CHAPTER 5: PARAMETER ESTIMATION

STATISTICS (FERM)

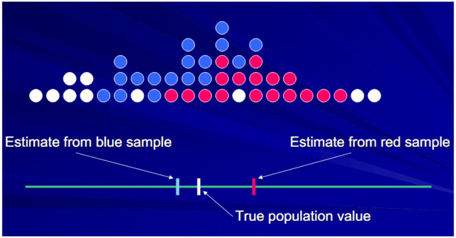
Lecturer: Nguyen Minh Quan, PhD
quannm@hcmiu.edu.vn

CONTENTS

1. Maximum likelihood estimators
2. Interval Estimates
3. Confidence Interval for a Normal Mean When the Variance Is Unknown
4. Confidence interval for a proportion
5. Confidence Interval for the Variance of a Normal Distribution

Statistical Inference estimator

- Statistical inference means drawing conclusions based on data. There are a many contexts in which inference is desirable, and there are many approaches to performing inference.
- It involves either estimation OR hypothesis testing.
- It is based on sample outcomes.
- In this chapter, we will learn about estimation.



Estimator of a parameter

- Reference: Chapter 7 in the textbook by S. Ross.
- Let $X_1, \cdots, X_n$ be a random sample from a distribution F_θ that is specified up to a vector of unknown parameters θ.
- For instance, the sample could be from a normal distribution having an unknown mean and variance.
- Any statistic used to estimate the value of an unknown parameter θ is called an estimator of θ.
- The observed value of the estimator is called the estimate.

Estimator of a parameter

Example

The usual estimator of the mean of a normal population, based on a sample $X_1, \cdots, X_n$ from that population, is the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

If a sample of size 4 yields the data $X_1 = 2, X_2 = 3.5, X_3 = 4, X_4 = 2.5$, then the estimate of the population mean, resulting from the estimator X, is the value 3.

Maximum Likelihood Method

- Suppose that the random variables $X_1, \cdots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed.

Example

The RVs X_i might be independent, exponential random variables each having the same unknown mean θ, Then, the joint density function of the random variables would be given by

$$f(x_1, x_2, .., x_n) = \frac{1}{\theta^n} \exp\left\{ -\sum_{i=1}^{n} x_i/\theta \right\}, 0 < x_i < \infty, i = 1, 2, \ldots, n$$

Maximum Likelihood Method

- Let $f(x_1, x_2, ..., x_n|\theta)$ denote the joint probability density (or mass) function of the random variables $X_1, X_2, ..., X_n$.

Maximum Likelihood Method

The maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, x_2, ..., x_n|\theta)$, where $x_1, ..., x_n$ are the observed values.

- The function $f(x_1, x_2, ..., x_n|\theta)$ is often referred to as the likelihood function of θ.

- We may also obtain $\hat{\theta}$ by maximizing $\ln(f(x_1, x_2, ..., x_n|\theta))$.

Maximum Likelihood Method

Example: Maximum Likelihood Estimator of a Bernoulli Parameter

Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p?

Solution

$$X_i = \begin{cases} 1, & \text{if trial is a success} \\ 0, & \text{othewise} \end{cases}$$

$$P\left(X_i = x\right) = p^x\left(1-p\right)^{1-x}; x = 0, 1$$

$$f\left(x_1, x_2, .., x_n | p\right) = P\left(X_1 = x_1, ..., X_n = x_n | p\right)$$

$$= p^{\sum\limits_{i=1}^{n} x_i}\left(1-p\right)^{n-\sum\limits_{i=1}^{n} x_i}$$

$$\Rightarrow \ln\left\{f\left(x_1, x_2, .., x_n | p\right)\right\} = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln\left(1-p\right)$$

Maximum Likelihood Method

Solution (Cont.)

