Estimation

Gonzalo G. Peraza Mues

March 24, 2017

A recent poll of 1500 randomly chosen people indicated that 22 percent of the entire population is presently dieting, with a margin of error of ± 2 percent.

- What does "margin of error" mean?
- How can the proportion of a population in the millions can be ascertained by sampling only 1500 people?

Estimators

When the distribution of the population is known but one or some of its parameters are missing one can use the results from a sample of the population to estimate these unknown parameters.

Definition: An estimator is a statistic whose value depends on the particular sample drawn. The value of the estimator, called the estimate, is used to predict the value of a population parameter.

Point estimators

- The sample mean $\bar{X}$.
- The population proportion $\hat{p}$.
- The population variance S^2.

Definition: An estimator whose expected value is equal to the parameter it is estimating is said to be an unbiased estimator of that parameter.

Point Estimator of the Population Mean.

- The estimator of the population mean μ is the sample mean $\bar{X}$.
- It is an unbiased estimator; $\mathrm{E}\left[\bar{X}\right] = \mu$.
- The standard error is $SEM = \sigma/\sqrt{n}$.

The estimate of the population mean will be correct to within ± 2 standard errors. To cut the standard error in half, we must increase the sample size by a factor of 4.

Point Estimator of a Population Proportion

- The estimator of the population proportion p is the sample proportion $\hat{p}$.
- It is an unbiased estimator; $\mathrm{E}[\hat{p}] = p$.
- The standard error is $SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$.

It can be shown that $p(1-p) \leq \frac{1}{4}$ which in turn means that $SD(\hat{p}) \leq \frac{1}{2\sqrt{n}}$.

Estimating a Population Variance

- The estimator of the population variance σ^2 is:
  - $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ if μ is known.
  - S^2 if μ is unknown.
- S^2 is an unbiased estimator; $\mathrm{E}\left[S^2\right] = \sigma^2$.

The population standard deviation σ is estimated by S.

Interval Estimators of the Mean with Known Population Variance

Definition: An interval estimator of a population parameter is an interval that is predicted to contain the parameter. The confidence we ascribe to the interval is the probability that it will contain the parameter.

To determine an interval estimator of a population parameter, we use the prob- ability distribution of the point estimator of that parameter.

Interval Estimators of the Mean with Known Population Variance

Let's obtain the $1-\alpha$ percent confidence interval estimator for μ:

$$P\left(\frac{\sqrt{n}}{\sigma}\left|\bar{X}-\mu\right| \leq z_{\alpha/2}\right) = 1-\alpha$$

Reorganizing

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sqrt{n}}{\sigma} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sqrt{n}}{\sigma}\right) = 1-\alpha$$

If the observed value of $\bar{X}$ is $\bar{x}$, then we call the interval

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

percent confidence interval estimate of μ.

Interval Estimators of the Mean with Known Population Variance. Example

Suppose that if a signal having intensity μ originates at location A, then the intensity recorded at location B is normally distributed with mean μ and standard deviation 3. To reduce the error, the same signal is independently recorded 10 times. If the successive recorded values are

$$17, 21, 20, 18, 19, 22, 20, 21, 16, 19$$

construct a 90, 95 and 99 percent confidence interval for μ, the actual intensity.

Interval Estimators of the Mean with Known Population Variance. How Large a Sample is Needed?

The length of the $100(1-\alpha)$ percent confidence interval estimator of the population mean will be less than or equal to b when the sample size n satisfies

$$n \geq \left(\frac{2z_{\alpha/2}\sigma}{b}\right)^2$$

Interval Estimators of the Mean with Known Population Variance. Example

From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 90 percent certain that our estimate of the mean weight of a salmon is correct to within ± 0.1 pounds, how large a sample is needed? What if we want to be 99 percent certain?

Interval Estimators of the Mean with Known Population Variance. Lower and Upper Confidence Bounds

A $100(1-\alpha)$ percent lower confidence bound for the population mean μ is given by

$$\bar{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}}$$

A $100(1-\alpha)$ percent upper confidence bound for the population mean μ is given by

$$\bar{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$$

Interval Estimators of the Mean with Unknown Population Variance

Since σ is no longer known, we replace it by its estimator S. We based our confidence interval in the t-distributed variable:

$$T_{n-1} = \sqrt{n}\frac{\bar{X} - \mu}{S}$$

A $100(1 - \alpha)$ percent confidence interval estimator for the population mean μ is given by the interval

$$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

Interval Estimators of the Mean with Unknown Population Variance

A $100(1-\alpha)$ percent lower confidence bound for the population mean μ is given by

$$\bar{X} - t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

A $100(1-\alpha)$ percent upper confidence bound for the population mean μ is given by

$$\bar{X} + t_{n-1,\alpha}\frac{S}{\sqrt{n}}$$

Interval Estimators of the Mean with Unknown Population Variance: Example

The Environmental Protection Agency (EPA) is concerned about the amounts of PCB, a toxic chemical, in the milk of nursing mothers. In a sample of 20 women, the amounts (in parts per million) of PCB were as follows:

$$16, 0, 0, 2, 3, 6, 8, 2, 5, 0, 12, 10, 5, 7, 2, 3, 8, 17, 9, 1$$

Use these data to obtain a (a) 95 percent confidence interval; (b) 99 percent confidence interval of the average amount of PCB in the milk of nursing mothers. (c) 95 percent upper confidence bound (d); 99 percent lower confidence bound

Interval Estimators of a Population Proportion

We can use the normal approximation to the binomial distribution to obtain the approximate $100(1 - \alpha)$ percent confidence interval estimator of p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The length of this interval is

$$2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

with the upper bound

$$\text{Length of interval} \leq \frac{z_{\alpha/2}}{\sqrt{n}}$$

Interval Estimators of a Population Proportion

A $100(1-\alpha)$ percent lower confidence bound for the population mean μ is given by

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A $100(1-\alpha)$ percent upper confidence bound for the population mean μ is given by

$$\hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Interval Estimators of a Population Proportion. Example

Out of a random sample of 100 students at a university, 82 stated that they were nonsmokers. Based on this, construct a 99 percent confidence interval estimate of p, the proportion of all the students at the university who are nonsmokers.

Interval Estimators of a Population Proportion. Example

How large a sample is needed to ensure that the length of the 90 percent confidence interval estimate of p is less than 0.01?