T-Test: One-Sample

Data Analysis for Psychology in R 1

Semester 2, Week 6

Dr Emma Waterston

Department of Psychology
The University of Edinburgh

Learning Objectives

- Understand when to use a one sample t-test
- Understand the null hypothesis for a one sample t-test
- Understand how to calculate the test statistic
- Know how to conduct the test in R

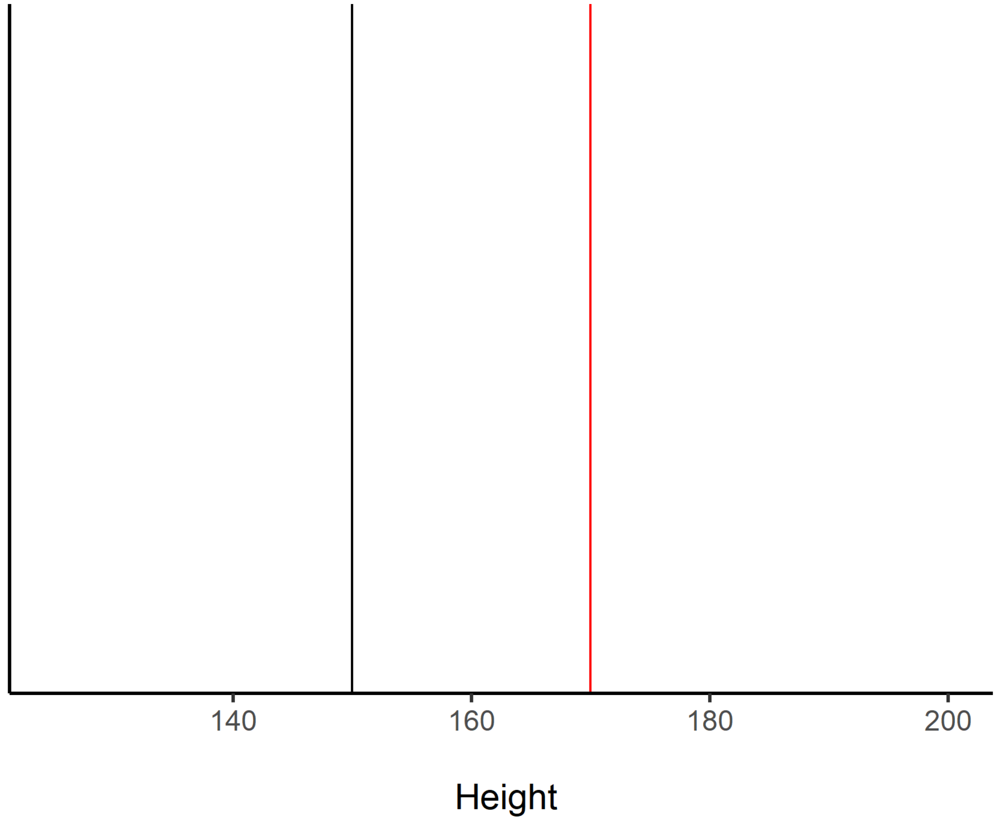
Topics for Today

- Introduce the three types of t-test
- One-sample t-test example
- Inferential tests for the one-sample t-test
- Assumptions and effect size

