, \( \mu \), \( \sigma^2 \)”.
Mean, $\mu$, affects location of mode. Variance, $\sigma^2$, affects spread.
Standard Normal Distribution

• A special case of the normal distribution occurs when the mean, $\mu$, is equal to 0 and the variance $\sigma^2$ is equal to 1.

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

• We say $x \sim N(0,1)$ – “$x$ is distributed normal, 0,1”.
Standard Normal Distribution
Standard Normal Distribution

- We can rescale any normal random variable so that it is distributed $N(0,1)$. If $x \sim N(\mu, \sigma^2)$ then $x'$

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

is $\sim N(0,1)$. 
The normal distribution is ubiquitous

The normal distribution is the most common probability density function (PDF) of all because of a theorem called the *Central Limit Theorem*

*Why do you think they call it the ‘normal’ distribution.*
Central Limit Theorem

Consider a sequence of $n$ independent random variables, $x_i$:

$$x_1, x_2, x_3, \ldots, x_{n-1}, x_n$$

with means and variances $\mu_i$ and $\sigma_i^2$. Then the PDF of the sum:

$$a_n = x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n$$

will tend to a $N(\mu, \sigma^2)$ for $n$ ‘large’. Where:

$$\mu = \sum_{i=1}^{n} \mu_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$
Central Limit Theorem

Consider a sequence of \( n \) independent random variables, \( x_i \):

\[
x_1, x_2, x_3, \ldots, x_{n-1}, x_n
\]

with means and variances \( \mu_i \) and \( \sigma_i^2 \). Then the PDF of the sum:

\[
a_n = x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n
\]

will tend to a \( N(\mu, \sigma^2) \) for \( n \) ‘large’. Where:

\[
\mu = \sum_{i=1}^{n} \mu_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2
\]
Central Limit Theorem

There are conditions:

1. $\sigma_i^2$ cannot be infinite.
2. $\sigma_i^2$ should be ‘small’ relative to $\sigma^2$

The PDF of the random variables can be anything that meets conditions 1 and 2. Some examples are shown on the next slide.
Exponential

Uniform

Gamma

Pareto

Chi-Square

Any Crazy Distribution You Can Make Up
The movie in the next slide shows what happens if we sum together random variables drawn from the funny looking PDF at right. The movie shows the PDF’s of the sums:

\[ a_1 = x_1 \ (n = 1) \]
\[ a_2 = x_1 + x_2 \ (n = 2) \]
\[ a_3 = x_1 + x_2 + x_3 \ (n = 3) \]
\[ a_4 = x_1 + x_2 + x_3 + x_4 \ (n = 4) \]
\[ \quad \quad \quad \quad \quad \quad \quad \quad \vdots \]
\[ a_7 = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 \ (n = 7) \]

By the time we get to \( n = 7 \) the PDF is essentially indistinguishable from normal. So the punch line of this movie is that even when the PDF of the individual \( x_i \) is very nonnormal, the sum of the \( x_i \) becomes normal very very rapidly.
Convergence to Normal Distribution is *really fast*!
A biological example of the Central Limit Theorem at work

Consider the phenotype for a ‘quantitative’ trait such as weight, height, etc. These traits are typically controlled by many genes. If the genes act additively (i.e. their effect on phenotype is summed together), we can represent the phenotype as:

\[ P = A_1 + A_2 + A_3 + A_4 + \ldots + A_n + E \]

where \( P \) is the phenotype, \( A_1 \ldots A_n \) are effects of the \( n \) genes that contribute to the expressed phenotype and \( E \) is an environmental effect (e.g. nutrition during development). Provided the genes have similar effects \( A_i \) (i.e. no \( A_j \) or \( E \) is much larger than any other effect, which could violate the assumption that \( \sigma_i^2 \) is small relative to \( \sigma^2 \) on slide 17), then as \( n \) becomes large, the phenotype should be normally distributed.

This is, in fact, what we see for many quantitative traits.
Skewness and Kurtosis

• When distributions differ from normality, it’s often useful to characterize how they differ. We use two quantities called skewness and kurtosis as aids in this characterization.
Skewness (3\textsuperscript{rd} moment about the mean)

\[ \gamma_1 = \frac{E[(x - E[x])^3]}{E[(x - E[x])^2]^{3/2}} = \frac{E[(x - \mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3} \]

\( k_3 \) is 0 for symmetric distributions
Kurtosis (4\textsuperscript{th} moment about the mean)

\[ \gamma_2 = \frac{E[(x - E[x])^4]}{E[(x - E[x])^2]^2} - 3 = \frac{E[(x - \mu)^4]}{\sigma^4} - 3 = \frac{\mu_4}{\sigma^4} - 3 \]

Figure 6.3 Symmetric frequency distributions. Distribution a is mesokurtic ("normal"), b is platykurtic, and c is leptokurtic.
To remember which is which

- **Platy** – flat [Greek from platus]
- **Lepto** – Slender; thin; fine [Greek, from leptos]

Although we talk of the differences in the tails of platy- and leptokurtic distributions, their names really refer to the mode:

Platykurtic: broad and flat.
Leptokurtic: Narrow and slender.

**Fig. 15.** Platykurtic curves have short tails like a platypus, while leptokurtic curves have long tails like kangaroos noted for ‘lepping’ (after ‘Student’).
Distribution of Sample Means

• Consider a sample of $n$ values all drawn from the same PDF.

$$X_1, X_2, X_3, \ldots, X_{n-1}, X_n$$

• We say that the $X_i$ in this sample are *i.i.d.* – “independent and identically distributed”.
Distribution of Sample Means

By identically distributed, we mean that all of the $X_i$ have:

1. The same PDF (which we already specified).
2. The same mean: $\mu$
3. The same variance: $\sigma^2$
Distribution of Sample Means

• Now consider the sum of the $X_i$.

$$a_n = X_1 + X_2 + X_3 + \ldots + X_{n-1} + X_n$$

• The mean and variance of this sum are:

$$\text{Mean} = \sum_{i=1}^{n} \mu = n\mu \quad \text{var}(a_n) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$$
Distribution of Sample Means

- Now consider the sample mean, $\bar{X}$.

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

- What is the variance of $\bar{X}$?
Distribution of Sample Means

- The variance of $\bar{X}$ is:

$$\text{var}(\bar{X}) = \text{var}\left(\frac{a_n}{n}\right) = \frac{1}{n^2} \text{var}(a_n)$$

$$= \frac{n \sigma^2}{n^2} = \frac{\sigma^2}{n}$$
Distribution of Sample Means

Likewise, the sample variance of $\bar{X}$ is:

$$s^2_{\bar{X}} = \frac{s^2}{n}$$

What happens as $n$ gets large?
Furthermore, by the Central Limit Theorem, the PDF of \( \bar{X} \) tends to:

\[
\bar{X} = N\left(\mu, \frac{\sigma^2}{n}\right)
\]

as \( n \) becomes large.
Hypothesis Testing

Consider an experiment:

1. Weight loss drug administered to 30 subjects.

2. \( \bar{X} \) weight loss = \(-1.1 \) kg.

3. \( s^2 \) in weight loss = \(4.0 \) kg\(^2\)
Questions:

• Is the drug effective?

• i.e., is it possible that we could have observed a 1.1 kg weight loss just by chance?
Formulate hypotheses in terms of the mean weight loss, $\mu$

$H_0 : \mu = 0 \text{ null hypothesis}$

$H_1 : \mu \neq 0 \text{ alternative hypothesis}$
Standardize the mean weight loss – i.e. convert to $N(0,1)$

- $s^2$ in weight loss = 4.0 kg$^2$
- 30 subjects in study

1. What is the variance of $\bar{X}$?
Step 1: Standardize the mean weight loss

- $s^2$ in weight loss = 4.0 kg$^2$
- 30 subjects in study

1. What is the variance of $\bar{X}$?

$$s_{\bar{X}}^2 = \frac{s^2}{n}$$
Standardize the mean weight loss

1. \( s^2_{X} = \frac{s^2}{n} = \frac{4.0}{30} = 0.13 \)
Standardize the mean weight loss

1. \[ s^2_{\bar{X}} = \frac{s^2}{n} = \frac{4.0}{30} = 0.13 \]

2. \[ s_{\bar{X}} = 0.37 \quad \text{(Standard Deviation of } \bar{X} \text{)} \]
Standardize the mean weight loss

1. \( s^2 \overline{X} = \frac{s^2}{n} = \frac{4.0}{30} = 0.13 \)

2. \( s_{\overline{X}} = 0.37 \)

3. \( Z = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{-1.1 - 0}{0.37} = -2.97 \)

\( Z \sim N(0,1) \)
Standardize the mean weight loss

1. \[ s^2_{\bar{X}} = \frac{s^2}{n} = \frac{4.0}{30} = 0.13 \]

2. \[ s_{\bar{X}} = 0.37 \]

3. \[ Z = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{-1.1 - 0}{0.37} = -2.97 \]

Under \( H_0 \), \( Z \sim N(0,1) \)
Rejection Region

• Clearly if $\bar{x}$ is $<< 0$, our $Z$ statistic will have a large negative value.