Differentiation w.r.t p yields

$$\frac{d}{dp}\ln\left\{f\left(x_1, x_2, .., x_n|p\right)\right\} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

Upon equating to zero, the maximum likelihood estimate $\hat{p}$ satisfies

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} - \frac{n - \sum_{i=1}^{n} x_i}{1-\hat{p}}$$

Thus, $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$. This is the proportion of the observed trials that result in successes!

Maximum Likelihood Method

Exercise

Suppose $X_1, ..., X_n$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

Hint

$$f(x_1, x_2, .., x_n|\lambda) = \frac{e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n} x_i}}{x_1!...x_n!}$$

$$\frac{d}{d\lambda}\ln\{f(x_1, x_2, .., x_n|\lambda)\} = -n + \frac{\sum\limits_{i=1}^{n} x_i}{\lambda}$$

Maximum Likelihood Method

Exercise

The number of traffic accidents in Berkeley, California, in 10 randomly chosen nonrainy days in 1998 is as follows:

$$4,\ 0,\ 6,\ 5,\ 2,\ 1,\ 2,\ 0,\ 4,\ 3.$$

Use these data to estimate the proportion of nonrainy days that had 2 or fewer accidents that year.

Solution Assume that the daily number of traffic accidents is a Poisson random variable.

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2.7$$

It follows that the maximum likelihood estimate of the Poisson mean is 2.7. Let X be the random number of accidents in a day.

$$P\left(X \leqslant 2\right) = e^{-2.7}\left(1 + 2.7 + (2.7)^2/2\right) = 0.4936$$

Maximum Likelihood Estimator in a Normal Population

Suppose $X_1, \cdots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ. The joint density is given by

$$f\left(x_1, x_2, .., x_n \middle| \mu, \sigma\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

The logarithm of the likelihood is thus given by

$$\ln f\left(x_1, x_2, .., x_n \middle| \mu, \sigma\right) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

It yields

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma} = \left[\sum_{i=1}^{n}(x_i-\hat{\mu})^2/n\right]^{1/2}$$

Maximum Likelihood Estimator in a Normal Population

Kolmogorov's law of fragmentation states that the size of an individual particle in a large collection of particles resulting from the fragmentation of a mineral compound will have an approximate lognormal distribution, where a random variable X is said to have a lognormal distribution if $\ln(X)$ has a normal distribution.

Example

Suppose that a sample of 10 grains of metallic sand taken from a large sand pile have respective lengths (in millimeters):

$$2.2, 3.4, 1.6, 0.8, 2.7, 3.3, 1.6, 2.8, 2.5, 1.9$$

Estimate the percentage of sand grains in the entire pile whose length is between 2 and 3 mm.

Maximum Likelihood Estimator in a Normal Population

Solution

Taking the natural logarithm of these 10 data values:

$$0.7885, 1.2238, 0.4700, -0.2231, 0.9933, 1.1939, 0.4700, 1.0296, 0.9163, 0.6419$$

The sample mean and sample standard deviation are $\bar{x} = 0.7504, s = 0.4351$. $Y := \ln X$ is a normal distribution.

$$P(2 < X < 3) = P(\ln 2 < Y < \ln 3)$$

$$= P\left(\frac{\ln 2 - 0.7504}{0.435} < \frac{Y - 0.7504}{0.435} < \frac{\ln 3 - 0.7504}{0.435}\right) \approx 0.3405$$

Interval Estimates

- Suppose that $X_1, \cdots, X_n$ is a sample from a normal population having unknown mean μ and known standard deviation σ.

- It is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that μ lies within.

- This method is so called interval estimator.

Interval Estimates

- Since the point estimator $\bar{X}$ is normal with mean μ and variance σ^2/n, we have
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- Therefore,
$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

- Thus, 95 percent of the time μ will lie within $1.96\sigma\sqrt{n}$ units of the sample average:
$$P\left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95$$

Interval Estimates

- If we now observe the sample $\bar{X} = x$, then we say that "with 95 percent confidence"

$$\bar{x} - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{1.96\sigma}{\sqrt{n}}$$

- The interval

$$\left(\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}}\right)$$

is called a 95 percent confidence interval estimate of μ.

