Lecture 4 - Confidence Intervals

Theophanis Tsandilas
Normal sampling distribution

If the **sampling distribution** is normal. Then, the CI can be constructed as follows:

\[
CI = \hat{\mu} \pm \frac{z_{\alpha/2}}{2} \times \sigma_{\hat{\mu}}
\]

For a 95% CI, we take the \(0.025\)th percentile of the standard normal distribution.

standard normal (z) distribution
Normal sampling distribution

standard normal (z) distribution  non-standardized normal distribution

\[ z_{0.025} \]  \[ z_{0.975} = -z_{0.025} \]

\[ \pm \sigma_{\hat{\mu}} \times z_{0.025} \]
Example

Imagine that a researcher takes a random sample of 30 people and keeps them awake for 24 hours prior to taking an IQ test. The researcher finds an IQ score equal to $M = 94.6$.

A common assumption is that the standard deviation of IQ scores in populations is $SD = 15$. Based on this assumption, construct a 95% CI of the mean.

$$CI = \hat{\mu} \pm z_{\alpha/2} \times \sigma_{\hat{\mu}} \quad \implies \quad CI = 94.6 \pm 5.4$$

\[
\begin{align*}
94.6 & \pm 1.96 \\
\frac{15}{\sqrt{30}} & = 2.74
\end{align*}
\]
Problem

We rarely know the standard deviation of the original population
When the population standard deviation is unknown, we need to estimate it from the sample:

\[ \hat{\sigma}_\mu = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{\mu})^2}{n-1}} \]
When the **standard error is estimated from the sample**, then the resulting **standardized sampling distribution** is not a z distribution.

It is a **t distribution**
The t distribution (Student’s t)

When the population standard deviation $\sigma$ is unknown and is estimated from the unbiased variance estimate:

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

then, the resulting **standardized sample mean** has a t distribution with $\nu = n - 1$ degrees of freedom.

published by William Gosset (1908) under the pseudonym « Student »
The t distribution (Student’s t)

A random variable $X$ following a $t$ distribution is denoted as:

$$X \sim t(\nu)$$
R code

```r
> par(mfrow=c(1,2), mar = c(4,4,.5,.5), pty='s')
> curve(dnorm(x), xlim=c(-4,4), ylab="Probability density", lty=2)
> curve(dt(x, 1), xlim=c(-4,4), add = TRUE)
> legend(2.25, .35, legend=c(expression(italic(t)), expression(italic(z))), lty=c(1,2), bty='n')
> curve(dnorm(x), xlim=c(-4,4), ylab="Probability density", lty=2)
> curve(dt(x, 29), xlim=c(-4,4), add = TRUE)
> legend(2.25, .35, legend=c(expression(italic(t)), expression(italic(z))), lty=c(1,2), bty='n')
```
CIs using the $t$ distribution

In most cases, the population SD is unknown, so we use the $t$ distribution as the standardized sampling distribution of the mean:

$$\hat{\mu} \pm t_{\nu, \alpha/2} \times \hat{\sigma}_\hat{\mu}$$

where $t_{\nu, \alpha/2}$ is the $\alpha/2 \times 100$ percentile of the $t$ distribution with $\nu = n - 1$ degrees of freedom
CIs using the $t$ distribution

$t_5 (n = 6)$

$\pm t_{5,.025} = \pm 2.57$

$t_{29} (n = 30)$

$\pm t_{29,.025} = \pm 2.05$
R code

Calculate the 2.5 and 97.5 percentiles of the \( t \) distribution for different degrees of freedom \( \nu \)

```r
> qt(.025, 5)
[1] -2.570582
> qt(.975, 5)
[1]  2.570582
> qt(.025, 11)
[1] -2.20985
> qt(.975, 11)
[1]  2.20985
> qt(.025, 29)
[1] -2.04523
> qt(.975, 29)
[1]  2.04523
> qt(.025, 100)
[1] -1.983972
> qt(.975, 100)
[1]  1.983972
```
Example

Consider the following IQ scores from a sample of 10 healthy people:

96  112  103  89  105  112  96  118  102  107

Construct a CI to estimate its mean, assuming that the population is normally distributed.
Example

