THE UNIVERSITY of EDINBURGH

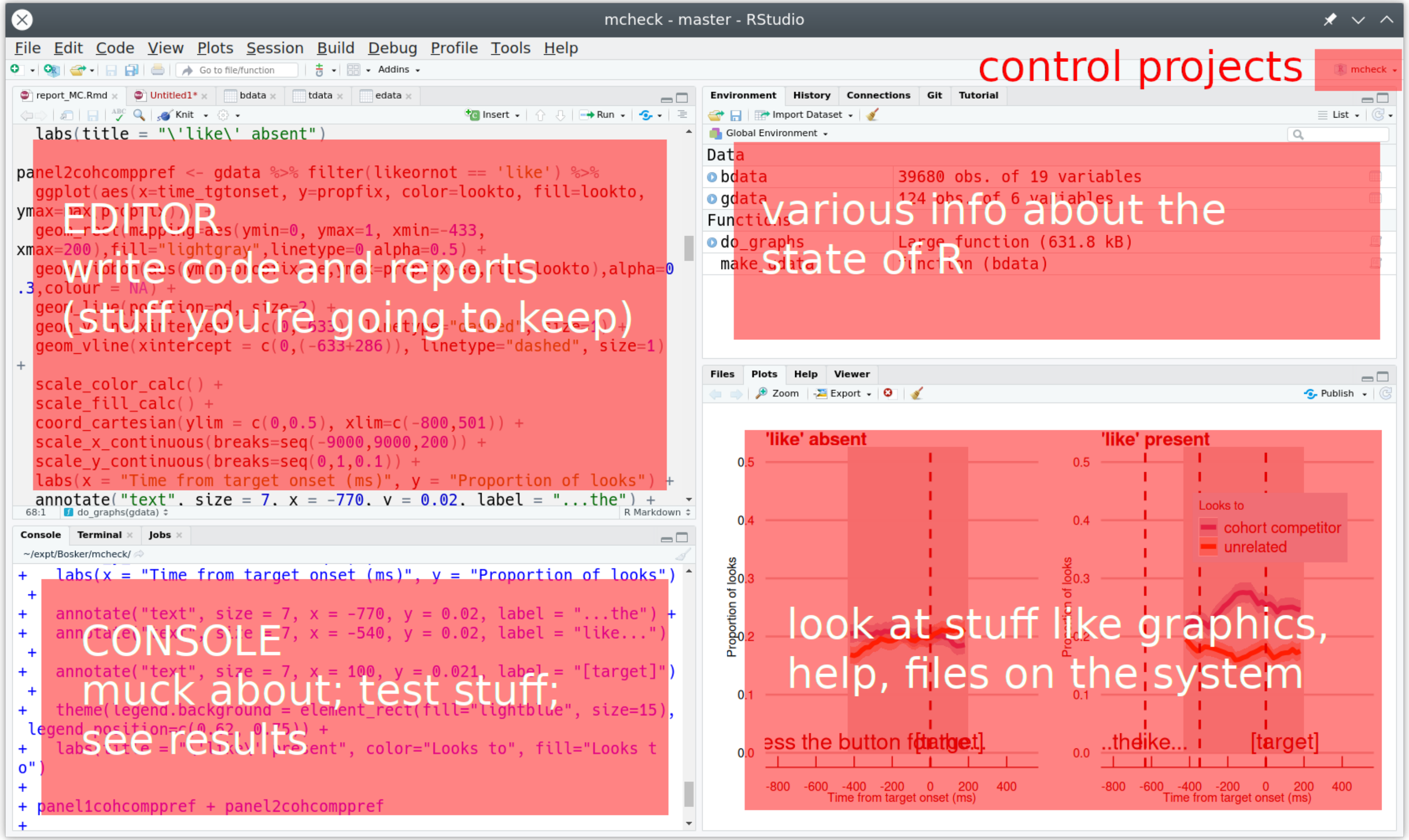
Week 7: Recap

Univariate Statistics and Methodology using R

Department of Psychology
The University of Edinburgh

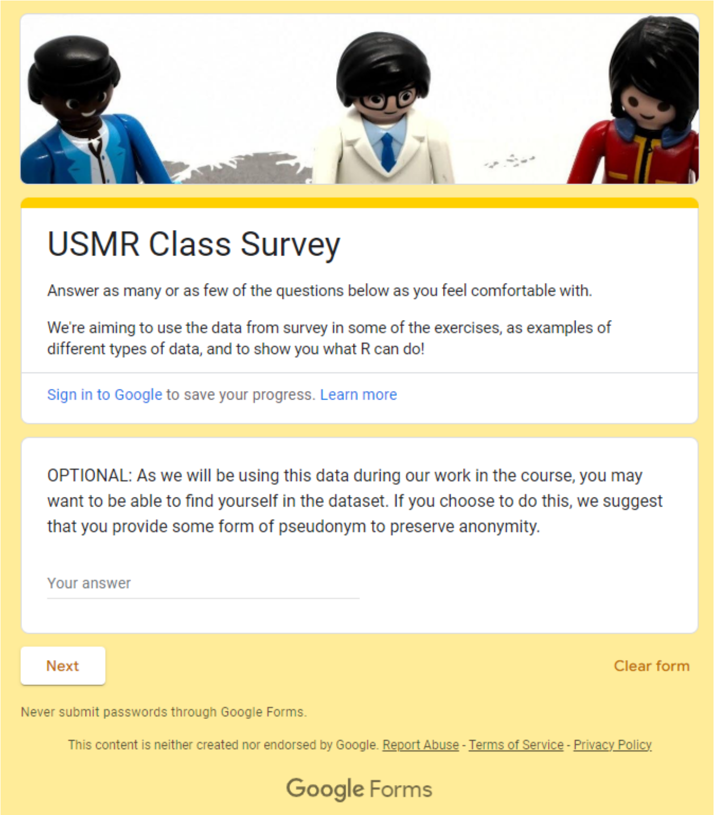
Week 1: R

Week 1: R


mcheck - master - RStudio
control projects
EDITOR
write code and reports
(stuff you're going to keep)
various info about the state of R
CONSOLE
muck about; test stuff;
see results
look at stuff like graphics,
help, files on the system

Week 2: Distributions and Repeated Sampling

We collect data



USMR Class Survey

Answer as many or as few of the questions below as you feel comfortable with.

We're aiming to use the data from survey in some of the exercises, as examples of different types of data, and to show you what R can do!

Sign in to Google to save your progress. Learn more

