Introduction to Empirical Methods

Lecture 2, Part 1: Statistical Inference and Bivariate Hypothesis Testing

Step By Step

- Observe puzzle or question of interest
  - Identify unit & variation of interest
- Create a causal theory (why & how x moves y)
  - Draw out testable hypotheses from theory
- Establish research design to test hypotheses
  - Define unit of observation
  - Define temporal/spatial domain of study
  - Operationalize variables
- Now, we examine our data and test our hypothesis

What We Are Doing?

- Seek to infer from sample (known data) to population (unknown data)
- Population: all occurrences of your phenomenon of interest
- Sample: a subset of the population of interest
- Take what is known about sample and infer that traits also apply to population
    - Statistical Inference
        - Statistical inference is always employed except in rare cases when population as a whole is surveyed.

Parameters

- Parameters
  - Traits that can be quantified such as averages, differences between groups, and relationships among variables.
- Roman Alphabet: Sample parameter
  - Sample Mean is $\bar{X}$
- Greek Alphabet: Population parameter
  - The *real* value
  - Population mean is μ
  - Our estimate of the population mean is $\hat{\mu}$
  - Typically, $\hat{\mu} = \bar{X}$

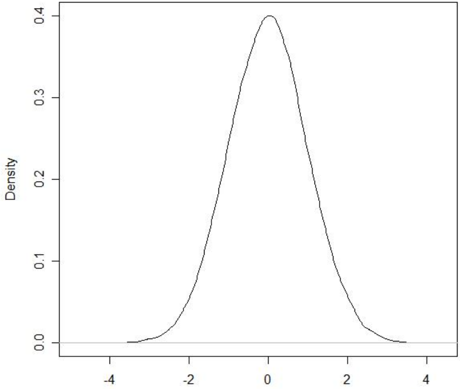
Samples to Populations

- Why do we take samples, if we care about the population?
  - Feasibility: cannot always survey or document every occurrence of a phenomenon.
  - Costs: survey all cases is almost always prohibitively expensive.
  - Practicality: no need to survey population.
- If "done well," sample parameters (traits) will accurately reflect population parameters (traits).

Random Samples

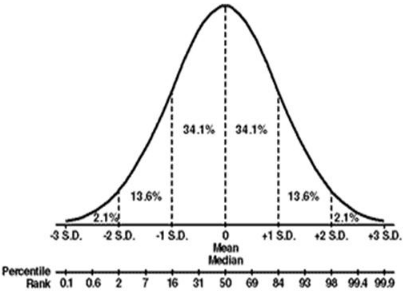
- Simple Random Sample
  - Ideal standard for a sample.
  - Every case in your population of interest has an equal chance of being selected as part of the sample.
  - On average, parameters from a simple random sample will reflect the population parameters.
- Simple random samples are useful because of the *central limit theorem*.
- Central Limit Theorem: all sampling distributions follow a normal distribution *in the limit* (i.e., as they get very large in size).

Standard Normal Distribution



- Bell-shaped and symmetric.
- Mean, median, and mode are all the same.
- Predictable area under the curve.

The "68-95-99" Rule



- 68% of cases fall within one standard deviation of the mean.
- 95% of cases fall within two standard deviation of the mean.
- 99% of cases fall within three standard deviations of the mean.

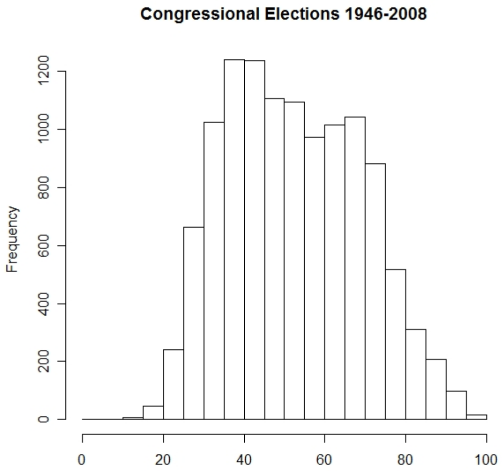
Back to the Central Limit Theorem

- Note: the Central Limit Theorem *does not* say that a variable in a sample is normally distributed.
- Any sample we draw is only one of a number of samples we could draw.
  - So, we could draw many samples and take the means of all of those samples.
- If we were to then graph those sample means, that would show us the *sampling distribution.*
  - The Central Limit Theorem says that the sampling distribution will be normally distributed.

An Example

- We want to know what share of the vote the Democratic candidate receives in congressional elections.
  - Between 1946-2008 there were 11,720 contested congressional elections – not counting special elections.
  - On average the Democratic candidate received 52.5% of the vote.
  - This is μ, the population parameter (note the absence of the "hat"). It is a value we can know: inference is not necessary.

All Congressional Elections

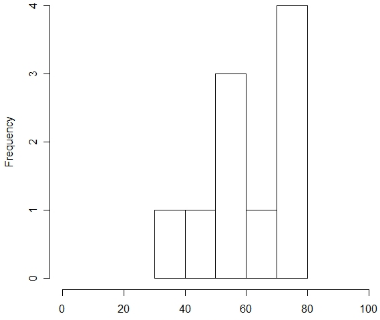


Hypothetically Speaking

- Imagine we did not have any data on congressional elections.
- We could collect the population data or we could save time and effort by taking a random sample.
- How large of a sample should we take?
  - Let's begin with one sample of 10.
  - Then we will compare ten samples of 10, ten samples of 100, and ten samples of 1,000.