• Conversely if $\bar{x}$ is $>> 0$, $Z$ will have a large positive value.

• So our strategy will be to reject $H_0$ when $Z$ has an improbable extreme (large or small) value.

But what’s extreme?
1. We pick a rejection region comprising the union of the events $Z >> 0$ and $Z << 0$.
2. We reject the null hypothesis if $Z$ falls within that rejection region.
The level of significance, \( \alpha \), is the probability that \( Z \) falls in the rejection region, given that the null hypothesis is true.
The level of significance, $\alpha$, is the probability that $Z$ falls in the rejection region, given that the null hypothesis is true.

Typically, we set $\alpha = 0.05$ (5%). This determines the $Z$ cutoff values, here $= \pm 1.96$. 

![Diagram showing the level of significance and Z values]
The level of significance, $\alpha$, is the probability that $Z$ falls in the rejection region, given that the null hypothesis is true.

Typically, we set $\alpha = 0.05$ (5%). This determines the $Z$ cutoff values, here $= \pm 1.96$.

$\alpha = 0.05$ is most common, but other values e.g. $\alpha = 0.01$ (1%) or 0.001 (0.1%) are also commonly used.
We reject $H_0$ for the weight loss experiment.

Our $Z$ score is $-2.97$, which clearly falls in the rejection region.
We reject $H_0$ for the weight loss experiment.

Our $Z$ score is $-2.97$, which clearly falls in the rejection region.

We reject $H_0$ and say that our result is significant at the $\alpha = 0.05$ level.
In other words...

• In other words, we are saying that a mean weight loss of 1.1 kg is *unlikely* to have occurred by chance alone.

• By saying the result is significant at the 0.05 level, we mean that there is a $\leq 5\%$ chance of obtaining our result if the null hypothesis were true.
Level of Significance

- 2.97 actually has a higher level of significance: $\alpha = 0.01$
- Most common $\alpha$ is 0.05 but others often used.
- Journals set $\alpha$ cutoffs: 0.05 (significant), 0.025, 0.01 (highly significant), 0.005, 0.001 (very highly significant), 0.0001
- $P$ normally used rather than alpha in reporting results in the scientific and medical literature we say $P < 0.05$ when reporting significance level.
Level of significance

• The smaller the value of $\alpha$, the higher the significance. The more reliable the result.

• By higher level of significance, we mean the results are less likely to be due to chance alone.
What if variance was greater?

1. \( s^2_x = \frac{s^2}{n} = \frac{10.0}{30} = 0.33 \)

2. \( s_x = 0.58 \)

3. \( Z = \frac{\bar{X} - \mu}{s_x} = \frac{-1.1 - 0}{0.58} = -1.91 \)

*So the more variable the data the less significant (reliable) the results.*
What if $n$ was smaller?

1. $s_{\bar{X}}^2 = \frac{s^2}{n} = \frac{4.0}{10} = 0.4$

2. $s_{\bar{X}} = 0.63$

3. $Z = \frac{\overline{X} - \mu}{s_{\bar{X}}} = \frac{-1.1 - 0}{0.63} = -1.74$

Reducing sample size also reduces significance.
Small Sample Hypothesis Tests

• $N(0,1)$ is not appropriate for small sample sizes, because unless $n \approx 30$, our estimate of the variance of $\overline{X}$, $s^2_X$, is not accurate.

• In this case we use an alternative distribution, called Student’s $t$ distribution.
Student’s $t$ Distribution

The derivation of the $t$-distribution was first published in 1908 by William Sealy Gosset, while he worked at a Guinness Brewery in Dublin. He was not allowed to publish under his own name, so the paper was written under the pseudonym **Student**.
William Sealy Gosset
AKA 'Student'

'Student' in 1908
Student’s $t$ Distribution

$$f(t) = \frac{\Gamma\left(\frac{\nu}{2} + \frac{1}{2}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right) \left(1 + \frac{t^2}{\nu}\right)^{\left(\frac{\nu}{2} + \frac{1}{2}\right)}}$$

Where $\nu = n - 1$ and is called the degrees of freedom
Student’s $t$ Distribution

$v = 1$

$v = 3$

$v = \infty$, $N(0,1)$

$t = \frac{\bar{X} - \mu}{S_{\bar{X}}}$
Student’s $t$ Distribution

• So Student’s $t$ Distribution is *leptokurtic* relative to the Normal Distribution.

• So if we use the $N(0,1)$ when the $t$ distribution is really appropriate, what happens to the probability of Type I error?
Student’s $t$ Distribution

• So Student’s $t$ Distribution is leptokurtic relative to the Normal Distribution.

• So if we use the $N(0,1)$ when the $t$ distribution is really appropriate, what happens to the probability of Type I error?

• We *underestimate* the Type I error.
Comparison of Normal and \( t \) distribution \((\nu = 1)\)

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \pm Z_{\alpha/2} )</th>
<th>( t ) distribution probability of Type I Error (actual ( \alpha ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>( \pm 1.96 )</td>
<td>0.30</td>
</tr>
<tr>
<td>0.01</td>
<td>( \pm 2.58 )</td>
<td>0.24</td>
</tr>
<tr>
<td>0.001</td>
<td>( \pm 3.29 )</td>
<td>0.18</td>
</tr>
<tr>
<td>0.0001</td>
<td>( \pm 3.89 )</td>
<td>0.16</td>
</tr>
</tbody>
</table>

The point of this slide is to emphasize that the Normal \( Z \) test really is inappropriate for small sample sizes.
Small sample rejection region

- $n = 4$
- $\nu = n - 1 = 3$
- $\alpha = 0.05$

To find the rejection region, we use the $t$-table in the text.
### t-test table

**Appendix B  Statistical Tables and Graphs**

**TABLE B.3  Critical Values of the t Distribution**

<table>
<thead>
<tr>
<th></th>
<th>$\alpha(2) : 0.50$</th>
<th>$0.20$</th>
<th>$0.10$</th>
<th>$0.05$</th>
<th>$0.02$</th>
<th>$0.01$</th>
<th>$0.005$</th>
<th>$0.002$</th>
<th>$0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>$\alpha(1) : 0.25$</td>
<td>$0.10$</td>
<td>$0.05$</td>
<td>$0.025$</td>
<td>$0.01$</td>
<td>$0.005$</td>
<td>$0.0025$</td>
<td>$0.001$</td>
<td>$0.0005$</td>
</tr>
<tr>
<td>3</td>
<td>0.765</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
<td>10.215</td>
<td>12.924</td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
<td>7.173</td>
<td>8.610</td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>6.869</td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
<td>5.208</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.408</td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.781</td>
</tr>
<tr>
<td>10</td>
<td>0.700</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
<td>4.144</td>
<td>4.587</td>
</tr>
<tr>
<td>11</td>
<td>0.697</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
<td>4.025</td>
<td>4.437</td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
<td>3.930</td>
<td>4.318</td>
</tr>
<tr>
<td>13</td>
<td>0.694</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.372</td>
<td>3.852</td>
<td>4.221</td>
</tr>
<tr>
<td>14</td>
<td>0.692</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
<td>3.787</td>
<td>4.140</td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
<td>3.733</td>
<td>4.073</td>
</tr>
</tbody>
</table>

Or you can use the functions TDIST and TINV in Excel.
The *t* distribution is asymptotically normal as the degrees of freedom tends to infinity.

These are the values from the table, 3 slides back.

<table>
<thead>
<tr>
<th>ν</th>
<th>α(2): 0.50</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>0.675</td>
<td>1.283</td>
<td>1.647</td>
<td>1.963</td>
<td>2.332</td>
<td>2.583</td>
<td>2.816</td>
<td>3.102</td>
<td>3.504</td>
</tr>
<tr>
<td>800</td>
<td>0.675</td>
<td>1.283</td>
<td>1.647</td>
<td>1.963</td>
<td>2.331</td>
<td>2.582</td>
<td>2.815</td>
<td>3.100</td>
<td>3.303</td>
</tr>
<tr>
<td>900</td>
<td>0.675</td>
<td>1.282</td>
<td>1.647</td>
<td>1.963</td>
<td>2.330</td>
<td>2.581</td>
<td>2.814</td>
<td>3.099</td>
<td>3.301</td>
</tr>
<tr>
<td>1000</td>
<td>0.675</td>
<td>1.282</td>
<td>1.646</td>
<td>1.962</td>
<td>2.330</td>
<td>2.581</td>
<td>2.813</td>
<td>3.098</td>
<td>3.300</td>
</tr>
<tr>
<td>∞</td>
<td>0.6745</td>
<td>1.2816</td>
<td>1.6449</td>
<td>1.9600</td>
<td>2.3263</td>
<td>2.5758</td>
<td>2.8070</td>
<td>3.0902</td>
<td>3.2905</td>
</tr>
</tbody>
</table>

The *t* distribution is asymptotically normal as the degrees of freedom tends to infinity.
Rejection region two-tailed t-test

