# Bush 631-607: Quantitative Methods

## Lecture 11 (11.16.2021): Uncertainty vol. II

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

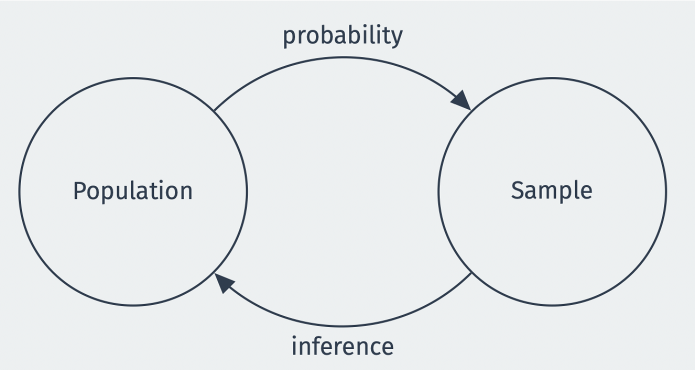
Fall 2021

# What is today's plan?

- Calculating uncertainty: beyond measures of spread.
- Hypotheses tests
- When our results are 'significant'?
- Errors in testing.
- Power analysis.
- R work: pnorm(),t.test(),power.prop.test(),power.t.test()

# Our data - our research interests

- Making inferences from data to population



# Uncertainty

- Our proposed results:
  - Regime type matters for the extent of aid provision.
  - Financial incentives increase support for anti-plastic campaign.
- Are these effects real or just noise?

# Estimation

- *Quantity of interest* in population.
- *Point estimation* $\rightarrow$ a 'best guess'.

$$\hat{\theta} = \bar{X}_n \rightarrow \text{population p}$$

# Estimation and uncertainty

- Is our estimate good?
- **Estimation error**: difference with 'true value'.
- Repetition $\rightarrow$ sampling distribution of $\hat{\theta}$
- Estimation error / bias using *expectations*
- $bias = E(est.Error) = E(Estimate - truth) = E(\bar{X}_n) - p = p - p = 0$
- **Unbiasedness**: Sample proportion is on average equal to the population proportion.

# Estimators in experiments

- *Estimator* $\rightarrow$ diff-in-means.
- Sample Average Treatment Effect (SATE):

$$SATE = \frac{1}{n} * \sum_{i=1}^{n} [Y_i(1) - Y_i(0)]$$

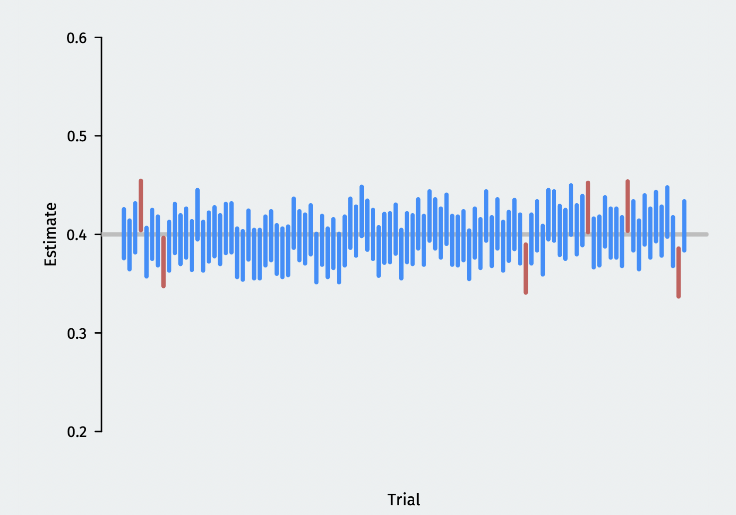
- Random assignment: estimator is unbiased.

# Uncertainty tools

- More ways to quantify uncertainty:
  - SD: variation of estimator (‘spread’ of distribution).
  - Only relevant for simulation.
  - Use SE for single sample.
- Confidence intervals:
  - Range of estimator true values.
  - The interpretation of CIs.

# Simulate CIs

- How many overlap with 'true' support?



# Testing our findings

- Describing uncertainty of results:
  - SD/SE.
  - Confidence intervals.
- Direct testing of results:
  - Hypothesis testing.
  - Significant effects.

