

Lecture 3

Confidence Intervals

Theophanis Tsandilas

Statistical inference

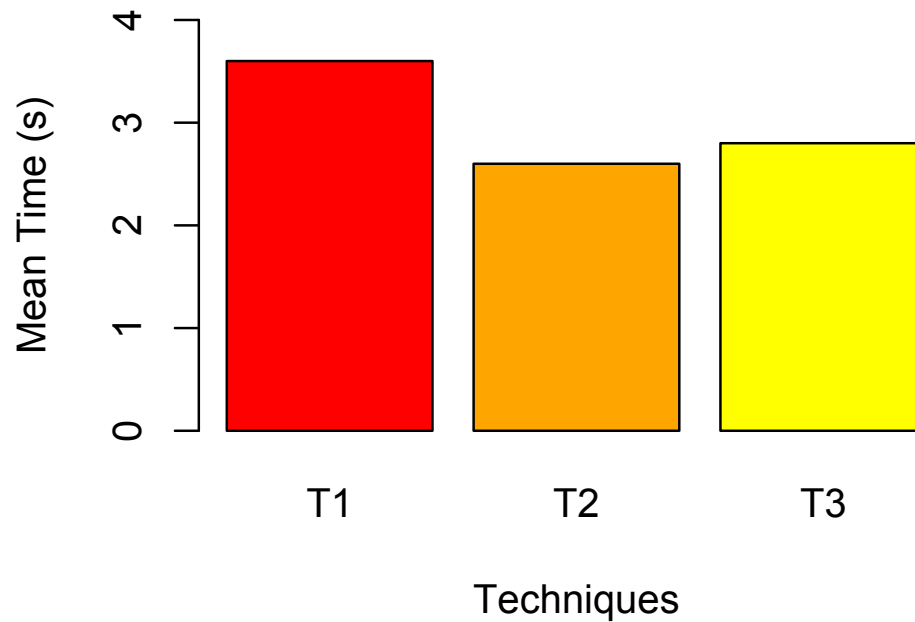
The process of deducing the parameters of an underlying probability distribution from a sample

Four broad types:

- ▶ point estimation
- ▶ interval estimation
- ▶ hypothesis testing
- ▶ prediction

Point estimates

How much informative is the following graph?



Point estimates

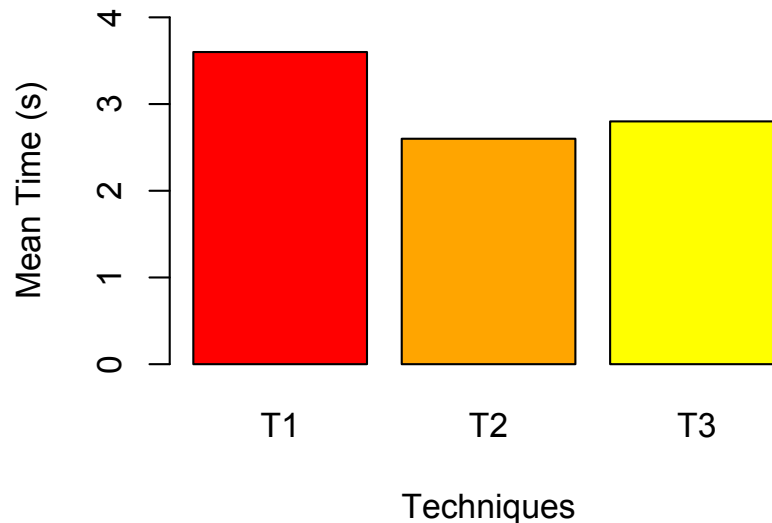
A point estimate can be thought of as a « best guess » of the true population parameter

Descriptive statistics such as the sample mean or the median are examples of point estimates

Question: What are the point estimates of a population's variance and standard deviation?

Point estimates

How much informative is the following graph?



A point estimate communicates no information about the uncertainty or quality of the estimate it provides

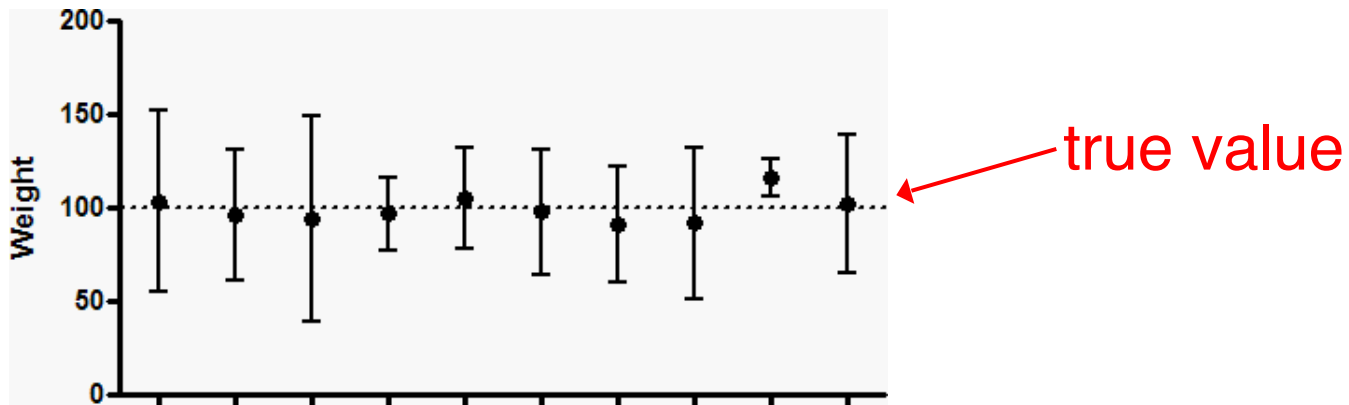
Interval estimate

An interval estimate does not provide an exact value, but rather a range of values that the parameter might plausibly take.

Most common method: constructing a **confidence interval (CI)**

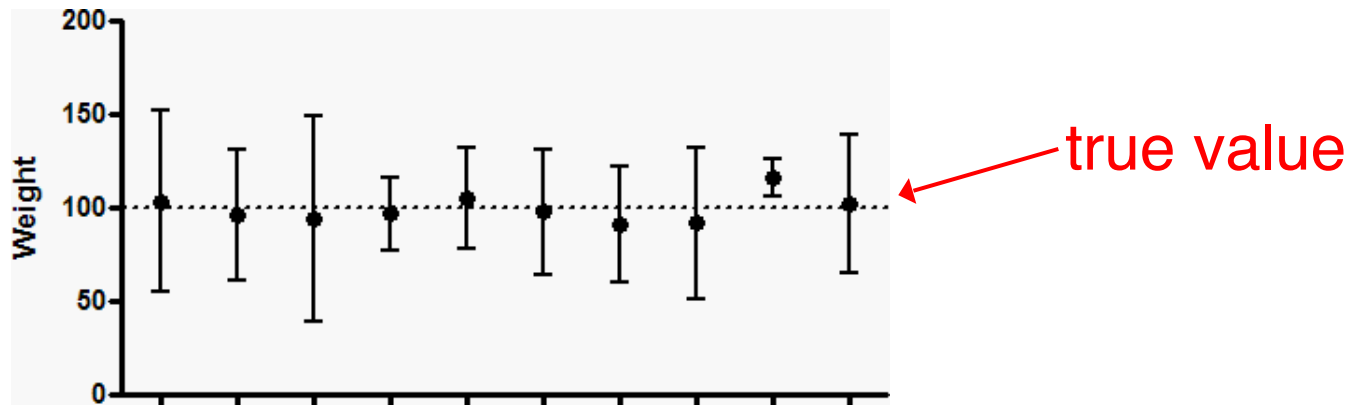
Confidence interval (CI)

It specifies a range of values that is expected to contain the true parameter value (but it may not)



Confidence interval (CI)

It specifies a range of values that is expected to contain the true parameter value (but it may not)

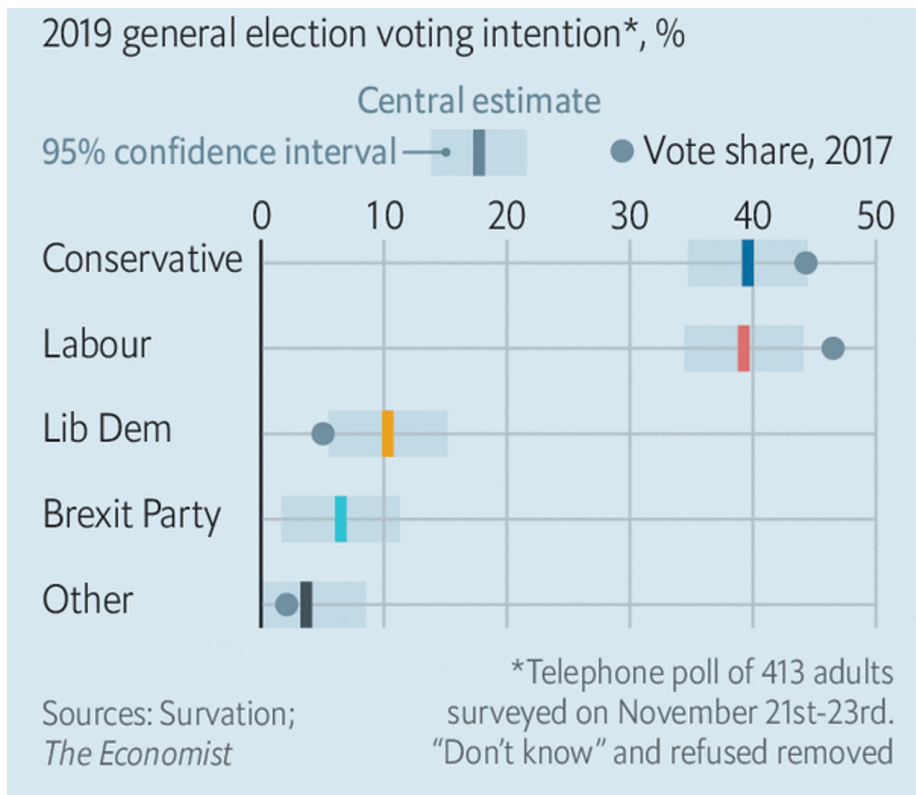


It is associated with a **confidence level**, usually expressed as a percentage

► e.g., 95% CI or 99% CI

Interpreting a confidence interval (CI)

