Statistics for the Biological, Environmental and Health Sciences

STAT 007

Estimating Parameters and Determining the Sample Size

Chapter 7

Point Estimates, Confidence Interval, and Samples Sizes

Sections 7-1 and 7-2

- In this sections we will:
  - Introduce estimates for population parameters (population mean and population proportion).
  - Discuss how to interpret these estimates.
  - Discuss how to determine the sample size to estimate a population parameter.

- In this section we will introduce the first statistical inference method: estimation.
- We use estimates to *guess* the value(s) of the population parameter based on observed data.
- We will focus on estimates in the following scenarios:
  a) we have data that describes individual trials that are independent successes or failures from n fixed trials and we want to find an estimate for the population proportion of successes: p.

  b) we have data that comes from a distribution that is bell-shaped for which, both the mean and standard deviation are unknown and we want to find an estimate for the population mean: μ.

Estimates

- A **point estimate** is a single value used to estimate a population parameter.
- A **confidence interval estimate** is a range (or an interval) of values used to estimate a population parameter.
- A $(1-\alpha)100\%$ **confidence interval estimate** is a confidence interval estimate that has a confidence level of $(1-\alpha)100\%$.
- The **confidence level** is the probability $1-\alpha$ that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Estimates

Example

A random sample of 860 births in New York State included 426 boys. It is of interest to estimate the population proportion of newborn boys.

a) A point estimate for the population proportion of newborn boys is 0.4953.

b) A 95% confidence interval estimate for the population proportion of newborn boys is $0.4619 < p < 0.5287$.

c) The confidence level of the above interval is 95% (or 0.95). This means that with probability 0.95 the confidence interval actually does contain the population proportion of newborn boys, assuming that the estimation process is repeated a large number of times.

d) From the **computed** confidence interval we can say: we are 95% **confident** that the interval from 0.4619 to 0.5287 actually does contain the true value of the population proportion of newborn boy.

e) A 99% confidence interval estimate for the population proportion of newborn boys is $0.4514 < p < 0.5392$. What is the confidence level of the above interval? How is this confidence level interpreted? How is the computed confidence interval estimate interpreted?

Confidence Interval Estimates

- We use *critical values* to compute confidence intervals.
- A **critical value** is the number on the borderline separating sample statistics that are significantly high (or low) from those that are not significant.
- For example, the number $z_{\alpha/2}$ is a critical value that is a z score with the property that it is at the border that separates and area of $\alpha/2$ in the right tail of the normal distribution.
- The following table shows the relationship between the confidence level, the corresponding value of α, and the critical value $z_{\alpha/2}$.

Most Common Confidence Level	Corresponding values of α	Critical value $z_{\alpha/2}$
90% (or 0.90) confidence level	$\alpha = 0.10$	1.645
95% (or 0.95) confidence level	$\alpha = 0.05$	1.960
99% (or 0.99) confidence level	$\alpha = 0.01$	2.575

Estimates for p

- The best **point estimate** for the population proportion, p is he sample proportion, $\hat{p}$
  - it is an unbiased estimator
  - is a consistent estimator: the standard deviation of the sampling distribution of $\hat{p}$ is smaller than that of other estimators for p.
- The $(1-\alpha)100\%$ **confidence interval estimate** for the population proportion is

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

- **Requirements** to use the above interval estimate:
  - The sample is a simple random sample.
  - The conditions for the binomial distribution are satisfied: there is a fixed number of trials, the trials are independent, there are two categories of outcomes, and the probabilities remain constant for each trial.
  - There are at least 5 successes and at least 5 failures (so the normal distribution is a suitable approximation to the binomial distribution.)

Determining Sample Size when estimating p

- The difference between a point estimate and the population parameter is an error. The **margin of error**, denoted E, is the maximum amount of that error.
- To determine how large the sample size n should be in order to estimate the population proportion, p, with a $(1-\alpha)100\%$ confidence interval and a desired E:
  a) If an estimate $\hat{p}$ is known: $n = \frac{[z_{\alpha/2}]^2 \hat{p}(1-\hat{p})}{E^2}$
  b) If no estimate $\hat{p}$ is known: $n = \frac{[z_{\alpha/2}]^2 0.25}{E^2}$

Confidence Intervals for p

Example

A random sample of 860 births in New York State included 426 boys. We are going to find estimates for the population proportion of newborn boys.

a) Find a point estimate for the population proportion of newborn boys.

b) Are the requirements for finding a confidence interval estimate for p reasonable?

c) Find the margin of error E that corresponds to a 95% confidence interval for p.

d) Find the 95% confidence interval estimate for the population proportion of newborn boys.

e) Make an interpretation of the confidence interval estimate you found.

f) It is believed that among all births, the proportion of boys is 0.512. Do these sample results provide strong evidence against that belief?

g) Make an interpretation of the confidence level of the interval you found.

