Poli 30D Political Inquiry
Normal Curve & Confidence Intervals

Shane Xinyang Xuan
ShaneXuan.com

November 17, 2016

Contact Information

Shane Xinyang Xuan
xxuan@ucsd.edu

We have someone to help you every day!

Professor Desposato	M	1330-1500 (Latin American Center)
Shane Xuan	Tu	1600-1800 (SSB332)
Cameron Sells	W	1000-1200 (SSB352)
Kelly Matush	Th	1500-1700 (SSB343)
Julia Clark	F	1200-1400 (SSB326)

Supplemental Materials

Our class oriented

ShaneXuan.com

UCLA SPSS starter kit

www.ats.ucla.edu/stat/spss/sk/modules_sk.htm

Princeton data analysis

http://dss.princeton.edu/training/

Road Map

This is our second last section! We are going to cover

- Normal curve
- Confidence interval

in today's section!

Normal Curve

The standard score of a raw score x is calculated by

$$z = \frac{x - \mu}{\sigma}$$

where μ is the population mean and σ is the population standard deviation.

Normal Curve

Now, say that I collect a sample from the population, calculate the mean, and want to know how far away the sample mean is from the population mean, I need to calculate the z-score first

$$z = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

We have $s/\sqrt{n}$ in the bottom because

$$\text{var}(\overline{X}) = \frac{s^2}{n}$$

We updated our z-score formula with the standard error of the mean because we are dealing with a sample! This z-score tells us how many standard errors there are between the sample mean and the population mean.

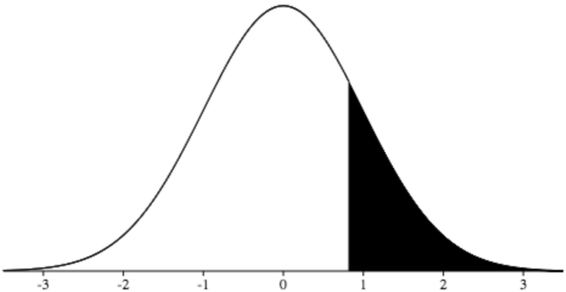
Normal Curve

We first look at (part of) a z-table:

First digit and first decimal of Z	Second decimal of Z									
	.00	*.01*	*.02*	*.03*	*.04*	*.05*	*.06*	*.07*	*.08*	*.09*
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Normal Curve

The corresponding figure to the z-table you just saw is



– Shaded on the right hand side
– Zero in the middle; positive on the right; negative on the left
– You might also see a z-table with a corresponding figure that is shaded on the left hand side in the exam. We will discuss how to deal with these tables as well.

How to read the z-table?

When in doubt, draw in out![1] Let's draw the following examples on board and solve them together!

- Read from table

$$P(z > 0.82) = 0.2061$$

[1]Quote from Dr. Ethan Hollander, Associate Professor of Political Science at Wabash College, who is also a UCSD alumnus!

How to read the z-table?

When in doubt, draw in out![1] Let's draw the following examples on board and solve them together!

- Read from table

$$P(z > 0.82) = 0.2061$$

- Minor calculation

$$P(z < 0.10) = 1 - 0.4602 = 0.539828$$

[1] Quote from Dr. Ethan Hollander, Associate Professor of Political Science at Wabash College, who is also a UCSD alumnus!

How to read the z-table?

When in doubt, draw in out![1] Let's draw the following examples on board and solve them together!

- Read from table

$$P(z > 0.82) = 0.2061$$

- Minor calculation

$$P(z < 0.10) = 1 - 0.4602 = 0.539828$$

- Some conceptual 'tweaks'

$$P(z < -0.10) = 0.4602$$

[1] Quote from Dr. Ethan Hollander, Associate Professor of Political Science at Wabash College, who is also a UCSD alumnus!

How to read the z-table?

When in doubt, draw in out![1] Let's draw the following examples on board and solve them together!

- Read from table

$$P(z > 0.82) = 0.2061$$

- Minor calculation

$$P(z < 0.10) = 1 - 0.4602 = 0.539828$$

- Some conceptual 'tweaks'

$$P(z < -0.10) = 0.4602$$

- Extra calculation

$$P(-0.70 < z < 0.80) = (1 - 0.2119) - 0.2420 = 0.5461$$

[1] Quote from Dr. Ethan Hollander, Associate Professor of Political Science at Wabash College, who is also a UCSD alumnus!

