

O. Estimation of Parameters.

1. Point and Interval Estimation.

Properties of Estimators. Let $\hat{\theta}$ be an estimator for θ .

a. **Unbiasedness** ($E(\hat{\theta}) = \theta$).

b. **Consistency** (As sample size gets larger, estimate gets better).

c. **Efficiency** ($\hat{\theta}$ has a small variance). Define BLUE.

d. **Maximum Likelihood** ($\hat{\theta}$ is the value of θ that is most likely to have produced the observed data).

2. A Confidence Interval for μ When σ is Known.

$\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$ You can only use this when you know the population variance. Don't forget that there are two formulas for the standard error depending on sample size! $\sigma_{\bar{x}}$ is the standard error.

An interval of this type is used in two situations:

(i) where the population variance, σ^2 , is in fact, known and the sample size is relatively large; or

(ii) where the variance is not known and the sample variance, s^2 , is used to replace σ^2 , but the degrees of freedom ($n - 1$) are so large that the appropriate value of $t^{(n-1)}$ is not very different from z .

The first of these situations is not very realistic, but serves as a good introduction to confidence intervals. The formula for this type of

confidence interval for the mean is, $\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$, where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Note: If $n > .05N$, use $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

(n is sample size and N is population size)

Example 1: Assume that a population is Normally distributed with an unknown mean and a population standard deviation of 36.

($X \sim N(?, 36)$) From a random sample of size $n = 9$, we get a sample mean of 62. (Because the population variance is known we can ignore any sample variance we might compute from the data.) Find a 95% confidence interval for the mean.

Step 1: State the confidence level and significance level.

The given confidence level of 95% represents the probability that the interval actually contains the mean and is stated as $1 - \alpha = .95$.

The significance level of 5% represents the probability of being

wrong and is $\alpha = .05$.

Step 2: Find the appropriate value of Z . Use the last line of Table 18 (or Table 17) in the Syllabus Supplement to find

$$z_{\alpha/2} = z_{.025} = 1.960 \text{ (the bottom number in the .025 column).}$$

Note that higher confidence levels give larger values of Z , and thus larger confidence intervals.

Step 3: Find the standard error. $\sigma_{\bar{x}} = \sigma / \sqrt{n} = \frac{36}{\sqrt{9}} = 12$.

Note that larger values of n make the standard error and the confidence interval smaller.

Step 4: Put it together. $\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}} = 62 \pm (1.960)(12)$

$= 62 \pm 23.52$. The last part of this expression means that the interval extends from $62 - 23.52 = 38.48$ to $62 + 23.52 = 85.52$.

The result can be written $P\{38.48 \leq \mu \leq 85.52\} = .95$.

Example 2: Assume that a population is Normally distributed with an unknown mean and a population standard deviation of 36.

($X \sim N(\mu, 36)$) From a random sample of size $n = 9$, we get a sample mean of 62. (Because the population variance is known we can ignore any sample variance we might compute from the data.) This time find a 99% confidence interval for the mean.

Step 1: State the confidence level and significance level.

The given confidence level of 99% represents the probability that the interval actually contains the mean and is stated as $1 - \alpha = .99$. The significance level of 1% represents the probability of being wrong and is $\alpha = .01$.

Step 2: Find the appropriate value of Z . Use the last line of Table 18 (or Table 17) in the Syllabus Supplement to find

$$z_{\alpha/2} = z_{.005} = 2.576 \text{ (the bottom number in the .005 column).}$$

Note that higher confidence levels give larger values of Z , and thus larger confidence intervals.

Step 3: Find the standard error. $\sigma_{\bar{x}} = \sigma / \sqrt{n} = \frac{36}{\sqrt{9}} = 12$.

No change from example 1.

Step 4: Put it together.

$$\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}} = 62 \pm (2.576)(12) = 62 \pm 30.91,$$

The result can be written $P\{31.09 \leq \mu \leq 92.91\} = .99$.

Or make a Normal curve with 62 in the middle and 31.09 and 92.91 on the sides. Label the area between 31.09 and 92.91 with 99%, the area below 31.09 with 0.5% and the area above 92.91 with 0.5%.

Definitions:

Note that if we are considering the possibility that the population mean is 50, we can now say that, since this value is on the confidence interval, that the mean is not significantly different from 50.

However the mean is significantly different from 20 or 100. Remember that a confidence level is the probability that a given confidence interval is correct. The usual interpretation of a confidence level of 95% is that if we take samples of n items and use the methods given here many times, 95% of the time the interval will contain the population mean. The significance level is the probability that the interval will not contain the population mean. If we say that the population mean is significantly different from 20 and our significance level is 5%, we are saying that there is a probability of 5% or less that the observed data could have been generated by a distribution with a population mean of 20.

3. A Confidence Interval for μ When σ is not known.

$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}}^{(n-1)} s_{\bar{x}} \quad \text{This is what you actually use most of the time!}$$

All that " σ unknown" means is that we do not have a value of the population variance. If you only have the sample variance, use the t table.

Finding degrees of freedom is easy. In most of the problems that we do the numbers of degrees of freedom is one less than the sample size or $n - 1$. The value of t that you need should be in Table 18 in the Syllabus Supplement. Each row represents the number of degrees of freedom given by the 'df' column. It is a good idea to take a ruler and put a line across the table after every 10th row. Note that the table skips values after 100 degrees of freedom, but a good guess is always possible, for example $t_{.05}^{110} \approx 1.659$.

"The variance is not known" implies no previous knowledge or assumption about the value of the population variance, σ^2 . Knowing the sample variance, s^2 is having a good guess as to what the variance is; it is not the same as knowing the variance.

If the population distribution is normal or approximately normal, the formula for a two-sided confidence interval for the mean is $\mu = \bar{x} \pm t_{\frac{\alpha}{2}}^{(n-1)} s_{\bar{x}}$,

$$\text{where } s_{\bar{x}} = \frac{s}{\sqrt{n}}. \quad \text{Note: If } n > .05N, \text{ use } s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

Be careful – It is a common error to think that a new population size is actually a sample size.

Note: this is the more common case – if you do not know the population variance and the sample size is not very large, using z instead of t is a very bad idea.

Example 1: We have a random sample of 10 homes. The sample mean of expenditures on maintenance is \$838 with a sample standard deviation of \$110. Construct a 95% confidence interval for the mean.

Step1: State the confidence level and significance level.

The given confidence level of 95% represents the probability that the interval actually contains the mean and is stated as $1 - \alpha = .95$.

The significance level of 5% represents the probability of being wrong and is $\alpha = .05$.

Step2: Find the appropriate value of t Use Table 18 in the