The First Sample

- We randomly select ten congressional elections.
- The mean of this sample is 57.38
  - This is $\bar{X}$



Nine More Samples of Ten

- Remember $\mu = 52.5$

Sample #	$\bar{X}$	Sample #	$\bar{X}$
1	**57.38**	6	52.06
2	49.49	7	**47.77**
3	51.73	8	48.19
4	**60.20**	9	51.12
5	51.21	10	52.36

Increase the Sample Size

- Ten samples of 100 elections
- Remember $\mu = 52.5$

Sample #	$\bar{X}$	Sample #	$\bar{X}$
1	54.52	6	53.05
2	53.18	7	54.01
3	55.07	8	52.63
4	53.40	9	52.54
5	53.85	10	54.33

Increase the Sample Size Again

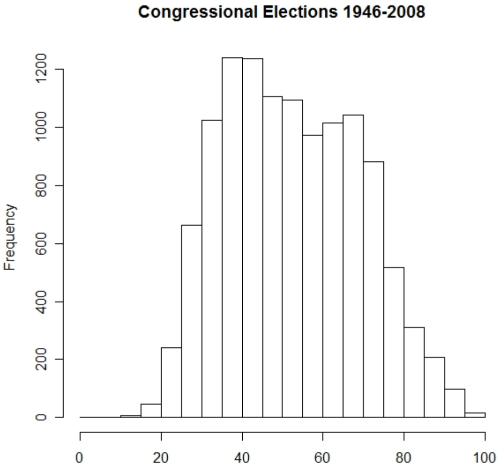
- Ten samples of 1000 elections
- Remember $\mu = 52.5$

Sample #	$\bar{X}$	Sample #	$\bar{X}$
1	51.76	6	51.79
2	52.66	7	52.59
3	52.19	8	53.38
4	52.90	9	52.56
5	52.46	10	51.72

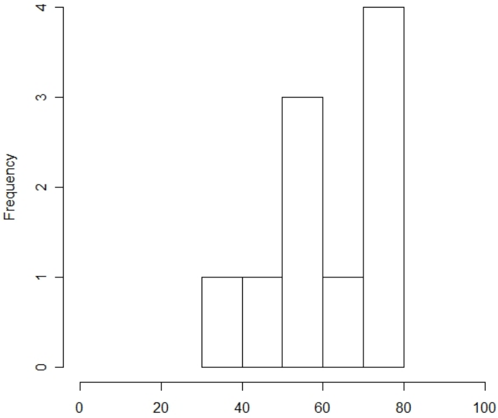
What do we notice?

- As our sample size gets larger our sample means are typically closer to the true mean.
    - And fewer are way off.
- What is happening?
    - For one thing, our larger samples look more like the population.

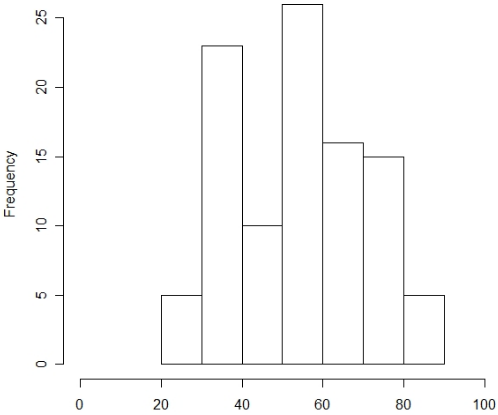
The Population



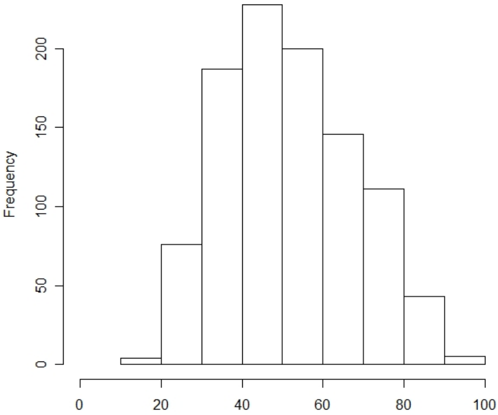
One Sample of 10



One Sample of 100



One Sample of 1,000



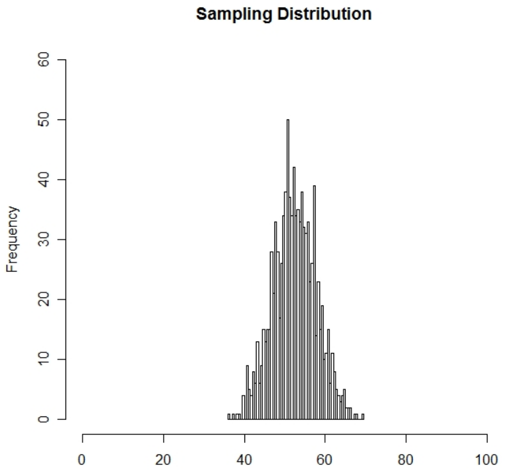
What else is going on?

- Our sampling distribution is becoming normal and narrower.
    - This is a good thing.
- The expected value of the sample mean equals the population mean.
- As our sample size increases:
    - The standard error of the mean decreases.
    - Our confidence that our inference is valid grows. Why?
    - We think our sample mean is the true mean, and if it is not, it will not be all that far off.

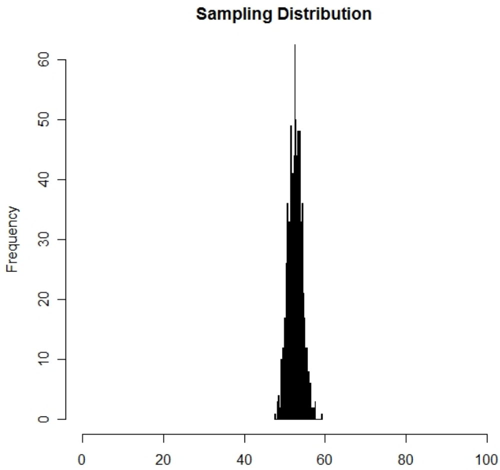
Sampling Distribution

- Let's draw 1,000 samples instead of ten samples.
    - A computer can do this easily enough for us.
    - 1,000 samples of ten elections.
    - 1,000 samples of one hundred elections.
    - 1,000 samples of one thousand elections.
    - 1,000 samples of five thousand elections.
- Watch what happens to the sampling distribution.
    - We don't actually have to take many samples because statisticians can tell us what the distribution looks like as the sample size gets larger.

