Lecture 30 Continuing confidence intervals.

What if the population is not normal?

In this case, provided the sample size is large enough, the CLT will kick in ie

$$\bar{X} \text{ is approx } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Also note that since σ is typically unknown we replace it with a sample standard deviation S. Thus the standardised r.v.

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \text{has approx } N(0,1) \text{ dist.}$$

which following our discussion from last time implies that

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}} \quad \text{is a large sample}$$

$100(1-\alpha)\%$ CI for μ regardless of the original distribution. (Your book says $n \geq 40$ is a reasonable rule of thumb).

Confidence intervals for p (the population proportion)

Recall that when n is large

$\hat{p}$ has approx $N(p, \sqrt{p(1-p)/n})$ distribution.

So

$$P\left(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \simeq 1-\alpha$$

Following the discussion/method from last lecture this leads to

$$P\left(\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n^2}}}{1+(z^2_{\alpha/2})/n} < p < \frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\alpha/2}}{4n}}}{1+(z^2_{\alpha/2})/n}\right) \simeq 1-\alpha$$

Note that as n becomes large $z^2_{\alpha/2}/2n$, $z^2_{\alpha/2}/n$ and $\frac{z^2_{\alpha/2}}{4n}$ all become quite small.

which means a satisfactory interval is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

This is the traditional formula given in older textbooks. Recent research has shown the first formula to be superior.

<u>One sided intervals</u> Sometimes we only care about one sides.

Form

$$(-\infty, u)$$

or

$$(l, \infty)$$

where u – upper bound
l – lower bound.

For the mean (large sample)

$$\mu < \bar{X} + z_\alpha \frac{s}{\sqrt{n}}$$ upper confidence bound

$$\mu > \bar{X} - z_\alpha \frac{s}{\sqrt{n}}$$ lower confidence bound.

What if σ is unknown and the population is normal?

Assume the sample is from a $N(\mu, \sigma)$ distribution, μ, σ unknown

If we have $X_1, \ldots, X_n$ from which we can calculate $\bar{X}$ and S. Large sample theory says that the r.v

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$ has approx $N(0,1)$ distribution.

What if n is small? There is another distribution which we use instead called the "t" distribution. In particular the r.v

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$ has the t distribution with $n-1$ degrees of freedom

Properties of the t-distribution

1. It is bell shaped and centered at 0
2. It is more spread out than the standard normal curve.

3. It has only one parameter ν (the degrees of freedom) which takes integer values 1, 2, 3, ...
4. As ν increases the spread becomes less
5. as $\nu \to \infty$ the density curve approaches the standard normal curve.

CI for μ with σ known.

A level $100(1-\alpha)\%$ CI for μ is given by

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2, n-1}$ means value of t distribution such that:

ie

$$P(T > t_{\alpha/2, n-1}) = \frac{\alpha}{2}$$