Hypothesis Testing and Confidence Intervals II.

(Two Sided Tests, 't' Tests, Tests for Variances, Goodness of Fit tests, P-Values.)

The hypotheses considered in the previous chapter addressed the issue of whether or not the true population mean was greater (less) than or no greater (less) than a particular value and the confidence intervals addressed a similar one sided issue. Similar principles can be applied to addressing the questions "Is the unknown population mean equal to a particular value or not?" or "What is the interval within which I could expect the unknown population mean to lay with some pre-specified probability?". In this process there will no longer be a question of which way round to pose the test, the null hypothesis will always be that the unknown population mean is equal to a particular value and the alternative will be that it is not. Then the testing issue becomes one of "Is the sample mean sufficiently far below, or sufficiently far above the null hypothesized population mean to warrant its rejection?". The notions of Type I and Type II Errors still remain, the former being to conclude that the population mean is not equal to the null hypothesised value when in truth it is, the latter being to conclude that the population mean is equal to the null hypothesised value when in truth it is not. Similarly the notions of power and size also retain their meanings.

Constructing the Test.

Now $\bar{X}$ will constitute evidence against the null if it is sufficiently far away from μ_0 which could be either less than or greater than the null hypothesized value. Effectively two critical values have to be determined, a lower one (C_L) and an upper one (C_U). The decision rule will be if $\bar{X} < C_L$ or if $\bar{X} > C_U$ reject H_0, otherwise accept H_0. Again, after choosing α for the probability of a type one error (which will occur when $\bar{X} < C_L$ or $\bar{X} > C_U$ when H_0 is true), the upper and lower critical values are determined as follows:

$$\alpha = P(\bar{x} < C_L\ ,\ \bar{x} > C_U \mid H_0\ true)$$

$$\rightarrow 1-\alpha = P(C_L \le \bar{x} \le C_U)$$

$$= P\left(\frac{C_L - \mu_0}{\sigma/\sqrt{n}} \le \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \le \frac{C_U - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{C_L - \mu_0}{\sigma/\sqrt{n}} \le Z \le \frac{C_U - \mu_0}{\sigma/\sqrt{n}}\right)$$

which in turn will be true when:

$$\frac{C_L - \mu_0}{\sigma/\sqrt{n}} = Z_\gamma$$

$$\frac{C_U - \mu_0}{\sigma/\sqrt{n}} = Z_{1-\delta}$$

$$\gamma + \delta = \alpha$$

In fact the critical values are usually based upon $\gamma = \delta = \alpha/2$ since this produces the shortest distance between the two critical values (the narrowest acceptance region) but generally any values such that $\gamma + \delta = \alpha$ would be appropriate. Making the upper and lower critical regions have equal probability yields upper and lower critical values as:

$$C_U,\ C_L = \mu_0 \pm \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}}$$

The two sided confidence interval follows in exactly the same fashion as with the one sided confidence intervals derived earlier so that the answer to the question "Given the sample mean and a known variance σ^2, what is an interval within which one could expect the unknown population mean to lay with $1-\alpha$ probability?" is given by:

$$\mu^*_U,\ \mu^*_L = \bar{x} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}}$$

which are the critical values for the correspondingly sized hypothesis test with the null hypothesized mean replaced by the sample mean.

"t"Tests.

In all of the hypothesis tests we have considered until now we have presumed to know the variance σ^2, but in practice invariably this will not be the case. We have available an estimate of the variance from our random sample given by:

$$\overline{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

Can it simply be substituted into the various formulae for the test critical values and confidence intervals we have derived? The answer is more or less but not exactly. Recall that in deriving our tests and confidence intervals we were transforming the sample mean into a standard normal random variate by subtracting its mean and dividing by its standard deviation. This is simply a linear transformation employing constants and one of the remarkable properties of normal random variables is that they remain normal under constant or non-stochastic linear transformations. However, the sample based estimate of the variance is itself a random variable and not a constant so that the transformation is no longer non-stochastic and the transformed mean does not retain its normality. Fortunately its distribution has been worked out by William Gossett to be a "t" distributed random variable with n-1 degrees of freedom[1]. Thus the one sided hypothesis test critical value and its corresponding confidence bound given in the previous chapter are respectively given by:

$$C_U = \mu_0 + \frac{\overline{\sigma}}{\sqrt{n}}.t_{1-\alpha}(n-1)$$

$$\mu* = \overline{X} + \frac{\overline{\sigma}}{\sqrt{n}}.t_{1-\alpha}(n-1)$$

---

[1]Details of the "t" distribution appear in the appendix to the previous chapter.

and the two sided test critical values and confidence intervals in this chapter are given by:

$$C_U, C_L = \mu_0 \pm \frac{\overline{\sigma}}{\sqrt{n}} t_{(1-\frac{\alpha}{2})}(n-1)$$

$$\mu^*_U, \mu^*_L = \bar{x} \pm \frac{\overline{\sigma}}{\sqrt{n}} t_{(1-\frac{\alpha}{2})}(n-1)$$

So the formulae essentially stay the same with the estimated standard deviation and t distributed variate being respectively substituted for the known standard deviation and standard normal variate. As pointed out in the appendix of the previous chapter, as the degrees of freedom become large the t distribution tends toward the standard normal. In fact for degrees of freedom above 50 the standard normal and t are pretty close already so that when large samples are involved there is very little difference between the two.

