One-Sample t-test

Data Analysis for Psychology in R 1

dapR1 Team

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

- Recording 1: Introduce the three types of t-test:

- Recording 2: One-sample t-test example

- Recording 3: Inferential tests for the one-sample t-test

- Recording 4: Assumptions and effect size.

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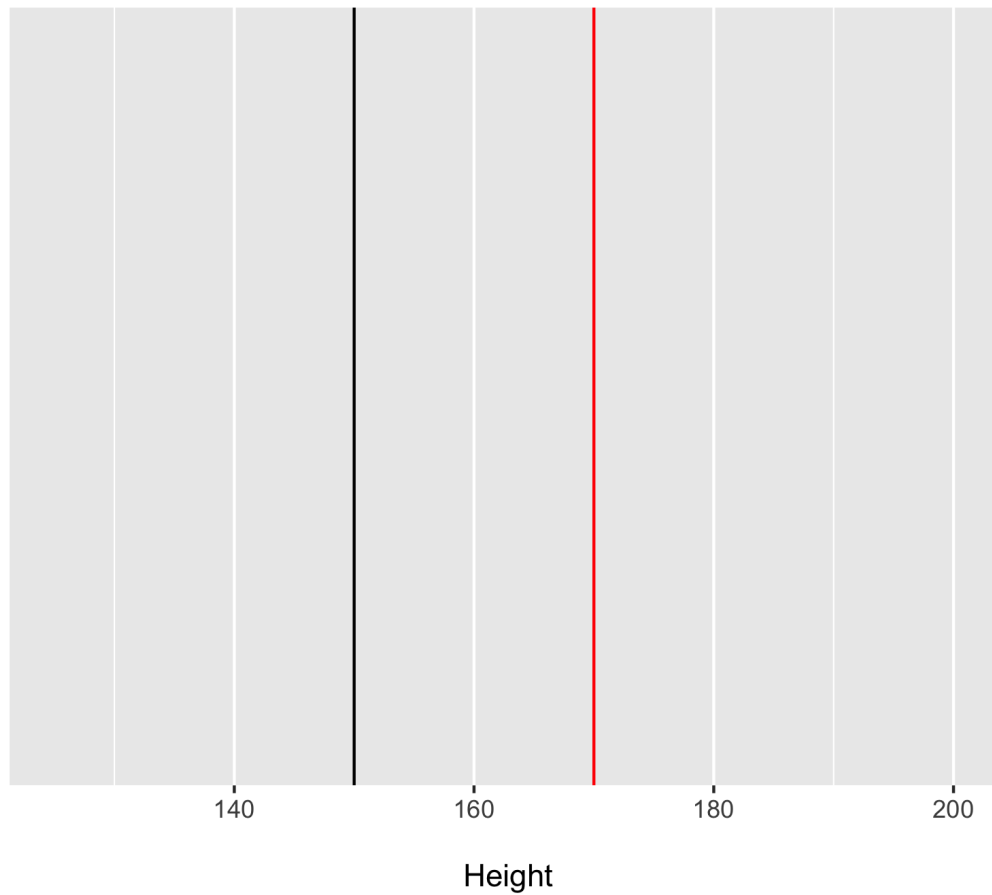
- Write down whether you think these means (two lines) are different. Write either:
  - Yes
 - No
  - It depends