T-Test: Purpose

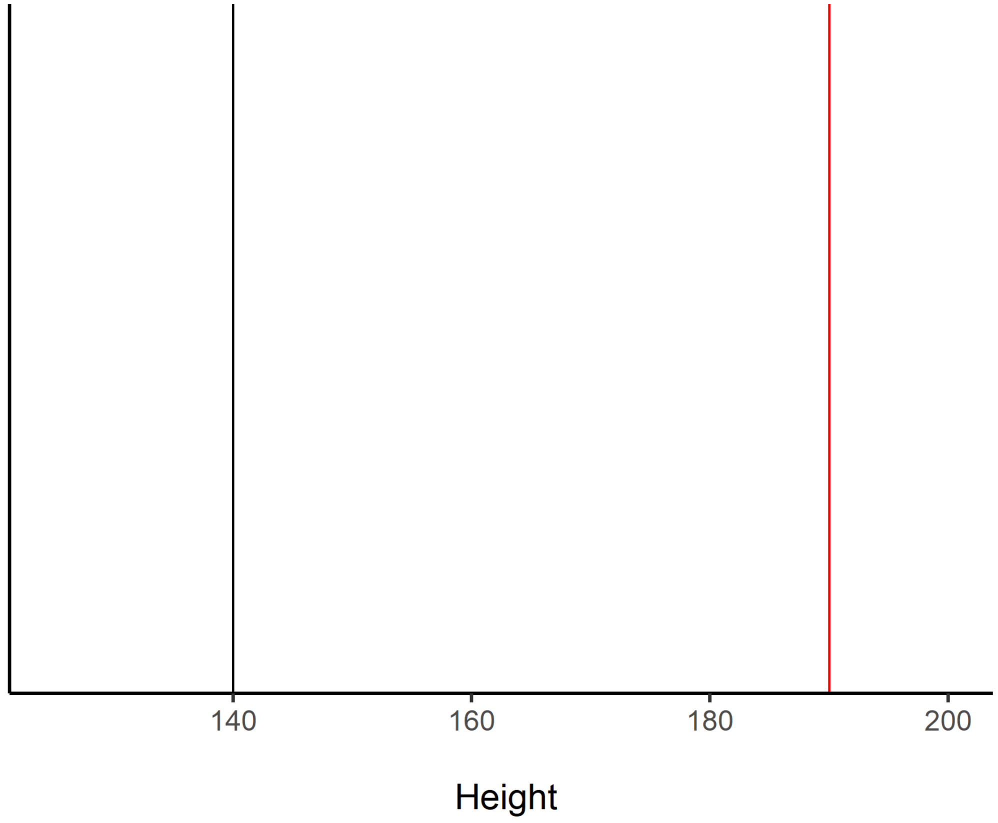
- t-tests (generally) concern testing the difference between two means.
  - Another way to state this is that the scores of two groups being tested are from the sample underlying population distribution.
- One-sample t-tests compare the mean in a sample to a known mean.
- Independent t-tests compare the means of two independent samples.
- Paired sample t-tests compare the mean from a single sample at two points in time (repeated measurements)
- We will look in more detail at these tests over the next three weeks.
  - But let's start by thinking a little bit about the logic t-tests.
  - For the next few slides, have a bit of paper and a pen handy.

Are these means different?



- Write down whether you think these means (two lines) are different. Write either:
    - Yes
    - No
    - It depends

What about these?



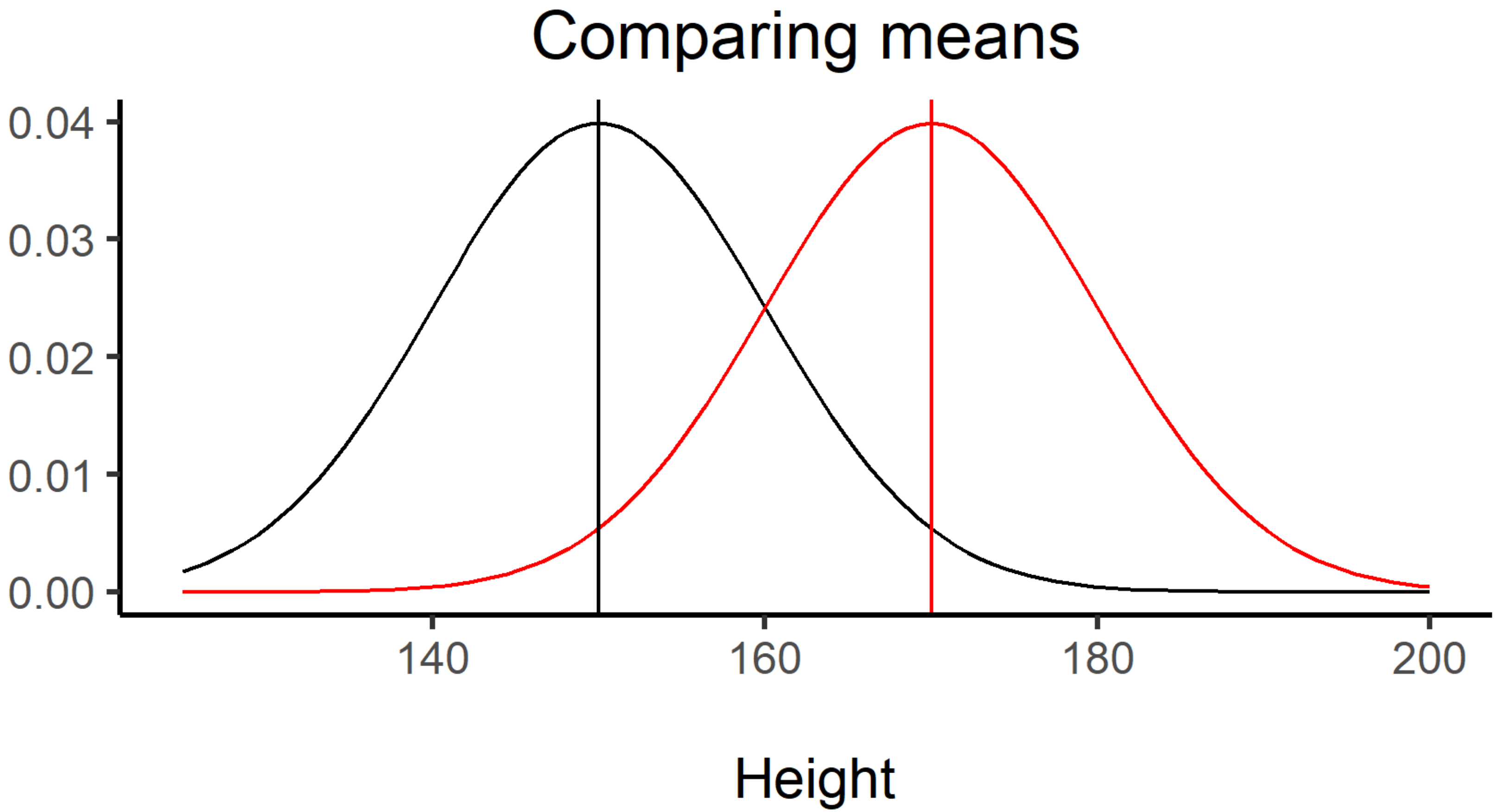
- Write down whether you think these means (two lines) are different. Write either:
  - Yes
  - No
  - It depends