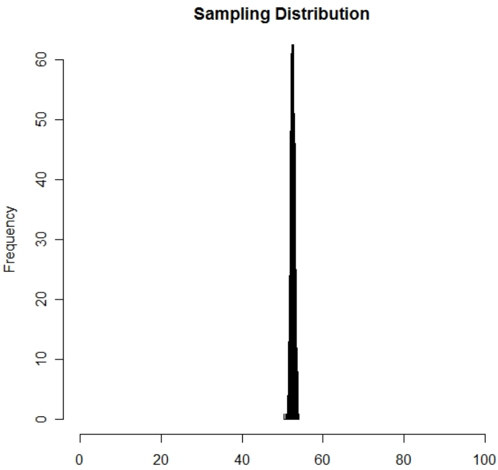
1,000 Samples of Ten Elections



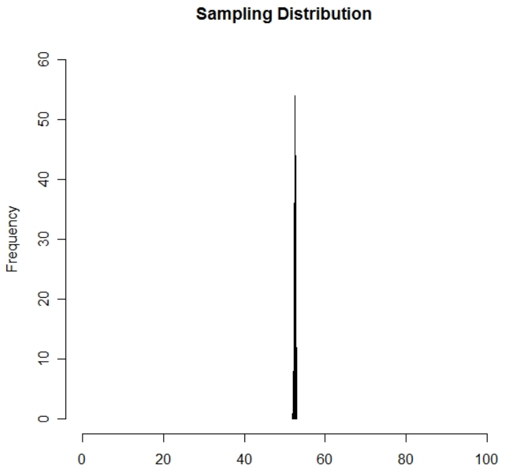
1,000 Samples of One Hundred Elections



1,000 Samples of One Thousand Elections



1,000 Samples of Five Thousand Elections



So, What Do We Know?

- We want to know the population parameters.
  - Often we can only get sample estimates of these parameters.
- As our sample size increases
  - The expected value of our estimate is more likely to equal the population value.
  - The sampling error decreases.
  - The sampling distribution takes on a known distribution.

Pay Attention to This

- *Central Limit Theorem*
  - For any trait or variable, even those that are not normally distributed in the population, if repeated random samples of size N are drawn from any population, with mean $\hat{\mu}$ (or $\bar{X}$) and standard deviation $\hat{\sigma}$ (or s), then, as N becomes large, the sampling distribution of sample means will approach normality with mean μ and standard deviation $\frac{\hat{\sigma}}{\sqrt{N}}$
    - Generally by "N becomes large" a rule of thumb is $N > 100$
- The importance of this will become clearer as we do more statistics.

Standard Deviation of Sampling Distribution

- *Central Limit Theorem*
  - The sampling distribution of sample means will approach normality with mean μ and standard deviation $\frac{\hat{\sigma}}{\sqrt{N}}$
- The standard error tells us the uncertainty in our estimate of the mean without pulling hundreds of repeated samples.
- Standard error of the mean equals the standard deviation of the sample over the square root of the sample size.
  - $\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$

What Does the Standard Error Tell Us?

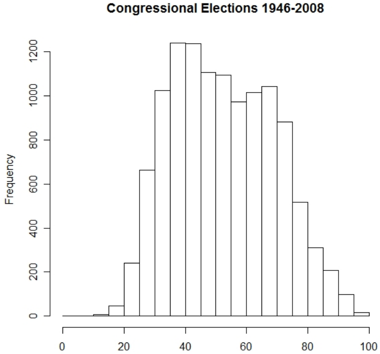
- Now, we can take a single sample and obtain:
  - An estimate of the population mean.
  - An estimate of how confident we are in that mean.
- We can also see that as our sample size gets larger our confidence will increase.
  - Because our standard error will decrease.

Big Sample=Small Standard Error

- The formula for the standard error has the sample size in the denominator.
- As the denominator increases, the fraction becomes smaller.
  - $8/2 > 8/4 > 8/6 > 8/8 > 8/10 > 8/12 > 8/14$ etc.
- So, the standard error decreases as the sample size increases.
- Remember, smaller standard error means greater confidence in our estimate.

Putting This into Practice

- Recall, Democratic candidates received 52.5% of the vote in congressional elections on average.
- The distribution looks like this:



Take a Sample of 1,000.

- The mean of this sample is 51.58
    - This is 0.92 off of the true mean.
        - Remember, we typically will not know the true value.
- The standard deviation of the sample is 16.035
- So, the standard error of the mean is $\frac{16.035}{\sqrt{1000}} = 0.51$.

What Does This Tell Us?

- Because we know that the sample means are from a normal distribution we can talk about a confidence interval.
- The standard error is the standard deviation of the sampling distribution.
- 95% of the sampling distribution lies plus or minus 1.96 standard deviations from the mean.
  - n.b. This assumes a large sample size.

Confidence Interval of a Mean

- To calculate the confidence interval of the mean we need to know:
  - Mean
  - Standard Deviation, N-Size
    - Or Standard Error
- $C.I. = \bar{X} \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}$
  - Or $[\bar{X} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}]$

Confidence Intervals Again

- Our sample mean was 51.58.
- The standard error was 0.51.
- So, we are 95% confident that the population mean lies between 50.59 and 52.57.
    - $51.58 \pm 1.96 * 0.51$
- There is a 2.5% chance the population mean is smaller than 50.59.
- There is a 2.5% chance the population mean is larger than 52.57.

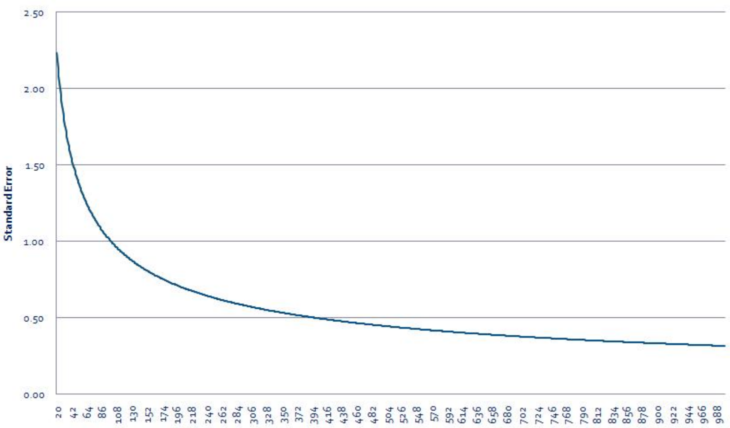
What about a Smaller Sample?

- Another sample, this time of 100 congressional elections.
  - $\bar{X} = 51.34$
  - $\hat{\sigma}_{\bar{X}} = 1.71$
- The 95% confidence interval is 47.99 to 54.68.

A Comparison

- The 95% confidence interval is 47.99 to 54.68.
  - Sample Size=100
- The 95% confidence interval is 50.59 and 52.57.
  - Sample Size=1000
- We are less confident of what the true mean is with a smaller sample size.
  - With a smaller sample size, the standard error is larger.

How Large a Sample Size?



- Sample mean held at 50 and standard deviation is held at 10.