\( \alpha = 0.05, \, \nu = 3 \)

\[ t = \frac{\bar{X} - \mu}{s / \sqrt{n}} \]

\[ -t_{\alpha(2), \nu=3} = 3.182 \]

\[ t_{\alpha(2), \nu=3} = 3.182 \]
# NORMDIST and TDIST

## Return Probabilities

<table>
<thead>
<tr>
<th>Z-test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NORMDIST</strong>($a, \mu, \sigma, C$)</td>
<td><strong>TDIST</strong>($a, \nu, \text{tails}$)</td>
</tr>
<tr>
<td>• $a$ – value of random variable you are calculating probability for.</td>
<td>$a$ – value of random variable you are calculating probability for.</td>
</tr>
<tr>
<td>• $\mu$ – mean</td>
<td>$\nu$ – degrees of freedom = $n - 1$.</td>
</tr>
<tr>
<td>• $\sigma$ – standard deviation.</td>
<td><strong>Notes:</strong> $a$ must be $\geq 0$.</td>
</tr>
</tbody>
</table>
| • $C$ – “Cumulative” = 0 for PDF and 1 for CDF. For probabilities you want $C = 1$. | tails = 1 for one-tailed test  
= 2 for two-tailed test                                                                                                                                         |
| NORMDIST($a, \mu, \sigma, 1$) = Pr[$ x \leq a$]                     | **TDIST**($a, \nu, 1$) = Pr[$t \geq a$]                                                                                                                                                                |
|                                                                        | **TDIST**($a, \nu, 2$) = 2 × Pr[$t \geq a$]                                                                                                                                                           |
# NORMINV and TINV

## Return Rejection Region Thresholds

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Z-test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-tailed</strong></td>
<td>NORMINV(α, μ, σ) Z_α(1)</td>
<td>TINV(2α, ν) t_α(1), ν</td>
</tr>
<tr>
<td>(right tail used as example)</td>
<td>Z_α(1)</td>
<td>t_α(1), ν</td>
</tr>
<tr>
<td><strong>Two-tailed</strong></td>
<td>± NORMINV(α/2, μ, σ)</td>
<td>± TINV(α, ν)</td>
</tr>
<tr>
<td></td>
<td>± Z_α(2)</td>
<td>± t_α(2), ν</td>
</tr>
</tbody>
</table>
Assumptions of $t$-tests

- Data normally distributed ($t$-test is ‘robust’ to violations of this assumption).
- Each data point is an independent sample.
1. Weight loss drug administered to 10 subjects.

2. $\overline{X}$ weight loss $= -1.5$ kg. Negative because it’s a loss in weight

3. $s^2$ in weight loss $= 4.0$ kg$^2$
One tailed test on our weight loss drug, $\mu$

- $H_0 : \mu = 0$ *null hypothesis*
- $H_1 : \mu < 0$ *alternative hypothesis*

So we are considering a more *restricted* alternative hypothesis now.
Standardize the mean weight loss

1. \[ s_{\bar{X}}^2 = \frac{s^2}{n} = \frac{4.0}{10} = 0.4 \]

2. Standard Deviation of \( \bar{X} \)

\[ s_{\bar{X}} = \sqrt{\frac{s^2}{n}} = \sqrt{0.4} = 0.63 \]
Standardize the mean weight loss

1. \[ s^2_X = \frac{s^2}{n} = \frac{4.0}{10} = 0.4 \]

2. \[ s_X = 0.63 \]

3. \[ t = \frac{\bar{X} - \mu}{s_X} = \frac{-1.5 - 0}{0.63} = -2.38 \]

Under \( H_0 \), \( t = t_{v=9} \)
Rejection region

So we reject $H_0$ at the $\alpha = 0.05$ but not at 0.01 significance level.
Two Sample Tests

Suppose that we want to test whether the means of two populations are different. Examples:

- CSI GPA vs. Hunter College.
- Weight of patients with heart disease vs. healthy control subjects.
- Batting averages of Rockies vs. Red Sox

We can do this with two sample $t$-tests
Assumptions of Two Sample $t$-tests

• Data in two samples are both normally distributed.

• The two populations have different means but equal variances.
Two Sample Tests

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

We assume that the variances are equal in the two populations.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$
Two Sample Tests

What do we mean by?

\[ S_{\bar{X}_1 - \bar{X}_2} \]
Two Sample Tests

The population variance of the difference in means is:

\[
\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}
\]

\[
= \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}
\]
Two Sample Tests

And since the variances in the two populations are assumed equal:

$$\text{var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$$
Two Sample Tests

But now we need an estimate of $\sigma^2$.

$$\text{var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$$
The standard deviations of populations 1 and 2, $s_1^2$ and $s_2^2$, are both estimates of $\sigma^2$. 
Two Sample Tests

Hence, we *pool* information on \( \sigma^2 \) from both populations into a single estimate.

\[
S_p^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2}
\]

\( s_p^2 \) is the pooled estimate of the variance, \( \sigma^2 \) which is a *weighted average*, where the ‘weights’ are the degrees of freedom \( \nu_1 \) and \( \nu_2 \).
Two Sample Tests

So our estimate of $\sigma^2_{X_1 - X_2}$ is:

$$S^2_{X_1 - X_2} = \frac{S^2_p}{n_1} + \frac{S^2_p}{n_2}$$
Two Sample Tests

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}}$$
Two Sample Tests

And our $t$ statistic becomes:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}}$$

Which has $\nu = \nu_1 + \nu_2 = n_1 + n_2 - 2$ degrees of freedom.
Two Sample Tests

Now what if we have a different hypothesis like:

\[ H_0: \mu_1 - \mu_2 = \mu_0 \]

\[ H_1: \mu_1 - \mu_2 \neq \mu_0 \] ?
Then $t$ becomes:

$$t = \frac{|\bar{X}_1 - \bar{X}_2| - \mu_0}{S_{\bar{X}_1 - \bar{X}_2}}$$

Where $S_{\bar{X}_1 - \bar{X}_2}$ is calculated the same way as before.
$t$-test is ‘robust’ to violation of these assumptions

• By *robust* we mean the test is often *still accurate* even if the assumptions of normally distributed data and equal variance are not true.

• This is especially true for large sample sizes.
Violation of normality assumptions

- Two-tailed tests are robust to skewed non-normal distributions.
- One tailed tests are less robust. More susceptible to skewness.
- For either one- or two-tailed tests, if data is strongly non-normal, $P$ (i.e. $\alpha$) values $< 0.01$ are not reliable.
Violation of normality assumptions

- Power of two-tailed tests also not strongly affected by skewness.
- Power of one-tailed tests is quite susceptible to skewness.
- Two-tailed test is less powerful for platykurtic distribution; more powerful for leptokurtic distributions.
Confidence Intervals

- We use hypothesis testing to help make decisions. In contrast, we use confidence intervals to tell us how reliable an estimate of some quantity is, for example a mean.
Confidence Limits for Population Mean $\mu$

$$\bar{X} - t_{\alpha(2)} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2)} \cdot s_{\bar{X}}$$
Confidence Intervals

*If* $H_0$ *is true then:*

$$
\Pr\left[-t_{\alpha(2)} \leq \frac{\bar{X} - \mu}{s_{\bar{X}}} \leq t_{\alpha(2)}\right] = 1 - \alpha
$$

$$
\Pr\left[-t_{\alpha(2)} \cdot s_{\bar{X}} \leq \bar{X} - \mu \leq t_{\alpha(2)} \cdot s_{\bar{X}}\right] = 1 - \alpha
$$

$$
\Pr\left[-\bar{X} - t_{\alpha(2)} \cdot s_{\bar{X}} \leq -\mu \leq -\bar{X} + t_{\alpha(2)} \cdot s_{\bar{X}}\right] = 1 - \alpha
$$

$$
\Pr\left[\bar{X} + t_{\alpha(2)} \cdot s_{\bar{X}} \geq \mu \geq \bar{X} - t_{\alpha(2)} \cdot s_{\bar{X}}\right] = 1 - \alpha
$$

$$
\Pr\left[\bar{X} - t_{\alpha(2)} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2)} \cdot s_{\bar{X}}\right] = 1 - \alpha
$$

Note that the probability $1 - \alpha$ on the right hand side of the equation never changes as we manipulate the inequality. This is because each step leaves the inequality in a form equivalent to:

$$
-t_{\alpha(2)} \leq \frac{\bar{X} - \mu}{s_{\bar{X}}} \leq t_{\alpha(2)}
$$
Confidence Interval Example

\( \alpha = 0.05 \)

So if we were to resample the study population many times, 95% of the sample means would lie in the interval:

\[
\bar{X} - t_{0.05(2)} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{0.05(2)} \cdot s_{\bar{X}}
\]
What happens if confidence interval overlaps with 0?
What happens if confidence interval overlaps with 0?