Differences in means

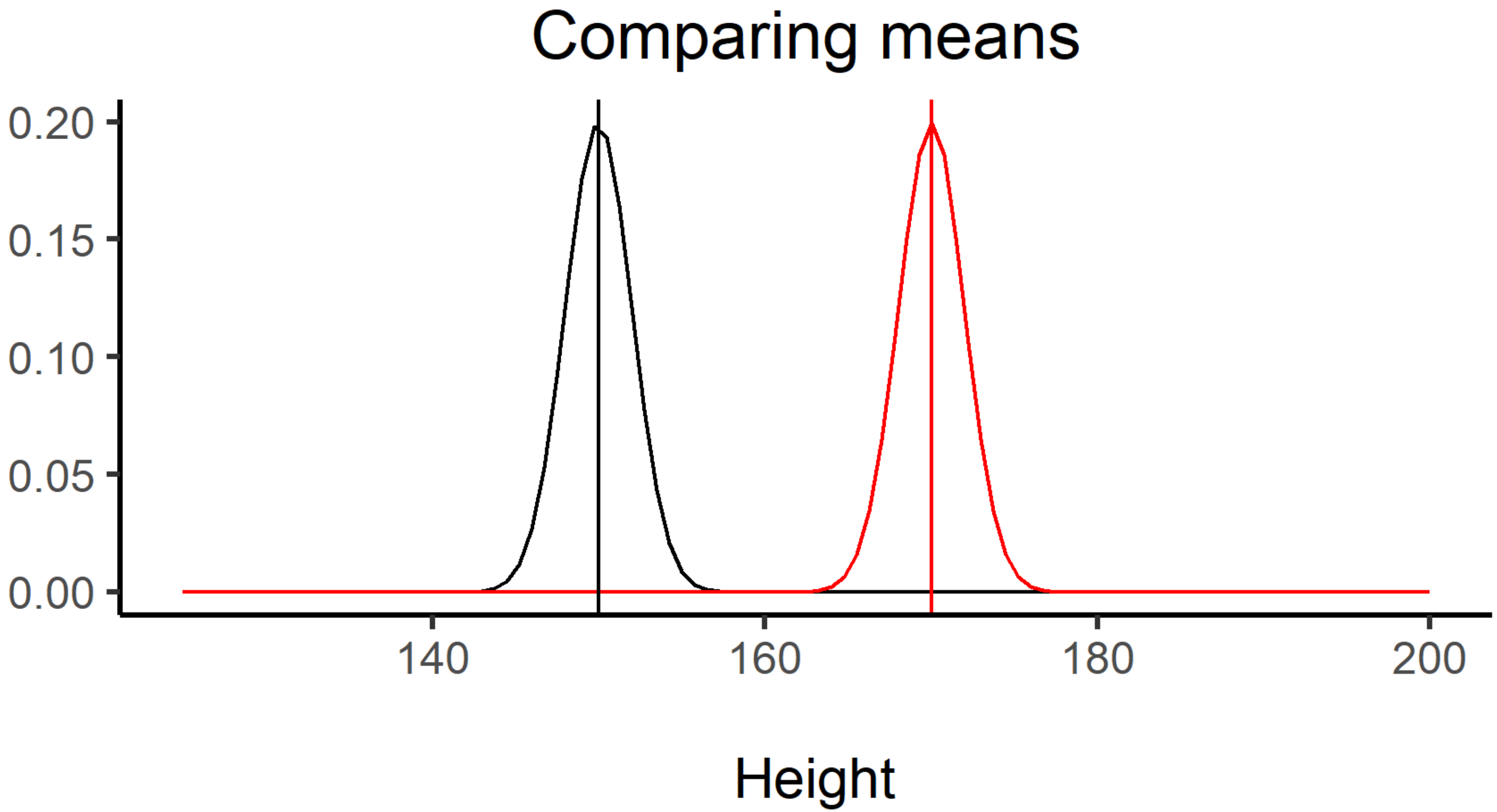
- OK, now please write down:

1. Why you wrote the answers you did?
2. If you wrote, "It depends", why can we not tell whether they are different or not?
3. What else might we want to know in order to know whether not the group means could be thought of as coming from the same distribution?

