POLI 30: Political Inquiry

Heidi McNamara

University of California, San Diego

h1hall@ucsd.edu

May 6, 2016

Reminders

Midterms handed back at the end of class today.

Email: h1hall@ucsd.edu

Office Hours: Wed. 11-1, SSB 343

Hannah's SPSS OH: Tuesdays 12:30-1:45, Sequoyah 142

Website: heidimcnamara.com

Quiz

What do all normal curves have in common (three things)?

What differentiates them from one another? i.e. How can we tell them apart (two things)?

Why are normal curves (and their particular properties) important?

Normal Curves

What do all normal curves have in common?

Normal Curves

What do all normal curves have in common?

- Symmetric
- Mean = Median
- Bell-shaped, with most of their density in the center and less in the tails

Normal Curves

What do all normal curves have in common?

- Symmetric
- Mean = Median
- Bell-shaped, with most of their density in the center and less in the tails

What differentiates normal curves from one another?

Normal Curves

What do all normal curves have in common?

- Symmetric
- Mean = Median
- Bell-shaped, with most of their density in the center and less in the tails

What differentiates normal curves from one another?

- Mean tells you where it is centered, and differs across curves
- Standard deviation tells you how thick or narrow the curve will be

Standard Deviations vs. Standard Errors

Standard Deviations vs. Standard Errors

Standard Deviations tell us about the level of dispersion *within* a sample—how far apart is each observation? This can help us determine the shape of the normal curve around the sample mean—is it narrow (small standard deviation) or wide (large standard deviation)? It is also used (in conjunction with the sample size (N)) to calculate the *sample error*.

Standard Deviations vs. Standard Errors

Standard Deviations tell us about the level of dispersion *within* a sample—how far apart is each observation? This can help us determine the shape of the normal curve around the sample mean—is it narrow (small standard deviation) or wide (large standard deviation)? It is also used (in conjunction with the sample size (N)) to calculate the *sample error*.

Standard Errors tell us about the level of dispersion *across* samples, if we were to take a bunch of them—how far apart would each sample's mean be from other sample means? This (along with the sample size (N)) can help us determine the margin of error around our sample mean—the larger the standard error, the larger the margin of error will be.

Calculating a Confidence Interval Around a Sample Mean

The 95% confidence interval around a sample mean is:

$$\bar{X} \pm 2 * \frac{\hat{\sigma}}{\sqrt{N}}$$

Where $\hat{\sigma}$ is:

$$\hat{\sigma} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}}$$

Let's start with some hypothetical data

I asked a random sample of 10 (for ease purposes) UCSD students their feelings toward Bernie Sanders on a scale of 1-100. Their answers are below.

Student	Feelings
1	89
2	95
3	40
4	65
5	80
6	72
7	20
8	68
9	99
10	57

Then Calculate the Mean

I asked a random sample of 10 (for ease purposes) UCSD students their feelings toward Bernie Sanders on a scale of 1-100. Their answers are below.

Student	Feelings
1	89
2	95
3	40
4	65
5	80
6	72
7	20
8	68
9	99
10	57
Mean	68.5

Then Take the Difference Between each observation and the mean...

Student	Feelings	$(X_i - \bar{X})$
1	89	20.5
2	95	26.5
3	40	-28.5
4	65	-3.5
5	80	11.5
6	72	3.5
7	20	-48.5
8	68	-0.5
9	99	30.5
10	57	11.5
Mean	68.5	

And then square those...

Student	Feelings	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	89	20.5	420.25
2	95	26.5	702.25
3	40	-28.5	812.25
4	65	-3.5	12.25
5	80	11.5	132.25
6	72	3.5	12.25
7	20	-48.5	2352.25
8	68	-0.5	0.25
9	99	30.5	930.25
10	57	11.5	132.25
Mean	68.5		

And then sum that...

