One-sample hypothesis tests

Today...

- One-sample t-tests

What are the steps of NHST?

1. Define null and alternative hypothesis.
2. Set and justify alpha level.
3. Determine which sampling distribution (z, t, or χ^2)
4. Calculate parameters of your sampling distribution under the null.
   - If z, calculate μ and σ_M
5. Calculate test statistic under the null.
   - If z, $\frac{\bar{X}-\mu}{\sigma_M}$
6. Calculate probability of that test statistic or more extreme under the null, and compare to alpha.

One-sample tests compare your given sample with a "known" population.

- Research question: does this sample come from this population?

Hypotheses

H_0: Yes, this sample comes from this population.

H_1: No, this sample comes from a different population.

Hypotheses

H_0: Yes, the means are equal

H_1: No, the means are not equal

Example

The sample data were obtained from Census at School, a website developed by the American Statistical Association to help students in the 4th through 12th grades understand statistical problem-solving.

- The site sponsors a survey that students can complete and a database that students and instructors can use to illustrate principles in quantitative methods.
- The database includes students from all 50 states, from grade levels 4 through 12, both boys and girls, who have completed the survey dating back to 2010.

Comparing Means

Comparing means is usually done with the *t*-test, of which there are several forms.

The one-sample *t*-test is appropriate when a **single sample mean** is compared to a **population mean** but the population standard deviation is unknown. A sample estimate of the population standard deviation is used instead. The appropriate sampling distribution is the *t*-distribution, with $N - 1$ degrees of freedom.

$$t_{df=N-1} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{N}}}$$

The heavier tails of the t-distribution, especially for small *N*, are the penalty we pay for having to estimate the population standard deviation from the sample.

One-sample *t*-tests

t-tests were developed by William Sealy Gosset, who was a chemist studying the grains used in making beer. (He worked for Guinness.)

- Specifically, he wanted to know whether particular strains of grain made better or worse beer than the standard.
- He developed the *t*-test, to test small samples of beer against a population with an unknown standard deviation.
    - Probably had input from Karl Pearson and Ronald Fisher
- Published this as "Student" because Guinness wouldn't let him share company secrets.

One-sample *t*-tests

	Z-test	***t*-test**
μ	known	known
σ	known	unknown
sem or σ_M	$\frac{\sigma}{\sqrt{N}}$	$\frac{\hat{\sigma}}{\sqrt{N}}$
Probability distribution	standard normal	t
DF	none	$N-1$
Tails	One or two	One or two
Critical value ($\alpha = .05, two - tailed$	1.96	Depends on DF

Always Assuming...

Normality. We assume the population distribution is normal. If we don't meet this assumption --> non-parametric tests. (Wilcoxon signed-rank test)

Independence. Observations in the dataset are not associated with one another. Put another way, collecting a score from Participant A doesn't tell me anything about what Participant B will say. They are generated independently.

A brief example

Using the same Census at School data, we find that Missouri students who participated in a memory game completed the game in an average time of 45.93 seconds. We know that the average US student completed the game in 45.05 seconds. How do our students compare?

We're using data from Missouri in 2021 only

```
library(tidyverse)
school %>%
  summarize(meanMemory = mean(Score_in_memory_game),
            sdMemory = sd(Score_in_memory_game))
```

```
##   meanMemory sdMemory
## 1   45.93098 13.05054
```

```
nrow(school)
```

```
## [1] 51
```

A brief example

Using the same Census at School data, we find that Missouri students who participated in a memory game completed the game in an average time of 45.93 seconds. We know that the average US student completed the game in 45.05 seconds. How do our students compare?

We're using data from Missouri in 2021 only

Hypotheses

$$H_0 : 45.05(\mu) = 45.93(\bar{X})$$

$$H_1 : 45.05(\mu) \neq 45.93(\bar{X})$$

$$\mu = 45.05$$

$$N = 51$$

$$\bar{X} = 45.93$$

$$s = 13.05$$

```
t.test(x = school$Score_in_memory_game, mu = 45.05,
       alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  school$Score_in_memory_game
## t = 0.48208, df = 50, p-value = 0.6318
## alternative hypothesis: true mean is not equal to 45.05
## 95 percent confidence interval:
##  42.26046 49.60150
## sample estimates:
## mean of x
##  45.93098
```

```r
lsr::oneSampleTTest(x = school$Score_in_memory_game,
                    mu = 45.05, one.sided = FALSE)
```

```
##
##    One sample t-test
##
## Data variable:   school$Score_in_memory_game
##
## Descriptive statistics:
##             Score_in_memory_game
##    mean                   45.931
##    std dev.               13.051
##
## Hypotheses:
##    null:        population mean equals 45.05
##    alternative: population mean not equal to 45.05
##
## Test results:
##    t-statistic:  0.482
##    degrees of freedom:  50
##    p-value:  0.632
##
## Other information:
##    two-sided 95% confidence interval:  [42.26, 49.602]
##    estimated effect size (Cohen's d):  0.068
```

Cohen's *D*

Cohen suggested one of the most common effect size estimates—the standardized mean difference—useful when comparing a group mean to a population mean or two group means to each other.

$$\delta = \frac{\mu_1 - \mu_0}{\sigma} \approx d = \frac{\bar{X} - \mu}{\hat{\sigma}}$$

Cohen's *d* is in the standard deviation (*Z*) metric. Why does this matter?