Interval Estimates

Example

Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

Solution

Since $\bar{x} = \frac{81}{9} = 9$. It follows, under the assumption that the values received are independent, that a 95 percent confidence interval for μ is

$$\left(\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}}\right) = (7.69, 10.31)$$

Hence, we are 95 percent confident that the true message value lies between 7.69 and 10.31.

Interval Estimates

- The interval $\left(\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}}\right)$ is called a two-sided confidence interval
- Next, we are interested in determining a value so that we can assert with, say, 95 percent confidence, that μ is at least as large as that value.
- That is, we find A such that

$$P\left(A < \mu\right) = 0.95$$

Interval Estimates

- To determine such value of A, note that if Z is a standard normal random variable then

$$P\left(Z < 1.645\right) = 0.95$$

$$\rightarrow P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.645\right) = 0.95$$

- Therefore,

$$P\left(\bar{X} - 1.645\frac{\sigma}{\sqrt{n}} < \mu\right) = 0.95$$

Interval Estimates

- Thus, a 95 percent one-sided upper confidence interval for μ is $\left(\overline{x} - 1.645\frac{\sigma}{\sqrt{n}}, \infty\right)$ where $\bar{x}$ is the observed value of the sample mean.

- Similarly, the 95 percent one-sided lower confidence interval for μ is $\left(-\infty, \overline{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$, where $\bar{x}$ is the observed value of the sample mean.

Interval Estimates

Example

Determine the upper and lower 95 percent confidence interval estimates of μ in previous example:

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$$

Solution We have $1.645\frac{\sigma}{\sqrt{n}} = 1.097$. Therefore,

$$\overline{x} - 1.645\frac{\sigma}{\sqrt{n}} = 9 - 1.097 = 7.903$$

Thus, the 95 percent upper confidence interval is $\left(\overline{x} - 1.645\frac{\sigma}{\sqrt{n}}, \infty\right) = (7.903, \infty)$.

Similarly, the 95 percent lower confidence interval is $(-\infty, 10.097)$

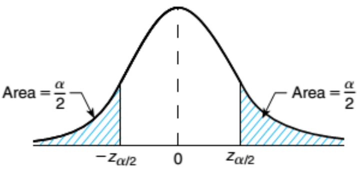
Interval Estimates with any specified level of confidence

- We can also obtain confidence intervals of any specified level of confidence.

- Recall that z_α is such that $P\left(Z > z_\alpha\right) = \alpha$, when Z is a standard normal random variable.

- On the other hand,

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$$



$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha.$

Interval Estimates with any specified level of confidence

- Thus,

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

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

- That implies

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

- Hence, a $100(1-\alpha)$ percent two-sided confidence interval for μ is

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

where $\bar{x}$ is the observed sample mean.

Interval Estimates with any specified level of confidence

- Similarly, knowing that $Z = \sqrt{n}\frac{\overline{X}-\mu}{\sigma}$ is a standard normal random variable, along with the identities

  $$P\left(Z > z_\alpha\right) = \alpha \text{ and } P\left(Z < -z_\alpha\right) = \alpha$$

  results in one-sided confidence intervals of any desired level of confidence.

- Specifically, we obtain

  $$\left(\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right) \text{ and } \left(-\infty, \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

  are, respectively, $100(1-\alpha)$ percent one-sided upper and $100(1-\alpha)$ percent one-sided lower confidence intervals for μ.

Interval Estimates with any specified level of confidence

Example

Use the below data to obtain a 99% confidence interval estimate of μ, along with 99 percent one-sided upper and lower intervals.
Data:

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$$

Solution Since $z_{0.005} = 2.58$, and

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.72$$

It follows that the 99 percent confidence interval estimate is

$$(9 - 1.72, 9 + 1.72) = (7.28, 10.72)$$

Interval Estimates with any specified level of confidence

Solution (cont.)