All the information



All the information



Questions?

t-statistic

- Recall when talking about hypothesis testing:
  - We calculate a test statistic that represents our question.
  - We compare our sample value to the sampling distribution under the null
- Here the test statistic is a t-statistic.

t-statistic

$$t = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \qquad \text{where} \qquad SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- The numerator = a difference in means
- The denominator = a estimate of variability
  - where
  - s = sample estimated standard deviation of x
  - n = sample size
- t = a standardized difference in means

Data Requirements: One-sample t-test

- A continuous variable
  - Remember we are calculating means
- A known mean that we wish to compare our sample to
- A sample of data from which we calculate the sample mean

Example

- Suppose I want to know whether the retirement age of Professors at my University is the same as the national average.
- The national average age of retirement for Prof's is 65.
- So I look at the age of the last 40 Prof's that have retired at Edinburgh and compare against this value.

Data

```
## # A tibble: 40 × 2
##    ID       Age
##    <chr>  <dbl>
##  1 Prof1     76
##  2 Prof2     66
##  3 Prof3     58
##  4 Prof4     68
##  5 Prof5     79
##  6 Prof6     74
##  7 Prof7     75
##  8 Prof8     50
##  9 Prof9     69
## 10 Prof10    70
## # … with 30 more rows
```

Hypotheses

- When we are testing whether the population mean (μ) is equal to a hypothesized value (μ_0).

$$H_0 : \mu = \mu_0$$

- Note this is identical to saying:

$$H_0 : \mu - \mu_0 = 0$$

Alternative Hypotheses

- Two-tailed:

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0$$

- One-tailed:

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

Hypotheses

- Let's assume a priori we have no idea of the ages the Prof's retired.
- So I specify a two-tailed hypothesis with α = .05.
- So I am simply asking, does my mean differ from the known mean.

Calculation

$$t = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \qquad \text{where} \qquad SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- Steps to calculate t:
  - Calculate the sample mean $(\bar{x})$.
  - Calculate the standard error of the mean $(\frac{s}{\sqrt{n}})$.
  - Calculate the sample standard deviation (s).
  - Check I know my sample size (n).
  - Use all this to calculate t.

Calculation

$$t = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \qquad \text{where} \qquad SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

```
dat %>%
  summarise(
    mu0 = 65,
    xbar = mean(Age),
    s = sd(Age),
    n = n()
  ) %>%
  mutate(
    se = s/sqrt(n)
  )  %>%
  kable(digits = 2) %>%
  kable_styling(full_width = FALSE)
```

mu0	xbar	s	n	se
65	66.3	10.01	40	1.58

Calculation

mu0	xbar	s	n	se
65	66.3	10.01	40	1.58

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{66.30 - 65.00}{\frac{10.01}{\sqrt{40.00}}} = \frac{1.30}{1.58} = 0.82$$

- So in our example $t = 0.82$

Questions?

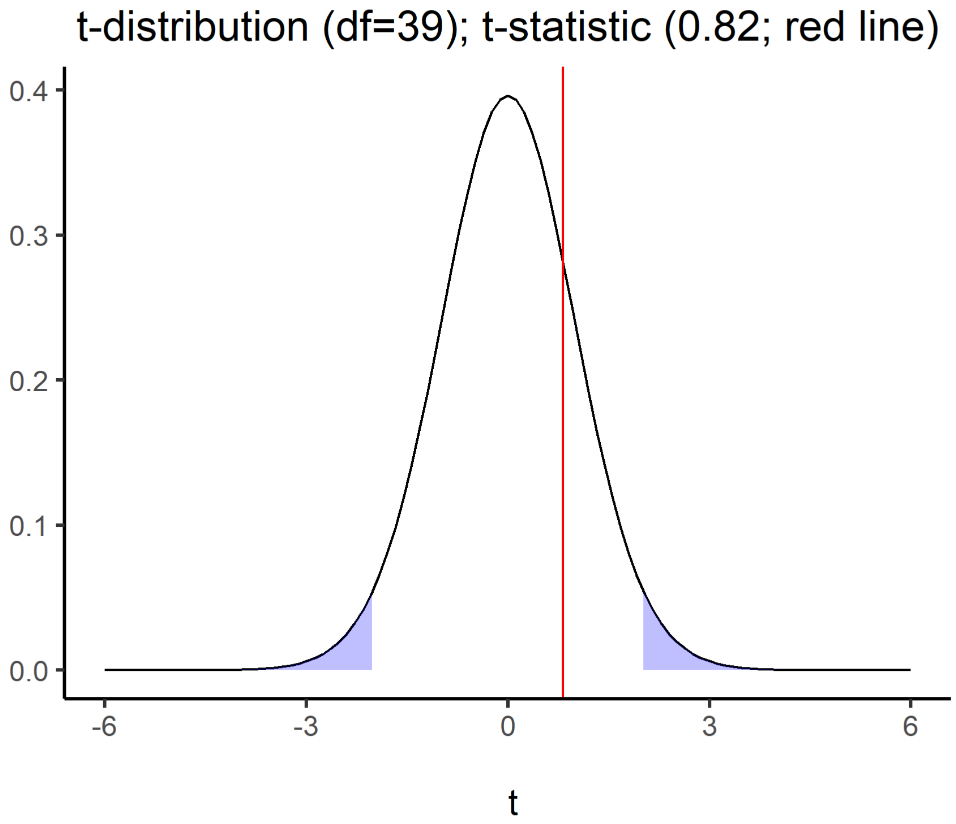
Is our test significant?