Introduction to Probability

- Subjective vs objective probability
- Randomness/Stochasticity
- Outcomes and sample spaces
- Events
    - Probability of occurrence
    - Independent
    - Mutually Exclusive
    - Collectively Exhaustive
    - Joint
    - Conditional

Distributions of Random Variables

- Random Variable
  - Definition
  - Distribution
  - Realization
  - Support
- Distribution
  - Sample/Empirical vs Theoretical
  - PMF/PDF
  - Data Generating Process (DGP)
  - CDF
  - Examples: Uniform and Normal

What is hypothesis testing?

- We will cover two basic comparisons:
  - Comparing our sample estimate to a hypothetical parameter.
  - Comparing sample estimates of two groups.
- We cannot just look at the difference between the estimates and the hypothetical parameter.
  - We have to deal with the uncertainty around our estimates.
  - We need to eliminate the possibility that any difference was just the result of random chance.

Research and Null Hypotheses

- H_1: The research hypothesis.
  - The difference we expect to see.
    - The difference between groups (treatment and control).
    - The difference between our estimate and the hypothesized parameter.
- H_0: The null hypothesis.
  - Any difference observed is just the result of random chance.
- Our goal is to *reject the null hypothesis.*
  - That is, we want to eliminate the possibility that any difference we see is just the result of chance.

The Five-Step Model for For Hypothesis Testing

1. Make assumptions and determine if those assumptions were met.
    - Typically we assume random sampling ($\leftarrow$ Central Limit Theorem).
2. State the null hypothesis.
    - Typically: there is no difference.
    - Exception: if we have a directional research hypothesis.
3. Determine the sampling distribution and critical region.
4. Calculate your test statistic.
    - The sampling distribution and the test statistic will depend on the particular hypothesis you're testing.
5. Interpret the results.
    - Do we reject the null hypothesis?
        - Is there a statistically significant relationship?

Types of Tests

- What kind of hypothesis test should we use?
    - What is the level of measurement for our independent and dependent variables?

		Ind.V. Type	
		Categorical	Continuous
D.V.	Categorical	Tabular Analysis	Probit/Logit
Type	Continuous	Difference of Means	Correlation Coefficient

Confidence in our Hypothesis

- Hypothesis tests approximate confidence in your hypothesis.
- Structured around the null hypothesis:
  - Statement of no relationship between x & y
- Never "prove" your hypothesis is correct:
  - Falsify that it is incorrect.
  - Seek to falsify null hypothesis.

Confidence in our Hypothesis

- Hypothesis testing: compare actual relationship between x & y in your sample data to what we expect to find if x & y unrelated (null)
- Norm: we want to be 95% confident
  - If relationship you observe would have occurred by chance less than 5 times out of 100, reject null hypothesis
- Hypothesis testing produces a *p-value* used to evaluate null hypothesis
- Relationship between x & y said to be statistically significant at the 95% confidence level.

Rejecting the Null

- Typically, we set our confidence level at 95%.
  - That is, if our test statistic has only a 5% chance of occurring if the null is true, then we will reject the null hypothesis.
  - The goal is to lessen the chance of Type I Error.
- Type I Error
  - Rejecting the null when the null is true.
  - Saying there is a relationship when there is not.
- Type II Error
  - Accepting the null when the null is false.
  - Saying there is not a relationship when there is.

More About *p-Values*

- p-value tells us the probability we would see the observed relationship in the sample data if there were truly no relationship in the population.
  - Ranges between 0-1.
  - Lower p-values indicate more confidence in the relationship.
  - With a larger sample size, we should get lower p-values.

Limitations of *p-Values* (aka Statistical Significance)

- Subject to manipulation:
  - Higher sample size reduces p-value (by reducing standard error)
- Not comparable (with other p-values):
  - Smaller p-value (Ex: $p < .05$ & $p < .04$) does not mean that one relationship is stronger than another.
- Silent as to the validity of our measures.
- Key Point: p-value is not conclusive evidence for causality (it is one piece of the larger puzzle):
  - Statistical significance only detects whether movement in x & y is likely to be due to chance.

A Cautionary Note

- Note the difference between a "test of statistical significance" & a "measure of association".
- Test for statistical significance: is the observed relationship between x & y due to chance?
- Measure of association: gages the strength or magnitude of the observed relationship between x & y.
  - A very large sample size will often result in a "statistically significant" effect, even if the actual relationship has little to no substantive significance.

Tabular Analysis (aka Cross Tabs)

		I.V. Type	
		Categorical	Continuous
D.V.	**Categorical**	**Tabular Analysis**	Probit/Logit
Type	Continuous	Difference of Means	Correlation Coefficient

Cross Tabs: the χ^2 test

- Two ordinal or categorical variables: study a cross tab.
  - That's short for cross tabulation.
- We can also use cross tabs to figure out if there are differences between groups.

Example

- Is there a difference in how Republicans and Democrats voted in Congressional Elections?
  - Our hypothesis: there is a difference.

	Democrats	**Republicans**
Vote for Dem. Cand.	233	43
Vote for Rep. Cand.	18	230

- We can look at this table and tell that there is a difference.
  - But what if it were a closer call?

A Harder Example

- Do people with college degrees vote differently than people without college degrees?

	No College Degree	College Degree
Vote for Dem. Cand.	209	225
Vote for Rep. Cand.	169	156

	No College Degree	College Degree
Vote for Dem. Cand.	55.3%	59.1%
Vote for Rep. Cand.	44.7%	40.9%

- From this table, I would say there is no difference, but don't quote me.
  - We can perform a statistical test to see.

The Hypothesis Test

- Our *research hypothesis*: There is a relationship between education and vote choice in US Congressional elections.
  - This is non-directional, but for cross tabs that is what we will do.
- The *null hypothesis*: There is no relationship between education and vote choice in US Congressional elections.
  - If we reject the null hypothesis, then we will say that we are fairly certain there is a relationship.

The Distribution

- In this particular example, we are looking at relationships in a table so we will use a Chi-Square Statistic (χ^2).
  - Strictly speaking this is a test of statistical independence.
  - This is one of the easiest hypothesis tests to perform.

χ^2

- The goal is to determine if there is a systematic relationship between our independent (education) and dependent (vote) variables.
- So, we are going to calculate the table we would have expected to have seen if there had been no relationship, and compare it to the table that we actually saw.
  - The greater the difference between the expected values and the observed values, the larger our chi-square statistic, the less likely the null is true, and the more likely there is a relationship.

The First Step

- You must first calculate the *marginals*.
    - The row and column totals.

