# Computing and Interpreting Chi-Squared

## Null hypothesis significance testing

**Jenine Harris**
**Brown School**

# Import the data

```
# import the April 17-23 Pew Research Center data
library(package = "haven")

# import the voting data
vote <- read_sav(file = "/Users/harrisj/Box/teaching/Teaching/Fall2020/d
```

# Data cleaning

```
# select variables of interest and clean them
vote.cleaned <- vote %>%
  select(pew1a, pew1b, race, sex, mstatus, ownhome, employ, polparty, ed
  zap_labels() %>%
  mutate(pew1a = recode_factor(.x = pew1a,
                               `1` = 'Register to vote',
                               `2` = 'Make easy to vote',
                               `5` = NA_character_,
                               `9` = NA_character_)) %>%
  rename(ease.vote = pew1a) %>%
  mutate(pew1b = recode_factor(.x = pew1b,
                               `1` = 'Require to vote',
                               `2` = 'Choose to vote',
                               `5` = NA_character_,
                               `9` = NA_character_)) %>%
  rename(require.vote = pew1b) %>%
  mutate(race = recode_factor(.x = race,
                              `1` = 'White non-Hispanic',
                              `2` = 'Black non-Hispanic',
                              `3` = 'Hispanic',
                              `4` = 'Hispanic',
                              `5` = 'Hispanic',
                              `6` = 'Other',
                              `7` = 'Other',
                              `8` = 'Other',
                              `9` = 'Other',
                              `10` = 'Other',
```

# Interpreting the chi-squared statistic

```
# chi-squared statistic for ease of voting
# and race
chisq.test(x = vote.cleaned$ease.vote,
           y = vote.cleaned$race)
```

```
##
##     Pearson's Chi-squared test
##
## data:  vote.cleaned$ease.vote and vote.cleaned$race
## X-squared = 28.952, df = 3, p-value = 2.293e-06
```

# Using Null Hypothesis Significance Testing (NHST) to organize statistical testing

One process to organize statistical tests like chi-squared is *Null Hypothesis Significance Testing* or NHST. The steps of NHST are...

- Write the null and alternate hypotheses
- Compute the test statistic
- Calculate the probability that your test statistic is at least as big as it is if there is no relationship (i.e., the null is true)
- If the probability that the null is true is very small, usually less than 5%, reject the null hypothesis
- If the probability that the null is true is not small, usually 5% or greater, retain the null hypothesis

# NHST Step 1: Write the null and alternate hypotheses

- The null hypothesis is usually a statement that claims there is *no difference* or *no relationship* between things.
- In this case, the null hypothesis stated that opinions on voter registration are no different across groups of race-ethnicity.
- The alternate hypothesis is the claim that there is *a difference* or *a relationship* between things.
- The null (H0) and alternate (HA) hypotheses are written *about the population* and are tested *using a sample from the population*.
- Here are the null and alternate for the voting data:
    - H0: People's opinions on voter registration are the same across race-ethnicity groups.
    - HA: People's opinions on voter registration are *NOT* the same across race-ethnicity groups.

# NHST Step 2: Compute the test statistic

- The test statistic to use when examining a relationship between two categorical variables is the chi-squared statistic, $\chi^2$ .

```
# chi-squared statistic for ease of voting
# and race
chisq.test(x = vote.cleaned$ease.vote,
           y = vote.cleaned$race)
```

```
##
##     Pearson's Chi-squared test
##
## data:  vote.cleaned$ease.vote and vote.cleaned$race
## X-squared = 28.952, df = 3, p-value = 2.293e-06
```

- The test statistic is: $\chi^2 = 28.95$.

# NHST Step 3: Calculate the probability that your test statistic is at least as big as it is if there is no relationship (i.e., the null is true)

- The probability of seeing a chi-squared as big as 28.95 in our sample *if there were no relationship in the population* between opinion on voting ease and race-ethnicity group would be 0.00000229255 or $p < .05$.

# NHST Step 4: If the probability that the null is true is very small reject the null hypothesis

- The probability that the null hypothesis, *People's opinions on voter registration are the same across race-ethnicity groups*, is true in the population based on what we see in the sample is 0.00000229255 or $p < .05$.
- This is a very small probability of being true and indicates that the null hypothesis is not likely to be true and should therefore be *rejected*.

**If the p is low, the null must go.**

# NHST Step 5: If the probability that the null is true is not small, usually 5% or greater, retain the null hypothesis

Step 5 does not apply in this situation.

# Report the results

We used the chi-squared test to test the null hypothesis that there was no relationship between opinions on voter registration by race-ethnicity group. We rejected the null hypothesis and concluded that there was a statistically significant association between views on voting ease and race-ethnicity [ $\chi^2$ (3) = 28.95; p < .05].