First estimate the mean and SD of the population:
\[ \hat{\mu} = 104.0 \quad \hat{\sigma} = 8.8 \]

Then estimate the standard error of the mean:
\[ \hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{8.8}{\sqrt{10}} = 2.77 \]

The margin of error (MOE) for the 95\% CI is:
\[ MOE = t_{0.975,9} \times \hat{\sigma}_{\hat{\mu}} \simeq 2.262 \times 2.77 = 6.27 \]
R code

```
> scores <- c(96, 112, 103, 89, 105, 112, 96, 118, 102, 107)
> m <- mean(scores)
> m
[1] 104
> se <- sd(scores)/sqrt(10)
> moe <- qt(.975,9)*se
> ci <- c(m - moe, m + moe)
> ci
[1] 97.7273 110.2727
```
t.test function

Alternatively, you can use the t.test function:

```r
> t.test(scores)

One Sample t-test

data:  scores
t = 37.506, df = 9, p-value = 3.378e-11
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 97.7273 110.2727
sample estimates:
mean of x
 104
```
t.test function

Alternatively, you can use the t.test function:

```r
> t.test(scores)$conf.int
[1]  97.7273 110.2727
attr("conf.level")
[1] 0.95
```

The t.test function returns a list. The list member “conf.int” gives the 95% CI as calculated by the t.test function.
Reporting confidence intervals

The American Psychological Association (APA) suggests reporting the full CI next to the point estimate:

\[ M = 165 \text{ cm}, \ 95\% \ CI \ [160 \text{ cm}, \ 172 \text{ cm}] \]

**Recommended**: It allows to report both symmetrical and asymmetrical CIs in the same way.
Our estimate for the mean IQ score of the population (based on our sample) is:

\[ M = 104.0, \ 95\% \ CI \ [97.7, \ 110.3] \]
Let’s play with R code!

Let’s repeatedly take samples \((n = 12)\) from a normally distributed population \((M = 100, \ SD = 10)\), and construct 95% CIs to estimate the mean, based on the *t* distribution.

Then, check how often the 95% CIs include the true mean value \(M = 100\). What do you expect?
R program

M <- 100  # True population mean
SD <- 10  # True population standard deviation
n <- 12   # Size of the samples

N <- 10000 # Number of sampling repetitions
count <- 0 # Counter of how many times the CI fails

for(i in 1:N){
    sample <- rnorm(n, mean = M, sd = SD)
    ci <- t.test(sample)$conf.int

}

cat("Average number of failures = ", count/N, "\n")
cat("Coverage probability = ", (1 - count/N)*100, "\%\n")
Let’s try a different population

Let’s repeat the experiment by sampling \((n = 12)\) from a population that follows a highly skewed log-normal distribution.

![Graph of log-normal distribution with mean 164.0 and median 99.5](image)

Mean = 164.0, Median = 99.5
R program

# Parameters of the log-normal distribution
m <- 4.6  # Mean of log-transformed values (normal distribution)
sd <- 1  # SD of log-transformed values (normal distribution)

# True population mean (mean of the original skewed distribution)
M <- exp(m + sd^2/2)

n <- 12  # Size of the samples
N <- 10000  # Number of sampling repetitions
count <- 0  # Counter of how many times the CI fails

for(i in 1:N)
{
    sample <- rlnorm(n, meanlog = m, sdlog = sd)
    ci <- t.test(sample)$conf.int

}

cat("Average number of failures = ", count/N, "\n")
cat("Coverage probability = ", (1 - count/N)*100, "\%\n")
# Parameters of the log-normal distribution
m <- 4.6  # Mean of log-transformed values (normal distribution)
sd <- 1   # SD of log-transformed values (normal distribution)

# True population mean (mean of the original skewed distribution)
M <- exp(m + sd^2/2)  # Proof: https://www.statlect.com/probability-distributions/log-normal-distribution

n <- 12  # Size of the samples
N <- 10000 # Number of sampling repetitions
count <- 0  # Counter of how many times the CI fails

for(i in 1:N){
    sample <- rlnorm(n, meanlog = m, sdlog = sd)
    ci <- t.test(sample)$conf.int

}

cat("Average number of failures = ", count/N, "\n")
cat("Coverage probability = ", (1 - count/N)*100, "\%\n")
The result...