- The sampling distribution for t-statistics is a t-distribution.
- The t-distribution is a continuous probability distribution very similar to the normal distribution.
  - Key parameter = degrees of freedom (df)
  - df are a function of n.
  - As n increases (and thus as df increases), the t-distribution approaches a normal distribution.
- For a one sample t-test, we compare our test statistic to a t-distribution with n-1 df.

Is our test significant?

- So we have all the pieces we need:
  - Degrees of freedom = n-1 = 40-1 = 39
  - We have our t-statistic (0.82)
  - Hypothesis to test (two-tailed)
  - α level (.05).
- So now all we need is the critical value from the associated t-distribution in order to make our decision.

Is our test significant?



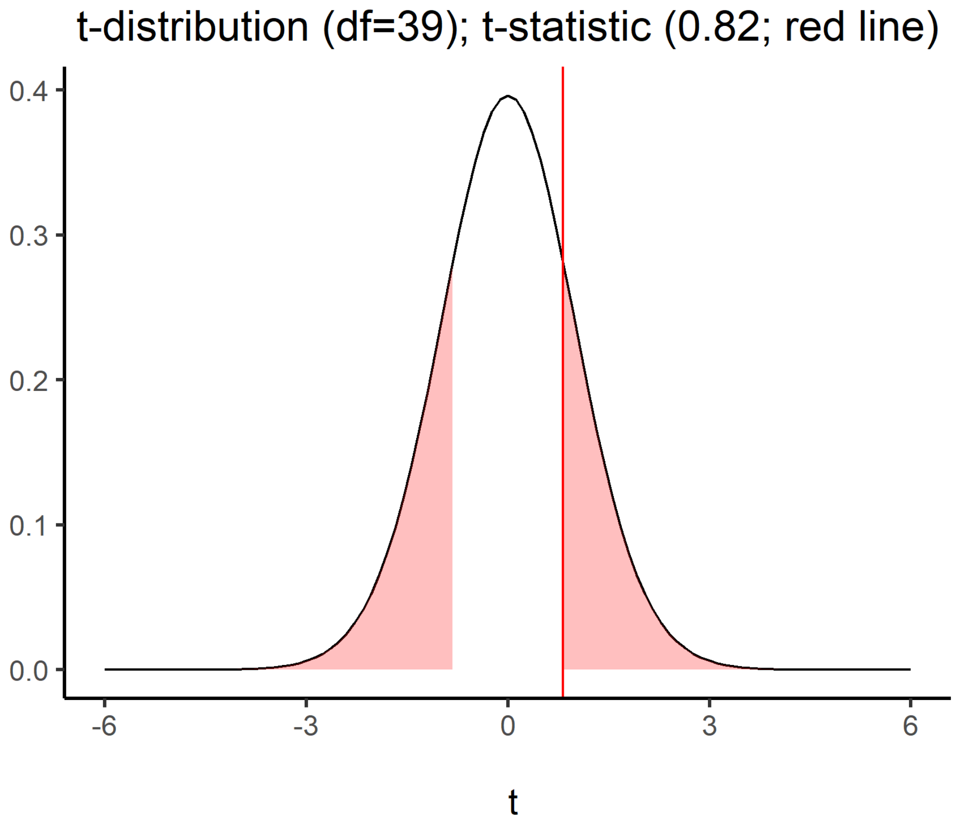

```
tibble(
  LowerCrit = round(qt(0.025, 39),2),
  UpperCrit = round(qt(0.975, 39),2),
)
```

```
## # A tibble: 1 × 2
##   LowerCrit UpperCrit
##       <dbl>     <dbl>
## 1     -2.02      2.02
```

Is our test significant?

- So our critical value is 2.02
  - Our t-statistic (0.82) is closer to 0 than this.
  - So we **fail to reject the null hypothesis**.
- $t(39) = 0.82, p > .05$, two-tailed.