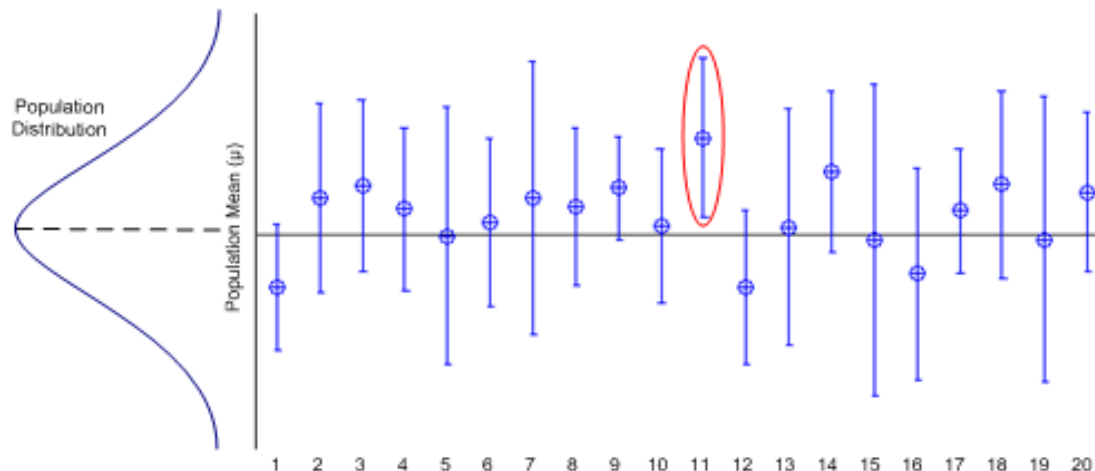
What do 95% confidence intervals represent here?



Formal interpretation of CIs

Classical **frequentists statistics** view a probability as a statement about the frequency with which events occur in the long run.

Of the many 95% CIs that might be constructed, 95% are expected to contain the true population parameter. The other 5% may completely fail!



Understanding probabilities

Suppose you know that if one opens a large number of cookie boxes, approximately 95% of the boxes will contain a surprise gift.

Question 1: What is the probability that a randomly selected cookie box will contain a surprise gift?

We can write this probability as **$\text{Pr}(\text{Gift}) = .95$**

Understanding probabilities

Suppose you know that if one opens a large number of cookie boxes, approximately 95% of the boxes will contain a surprise gift.

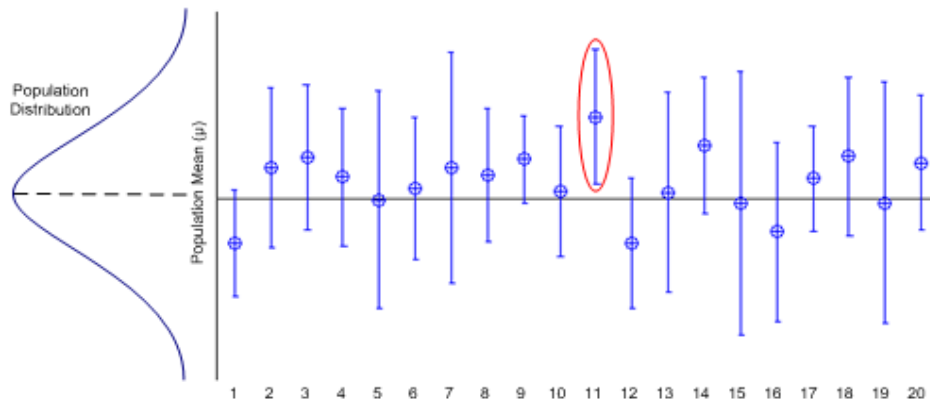
Suppose you randomly choose to open box A.

Question 2: Given that you opened box A, what is the probability the box contains a surprise gift?

$\Pr(\text{Gift} \mid A) = ?$

Understanding probabilities

Suppose that many research teams run independent experiments and construct a 95% CI to estimate a mean.



Question 1: What is the probability that the CI of a random experiment will include the true population value?

Question 2: Given an experiment (e.g., the one in red), what is the probability that its range includes the true population value?

CIs and probabilities

Formally speaking, a given CI does not specify a probability range. The true parameter is considered as fixed and the CI may or may not contain it.

We cannot claim that a 95% CI contains the true population parameter with a 95% probability or with 95% confidence

Classical frequentists statistics do not allow for such probabilistic reasoning.

Informal interpretation of CIs

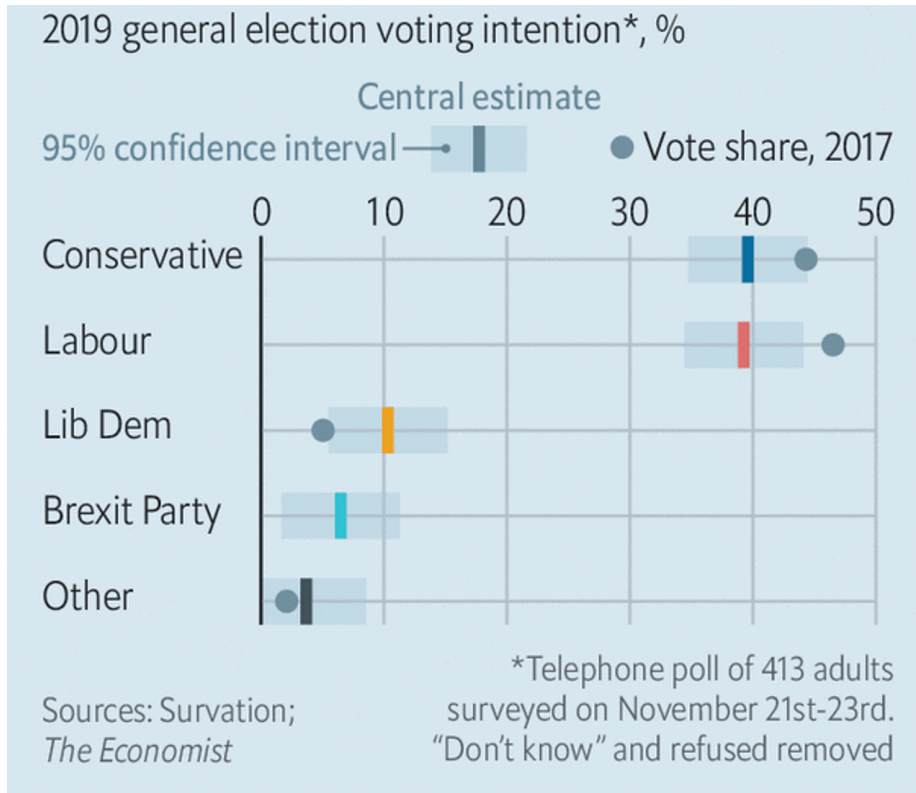
However, it is often reasonable to treat a CI as an expression of confidence or belief that it does contain the true value.

See [Baguley] and [Cumming and Finch, 2005]

Attention: This view has its critics.

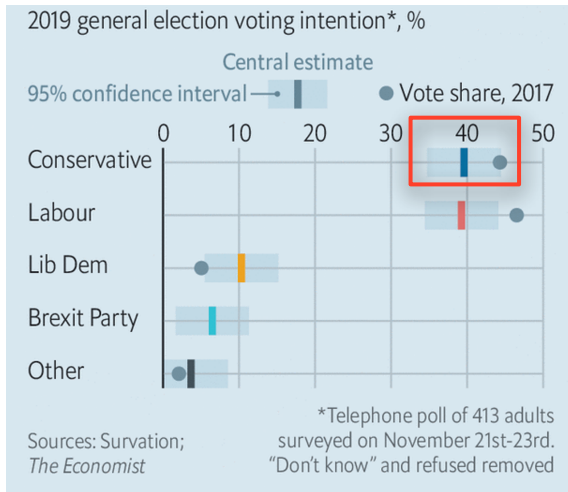
Interpreting a confidence interval (CI)

What do 95% confidence intervals represent here?



Interpreting a confidence interval (CI)

What do 95% confidence intervals represent here?



The Economist

Conservatives will gain between 35% and 45% of the votes with 95% probability.

There is a 95% probability that between 35% and 45% of the voters intend to vote for Conservatives.

We are 95% confident that 35% and 45% of the voters intend to vote for Conservatives.

We expect that the true voting intention for Conservative will be included by the CI of 95% of the polls.



Confidence level

A 100% CI will include the whole range of possible values

A 0% CI reduces to a point estimate

A 95% CI is the most common choice (**by tradition**)

alpha level

If C is the confidence level of a confidence interval, then:

$$C = 100 (1 - \alpha)$$

where α (or **alpha**) represents the number of times that a $C\%$ CI is expected to fail:

$$\text{If } C = 95, \text{ then } \alpha = .05$$

Structure of a confidence interval

It is defined by two points that form its limits, i.e., its lower and upper bounds

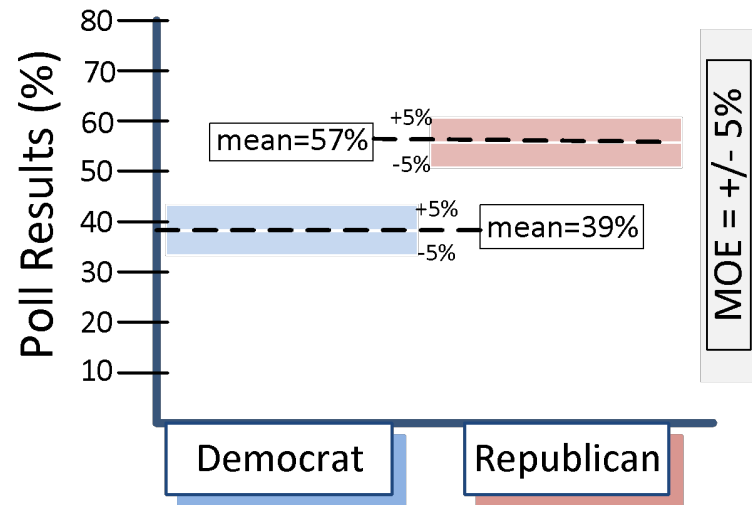
It can be **symmetrical**, where the point estimate lies in the center of the CI

...or **asymmetrical**, where the point estimate is not at the center of the CI

Symmetrical CIs

The intervals can be described by the point estimate plus or minus half of the interval, e.g.,
 165 ± 6 cm

This half width of the interval is known as the **margin of error (MOE)**



Width of a CI

Depends on the confidence level:

- ▶ 99% CIs are wider than 95% CIs

It also depends on the size of the sample:

- ▶ small samples produce wide CIs

Sampling distributions and CIs

To derive the $C\%$ **CI** of a statistic (e.g., the mean), we first need to approximate the **sampling distribution** of this statistic.