# The tea tasting experiment

- The scenarios:

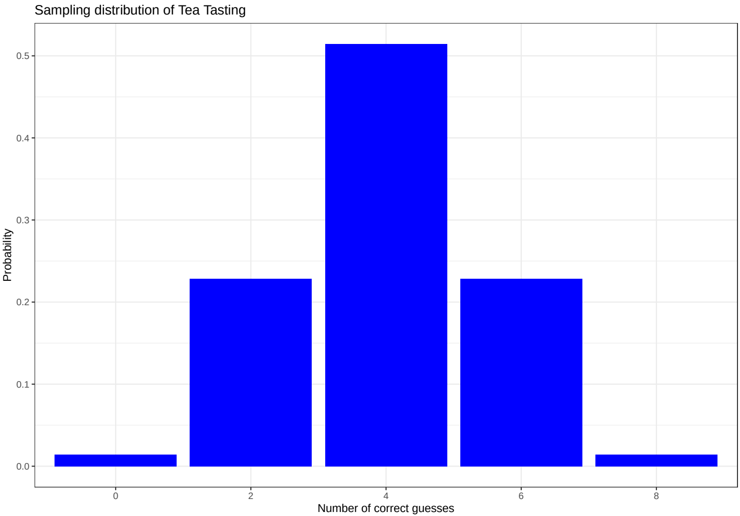
| *Cups* | *Lady's guess* | *Actual order* | *Scenarios* | | | ... |
|---|---|---|---|---|---|---|
| 1 | M | M | T | T | T | |
| 2 | T | T | T | T | M | |
| 3 | T | T | T | T | M | |
| 4 | M | M | T | M | M | |
| 5 | M | M | M | M | T | |
| 6 | T | T | M | M | T | |
| 7 | T | T | M | T | M | |
| 8 | M | M | M | M | T | |
| Number of correct guesses | | 8 | 4 | 6 | 2 | ... |

# Guessing?

- Only one way to choose **all 4 cups** correctly.
- But 70 ways to choose 4 among 8: ${}_8P_4 = \frac{8!}{4!*(8-4)!}$
- Assume random choice: equal probability to each combination.
- Chances for guessing all 4 correctly: $\frac{1}{70} = 0.014 = 1.4\%$.

*Does not seem like a guess...*

# Probabilities and guesses



# Simulating tea tasting

- What are the odds of n correct cups?

```r
# Simulations
sims <- 1000
guess <- c("M","T","T","M","M","T","T","M")
correct <- rep(NA, sims)

for (i in 1:sims){
  cups <- sample(c(rep("T",4), rep("M",4)), replace = FALSE)
  correct[i] <- sum(guess == cups)
}

head(correct)
```

```
## [1] 4 4 4 4 4 2
```

```r
prop.table(table(correct))
```

```
## correct
##     0     2     4     6     8
## 0.015 0.238 0.523 0.211 0.013
```

# Statistical hypothesis testing

- **A thought experiment** - meaningful or random chance?
- Example:
  - You work in a research firm, study social equity.
  - Senior analyst claims 20% of Houston households are poor.
  - Data collection, sample of 950 households in Houston.
  - Mean in sample: 23% under poverty line.

*Was the senior analyst wrong?*
*How certain are we that the sample is correct?*

# Statistical hypothesis testing

- Probabilistic *proof by contradiction*
- Assume the contrast to our expectations is not possible.

# Proof by contradiction

- Houston research firm example
- Assume $\rightarrow$ difference (sample and analyst) are zero.
- Incorrect? $\rightarrow$ differences exist.
- Senior analyst may have been wrong.
- We can never **fully** reject a hypothesis (no 100% certainty).

# Procedure for hypothesis tests

- (1) Generate a **null hypothesis** ($H_0$).
- A statement we want to refute:
  - A devil's advocate position.
  - Red teams in strategic thinking/wargaming.
- (2) Define **alternative hypothesis** $H_1/H_a$.
- Statement of what we actually hope is true.
- Alternative hypothesis is *opposite* of null.
- (3) Show that differences are not due to chance.

# Illustrating hypothesis testing

- 2016 elections polls

**Trump consistently outperformed polls in key states**

Trump Clinton

| | POLL AVERAGE | POPULAR VOTE COUNT | DIFFERENCE |
|---|---|---|---|
| National polls | +3 | Even | +3 |
| Missouri | +9 | +19 | +10 |
| Wisconsin | +6 | +1 | +7 |
| Ohio | +2 | +9 | +7 |
| Iowa | +3 | +10 | +7 |
| Maine | +9 | +3 | +6 |
| Minnesota | +6 | +1 | +5 |
| Michigan | +3 | +1 | +4 |
| North Carolina | Even | +4 | +4 |
| Pennsylvania | +2 | +1 | +3 |
| New Hampshire | +2 | Even | +2 |
| Colorado | +4 | +2 | +2 |
| Georgia | +4 | +6 | +2 |
| Florida | +1 | +1 | +2 |
| Arizona | +3 | +4 | +1 |
| Virginia | +5 | +5 | No difference |

# Illustrating hypothesis testing

- Now for 2020 Elections
- Finals polls: Trump support at 44%.
- Actual results: Trump support at 47.5%.
- Is 3% a real difference?