Exact p-values




```
## # A tibble: 1 × 1
##   Exactp
##    <dbl>
## 1   0.42
```

In R: Types of Hypothesis

- `alternative =` refers to the direction of our alternative hypothesis (H_1)
  - $\mu < \mu_0$: `alternative="less"`
  - Our Edinburgh Prof's have a lower retirement age than the national average
  - $\mu > \mu_0$: `alternative="greater"`
  - Our Edinburgh Prof's have a higher retirement age than the national average
  - $\mu \neq \mu_0$: `alternative="two-sided"`
  - Our Edinburgh Prof's have a different retirement age than the national average

```
t.test(dat$Age, mu=65, alternative="______")
```

Our test: In R

```r
t.test(dat$Age, mu=65, alternative="two.sided")
```

```
##
##      One Sample t-test
##
## data:  dat$Age
## t = 0.82, df = 39, p-value = 0.4
## alternative hypothesis: true mean is not equal to 65
## 95 percent confidence interval:
##  63.1 69.5
## sample estimates:
## mean of x
##      66.3
```

Write up

A one-sample t-test was conducted to determine there was a statistically significant $(\alpha = .05)$ mean difference between the average retirement age of Professors and the age at retirement of a sample of 40 Edinburgh Professors. Although the sample had a higher average age of retirement (Mean=66.3, SD=10.01) than the population (Mean = 65), this difference was not statistically significant $(t(39) = 0.82, p > .05, two - tailed)$.

Questions?

Assumption checks summary

	Description	One-Sample t-test	Independent Sample t-test	Paired Sample t-test
Normality	Continuous variable (and difference) is normally distributed.	Yes (Population)	Yes (Both groups/ Difference)	Yes (Both groups/ Difference)
Tests:	Descriptive Statistics; Shapiro-Wilks Test; QQ-plot			
Independence	Observations are sampled independently.	Yes	Yes (within and across groups)	Yes (within groups)
Tests:	None. Design issue.			
Homogeneity of variance	Population level standard deviation is the same in both groups.	NA	Yes	Yes
Tests:	F-test			
Matched Pairs in data	For paired sample, each observation must have matched pair.	NA	NA	Yes
Tests:	None. Data structure issue.			

Assumptions

- As noted above, we have some requirements of the data, and we have model assumptions for the test to be valid:
  - DV is continuous
  - Independence - the data are independent
  - Normality - The data are normally distributed **OR** the sample size is sufficiently large (rule of thumb $n = 30$)

If any of these assumptions are not met, the results of the test are unreliable

Assumptions: How to check/test

- DV is continuous
  - The dependent variable should be measured at the interval or ratio level
- Independence
  - More of a study design issue, and cannot directly test
- Normality
  - Can be checked visually with plots, as well as with descriptive statistics, and a Shapiro-Wilks Test

Assumption checks: Normality

- Descriptive statistics:
  - Skew:
    - Below are some rough guidelines on how to interpret skew.
    - No strict cuts for skew - these are loose guidelines.

Verbal label	**Magnitude of skew in absolute value**
Generally not problematic	$\lvert Skew \rvert < 1$
Slight concern	$1 > \lvert Skew \rvert < 2$
Investigate impact	$\lvert Skew \rvert > 2$

Skew

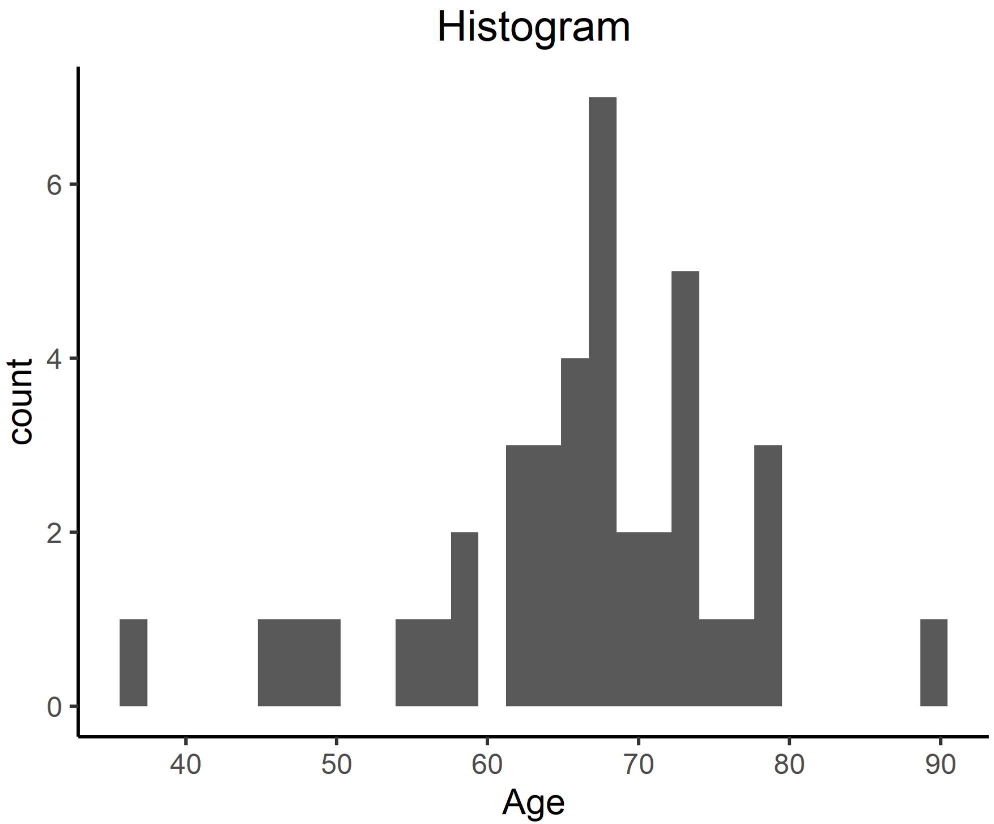
```
library(psych)
dat %>%
  summarise(
    skew = round(skew(Age),2)
  )
```

```
## # A tibble: 1 × 1
##    skew
##   <dbl>
## 1 -0.63
```