Average number of failures = 0.1549
Coverage probability = 84.51 %
What about larger samples…

n = 30, Coverage probability = 88%

n = 100, Coverage probability = 92%

n = 500, Coverage probability = 94%

However, notice that my distribution is extremely skewed.
Monte Carlo methods

All these experiments are **Monte Carlo** simulations.

They are very useful for assessing the performance of statistical models and procedures.

Very often used to explore what happens when the assumptions of a statistical procedure are violated.
Estimating differences

We are often interested in estimating a difference between two populations:

- What is the difference in the mean temperature between June and July?
- What is the difference in math performance between women and men?
- How much slower (or faster) is human’s reaction time under alcohol use?
Confidence intervals for a difference

Given two independent sample statistics, we could construct the confidence interval of their difference.

**Example:** The mean difference in speed between techniques A and B is: 1.22s, 95% CI [1.12s, 1.34s]
Confidence intervals for a difference

How do we go from the CI of individual parameters to the CI of their difference?
Difference between normal distributions

The difference between two normal distributions is itself normal.
Difference between normal distributions

Given two independent samples (1 and 2) from two normal populations, their difference in means can be estimated as:

\[
\hat{\mu}_1 - \hat{\mu}_2 \pm t_{\nu, \alpha/2} \times \hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2}
\]

estimated difference in means

estimated standard error of the difference
Degrees of freedom

\[ \hat{\mu}_1 - \hat{\mu}_2 \pm t_{\nu, \alpha/2} \times \hat{\sigma} \hat{\mu}_1 - \hat{\mu}_2 \]

If the size of the two samples is \( n_1 \) and \( n_2 \), then the degrees of freedom of the \( t \) distribution are:

\[ \nu = (n_1 - 1) + (n_2 - 1) = (n_1 + n_2) - 2 \]
Standard error of the difference

\[ \hat{\mu}_1 - \hat{\mu}_2 \pm t_{\nu, \alpha/2} \times \hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} \]

How do we calculate the standard deviation of a difference?
The variance sum law

\[ \sigma_{X_1 \pm X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 \pm 2\sigma_{X_1, X_2} \]

Covariance (shared variance) between \( X_1 \) and \( X_2 \)

If the two variables are independent, they are uncorrelated, and their covariance is zero. Thus:

\[ \sigma_{X_1 \pm X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 \]
Applying the law to the distributions of means:

\[ \sigma_{\hat{\mu}_1 - \hat{\mu}_2}^2 = \sigma_{\hat{\mu}_1}^2 + \sigma_{\hat{\mu}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \]

\[ \sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \]

If we further assume equal variances (\(\sigma_1 = \sigma_2\)), then:

\[ \sigma_{\hat{\mu}_1 - \hat{\mu}_2} = \sigma_1 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sigma_2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]
Unknown population variances

When the common variance is not known, it is better to estimate it by taking into account all available data, from both samples:

\[
\hat{\sigma}_{\mu_1 - \mu_2} = \hat{\sigma}_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

where the pooled estimate is calculated as follows:

\[
\hat{\sigma}_{pooled} = \sqrt{\frac{\sum_{i=1}^{n_1} (x_i - \hat{\mu}_1)^2 + \sum_{i=1}^{n_2} (x_i - \hat{\mu}_2)^2}{n_1 + n_2 - 2}}
\]
Summary of assumptions

1. Both samples are drawn from population with normal distributions

2. The two samples are independent

3. Their populations have equal variances (homogeneity of variance assumption)
Homogeneity of variance

This assumption makes calculation easier.

If variances are not equal, then the sampling distribution of the standardized mean is no longer a $t$ distribution!
Unequal variances

We can use the same formula:

\[ \hat{\mu}_1 - \hat{\mu}_2 \pm t_{\nu', \alpha/2} \times \hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} \]

with “corrected” degrees of freedom:

\[ \nu' = \frac{\left( \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2} \right)^2}{\frac{\hat{\sigma}_1^4}{n_1^2(n_1-1)} + \frac{\hat{\sigma}_2^4}{n_2^2(n_2-1)}} \] (Welch-Satterthwaite correction)

and:

\[ \hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \]
Unequal variances

Please, don’t panic! You don’t have to remember all these equations. Software can do these corrections for you.