- Our Q: true population support for Trump (p).
- Null hypothesis: $H_0 : p = 0.475$
- Alternative hypothesis: $H_1 : p \neq 0.475$
- Gather data: sample of about 1300 people.
- Mean proportion: $\bar{X} = 0.44$

# Testing our null hypothesis

- If null is true, what the data distribution?
- $X_i ... X_n$ is Bernoulli with $p = 0.475$
- $X_i ... X_n = 1$ if support Trump, 0 otherwise.
- Check with simulated data: share of votes in each sample.

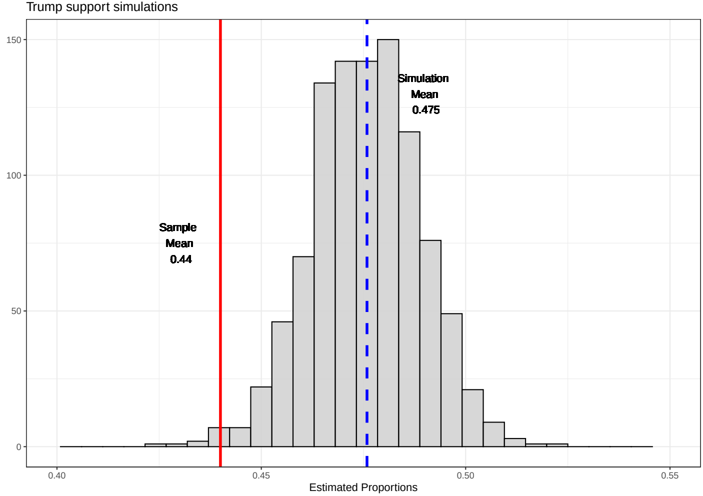
```r
### Trump support simulation
t_vote <- rbinom(n = 1000, size = 1363, prob = 0.475)
t_share <- t_vote / 1363

head(t_share)
```

```
## [1] 0.4893617 0.4651504 0.4864270 0.4739545 0.4856933 0.4820249
```

# Distribution of estimator

- Sample mean vs. simulated proprortion of Trump support



# Testing our null hypothesis

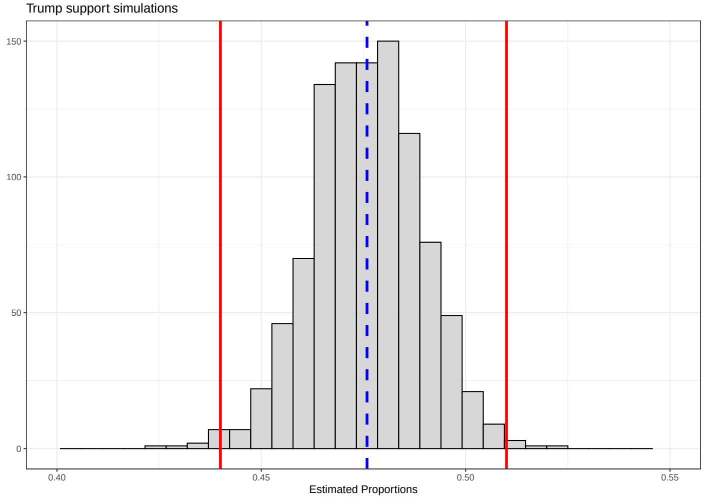
- Define *p-value*.
- Probability that observed data is as extreme as null.
- If Trump null is true...
- How often we get survey with big polling error?
  - Smaller *p-values* $\rightsquigarrow$ stronger evidence against null.
  - **Not** probability that null is true!
- p-values are **two-sided**:
  - Error is 3.5% ($0.44 - 0.475 = -0.035$).
  - Check **both sides** of true mean ($p = 0.475$).

```
# Error (0.035) in both sides of mean (0.475)
mean(t_share < 0.44) + mean(t_share > 0.51)
```

```
## [1] 0.01
```

# Polling error odds

- 1% chance of polling error $\rightarrow$ not too likely



# Hypothesis testing expanded

- **Directional** hypothesis
- Null is incorrect from 'one direction' only.
- Define hypotheses:
  - Null: $H_0 : p = 0.475$
  - **one-sided alternative**: $H_1 : p < 0.475$
  - $H_1 \rightarrow$ polls underestimate Trump support
- Define p-value as one-sided:
  - Probability that random sample underestimate Trump support as we see in our sample?
  - Smaller than two-sided p-value (only ;eft side of distribution!)

```
# Error (0.035) is one-sided
mean(t_share < 0.44)
```

```
## [1] 0.007
```

# Final step in hypothesis tests

- Reject the null or not?
- Decide on *threshold* for rejection (test level $\alpha$)
- **Decision rule**:
  - **Reject null** if p-value is *below* $\alpha$
  - Otherwise **retain the null** or **fail to reject**.
- Common thresholds:
  - $p \geq 0.1$: "not statistically significant".
  - $p < 0.05$: "statistically significant".
  - $p < 0.01$: "highly significant".

