EXAMPLES INTERVAL ESTIMATION FOR ONE POPULATION

Example 1: Unoccupied seats on flights cause airlines to lose revenue (gelir, hasılat). Suppose a large airplane wants to estimate its average number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected, and the number of unoccupied seats is noted for each of the sampled flights. Estimate μ, the mean number of unoccupied seats per flight during the past year, using given $\bar{x} = 11.6$, s=4.1 and a 90 % confidence interval.

Solution: The (*1-α*)100% confidence interval for a population mean (based on the z-statistic):

$P\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha$, ***since population variance** (σ^2) **is unknown and n=225>30, we will use this formula.***

Substituting** $\bar{x} = 11.6$, **n=225, s=4.1 and** $z_{0.05} = 1.65$ **into the formula, we get;

$$P\left(11.6 - 1.65 \times \frac{4.1}{\sqrt{225}} < \mu < 11.6 + 1.65 \times \frac{4.1}{\sqrt{225}}\right) = 0.90$$

$$P(11.6 - 0.45 < \mu < 11.6 + 0.45) = 0.90$$

The 90 % confidence interval for population mean is: $P(11.15 < \mu < 12.05) = 0.90$.

Comment: That is, at the 90 % confidence level, we estimate the true *mean number of unoccupied seats per flight* (μ) to be between 11.15 and 12.05 dring the sampled year.

Important: We stress that the confidence level for this example, 90%, refers to the procedure used. If we were apply that procedure repeatedly to different samples, approximately 90% of the intervals would contain μ. Although we do not know for sure whether this particular interval (11.15, 12.05) is one of the 90% that contain μ or one of the 10% that do not, our knowledge of probability gives us "confidence" that the interval contains μ.

Example 2: Consider the pharmaceutical company that desires an estimate of the mean increase in blood pressure of pateints who take a new drug. The increase in blood pressure *associated with the new drug for all patients in the population are assumed to be normally distrubuted.* The blood pressure increases (measured in points) measured for the n=6 patients in the human testing phase and the mean and the standard deviation of blood pressure increases for this sample of patients are determined to be $\bar{x} = 2.283$ and s=0.950. Use this information to construct a 95 % confidence interval for μ, the mean increase in blood pressure associated with the new drug for all patients in the population.

Solution: We do not get the normal distribution of $\overline{x}$ “automically” from the Central Limit (CLT) Thorem since here n=6 < 30 (the sample size is small). Instead, we must assume that the measured variable, the increase in blood pressure, is normally distributed in order to write that the distribution of $\overline{x}$ to be normal.

The (*1-α*)100% confidence interval for a population mean (based on the t-statistic):

$P\left(\overline{x} - t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}}\right) = 1-\alpha$, ***since population variance*** (σ^2) ***is unknown and n=6<30, we will use this formula.***

Substituting $\overline{x} = 2.283$, ***n=6, s=0.950 and*** $t_{0.05/2,(6-1)} = t_{0.025,5} = 2.571$ ***into the formula, we get;***

$$P\left(2.283 - 2.571\times\frac{0.950}{\sqrt{6}} < \mu < 2.283 + 2.571\times\frac{0.950}{\sqrt{6}}\right) = 0.95$$

$$P(2.283 - 0.997 < \mu < 2.283 + 0.997) = 0.95$$

The 95 % confidence interval for population mean is: $P(1.286 < \mu < 3.280) = 0.95$.

Comment: We can be 95 % confident (or we estimate with 95% confidence) that the true mean increase in blood pressure associated with taking this new drug (μ) between 1.286 and 3.280 points.

Example 3: A group of Engineers study of contaminated fish in a River. The Engineers has collected data for a random sample of 30 fish contaminated with **Dikloro Difenil Trikloroethan-DDT** (çok zehirli ve inatçı bir böcek öldürücü). The engineers made sure to capture contaminated fish in several different randomly selected streams and tributaries of the river. The standart deviation of fish weights for this sample of fishes (n=30) is measured to be s=9.2. Engineers wants to estimate the true variation in fish weights in order to determine whether the fish are stable enough to allow further testing for DDT contamination. *Use the sample data to find a 95 % confidence interval for the true variation in fish weights (*σ^2*).*

Solution: The (*1-α*)100% confidence interval for a population variance (based on the χ^2 -statistic): $P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,(n-1)}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,(n-1)}}\right) = 1-\alpha$

We calculate the interval by using the the values $\chi^2_{0.025,29} = 45.722$ and $\chi^2_{0.975,29} = 16.047$ obtained from the chi-square table.

$$P\left(\frac{(30-1)(9.2)^2}{45.722} < \sigma^2 < \frac{(30-1)(9.2)^2}{16.047}\right) = 0.95$$

The 95 % confidence interval for population variance: $P(53.7 < \sigma^2 < 152.9) = 0.95$

Comment: Thus, the Engineers can be 95 % confident that the true variance (σ^2) in weights of the population of contaminated fish ranges between 53.7 and 152.9.

Example 4: Public-opinion polls are conducted regularly to estimate the fraction of U.S. citizens who trust the president. Suppose 1000 people are randomly chosen and 637 answer that they trust the president. How would you estimate the true fraction of all U.S. citizens who trust the president? Construct a 95 % confidence interval for the true percentage of all U.S. citizens who trust the president.

Solution: The $(1-\alpha)100\%$ confidence interval for a population proportion, *p*:

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

Substituting n=1000, $\hat{p} = \frac{637}{1000} = 0.637$ and $z_{0.025} = 1.96$ into the formula, we obtain:

$$P\left(0.637 - 1.96\sqrt{\frac{0.637(0.363)}{1000}} < p < 0.637 + 1.96\sqrt{\frac{0.637(0.363)}{1000}}\right) = 0.95$$

$$P(0.637 - 0.030 < p < 0.637 + 0.030) = 0.95$$

The 95% confidence interval for population proportion: $P(0.607 < p < 0.667) = 0.95$

Comment: Then we can be 95 % confident that the interval from 60.7 % to 66.7 % contains the true percentage/proportion of all U.S. citizens who trust the president.

That is, in repeated constrructions of confidence intervals, approximately 95 % of all samples would produce confidence intervals that enclose p.