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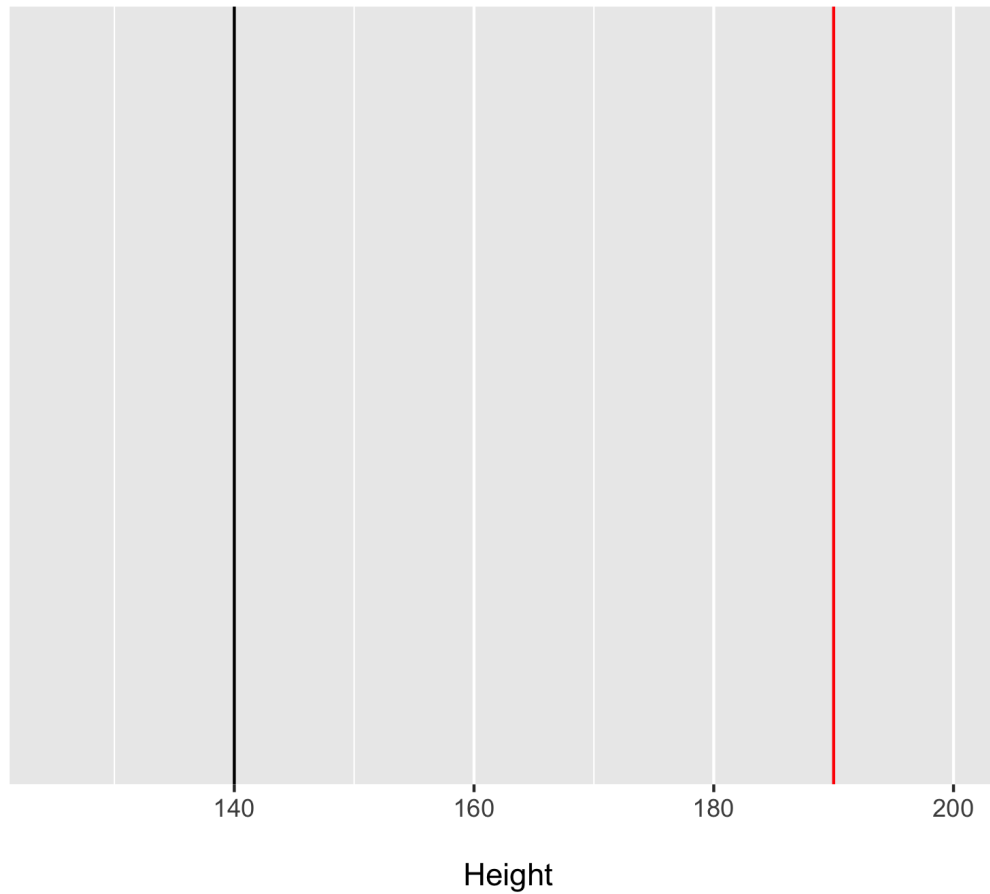
- Write down whether you think these means (two lines) are different. Write either:
  - Yes
 - No
  - It depends