	No Degree	Degree	Total
Vote for Dem. Cand.	209	225	
Vote for Rep. Cand.	169	156	
Total			

The First Step

- You must first calculate what are called the *marginals*:
  - The row and column totals.

	No Degree	**Degree**	**Total**
Vote for Dem. Cand.	209	225	*434*
Vote for Rep. Cand.	169	156	*325*
Total	*378*	*381*	*759*

Using the Marginals (i.e., row and column totals)

- We now use the marginal totals to calculate the table we would have observed had there been no relationship between the variables.
  - We know how many Democratic and Republican votes there were.
  - We know how many people had college degrees and how many people did not have college degrees.
  - If there had been no relationship, the probability of someone without a degree voting for the Democrat would be the same as someone with a degree voting for the Democrat (and so on).

Expected Table if There were No Relationship #1

- To calculate the expected values, we multiply the marginals for each cell, and then divide by the total number of observations.

	No Degree	Degree	Total
Vote for Dem. Cand.			434
Vote for Rep. Cand.			325
Total	378	381	759

Expected Table if There were No Relationship #2

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	**Degree**	**Total**
Vote for Dem. Cand.	$\frac{434*378}{759}$		**434**
Vote for Rep. Cand.			325
Total	**378**	381	**759**

Expected Table if There were No Relationship #3

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	Degree	Total
Vote for Dem. Cand.	216.14		434
Vote for Rep. Cand.			325
Total	378	381	759

Expected Table if There were No Relationship #4

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	**Degree**	**Total**
Vote for Dem. Cand.	216.14		434
Vote for Rep. Cand.	$\frac{325*378}{759}$		**325**
Total	**378**	381	**759**

Expected Table if There were No Relationship #5

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	Degree	Total
Vote for Dem. Cand.	216.14		434
Vote for Rep. Cand.	161.86		325
Total	378	381	759

Expected Table if There were No Relationship #6

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	Degree	Total
Vote for Dem. Cand.	216.14	$\frac{434*381}{759}$	**434**
Vote for Rep. Cand.	161.86	$\frac{325*381}{759}$	**325**
Total	378	**381**	**759**

Expected Table if There were No Relationship #7

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	**Degree**	**Total**
Vote for Dem. Cand.	216.14	217.86	434
Vote for Rep. Cand.	161.86	163.14	325
Total	378	381	759

Expected Table if There were No Relationship #7

- To calculate the expected values, we multiply the marginals for each cell and divide by the total number of observations.

	No Degree	Degree	Total
Vote for Dem. Cand.	216.14	217.86	434
Vote for Rep. Cand.	161.86	163.14	325
Total	378	381	759

	No Degree	Degree
Vote for Dem. Cand.	57.2%	57.2%
Vote for Rep. Cand.	42.8%	42.8%

Observed vs. Expected: calculating χ^2

- We now have our observed values and our expected values. We can compare those values to calculate our chi-square statistic.
  - We do this by going cell by cell.
- Within a cell, we take the observed value and subtract the expected value.
- We square this difference and then divide that by the expected value.
- We add up all those values for all of our cells and that is our chi-square statistic.

Observed vs. Expected: calculating χ^2

Observed

	No Degree	Degree
Vote for Dem. Cand	**209**	225
Vote for Rep. Cand	169	156

Expected

	No Degree	Degree
Vote for Dem. Cand	**216.14**	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	Degree
Vote for Dem. Cand	$\frac{(209-216.14)^2}{216.14}$	
Vote for Rep. Cand		

Observed vs. Expected: calculating χ^2

Observed

	No Degree	**Degree**
Vote for Dem. Cand	209	225
Vote for Rep. Cand	169	156

Expected

	No Degree	**Degree**
Vote for Dem. Cand	216.14	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	**Degree**
Vote for Dem. Cand	**.24**	
Vote for Rep. Cand		

Observed vs. Expected: calculating χ^2

Observed

	No Degree	Degree
Vote for Dem. Cand	209	225
Vote for Rep. Cand	169	156

Expected

	No Degree	Degree
Vote for Dem. Cand	216.14	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	Degree
Vote for Dem. Cand	.24	$\frac{(225-217.86)^2}{217.86}$
Vote for Rep. Cand		

Observed vs. Expected: calculating χ^2

Observed

	No Degree	Degree
Vote for Dem. Cand	209	225
Vote for Rep. Cand	169	156

Expected

	No Degree	Degree
Vote for Dem. Cand	216.14	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	Degree
Vote for Dem. Cand	.24	.23
Vote for Rep. Cand		

Observed vs. Expected: calculating χ^2

Observed

	No Degree	**Degree**
Vote for Dem. Cand	209	225
Vote for Rep. Cand	169	156

Expected

	No Degree	**Degree**
Vote for Dem. Cand	216.14	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	**Degree**
Vote for Dem. Cand	.24	.23
Vote for Rep. Cand	$\frac{(169-161.86)^2}{161.86}$	$\frac{(156-163.14)^2}{163.14}$

Observed vs. Expected: calculating χ^2

Observed

	No Degree	Degree
Vote for Dem. Cand	209	225
Vote for Rep. Cand	169	156

Expected

	No Degree	Degree
Vote for Dem. Cand	216.14	217.86
Vote for Rep. Cand	161.86	163.14

Observed vs. Expected

	No Degree	Degree
Vote for Dem. Cand	.24	.23
Vote for Rep. Cand	.32	.31

Putting it Together

- We add up the values in those cells to calculate our chi-square statistic.

Observed vs. Expected

	No Degree	Degree
Vote for Dem. Cand	.24	.23
Vote for Rep. Cand	.32	.31

- $\chi^2 = .24 + .23 + .32 + .31 = 1.098$
  - Is this χ^2 large enough to reject the null hypothesis?

Critical Value of χ^2

- We first need to calculate the degrees of freedom.
  - (# of Columns - 1) * (# of Rows - 1)
  - (2-1)*(2-1)=1*1=1 degree of freedom
- With our degrees of freedom and our χ^2 statistic we go to the table (Appendix A in Kellstedt & Whitten).

Degrees of Freedom	Probability										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Nonsignificant								Significant		

Table of χ^2 Values

- The table consists of critical values.
  - A χ^2 value must be at least this large to reject the null at that level.
  - Our χ^2 is 1.098.

df	.10	.05	.02	.01
1	2.706	3.841	5.412	6.635
2	4.605	5.991	7.824	9.210
3	6.251	7.815	9.837	11.341

- Since our χ^2 statistic is smaller than any of the critical values, we cannot reject the null.
  - Therefore, we cannot state that there is any relationship between having a college degree and how one voted in the 2006 US Congressional election.