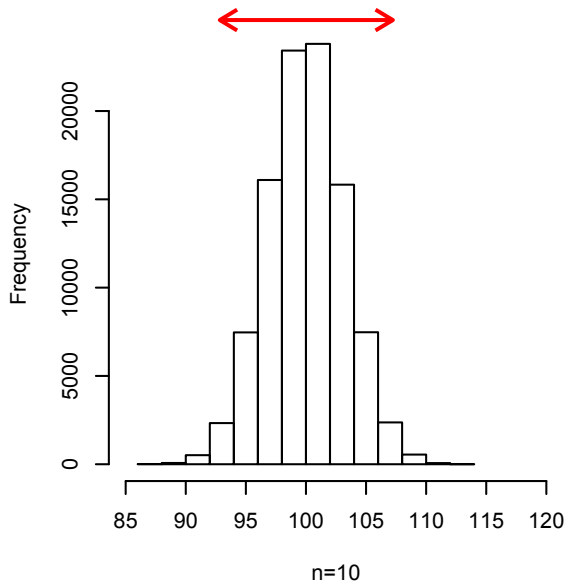
Why?

Because the sampling distribution provides the probability distribution of all possible values of the statistic. Our goal is to identify the $C\%$ of these possible values.

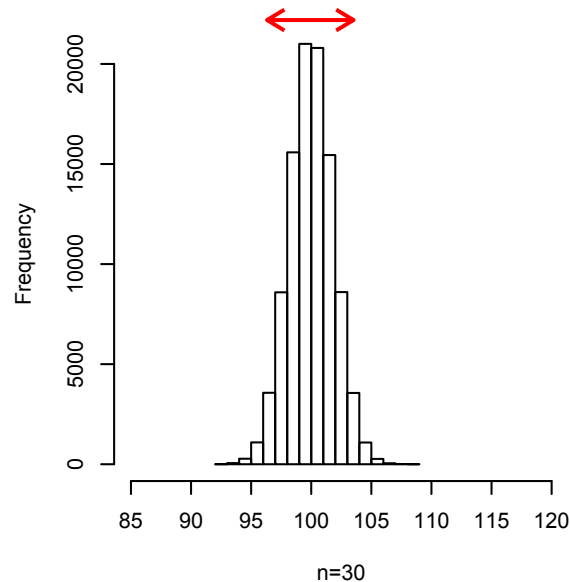
Example

Consider the **sampling distribution** of the mean for a normally distributed population ($M = 100$, $SD = 10$)

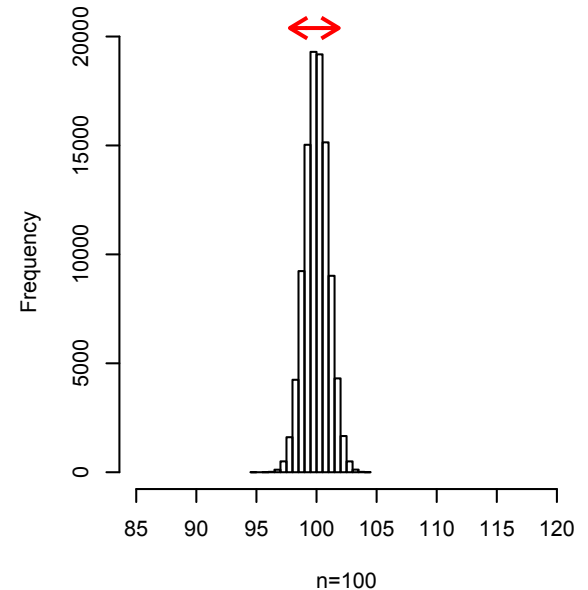
95% of the probable mean values



95% of the probable mean values



95% of the probable mean values



The sampling distribution becomes narrower as more samples are added. Thus, CIs should also become narrower.

Intuition of how CIs work

explained for means

Consider the sampling distribution of the mean ($n = 20$).
Let's take the range between its 10th and its 90th percentile.

Question 1: How many of the distribution's values does this range contain?

Imagine that we repeat an experiment by drawing random samples ($n = 20$) a large number of times. Every time, we calculate the mean of the sample.

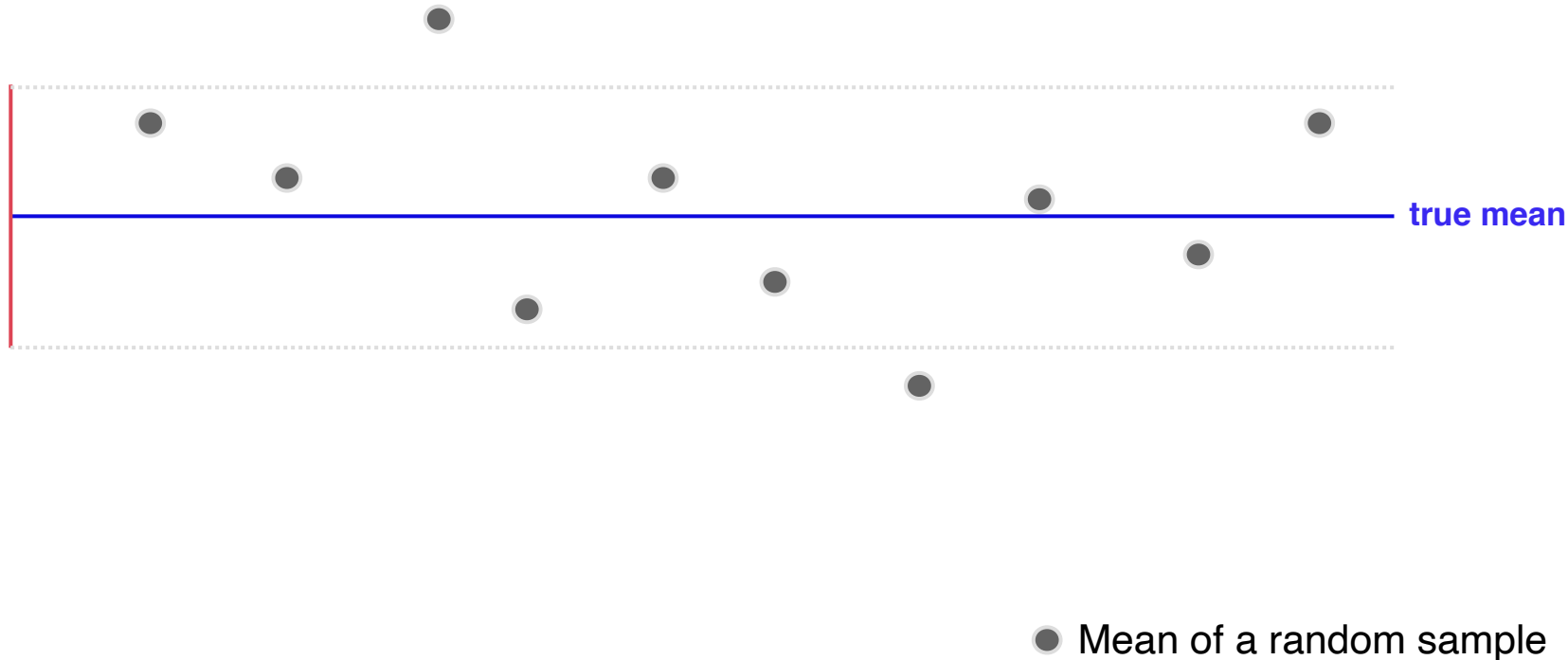
Question 2: On average, how many of these sample means does the above range contain?

Intuition of how CIs work

explained for means

On average, how many sample means will be included in this range?

