Lecture 12

Other hypothesis tests and Confidence intervals

In the previous lectures we looked at testing hypotheses about means. The test procedure is the same for other quantities but the distributions might be different or there might be some slight variation.

Proportions

We have a single sample of size n, X of which possess the characteristic. The model here is X~B(n,p). p is unknown.

Example.

Each of 400 circuit boards was tested. 27 were found to have faults on them. The process is "in control" if no more 5% have faults. Is there evidence that the process is out of control?

The proportion is actually a mean. If we code the faulty boards as 1 and the OK as 0. $\bar{x}$ is the proportion of 1s in the sample.

However with n known the Binomial is a one parameter distribution. The variance is *np(1-p)*.

$$\hat{p} = \frac{X}{n} \text{ so } \qquad V(\hat{p}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Thus

H_0: $p = p_0$

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \text{ as N(0,1).}$$

and reject H_0 if $T > c_\alpha$, $T < c_\alpha$ or $|T| > c_\alpha$ depending on H_1.

Note: That we use the hypothesised $p_{0,}$ (not $\hat{p}$), in the estimate of variance (denominator) . That is because the T statistic is computed on the assumption that H_0 is true.

On the other hand when constructing a confidence interval we use $\hat{p}$, because that is our best estimate of p.

$$\text{CI} \quad \hat{p} \pm C_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Example:

$$T = \frac{\frac{27}{400} - 0.05}{\sqrt{\frac{0.05 \times 0.95}{400}}} = 1.606 \qquad \text{p-value-0.0541}$$

Just fail to reject H_0 at 5%.

For the confidence interval $\sqrt{\frac{0.0675 \times 0.9325}{400}} = 0.0125$

So CI = 0.0675 +/- 2* 0.0125 (95 % confidence)

(0.0425,0.0925) roughly between 4% and 9%.

You need a lot of observations to estimate proportions accurately!

The procedure here is based on the CLT approximation of the Normal to the Binomial. For $0.2<p<0.8$ the approximation is very good but here the p are much smaller so we might have a problem.

With new algorithms and computers (Excel) it is no problem to compute exact tests and confidence intervals.

If X~B(400,0.05) then

$P(X \leq 26) =$ BINOMDIST(26,400,0.05,1) = 0.927

Thus $P(X \geq 27) = 1 - 0.927 = 0.073$

this is the exact p-value for the test

The exact CI is trickier:

The confidence interval <u>excludes</u> those low p value for which:

$$\text{BINOMDIST}(27,400,p,1) \leq 0.025$$

p=0.047 gives 0.02491 lower limit.

and those high values of p for which :

$$1- \text{BINOMDIST}(27,400,p,1) \leq 0.025$$

p=0.094 gives 0.02386

$$1 – 0.02491-0.02386=0.95123$$

So (0.047,0.094) is a 95.1% CI – not worth the bother for extra accuracy.

If n was smaller or p even more extreme then it might be important.

Difference of two proportions.

Two opinion poles were conducted one year apart. The sample sizes in each of the poles were 1000.
In the first pole 40.2% supported a certain party in the second 37.1% supported this party.
Has the support eroded in the intervening period?

H_0: $p_1 = p_2$ or $p_1 - p_2 = 0$.

Is it one tail or two tail?
The question to be answered suggests a one tailed test $H_1 : p_1 > p_2$.
But ... Has the question been suggested by the data? Would we still ask the same question if the second poll came out as 43%? If the answer to this is NO. Then two tailed test –otherwise you are using the data twice!

We should test has the support *changed* over the year.

H_1: $p_1 \neq p_2$

$$T = \frac{\hat{p}_1 - \hat{p}_2}{sd(\hat{p}_1 - \hat{p}_2)}$$ then same format as before.

The calculation of the standard deviation is again a little different.

If H_0 is true then – the p equal in both polls. The two samples are basically a sample of size 2000 with 402 + 371 = 773 supporting the party.

We get a pooled estimate of p:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$ (note this will only equal the "average p" if the sample sizes are equal.

$$sd(\hat{p}_1 - \hat{p}_2) = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1 - \hat{p})}$$

If we want to construct a CI for the difference we have yet another variation:

$$sd(\hat{p}_1 - \hat{p}_2) = \sqrt{\left(\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}\right)}$$

But note that looking at the difference alone when the p's are not the same is not a good idea. Suppose a difference is equal to 0.1. The two cases

p_1=0.5 and p_2=0.4 - the p's are not very different.

p_1=0.001 and p_2=0.1001 the p's are very different

(1) – virtually nobody,
(2) approximately 1/10 reasonably common.