OPTIONAL: As we will be using this data during our work in the course, you may want to be able to find yourself in the dataset. If you choose to do this, we suggest that you provide some form of pseudonym to preserve anonymity.

Your answer

Next

Clear form

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Heights and Eye-Colours of USMR students

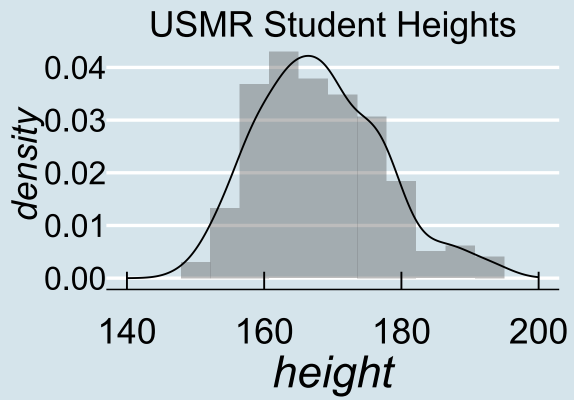
```
library(tidyverse)
demo <- read_csv("https://uoepsy.github.io/data/surveydata_allcourse22
  filter(course=="usmr") %>%
  select(height, eyecolour) %>%
  na.omit()

dim(demo)
```

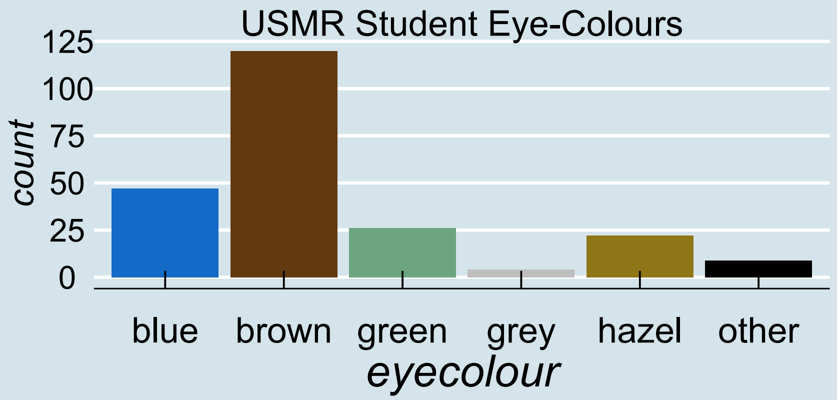
```
## [1] 228   2
```

```
head(demo)
```

```
## # A tibble: 6 × 2
##   height eyecolour
##    <dbl> <chr>
## 1   180  brown
## 2   162  brown
## 3   175  brown
## 4   175  brown
## 5   168. brown
## 6   157. brown
```

