Weekly Review 7

Last week we discussed the question: Is $\mu \neq \mu_0$? This leads to the *two-sided* alternative hypothesis:

$$\begin{aligned} \mathrm{H}_0 &: \mu = \mu_0 \\ \mathrm{H}_A &: \mu \neq \mu_0 \end{aligned}$$

by using z- or t-table for the case that σ is known and that σ is unknown and estimated by s. Whenever a testing procedure uses t-table, it is often called a *t-test*.

Now consider the question: Is μ greater than μ_0? Statistically, we test a *one-sided* alternative hypothesis:

$$\begin{aligned} \mathrm{H}_0 &: \mu \leq \mu_0 \\ \mathrm{H}_A &: \mu > \mu_0 \end{aligned}$$

I have emphasised that **the null hypothesis must contain the equality sign**. Thus, no matter the question is asking: "Is $\mu > \mu_0$?" or asking: "Is $\mu \leq \mu_0$?", the null hypothesis must be $\mu \leq \mu_0$. Also, this means that we have to be very careful with the English language here: the phrase "greater than/less than" does not contain the equality sign and hence is the alternative hypothesis, while "not less than/not greater than" contains the equality sign and hence is the null hypothesis. In other words, when you have two hypotheses, the one containing the equality sign will be the null, and the other one will then be the alternative.

In the same spirit as the argument used above, we first assume H_0 is true and then check whether the test statistic calculated from the given sample is too unusual under H_0. If yes, then we reject H_0. How unusual is too unusual? It is too unusual when 95% of the time [or $100(1-\alpha)\%$ of the time] it would not happen. In this story, depending on whether σ is known or is unknown respectively, it is too unusual (or too extreme) if

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{or} \quad t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

is too large. Although we say we assume H_0 is true, we use only μ_0 in the calculation because if using μ_0 in the test statistic, it is still too large, then any other values less than μ_0 (i.e. any other values satisfying H_0) would only give us even larger test statistic. Thus, assuming $\mu = \mu_0$ is assuming the worst scenario or assuming the value closest to H_A. This will lead to the most conservative way to measure the strength of the evidence. If under this worst scenario (i.e. evaluated in the most conservative way) we still find this evidence strong enough to reject H_0, then we can comfortably reject H_0, because using other values satisfying H_0 in the test statistic will result in even larger test statistic values.

OK, I still have not yet answered how large is too large. To answer this question, we have to know the distribution of the test statistic and once we know it, then it is straightforward: if it follows the standard normal (which happens when σ is known), then reject H_0 at the α significance level whenever $z \geq z_\alpha$; if it follows the t-distribution with degrees of freedom $n-1$ (it happens when σ is unknown and is estimated by s), then reject H_0 at the α significance level whenever $t \geq t_\alpha$. It is because by the definition of z_α and t_α, we know that $100(1-\alpha)\%$ of the time the test statistic would be on the left-hand side of the critical value z_α or t_α. Thus, if the event $\{z \geq z_\alpha\}$ or $\{t \geq t_\alpha\}$ really happens, then it is too unusual (having chance no more than α) under H_0, and so it is reasonable to conclude that the calculation of the probability is actually incorrect, because if the probability were really so small, the event should not have happened. The only place where the calculation can be wrong is the assumption that H_0 is true. Thus, rejecting H_0 is a very reasonable conclusion.

Notice that in the above procedure when we say we assume H_0: $\mu \leq \mu_0$, as a matter of fact the calculation assumes only $\mu = \mu_0$ and so the testing procedure for a *composite* null hypothesis hypothesis (specifying a range of possible values)

$$\begin{aligned} H_0 &: \mu \leq \mu_0 \\ H_A &: \mu > \mu_0 \end{aligned}$$

will be identical to the testing procedure for a *simple* null hypothesis (specifying only one single possible value)

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_A &: \mu > \mu_0 \end{aligned}$$

The only difference is in the meaning of the significance level. For the latter (simple null hypothesis), the significance level α is equal to the probability of committing type I error, while for the former (composite null hypothesis), then using μ_0 in the calculation is only the worst scenario and so the worst (largest) probability of committing type I error is α, i.e. α is the upper bound of the type I error probability.