A linear scale is not good for proportions. (close to 0 or 1) a more sophisticated approach is required.

Testing Variances.

Testing whether the variance of a quantity has a certain value is quite uncommon in practice. Variance measures the stability of a production process, its precision. The context might be a situation where the mean is not at issue - it is known to be correct does the precision meet specs.
Confidence intervals for variance express the extent of precision achieved by a process.

In test (for mean) power calculations based on estimated variances we might take a pessimistic view and use the upper limit on the variance rather than the estimated value. (to be sure, to be sure).

Comparing variances would also be pretty rare. Suppose the means are equal the smaller variance process is better.
However a most important application of this arises in testing the equality of more than two means – this is done by comparing estimates of variance, but again this beyond a first course.

The tests are based on the distribution of the estimate of variance.

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

The result is $(n-1)\dfrac{\hat{\sigma}^2}{\sigma^2} \sim \mathrm{X}_{n-1}^2$ pronounced Chi-squared (ki) . The distribution is a special case of the Gamma distribution X_n^2 is $Gamma(\frac{1}{2},\frac{n}{2})$.

When the <u>mean is known </u>the appropriate estimate is:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \text{ and then } \; n\frac{\tilde{\sigma}^2}{\sigma^2} \sim \mathrm{X}_n^2$$

The distribution is tabulated and in Excel

$P(\mathrm{X}_n^2 > x) =$CHIDIST(x,n)

CHIINV(p,n) returns x such that $P(\mathrm{X}_n^2 > x) = p$.

Example
A machine produces ball bearings 5 mm in diameter. Applications require that practically all the ball bearings should be within 4.91mm and 5.09 mm. Data have been collected and produced the following summaries:
$\sum x^2$ =2500.124 , $\sum x$ =499.97, n=100.

Assume normal distribution of bearing diameter.
Assume mean = 5mm.
The requirement "all the bearings" within 0.09 of the mean we translate into $3\sigma \le 0.09$ or $\sigma^2 \le 0.03^2 = 0.0009$

H_0 : $\sigma^2 = 0.0009$ $\qquad$ H_1 $\sigma^2 > 0.0009$

Estimate with known mean

$$\tilde{\sigma}^2 = \frac{2500.124 - 100 \times 5^2}{100} = 0.00124$$

Thus the estimate is greater than the hypothesised value.

The statistic is

$$s = n\frac{\tilde{\sigma}^2}{\sigma^2} = 100 \times \frac{12.4}{9} = 137.78$$

CHIDIST(137.78,100)=0.007347

we reject H_0 at a significance level of 0.7%

Confidence interval for the variance.
Same data as for test but just for variety we will use the estimated mean.

$$\hat{\sigma}^2 = \frac{1}{99}(2500.124 - 100 \times 4.9997)=0.003475$$

Construction:

$$\frac{\hat{\sigma}^2}{\sigma^2}(n-1) \sim \mathrm{X}^2_{n-1}$$

For a $1-\alpha$ confidence interval :

$$CHIINV(1-\alpha/2, n-1) \le \frac{\hat{\sigma}^2}{\sigma^2}(n-1) \le CHIINV(\alpha/2, n-1)$$

Note: CHIINV here is as defined in Excel P(X>x) .

More usually it would be defined as P(X<x) so the 1-$\alpha/2$ and $\alpha/2$ would swap places.

$$\frac{CHIINV(1-\alpha/2, n-1)}{n-1} \le \frac{\hat{\sigma}^2}{\sigma^2} \le \frac{CHIINV(\alpha/2, n-1)}{n-1}$$

Taking reciprocals reverses the direction of the inequalities:

$$\frac{n-1}{CHIINV(1-\alpha/2, n-1)}\hat{\sigma}^2 \ge \sigma^2 \ge \frac{n-1}{CHIINV(\alpha/2, n-1)}\hat{\sigma}^2$$

For a 95% confidence interval with n=100 we have:

CHIINV(0.975,99)=73.3611
CHIINV(0.025,99)=128.4219

$$1.35 * 0.003475 \geq \sigma^2 \geq 0.77*0.003475$$

$$0.2676*10^{-2} \leq \sigma^2 \leq 0.4691*10^{-2}$$

Although rarely given the confidence interval for the standard deviation is probably more interesting.

$$0.052 \leq \sigma \leq 0.068.$$

We can see from this that the SD is nearly twice of what is required (0.03).