We can describe data






```
demo %>% summarise(
  mean_height = mean(height),
  sd_height = sd(height)
)
```

```
## # A tibble: 1 × 2
##   mean_height sd_height
##         <dbl>     <dbl>
## 1        168.      9.10
```

```
demo %>% count(eyecolour) %>%
  mutate(prop = n/sum(n))
```

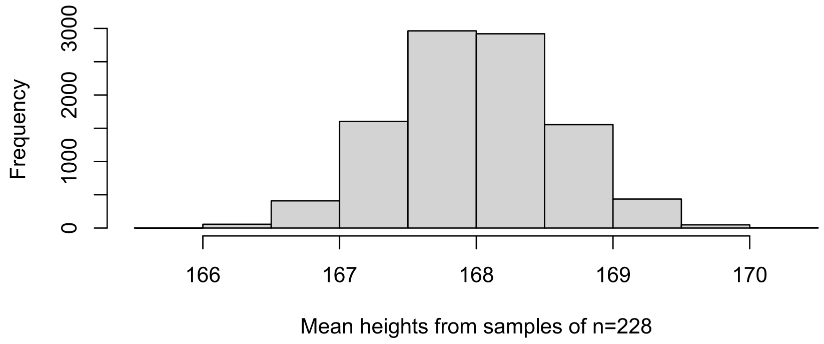
```
## # A tibble: 6 × 3
##   eyecolour     n   prop
##   <chr>     <int>  <dbl>
## 1 blue         47 0.206
## 2 brown       120 0.526
## 3 green        26 0.114
## 4 grey          4 0.0175
## 5 hazel        22 0.0965
## 6 other         9 0.0395
```

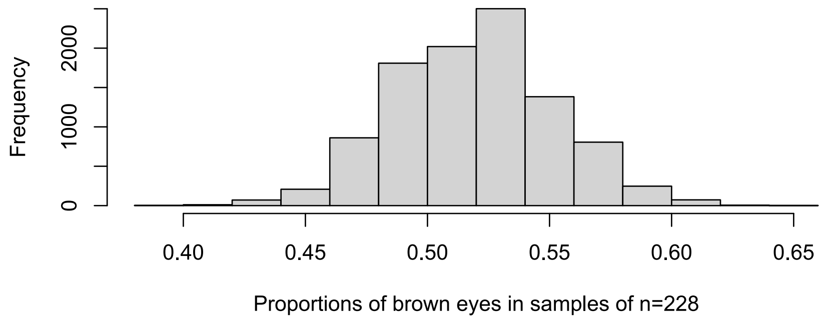
what if… we had collected different data?

- Statistics we have observed from $n = 228$:
  - mean height: 168
  - standard deviation of heights: 9.1
  - proportion of people with brown eyes: 52%
- Statistics that we *might have* computed would be different.

what if... we had collected different data?

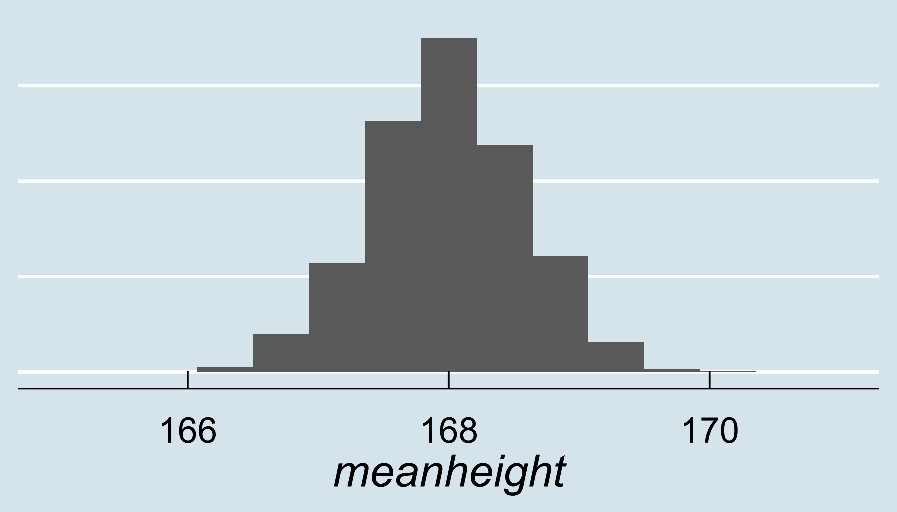
- Statistics we have observed from $n = 228$:
  - mean height: 168
  - standard deviation of heights: 9.1
  - proportion of people with brown eyes: 52%
- Statistics that we *might have* computed would be different.
- Demonstration: lecture7_demo.R





