Confidence intervals for normal distributions (individuals and means)

We have previously discussed the two tailed probability statement of the form

$$P(-Z_0 < Z < Z_0) = 1 - \alpha$$

and we know that $Z_i = \frac{Y_i - \mu}{\sigma}$. If we substitute this into the probability statement above, we can show that,

$$P(-Z_0 \le Z \le Z_0) = P(-Z_0 \le \frac{Y-\mu}{\sigma} \le Z_0) =$$
$$P(-Z_0\sigma \le Y - \mu \le Z_0\sigma) = P(-Y - Z_0\sigma \le -\mu \le -Y + Z_0\sigma) =$$
$$P(Y + Z_0\sigma \ge \mu \ge Y - Z_0\sigma) = P(Y - Z_0\sigma \le \mu \le Y + Z_0\sigma) = 1 - \alpha$$

The last statement gives a function that indicates that the value $Y - Z_0\sigma$ is smaller than, or equal to, μ and $Y + Z_0\sigma$ is larger than, or equal to, μ. This gives an interval which we expect, on average, would encompass the value of μ 100*(1 – α)% of the time. For example if α =0.05, we would expect that with repeated sampling, 95% of the sample estimates of μ would fall between $Y - Z_0\sigma$ and $Y + Z_0\sigma$. This same derivation follows for working with both means and individual observations, so we would expect 100*(1 – α)% of the individual points in a normal distribution to fall in the interval,

$$P(Y - Z_0\sigma \le \mu \le Y + Z_0\sigma) = 1 - \alpha .$$

When working with sample means we may want to know "how good is our estimate of the mean?" We can apply the same interval to means with the formula,

$$P(\overline{Y} - Z_0\sigma_{\overline{Y}} \le \mu \le \overline{Y} + Z_0\sigma_{\overline{Y}}) = P(\overline{Y} - Z_0\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{Y} + Z_0\frac{\sigma}{\sqrt{n}}) = 1 - \alpha .$$

We would interpret this as indicating that, on average, the true mean would fall within this interval 100*(1 – α)% of the time. The value 100*(1 – α)% is called the "confidence value" and the interval is called the confidence interval.

The same derivation works for the t distribution, so for samples we expect 100*(1 – α)% of the observations to fall within the interval

$$P(Y - t_0 s \le \mu \le Y + t_0 s) = 1 - \alpha ,$$

We would expect that, on the average, the true mean would fall within the interval

$$P(\overline{Y} - t_0 s_{\overline{Y}} \le \mu \le \overline{Y} + t_0 s_{\overline{Y}}) = P(\overline{Y} - t_0\frac{s}{\sqrt{n}} \le \mu \le \overline{Y} + t_0\frac{s}{\sqrt{n}}) = 1 - \alpha$$

100*(1 – α)% of the time.

The t distribution would be used when the variance, S^2 is estimated from the sample and the Z distribution used when a known variance value was available (σ^2).

Examples of interval applications on means.

What interval on a population which has $\mu = 22$ and $\sigma^2 = 81$ would include 90% of the individuals in the population?

Answer: P(7.196 < Y_i < 36.804) = 0.900

Place a 70% confidence interval on the estimate of a mean ($\bar{Y} = 43$) from a sample of 20 individuals drawn from a population with known variance $\sigma^2 = 441$ (note that d.f. = 19).

Answer: P(38.133 < $\mu_{\bar{Y}}$ < 47.867) = 0.700

Given a sample with $\bar{Y} = 52$, $S^2 = 81$ and d.f.=25, what interval would you expect to include 98% of the individuals in the population?

Answer: P(29.634 < Y_i < 74.366) = 0.980

Given a sample with $\bar{Y} = 24$, $S^2 = 324$ and d.f. = 25, place an 98% confidence interval on the estimate of the mean.

Answer: P(15.227 < Y_i < 32.773) = 0.980

Confidence intervals for variances

The distribution of variances is described by the Chi square distribution. We have discussed the interval on this distribution where

$$\mathrm{P}(\chi_1^2 \le \chi^2 \le \chi_2^2) = 1 - \alpha$$

Due to the asymmetry and non-negative nature of the Chi square distribution the upper and lower values are not equal with reversed signs as with the t distribution interval. Separate values for χ_1^2 and χ_2^2 are needed.

We also know that $\chi^2 = \frac{(Y_i - \mu)^2}{\sigma^2} = \frac{SS}{\sigma^2}$. This can be substituted into the probability statement above such that,

$$\mathrm{P}(\chi_1^2 \le \chi^2 \le \chi_2^2) = \mathrm{P}(\chi_1^2 \le \frac{SS}{\sigma^2} \le \chi_2^2) = \mathrm{P}(\frac{1}{\chi_1^2} \ge \frac{\sigma^2}{SS} \ge \frac{1}{\chi_2^2})$$

$$= \mathrm{P}(\frac{SS}{\chi_1^2} \ge \sigma^2 \ge \frac{SS}{\chi_2^2}) = \mathrm{P}(\frac{SS}{\chi_2^2} \le \sigma^2 \le \frac{SS}{\chi_1^2}) = 1 - \alpha$$

This last statement gives a confidence interval for variances.

Examples of interval applications on variances.

Place a 95% confidence interval on the estimate of the variance for a sample with $\bar{Y} = 21$, $S^2 = 196$ and d.f.=18. (Note : SS = 196*18)

Answer: P(111.906 < σ^2 < 428.637) = 0.950