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

Comparing means

All the information

Comparing means
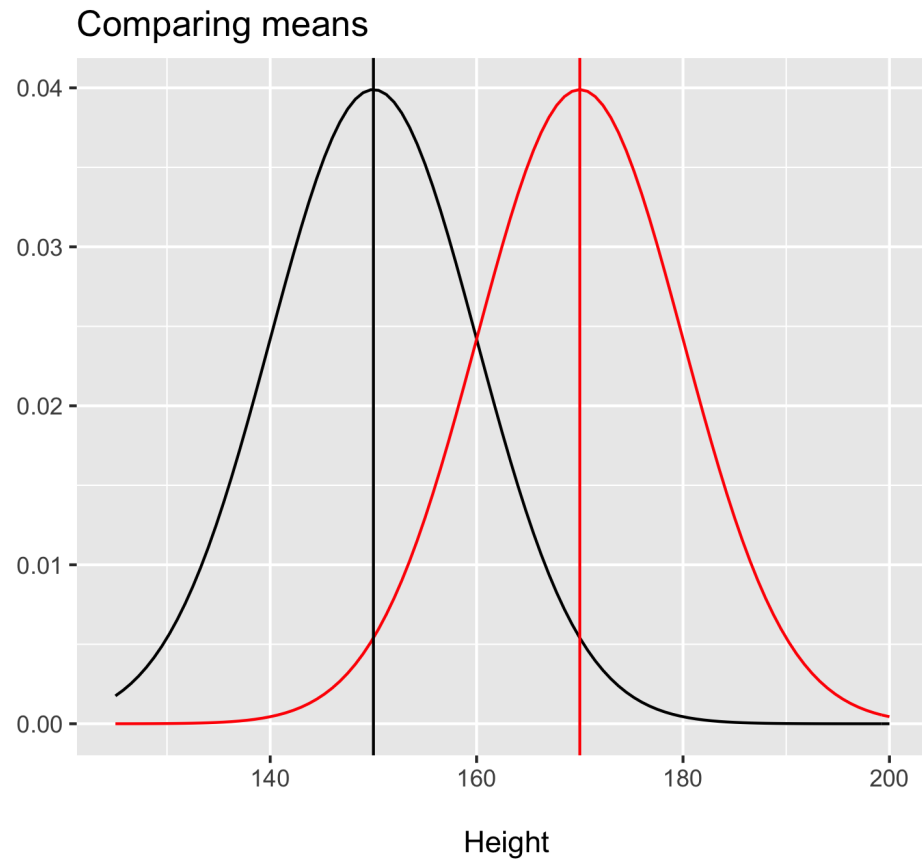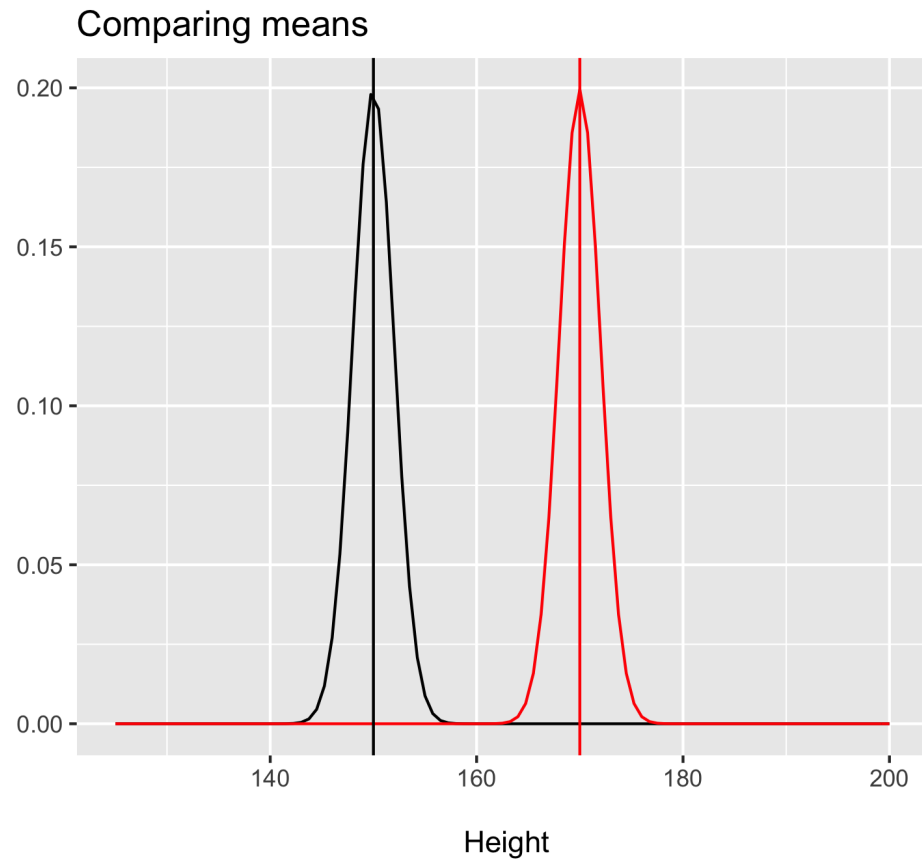
t-statistic

- 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

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}}$$

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

Time for a break

Welcome Back!

Now we have introduced the general principle of t-tests, we will consider the one-sample test in more detail

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 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

- We are comparing a single sample mean $\bar{x}$ to a known mean μ

$$H_0 : \mu = \bar{x}$$

- Note this is identical to saying:

$$H_0 : \mu - \bar{x} = 0$$

Alternative Hypotheses

- Two-tailed:

$$H_0 : \mu = \bar{x}$$
$$H_1 : \mu \neq \bar{x}$$

- One-tailed:

$$H_0 : \mu = \bar{x}$$
$$H_1 : \mu < \bar{x}$$
$$H_1 : \mu > \bar{x}$$

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 $\alpha = 0.05$.

- So I am simply asking, does my mean differ from the known mean.