quantifying sampling variation

Simulations

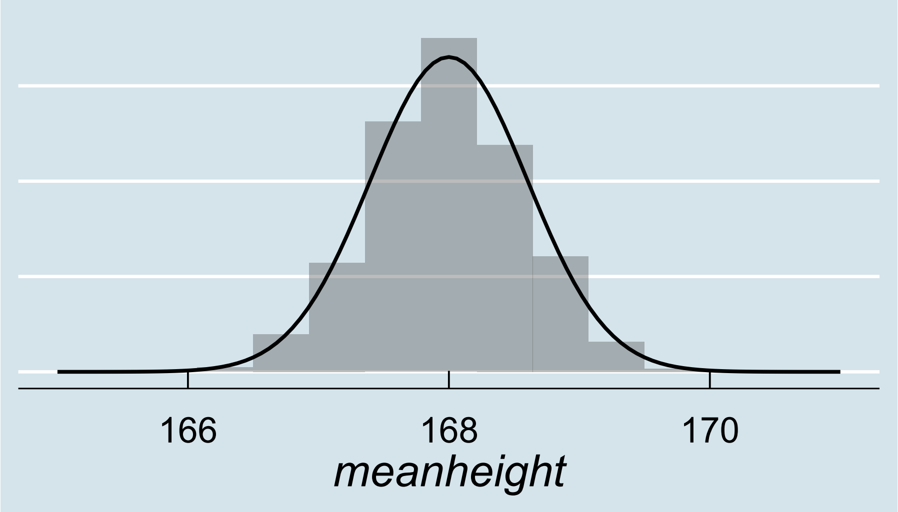



```
mheights <- replicate(1000, mean(rnorm(n = 228, mean = 168, sd = 9.1))
sd(mheights)
```

```
## [1] 0.6028
```

quantifying sampling variation (2)

Simulations




```
mheights <- replicate(1000, mean(rnorm(n = 228, mean = 168, sd = 9.1)))
sd(mheights)
```

```
## [1] 0.6028
```

Maths

sampling distribution is normally distributed with a standard deviation of:

$$\frac{\sigma}{\sqrt{n}}$$

Where:
n = sample size
σ = population standard deviation

```
9.1 / sqrt(228)
```