# Decision rule

- Threshold $\rightarrow$ arbitrary (not a 'magic cut-off' points).
- $p = 0.051$ vs. $p = 0.049$ ??
- Function of data and sampling procedure.

Ronald Fisher. Journal of the Ministry of Agriculture of Great Britain (1926)

If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 per cent. point), or one in a hundred (the 1 per cent. point). Personally, the writer prefers to set a low standard of significance at the 5 per cent. point, and ignore entirely all results which fail to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment *rarely fails* to give this level of significance. The very high odds sometimes claimed for experimental results should usually be discounted, for inaccurate methods of estimating error have far more influence than has the particular standard of significance chosen.

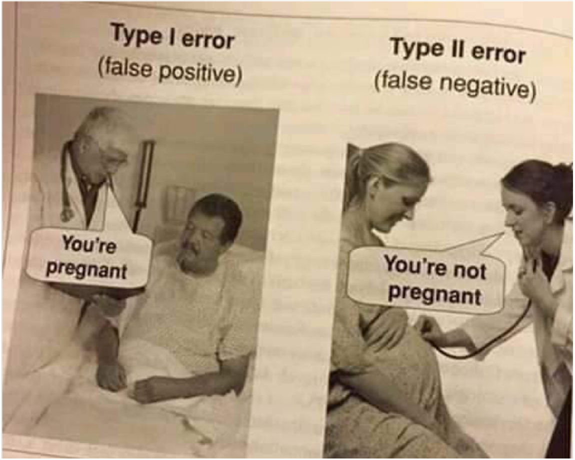
# Test errors

- $p = 0.05 \rightarrow$ extreme data only happen in 5% of repeated samples (if null is true).
- $\rightsquigarrow$ 5% of time we reject null that is true!
- Types of errors:

| | $H_0$ True | $H_0$ False |
|---|---|---|
| Retain $H_0$ | Awesome! | Type II error |
| Reject $H_0$ | Type I error | Good stuff! |

# Test errors

- What does these errors mean?



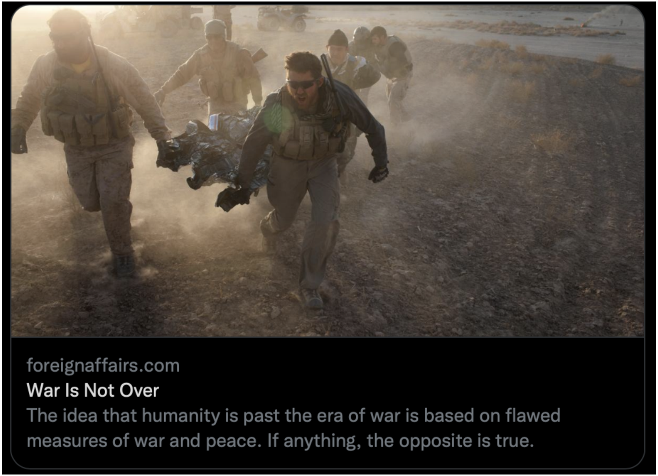
# Error types

- Trade-offs between error types:
  - What if we never reject the null?
  - P(type I) $= 0$.
  - Yet $\rightarrow$ P(type II) $= 1$.
- Rejecting null: yes/no decision.
- p-value $\rightarrow$ refute the null.
- Not a test of quantity of interest.

*Statistical! Not scientific significance*

# Figuring p-values

THE DECLINE OF WAR?? (Fazal and Poast 2019)