Confidence Interval

95% of the area under the normal distribution lies within 1.96 standard deviations of the mean. That is,

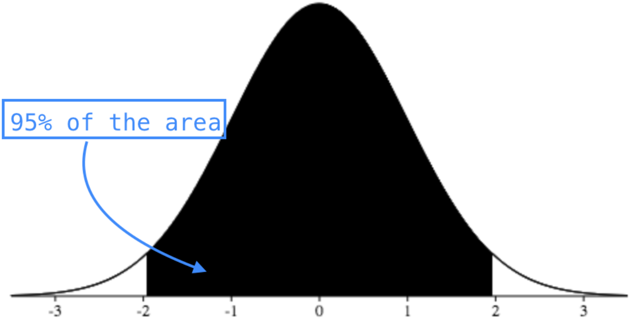
$$P(-1.96 < z < 1.96) = 95\%$$

Confidence Interval

95% of the area under the normal distribution lies within 1.96 standard deviations of the mean. That is,

$$P(-1.96 < z < 1.96) = 95\%$$

When in doubt, draw it out!



Announcements

- I will not be here on 11/30. Cameron Sells will cover for me. He will also give you a quiz in section.
- Today's section is the last section before HW4 is due. You should start early, and take use of my office hours if you have questions.
- I will try to hand HW back to you before Thanksgiving after lectures. Make sure you go to lectures if you want feedback on HW2 before Thanksgiving.

Quiz: confidence interval

68% of normally distributed data is within one standard deviation of the mean. Show me why.

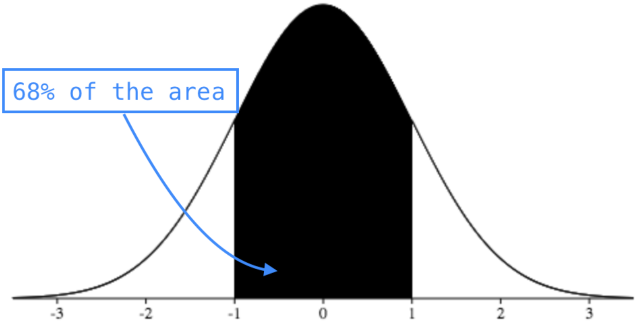
Write down you name and email address on the quiz.

Quiz: confidence interval

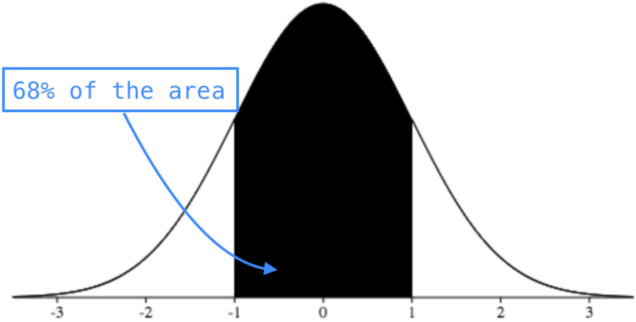
68% of normally distributed data is within one standard deviation of the mean. Show me why.

Write down you name and email address on the quiz.

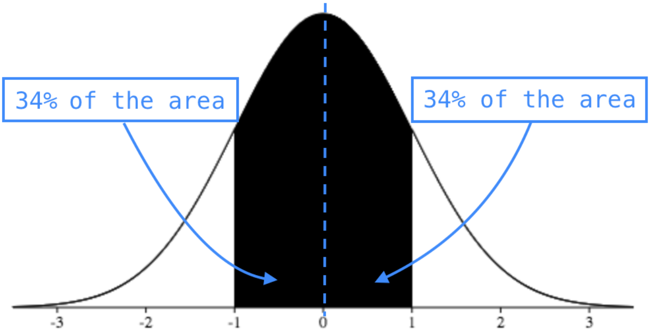
Solution:
Here is the figure that you should have in mind (or written down) when you are solving this problem!