Calculation

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}}$$

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

  - Use all this to calculate t.

Calculation

```
dat %>%
  summarise(
    PopMean = 65,
    Mean = mean(Age),
    SD = sd(Age),
    N = n()
  ) %>%
  mutate(
    SE = SD/sqrt(N)
  )
```

```
## # A tibble: 1 × 5
##   PopMean  Mean    SD     N    SE
##     <dbl> <dbl> <dbl> <int> <dbl>
## 1      65  66.3  10.0    40  1.58
```

Calculation

```
## # A tibble: 1 × 5
##   PopMean  Mean    SD     N    SE
##     <dbl> <dbl> <dbl> <int> <dbl>
## 1      65  66.3  10.0    40  1.58
```

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}} = \frac{66.3 - 65}{\frac{10.01}{\sqrt{40}}} = \frac{1.3}{1.583} = 0.821$$

- So in our example $t = 0.821$

Time for a break

Welcome Back!

**Now we have calculated our test statistic, it is time to conduct an inferential test**

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 (0.821)
    - Hypothesis to test (two-tailed)
    - α level (0.05).

- So now all we need is the critical value from the associated t-distribution in order to make our decision.

Is our test significant?

t-distribution (df=39); t-statistic (0.821; red line)
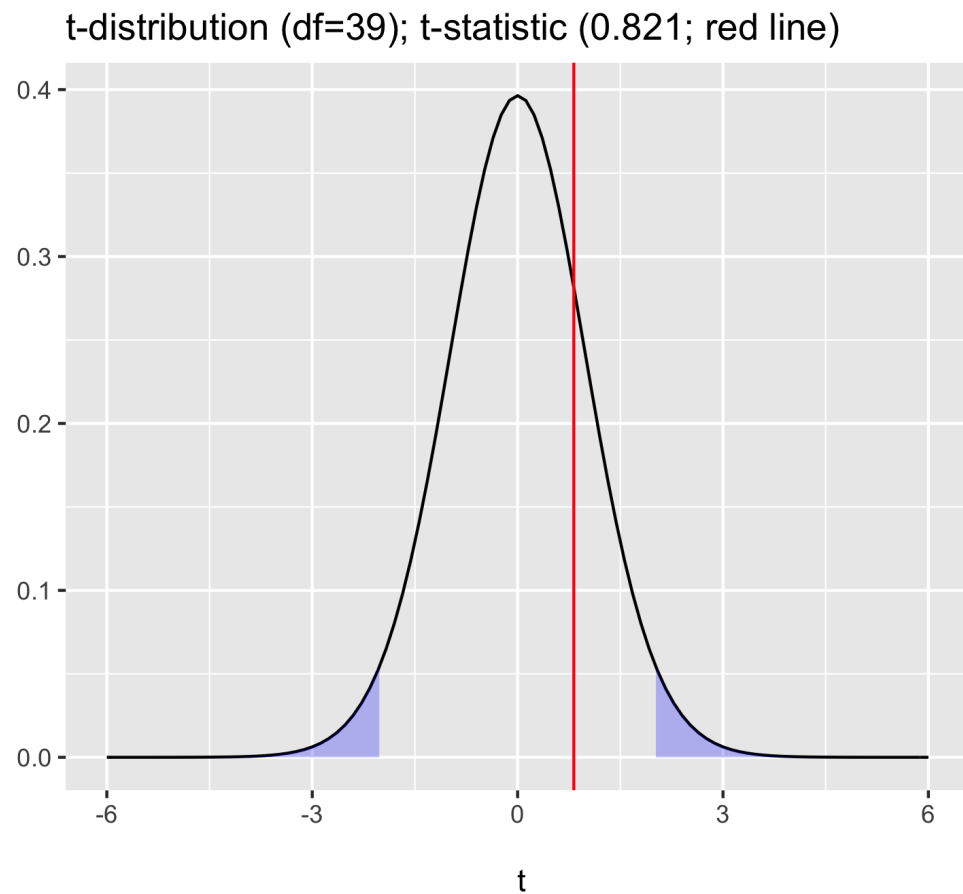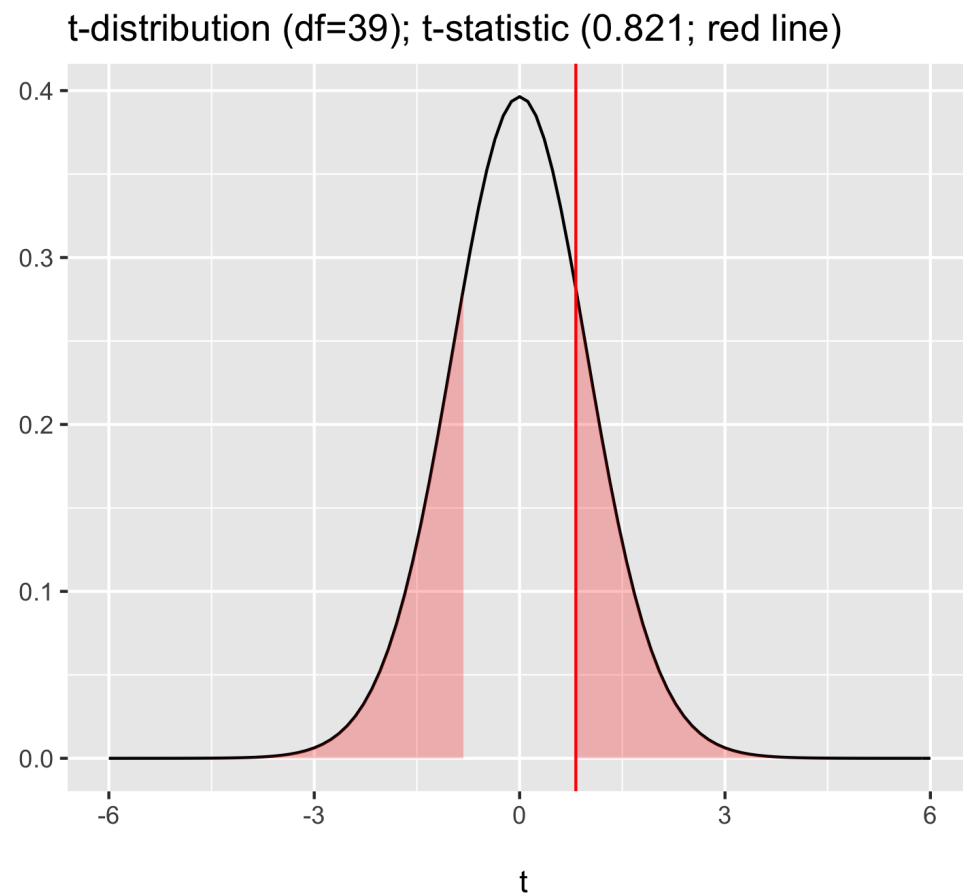


```
## # A tibble: 1 × 2
##   LowerCrit UpperCrit
##       <dbl>     <dbl>
## 1     -2.02      2.02
```

Is our test significant?

- So our critical value is 2.02

  - Our t-statistic (0.821) is closer to 0 than this.
  - So we fail to reject the null hypothesis.

- t(39)=0.821, p > .05, two-tailed.

Exact p-values

t-distribution (df=39); t-statistic (0.821; red line)


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

In R

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

```
##
##      One Sample t-test
##
## data:  dat$Age
## t = 0.82152, df = 39, p-value = 0.4163
## alternative hypothesis: true mean is not equal to 65
## 95 percent confidence interval:
##  63.09922 69.50078
## sample estimates:
## mean of x
##      66.3
```

Write up

A one-sample t-test was conducted in order to determine if a statistically significant (α =.05) mean difference existed between the average retirement age of Professors, and the age at retirement of a sample of 5 psychology Professors. The sample scored higher (Mean=66.3, SD=10.01) than the population (Mean = 65), however the difference was not statistically significant (t(39)=0.821, p > .05, two-tailed).

Time for a break

Welcome Back!

Every inferential set comes with a set of assumptions. These need to be checked in order to make sure results are valid. So let's look at t-test assumptions, as well as calculating effect size measures

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:

    - DV is continuous.