Student	Feelings	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	89	20.5	420.25
2	95	26.5	702.25
3	40	-28.5	812.25
4	65	-3.5	12.25
5	80	11.5	132.25
6	72	3.5	12.25
7	20	-48.5	2352.25
8	68	-0.5	0.25
9	99	30.5	930.25
10	57	11.5	132.25
Mean:	68.5	Sum:	5506.5

And divide by (N-1)...

Student	Feelings	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	89	20.5	420.25
2	95	26.5	702.25
3	40	-28.5	812.25
4	65	-3.5	12.25
5	80	11.5	132.25
6	72	3.5	12.25
7	20	-48.5	2352.25
8	68	-0.5	0.25
9	99	30.5	930.25
10	57	11.5	132.25
Mean:	68.5	Sum:	5506.5

$$\frac{5506.5}{(10-1)} = 611.8333$$

And then take the square root...

Student	Feelings	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	89	20.5	420.25
2	95	26.5	702.25
3	40	-28.5	812.25
4	65	-3.5	12.25
5	80	11.5	132.25
6	72	3.5	12.25
7	20	-48.5	2352.25
8	68	-0.5	0.25
9	99	30.5	930.25
10	57	11.5	132.25
Mean:	68.5	Sum:	5506.5

$$\frac{5506.5}{(10-1)} = 611.8333$$

$$\hat{\sigma} = \sqrt{611.8333} = 24.735$$

Now to use the $\hat{\sigma}$ to calculate the confidence interval...

Formula reminder for the margin of error:

$$\bar{X} \pm 2 * \frac{\hat{\sigma}}{\sqrt{N}}$$

If we plug in everything we now know:

$$68.5 \pm 2 * \frac{24.735}{\sqrt{10}} =$$

$$68.5 \pm 2 * 7.82 =$$

$$68.5 \pm 15.64$$

Now to use the $\hat{\sigma}$ to calculate the confidence interval...

Formula reminder for the margin of error:

$$\bar{X} \pm 2 * \frac{\hat{\sigma}}{\sqrt{N}}$$

If we plug in everything we now know:

$$68.5 \pm 2 * \frac{24.735}{\sqrt{10}} =$$

$$68.5 \pm 2 * 7.82 =$$

$$68.5 \pm 15.64$$

We can say with 95% certainty that the true population mean for how UCSD students feel about Bernie Sanders is within this range: **[52.86, 84.14]**.

What if we wanted to be 99.7% sure?

What if we wanted to be 99.7% sure?

Then our formula for the margin of error would be:

$$\bar{X} \pm 3 * \frac{\hat{\sigma}}{\sqrt{N}}$$

If we plug in everything we now know:

$$68.5 \pm 3 * \frac{24.735}{\sqrt{10}} =$$

$$68.5 \pm 3 * 7.82 =$$

$$68.5 \pm 23.46$$

What if we wanted to be 99.7% sure?

Then our formula for the margin of error would be:

$$\bar{X} \pm 3 * \frac{\hat{\sigma}}{\sqrt{N}}$$

If we plug in everything we now know:

$$68.5 \pm 3 * \frac{24.735}{\sqrt{10}} =$$

$$68.5 \pm 3 * 7.82 =$$

$$68.5 \pm 23.46$$

We can say with 99.7% certainty that the true population mean for how UCSD students feel about Bernie Sanders is within this range: **[45.04, 91.96]**

Calculating a Confidence Interval Around a Sample Proportion

The 95% confidence interval around a sample proportion is:

$$\hat{p} \pm 2 * \frac{\sqrt{(\hat{p})(1-\hat{p})}}{\sqrt{N}}$$

Let's begin with more hypothetical data...

Suppose I asked a random sample of 10 (for ease) UCSD students if they had fun at Sun God last weekend. Their answers are below. "Yes" is notated with a "1", "No" with a "0". Working alone to begin with, and then in groups if/when you run into trouble, calculate the 95% confidence interval around this sample's proportion of Sun God enjoyment.