Another Example

- The 2000 Nat'l Election Survey asked if respondents believed U.S. National Security had gotten better or worse since Pres. Clinton took office in 1993.

	Dem.	Ind.	Rep.
Worse	132	44	296
Same	432	98	233
Better	258	37	117

Calculate the Marginals

	Dem.	Ind.	Rep.	Total
Worse	132	44	296	472
Same	432	98	233	763
Better	258	37	117	412
Total	822	179	646	1647

Calculate the Expected Table

	Dem.	Ind.	Rep.	Total
Worse	$\frac{472*822}{1647}$	$\frac{472*179}{1647}$	$\frac{472*646}{1647}$	472
Same	$\frac{763*822}{1647}$	$\frac{763*179}{1647}$	$\frac{763*646}{1647}$	763
Better	$\frac{412*822}{1647}$	$\frac{412*179}{1647}$	$\frac{412*646}{1647}$	412
Total	822	179	646	1647

The Expected Table

	Dem.	Ind.	Rep.	Total
Worse	235.6	51.3	185.1	472
Same	380.8	82.9	299.3	763
Better	205.6	44.8	161.6	412
Total	822	179	646	1647

Observed vs. Expected

Observed

	Dem.	Ind.	Rep.
Worse	132	44	296
Same	432	98	233
Better	258	37	117

Expected

	Dem.	Ind.	Rep.
Worse	235.6	51.3	185.1
Same	380.8	82.9	299.3
Better	205.6	44.8	161.6

Observed vs. Expected

	Dem.	Ind.	Rep.
Worse	$\frac{(132-235.6)^2}{235.6}$	$\frac{(44-51.3)^2}{51.3}$	$\frac{(296-185.1)^2}{185.1}$
Same	$\frac{(432-380.8)^2}{380.8}$	$\frac{(98-82.9)^2}{82.9}$	$\frac{(233-299.3)^2}{299.3}$
Better	$\frac{(258-205.6)^2}{205.6}$	$\frac{(37-44.8)^2}{44.8}$	$\frac{(117-161.6)^2}{161.6}$

Observed vs. Expected

Observed

	Dem.	Ind.	Rep.
Worse	132	44	296
Same	432	98	233
Better	258	37	117

Expected

	Dem.	Ind.	Rep.
Worse	235.6	51.3	185.1
Same	380.8	82.9	299.3
Better	205.6	44.8	161.6

Observed vs. Expected

	Dem.	Ind.	Rep.
Worse	45.5	1.0	66.4
Same	6.9	2.7	14.7
Better	13.3	1.4	12.3

Final Steps

- $\chi^2 = 45.5 + 1 + 66.4 + 6.9 + 2.7 + 14.7 + 13.3 + 1.4 + 12.3$
  - $\chi^2 = 164.3$
- $d.f. = (3-1)*(3-1) = 2*2 = 4$

d.f.	0.1	0.05	0.02	0.01
1	2.71	3.84	5.41	6.63
2	4.61	5.99	7.82	9.21
3	6.25	7.81	9.49	11.34
4	7.78	9.49	11.67	13.28
5	9.24	11.07	13.39	15.09

- $\chi^2 = 164.3$
- Critical Value at .05, 4 d.f. $= 9.49$
- What does this test tell us?
  - These two variables are not independent.
  - We can say with 95% confidence that there is a relationship between partisanship and opinion.

The Limitations of Cross-Tabs (the χ^2 test)

- We cannot describe the direction or the size of the relationship.
  - Are independents different from Democrats and/or Republicans?
  - What is the size of this relationship?
    - N matters: a small relationship with a large sample size will be statistically significant.

What kind of hypothesis test should we run?

- We'll next move to continuous variables.
- What is the level of measurement for our independent and dependent variables?

		Ind.V. Type	
		Categorical	**Continuous**
D.V.	Categorical	Tabular Analysis	Probit/Logit
Type	**Continuous**	Difference of Means	**Correlation Coefficient**

r: Correlation as a Measure of Covariation

- With a continuous independent variable and a continuous dependent variable we calculate the correlation coefficient.
- To what extent are changes in the values of X associated with changes in the value of Y?
  - Can you predict the value of our dependent variable from the value of our independent variable?
- To describe this we calculate the (bivariate) correlation coefficient also known as "*r*".
  - Measures the strength and direction of the relationship.

Before We Do That...

- The extent to which the values of two variables move together is called "covariation."
  - The statistic that measures the amount of covariation is called the *covariance*.
  - $cov_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$
  - Positive covariance
    - As X increases (decreases), Y tends to increase (decreases).
  - Negative covariance
    - As X increases (decreases), Y tends to decrease (increases).

An Example

- Hypothesis: In states with low per-capita incomes, state governments will employ more workers.

State	Per-Capita Income	% Gov't Employees
Alabama	$24,028	19.2%
Florida	$30,446	14.5%
Georgia	$29,442	16.4%
Mississippi	$23,448	21.8%
North Carolina	$28,235	17.3%
South Carolina	$26,132	18.2%
Tennessee	$28,455	15.5%

Step One

- X=Per-Capita Income
- Y=% Gov't Employees
- First, calculate the mean for each variable.
  - $\bar{X} = \$27,169.43$
  - $\bar{Y} = 17.557\%$

Step Two

- Calculate the deviations from the mean.