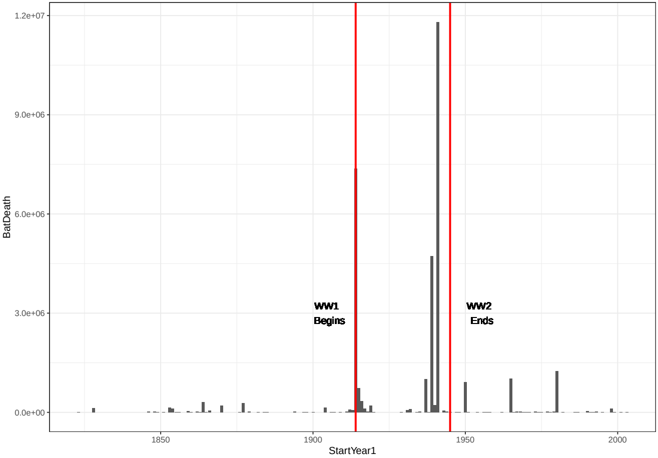
# The decline of war

- An empirical question
- Testing casualties data: COW war data (1816-2014)

```
##   WarNum              WarName WarType ccode                StateName StartMon
## 1      1  Franco-Spanish War       1   230                    Spain
## 2      1  Franco-Spanish War       1   220                   France
## 3      4 First Russo-Turkish       1   640           Ottoman Empire
## 4      4 First Russo-Turkish       1   365                   Russia
## 5      7    Mexican-American       1    70                   Mexico
## 6      7    Mexican-American       1     2 United States of America
## 7     10    Austro-Sardinian       1   337                  Tuscany
## 8     10    Austro-Sardinian       1   325                    Italy
##   StartYear1 BatDeath
## 1       1823      600
## 2       1823      400
## 3       1828    80000
## 4       1828    50000
## 5       1846     6000
## 6       1846    13283
## 7       1848      100
## 8       1848     3400
```

# War death data

- Outlier years (1914-1945)
- What about pre-1914, and post-1945?



# War death data

- Compare battle deaths

```
. sum total_deaths if pre_1914==1
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| total_deaths | 38 | 45615 | 84600.76 | 1000 | 310000 |

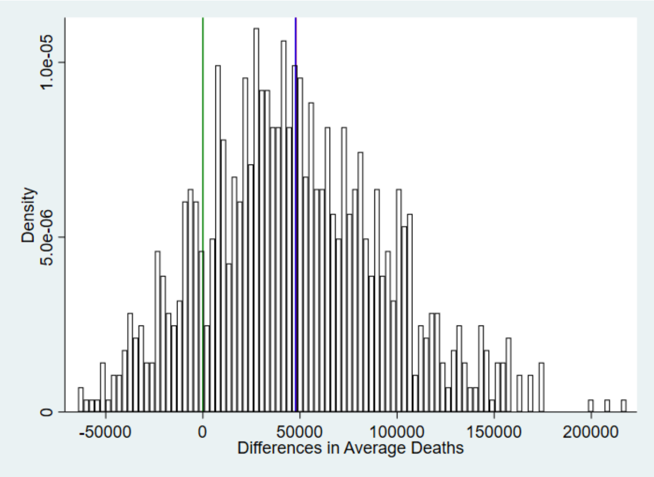
```
.
. sum total_deaths if post_1945==1
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| total_deaths | 38 | 93334.74 | 290449.9 | 1001 | 1250000 |

- Data issue?
- Correct using bootstrap method.

# Uncertainty in war death data

- Simulating battle deaths
- Difference is 16%, enough?



# One sample tests

- Our sample mean - meaningful or random?
- Not just for binary r.v.s (Trump support).
- Steps for testing:
  1. Define null and alternative hyps ($H_0; H_1$).
  2. Select *test statistic* and level of test ($\alpha$).
  3. Derive reference distribution of measure under null.
  4. Calculate p-values.
  5. Make a decision: reject/retain.

# One sample tests

- **Test statistic**:
  - Function of data.
  - Used to assess both hypotheses.
  - Most common $\rightarrow$ *z-score*
- The z-statistic:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Or:

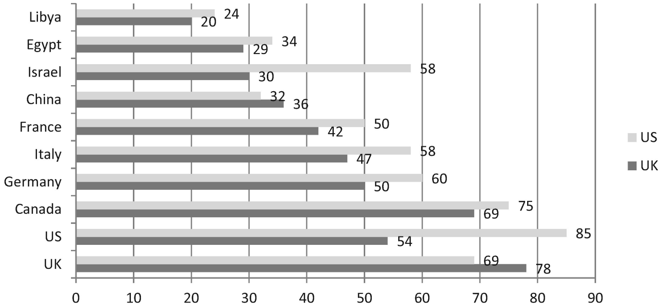
$$Z = \frac{observed-null}{SE}$$

- How many SEs away from the null guess is the sample mean?

# Testing sample means

## Thermometer scores

- Attitudes towards groups/concepts (0 ('cold') - 100 ('warm') scale).
- American attitudes to other nations (2017)



# Americans attitudes: My Homecoming

- Focus on Israel.
- Score of 58 $\rightarrow$ in survey (Scotto and Reifler 2017).
- Recent surveys (Pew, 2019) $\rightarrow$ mean score of 55.
- Did attitudes shift? How 'real' are our survey results?
- Use hypothesis test!!