Determining Sample Size for p

Example

If we were to conduct a survey to determine the proportion of children (older than 1 year) who have received measles vaccinations, how many children must be surveyed in order to be 95% confident that the sample proportion is in error by no more than three percentage points?

a) Assume that a recent survey showed that the proportion of children that have received measles vaccinations is 0.9.

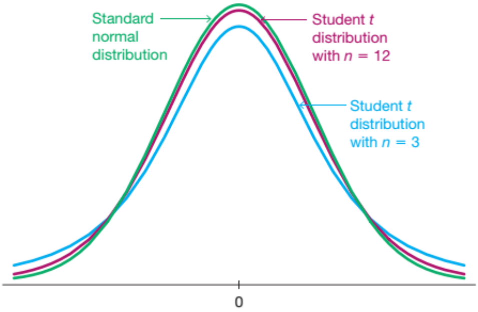
b) Assume that we have no prior information suggesting a possible value of the population proportion.

Estimates for μ

- In what follows we will provide estimates for the population mean, μ.
- We will consider the more realistic case in which the population standard deviation, σ, is also unknown.
- Critical values from a $(1-\alpha)100\%$ confidence interval estimate for μ are obtained from the Student *t* distribution and denoted $t_{\alpha/2}$.
- The Student *t* distribution has a parameter called *degrees of freedom*, denoted *df*.
- When estimating μ, the critical value $t_{\alpha/2}$ separates an area of $\alpha/2$ in the right tail of the Student *t* distribution with $df = n-1$, where *n* is the size of the sample.

Properties of the Student *t* distribution

- It is different for different sample sizes.
- It has the same general bell shape as the standard normal distribution. The wider shape reflects the greater variability that is expected when σ is estimated by s.
- It has a mean equal to 0.
- It has a standard deviation that varies with the sample size and is larger than 1.
- As the sample size increases, it gets closer to the standard normal distribution.



Estimates for μ (unknown σ)

- The sample mean, $\overline{x}$, is the best **point estimate** of the population mean, μ:
  - it is an unbiased estimator
  - is a consistent estimator: the standard deviation of the sampling distribution of $\overline{x}$ is smaller than that of other estimators for μ.
- The $(1-\alpha)100\%$ **confidence interval estimate** for the population mean is

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

  where $s = \sqrt{\sum \frac{(x_i-\overline{x})^2}{n-1}}$ is the sample standard deviation.
- **Requirements** to use the above interval estimate:
  - The sample is a simple random sample.
  - Either or both of these conditions are satisfied: the population is normally distributed or $n > 30$.

Determining Sample Size when estimating μ

- To determine how large the sample size n should be in order to estimate the population mean μ with a $(1-\alpha)100\%$ confidence interval and a desired E:
  - if σ is known use $n = \left[\frac{z_{\alpha/2}\ \sigma}{E}\right]^2$
  - When σ is not known (most regular case) use $n = \left[\frac{z_{\alpha/2}\ \sigma}{E}\right]^2$, with
    a) $\sigma \approx range/4$
    b) $\sigma \approx s$
    c) σ a value from results of some other earlier studies.

Estimates for μ

Example

Listed below are weights (in hectograms or hg) of randomly selected girls at birth and corresponding normal quantile plot, based on data from the National Center for Health Statistics. Here are the summary statistics: $n = 15$, $\overline{x} = 30.9$ hg, $s = 2.9$ hg.

Data: `33 28 33 37 31 32 31 28 34 28 33 26 30 31 28`

a) Find a point estimate for the mean birth weight of girls.

b) Based on the quantile plot in next slide, are the requirements to find a confidence interval estimate for μ reasonable?

c) Find the margin of error E that corresponds to a 95% confidence interval for μ.

d) Find the 95% confidence interval estimate of the population mean birth weight of girls.

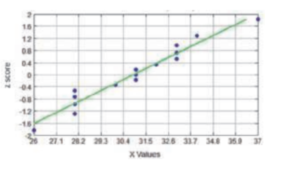
e) Make an interpretation of the confidence interval estimate you found.

f) Make an interpretation of the confidence level of the interval you found.

Estimates for μ

Example

Quantile plot:



Determining Sample Size for μ

Example

Assume that we want to estimate the mean IQ score for the population of adults who smoke, for which the population standard deviation is 15.

How many smokers must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

Practice

Look at the exercises at the end of Section 7-1 in page 295

Specially, look at exercises: 1 to 8, 13, 15, 18, 21, 26, 29, 31, 33

Look at the exercises at the end of Section 7-2 in page 309

Specially, look at exercises: 9, 10, 11, 13, 17, 18, 24, 28, 30, 34