However, you need to be aware of the underlying assumptions, their problems, and their corrections. And you need to be aware of which methods you use.
Unequal variances

Experts recommend adopting the Welch-Satterthwaite correction by default.

Even if the assumption of the homogeneity of variance is true, the loss of accuracy when applying the correction is considered negligible.

(see discussion and references provided by Baguley)
Example

A research team is interested in comparing the performance in an IQ test between two groups:

\textbf{G1}: adults who have, and
\textbf{G2}: adults who have not completed any graduate studies

From each group, they test 15 participants.


\textbf{S_{G2}}: 96, 108, 90, 104, 97, 103, 95, 102, 93, 107, 101, 88, 99, 104, 97
R code

We are going to use the `t.test` R function:

```r
> g2<-c(96, 108, 90, 104, 97, 103, 95, 102, 93, 107, 101, 88, 99, 104, 97)
> t.test(g1)$conf.int
[1] 102.0595 107.6738
attr(,"conf.level")
[1] 0.95
> t.test(g2)$conf.int
[1] 95.64067 102.22600
attr(,"conf.level")
[1] 0.95
> t.test(g1,g2)$conf.int
[1] 1.796244 10.070423
attr(,"conf.level")
[1] 0.95
> t.test(g1,g2, var.equal = TRUE)$conf.int
[1] 1.800908 10.065759
attr(,"conf.level")
[1] 0.95
```

CIs of the mean of individual samples

CI of the mean difference (corrected for unequal variances)

CI of the mean difference (equal variances assumed)
Plots

95% CIs

IQ Score

0 2 4 6 8 10

G1  G2

95% CI

Diff IQ Score (G1 - G2)
R code

Plotting (*plotrix* package)

```r
> require(plotrix)
> ci1 <- t.test(g1)$conf.int
> ci2 <- t.test(g2)$conf.int
> cidiff <- t.test(g1,g2)$conf.int
> m1 <- mean(g1)
> m2 <- mean(g2)
> plotCI(c(1,2), c(m1,m2), ui=c(ci1[2],ci2[2]), li=c(ci1[1],ci2[1]), pch=19, xlab=NA, xaxt='n', xlim=c(.5, 2.5), ylab="IQ Score")
> axis(1, at = 1:2, labels = c("G1", "G2"))
> plotCI(1, m1 - m2, ui=cidiff[2], li=cidiff[1], pch=19, xlab=NA, xaxt='n', ylim=c(-1, 10), ylab="Diff IQ Score (G1 - G2)")
> segments(0, 0, 2,0, lty=2)
```
Example’s conclusions

The mean IQ score of group G1 is:

\[ M = 104.9, \text{ 95\% CI [102.1, 107.7]} \]

The mean IQ score of group G2 is:

\[ M = 98.9, \text{ 95\% CI [95.6, 102.2]} \]

The mean score difference between group G1 and group G2 is:

\[ M = 5.9, \text{ 95\% CI [1.8, 10.1]} \]
Exercise

Write a Monte Carlo simulation in R to evaluate the effect of unequal variances to the accuracy of CIs for mean differences.

Test whether and how results change if the size of the samples are not equal: $n_1 \neq n_2$
A research team is interested in assessing the effect of taking a Geometry course on students’ performance in an IQ test.

20 students are randomly split into two groups:

- $G_{\text{control}}$. This is the control group.
- $G_{\text{geometry}}$. The group that follows the Geometry course.