- Skew is low (< 1), so we would conclude that it is not problematic.

Histograms

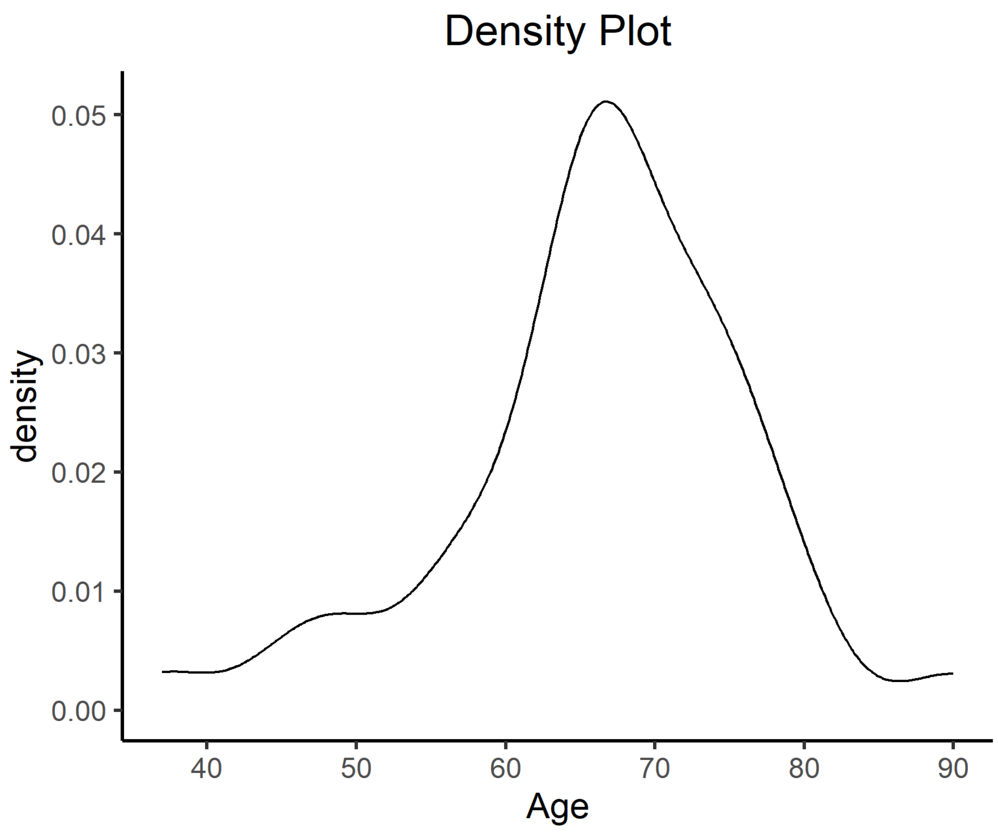
```r
ggplot(dat, aes(x=Age)) +
  geom_histogram() +
  labs(title = "Histogram")
```



- Our histogram looks "lumpy", but we have relatively low n for looking at these plots.