If confidence interval overlaps with 0, μ is not significantly different from 0.
If confidence interval overlaps with 0, then:

\[
\overline{X} - t_{\alpha(2)} \cdot s_{\overline{X}} \leq 0 \leq \overline{X} + t_{\alpha(2)} \cdot s_{\overline{X}}
\]

Subtract \( \overline{X} \) from all parts of the inequality

\[
-t_{\alpha(2)} \cdot s_{\overline{X}} \leq -\overline{X} + 0 \leq t_{\alpha(2)} \cdot s_{\overline{X}}
\]

Divide through by \( s_{\overline{X}} \)

\[
-t_{\alpha(2)} \leq -\frac{\overline{X}}{s_{\overline{X}}} + 0 \leq t_{\alpha(2)}
\]

Multiply inequality by \(-1\).
Change direction inequality.

\[
t_{\alpha(2)} \geq \frac{\overline{X} - 0}{s_{\overline{X}}} \geq -t_{\alpha(2)}
\]

Flip around

\[
-t_{\alpha(2)} \leq \frac{\overline{X} - 0}{s_{\overline{X}}} \leq t_{\alpha(2)}
\]

If confidence interval overlaps with 0 it implies mean \( \mu \) not significantly different from 0.
Note that final inequality is just what we get when \( \mu \) not significantly different from 0

\[
\begin{align*}
H_0 & : \mu = 0 \\
H_1 & : \mu \neq 0 \\
\bar{X} - \mu & \quad \bar{X} - 0 = \bar{X} \\
s_\bar{X} & \\
\end{align*}
\]

For lack of significance:

\[
- t_{\alpha(2)} \leq \frac{\bar{X} - 0}{s_\bar{X}} \leq t_{\alpha(2)}
\]

This is the final inequality on last page

Remember we use \( \mu \) value for \( H_0 \)
## Confidence interval example

\[
\bar{X} - t_{\alpha(2),\nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2),\nu} \cdot s_{\bar{X}}
\]

\[
\bar{X} = 4.3 \quad s = 3.1 \quad \nu = 9 \quad s_{\bar{X}} = ?
\]

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>CI %</th>
<th>(t_{\alpha(2),\nu})</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Confidence interval example

\[
\bar{X} - t_{\alpha(2),\nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2),\nu} \cdot s_{\bar{X}}
\]

\[
\bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0
\]

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>CI %</th>
<th>(t_{\alpha(2),\nu})</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE B.3  Critical Values of the $t$ Distribution

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\alpha(2)$: 0.50</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha(1)$: 0.25</td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
<td>0.005</td>
<td>0.0025</td>
<td>0.001</td>
<td>0.0005</td>
</tr>
<tr>
<td>3</td>
<td>0.765</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
<td>10.215</td>
<td>12.924</td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
<td>7.173</td>
<td>8.610</td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>6.869</td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
<td>5.208</td>
<td>5.959</td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.408</td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>1.383</td>
<td>1.833</td>
<td><strong>2.262</strong></td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.781</td>
</tr>
<tr>
<td>10</td>
<td>0.700</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
<td>4.144</td>
<td>4.587</td>
</tr>
<tr>
<td>11</td>
<td>0.697</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
<td>4.025</td>
<td>4.437</td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
<td>3.930</td>
<td>4.318</td>
</tr>
<tr>
<td>13</td>
<td>0.694</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.372</td>
<td>3.852</td>
<td>4.221</td>
</tr>
<tr>
<td>14</td>
<td>0.692</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
<td>3.787</td>
<td>4.140</td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
<td>3.733</td>
<td>4.073</td>
</tr>
</tbody>
</table>
Confidence interval example

$$\bar{X} - t_{\alpha(2), \nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2), \nu} \cdot s_{\bar{X}}$$

$$\bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0$$

<table>
<thead>
<tr>
<th>22</th>
<th>CI %</th>
<th>$t_{\alpha(2), \nu}$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td>2.262</td>
<td>$4.3 - 2.262 \times 1.0$</td>
<td>$= 2.038$</td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Confidence interval example

\[
\bar{X} - t_{\alpha(2),\nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2),\nu} \cdot s_{\bar{X}}
\]

\[
\bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0
\]

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>CI %</th>
<th>(t_{\alpha(2),\nu})</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td>2.262</td>
<td>4.3 - 2.262 \times 1.0 = 2.038</td>
<td>4.3 + 2.262 \times 1.0 = 6.562</td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**t-test table**

**TABLE B.3  Critical Values of the t Distribution**

<table>
<thead>
<tr>
<th>ν</th>
<th>0.50</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.765</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
<td>10.215</td>
<td>12.924</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
<td>7.173</td>
<td>8.610</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>6.869</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
<td>5.208</td>
<td>5.959</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.408</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.781</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.700</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
<td>4.144</td>
<td>4.587</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.697</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
<td>4.025</td>
<td>4.437</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
<td>3.930</td>
<td>4.318</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.694</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.372</td>
<td>3.852</td>
<td>4.221</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.692</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
<td>3.787</td>
<td>4.140</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
<td>3.733</td>
<td>4.073</td>
<td></td>
</tr>
</tbody>
</table>
## Confidence interval example

\[
\bar{X} - t_{\alpha(2),\nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2),\nu} \cdot s_{\bar{X}}
\]

\[
\bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0
\]

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>CI %</th>
<th>(t_{\alpha(2),10})</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td>2.262</td>
<td>4.3 - 2.262\times 1.0 =</td>
<td>4.3 + 2.262\times 1.0 =</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.038</td>
<td>6.562</td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td>3.250</td>
<td>4.3 - 3.250\times 1.0 =</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Confidence interval example

\[ \bar{X} - t_{\alpha(2),\nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2),\nu} \cdot s_{\bar{X}} \]

\( \bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0 \)

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>CI %</th>
<th>( t_{\alpha(2),\nu} )</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td>2.262</td>
<td>( 4.3 - 2.262 \times 1.0 = 2.038 )</td>
<td>( 4.3 + 2.262 \times 1.0 = 6.562 )</td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td>3.250</td>
<td>( 4.3 - 3.250 \times 1.0 = 1.05 )</td>
<td>( 4.3 + 3.250 \times 1.0 = 7.55 )</td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B
Statistical Tables and Graphs

**TABLE B.3** Critical Values of the *t* Distribution

<table>
<thead>
<tr>
<th>υ</th>
<th>α(2): 0.50</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.002</th>
<th>0.001</th>
<th>0.005</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.765</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
<td>10.215</td>
<td>12.924</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
<td>7.173</td>
<td>8.610</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>6.869</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
<td>5.208</td>
<td>5.959</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.700</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
<td>4.144</td>
<td>4.587</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.697</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
<td>4.025</td>
<td>4.437</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.695</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
<td>3.930</td>
<td>4.318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.694</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.372</td>
<td>3.852</td>
<td>4.221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.692</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
<td>3.787</td>
<td>4.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
<td>3.733</td>
<td>4.073</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Confidence interval example

\[ \bar{X} - t_{\alpha(2), \nu} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2), \nu} \cdot s_{\bar{X}} \]

\[ \bar{X} = 4.3 \quad s = 3.16 \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.16}{\sqrt{10}} \approx 1.0 \]

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>CI %</th>
<th>( t_{\alpha(2),10} )</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>95%</td>
<td>2.262</td>
<td>[ 4.3 - 2.262 \times 1.0 = 2.038 ]</td>
<td>[ 4.3 + 2.262 \times 1.0 = 6.562 ]</td>
</tr>
<tr>
<td>0.01</td>
<td>99%</td>
<td>3.250</td>
<td>[ 4.3 - 3.250 \times 1.0 = 1.05 ]</td>
<td>[ 4.3 + 3.250 \times 1.0 = 7.55 ]</td>
</tr>
<tr>
<td>0.001</td>
<td>99.9%</td>
<td>4.781</td>
<td>[ 4.3 - 4.781 \times 1.0 = -0.481 ]</td>
<td>[ 4.3 + 4.781 \times 1.0 = 9.081 ]</td>
</tr>
</tbody>
</table>
Experimental Design
Experiment I
Test drug to see if it causes allergic reaction.

• Inject right arm of one group with drug. Compare inflammation in right arm to second group.
Experimental Design

Experiment II

Test drug to see if it causes allergic reaction.

• Inject right arm of one group with drug.
• Inject right arm of 2nd group with placebo.
• Compare inflammation in right arm among groups.
Experimental Design
Experiment III

Test drug to see if it causes allergic reaction.

- One group only.
- Inject right arm of each subject with drug.
- Inject left arm of each subject with placebo.
- Compare inflammation in right and left arms of each subject.
Experimental Design

Experiment III

Test drug to see if it causes allergic reaction.

- One group only.
- Inject right arm of each subject with drug.
- Inject left arm of each subject with placebo.
- Compare inflammation in right and left arms of each subject.

*This is called a paired comparison.*
Experimental Design  
Experiment IV  

Test drug to see if it causes allergic reaction.  

• One group only.  
• For half of subjects inject right arm with drug and left arm with placebo.  
• For other half do the reverse.  
• Assign subjects to these groups (right drug; left placebo) and (right placebo; left drug) randomly.  

*This is still a paired comparison.*
Other Examples of Paired Sample Tests

- Before and after measurements of patients undergoing some treatment.
  - Comparisons of crop yields under two fertilizer regimes in similar plots within fields.
  - Comparison of automobile gas mileage for cars using two types of gasoline. Assemble a fleet of cars. Measure mileage on each car using both fuels. Then compare mileage for each car.
Paired Sample Tests

• We could apply the typical $t$-test and compare means in two samples.