Cohen's *D*

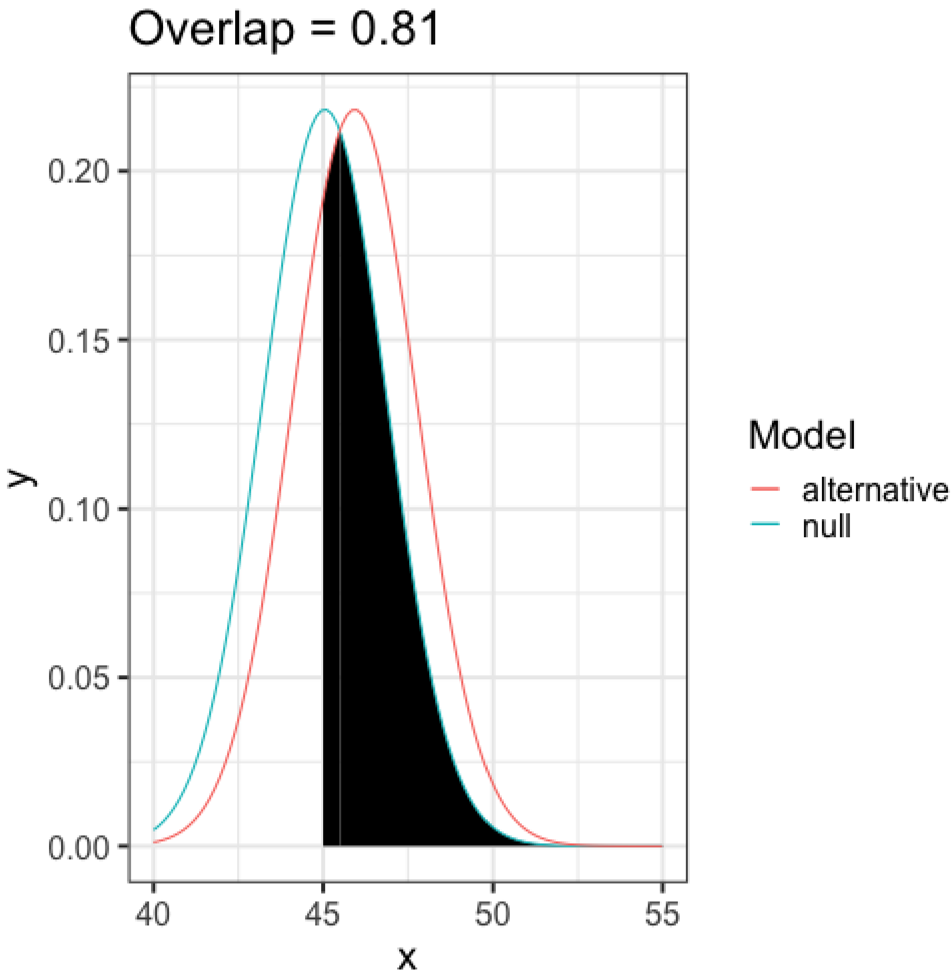
Cohens’s *d* for these data is .068. In other words, the sample mean differs from the population mean by .068 standard deviation units.

- .2 = Small
- .5 = Medium
- .8 = Large

Cohen, J. (1988), Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale: Lawrence Erlbaum.

Not bounded by 1 the way a correlation is bounded

Another useful metric is the overlap between the two distributions -- the smaller the overlap, the farther apart the distributions



You Try!

Download the data: sample data

```
library(tidyverse)
school_pollution = read.csv(here("data/cenus_at_school_missouri.csv"))
school_pollution = school %>%
  filter(!is.na(Importance_reducing_pollution))
```

You'll need to change the working directory to wherever you downloaded the file

You Try!

Students were asked the following: *"How important is reducing pollution to you?"*. They were given options of 0 (not important) to 1000 (very important).

The average rating across the nation is 762.52.

You Try!

Using the same data:

Step 1: Calculate the t statistic "by hand".

Step 2: Calculate the p-value and 95% CI. Interpret! Use R to get the p-value, but calculate 95% CI "by hand".

Step 3: Calculate Cohen's *D* "by hand". Interpret!