- But we also have some additional model assumptions for the test to be valid.

    1. The data are normally distributed.
    2. The data are an independent random sample.

- (2) we can not directly test.

- (1) we can look at descriptive statistics, QQplots, histograms and a Shapiro-Wilks Test.

Assumption checks: Normality

- Descriptive statistics:
  - Skew: No strict cuts for skew.
    - Skew < |1| generally not problematic
    - |1| < skew > |2| slight concern
    - Skew > |2| investigate impact

Skew

```r
library(moments)
dat %>%
  summarise(
    skew = round(skewness(Age),2)
  )
```
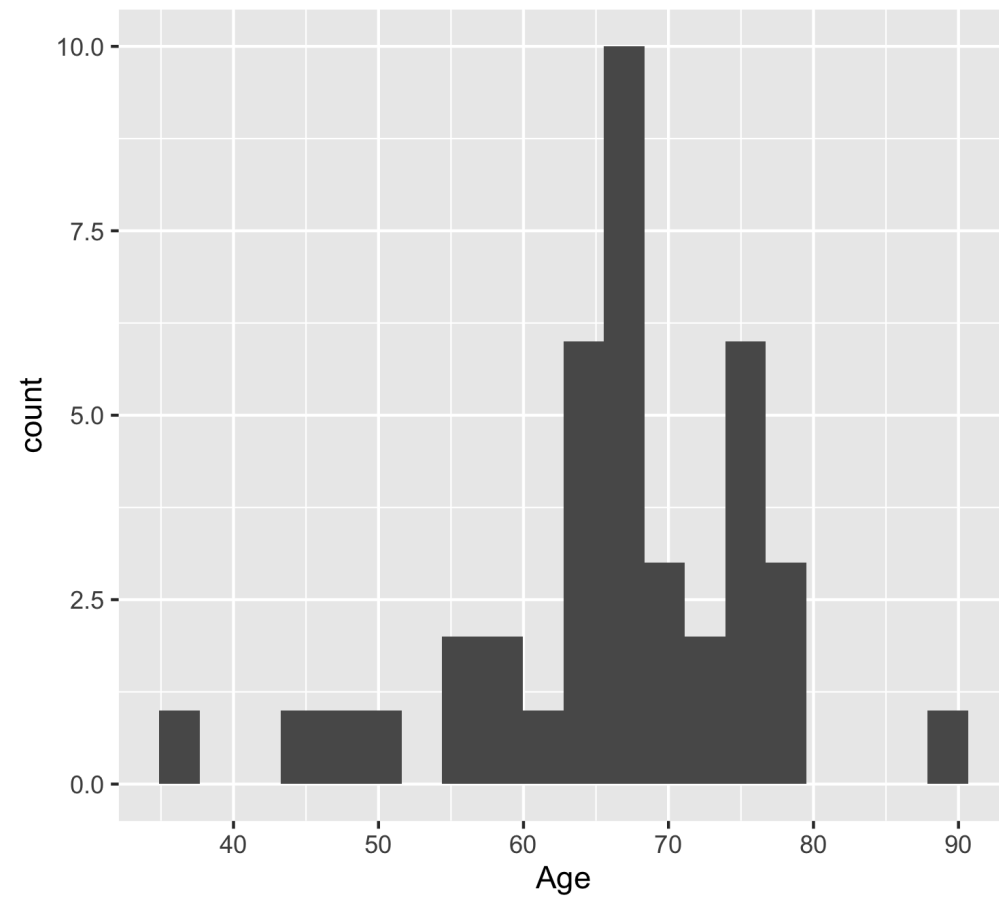
```
## # A tibble: 1 × 1
##     skew
##    <dbl>
## 1 -0.66
```

- Skew is low.