• However, each experimental subject will have their own idiosyncrasies that will inflate $\frac{1}{\sigma^2} \left( \bar{X}_1 - \bar{X}_2 \right)^2$ reducing the power of the test.

• It’s more efficient to compare results for samples 1 and 2 within each subject.
Paired Sample Tests

- $H_0 : \mu_d = \mu_0$ *null hypothesis*
- $H_A : \mu_d \neq \mu_0$ *alternative hypothesis*

The test statistic is:

$$t = \frac{\bar{d} - \mu_0}{S_{\bar{d}}}$$
Paired Sample Tests

Where:

\[ s_d^2 = \frac{s^2}{n} \]

and

\[ s_d^2 = \text{var}(d) \]

\[ \nu = n - 1 \text{ degrees of freedom} \]
Paired Sample Tests

Data from EXAMPLE 9.2. Text page 163.

<table>
<thead>
<tr>
<th>Plot ($j$)</th>
<th>New Fertilizer ($X_{1j}$)</th>
<th>Old Fertilizer ($X_{2j}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2250</td>
<td>1920</td>
</tr>
<tr>
<td>2</td>
<td>2410</td>
<td>2020</td>
</tr>
<tr>
<td>3</td>
<td>2260</td>
<td>2060</td>
</tr>
<tr>
<td>4</td>
<td>2200</td>
<td>1960</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

*Each subject serves as its own control.*
Paired Sample Tests

Data from EXAMPLE 9.2. Text page 163.

<table>
<thead>
<tr>
<th>Plot ($j$)</th>
<th>New Fertilizer ($X_{1j}$)</th>
<th>Old Fertilizer ($X_{2j}$)</th>
<th>Difference ($d_j$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2250</td>
<td>1920</td>
<td>330</td>
</tr>
<tr>
<td>2</td>
<td>2410</td>
<td>2020</td>
<td>390</td>
</tr>
<tr>
<td>3</td>
<td>2260</td>
<td>2060</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>2200</td>
<td>1960</td>
<td>240</td>
</tr>
</tbody>
</table>

Each subject serves as its own control.
Paired Sample Tests

- $H_0: \mu_d \leq \mu_0 = 250$ null hypothesis
- $H_A: \mu_d > \mu_0 = 250$ alternative hypothesis

One tailed or two-tailed?
Paired Sample Tests

\[ t = \frac{\bar{d} - \mu_0}{s_d} = \frac{\bar{d} - 250}{s_d} \]

\[ \bar{d} = 295.6 \quad s_d = 80.6 \quad s_{\bar{d}} = 26.9 \]

\[ t = 1.695 \quad t_{0.05(2),8} = 1.860 \]
Type I Error:
Type I error is the mistake we make when we mistakenly reject the null hypothesis when it is true.

What is the probability of Type I Error?
Type I Error

The probability of Type I error is just $\alpha$, the significance level, the area of the rejection region.
Type I & II Error

Type II Error:

A Type II error occurs when we fail to reject the null hypothesis when it is false i.e. when the alternative, $H_1$, is true.

We typically call the probability of Type II Error $\beta$. 
Type I Error:

Type I error is the mistake we make when we mistakenly reject the null hypothesis when it is *true.*
Type I & II Error

Type II Error:

A Type II error occurs when we fail to reject the null hypothesis when it is false i.e. when the alternative, $H_1$, is true.

We typically call the probability of Type II Error, $\beta$
Type II Error

$H_0$ True

$H_1$ True

Type II Error $\beta$
## Type I & II Error

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ Rejected</td>
<td>Type I Error</td>
<td>No Error</td>
</tr>
<tr>
<td>$H_0$ Not Rejected</td>
<td>No Error</td>
<td>Type II Error</td>
</tr>
</tbody>
</table>
Statistical Power

The Power of a statistical test is:

$$1 - \Pr(\text{Type II Error})$$

$$= 1 - \beta$$

The power is the probability that we will correctly reject $H_0$ when it is false.
Statistical Power

$H_0$ True

$H_0$ False

Power

Type II Error
One-Tailed Tests

- $H_0: \mu = 0$ null hypothesis
- $H_1: \mu > 0$ alternative hypothesis

So we are considering a more restricted alternative hypothesis now.
One-tailed test

\[ \alpha = 0.05 \]
This graph shows the rejection regions for both one- and two-tailed tests, so they can be compared.
Sample of 10,000,000 normal random variables from random number generator.

\( \mu = 0 \)

\( \sigma = 1 \)

Histogram bar width = 1

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Bar width = 1

Probability $x$ falls in any given bar is equal to area of that bar (height $\times$ width).

Probability $x > 0$ is represented by area shaded yellow.
Bar width = 1/2
Probability Density Function – PDF

Bar width = 1/4
Bar width = 1/8
Bar width = 1/16
Probability Density Function – PDF

Height of the curve $f(x)$ at $x$ gives relative likelihood the value $x$ will occur.

You can think of the PDF as a histogram with infinitely thin bars.
\[ \Pr(x \leq -1) \]

\[ \Pr(x \leq 1) = \sum_{i=-\infty}^{n} f(x_i) \Delta x \]

Where \( f(x_i) = \) height of histogram bar \( i \).

\( \Delta x = \) bar width.

All we are doing is summing the area of the bars.
\[ \Pr(x \leq -1) \]

\[
\lim_{\Delta x \to 0} f(x_i) \Delta x \to f(x)dx
\]

\[
\Pr(x \leq 1) = \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\]
Cumulative Distribution Function (CDF)

\[ F(x) = \int_{-\infty}^{x} f(\xi) \, d\xi \]

\[ = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} \, d\xi \]
Calculating probabilities for the normal distribution

\[ \Pr(x \leq 1) = F(-1) = \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]
Pr\((x \leq 1)\)

Pr\((x \leq 1) = F(1)\)

\[
= \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]
Calculating probabilities for the normal distribution

\[ F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \]

\[ F(1) \]

\[ \Pr(x \leq 1) = F(1) \]
Calculating probabilities for the normal distribution

\[ \Pr(-1 \leq x \leq 1) \]

\[ \Pr(-1 \leq x \leq 1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]
Calculating probabilities for the normal distribution

\[ \Pr(-1 \leq x \leq 1) = F(1) - F(-1) \]
Formula for probability

\[ \Pr(-1 \leq x \leq 1) = F(1) - F(-1) \]

\[ F(1) - F(-1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]
More generally:

$$\Pr(a \leq x \leq b) = F(b) - F(a)$$
\[ \Pr(-\infty \leq x \leq \infty) = F(\infty) - F(-\infty) = \int_{-\infty}^{\infty} f(x) \, dx \]
\[ \Pr(-\infty \leq x \leq \infty) = F(\infty) - F(-\infty) = \int_{-\infty}^{\infty} f(x) \, dx = 1 \]
Expectations

\[ E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx \]  
Expectation of a function \( g(x) \)

\[ E[x] = \int_{-\infty}^{\infty} x f(x) \, dx \]  
Mean

\[ E[(x - E[x])^2] = \int_{-\infty}^{\infty} (x - E[x])^2 f(x) \, dx \]  
Variance
Factorial Analysis of Variance

• One-way ANOVA for Single Factor A

\[ X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \]

• Two-way Factorial ANOVA for factors A & B

\[ X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + \varepsilon_{ijk} \]
Factorial Analysis of Variance (Example 12.1 Zar)
Effect of Hormone and Sex on serum Ca levels

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B_1$</td>
<td>$B_2$</td>
</tr>
<tr>
<td>No Hormone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td>$X_{11k=1..n}$</td>
<td>$X_{12k=1..n}$</td>
</tr>
<tr>
<td>Hormone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_2$</td>
<td>$X_{21k=1..n}$</td>
<td>$X_{22k=1..n}$</td>
</tr>
<tr>
<td>$\bar{X}_{.1}$</td>
<td>$\bar{X}_{.2}$</td>
<td>$\bar{X}$</td>
</tr>
</tbody>
</table>

\[ X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + \varepsilon_{ijk} \]
## Factorial Analysis of Variance (Example 12.1 Zar)

### Effect of Hormone and Sex on serum Ca levels

<table>
<thead>
<tr>
<th>Levels</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Hormone</td>
<td>$X_{11k}=1..n$</td>
<td>$X_{12k}=1..n$</td>
</tr>
<tr>
<td>Hormone</td>
<td>$X_{21k}=1..n$</td>
<td>$X_{22k}=1..n$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A&lt;sub&gt;1&lt;/sub&gt;</th>
<th>B&lt;sub&gt;1&lt;/sub&gt;</th>
<th>B&lt;sub&gt;2&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_{1.}$</td>
<td>$\bar{X}_{1.}$</td>
<td>$\bar{X}_{1.}$</td>
</tr>
<tr>
<td>$\bar{X}_{.1}$</td>
<td>$\bar{X}_{.2}$</td>
<td>$\bar{X}$</td>
</tr>
</tbody>
</table>

### Model Equation

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + e_{ijk}$$
Factorial Analysis of Variance

SS Total = \[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (X_{ijk} - \bar{X})^2 \]