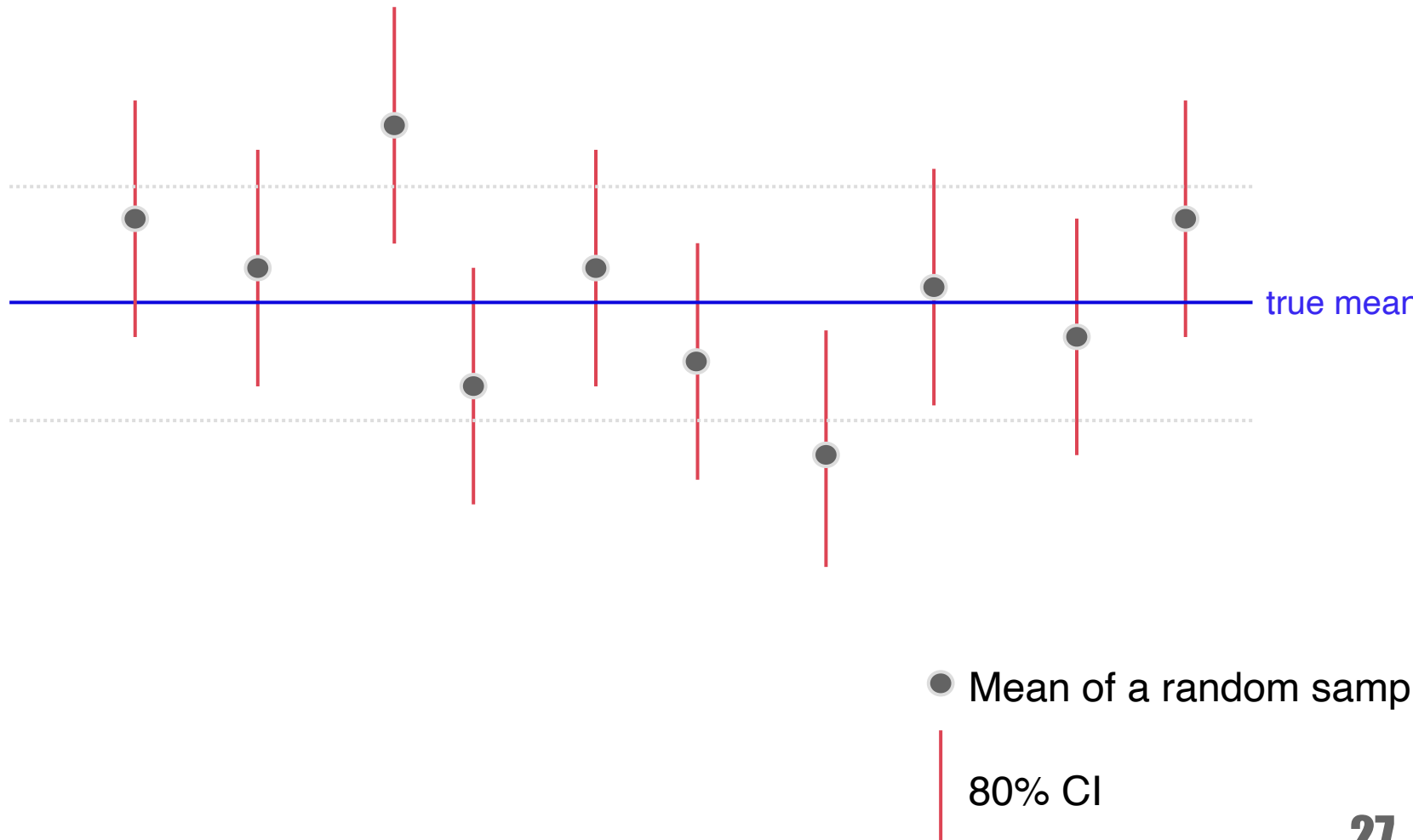
Range between 10th and 90th percentile
of sampling distribution of the mean



Intuition of how CIs work

explained for means

On average, how many 80% CIs will include the true mean?



Standard error (CE) of a statistic

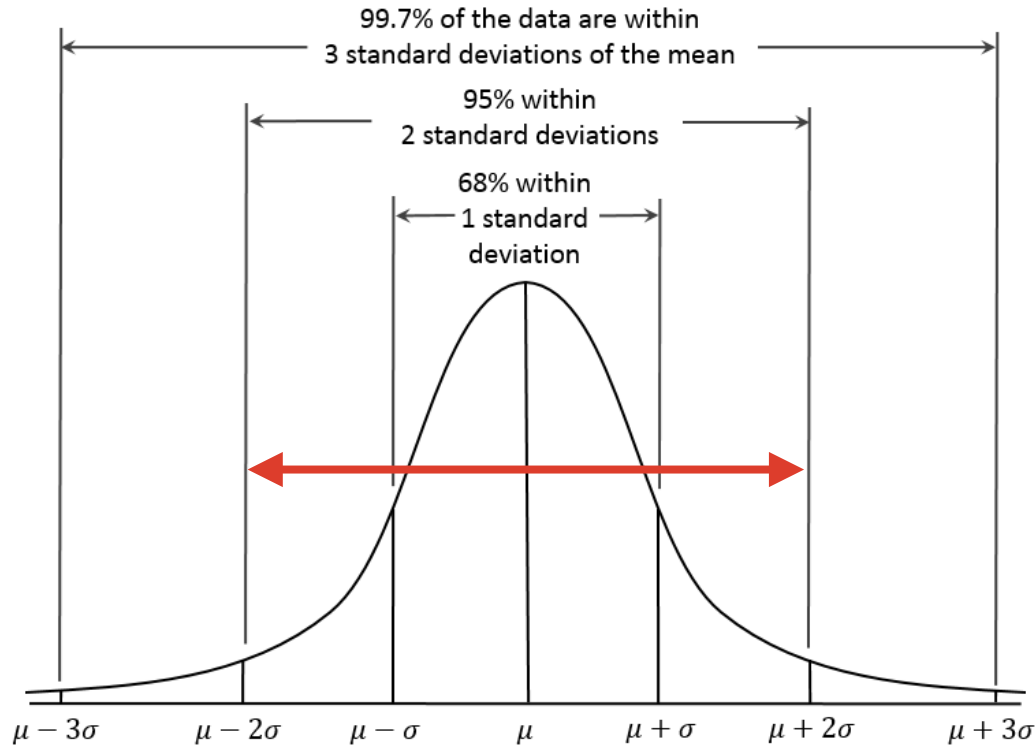
There is a clear connection between the **standard deviation** of the **sampling distribution** of a statistic and a **CI**

Standard deviation of the sampling distribution

=

Standard error (SE) of the statistic

Remember (normal distributions)...

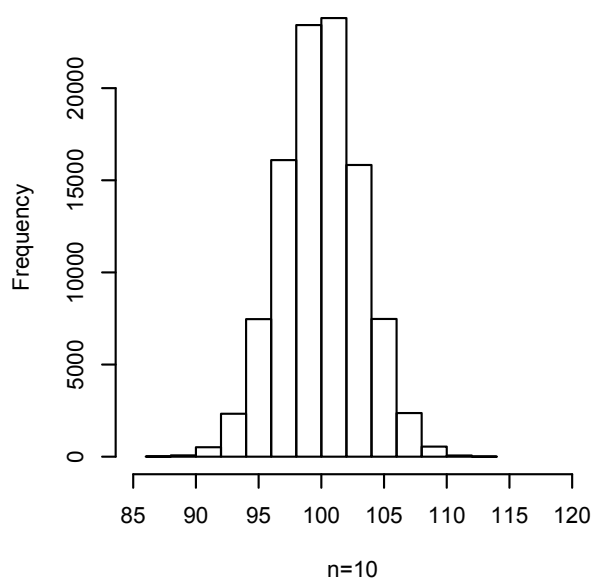


Suppose this represents the sampling distribution of the mean.

How could we derive the **95% confidence interval** of the mean given the standard error?

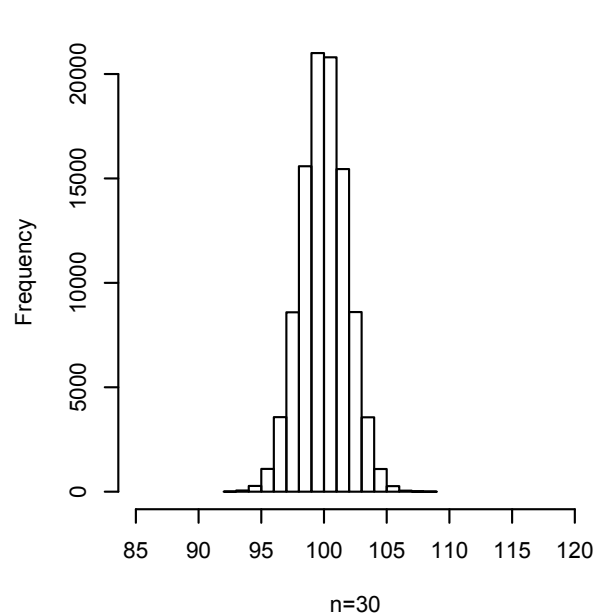
Example

Standard error of the mean (SEM) for a normal population ($M = 100$, $SD = 10$), when $n = 10$, 30 , and 100



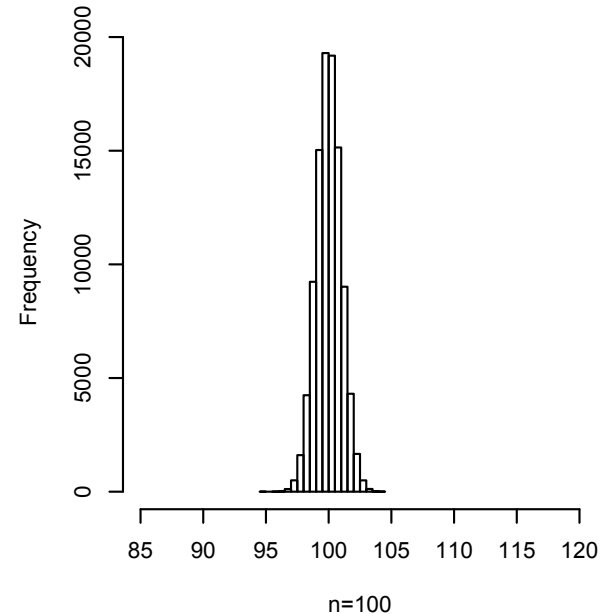
SEM = 3.16

95% CI = ?



SEM = 1.24

95% CI = ?



SEM = 1.00

95% CI = ?

SEM calculation

The **standard error of the mean** ($\sigma_{\hat{\mu}}$) derives from the standard deviation (σ) of the original population and the sample size n :

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma^2}{n} \quad \textbf{(variances)}$$

$$\Rightarrow \sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \quad \textbf{(standard deviations)}$$

From standard errors to CIs

The generic form of a symmetric CI is:

$$CI = \hat{\mu} \pm \phi_{\alpha/2} \times \sigma_{\hat{\mu}}$$

where $\phi_{\alpha/2}$ is the $\alpha/2 \times 100$ percentile of a symmetric standardized distribution.