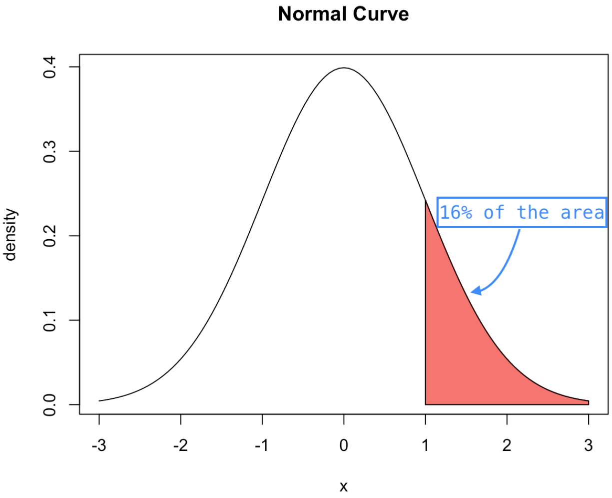
z-score for 68% CI



z-score for 68% CI



z-score for 68% CI



Remember that you are reading the area that is red!

$$P(z > 1) = 50\% - 34\% = 16\%$$

z-score for 68% CI

First digit and first decimal of Z	Second decimal of Z									
	.00	*.01*	*.02*	*.03*	*.04*	*.05*	*.06*	*.07*	*.08*	*.09*
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

$$P(z > 1) = 16\% \rightsquigarrow z \approx 1.0$$

Confidence Interval: Sample Mean

Confidence intervals provide more information than point estimates. The 95% confidence interval of the sample mean will contain the population mean 95% of the time.

Confidence Interval: Sample Mean

Confidence intervals provide more information than point estimates. The 95% confidence interval of the sample mean will contain the population mean 95% of the time.

To calculate the confidence interval of the mean, here is the formula that we will be using

$$\overline{X} \pm z * \underbrace{\frac{s}{\sqrt{n}}}_{\text{s.e.}}$$

where z is the statistic for the selected confidence level, and s is the standard deviation of the sample. Also, commonly used confidence level includes 68% ($z = 1$), 90% ($z = 1.645$), and 95% (z=1.96).

Confidence Interval: Sample Mean

Suppose the sample mean is 127. The sample standard deviation is 19. The sample size is 1000. Please write down the 95% confidence interval.

Confidence Interval: Sample Mean

Suppose the sample mean is 127. The sample standard deviation is 19. The sample size is 1000. Please write down the 95% confidence interval.

The solution is

$$127 \pm 1.96 * \frac{19}{\sqrt{1000}}$$

Confidence Interval: Sample Mean

Suppose the sample mean is 127. The sample standard deviation is 19. The sample size is 1000. Please write down the 95% confidence interval.

The solution is

$$127 \pm 1.96 * \frac{19}{\sqrt{1000}}$$

Note that standard deviation is calculated in the following way

$$s = \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{n - 1}}$$

because we are dealing with the sample!

Confidence Interval: Sample Proportion

To calculate the confidence interval for a percentage, here is the formula that we will be using

$$\hat{p} \pm z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval: Sample Proportion

To calculate the confidence interval for a percentage, here is the formula that we will be using

$$\hat{p} \pm z * \underbrace{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{standard error}}$$

Difference between means/proportions

Standard error

$$se_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{1}$$

$$se_{\hat{p_1} - \hat{p_2}} = \sqrt{\frac{\hat{p_1}(1 - \hat{p_1})}{n_1} + \frac{\hat{p_2}(1 - \hat{p_2})}{n_2}} \tag{2}$$

Confidence interval

$$(\bar{X}_1 - \bar{X}_2) \pm z * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{3}$$

$$(\hat{p_1} - \hat{p_2}) \pm z * \sqrt{\frac{\hat{p_1}(1 - \hat{p_1})}{n_1} + \frac{\hat{p_2}(1 - \hat{p_2})}{n_2}} \tag{4}$$

Difference between means/proportions

Standard error

$$se_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (1)$$

$$se_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (2)$$

Confidence interval

$$(\bar{X}_1 - \bar{X}_2) \pm z * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3)$$

$$(\hat{p}_1 - \hat{p}_2) \pm z * \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (4)$$

Difference between means/proportions

Standard error

$$se_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (1)$$

$$se_{\hat{p_1} - \hat{p_2}} = \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} \quad (2)$$