Since $z_\alpha = z_{0.01} = 2.33$, a 99 percent upper confidence interval is

$$(9 - z_{0.1}(2/3), \infty) = (7.447, \infty)$$

Similarly, a 99 percent lower confidence interval is

$$(-\infty, 9 + z_{0.1}(2/3)) = (-\infty, 10.553)$$

Interval Estimates with any specified level of confidence

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 95 percent certain that our estimate of the present season's mean weight of a salmon is correct to within ± 0.1 pounds, how large a sample is needed?

Hint

$$1.96\frac{\sigma}{\sqrt{n}} \leqslant 0.1 \rightarrow n \geqslant 34.57$$

Interval Estimates with any specified level of confidence

Exercises

A sample survey of 54 discount brokers showed that the mean price charged for a trade of 100 shares at \$50 per share was \$33.77 (AAII Journal, February 2006). The survey is conducted annually. With the historical data available, assume a known population standard deviation of \$15.

a. Using the sample data, what is the margin of error associated with a 95% confidence interval?

b. Develop a 95% confidence interval for the mean price charged by discount brokers for a trade of 100 shares at \$50 per share.

Confidence Interval for μ when σ is unknown

- Suppose now that $X_1, \cdots, X_n$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1-\alpha)$ percent confidence interval for μ.

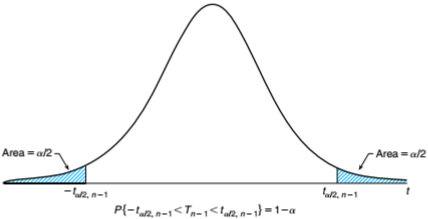
- Let $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$ denote the sample variance.

- It follows that (from Corollary 6.5.2)

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

is a t-random variable with $n-1$ degrees of freedom.

Confidence Interval for μ when σ is unknown



- From the symmetry of the t-density function, we have that for any $\alpha \in (0, 1/2)$,

$$P\left(-t_{\alpha/2,n-1} < \sqrt{n}\frac{\bar{X}-\mu}{S} < t_{\alpha/2,n-1}\right) = 1-\alpha$$

- Therefore, if it is observed that $\bar{X} = \bar{x}$ and $S = s$, then we can say that "with $100(1-\alpha)$ percent confidence":

$$\mu \in \left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

Confidence Interval for μ when σ is unknown

Example

Let us again consider previous example (Example 7.3a) but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

Solution

We have $\bar{x} = 9$, $s^2 = 9.5$, $s = 3.082$.
Hence, as $t_{0.025,8} = 2.306$, a 95 percent confidence interval for μ is

$$\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = (6.63, 11.37)$$

Confidence Interval for μ when σ is unknown

If it is observed that $\bar{X} = \bar{x}$, $S = s$, then we can assert:

Upper confidence interval

"with $100(1-\alpha)$ percent confidence" that

$$\mu \in \left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha, n-1}, \infty\right)$$

Lower confidence interval

"with $100(1-\alpha)$ percent confidence" that

$$\mu \in \left(-\infty, \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha, n-1}\right)$$

Confidence Interval for μ when σ is unknown

Example

Determine a 95 percent confidence interval for the average resting pulse of the members of a health club if a random selection of 15 members of the club yielded the data

$$54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77.$$

Also determine a 95 percent lower confidence interval for this mean.

Interval Estimates with any specified level of confidence

Exercises

A survey conducted by the American Automobile Association showed that a family of four spends an average of \$215.60 per day while on vacation. Suppose a sample of 64 families of four vacationing at Niagara Falls resulted in a sample mean of \$252.45 per day and a sample standard deviation of \$74.50.

a. Develop a 95% confidence interval estimate of the mean amount spent per day by a family of four visiting Niagara Falls.

b. Based on the confidence interval from part (a), does it appear that the population mean amount spent per day by families visiting Niagara Falls differs from the mean reported by the American Automobile Association? Explain.