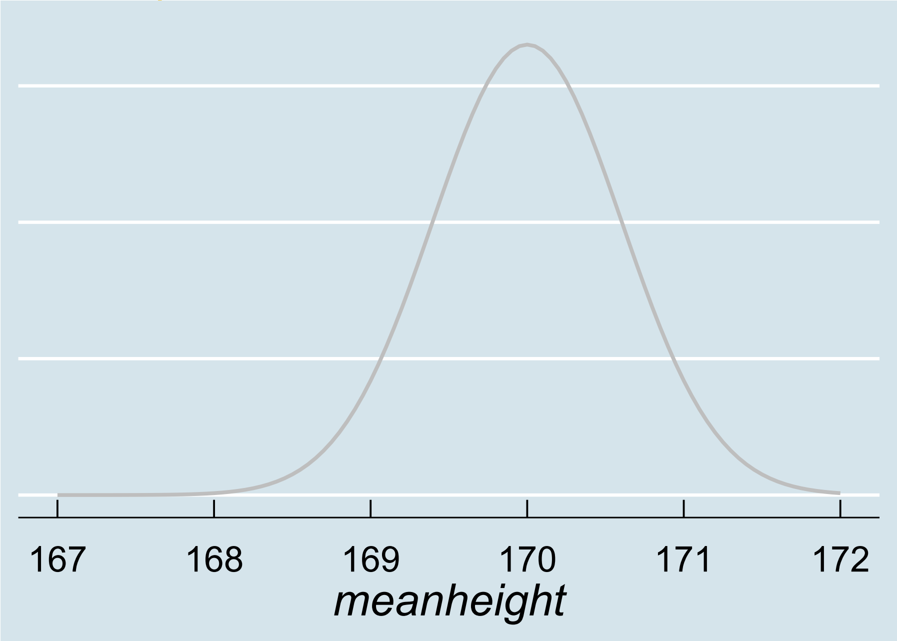
```
## [1] 0.6027
```

Week 3 - Test Statistics

What we expect vs What we observe

e.g. "If the population mean height is 170, is it unlikely to see our sample's mean height of 168?"

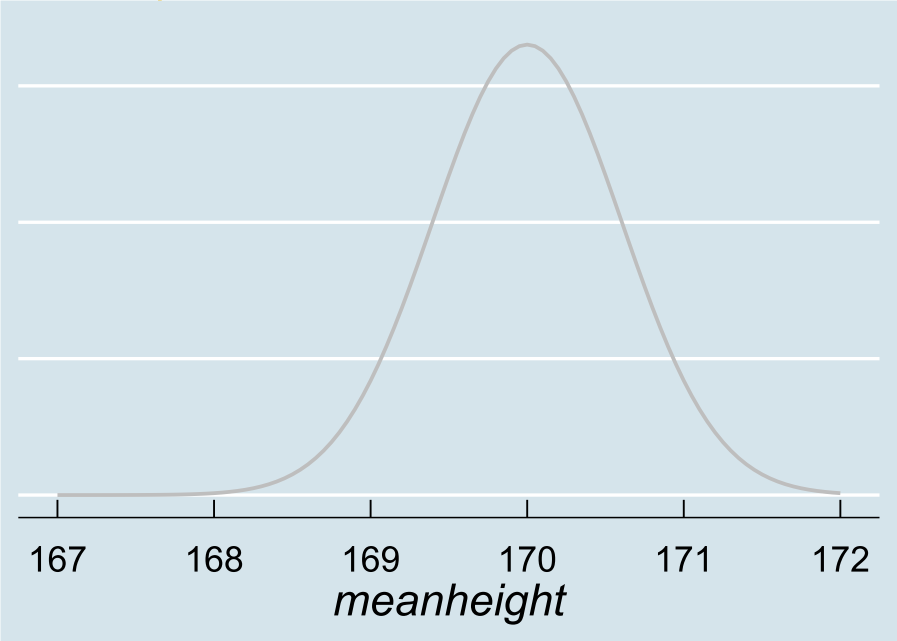
What we expect:



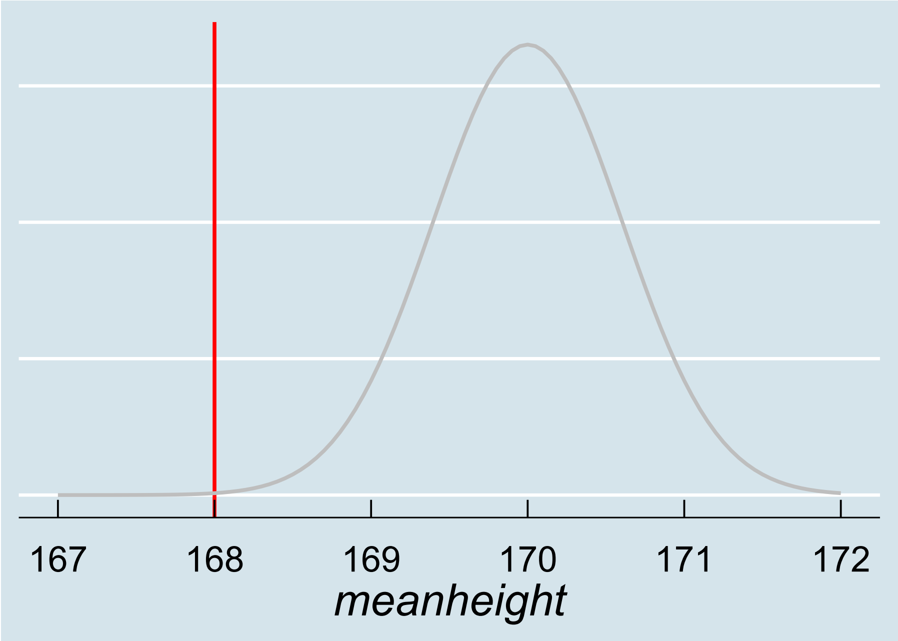
What we expect vs What we observe

e.g. "If the population mean height is 170, is it unlikely to see our sample's mean height of 168?"