Confidence interval

$$(\bar{X}_1 - \bar{X}_2) \pm z * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3)$$

$$(\hat{p_1} - \hat{p_2}) \pm z * \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} \quad (4)$$

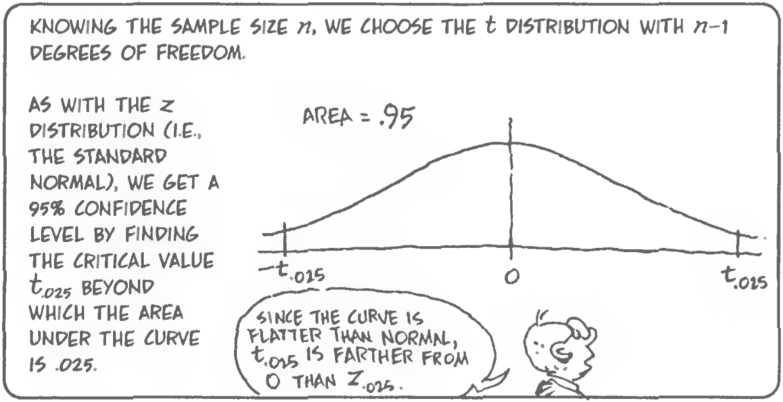
Good news: you don't need to know this for this class.

t-distribution

What we have been talking about relies on a strong assumption: the data follow a normal curve. However, this is not necessarily the case all the time. Sometime, we need to use a t-distribution. The rule of thumb is that: use z-table when $n \geq 30$, and t-table when $n < 30$. The t-statistics is calculated by

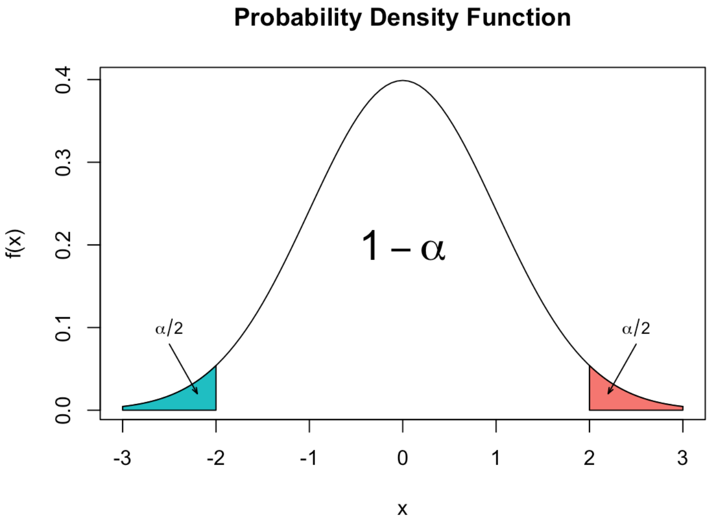
$$t_{\overline{X}} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

t-distribution



The cartoon guide to statistics (Larry Gonick)

t-distribution



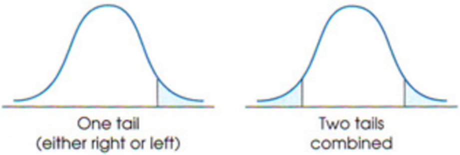
t-distribution

FOR A $(1-\alpha)\cdot 100\%$ CONFIDENCE INTERVAL, WE FIND THE CRITICAL VALUE $t_{\frac{\alpha}{2}}$ SUCH THAT $Pr(t \geq t_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. HERE IS A SHORT TABLE OF CRITICAL VALUES FOR THE t DISTRIBUTION:

	$1-\alpha$	.80	.90	.95	.99
	α	.20	.10	.05	.01
	$\alpha/2$	.10	.05	.025	.005
DEGREES OF FREEDOM	1	3.09	6.31	12.71	63.66
	10	1.37	1.81	2.23	4.14
	30	1.31	1.70	2.04	2.75
	100	1.29	1.66	1.98	2.63
	∞	1.28	1.65	1.96	2.58

The cartoon guide to statistics (Larry Gonick)

t-distribution



	PROPORTION IN ONE TAIL					
	0.25	0.10	0.05	0.025	0.01	0.005
	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055