X	$\bar{X}$	$X - \bar{X}$	Y	$\bar{Y}$	$Y - \bar{Y}$
24,028	27,169.43	-3,141.43	0.192	0.1756	0.016
30,446	27,169.43	3,276.57	0.145	0.1756	-0.031
29,442	27,169.43	2,272.57	0.164	0.1756	-0.012
23,448	27,169.43	-3,721.43	0.218	0.1756	0.042
28,235	27,169.43	1,065.57	0.173	0.1756	-0.003
26,132	27,169.43	-1,037.43	0.182	0.1756	0.006
28,455	27,169.43	1,285.57	0.155	0.1756	-0.021

Putting it Together

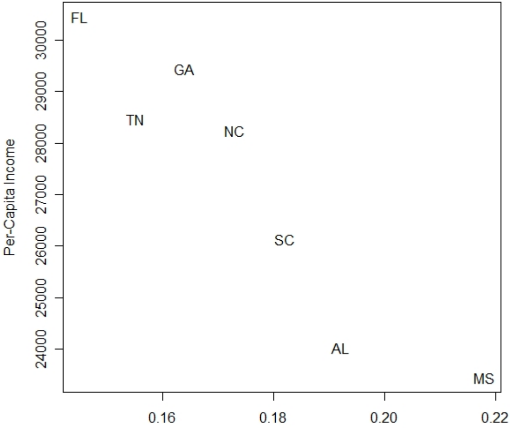
State	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
Alabama	-3141.43	0.016	-51.609
Florida	3276.57	-0.031	-100.169
Georgia	2272.57	-0.012	-26.297
Mississippi	-3721.43	0.042	-157.895
North Carolina	1065.57	-0.003	-2.740
South Carolina	-1037.43	0.006	-6.669
Tennessee	1285.57	-0.021	-26.446
			$\sum = -371.826$

- $cov_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{-371.826}{7-1} = -53.118$

Why a Negative Covariance?

State	Per-Capita Income	% Gov't Employees	Product
Alabama	Below Mean	Above Mean	Negative
Florida	Above Mean	Below Mean	Negative
Georgia	Above Mean	Below Mean	Negative
Mississippi	Below Mean	Above Mean	Negative
North Carolina	Above Mean	Below Mean	Negative
South Carolina	Below Mean	Above Mean	Negative
Tennessee	Above Mean	Below Mean	Negative

Negative Covariation Visually



Calculating the Correlation Coefficient

- Our measure of correlation is Pearson's correlation coefficient or r
- $r = \frac{cov_{XY}}{\sqrt{var_X var_Y}}$
  - We have the covariance between X & Y
  - We need to calculate the variance for each variable.

Correlation in the Example

- The covariance between X & Y is -53.118.
- The variance of X=6211785.673
- The variance of Y=0.000515673
  - Use the formula for the sample variance in this case.
  - $r = \frac{-53.118}{\sqrt{6211785.673 * 0.0005157}} = -0.936$

Important Things to Remember

- r has a range between 1 and -1.
  - If $r = 1$, then there is a perfect positive correlation.
  - If $r = -1$, then there is a perfect negative correlation.
    - We can exactly predict the dependent variable value from the independent variable.
  - There is absolutely no relationship when $r = 0$.
- The closer r is to either extreme, the stronger the relationship.
  - $-.8$ is a stronger relationship than .3

More Things to Remember...

- The correlation coefficient will always calculate the *linear* relationship between the two variables.
  - So, it is inappropriate if the relationship is not linear.
- Just because there is a strong linear relationship, it does not mean that our independent variable is causing our dependent variable.
  - Our correlation coefficient would be exactly the same if we reversed what we labeled "independent" and "dependent" (i.e., switched X and Y).

Correlation and Uncertainty

- There is uncertainty around our calculation of the correlation coefficient in our sample.
- We want to reject the null hypothesis that $r = 0$.
  - Remember, if $r = 0$, there is no association between the variables.
  - Rejecting the null means that we are confident there is an association between the variables.

The t-distribution

- The key here is the Student's t-distribution which is very similar to the normal distribution.
  - The procedure is going to be similar to the χ^2 test.
    1. Calculate test statistic.
    2. Determine the degrees of freedom and look up the critical value.
    3. Compare test statistic to critical value.

Brilliant!

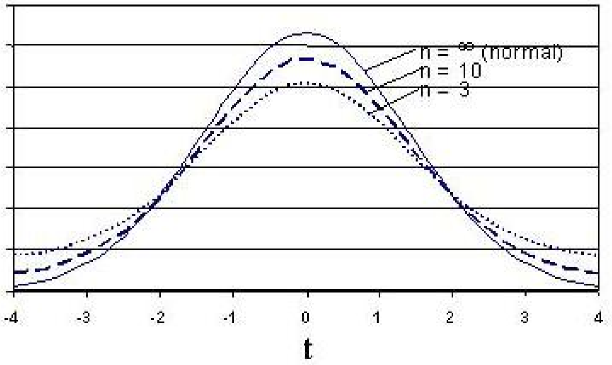
- Brewers do not just brew beer and send it out to the public.
  - They do quality control.

Brilliant!

- So, a worker at the Guinness brewery developed a method of using small samples of beer to test its quality.
- His method used the t-distribution, which he published anonomously under the pseudonym A. Student.

Student's t-Distribution

- The t-distribution is very similar to a standard normal distribution



The t-statistic

- $t_r = \sqrt{|r| * \frac{n-2}{1-r^2}}$
- Degrees of Freedom$=n-1$
  - For our example...
  - $t_r = \sqrt{|-0.936| * \frac{7-2}{1-(-0.936)^2}}$
  - $t_r = 6.15$ with 6 degrees of freedom

Critical Values of t

- $t_r = 6.15$ with 6 degrees of freedom
- Appendix B of Kellstedt and Whitten

	Level of Significance			
df	.10	.05	.025	.01
5	1.476	2.015	2.571	3.365
6	**1.440**	**1.943**	**2.447**	**3.143**
7	1.415	1.895	2.365	2.998

- So, we can reject the null at any level on this table.

What kind of hypothesis test should we run?

- What if the dependent variable is continuous, but the independent variable is categorical?
- What is the level of measurement for our independent and dependent variables?

		Ind.V. Type	
		Categorical	Continuous
D.V.	Categorical	Tabular Analysis	Probit/Logit
Type	**Continuous**	**Difference of Means**	Correlation Coefficient

- Now we're going to look at a difference of means test.

Research and Null Hypotheses

- H_1: The research hypothesis.
  - The difference between groups that we expect to see.
- H_0: The null hypothesis.
  - Any difference observed is just the result of random chance.
- Our goal is to *reject the null hypothesis.*
  - That is, we want to eliminate the possibility that the results we see are just the result of chance.

Confidence in our Hypothesis: Reminder

- Hypothesis testing: compare actual relationship between x & y in your data to what we would expect to find if x & y unrelated (null).
- Norm: want to be 95% confident
  - If relationship you observe would have occurred by chance less than 5 times out of 100, reject null hypothesis.
- Hypothesis testing produces a *p-value* used to evaluate null hypothesis.
- Relationship between x & y said to be statistically significant at the 95% confidence level.