What we expect:



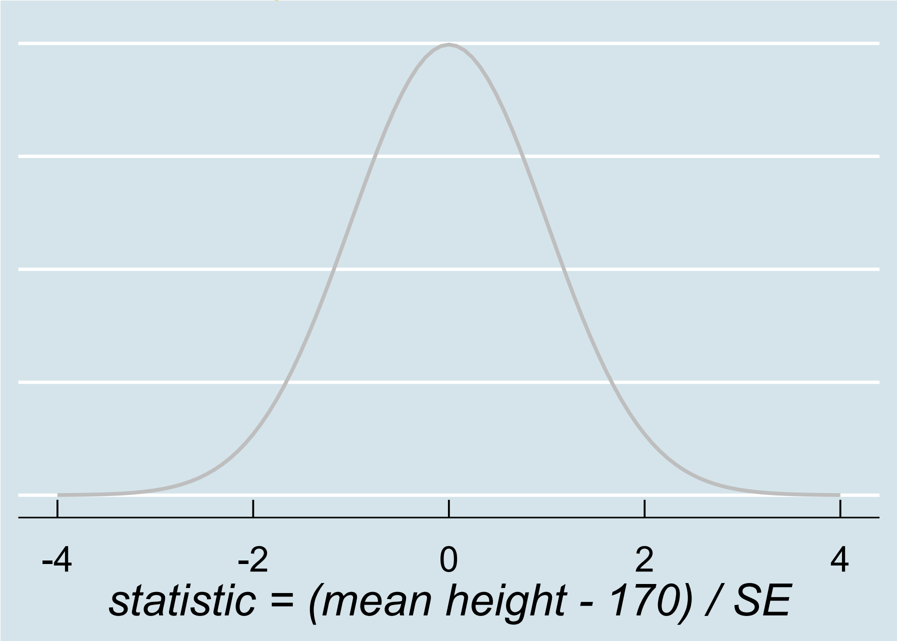
What we observe:



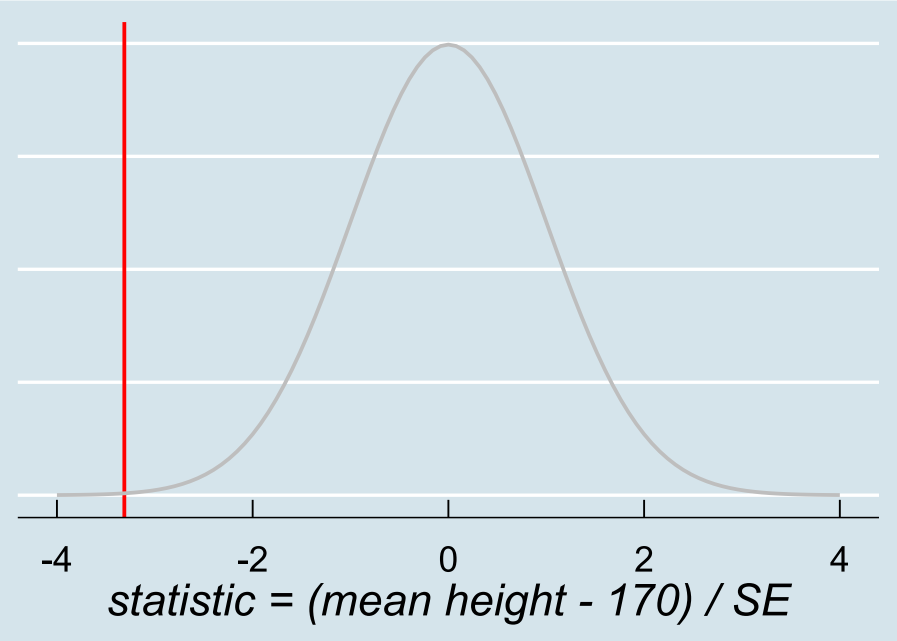
Standardised Test Statistics

e.g. "Given the statistics we expect, how unlikely is the statistic we observe?"