# Testing public attitudes

- The set-up:
  - $H_0 : \mu = 55$
  - $H_1 : \mu \neq 55$
- Calculate the z-stat:

$$Z = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{58 - 55}{30/\sqrt{800}} = 2.82$$

- Observed score is about 3 SEs from the null!!
- How unlikely is it??

# Testing public attitudes data

$$Z = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}$$

- Check the distribution of z under null.
- Not binomial, use CLT in large samples.
- X $\sim$ N($\mu$, $\sigma^2$)
- z is normal and follows the standard normal (mean 0, SD 1).

# Testing public attitudes data in R

- What is the p-value of our test?

```r
### Israel thermometer
# Define all values
n <- 800
x_bar <- 58
mu <- 55
sd <- 30
se <- (sd/sqrt(n))

# Calculate z-score (2.82)
z_ISR <- (x_bar - mu)/(sd/sqrt(n))

# What is the p-value?
pnorm(-z_ISR)
```

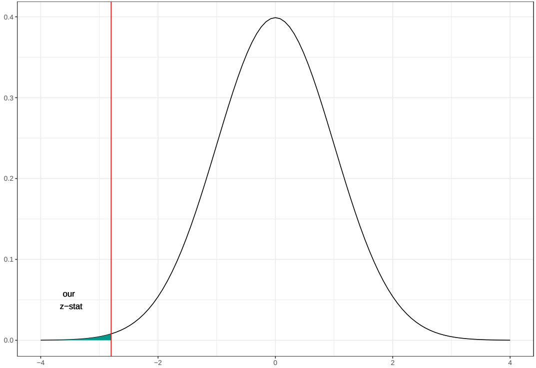
```
## [1] 0.002338867
```

# Our survey plotted

- z-score (-2.8) and p-value (0.002)

```r
# Create standard normal distribution data
p.norm <- data.frame(x = c(-4,4))

ggplot(p.norm, aes(x=x)) + stat_function(fun = dnorm) +
  geom_area(stat = "function",  fun = dnorm, fill = "#00998a", xlim = c(-4,-2.8)) +
  geom_vline(xintercept = -2.8, color = "red") + xlab("") + ylab("") +
  geom_text(aes(x=-3.5, y=0.05, label = "our \n z-stat")) + theme_bw()
```



# Small samples problem

- z-score is useful for large samples under CLT.
- Small sample $\rightarrow$ uncertainty about $\bar{X}$ distribution.
- Assume $X_i$ is normally distributed (not likely!)
- Find t-statistic instead:

$$T = \frac{\bar{X} - \mu}{\hat{SE}} \approx t_{n-1}$$

- t-distribution with n-1 degrees of freedom.
- Centered around zero.

# Test with t-distribution

- t-test vs. z-test:
  - z $\rightarrow$ normal distribution (large samples, n $>$ 30).
  - t $\rightarrow$ t-distribution (smaller samples, more conservatives).
  - In t-distribution: larger p-values (less likely to reject null).
- One more testing option: construct CIs
- If $\mu = 55$ is not in CIs $\rightarrow$ **reject null!**

```r
# Construct 95% and 99% CIs for x_bar = 58
c(x_bar - qnorm(0.995) * se, x_bar + qnorm(0.995) * se)
```

```
## [1] 55.26792 60.73208
```

```r
c(x_bar - qnorm(0.975) * se, x_bar + qnorm(0.975) * se)
```

```
## [1] 55.92114 60.07886
```

# Test for proportions

- Trump support numbers again: survey (0.44); actual (0.475).
- prop.test(): calculate p-value and 95% CIs.

```
# 600 supports in 1363 sample (x-bar)
prop.test(600, n = 1363, p = 0.475, correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  600 out of 1363, null probability 0.475
## X-squared = 6.6171, df = 1, p-value = 0.0101
## alternative hypothesis: true p is not equal to 0.475
## 95 percent confidence interval:
##  0.4140563 0.4666907
## sample estimates:
##         p
## 0.4402054
```

# Two sample tests

- Goal: learn about population differnece in means.
- Compare differences b-w multiple groups: same testing procedures.
- Define:
  - Null PATE: $H_0 : \mu_T - \mu_C = 0$
  - Alt. PATE: $H_1 : \mu_T - \mu_C \neq 0$
  - Test statistic: diff-in-means estimator.
  - z-score for *two sample z-test*.
- Are the differences in sample means just random chance?