SS Cells = SSC = \[ \sum_{i=1}^{a} \sum_{j=1}^{b} n(\bar{X}_{ij} - \bar{X})^2 \]

\[ X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + e_{ijk} \]

\( a, b \) are numbers of levels of factors A and B

(Treatments)
Factorial Analysis of Variance

\[ SS \text{ Factor A} = \sum_{i=1}^{a} bn(\bar{X}_{i \cdot} - \bar{X})^2 \]

\[ SS \text{ Factor B} = \sum_{j=1}^{b} an(\bar{X}_{\cdot j} - \bar{X})^2 \]

\[ SS (A \times B) = SSC - SSA - SSB \]

\[ X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + \varepsilon_{ijk} \]

\( a, b \) are numbers of levels of factors A and B

Interaction
Factorial Analysis of Variance

\[
SS \text{ Total} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (X_{ijk} - \bar{X})^2
\]

\[
SS \text{ Error} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (X_{ijk} - \bar{X}_{ij})^2 \quad \text{(Within Cells)}
\]
# Factorial Analysis of Variance

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>$a \times b \times (n - 1)$</td>
</tr>
<tr>
<td>Factor A</td>
<td>$(a-1)$</td>
</tr>
<tr>
<td>Factor B</td>
<td>$(b-1)$</td>
</tr>
<tr>
<td>$A \times B$</td>
<td>(DF Factor A) × (DF Factor B)</td>
</tr>
<tr>
<td></td>
<td>$(a-1) \times (b-1)$</td>
</tr>
<tr>
<td>Cells</td>
<td>$a \times b - 1$</td>
</tr>
<tr>
<td>Total</td>
<td>$(a \times b \times n) - 1 = N - 1$</td>
</tr>
</tbody>
</table>
F test – factor A

\[ F = \frac{\text{factor A MS}}{\text{error MS}} \]

\[ F = \frac{\text{factor B MS}}{\text{error MS}} \]

\[ F = \frac{A \times B \text{ MS}}{\text{error MS}} \]

- \( a - 1 \) df
- \( b - 1 \) df
- \( (a - 1) \times (b - 1) \) df
Suppose we have two factors A and B that we want to investigate. Let’s compare the efficiency of sequential one-way and factorial ANOVAs. Let’s assume for the moment that there’s no interaction between the factors. Also let’s assume there are two levels for both factors A and B.
Factorial Analysis of Variance

- Factorial ANOVA is cheaper and easier to conduct than sequential one-way ANOVAs.
- Can estimate the error variance to same precision with half the sample size.
- With three factors, factorial ANOVA needs only $\frac{1}{2^2} = \frac{1}{4}$ the sample size, with four factors $\frac{1}{2^3} = \frac{1}{8}$ the sample size and so forth. So efficiency increases with complexity.
Factorial Analysis of Variance

• In addition, factorial ANOVA can estimate interactions; sequential one-way ANOVAs cannot.

• Sequential one-way ANOVAs also have Type I error inflation problem that pair wise t-tests did.
Interactions (Fig. 12.2 Zar)

Mean of Response Variable

Level of Factor A
Figure 12.2 Means in a two-factor ANOVA, showing various effects of the two factors and their interaction. (a) No effect of factor A, small effect of factor B (and if there were no effect of B the two lines would coincide), and no interaction of A and B. (b) Large effect of factor A, small effect of factor B, and no interaction (which is the situation in Fig. 12.1). (c) No effect of A, large effect of B, and no interaction. (d) Large effect of A, large effect of B, and no interaction. (e) No effect of A, no effect of B, but interaction between A and B. (f) Large effect of A, no effect of B, with slight interaction. (g) No effect of A, large effect of B, with large interaction. (h) Effect of A, large effect of B, with large interaction.
Fixed vs. Random Effects

- Fixed – levels of treatments determined by investigator.
- Random – levels of treatments sampled from typical variation in treatment variable.
- Random effect example: Gas mileage.
  - Tires: standard, radial
  - Car: five models sampled at random
Fixed vs. Random Effects – Examples

1. Four different cancer drugs are administered to a group of patients in four different age groups.

2. The effects of mercury, cadmium and lead pollution on marsh invertebrates are investigated at a set of polluted sites chosen randomly surrounding Arthur Kill.

3. You study insect damage by the corn rootworm on different varieties of corn in four agricultural plots. There are too many varieties of corn to test all of them. Hence, you select 5 from the total and see how much the damage varies among these varieties.

4. You test the impact of 4 pesticides (compounds 1-4) and 5 fertilizers (chemicals 1-5) on cotton production.

5. You investigate three types of teaching methods in a representative group of public schools in NYC.
Fixed vs. Random Effects

- Model I ANOVA: all treatments fixed.
- Model II ANOVA: all treatments random.
- Model III ANOVA: both fixed and random effects
Mean Squares

• Differs between Models I, II and III
Mean Squares

- Differs between Models I, II and III

**Section 12.1 Two-Factor Analysis of Variance with Equal Replication**

**TABLE 12.3** Computation of the $F$ Statistic for Tests of Significance in a Two-Factor ANOVA with Replication

<table>
<thead>
<tr>
<th>Hypothesized effect</th>
<th>Model I (factors $A$ and $B$ both fixed)</th>
<th>Model II (factors $A$ and $B$ both random)</th>
<th>Model III (factor $A$ fixed: factor $B$ random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor $A$</td>
<td>factor $A$ MS</td>
<td>factor $A$ MS</td>
<td>factor $A$ MS</td>
</tr>
<tr>
<td></td>
<td>error MS</td>
<td>$A \times B$ MS</td>
<td>$A \times B$ MS</td>
</tr>
<tr>
<td>Factor $B$</td>
<td>factor $B$ MS</td>
<td>factor $B$ MS</td>
<td>factor $B$ MS</td>
</tr>
<tr>
<td></td>
<td>error MS</td>
<td>$A \times B$ MS</td>
<td>$A \times B$ MS</td>
</tr>
<tr>
<td>$A \times B$ interaction</td>
<td>$A \times B$ MS</td>
<td>$A \times B$ MS</td>
<td>$A \times B$ MS</td>
</tr>
<tr>
<td></td>
<td>error MS</td>
<td>error MS</td>
<td>error MS</td>
</tr>
</tbody>
</table>

MS (see Table 12.3). We test for the interaction effect, as before, by $F = \text{interaction MS/error MS}$, and it is generally not useful to declare factor effects significant if there is a significant interaction effect. The Model II ANOVA for designs with more than two factors will be discussed in Chapter 15.
Randomized Complete Blocks Design

- The conceptual extension of the paired $t$-test.
- Example: growth of guinea pigs on 4 diets in 4 rearing rooms. Each room is a block.
- Conditions vary among pens. Climate control could shift. This design compares growth among diets within each block.
Randomized Complete Blocks Design

I

1 2
3 4

II

4 1
3 2

III

3 2
4 1

IV

1 3
4 2

*Each pen has all four diets represented*
## Randomized Complete Blocks Design (Example 12.4 Zar)

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Diets</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>5.3</td>
<td>4.9</td>
<td>8.8</td>
</tr>
<tr>
<td>2</td>
<td>9.9</td>
<td>5.7</td>
<td>7.6</td>
<td>8.9</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
<td>4.7</td>
<td>5.5</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>3.5</td>
<td>2.8</td>
<td>3.3</td>
</tr>
<tr>
<td>5</td>
<td>10.3</td>
<td>7.7</td>
<td>8.4</td>
<td>9.1</td>
</tr>
</tbody>
</table>
Randomized Complete Blocks Design

- Analyzed as a mixed factorial model (Model II) without replication – notice there’s just one datum per cell.
- Treatment is a fixed effect (e.g. diet in Example 12.4).
- Block is a random effect.
One additional point on the Randomized Complete Blocks Design (Example 12.4 Zar)

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Diets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>9.9</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
</tr>
<tr>
<td>5</td>
<td>10.3</td>
</tr>
</tbody>
</table>
Estimated Marginal Means of WEIGHT_G

This looks like an interaction, but...
### Randomized Complete Blocks Design

Tests of Between-Subjects Effects

**Dependent Variable: Weight Gain**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Hypothesis</td>
<td>912.601</td>
<td>1</td>
<td>912.601</td>
<td>58.269</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>62.647</td>
<td>4</td>
<td>15.662</td>
<td></td>
</tr>
<tr>
<td>Diet</td>
<td>Hypothesis</td>
<td>27.426</td>
<td>3</td>
<td>9.142</td>
<td>11.825</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>9.277</td>
<td>12</td>
<td>.773</td>
<td></td>
</tr>
<tr>
<td>Block</td>
<td>Hypothesis</td>
<td>62.647</td>
<td>4</td>
<td>15.662</td>
<td>20.259</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>9.277</td>
<td>12</td>
<td>.773</td>
<td></td>
</tr>
<tr>
<td>Diet * Block</td>
<td>Hypothesis</td>
<td>9.277</td>
<td>12</td>
<td>.773</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>.000</td>
<td>0</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

- a. MS(Block)
- b. MS(Diet * Block)
- c. MS(Error)

No F-test for interaction
Randomized Complete Blocks Design (Example 12.4)

<table>
<thead>
<tr>
<th>Blocks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5.3</td>
<td>4.9</td>
<td>8.8</td>
</tr>
<tr>
<td>2</td>
<td>9.9</td>
<td>5.7</td>
<td>7.6</td>
<td>8.9</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
<td>4.7</td>
<td>5.5</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>3.5</td>
<td>2.8</td>
<td>3.3</td>
</tr>
<tr>
<td>5</td>
<td>10.3</td>
<td>7.7</td>
<td>8.4</td>
<td>9.1</td>
</tr>
</tbody>
</table>

\[
X_{ij} = \mu + A_i + B_j + (A \times B)_{ij} + e_{ij}
\]
Correlation

$H_0$: $X$ and $Y$ unrelated.