Test Statistics we expect:



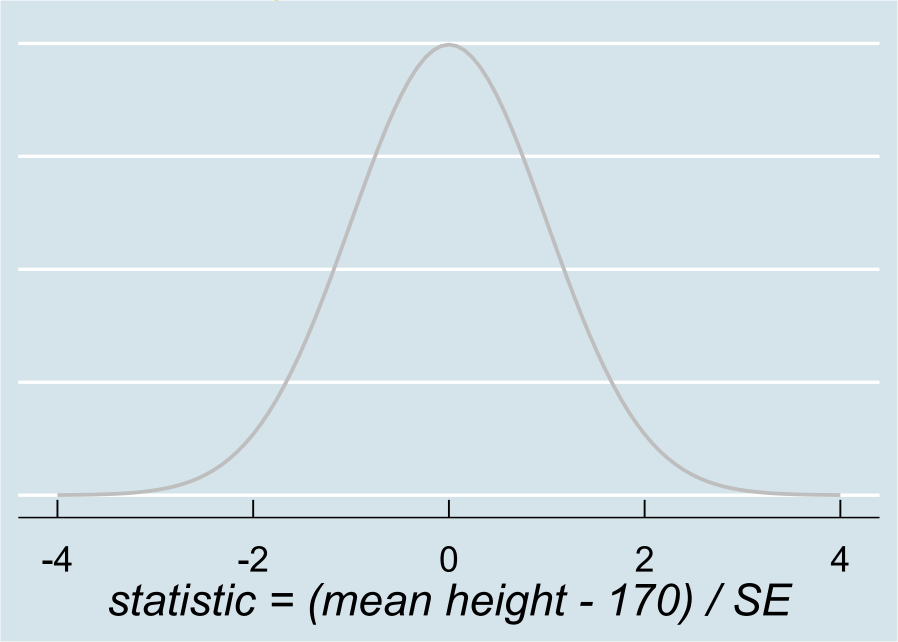
Test Statistics we observe:



sampling variation in theory

e.g. "Given the statistics we expect, how unlikely is the statistic we observe?"

Test Statistics we expect:



$$SE = \frac{\sigma}{\sqrt{n}}$$

Where:

$n =$ sample size

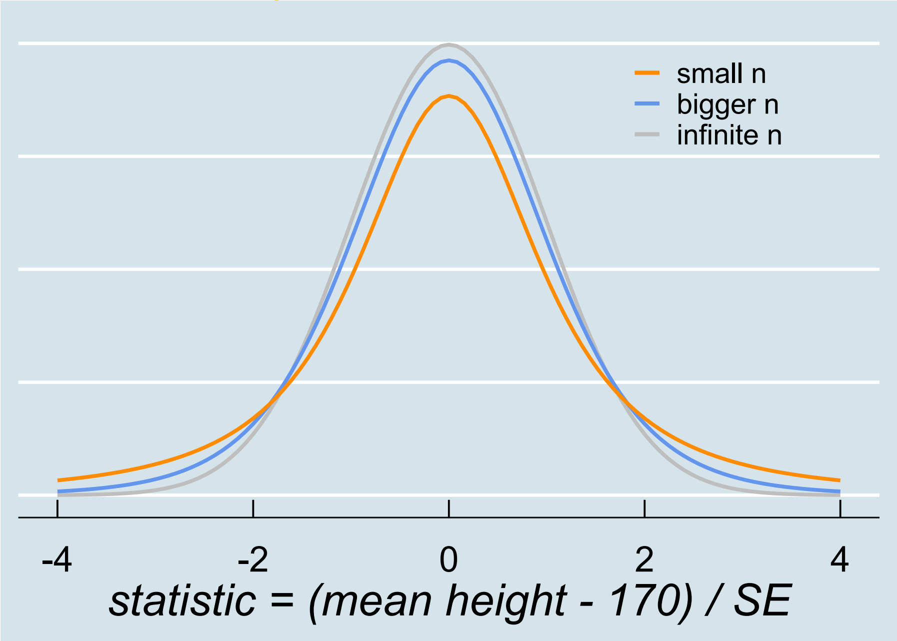
$\sigma =$ population standard deviation

$$z = \frac{168 - 170}{\frac{??}{\sqrt{228}}}$$

sampling variation in practice

e.g. "Given the statistics we expect, how unlikely is the statistic we observe?"