Confidence interval for a proportion

- **Recall the Central Limit Theorem:** For n large, the sample proportion $\hat{p}$ follows a normal distribution with the following mean and standard error (standard deviation):

$$\mu_{\hat{p}} = p, \quad SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- In order for the Central Limit Theorem to hold, the sample size is typically considered sufficiently large when $np \geq 10$ and $n(1-p) \geq 10$, which is called the success-failure condition.

- In many situations, we do not know p. We can use the **approximate condition** for the Success-failure condition: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$ to estimate $SE_{\hat{p}}$ or one can use T-distribution if this condition is not satisfied.

Confidence interval for a proportion p

- The interval $(\hat{p} - 1.96 \times SE_{\hat{p}}, \hat{p} + 1.96 \times SE_{\hat{p}})$ is called the 95% confidence interval for the proportion p.
- The value of $1.96 \times SE_{\hat{p}} = z_{0.025} \times SE_{\hat{p}}$ is called the **margin error.**
- In general, the interval $(\hat{p} - z_{\alpha/2} \times SE_{\hat{p}}, \hat{p} + z_{\alpha/2} \times SE_{\hat{p}})$ is called the $100(1-\alpha)\%$ confidence interval for the proportion p.
- For example, below is a graph for 25 intervals with confidence level of 95% for the proportion which is $p = 0.88$. Twenty-four intervals do capture $p = 0.88$ and one interval does not capture $p = 0.88$:

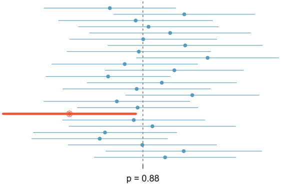


Figure 5.6: Twenty-five point estimates and confidence intervals from the simulations in Section 5.1.2. These intervals are shown relative to the population proportion $p = 0.88$. Only 1 of these 25 intervals did not capture the population proportion, and this interval has been bolded.

Confidence interval for a proportion p

- The interval $(\hat{p} - 1.96 \times SE_{\hat{p}}, \hat{p} + 1.96 \times SE_{\hat{p}})$ is called the 95% confidence interval for the proportion p.
- The value of $1.96 \times SE_{\hat{p}} = z_{0.025} \times SE_{\hat{p}}$ is called the **margin error.**
- In general, the interval $(\hat{p} - z_{\alpha/2} \times SE_{\hat{p}}, \hat{p} + z_{\alpha/2} \times SE_{\hat{p}})$ is called the $100(1-\alpha)\%$ confidence interval for the proportion p.
- For example, below is a graph for 25 intervals with confidence level of 95% for the proportion which is $p = 0.88$. Twenty-four intervals do capture $p = 0.88$ and one interval does not capture $p = 0.88$:

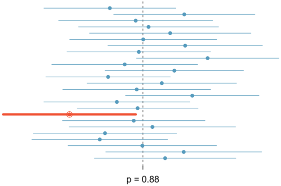


Figure 5.6: Twenty-five point estimates and confidence intervals from the simulations in Section 5.1.2. These intervals are shown relative to the population proportion $p = 0.88$. Only 1 of these 25 intervals did not capture the population proportion, and this interval has been bolded.

Confidence interval for a proportion p

Example

In a Pew Research poll, there is $\hat{p}$ =88.7% of a random sample of $n = 1000$ American adults supported expanding the role of solar power.
(a) Compute and interpret a 95% confidence interval for the population proportion p.
(b) Assume $\hat{p}$ =88.7% with $n = 1000$, construct a 99% confidence interval for p.
(c) Comment your observation on length of the two intervals in part (a) and (b) corresponding to the 95% and 99% confidence level.