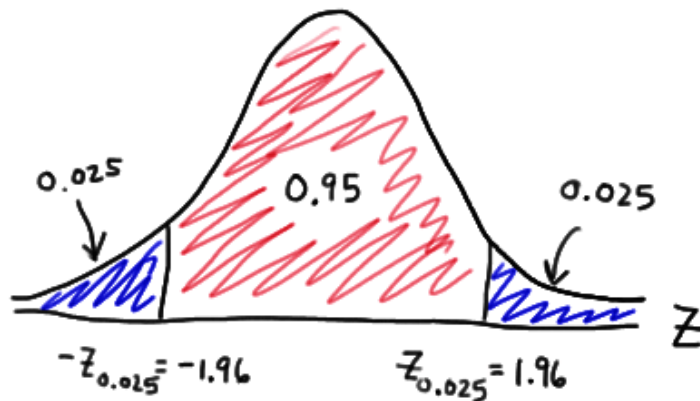
Normal sampling distribution

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

$$CI = \hat{\mu} \pm z_{\alpha/2} \times \sigma_{\hat{\mu}}$$

$\alpha/2 \times 100$ percentile
of the standard normal distribution

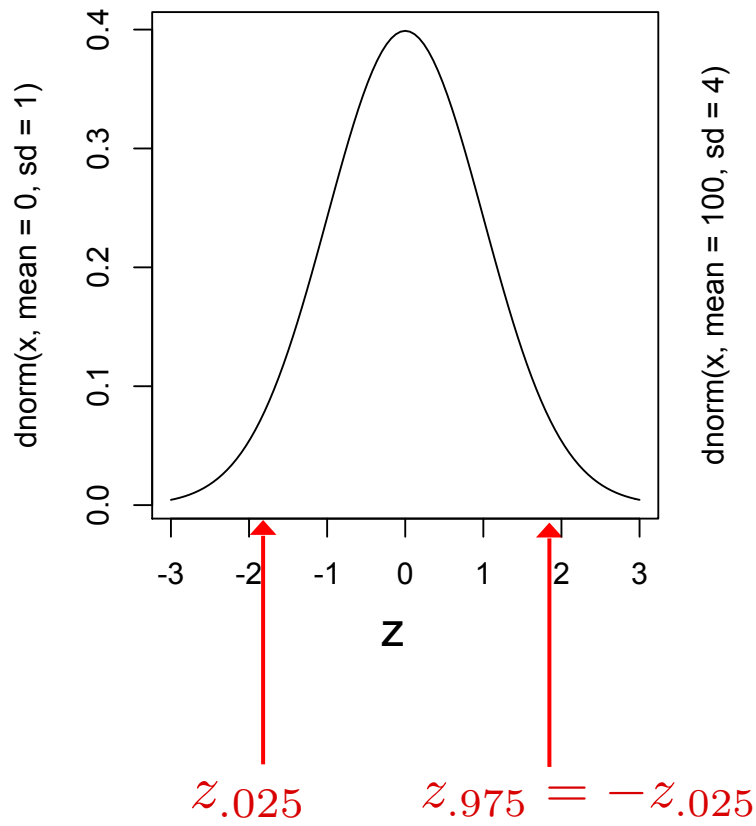
For a 95% CI, we take the 2.5th percentile



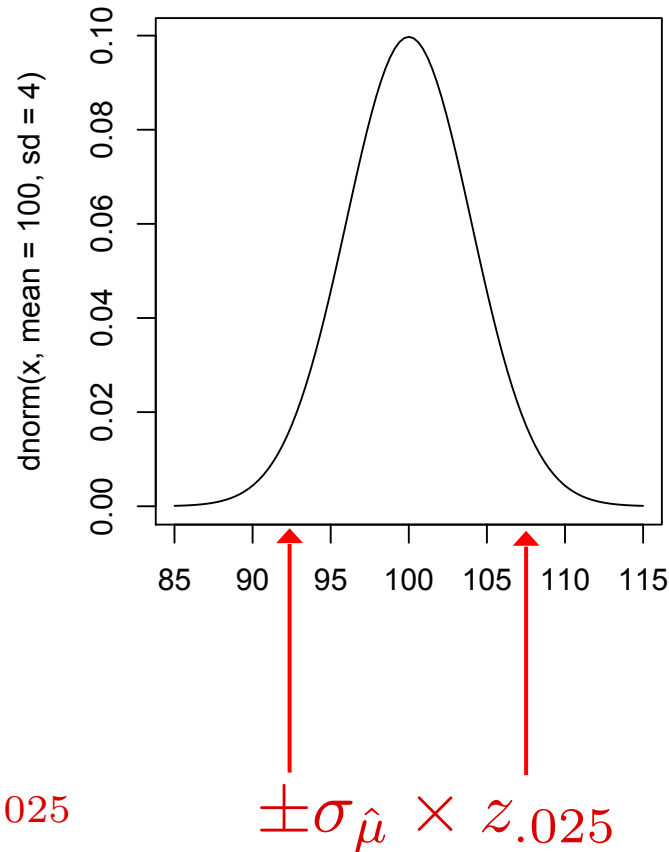
standard normal (z) distribution

Normal sampling distribution

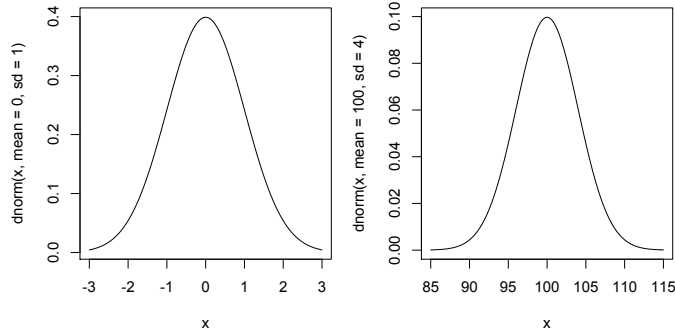
standard normal (z) distribution



non-standardized normal distribution



R code



Calculate the 2.5 percentile of the z distribution:

```
> qnorm(.025)  
[1] -1.959964
```

Calculate the 97.5 percentile of the z distribution

```
> qnorm(.975)  
[1] 1.959964
```

Calculate the 2.5 and 97.5 percentiles of the normal distribution with $M = 100$ and $SD = 4$

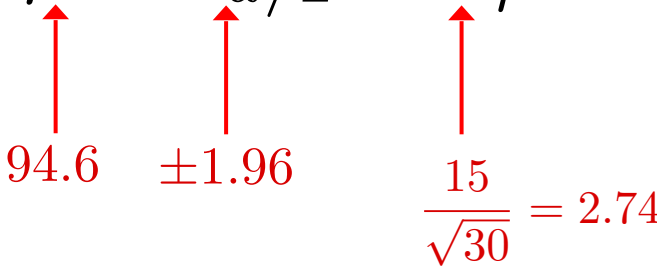
```
> qnorm(.975, 100, 4)  
[1] 107.8399  
> qnorm(.025, 100, 4)  
[1] 92.16014
```

Example

Imagine that a researcher takes a random sample of **30 people** and keeps them awake for 24 hours prior to taking awn 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**.

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


94.6 ±1.96 $\frac{15}{\sqrt{30}} = 2.74$

Problem

We rarely know the standard deviation of the original population

SEM estimation

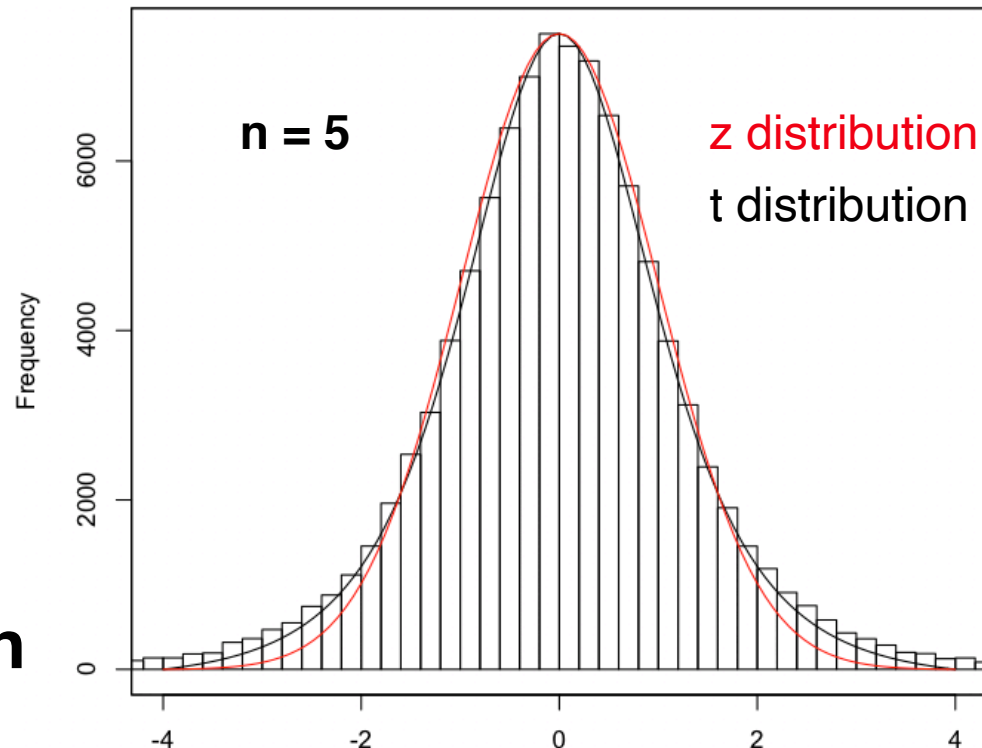
When the population standard deviation is unknown, we need to estimate it from the sample:

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

unbiased estimate of population stand. deviation

But...