# Two sample test: generating data

- Generate experimental data (recipient regime and foreign aid)
- Focus on treatment 1 (regime type) & continuous DV (extent of aid)

```
head(exp.dat, n=8)
```

```
##     ID trt1 trt2 dv_cor1 dv_cor2 cont_cor1 cont_cor2
## 1 0001    1    0       1       0  1523.100  1395.533
## 2 0002    1    1       1       1  1492.402  1466.578
## 3 0003    1    0       1       0  1500.165  1431.904
## 4 0004    1    0       1       0  1510.011  1406.666
## 5 0005    0    1       0       1  1515.649  1442.158
## 6 0006    0    0       0       0  1512.053  1430.640
## 7 0007    0    0       0       1  1474.265  1451.380
## 8 0008    1    1       1       0  1498.759  1443.719
```

# Two sample test

- Define groups to compare in the data
- Run a **two sample t-test**

```
t.test(exp.dat$cont_cor1[exp.dat$trt1 == 0],
       exp.dat$cont_cor1[exp.dat$trt1 == 1])
```

```
##
##  Welch Two Sample t-test
##
## data:  exp.dat$cont_cor1[exp.dat$trt1 == 0] and exp.dat$cont_cor1[exp.dat$tr
## t = -13.697, df = 993.53, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -23.59653 -17.68267
## sample estimates:
## mean of x mean of y
##  1489.333  1509.973
```

# Two sample test

- Hawk/Dove leader and FP actions (Mattes and Weeks 2019)

```r
# The data
head(hawksdata, n=3)
```

```
## # A tibble: 3 x 32
##      caseid  hawk_t party_t rapproche_t success_t    hawk    intl   trust voted16
##       <dbl> <dbl+l> <dbl+l>   <dbl+lbl> <dbl+lbl> <dbl+l> <dbl+l> <dbl+l> <dbl+l>
## 1 329144398 2 [Dip~ 1 [Rep~ 1 [Reducin~ 1 [Pulls~ 4 [Agr~ 4 [Agr~ 2 [The~ 4 [Yes~
## 2 329105048 1 [Mil~ 1 [Rep~ 1 [Reducin~ 1 [Pulls~ 2 [Dis~ 4 [Agr~ 2 [The~ 4 [Yes~
## 3 328964530 1 [Mil~ 2 [Dem~ 1 [Reducin~ 1 [Pulls~ 2 [Dis~ 2 [Dis~ 2 [The~ 4 [Yes~
## # ... with 23 more variables: polact_1 <dbl+lbl>, polact_2 <dbl+lbl>,
## #   polact_3 <dbl+lbl>, polact_4 <dbl+lbl>, hddv1 <dbl+lbl>,
## #   hdmed1_strat <dbl+lbl>, hdmed1_pacifist <dbl+lbl>,
## #   hdmed1_warmonger <dbl+lbl>, hddv2 <dbl+lbl>, hdmed2_strat <dbl+lbl>,
## #   hdmed2_pacifist <dbl+lbl>, hdmed2_warmonger <dbl+lbl>, birthyr <dbl>,
## #   gender <dbl+lbl>, educ <dbl+lbl>, pid3 <dbl+lbl>, pid7 <dbl+lbl>,
## #   ideo5 <dbl+lbl>, newsint <dbl+lbl>, pew_religimp <dbl+lbl>, ...
```

```r
# Groups of support by Hawk/Dove treatment
table(hawksdata$hawk_t,hawksdata$hddv1)
```

```
##
##       1   2   3   4   5
##   1  59 132 148 187  74
##   2  73  83  83 217 143
```

# Two sample test

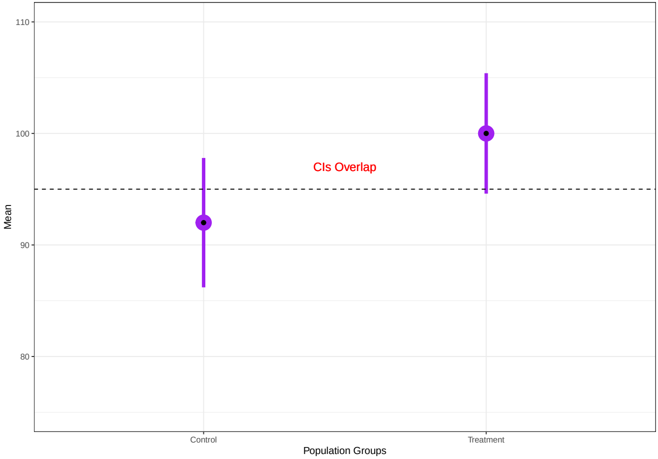
- Does type (hawk/dove) makes a difference?
- Use **t-test**: compare support (1-5 scale) b-w groups

```r
# t-Test
t.test(hawksdata$hddv1[hawksdata$hawk_t == 1],
       hawksdata$hddv1[hawksdata$hawk_t == 2])
```

```
##
##  Welch Two Sample t-test
##
## data:  hawksdata$hddv1[hawksdata$hawk_t == 1] and hawksdata$hddv1[hawksdata$
## t = -4.3646, df = 1183, p-value = 1.385e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4577038 -0.1738210
## sample estimates:
## mean of x mean of y
##  3.141667  3.457429
```