Let's do it together

Step 1: Get the test statistic

```
school_pollution %>%
  summarize(meanPollution = round(mean(Importance_reducing_pollution),
                                  digits = 3),
            sdPollution = round(sd(Importance_reducing_pollution),
                                digits = 3))
```

```
##   meanPollution sdPollution
## 1       718.673     297.403
```

```
nrow(school_pollution)
```

```
## [1] 49
```

$$\frac{718.673 - 762.52}{\frac{297.403}{\sqrt{49}}}$$

Let's do it together

Step 1: Get the test statistic

$$\frac{718.673 - 762.52}{\frac{297.403}{\sqrt{49}}}$$

```
(718.673 - 762.52) / (297.403/sqrt(49))
```

```
## [1] -1.032031
```

Let's do it together

Step 2a: Get the p-value. Interpret

```r
t = (718.673 - 762.52) / (297.403/sqrt(49))
N = nrow(school_pollution)
2 * pt(q = t, df = N-1, lower.tail = TRUE)
```

```
## [1] 0.3072299
```

The probability of obtaining a test statistic of -1.032 or more extreme, given the null hypothesis is true is .307. This is above our standard threshold of .05. We therefore cannot reject the null hypothesis. We retain the null hypothesis that states that our sample mean and population mean are the same.

Step 2b: Get the 95% CIs. Interpret

$$\bar{X} \pm [t_{.975,df} \frac{\hat{\sigma}}{\sqrt{N}}]$$

```
# we have N stored from previous slide
xbar = 718.673
sem = 297.403/sqrt(49)

ci_lb_t = xbar - sem * qt(p = .975, df = N-1)
ci_ub_t = xbar + sem * qt(p = .975, df = N-1)

ci_lb_t
```

```
## [1] 633.2489
```

```
ci_ub_t
```

```
## [1] 804.0971
```

If we carried out random sampling a large number of times, and each time we calculated the 95% confidence interval, we would expect 95% of those intervals to contain the true population mean, which is likely between 633 and 804

Step 3: Get Cohen's *d*. Interpret

$$d = \frac{\bar{X} - \mu}{\hat{\sigma}}$$

```
# we have xbar from before
stdev = 297.403
mu = 762.52

d = abs((xbar - mu) / stdev)
d
```

```
## [1] 0.1474329
```

$$t(48) = -1.032$$
$$CI_{95}[633.249, 804.097]$$
$$\text{Cohen's d} = 0.147$$

The standardized difference in means is quite small, with a Cohen's *d* equal to 0.147. According to Cohen (1988), this is considered a small effect size.

```r
lsr::oneSampleTTest(x = school_pollution$Importance_reducing_pollution,
                    mu = 762.52, one.sided = FALSE)
```

```
##
##    One sample t-test
##
## Data variable:   school_pollution$Importance_reducing_pollution
##
## Descriptive statistics:
##             Importance_reducing_pollution
##    mean                           718.673
##    std dev.                       297.403
##
## Hypotheses:
##    null:        population mean equals 762.52
##    alternative: population mean not equal to 762.52
##
## Test results:
##    t-statistic:  -1.032
##    degrees of freedom:  48
##    p-value:  0.307
##
## Other information:
##    two-sided 95% confidence interval:  [633.249, 804.097]
##    estimated effect size (Cohen's d):  0.147
```

The usefulness of the one-sample *t*-test

How often will you conducted a one-sample *t*-test on raw data?

- (Probably) never

How often will you come across one-sample *t*-tests?

- A lot!

The one-sample *t*-test is used to test coefficients in a model.

Let's compare to 0

```
lsr::oneSampleTTest(x = school_pollution$Importance_reducing_pollution,
                    mu = 0, one.sided = FALSE)
```

```
##
##    One sample t-test
##
## Data variable:   school_pollution$Importance_reducing_pollution
##
## Descriptive statistics:
##             Importance_reducing_pollution
##    mean                           718.673
##    std dev.                       297.403
##
## Hypotheses:
##    null:        population mean equals 0
##    alternative: population mean not equal to 0
##
## Test results:
##    t-statistic:  16.915
##    degrees of freedom:  48
##    p-value:  <.001
##
## Other information:
##    two-sided 95% confidence interval:  [633.249, 804.097]
##    estimated effect size (Cohen's d):  2.416
```

Let's compare to 0

```
model = lm(Importance_reducing_pollution ~ 1, data = school_pollution)
summary(model)
```

```
##
## Call:
## lm(formula = Importance_reducing_pollution ~ 1, data = school_pollution)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -717.7 -168.7  105.3  231.3  281.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   718.67      42.49   16.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 297.4 on 48 degrees of freedom
```

```r
model = lm(Petal.Width ~ Species, data = iris)
summary(model)
```

```
##
## Call:
## lm(formula = Petal.Width ~ Species, data = iris)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.626 -0.126 -0.026  0.154  0.474
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.24600    0.02894    8.50 1.96e-14 ***
## Speciesversicolor  1.08000    0.04093   26.39  < 2e-16 ***
## Speciesvirginica   1.78000    0.04093   43.49  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2047 on 147 degrees of freedom
## Multiple R-squared:  0.9289,    Adjusted R-squared:  0.9279
## F-statistic:   960 on 2 and 147 DF,  p-value: < 2.2e-16
```

If time...

Accessing your results in `R`

- Go through list objects
- Go through `broom`

Next time...

Comparing two means