Student	Feelings
1	1
2	0
3	0
4	1
5	0
6	0
7	0
8	1
9	1
10	0

Then we calculate the proportion who said "Yes"....

Suppose I asked a random sample of 10 (for ease) UCSD students if they had fun at Sun God last weekend. Their answers are below. "Yes" is notated with a "1", "No" with a "0".

Student	Feelings
1	1
2	0
3	0
4	1
5	0
6	0
7	0
8	1
9	1
10	0
Proportion "yes"	$\frac{4}{10} = 0.4$

Now to calculate the margin of error...

Formula for the 95% CI around a sample proportion:

$$\hat{p} \pm 2 * \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Now to calculate the margin of error...

Formula for the 95% CI around a sample proportion:

$$\hat{p} \pm 2 * \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Plugging in what we know:

$$0.4 \pm 2 * \sqrt{\frac{0.4(1-0.4)}{10}} =$$

$$0.4 \pm 2 * \sqrt{\frac{0.24}{10}} =$$

$$0.4 \pm 2 * \sqrt{0.024} =$$

$$0.4 \pm 2 * 0.15 =$$

$$0.4 \pm 0.30$$

Now to calculate the margin of error...

Formula for the 95% CI around a sample proportion:

$$\hat{p} \pm 2 * \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Plugging in what we know:

$$0.4 \pm 2 * \sqrt{\frac{0.4(1-0.4)}{10}} =$$

$$0.4 \pm 2 * \sqrt{\frac{0.24}{10}} =$$

$$0.4 \pm 2 * \sqrt{0.024} =$$

$$0.4 \pm 2 * 0.15 =$$

$$0.4 \pm 0.30$$

Thus, we can say with 95% certainty that the true proportion of UCSD students who enjoyed Sun God is somewhere between 10% and 70%.

Null Hypotheses

A **Null Hypothesis** (H_0) is the hypothesis that there is no significant difference between two groups and that any observed difference is due to sampling or experimental error.

For instance, we could examine the rate of support for keeping troops in Iraq between Republicans and Democrats (hypothetical data). We pull a random sample of 1500 Americans and find that 70% of Republicans support maintaining a ground force in Iraq, while only 40% of Democrats do. By random chance, we sampled 800 Republicans, and 700 Democrats.

What would our null hypothesis? Alternate hypothesis?

What we observe:

$$p_{Rep} = 70$$
$$p_{Dem} = 40$$
$$p_{Rep} - p_{Dem} = 30$$

What would our null hypothesis? Alternate hypothesis?

What we observe:

$$p_{Rep} = 70$$
$$p_{Dem} = 40$$
$$p_{Rep} - p_{Dem} = 30$$

What would our null hypothesis be?

What would our null hypothesis? Alternate hypothesis?

What we observe:

$$p_{Rep} = 70$$
$$p_{Dem} = 40$$
$$p_{Rep} - p_{Dem} = 30$$

What would our null hypothesis be?

$$H_0 : p_{Rep} - p_{Dem} = 0$$

What would our null hypothesis? Alternate hypothesis?

What we observe:

$$p_{Rep} = 70$$
$$p_{Dem} = 40$$
$$p_{Rep} - p_{Dem} = 30$$

What would our null hypothesis be?

$$H_0 : p_{Rep} - p_{Dem} = 0$$

What would our alternate hypothesis be?

What would our null hypothesis? Alternate hypothesis?

What we observe:

$$p_{Rep} = 70$$
$$p_{Dem} = 40$$
$$p_{Rep} - p_{Dem} = 30$$

What would our null hypothesis be?

$$H_0 : p_{Rep} - p_{Dem} = 0$$

What would our alternate hypothesis be?

$$H_A : p_{Rep} - p_{Dem} \neq 0$$
$$H_A : p_{Rep} - p_{Dem} > 0$$

Difference of Proportions Test

To test our null hypothesis, we need to conduct a difference of proportions test. To do this, we need to calculate the confidence interval around this *difference of proportions*.

$$(\hat{p}_{Rep} - \hat{p}_{Dem}) \pm 2 * \sqrt{\left(\sqrt{\frac{p_{Rep}(1-p_{Rep})}{N_{Rep}}}\right)^2 + \left(\sqrt{\frac{p_{Dem}(1-p_{Dem})}{N_{Dem}}}\right)^2}$$

This looks horrible, but I promise it isn't.