A New Question

- What if what we care about whether two groups have different means?
  - For example...
  - Do Democrats have a different opinion on taxes than Republicans?
  - Do men make more money than women?
- Difference of Means Test
  - $t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{(\bar{X}_1 - \bar{X}_2)}}$

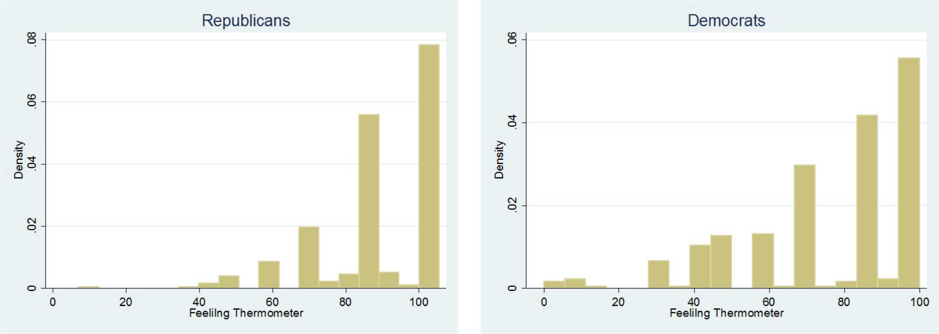
The Background

- For example, if you're looking to see if there is a difference between men and women...
- You could draw a sample of men and a sample of women.
- Or you could draw a sample of people and code a variable that distinguishes men and women, and then split the sample into one of men and one of women.
  - What is most important is that the selection of a case in one group has no effect on the probability a case will be chosen in the other group.
    - For example, if I had picked a particular woman, this should not change the probability her brother were then chosen.
- It is important to note that all of these formulas assume that every case was chosen with equal probability.

An example

- Research Question
  - Do Republicans and Democrats have different feelings toward the military?
- Data
  - 2004 NES
  - Feeling thermometer ratings for military
- Research Hypothesis
  - The mean rating for Republicans is different from the mean rating for Democrats.
  - $H_1 : \bar{X}_R \neq \bar{X}_D$
- Null Hypothesis
  - The mean rating for Republicans is the same as the mean rating for Democrats.
  - $H_0 : \bar{X}_R = \bar{X}_D$
- Rejecting the null...
  - If the means are not the same, then groups must be different.

Comparing Republicans and Democrats



Descriptive Statistics

- This is all the information we need to calculate our test statistic.
- Republicans
  - $\bar{X}_R = 86.99$
  - $\hat{\sigma}_R = 14.95$
  - $N = 317$
- Democrats
  - $\bar{X}_D = 76.07$
  - $\hat{\sigma}_D = 23.47$
  - $N = 327$
- What do the standard deviations tell us?

Difference of Means Test

- $t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}}$
- $\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{N_1 - 1} + \frac{\hat{\sigma}_2^2}{N_2 - 1}}$
  - This is a *pooled estimate* because it includes information from both samples.
- If $t = 0$, then $\bar{X}_1 = \bar{X}_2$
  - This is the null hypothesis.

Putting It Together

- $\bar{X}_1 - \bar{X}_2$
  - 86.99-76.07=10.92
- $\sqrt{\frac{\hat{\sigma}_1^2}{N_1 - 1} + \frac{\hat{\sigma}_2^2}{N_2 - 1}}$
  - $\sqrt{\frac{14.95^2}{317-1} + \frac{23.47^2}{327-1}} = 1.55$
- Put it together.
- $t = \frac{10.92}{1.55} = 7.06$

Critical Values

- With a large sample size...
  - .05 level: t(critical)=1.960
  - .10 level: t(critical)=1.645
    - Remember these.
- In this case... $t = 7.06$
- There is a difference in how members of the two parties view the military.

Another Way of Looking at It

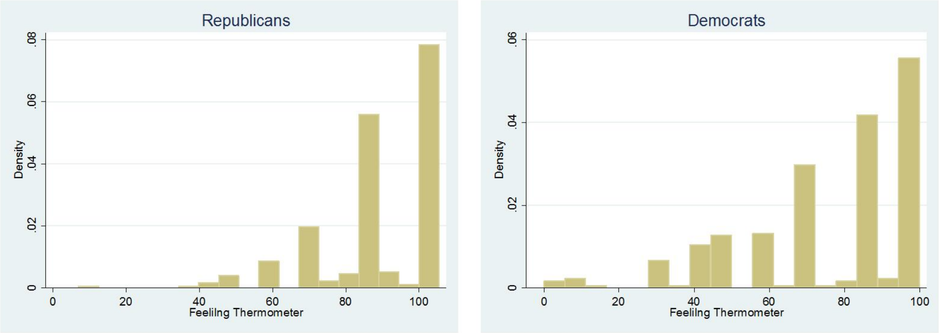
- Calculate the 95% confidence intervals around the mean.
- Republicans
  - $86.99 \pm 1.96 * \frac{14.95}{\sqrt{317}}$
- Democrats
  - $76.07 \pm 1.96 * \frac{23.47}{\sqrt{327}}$

	Republicans	Democrats
Lower Bound	85.35	73.53
Upper Bound	88.64	78.62

- What do we notice about the confidence intervals?

Comparing Republicans and Democrats

- What is causing the difference?



Sample Size of 200

- You are interested in determining whether men and women earn different incomes.
  - You survey 100 men and 100 women.
    - For men, the hourly wage was \$23 with a variance of \$195.
    - For women, the hourly wage was \$20 with a variance of \$216.
- $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1 - 1} + \frac{\hat{\sigma}_2^2}{N_2 - 1}}}$
- Answer: 1.472
  - Is there a statistically significant difference between men and women?

Sample Size of 4,000

- You are interested in determining whether men and women earn different incomes.
  - You survey 2000 men and 2000 women.
    - For men, the hourly wage was \$21 with a variance of \$195.
    - For women, the hourly wage was \$20 with a variance of \$216.
- $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1 - 1} + \frac{\hat{\sigma}_2^2}{N_2 - 1}}}$
- Answer: 2.205
  - Is there a statistically significant difference between men and women?

What About the Last Two Examples

- In the first case, the difference between men and women was \$3.
    - But there wasn't a statistically significant difference.
- In the second case, the difference between men and women was \$1.
    - And there was a statistically significant difference.
- What was different about the two problems?
- Statistically significant vs. substantively significant.