Results are as follows:

- $G_{\text{control}}$: 102, 94, 90, 104, 95, 100, 101, 96, 100, 96
- $G_{\text{geometry}}$: 106, 108, 93, 103, 100, 100, 105, 98, 103, 96
Example

The two groups can be assumed as independent:
Example (variation of the previous)

A research team is interested in comparing the performance in an IQ test before and after attending a Geometry course:

10 participants are tested before and after the completion of the course.

\[ S_{\text{Before}}: 102, 94, 90, 104, 95, 100, 101, 96, 100, 96 \]

\[ S_{\text{After}}: 106, 108, 93, 103, 100, 100, 105, 98, 103, 96 \]

The two samples are no longer independent!
Repeated-measures designs

Such experimental designs, where repeated observations are made over the same individuals are known as repeated measures (or within-participants designs).

Observations are related (paired), and thus, they are not independent.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>101</td>
<td>96</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>After</td>
<td>106</td>
<td>108</td>
<td>93</td>
<td>103</td>
<td>100</td>
<td>100</td>
<td>105</td>
<td>98</td>
<td>103</td>
<td>96</td>
</tr>
</tbody>
</table>
CIs for related samples

The solution is trivial. We calculate the difference in performance for each participant.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>101</td>
<td>96</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td>106</td>
<td>108</td>
<td>93</td>
<td>103</td>
<td>100</td>
<td>100</td>
<td>105</td>
<td>98</td>
<td>103</td>
<td>96</td>
</tr>
<tr>
<td><strong>Diff.</strong></td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>-1</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

We then construct the CI that estimates the mean of these differences.
R Code

```r
> before = c(102, 94, 90, 104, 95, 100, 101, 96, 100, 96)
> after = c(106, 108, 93, 103, 100, 100, 105, 98, 103, 96)
> t.test(after-before)$conf.int
[1] 0.3800225 6.4199775
attr(,"conf.level")
[1] 0.95
> t.test(after, before, paired=TRUE)$conf.int
[1] 0.3800225 6.4199775
attr(,"conf.level")
[1] 0.95
```

Alternative
Results

The CI of the mean difference is now narrower.

In general, repeated measures (within-participants) designs produce more precise estimations than experimental designs with independent participant groups (between-participants designs).
Pros and cons of repeated measures

Pros:

- Require a smaller number of participants
- More powerful in estimating differences

Cons:

- Participation is longer for each participant
- Can introduce undesirable *carryover effects* (e.g., fatigue or learning)
Example

Consider the previous example, where students took the IQ test before and after following the Geometry course. The mean IQ score improved, **but what caused this improvement?**

- Was it the Geometry course?
- Or was it the fact that students were now more familiar with IQ tests?
Complex designs

Experimental designs can be more complex in different aspects:

- Control for multiple factors *(independent variables)*
- Study more than two levels per factor
- Combine between-participants and repeated-measures designs
Example

A research team is interested in assessing the effect of a Geometry course on students’ IQ performance. They randomly create two groups of students (Control vs. Geometry).

Each student takes three IQ tests over three weeks.

<table>
<thead>
<tr>
<th>Control Group</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Week 2</td>
<td>105</td>
<td>97</td>
<td>93</td>
<td>96</td>
<td>99</td>
<td>105</td>
<td>100</td>
<td>93</td>
<td>103</td>
<td>98</td>
</tr>
<tr>
<td>Week 3</td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>106</td>
<td>98</td>
<td>105</td>
<td>102</td>
<td>98</td>
<td>104</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geometry Class Group</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>105</td>
<td>98</td>
<td>100</td>
<td>97</td>
<td>103</td>
<td>104</td>
<td>103</td>
<td>103</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Week 2</td>
<td>106</td>
<td>100</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>105</td>
<td>100</td>
<td>95</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>Week 3</td>
<td>103</td>
<td>101</td>
<td>98</td>
<td>106</td>
<td>98</td>
<td>100</td>
<td>101</td>
<td>95</td>
<td>104</td>
<td>97</td>
</tr>
</tbody>
</table>
Example

This is a **mixed design** that combines:

- a between-participants factor (Geometry Course) with two levels: **No** (Control) and **Yes** (Geometry)
- a within-participants (repeated-measures) factor (Week) with three levels: **Week 1**, **Week 2**, **Week 3**
Example

**Question 1**: What’s the mean difference in IQ scores between students who took and students who did not take the geometry course?