Density

```
ggplot(dat, aes(x=Age)) +
  geom_density() +
  labs(title = "Density Plot")
```



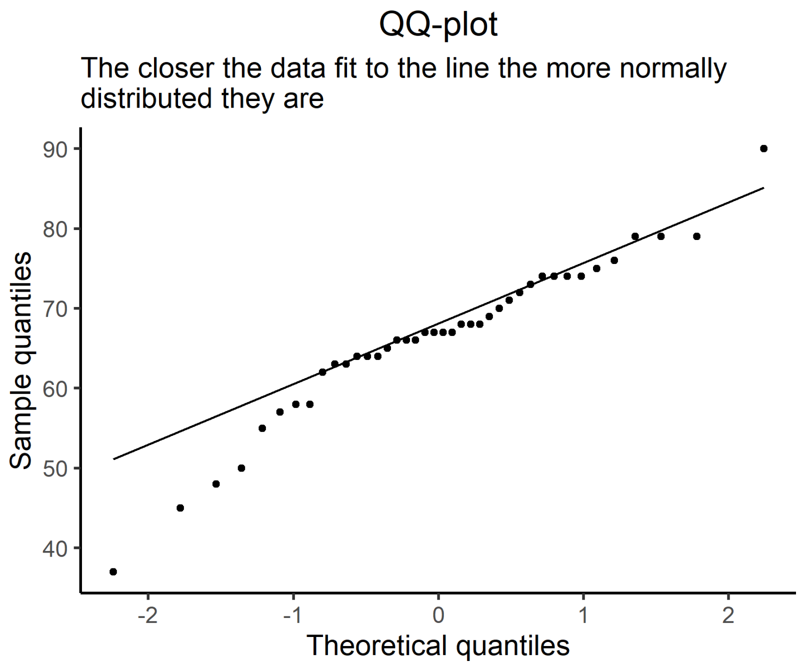
- Our density plot looks relatively normal.

Assumption checks: Normality

- QQ-plots (Quantile-Quantile plot):
    - Plots the sorted quantiles of one data set (distribution) against sorted quantiles of data set (distribution).
    - Quantile = the percent of points falling below a given value.
    - For a normality check, we can compare our own data to data drawn from a normal distribution

QQ-plots

```r
ggplot(dat, aes(sample = Age)) +
  geom_qq() +
  geom_qq_line() +
      labs(title="QQ-plot",
       subtitle="The closer the data fit to the
       x = "Theoretical quantiles",
       y = "Sample quantiles")
```



- This looks a little concerning.
- We have some deviation in the lower left corner.
- This is showing we have more lower values for age than would be expected.

Assumption checks: Normality

- Shapiro-Wilks test:
  - Checks properties of the observed data against properties we would expected from normally distributed data.
  - Statistical test of normality.
  - H_0: data = the sample came from a population that is normally distributed.
  - p-value $< \alpha$ = reject the null, data are not normal.
    - Sensitive to n as all p-values will be.
    - In very large n, normality should also be checked with QQ-plots alongside statistical test.

Shapiro-Wilks in R

```
shapiro.test(dat$Age)
```

```
##
##     Shapiro-Wilk normality test
##
## data:  dat$Age
## W = 0.95, p-value = 0.08
```

- Fail to reject the null, $p = .08$, which is > .05
- Taken collectively, it looks like our assumption of normality is met.

Effect Size: Cohen's D

- Cohen's-D is the standardized difference in means.
  - Having a standardized metric is useful for comparisons across studies.
  - It is also useful for thinking about power calculations
- The basic form of D is the same across the different t-tests:

$$D = \frac{Differece}{Variation}$$

Interpreting Cohen's D

- Below are some rough guidelines on how to interpret the size of the effect.
- These are not exact labels, but a loose guidance based on empirical research.
- Perhaps the most common "cut-offs" for D-scores:

Verbal label	Magnitude of D in absolute value
Small (or weak)	≤ 0.20
Medium (or moderate)	≈ 0.50
Large (or strong)	≥ 0.80

Cohen's D: One-sample t-test

- One-sample t-test:

$$D = \frac{\bar{x} - \mu_0}{s}$$

- μ_0 = hypothesised mean
- $\bar{x}$ = sample mean
- s = sample standard deviation

Cohen's D in R

```r
library(effectsize)
cohens_d(dat$Age, mu=65, alternative="two.sided")
```

```
## Cohen's d |         95% CI
## -------------------------
## 0.13      | [-0.18, 0.44]
##
## - Deviation from a difference of 65.
```

Write up: Assumptions

The DV of our study, Age, was measured on a continuous scale, and data were independent (based on study design). The assumption of normality was visually assessed (via histograms, density plots, and a QQplot) as well as statistically via a Shapiro-Wilks test. Whilst the QQplot did show some deviation from the diagonal line, the Shapiro-Wilks test suggested that the sample came from a population that was normally distributed $(W = 0.95, p = .08)$. This was inline with the histogram and density plot, which suggested that Age was normally distributed (and where skew < 1). The size of the effect was found to be small D = 0.13 [-0.18, 0.44].

Summary

- Today we have covered:
  - Basic structure of the one-sample t-test
  - Calculations
  - Interpretation
  - Assumption checks
  - Effect size measures (Cohen's D)