$H_1$: $X$ and $Y$ have linear relationship.

\[
    r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} = \frac{\text{cov}(X,Y)}{S_X S_Y}
\]
Correlation

\[ H_0: X \text{ and } Y \text{ unrelated.} \]

\[ H_1: X \text{ and } Y \text{ have linear relationship.} \]

\[ r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} = \frac{\text{cov}(X,Y)}{s_X s_Y} \]

\[ \text{cov}(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \]
Correlation

\[ r \text{ ranges from } -1 \text{ to } 1. \]

\[
r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}
\]

\[-1 \leq r \leq 1\]
\[
 r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}
\]

\[
 r = 0.87
\]
The correlation coefficient $r$ is calculated as:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

When evaluated, we find $r = -0.87$. This indicates a strong negative correlation between the variables $X$ and $Y$. The scatter plot shows a clear linear relationship with the data points clustering around the line of best fit.
$r = 0.53$
$r = 0.99$
$r = 1.00$
\[ r = 0.00 \]
Hypothesis Testing on Correlation Coefficient

$H_0$: $X$ and $Y$ are unrelated.  $H_1$: $X$ and $Y$ are correlated.

$$t = \frac{r}{s_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Reject if:  $|t| \geq t_{\alpha(2),(n-2)}$
Linear Regression

**Objective:** fit a line to data.
Fitting a line to this data by eye might give a pretty accurate result.
It would be much harder to fit a line by eye to these data. Especially one that would be accurate if we extrapolated beyond the bounds of the data.
Variation in $Y$ is what we are trying to explain.
We explain some of the variation in $Y$ with $X$. 
Residual is variation in $Y$ not explained by model (linear equation).
Linear Regression

Simple linear model:

\[ Y_i = b_1 X_i + b_0 + \varepsilon_i \]

- \( Y_i \) = dependent (response) variable
- \( X_i \) = independent (explanatory, predictor) variable
- \( b_1 \) = slope
- \( b_0 \) = \( Y \) intercept
- \( \varepsilon_i \) = residual
Linear Regression

Least Squares

\[ Y_i = b_1 X_i + b_0 + \varepsilon_i \]

\[ \hat{Y}_i = b_1 X_i + b_0 \]

Here \( \hat{Y} \) is the value of \( Y \) predicted by the ‘regression’ line.

\[ Y_i - \hat{Y}_i = \varepsilon_i \]

\[ RSS = \sum_{i=1}^{n} \left( Y_i - \hat{Y}_i \right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 \]

We minimize this to find the line that best fits the data.

\( RSS \) stands for ‘Residual Sum of Squares’
We explain some of the variation in $Y$ with $X$. 
Linear Regression

This gives:

\[ \hat{b}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)} \]

\[ \hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \]

Here, \( \hat{b}_0, \hat{b}_1 \) are estimates of the ‘true’ intercept and slope parameters \( \beta_1 \) and \( \beta_0 \).
Variance and Covariance

\[
\text{cov}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}
\]

\[
\text{var}(X) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}
\]
Sums of Squares in Regression

Total $SS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

Regression $SS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$

Residual $SS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$

$= \text{Total } SS - \text{Regression } SS$
Each Sum of Squares has Degrees of Freedom

<table>
<thead>
<tr>
<th></th>
<th>$SS$</th>
<th>$DF$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>$n - 1$</td>
</tr>
<tr>
<td>Regression</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>$n - 2$</td>
</tr>
</tbody>
</table>
Linear Regression – Hypothesis Testing – *F*-test

*H₀*: X and Y unrelated.

*H₁*: X and Y have linear relation.

\[ F = \frac{\text{Regression } MS}{\text{Residual } MS} \]

Regression \( MS = \frac{\text{Regression } SS}{1} \)

Residual \( MS = \frac{\text{Residual } SS}{n - 2} \)
Linear Regression – Hypothesis Testing – $t$-test

$H_0$: $X$ and $Y$ unrelated.

$H_1$: $X$ and $Y$ have linear relationship.

$$t = \frac{\hat{b}_1 - \beta_1}{s_{\hat{b}_1}},$$

$\beta$ is the hypothesized value of $b$

$$s_{\hat{b}_1}^2 = \frac{\text{Residual } MS}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

In the case of simple linear regression a $t$-test that $\hat{b}$ is significantly different from 0 ($\beta = 0$) is equivalent to the $F$ test. This is not true for multiple and polynomial regression however.
$Y \sim N(0,1)$

No Slope
Linear Regression – Model Fit

How good does the model fit the data? The coefficient of determination:

\[
r^2 = \frac{\text{Regression } SS}{\text{Total } SS} = \frac{n \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})^2}{n \sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}
\]

Gives the proportion of the variation in \( Y \) that is explained by \( X \).

\( r \) is called the correlation coefficient. More about that soon.
Linear Regression – Hypothesis Testing – Model Fit

$r^2$ ranges from 0 (X and Y not related) to 1 (perfect linear fit).

$$r^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}}$$

$r^2$ is often called the coefficient of determination.
Linear Regression – Hypothesis Testing – $F$-test

$H_0$: $X$ and $Y$ unrelated.

$H_1$: $X$ and $Y$ have linear relationship.

$$F = \frac{\text{Regression } MS}{\text{Residual } MS}$$

Regression $MS = \frac{\text{Regression } SS}{1}$

Residual $MS = \frac{\text{Residual } SS}{n-2}$
confidence interval = statistic \pm (t)(SE of statistic)

We can place a confidence interval around the value of the regression slope:

\[ \hat{b} \pm t_{\alpha(2),(n-2)} \times S_{\hat{b}} \]

**What does it mean if these confidence intervals overlap with 0?**
$Y \sim N(0,1)$

No Slope
Linear Regression – Confidence Intervals

confidence interval = statistic ± (t)(SE of statistic)

We can also calculate a confidence interval around the predicted value of $Y_i$, $\hat{Y}_i$ – i.e. around the regression line.

$$\hat{Y}_i \pm t_{\alpha(2),(n-2)} s_{\hat{Y}}$$

$$s_{\hat{Y}} = \sqrt{s^2_{Y\cdot X} \left[ \frac{1}{n} + \frac{(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right]}$$

$$s^2_{Y\cdot X} = \text{residual } MS$$
Repeated Measures Analysis of Variance

• Situations in which biologists would make repeated measurements on same individual
  – Change in a trait or variable measured at different times
    • E.g., clutch size variation over time
    • Change in survivorship over time among populations
  – Individual is exposed to different level of a same treatment
    • E.g., same plants exposed to varying [CO₂]
How to Analyze Repeated Measures Designs

• Univariate and Multivariate Approaches
  – Univariate
    • Randomized Designs
    • Split-plot designs
  – Multivariate:
    • MANOVA
  – Mixed Model Analysis
    • GLMM (General Linear Mixed Model)
• Mixed Model Approach is preferred method
Advantages of Repeated Measures

• Recall that experimental design has goal of reducing error and minimizing bias
  – E.g., use randomized blocks
• In repeated measures individuals are “blocks”
• Assume within-subject variation lower than among subjects
• Advantage – can conduct complex designs with fewer experimental units
Basic Repeated Measures Design

• Completely Randomized Design (CRD)
• Data collected in a sequence of evenly spaced points in time
• Treatments are assigned to experimental units
  – I.e., subjects
• Two factors:
  – Treatment
  – Time
All repeated measures experiments are factorial.

**Treatment** is called the **between-subjects** factor. Levels change only between treatments. Measurements on the same subject represent the same treatment.

**Time** is called **within-subjects** factor. Different measurements on the same subject occur at different times.
Hypotheses

• How does the treatment mean change over time?

• How do treatment differences change over time?
What do hypotheses mean?

- Is there a **Time main effect**?

- Is there a **Treatment \times Time** interaction?
Why is Repeated Measures ANOVA unique?