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 σ is unknown and is estimated from the unbiased variance estimate:

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

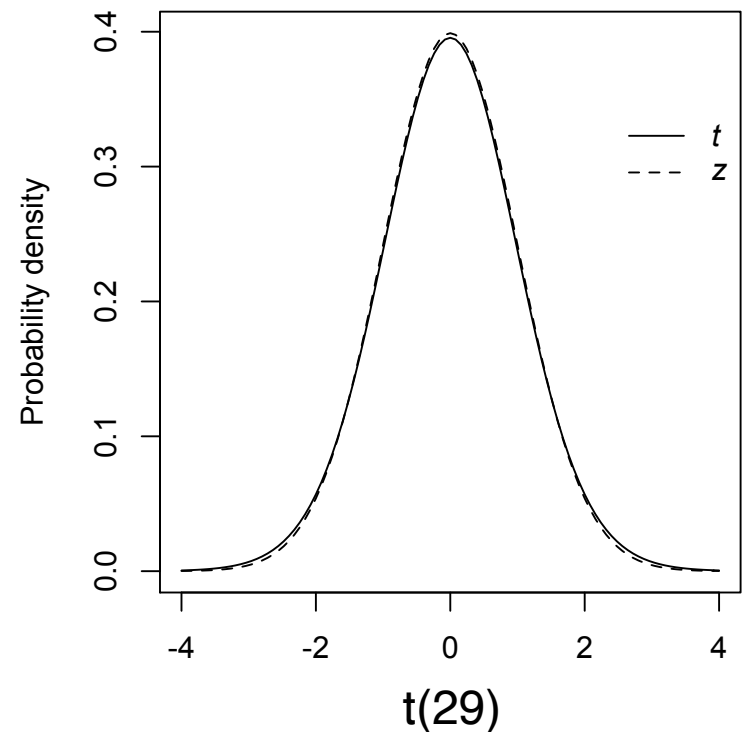
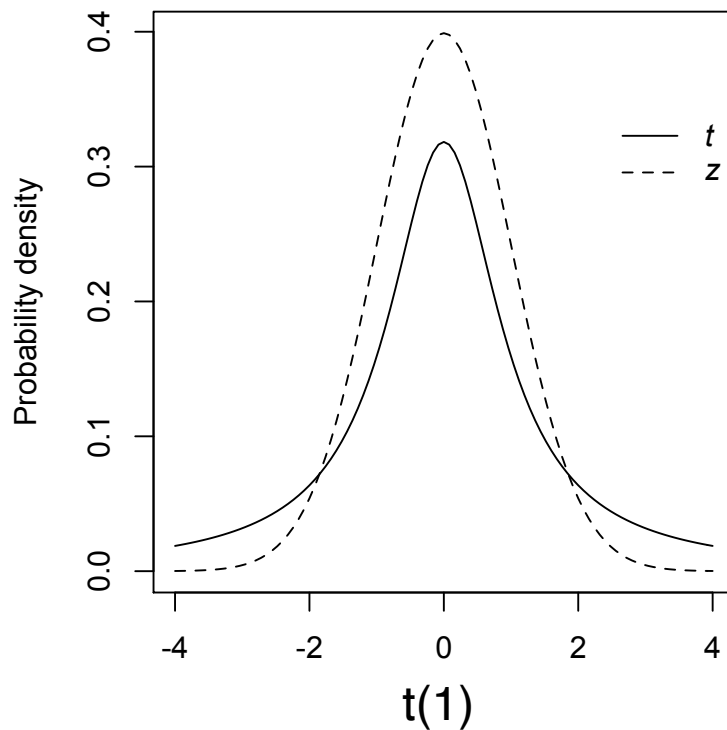
then, the resulting **standardized sample mean** has a t distribution with $\nu = n - 1$ degrees of freedom.

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

```
> 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')
```

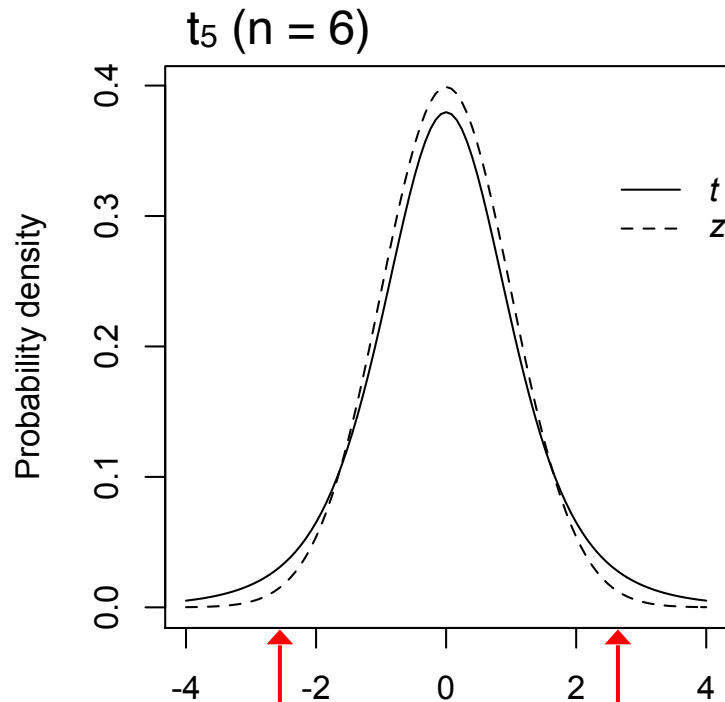
CIs using the t distribution

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

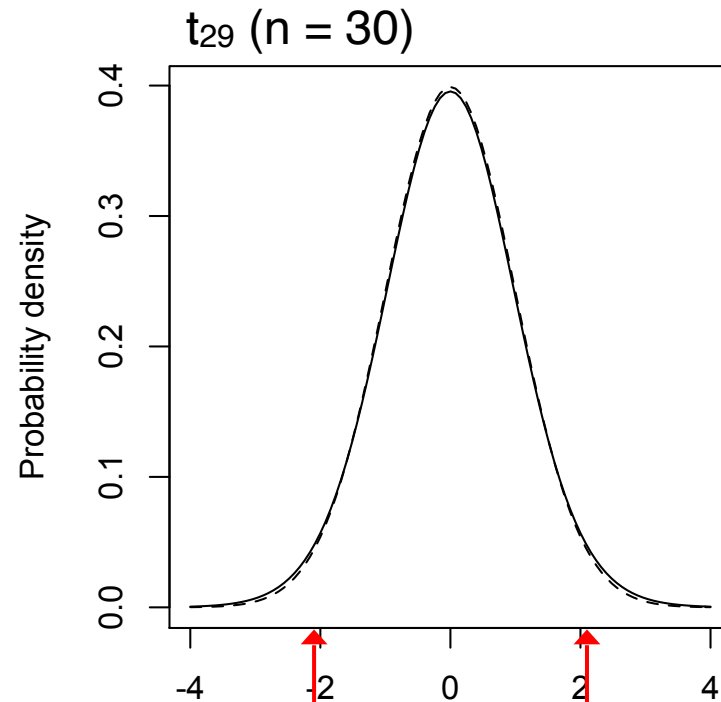
$$\hat{\mu} \pm t_{\nu, \alpha/2} \times \hat{\sigma}_{\hat{\mu}}$$

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



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



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

R code

Calculate the 2.5 and 97.5 percentiles of the t distribution for different degrees of freedom ν

```
> qt(.025, 5)
[1] -2.570582
> qt(.975, 5)
[1] 2.570582
> qt(.025, 11)
[1] -2.200985
> qt(.975, 11)
[1] 2.200985
> 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

96 112 103 89 105 112 96 118 102 107

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:

$$MOE = t_{9,.975} \times \hat{\sigma}_{\hat{\mu}} \simeq 2.262 \times 2.77 = 6.27$$

Reporting confidence intervals

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

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

Recommended: It allows to report both symmetrical and asymmetrical CIs in the same way.

Example

96 112 103 89 105 112 96 118 102 107

Our estimate for the mean IQ score of the population (based on our sample) is:

$M = 104.0, 95\% \text{ CI } [97.7, 110.3]$

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:

```
> 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:

```
> 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.

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

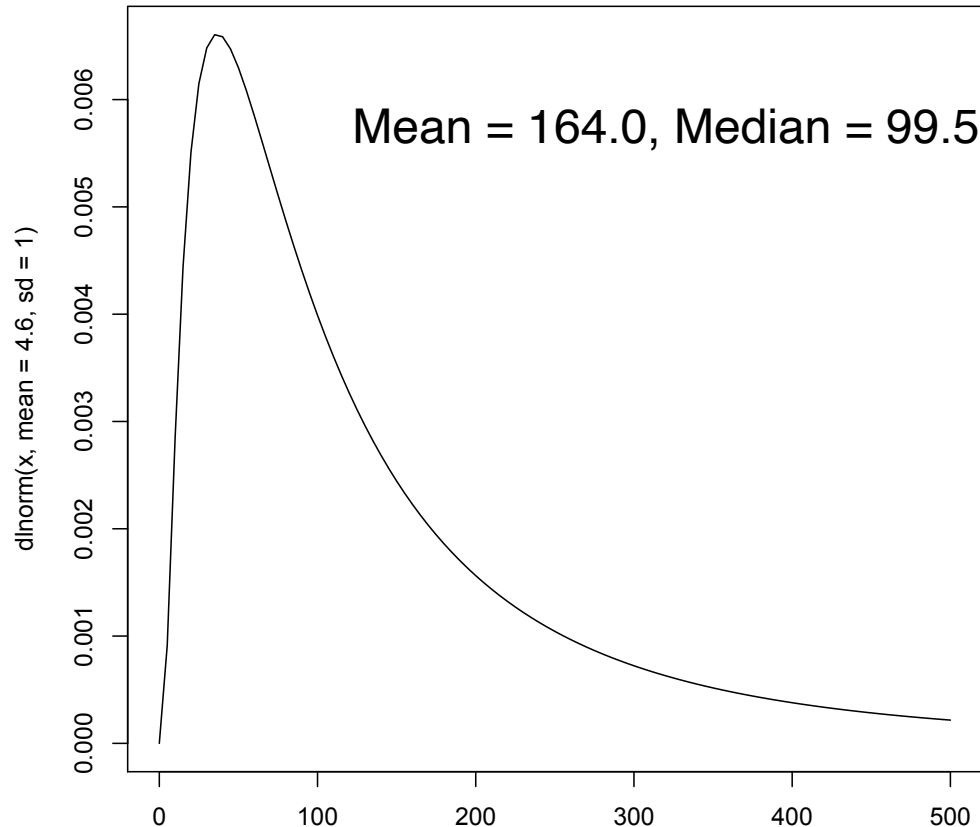
for(i in 1:N){
  sample <- rnorm(n, mean = M, sd = SD)
  ci <- t.test(sample)$conf.int

  if(ci[2] < M || ci[1] > M) count <- count + 1
}

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.



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

  if(ci[2] < M || ci[1] > M) count <- count + 1
}

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

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) 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

  if(ci[2] < M || ci[1] > M) count <- count + 1
}

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?

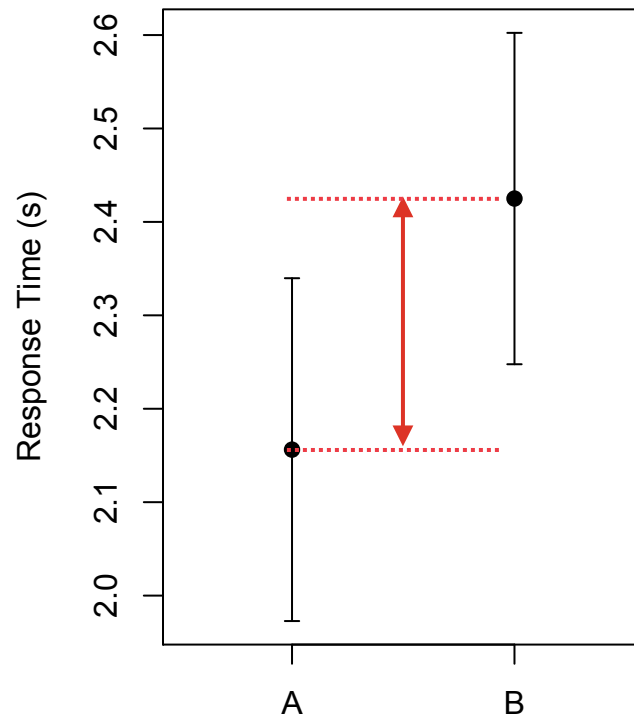
Confidence intervals for a difference

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]

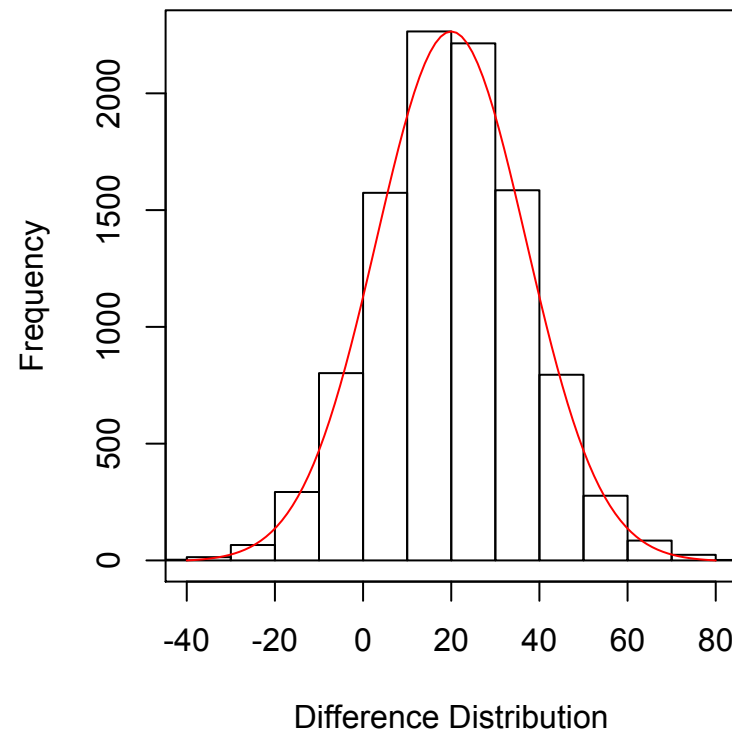
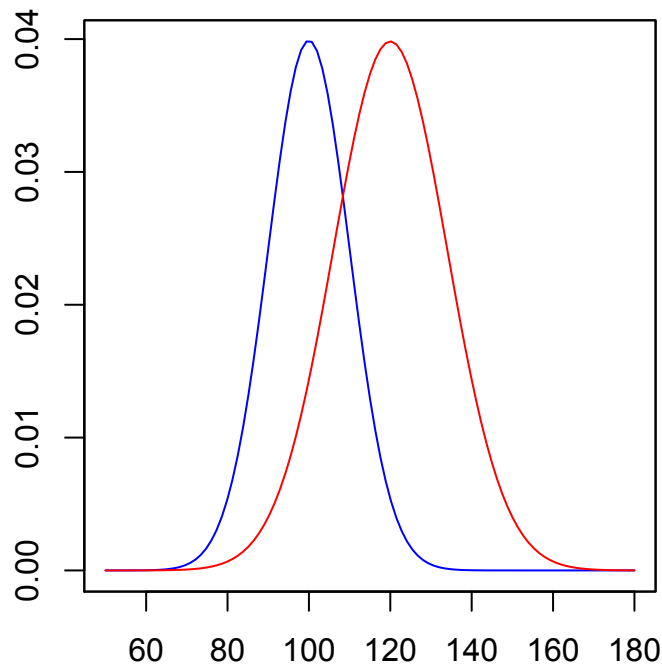
Confidence intervals for a difference

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:

$$\sigma_{\hat{\mu}_1 - \hat{\mu}_2}^2 = \sigma_{\hat{\mu}_1}^2 + \sigma_{\hat{\mu}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad \sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}}$$

Population SD
↙

If we further assume equal variances ($\sigma_1^2 = \sigma_2^2$), 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:

$$\hat{\sigma}_{\hat{\mu}_1 - \hat{\mu}_2} = \hat{\sigma}_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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:

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

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.

on)

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:

G1: adults who have, and

G2: adults who have not completed any graduate studies

From each group, they test 15 participants.

S_{G1}: 102, 104, 103, 106, 95, 108, 101, 108, 113, 96, 112,
106, 105, 109, 105

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:

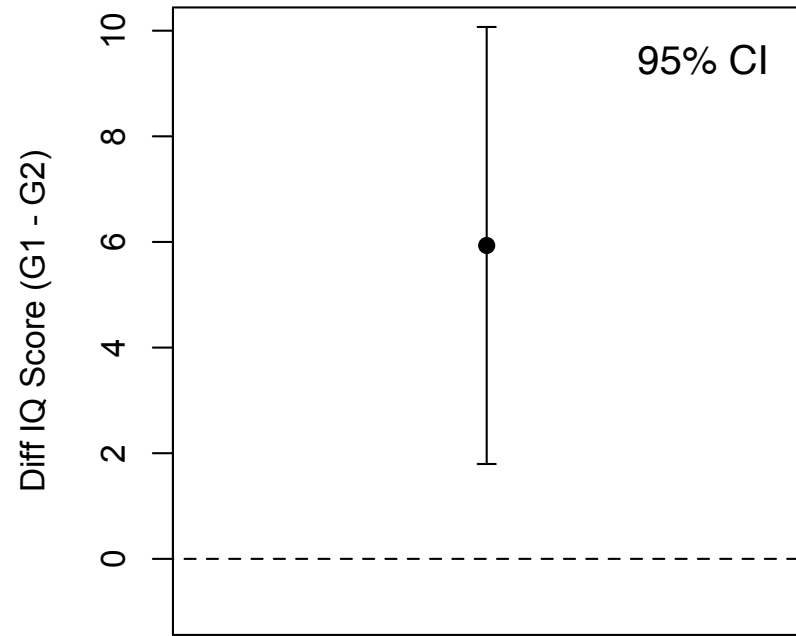
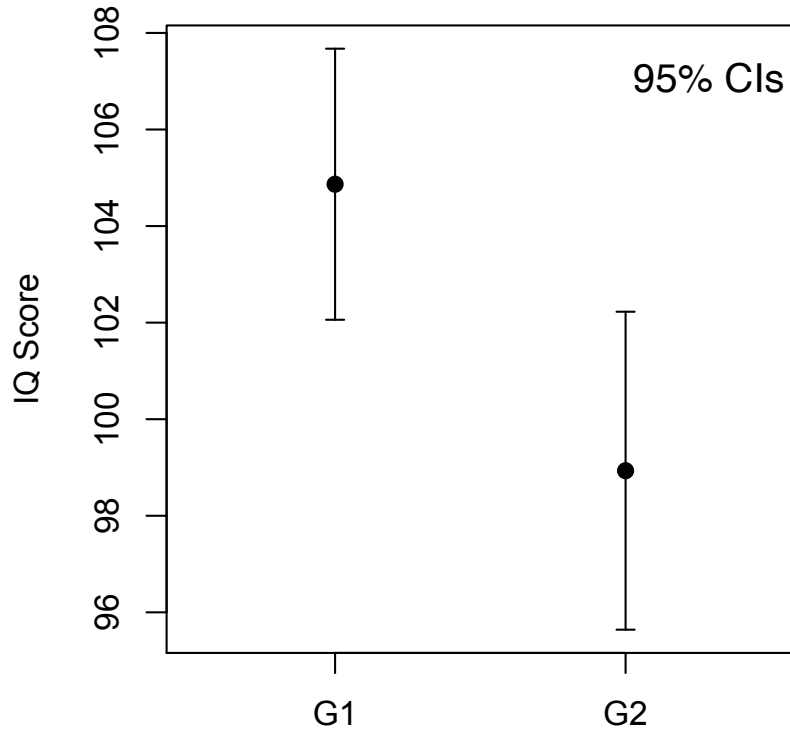
```
> g1<-c(102, 104, 103, 106, 95, 108, 101, 108, 113, 96, 112, 106, 105, 109, 105)
> g2<-c(96, 108, 90, 104, 97, 103, 95, 102, 93, 107, 101, 88, 99, 104, 97)
> t.test(g1)$conf.int
[1] 102.0595 107.6738
attr(,"conf.level")
[1] 0.95
> t.test(g2)$conf.int
[1] 95.64067 102.22600
attr(,"conf.level")
[1] 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