Test Statistics we expect:



$$SE = \frac{s}{\sqrt{n}}$$

Where:

n = sample size

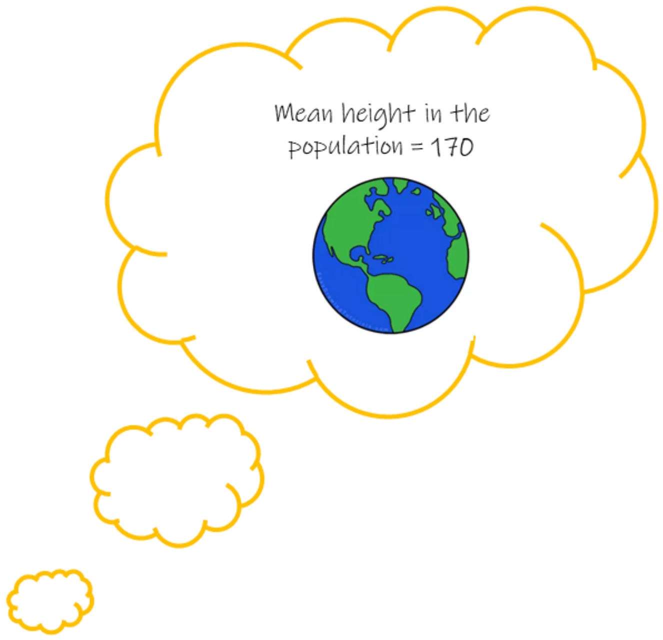
s = sample standard deviation

$$t = \frac{168 - 170}{\frac{9.01}{\sqrt{228}}}$$

statistical testing

1. Assume the null hypothesis is true

2. How likely would we be to obtain our statistic in a universe where the null hypothesis is true?



statistical testing

1. Assume the null hypothesis is true
2. How likely would we be to obtain our statistic in a universe where the null hypothesis is true?


```
usmr <- read_csv("https://uoepsy.github.io/data/surveydata_allcourse22.csv") %>%
  filter(course=="usmr") %>%
  filter(!is.na(height), !is.na(eyecolour))
```

t-tests

One sample t-test

- how far the sample mean is from some number:

$$t = \frac{\bar{x}_1 - \mu_0}{SE_{\bar{x}}}$$

```
t.test(usmr$height, mu = 170)
```

```
##
##     One Sample t-test
##
## data:  usmr$height
## t = -3, df = 227, p-value = 0.003
## alternative hypothesis: true mean is not equal to 170
## 95 percent confidence interval:
##  167.0 169.4
## sample estimates:
## mean of x
##     168.2
```

t-tests (2)

Two sample t-test

- how far the difference in means is from zero:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE_{\bar{x}_1 - \bar{x}_2}}$$

```
t.test(height ~ catdog, data = usmr)
```

```
##
##     Welch Two Sample t-test
##
## data:  height by catdog
## t = -1.5, df = 195, p-value = 0.1
## alternative hypothesis: true difference in means between group cat and group dog is not equal to 0
## 95 percent confidence interval:
##  -4.2254  0.6117
## sample estimates:
## mean in group cat mean in group dog
##             167.1             168.9
```

Week 4 & 5 - more test statistics

Chi-square tests

Chi-squared: Goodness of Fit

$$\chi^2 = \Sigma \frac{(Observed - Expected)^2}{Expected}$$

```
table(usmr$eyecolour)
```

```
##
##   blue brown green  grey hazel other
##     47   120    26     4    22     9
```

```
chisq.test(table(usmr$eyecolour))
```

```
##
##      Chi-squared test for given probabilities
##
## data:  table(usmr$eyecolour)
## X-squared = 242, df = 5, p-value <2e-16
```

Chi-square tests (2)

Chi-squared: Test of Independence

$$\chi^2 = \Sigma \frac{(Observed - Expected)^2}{Expected}$$

```
table(usmr$ampm, usmr$catdog)
```

```
##
##                  cat dog
##   Evening person  33  45
##   Morning person  18  27
```

```
chisq.test(table(usmr$ampm, usmr$catdog))
```

```
##
##     Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(usmr$ampm, usmr$catdog)
## X-squared = 0.0036, df = 1, p-value = 1
```

Correlation tests

Correlation

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

```
cor.test(usmr$sleeprating, usmr$loc)
```

```
##
##     Pearson's product-moment correlation
##
## data:  usmr$sleeprating and usmr$loc
## t = 3.4, df = 74, p-value = 0.001
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1588 0.5504
## sample estimates:
##   cor
## 0.371
```

and here we are!