GPCO 453: Quantitative Methods I

Sec 07: Samples

Shane Xinyang Xuan[1]
ShaneXuan.com

November 13, 2017

[1]Department of Political Science, UC San Diego, 9500 Gilman Drive #0521.

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Garg	Tu	1300-1500 (RBC 1303)
Shane Xuan	M	1100-1200 (SSB 332)
	M	1530-1630 (SSB 332)
Joanna Valle-luna	Tu	1700-1800 (RBC 3131)
	Th	1300-1400 (RBC 3131)
Daniel Rust	F	1100-1230 (RBC 3213)

Roadmap

In this section, we cover the basics for probability:

- Sample proportion

Roadmap

In this section, we cover the basics for probability:

- Sample proportion
- Sample mean, known σ

Roadmap

In this section, we cover the basics for probability:

- Sample proportion
- Sample mean, known σ
- Sample mean, unknown σ

Standard Errors

Statistic	Population	Sample
Mean	μ	$\overline{x}$
Proportion	p	$\hat{p}$
Standard Deviaion	σ	s

Standard Errors

Statistic	Population	Sample
Mean	μ	$\overline{x}$
Proportion	p	$\hat{p}$
Standard Deviaion	σ	s

- Standard error is the standard deviation of $\overline{x}$ or $\hat{p}$

Standard Errors

Statistic	Population	Sample
Mean	μ	$\overline{x}$
Proportion	p	$\hat{p}$
Standard Deviaion	σ	s

- Standard error is the standard deviation of $\overline{x}$ or $\hat{p}$
- Standard deviation of $\bar{x}$ is $\sigma_{\overline{x}}$:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \tag{1}$$

Standard Errors

Statistic	Population	Sample
Mean	μ	$\overline{x}$
Proportion	p	$\hat{p}$
Standard Deviaion	σ	s

- Standard error is the standard deviation of $\overline{x}$ or $\hat{p}$
- Standard deviation of $\bar{x}$ is $\sigma_{\overline{x}}$:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \tag{1}$$

- Standard deviation of $\hat{p}$ is $\sigma_{\hat{p}}$:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{2}$$

Example: Sampling distribution of $\hat{p}$

- Assume that the population proportion is $p = 0.76$, and the sample of 400 people is selected from the population

Example: Sampling distribution of $\hat{p}$

- Assume that the population proportion is $p = 0.76$, and the sample of 400 people is selected from the population
- Show the sampling distribution of sample proportion $\hat{p}$

Example: Sampling distribution of $\hat{p}$

- Assume that the population proportion is $p = 0.76$, and the sample of 400 people is selected from the population
- Show the sampling distribution of sample proportion $\hat{p}$
- The sampling distribution of $\hat{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $np(1-p) \geq 5$

Example: Sampling distribution of $\hat{p}$

- Assume that the population proportion is $p = 0.76$, and the sample of 400 people is selected from the population
- Show the sampling distribution of sample proportion $\hat{p}$
- The sampling distribution of $\hat{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $np(1-p) \geq 5$
- We first calculate that

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

$$= \sqrt{\frac{0.76(1-0.76)}{400}} = 0.0214 \tag{4}$$

Example: Sampling distribution of $\hat{p}$

- Assume that the population proportion is $p = 0.76$, and the sample of 400 people is selected from the population
- Show the sampling distribution of sample proportion $\hat{p}$
- The sampling distribution of $\hat{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $np(1-p) \geq 5$
- We first calculate that

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

$$= \sqrt{\frac{0.76(1-0.76)}{400}} = 0.0214 \tag{4}$$

- The distribution is hence $\hat{p} \sim \mathcal{N}(0.76, 0.0214^2)$

Sample Mean with known σ

- 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.

Sample Mean with known σ

- 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:

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

Sample Mean with known σ

- 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:

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

  - z is the statistic for the selected confidence level

Sample Mean with known σ

- 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:

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

  - z is the statistic for the selected confidence level
  - σ is the population standard deviation (if unknown, use s)

Sample Mean with known σ

- 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:

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

  - z is the statistic for the selected confidence level
  - σ is the population standard deviation (if unknown, use s)
  - $\frac{\sigma}{\sqrt{n}}$ is the standard error (that is, standard deviation of $\overline{x}$)

Sample Mean with known σ

- 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:

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

  - z is the statistic for the selected confidence level
  - σ is the population standard deviation (if unknown, use s)
  - $\frac{\sigma}{\sqrt{n}}$ is the standard error (that is, standard deviation of $\overline{x}$)
  - $z\frac{\sigma}{\sqrt{n}}$ is the margin of error

Sample Mean with known σ

- 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:

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

  - z is the statistic for the selected confidence level
  - σ is the population standard deviation (if unknown, use s)
  - $\frac{\sigma}{\sqrt{n}}$ is the standard error (that is, standard deviation of $\overline{x}$)
  - $z\frac{\sigma}{\sqrt{n}}$ is the margin of error
- Commonly used confidence level includes 68% ($z = 1$), 90% ($z = 1.645$), and 95% ($z = 1.96$).

Example: Sample Mean with known σ

- Suppose the population standard deviation is 19. I selected a random sample. The sample size is 1000, and the sample mean is 127. Please write down the 95% confidence interval.