R code

Plotting (*plotrix* package)

```
> 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, yaxt='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, 95% CI [102.1, 107.7]

The mean IQ score of group G2 is:

M = 98.9, 95% CI [95.6, 102.2]

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

M = 5.9, 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: $n_1 \neq n_2$

Example

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_{control}. This is the control group.

G_{geometry}. The group that follows the Geometry course.

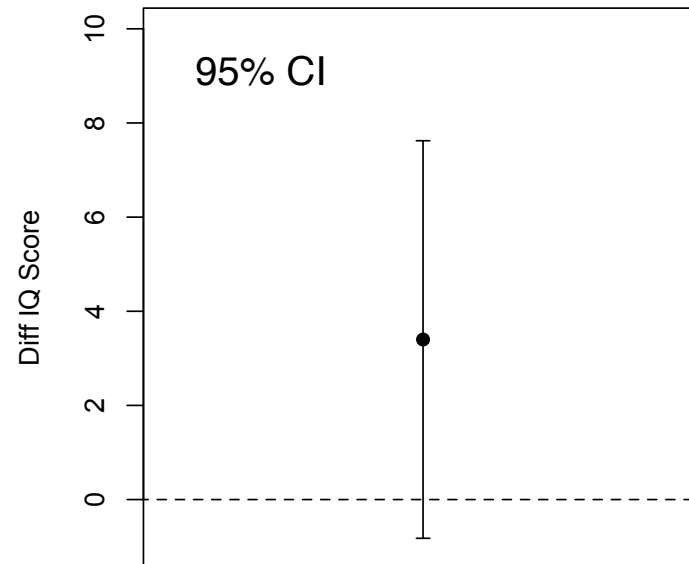
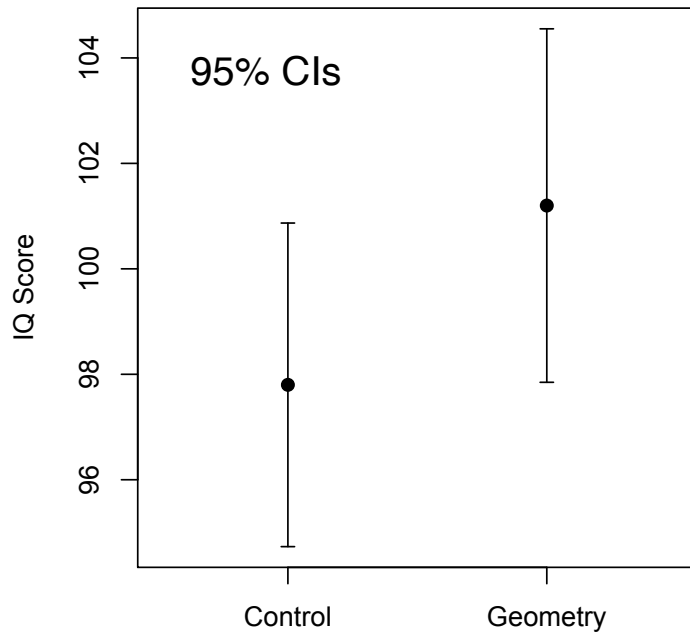
Results are as follows:

G_{control}: 102, 94, 90, 104, 95, 100, 101, 96, 100, 96

G_{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_{Before}: 102, 94, 90, 104, 95, 100, 101, 96, 100, 96

S_{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.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Before	102	94	90	104	95	100	101	96	100	96
After	106	108	93	103	100	100	105	98	103	96

CIs for related samples

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

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Before	102	94	90	104	95	100	101	96	100	96
After	106	108	93	103	100	100	105	98	103	96
Diff.	4	14	3	-1	5	0	4	2	3	0

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

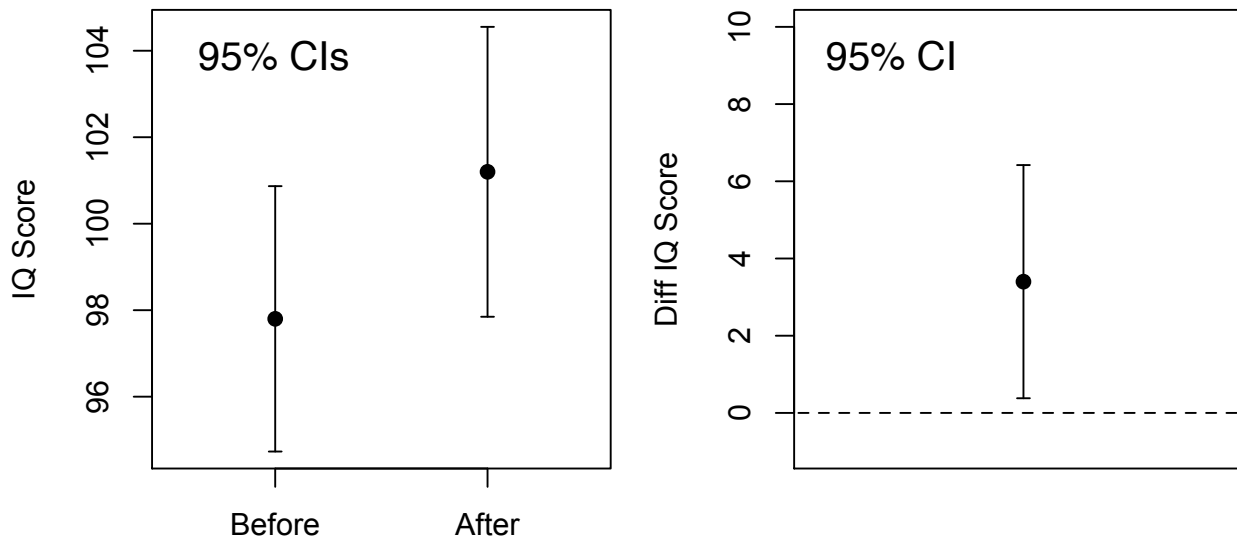
R Code

```
> 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 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**.

Control Group

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Week 1	102	94	90	104	95	100	97	96	96	96
Week 2	105	97	93	96	99	105	100	93	103	98
Week 3	101	100	93	106	98	105	102	98	104	93

Geometry Class Group

	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Week 1	105	98	100	97	103	104	103	103	100	99
Week 2	106	100	93	100	100	105	100	95	96	98
Week 3	103	101	98	106	98	100	101	95	104	97

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?

Control Group

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Week	102	94	90	104	95	100	97	96	96	96
Week	105	97	93	96	99	105	100	93	103	98
Week	101	100	93	106	98	105	102	98	104	93
Mean:	102.7	97.0	92.0	102.0	97.3	103.3	99.7	95.7	101.0	95.7

Geometry Class Group

	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Week	105	98	100	97	103	104	103	103	100	99
Week	106	100	93	100	100	105	100	95	96	98
Week	103	101	98	106	98	100	101	95	104	97
Mean:	104.7	99.7	97.0	101.0	100.3	103.0	101.3	97.7	100.0	98.0

Example

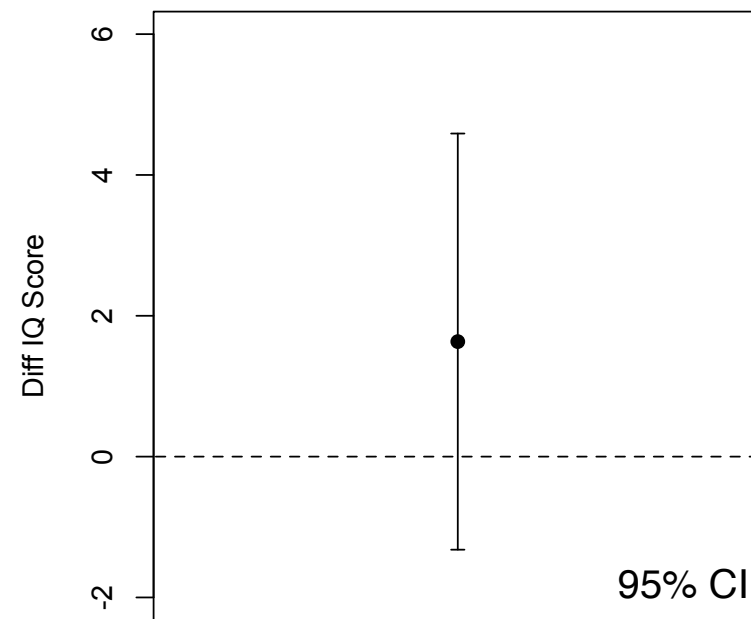
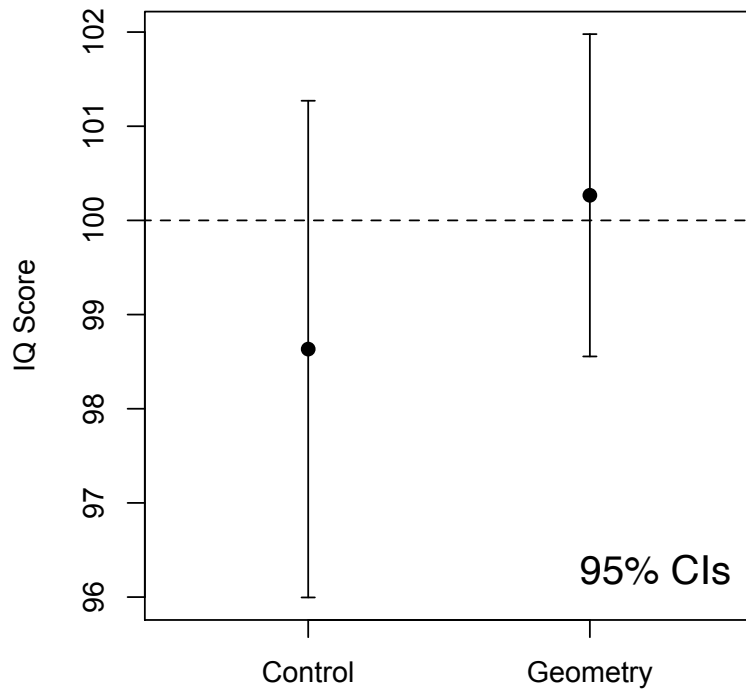
Question 1: What's the mean difference in IQ scores between students who took and students who did not take the geometry course?

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

Week	105	98	100	97	103	104	103	103	100	99
Week	106	100	93	100	100	105	100	95	96	98
Week	103	101	98	106	98	100	101	95	104	97
Mean:	104.7	99.7	97.0	101.0	100.3	103.0	101.3	97.7	100.0	98.0

Results



Next lecture

We will show how to deal with non-normal distributions...