Estimating mean and proportion

Nikos Apostolakis

April 10, 2024

1 Confidence Intervals

We won't need to do that very often but if you are given a confidence level c and you want to find the corresponding critical value z_c you need to find what right tail $\alpha/2$ corresponds to c. In general c is the area in the middle, α is the area of both tails, and $\alpha/2$ is the area of only one of the tails. This is shown in Figure 1 for confidence $c = 90\%$. Then $\alpha = 10\% = 0.1$ and $\alpha/2 = 5\% = 0.05$. So we would need to find 0.05000 inside the z–table.

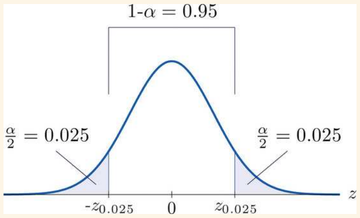

Figure 1: The relation between α and c.

But we rarely do that since we have the inverse tables. In the t–tables the z–values correspond to infinite many degrees of freedom $\nu = \infty$. For convenience, here is the table of critical values of z.

α	$\frac{\alpha}{2}$	c	z_c
0.25	0.125	75%	1.15
0.20	0.100	80%	1.28
0.15	0.750	85%	1.44
0.10	0.050	90%	1.645
0.05	0.025	95%	1.96
0.02	0.010	98%	2.33
0.01	0.005	99%	2.58

1.1 Estimating μ.

The confidence interval is $\bar{x} \pm E$, that is

$$\bar{x} - E \leq \mu \leq \bar{x} + E$$

where how we compute the *standard error* E depends on whether σ is known, and whether we have a large sample.

Confidence interval for μ

For large sample ($n \geq 30$),

$$E = \begin{cases} z_c \dfrac{\sigma}{\sqrt{n}} & \sigma \text{ is known} \\ \\ z_c \dfrac{s}{\sqrt{n}} & \sigma \text{ is unknown.} \end{cases}$$

For small sample ($n < 30$) we use critical values of the t–distribution with $n-1$ degrees of freedom.

$$E = \begin{cases} t_c \dfrac{\sigma}{\sqrt{n}} & \sigma \text{ is known} \\ \\ t_c \dfrac{s}{\sqrt{n}} & \sigma \text{ is unknown.} \end{cases}$$

1.2 Estimating p.

When is a sample large enough for estimating p

We need to have that there are at least 5 observations that fit the characteristic we and at least 5 observations that don't fit:

$$n\widehat{p} > 5$$
$$n\widehat{q} > 5.$$

The confidence interval is $\widehat{p} \pm E$, that is

$$\widehat{p} - E \leq p \leq \widehat{p} + E$$

where

Confidence interval for p

For large sample ($n \geq 30$),

$$E = z_c \sqrt{\frac{\widehat{p}\,\widehat{q}}{n}}.$$

2 Exercises

1. The amount of a particular biochemical substance related to bone breakdown was measured in 30 healthy women. The sample mean and standard deviation were 3.3 nanograms per milliliter (ng/mL) and 1.4 ng/mL. Construct an 80% confidence interval for the mean level of this substance in all healthy women.

2. In order to estimate the mean FICO credit score of its members, a credit union samples the scores of 95 members, and obtains a sample mean of 738.2 with sample standard deviation 64.2. Construct a 99% confidence interval for the mean FICO score of all of its members.

3. For all settings a packing machine delivers a precise amount of liquid; the amount dispensed always has standard deviation 0.07 ounce. To calibrate the machine its setting is fixed and it is operated 50 times. The mean amount delivered is 6.02 ounces with sample standard deviation 0.04 ounce. Construct a 99.5% confidence interval for the mean amount delivered at this setting.

   Hint. Not all the information provided is needed. You have to identify the relevant information.

4. To test a new tread design with respect to stopping distance, a tire manufacturer manufactures a set of prototype tires and measures the stopping distance from 70 mph on a standard test car. A sample of 25 stopping distances yielded a sample mean 173 feet with sample standard deviation 8 feet. Construct a 98% confidence interval for the mean stopping distance for these tires. Assume a normal distribution of stopping distances.

5. A sample of 26 women's size 6 dresses had mean waist measurement 25.25 inches with sample standard deviation 0.375 inch. Construct a 95% confidence interval for the mean waist measurement of all size 6 women's dresses. Assume waist measurements are normally distributed.

6. Nutritionists are investigating the efficacy of a diet plan designed to increase the caloric intake of elderly people. The increase in daily caloric intake in 12 individuals who are put on the plan is (a minus sign signifies that calories consumed went down):

121	284	−94	295	183	312
188	−102	259	226	152	167

   Construct a 99.8% confidence interval for the mean increase in caloric intake for all people who are put on this diet. Assume that population of differences in intake is normally distributed.

   Hint. You need first to calculate the mean $\bar{x}$ and standard deviation s of the sample. Remember how to do that?

7. In a random sample of 2,300 mortgages taken out in a certain region last year, 8% were adjustable-rate mortgages.
   (a) Verify that this is a large enough sample.
   (b) Construct a 99% confidence interval for the proportion of all mortgages taken out in this region last year that were adjustable-rate mortgages.

8. In a research study in cattle breeding, 159 of 273 cows in several herds that were in estrus were detected by means of an intensive once a day, one-hour observation of the herds in early morning.
   (a) Verify that this is a large enough sample.
   (b) C Assuming that the sample is sufficiently large, construct a 90% confidence interval for the proportion of all cattle in estrus who are detected by this method.

9. A survey of 50 randomly selected adults in a small town asked them if their opinion on a proposed "no cruising" restriction late at night. Responses were coded 1 for in favor, 0 for indifferent, and 2 for

opposed, with the results shown in the table.

1	0	2	0	1	0	0	1	1	2
0	2	0	0	0	1	0	2	0	0
0	2	1	2	0	0	0	2	0	1
0	2	0	2	0	1	0	0	2	0
1	0	0	1	2	0	0	2	1	2

(a) Verify that this is a large enough sample.

(b) Construct a 90% confidence interval for the proportion of all adults in the community who are indifferent concerning the proposed restriction.