<table>
<thead>
<tr>
<th>Control Group</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Week</td>
<td>105</td>
<td>97</td>
<td>93</td>
<td>96</td>
<td>99</td>
<td>105</td>
<td>100</td>
<td>93</td>
<td>103</td>
<td>98</td>
</tr>
<tr>
<td>Week</td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>106</td>
<td>98</td>
<td>105</td>
<td>102</td>
<td>98</td>
<td>104</td>
<td>93</td>
</tr>
<tr>
<td>Mean:</td>
<td>102.7</td>
<td>97.0</td>
<td>92.0</td>
<td>102.0</td>
<td>97.3</td>
<td>103.3</td>
<td>99.7</td>
<td>95.7</td>
<td>101.0</td>
<td>95.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geometry Class Group</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>105</td>
<td>98</td>
<td>100</td>
<td>97</td>
<td>103</td>
<td>104</td>
<td>103</td>
<td>103</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Week</td>
<td>106</td>
<td>100</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>105</td>
<td>100</td>
<td>95</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>Week</td>
<td>103</td>
<td>101</td>
<td>98</td>
<td>106</td>
<td>98</td>
<td>100</td>
<td>101</td>
<td>95</td>
<td>104</td>
<td>97</td>
</tr>
<tr>
<td>Mean:</td>
<td>104.7</td>
<td>99.7</td>
<td>97.0</td>
<td>101.0</td>
<td>100.3</td>
<td>103.0</td>
<td>101.3</td>
<td>97.7</td>
<td>100.0</td>
<td>98.0</td>
</tr>
</tbody>
</table>
Example

**Question 1**: What’s the mean difference in IQ scores between students who took and students who did not take the geometry course?

Taking the mean is the most common way of aggregating data over an independent variable.

However, it is not the only way. Depending on their research goals, the researchers may, for example, decide to use the median to reduce the influence of extreme values. Or they may decide to take the maximum score, or just the final score.

<table>
<thead>
<tr>
<th>Week</th>
<th>Control Group</th>
<th>Geometry Class Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>105 98 100 97 103 104 103 103 100 99</td>
<td>106 100 93 100 100 105 100 95 96 98</td>
</tr>
<tr>
<td>Week 2</td>
<td>106 100 93 100 100 105 100 95 96 98</td>
<td>103 101 98 106 98 100 101 95 104 97</td>
</tr>
<tr>
<td>Mean:</td>
<td>104. 99.7 97.0 101.0 100.3 103.0 101.3 97.7 100.0 98.0</td>
<td></td>
</tr>
</tbody>
</table>
Results

IQ Score

95% CIs

Control  Geometry

Diff IQ Score

95% CI
Storing data on CSV files

```
<table>
<thead>
<tr>
<th>group</th>
<th>participant</th>
<th>week</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;control&quot;</td>
<td>1</td>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>1</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>1</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>2</td>
<td>1</td>
<td>94</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>2</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>2</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>3</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>&quot;control&quot;</td>
<td>3</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>11</td>
<td>1</td>
<td>105</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>11</td>
<td>2</td>
<td>106</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>11</td>
<td>3</td>
<td>103</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>12</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>12</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>12</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>13</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>13</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>13</td>
<td>3</td>
<td>98</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>14</td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>14</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>&quot;geometry&quot;</td>
<td>14</td>
<td>3</td>
<td>106</td>
</tr>
</tbody>
</table>
```

Each raw represents a unique observation.
#read the data from a csv file into a data frame
data <- read.csv(file="IQ-tests.csv", header=TRUE, sep="",""

# aggregate the scores by group and participant
data.aggr <- aggregate(score~group+participant, data, 
  FUN = mean)

# split the aggregated data into two groups
data.control <- data.aggr[data.aggr$group=="control",]
data.geometry <- data.aggr[data.aggr$group=="geometry",]

# calculate the means
mean.control <- mean(data.control$score)
mean.geometry <- mean(data.geometry$score)
Some bad practices

A **bad practice** is to construct and graph confidence intervals over the full set of data.

This approach is not appropriate as it treats all observations as independent, i.e., as observations from different participants.
Some bad practices

A bad practice is to construct and graph confidence intervals over the full set of data.