Let's break it down... First calculate the standard error for the Republican proportion

Formula:

$$\sqrt{\frac{p_{Rep}(1-p_{Rep})}{N_{Rep}}}$$

With our numbers:

$$\sqrt{\frac{0.7(1-0.7)}{800}} = 0.016$$

Now for the Democratic proportion...

Formula:

$$\sqrt{\frac{p_{Dem}(1-p_{Dem})}{N_{Dem}}}$$

With our numbers:

$$\sqrt{\frac{0.4(1-0.4)}{700}} = 0.018$$

Now take these standard errors and plug them back into the confidence interval formula...

Formula:

$$(\hat{p}_{Rep} - \hat{p}_{Dem}) \pm 2 * \sqrt{(\sqrt{\frac{p_{Rep}(1-p_{Rep})}{N_{Rep}}})^2 + (\sqrt{\frac{p_{Dem}(1-p_{Dem})}{N_{Dem}}})^2}$$

With our numbers:

$$(0.7 - 0.4) \pm 2 * \sqrt{(0.016)^2 + (0.018)^2} =$$
$$(0.3) \pm 2 * 0.024 =$$
$$(0.3) \pm 0.048$$

Now take these standard errors and plug them back into the confidence interval formula...

Formula:

$$(\hat{p}_{Rep} - \hat{p}_{Dem}) \pm 2 * \sqrt{(\sqrt{\frac{p_{Rep}(1-p_{Rep})}{N_{Rep}}})^2 + (\sqrt{\frac{p_{Dem}(1-p_{Dem})}{N_{Dem}}})^2}$$

With our numbers:

$$(0.7 - 0.4) \pm 2 * \sqrt{(0.016)^2 + (0.018)^2} =$$
$$(0.3) \pm 2 * 0.024 =$$
$$(0.3) \pm 0.048$$

So our confidence interval is: [0.252, 0.348].

So... can we reject the null hypothesis?

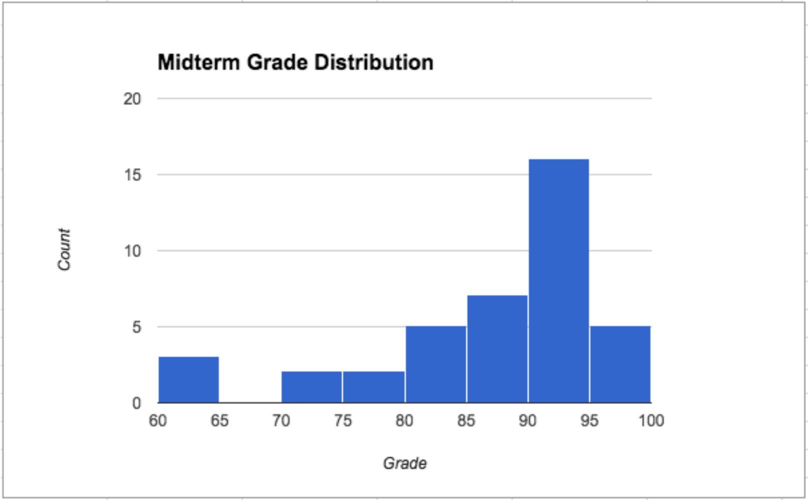
So... can we reject the null hypothesis?

Yes! Our confidence interval does not include 0, so we can reject the null hypothesis that the difference in our sample proportions is due to chance with 95% certainty.

Midterm Distribution

- Median: 90
- Mean: 86
- Standard Deviation: 9

Midterm Distribution



Median: 90, Mean: 86, Standard Deviation: 9