• Problem involves the covariance structure
  – Particularly the error variance covariance structure

• ANOVA and MANOVA assume independent errors
  – All observations are equally correlated
  – However, in repeated measures design, adjacent observations are likely to be more correlated than more distant observations
Objectives of R ANOVA

• Compare treatment means over time
• Compare regression lines over time

• Critical to assess the covariance structure of the data
  – Assessing covariance structure is not the main interest
  – Assessing covariance structure required for obtaining valid inferences about the treatment means
Overview of Univariate Approaches

• Based on CRD or split-plot designs

• Hypothesis: do different treatment levels applied to same individuals have a significant effect

• CRD: Individual is the block
  – Blocking increases the precision of the experiment
  – Measurements made on different time periods comprise the *within-subject* factor
Univariate Approach

• Split-Plot Design:
  – Treatment factor corresponds to main-plot factor
    • I.e., between-subjects factor is main plot factor
  – Time factor is the sub-plot factor
    • I.e., within-subjects is the sub-plot factor

• Problem: in true split-plot design
  – Levels of sub-plot factor are randomly assigned
  – Equal correlation among responses in sub-plot unit
  – Not true in repeated measures design – measurements made at adjacent times are more correlated with one another than more distant measurements
Univariate Approach

- Assumptions must be made regarding the covariance structure for the within-subject factor
  - Circularity
    - Circular covariance matrix: difference between any two levels of within-subject factor has same constant value
  - Compound Symmetry
    - All variances are assumed to be equal
    - All covariances are assumed to be equal
  - Sphericity may be used to assess the circularity of covariance matrix
- One must test for these assumptions otherwise F-ratios are biased
Repeated Measures Model

- **Univariate Model**

\[ X_{ijklm} = \mu + \nu_i + \psi_{k(i)} + \tau_j + \nu\tau_{ij} + \psi\tau_{jk(i)} + \varepsilon_{m(ijk)} \]

- \( \mu \) - grand mean
- \( \nu_i \) - effect of treatment on response variable
- \( \psi_{k(i)} \) - subject effect nested within treatment
- \( \tau_j \) - Time effect
- \( \nu\tau_{ij} \) - Treatment x Time interaction
- \( \psi\tau_{jk(i)} \) - Subject x Time interaction
- \( \varepsilon \) - error term
  - m is a dummy subscript – indicates error is nested within individual observation
MANOVA Approach

• Successive response measurements made over time are considered correlated dependent variables
  – That is, response variables for each level of within-subject factor is presumed to be a different dependent variable

• MANOVA assumes there is an unstructured covariance matrix for dependent variables

• Entails using Profile Analysis
Concerns using MANOVA

- **Sample size requirements**
  - $N - M > k$
  - I.e., the number of samples (subjects) $(N)$ less the number of between-subjects treatment levels (groups) $(M)$ must be greater than the number of dependent variables

- **Low sample sizes have low power**
- **Power increases as ratio $n:k$ increases**
- **May have to increase $N$ and reduce $k$ to obtain reasonable analysis using MANOVA**
What does MANOVA test

• Performs a simultaneous analysis of response curves
  – Evaluates differences in shapes of response curves
  – Evaluates differences in levels of response curves

• Based on profile analysis – combines multivariate and univariate approaches
Profile Analysis

• Test that lines are parallel
• Test of lines equal elevation
• Test that lines are flat
Test of Assumptions: Univariate Approach

- Sphericity and Compound Symmetry
  - Mauchley’s Test for Sphericity
- Box (1954) found that F-ratio is positively biased when sphericity assumption is not met
  - Tend to reject falsely
- How far does covariance matrix deviate from sphericity?
  - Measured by $\varepsilon$
    - If sphericity is met, then $\varepsilon = 1$
- Adjustments for positive bias:
  - Greenhouse-Geiser
  - Huynh-Feldt condition:
Adjustments made to degree’s of freedom

• **Greenhouse-Geiser**
  – (k - 1), (k - 1)(n - 1) instead of 1, (n - 1)
  – Very conservative adjustment

• **Huynh-Feldt**
  – Better to estimate $\varepsilon$ and adjust df with the estimated $\varepsilon$.

• If $\varepsilon$ is above 0.7 then use the Huynh-Feldt correction
Repeated Measures Analysis as a Mixed Model

• Repeated measures analysis is a mixed model
• Why?
• First, we have a treatment, which is usually considered a fixed effect
• Second, the subject factor is a random effect
• Models with fixed and random effects are mixed models
  – A model with heterogeneous variances (more than one parameter in covariance matrix) is also a mixed model
Random Effects

- Random-effects are factors where the levels of the factor in experiment are a random sample from a larger population of possible levels.

- Models in which all factors are random are random effects, or nested, or hierarchical models.
Defining Mixed Models

- Recall the GLM
  \[ Y = X\beta + \varepsilon \]

- Assumptions:
  - \( \mathbb{E}[Y] = X\beta \)
  - \( \text{Var}[Y] = \text{var}[\varepsilon] = \sigma^2 I \)

- We have structures:
  - Mean
  - Variance
The mixed model

• GLMM is defined as:

\[ Y = X\beta + Z\nu + \varepsilon \]

• \( Y, X, \) and \( \beta \) are as in GLM

• \( Z \) is a known design matrix for the random effects

• \( \nu \) - vector of unknown random effects parameters

• \( \varepsilon \) - vector of unobserved random errors
The terms explained

• $X\beta$ denotes fixed effects
• $Z\nu$ denotes the random effects
• $\varepsilon$ denotes repeated measures effects
Assumptions of GLMM

- $\nu$ is $N_p(0, G)$  
  - i.e., multivariate normal with mean vector 0 and covariance matrix $G$
- $\varepsilon$ is $N_p(0, R)$  
  - i.e., multivariate normal with mean vector 0 and covariance matrix $R$ (repeated measures structure)
- $\nu$, $\varepsilon$ are uncorrelated
Assumptions

\[ E \begin{bmatrix} \nu \\ \varepsilon \end{bmatrix} = 0 \]

\[ \text{Var} \begin{bmatrix} \nu \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \]
Assumptions

\[ E[Y] = X\beta \]

\[ \text{var}[Y] = ZGZ' + R \]
GLM vs GLMM

- GLM is special case of GLMM
  - $Z=0$
  - $R=\sigma^2 I$
    - i.e., no (additional) random effects
    - Independent random errors
Why use Mixed Models to analyze Repeated Measure Designs?

• Can estimate a number of different covariance structures
  – Key because each experiment may have different covariance structure
  – Need to know which covariance structure best fits the random variances and covariance of data
SAS Mixed Repeated Measures Syntax

PROC MIXED data= <dataset> cl;

CLASS <independent classification variables>;

MODEL <dependent variable> = <fixed sources>/ddfm=satterth;

REPEATED/subject = <EU id>
    Type=<covariance structure> r rcorr;
SAS Mixed Model

- PROC MIXED … cl
- CLASS
- MODEL <dependent variable> = <fixed sources>

- cl requests confidence limits for variance covariance estimates
- Identifies variables used as sources of variation and subject option of REPEATED statement
- Specifies dependent variable and all fixed sources of variation (includes treatment, time and their interaction. The ddfm option computes the correct degrees of freedom for the various terms.
SAS Mixed Model

- **REPEATED/ subject = <EU id> type=<covariance structure> r rcorr;**
  - `subject` identifies the experimental unit in the data set which represents the repeated measure.
  - `type` identifies the covariance structure
  - `r` requests printing of the covariance matrix for the repeated measures
  - `rcorr` requests printing of the correlation matrix for the repeated measures
Covariance Structures:  Simple

- Equal variances along main diagonal
- Zero covariances along off diagonal
- Variances constant and residuals independent across time.
- The standard ANOVA model
- Simple, because a single parameter is estimated: the pooled variance
Covariance Structures: Unstructured

- Separate variances on diagonal
- Separate covariances on off diagonal
- Multivariate repeated measures
- Most complex structure
- Variance estimated for each time, covariance for each pair of times
- Need to estimate 10 parameters
- Leads to less precise parameter estimation

\[
\begin{align*}
\text{Time}_1 & \quad \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix} \\
\text{Time}_2 & \\
\text{Time}_3 & \\
\text{Time}_4 & 
\end{align*}
\]
Covariance Structures: compound symmetry

- Equal variances on diagonal; equal covariances along off diagonal (equal correlation)
- Simplest structure for fitting repeated measures
- Split-plot in time analysis
- Used for past 50 years
- Requires estimation of 2 parameters
Covariance Structures: First order Autoregressive

- Equal variances on main-diagonal
- Off diagonal represents variance multiplied by the repeated measures coefficient raise to increasing powers as the observations become increasingly separated in time. Increasing power means decreasing covariances. Times must be equally ordered and equally spaced.

Estimates 2 parameters

\[
\begin{bmatrix}
\sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 & \rho^3 \sigma^2 \\
\rho \sigma^2 & \sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 \\
\rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2 & \rho \sigma^2 \\
\rho^3 \sigma^2 & \rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2 \\
\end{bmatrix}
\]
Strategies for finding suitable covariance structures

• Run unstructured first
• Next run compound symmetry – simplest repeated measures structure
• Next try other covariance structures that best fit the experimental design and biology of organism
Criteria for Selecting “best” Covariance Structure

• Need to use model fitting statistics:
  – AIC – Akaike’s Information Criteria
  – SBC – Schwarz’s Bayesian Criteria

• Larger the number the better
  – Usually negative so closest to 0 is best

• Goal: covariance structure that is better than compound symmetry