Confidence Intervals

Confidence Intervals are an extension of the concept of Margin of Error which we met earlier in this course.

Remember we saw:

- The sample proportion will differ from the population proportion by more than the margin of error less than 5% of the time.

- The Margin of Error for a sample of size n is $1/\sqrt{n}$

Suppose we wanted to estimate a Population Mean rather than a Population Proportion, can we compute a Margin of Error for the Mean?

Confidence Intervals provide us with the answer.

A Confidence Interval is an interval of numbers containing the most plausible values for our Population Parameter.

The probability that this procedure produces an interval that contains the actual true parameter value is known as the Confidence Level and is generally chosen to be 0.9, 0.95 or 0.99.

Confidence Intervals take the form:

Point Estimate +/- Critical Value x Standard Error

LARGE SAMPLE CONFIDENCE INTERVAL FOR A POPULATION MEAN

GENERAL FORMULA

$$\bar{x} \pm (z \text{ critical value}) \frac{\sigma}{\sqrt{n}}$$

The level of confidence determines the *z* critical value.

99%	2.58
95%	1.96
90%	1.645

Since *n* is large the unknown σ can be replaced by the sample value *s*.

$$\bar{x} \pm (z \text{ critical value}) \frac{s}{\sqrt{n}}$$

Example

A random sample of 225 1st year statistics tutorials was selected from the past 5 years and the number of students absent from each one recorded. The results were $\bar{x}$ =11.6 and *s*=4.1 absences.

Estimate the mean number of absences per tutorial over the past 5 years with 90% confidence.

90% CI for μ is

$$\bar{x} \pm 1.645\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$11.6 \pm 1.645\left(\frac{4.1}{\sqrt{225}}\right) = 11.6 \pm 0.45 = (11.15, 12.05)$$

INTERPRETATION:

It is incorrect to say that there is a probability of 0.90 that μ is between 11.15 and 12.05. In fact this probability is either 1 or 0 (μ either **is** or **is not** in the interval).

The 90% refers to the percentage of all possible intervals that contain μ i.e. to the estimation process rather than a particular interval.

It is also incorrect to say that 90% of all tutorials had between 11.15 and 12.05 missing students.

SMALL SAMPLE CONFIDENCE INTERVAL FOR A POPULATION MEAN

- Consider only samples from **populations which are (approx) normal.**
- Use the distribution of $\dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ which is known as the *t* distribution with *(n-1)* **degrees of freedom (df).**

Note: As the df increases the *t* curve approaches the *z* curve.

Areas under the t curve are tabulated in tables 9 & 10 of NCST (note ν=df)

A small sample confidence interval for μ is

$$\bar{x} \pm (t \text{ critical value})\frac{s}{\sqrt{n}} \qquad \text{with df} = n-1$$

NOTE:
This confidence interval is appropriate for small samples **ONLY** when the population distribution is normal.

Example

Sample of 15 test-tubes tested for number of times they can be heated on Bunsen burner before they cracked gave $\bar{x}$ =1,230, *s*=270 . Construct 99% confidence interval for μ.

df=n-1=15-1=14

for 99% confidence t=2.977

$$\bar{x} \pm (t \text{ critical value}) \frac{s}{\sqrt{n}} = 1{,}230 \pm 2.977 \frac{270}{\sqrt{15}} = (1{,}020, 1{,}440)$$

LARGE SAMPLE CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

In many populations each item belongs to one of two categories (S & F). We are interested in estimating the proportion (or percentage) of the population who belong to each category.
e.g. proportion of adults who smoke cigarettes, proportion who vote for FF, proportion who drink Budweiser etc.

Example

Each year 1st year science students may (S) or may not (F) choose to study Statistics. To estimate the fraction who do study Statistics a sample of 1000 students was chosen from the past 10 years and 637 had chosen Statistics as a 1st year subject.

If the fraction of students who choose Statistics is *p* then randomly selecting somebody gives a probability *p* of S and *1-p* of F. The obvious solution is to use the sample proportion ($\hat{p}$) to estimate the population proportion (*p*) where

$$\hat{p} = \frac{\text{number of S's in sample}}{\text{Sample size}} = \frac{637}{1000} = .637$$

The sampling distribution of $\hat{p}$ is approx normal with

$\mu_{\hat{p}}=p$ and $\sigma^2_{\hat{p}}$=p(1-p)/n

Consequently the confidence interval for p is

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{p(1-p)}{n}}$$

Approximate p with $\hat{p}$ to get

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

for the approximate confidence interval.

For our example the 95% confidence interval for p is

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (.607, .667)$$

Example

The Graduate Schools of Business and Law and Human Sciences take a random sample of recent PhD graduates (n=135) and finds that 12 of these are unemployed. Compute a 90% confidence interval for the proportion of all PhD graduates who fail to find a job.

$\hat{p}$ =12/135=.089

Check for large sample

$$\hat{p} \pm 3\sigma_{\hat{p}} \approx \hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (.015, .163)$$

90% confidence interval for *p*

$$\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}} \approx \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (.049, .129)$$

Suppose we know that the unemployment rate in the country is 3%. Can we conclude that PhD Graduates are more likely to be unemployed than the population in general?

Since .03 lies outside the confidence interval it is not consistent with our data and we would conclude that the unemployment rate among PhD graduates is above that of the population as a whole.