Tests and confidence intervals for Variances.

The magnitude of the variance may be of interest and, just as in the case of the mean, tests and confidence intervals of both the one sided and two sided variety are available. In this case the χ^2 distribution will provide the basis for the comparisons, since it can be shown that for $X \sim N(\mu\sigma^2)$:

$$\frac{(n-1)\overline{\sigma^2}}{\sigma^2} = \frac{\sum_{i=1}^{n}(X - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

So that for an upper tailed test of size α for H_0: $\sigma^2 \leq \sigma^2_0$ against H_1: $\sigma^2 > \sigma^2_0$, a critical value C may be obtained from:

$$\alpha = P(Type\ I) = P(\overline{\sigma^2} > C_{|H_0\ is\ true})$$

which will be true when:

$$C = \frac{\sigma_0^2}{n-1}\chi_{1-\alpha}(n-1)$$

In a similar fashion a lower tailed test of size α for H_0: $\sigma^2 \geq \sigma^2_0$ against H_1: $\sigma^2 < \sigma^2_0$, a critical value C may be obtained from:

$$\alpha = P(Type\ I) = P(\overline{\sigma^2} < C_{|H_0\ is\ true})$$

which will be true when:

$$C = \frac{\sigma_0^2}{n-1}\chi_\alpha(n-1)$$

Again following the logic of previous arguments lower and upper 1-α Confidence Bounds for the value of the variance would be given by when:

$$\frac{\overline{\sigma^2}}{n-1}\chi_{\alpha/2}(n-1) \ , \ \frac{\overline{\sigma^2}}{n-1}\chi_{1-\alpha/2}(n-1)$$

In a similar fashion a two sided test of size α for H_0: $\sigma^2 = \sigma^2_0$ against H_1: $\sigma^2 \neq \sigma^2_0$, two critical values C_L and C_U may be obtained from:

$$C_L = \frac{\sigma_0^2}{n-1}\chi_{\alpha/2}(n-1) \ , \ C_U = \frac{\sigma_0^2}{n-1}\chi_{1-\alpha/2}(n-1)$$

Goodness of Fit Tests.

Sometimes rather than being interested in population means and variances an investigator is interested in whether or not the population distribution takes a particular form f(x) (where f(x) is known). In this case we resort to goodness of fit tests. Given a random sample X_i, i = 1,..,n drawn from a distribution f(X), the domain of X is split up into K mutually exclusive and exhaustive intervals and the two basic formulae for the goodness of fit test are then given by:

$$GF = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = n\sum_{k=1}^{K} \frac{(p^O{}_k - p^E{}_k)^2}{p^E{}_k}$$

where O_k is the observed number of elements in the sample in the k'th interval ($p^0{}_k$ being the corresponding proportion) and E_k is the expected number of elements in the k'th interval under the null hypothesis ($p^E{}_k$ being the corresponding proportion). Obviously O_k is obtained by simply counting how many observations arrise in the k'th interval, E_k is evaluated by computing:

$$E_k = n(F(b) - F(a))$$

where a and b are the upper and lower limits of the k'th interval and F( ) is the Cumulative distribution function of f( ). Note that $p^0{}_k$ and $p^E{}_k$ can be respectively obtained by dividing O_k and E_k by n.

GF is a χ^2(K-1-m) statistic where K is the number of intervals and m is the number of parameters to be estimated when the null distribution f( ) is not fully specified. Clearly the lower bound for GF is 0, it occurs when what is observed is perfectly coincident with what is expected in all categories. On the other hand when there is a large divergence between what is observed and what is predicted by the theory GF will be a large positive number. Consequently GF will be used in the context of a one sided upper tailed test where H_0: f(X) is the true distribution and H_1: f(X) is NOT the true distribution a test of size α would reject the null if GF > $\chi^2_{(1-\alpha)}$(K-1-m) otherwise the null is to be accepted.

As for implementing the test, there are some practical considerations. Regarding choosing the mutually exclusive and exhaustive partitions two basic rules of thumb should be acknowledged. The partitions should be chosen so that $E_k \geq 5$ for all k and they should be chosen so that the various intervals are close to equiprobable. Clearly this means that no more than n/5 intervals should be entertained and generally n/10 would be a better guide.

"P" Values.

Many standard statistical software packages provide "P" values for the standard statistics they calculate. Basically the "P" value for a given statistic (be it a sample mean or a sample variance) is the probability under the null that you would observe a statistic greater than that value so that for a sample mean under the H_0: $\mu \leq \mu_0$ the "P" value is given by:

$$P = P(a \ sample \ mean \geq \overline{X} \mid H_0)$$

so that if P is less than the size of the test then the null would be rejected.