Confidence Interval for σ^2

If $X_1, \cdots, X_n$ is a sample from a normal distribution having unknown parameters μ and σ^2, then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1)\,\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Thus,

$$P\left\{\chi^2_{1-\alpha/2,n-1} \leqslant (n-1)\,\frac{S^2}{\sigma^2} \leqslant \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$

or, equivalently,

$$P\left\{\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,n-1}} \leqslant \sigma^2 \leqslant \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,n-1}}\right\} = 1-\alpha$$

Confidence Interval for σ^2

Hence when $S^2 = s^2$, $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\left(\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

Example

A standardized procedure is expected to produce washers with very small deviation in their thicknesses. Suppose that 10 such washers were chosen and measured. If the thicknesses of these washers were, in inches,

$$.123\ .133\ .124\ .125\ .126\ .128\ .120\ .124\ .130\ .126$$

what is a 90 percent confidence interval for the standard deviation of the thickness of a washer produced by this procedure?

Confidence Interval for σ^2

Solution

A computation gives that $S^2 = 1.366 \times 10^{-5}$. Note that

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.05,9} = 16.917; \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.95,9} = 3.334$$

$$\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,n-1}} = 7.267 \times 10^{-6}; \quad \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,n-1}} = 36.875 \times 10^{-6}$$

Taking square roots, it follows that, with confidence 0.9,

$$\sigma \in (2.696 \times 10^{-3}, 6.072 \times 10^{-3})$$

Confidence Interval for μ or σ^2: Sums up the results

TABLE 7.1 *100(1 −α) Percent Confidence Intervals*

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \sum_{i=1}^{n} X_i/n, \qquad S = \sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2/(n-1)}$$

Assumption	Parameter	Confidence Interval	Lower Interval	Upper Interval
σ^2 known	μ	$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	$\left(-\infty, \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$	$\left(\overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right)$
σ^2 unknown	μ	$\overline{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$	$\left(-\infty, \overline{X} + t_{\alpha, n-1} \frac{S}{\sqrt{n}}\right)$	$\left(\overline{X} - t_{\alpha, n-1} \frac{S}{\sqrt{n}}, \infty\right)$
μ unknown	σ^2	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}\right)$	$\left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha, n-1}}\right)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}, \infty\right)$

Estimating the Difference in Means of Two Normal Populations

TABLE 7.2 *$100(1-\sigma)$ Percent Confidence Intervals for $\mu_1 - \mu_2$*

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

$$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

$$\overline{X} = \sum_{i=1}^{n} X_i/n, \qquad S_1^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2/(n-1)$$

$$\overline{Y} = \sum_{i=1}^{m} Y_i/n, \qquad S_2^2 = \sum_{i=1}^{m} (Y_i - \overline{Y})^2/(m-1)$$

Difference in Means of Two Normal Populations

Assumption	Confidence Interval
σ_1, σ_2 known	$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\sigma_1^2/n + \sigma_2^2/m}$
σ_1, σ_2 unknown but equal	$\overline{X} - \overline{Y} \pm t_{\alpha/2,n+m-2}\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$

Assumption	Lower Confidence Interval
σ_1, σ_2 known	$(-\infty, \overline{X} - \overline{Y} + z_{\alpha}\sqrt{\sigma_1^2/n + \sigma_2^2/m})$
σ_1, σ_2 unknown but equal	$\left(-\infty, \overline{X} - \overline{Y} + t_{\alpha,n+m-2}\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}\right)$

Note: Upper confidence intervals for $\mu_1 - \mu_2$ are obtained from lower confidence intervals for $\mu_2 - \mu_1$.

Difference in Means of Two Normal Populations

Example

A civil engineer wishes to measure the compressive strength of two different types of concrete. A random sample of 10 specimens of the first type yielded the following data (in psi).
Type 1: 3250, 3268, 4302, 3184, 3266 3297, 3332, 3502, 3064, 3116

whereas a sample of 10 specimens of the second yielded the data
Type 2: 3094, 3106, 3004, 3066, 2984, 3124, 3316, 3212, 3380, 3018

If we assume that the samples are normal with a common variance, determine a 95 percent two-sided confidence interval for the difference in means.

End of Chapter 5