This approach is not appropriate as it treats all observations as independent, i.e., as observations from different participants.

Recommendation: Be always clear about what reported CIs represent and how they were calculated.
Example

**Question 2:** How does student performance evolve over time, from Week 1 to Week 3?

<table>
<thead>
<tr>
<th>Control Group</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td>105</td>
<td>97</td>
<td>93</td>
<td>96</td>
<td>99</td>
<td>105</td>
<td>100</td>
<td>93</td>
<td>103</td>
<td>98</td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>106</td>
<td>98</td>
<td>105</td>
<td>102</td>
<td>98</td>
<td>104</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geometry Class Group</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td>105</td>
<td>98</td>
<td>100</td>
<td>97</td>
<td>103</td>
<td>104</td>
<td>103</td>
<td>103</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td>106</td>
<td>100</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>105</td>
<td>100</td>
<td>95</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td>103</td>
<td>101</td>
<td>98</td>
<td>106</td>
<td>98</td>
<td>100</td>
<td>101</td>
<td>95</td>
<td>104</td>
<td>97</td>
</tr>
</tbody>
</table>
Example: Approach

We examine how IQ scores evolve for all 20 students.

We forget the student groups for now and focus on the repeated-measures variable, the Week.
Example: Results
### Control Group

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td>102</td>
<td>94</td>
<td>90</td>
<td>104</td>
<td>95</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td>105</td>
<td>97</td>
<td>93</td>
<td>96</td>
<td>99</td>
<td>105</td>
<td>100</td>
<td>93</td>
<td>103</td>
<td>98</td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>106</td>
<td>98</td>
<td>105</td>
<td>102</td>
<td>98</td>
<td>104</td>
<td>93</td>
</tr>
<tr>
<td><strong>Week 3 -1</strong></td>
<td>-1</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>-3</td>
</tr>
</tbody>
</table>

### Geometry Class Group

<table>
<thead>
<tr>
<th></th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td>105</td>
<td>98</td>
<td>100</td>
<td>97</td>
<td>103</td>
<td>104</td>
<td>103</td>
<td>103</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td><strong>Week 2</strong></td>
<td>103</td>
<td>101</td>
<td>98</td>
<td>106</td>
<td>98</td>
<td>100</td>
<td>101</td>
<td>95</td>
<td>104</td>
<td>97</td>
</tr>
<tr>
<td><strong>Week 3</strong></td>
<td>103</td>
<td>101</td>
<td>98</td>
<td>106</td>
<td>98</td>
<td>100</td>
<td>101</td>
<td>95</td>
<td>104</td>
<td>97</td>
</tr>
<tr>
<td><strong>Week 3 -1</strong></td>
<td>-2</td>
<td>3</td>
<td>-2</td>
<td>9</td>
<td>-5</td>
<td>-4</td>
<td>-2</td>
<td>-8</td>
<td>4</td>
<td>-2</td>
</tr>
</tbody>
</table>
Non-normal distributions

How do we construct confidence intervals for non-normal distributions?
CIs for binomial proportions

**Example:** A basketball player has attempted a total of $n = 25$ three point shots and has succeeded in $x = 9$. Can you estimate the player’s average success rate?

There are several alternatives for constructing CIs for binomial proportions, where some of them work well only for large samples (Control Limit Theorem) or proportions in the region of $.5$.

Baguley recommends the exact Blaker CI under most circumstances.
R Code

```r
install.packages("exactci")
library(exactci)

ci <- binom.exact(9, 25, tsmethod='blaker',
                   conf.level=.95)$conf.int

cat(ci[1]*100, ci[2]*100, "\n")
```

The success rate of the player is: 36%, 95% CI [19%, 56%]
CI for a difference in proportions

**Example:** A second basketball player has attempted a total of 20 three-point shots and has succeeded in 10. Estimate the difference between the success rate of the two players.

Baguley recommends the *continuity corrected version of the Wilson CI.*
R Code

\[
\text{prop.test}(c(25, 20), c(9, 8))$conf.int
\]

The success rate of the 2nd player is higher by a

14\%, 95\% \text{ CI } [-19\%, 47\%]