Example: Sample Mean with known σ

- Suppose the population standard deviation is 19. I selected a random sample. The sample size is 1000, and the sample mean is 127. Please write down the 95% confidence interval.
- The solution is

$$\overline{x} \pm z * \frac{\sigma}{\sqrt{n}} \tag{5}$$

$$= 127 \pm 1.96 * \frac{19}{\sqrt{1000}} \tag{6}$$

$$= (125.82, 128.18)$$

What if we do not know σ?

- There are cases in which we do not know the population standard deviation σ. We use sample standard deviation s to estimate σ.

What if we do not know σ?

- There are cases in which we do not know the population standard deviation σ. We use sample standard deviation s to estimate σ.
- A little tweak on the formula

$$\overline{x} \pm z * \frac{\sigma}{\sqrt{n}} \quad \Longrightarrow \quad \overline{x} \pm t * \frac{s}{\sqrt{n}}$$

What if we do not know σ?

- There are cases in which we do not know the population standard deviation σ. We use sample standard deviation s to estimate σ.
- A little tweak on the formula

$$\overline{x} \pm z * \frac{\sigma}{\sqrt{n}} \quad \Longrightarrow \quad \overline{x} \pm t * \frac{s}{\sqrt{n}}$$

  - Use t-score instead of z-score

What if we do not know σ?

- There are cases in which we do not know the population standard deviation σ. We use sample standard deviation s to estimate σ.
- A little tweak on the formula

$$\overline{x} \pm z * \frac{\sigma}{\sqrt{n}} \quad \Longrightarrow \quad \overline{x} \pm t * \frac{s}{\sqrt{n}}$$

  - Use t-score instead of z-score
  - Use s instead of σ

What if we do not know σ?

- There are cases in which we do not know the population standard deviation σ. We use sample standard deviation s to estimate σ.
- A little tweak on the formula

$$\overline{x} \pm z * \frac{\sigma}{\sqrt{n}} \quad \Longrightarrow \quad \overline{x} \pm t * \frac{s}{\sqrt{n}}$$

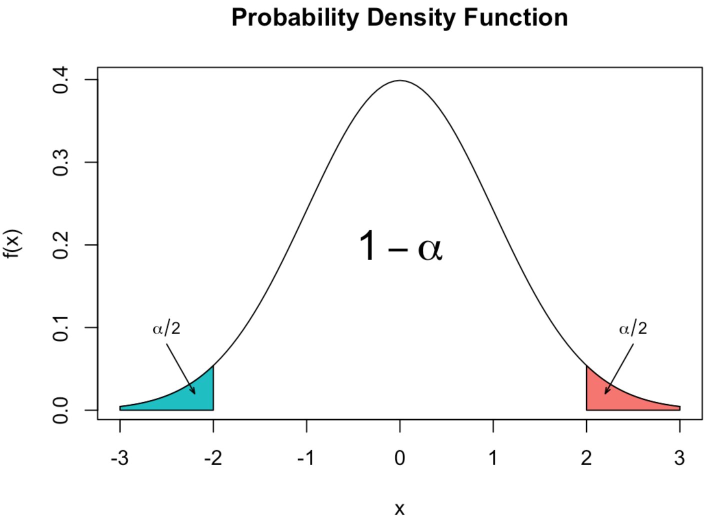
  - Use t-score instead of z-score
  - Use s instead of σ
- That is, in the case in which σ is unknown, we can bootstrap σ with s, as long as we use t instead of z

t-distribution

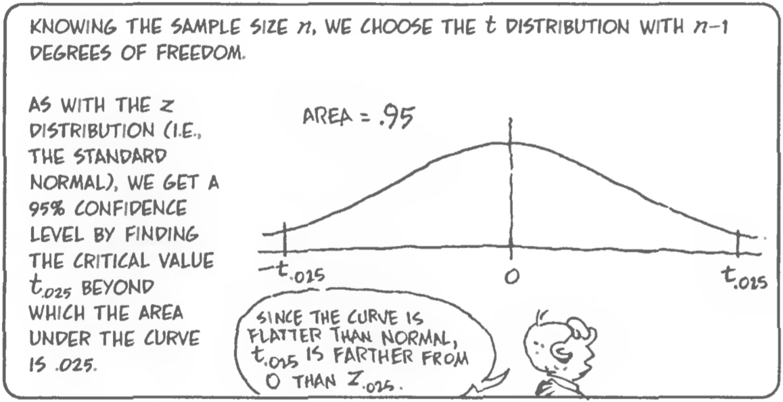
- (Rule of thumb) We use z-table when $n \geq 30$, and t-table when $n < 30$. The t-statistics is calculated by

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

t-distribution



t-distribution



The cartoon guide to statistics (Larry Gonick)

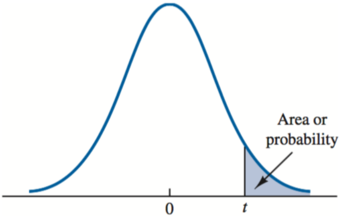
t-distribution

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



Degrees of Freedom	Area in Upper Tail .20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250

t-distribution

Table: When t-table only gives you information about Area in Upper Tail

$\frac{\alpha}{2}$	α	$1-\alpha$
0.05	0.1	90%
0.025	0.05	95%
0.01	0.02	98%
0.005	0.01	99%