# Testing failures

- We fail to reject null, but null is still untrue. Why?
- Sample is too small → 'inflates' SEs.



# Power analysis

- **Power Analysis**:
  - Power = 1 - P(type II error).
  - Type II error: *false negative*.
  - Maximize test power!!!
- Calculate smallest sample size to identify differences.
- Like MOE discussion - initial step in survey design
- Minimum sample size to reject the null:
  - Surveys: similar proportion b-w groups (candidates).
  - Experiments: reject zero ATE.

# Finding the power...

- Define:
  1. Hypothetical sample size.
  2. Probability of rejecting null (p-value, $\alpha$).
  3. Proportions of sample and 'population'.
- Calculate power of test (prefered power $> 0.8$).

```r
# Define parameters and values
n <- 250
p_bar <- 0.47
p <- 0.5
alpha <- 0.05
cr_value <- qnorm(1 - alpha/2)
se_bar <- sqrt(p_bar * (1-p_bar) / n)
se_p <- sqrt(p * (1-p) / n)

# power
pnorm(p - cr_value * se_p, mean = p_bar, sd = se_bar) +
  pnorm(p + cr_value * se_p, mean = p_bar, sd = se_bar, lower.tail = F)
```

```
## [1] 0.1572895
```

# Now for the power analysis

- Find sample size: power.prop.test() function.
- Define: Groups proportions ($p_1, p_2$); significance level ($\alpha$)' power of test.

```
# What is minimum group size? (power = 0.9)
power.prop.test(p1 = 0.5, p2 = 0.45, sig.level = 0.05, power = 0.9)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 2094.153
##              p1 = 0.5
##              p2 = 0.45
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Power analysis

- Smaller proportions differences

```r
# What is minimum group size? (power = 0.9)
power.prop.test(p1 = 0.05, p2 = 0.1, sig.level = 0.05, power = 0.8)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 434.432
##              p1 = 0.05
##              p2 = 0.1
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Power analysis

- Using power.prop.test() to find the test power

```r
# Define sample size and groups proprtions
power.prop.test(n = 100, p1 = 0.05, p2 = 0.1, sig.level = 0.05)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 100
##              p1 = 0.05
##              p2 = 0.1
##       sig.level = 0.05
##           power = 0.2674798
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

# Power analysis

- Continuous variables: *t-test* version of power analysis.
- One sample: define mean (delta) and SD of r.v. in sample.
- Null: $H_0 : \bar{X} = 0$

```r
# Minimum sample size
power.t.test(delta = 0.3, sd = 1, type = "one.sample", power = 0.9)
```

```
##
##      One-sample t test power calculation
##
##               n = 118.6865
##           delta = 0.3
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
```

# Power analysis

- Two sample: define diff-in-means (delta) and SD.
- Null: $H_0 : \bar{X}_T - \bar{X}_C = 0$

```
# Minimum sample size
power.t.test(delta = 0.3, sd = 1, type = "two.sample",
             alternative = "one.sided", power = 0.85)
```

```
##
##      Two-sample t test power calculation
##
##               n = 160.443
##           delta = 0.3
##              sd = 1
##       sig.level = 0.05
##           power = 0.85
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```

# Power analysis

- Two sample: define diff-in-means (delta) and SD.
- For small effects → more 'demanding' sample

```r
# Minimum sample size
power.t.test(delta = 0.03, sd = 1, type = "two.sample",
             alternative = "one.sided", power = 0.85)
```

```
##
##      Two-sample t test power calculation
##
##               n = 15976.9
##           delta = 0.03
##              sd = 1
##       sig.level = 0.05
##           power = 0.85
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```

# Wrapping up Week 12

- Summary:
  - Testing uncertainty: beyond measures of spread.
  - Proof by contradiction.
  - The Null and Alternative hypotheses.
  - p-value, one-sided or two-sided test.
  - When to reject the null? (obtain it?)
  - Type I & II errors.
  - Thermometer scores and public attitudes (continuous measure).
  - t-test for sample mean or two groups.
  - Power analysis.