Histograms

```
dat %>%
  ggplot(., aes(x=Age)) +
  geom_histogram(bins = 20)
```

- Our histogram looks "lumpy", but we have relatively low N for looking at these plots.
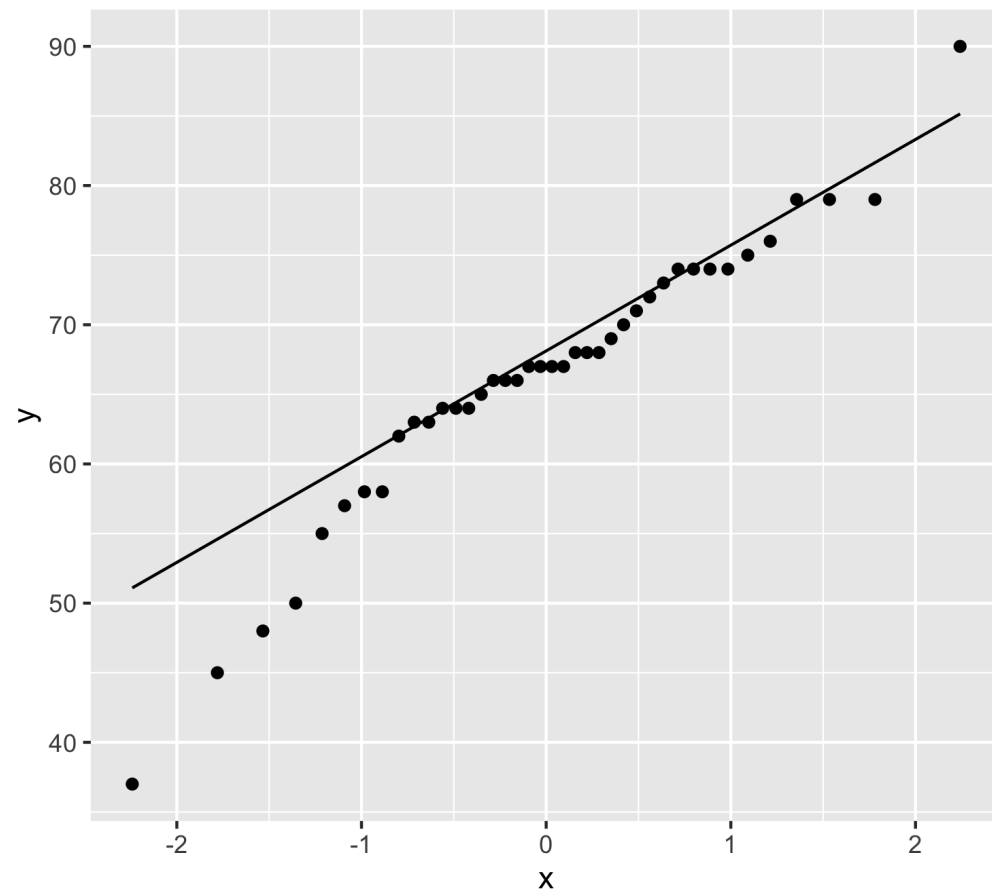
Assumption checks: Normality

- QQ-plots:
  - 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

```
dat %>%
  ggplot(., aes(sample = Age)) +
  stat_qq() +
  stat_qq_line()
```

- 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 = a normal distribution.
  - 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 R

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

```
##
##      Shapiro-Wilk normality test
##
## data:  dat$Age
## W = 0.95122, p-value = 0.08354
```

- Fail to reject the null, $p > .05$

- 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 (more in a couple of weeks)

- The basic form of D is the same across the different t-tests:

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

Interpreting Cohen's D

- There are a number of guides for interpreting Cohen's D.

    - These are not set in stone, and are intended as heuristics.

- Perhaps the most common "cut-offs" for D-scores:

    - ~ 0.2 = small effect
    - ~ 0.5 = moderate effect
    - ~ 0.8 = large effect

Cohen's D: One-sample t

- One-sample t-test:

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

- μ = population mean

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

Cohen's D in R

```
library(effsize)
cohen.d(dat$Age, NA, mu=65, conf.level = .95)
```

```
##
## Cohen's d (single sample)
##
## d estimate: 0.1298935 (negligible)
## Reference mu: 65
## 95 percent confidence interval:
##      lower     upper
## -0.5104117  0.7701986
```

Summary

- Today we have covered:
    - Basic structure of the one-sample t-test
    - Calculations
    - Interpretation
    - Assumption checks
    - Effect size measures