Comparing the procedure for testing a one-sided alternative hypothesis with that for testing a two-sided alternative, we can see that the formulae for the test statistic are the same but the critical values are not. Therefore, it is also important to write down both the null hypothesis and the alternative hypothesis explicitly beforehand, especially the latter, because, for instance, in the case σ is known and $\alpha = 0.05$, if the test statistic z is say 1.75, then for the two-sided alternative H_A: $\mu \neq \mu_0$ we will not reject H_0 because $1.75 < 1.96$ and so is not too large. However, for the one-sided alternative H_A: $\mu > \mu_0$, then we will reject H_0 because $1.75 > 1.645$ and so is too large. Thus, if you write down the alternative hypothesis (or you fix α) after calculating the test statistic, you may be able to choose a conclusion according to your own wish (i.e. you can conclude whatever you want to conclude)!

Remember: We are checking whether our evidence is strongly supporting the alternative hypothesis, and hence we have to know what the alternative hypothesis is beforehand so that we know how to measure the strength of our evidence. That is to say, we have to fix our decision rule before we collect our data.

By the same token, if the alternative hypothesis is $\mathrm{H}_A : \mu < \mu_0$, we should reject the null hypothesis H_0 when the observed sample mean is much smaller than the value μ_0 assumed in the null hypothesis. We reject the null hypothesis at the α significance level whenever the test statistic is too negative, i.e. whenever $z \leq -z_\alpha$ if we know σ. If we do not know the population standard deviation σ, we estimate it by s and so the test statistic will be t. We reject the null hypothesis at the α significance level whenever $t \leq -t_\alpha$ where the degrees of freedom is $n-1$.

To summarise, the idea of the decision described above is, therefore, to investigate whether the sample provides strong evidence to support the alternative hypothesis. More precisely, we reject the null hypothesis if the test statistic is *more extreme* than the critical value, where the critical value depends on the alternative hypothesis. A value is more extreme if it suggests that H_A is more likely to be true. For testing hypotheses concerning a population mean, a two-side alternative hypothesis will lead to two critical values, and so we have a two-sided (or *two-tailed*) test; a one-sided alternative hypothesis leads to one critical value, and so we have a one-sided (or *one-tailed*) test.

In conclusion, suppose we are testing the null hypothesis

$$\mathrm{H}_0 : \mu = \mu_0,$$

against one of the three different alternative hypotheses:

$$\mathrm{H}_A : \mu \neq \mu_0, \qquad \mathrm{H}_A : \mu > \mu_0, \qquad \text{or} \qquad \mathrm{H}_A : \mu < \mu_0.$$

The test statistic is simply

$$\begin{aligned} z &= \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}, && \text{if we know } \sigma, \\ t &= \frac{\overline{x} - \mu_0}{s/\sqrt{n}}, && \text{if we do not know } \sigma, \end{aligned}$$

and the decision rule is:

H_0	H_A	(Critical region) Reject H_0 at the α significance level if
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z\rvert \geq z_{\frac{\alpha}{2}}$ if we know σ, $\lvert t\rvert \geq t_{\frac{\alpha}{2}}$ if we do not know σ,
$\mu = \mu_0$ or $\mu \leq \mu_0$	$\mu > \mu_0$	$z \geq z_\alpha$ if we know σ, $t \geq t_\alpha$ if we do not know σ,
$\mu = \mu_0$ or $\mu \geq \mu_0$	$\mu < \mu_0$	$z \leq -z_\alpha$ if we know σ, $t \leq -t_\alpha$ if we do not know σ,

where $t_{\frac{\alpha}{2}}$ and t_α have degrees of freedom $n-1$.

If we find strong evidence against the null hypothesis, we reject it; after rejection, we conclude that the true mean μ is *significantly different from* μ_0 (when H_A is $\neq$), *significantly greater than* μ_0 (when H_A is $>$), or *significantly less than* μ_0 (when H_A is $<$). In particular, when $\mu_0 = 0$, then "*μ is significantly different from zero*" will often be shortened to just "*μ is significant*".

Sometimes when we are testing hypotheses, we may want to calculate the *p-value* which is the lowest significance level for which the null hypothesis could have been rejected. It can also be considered as the probability (if we take another random sample) of getting the observed value or "more extreme" values; a value is "more extreme" when it provides even stronger evidence than the observed value to support the alternative hypothesis. (Graphically, it will be equal to the area to the right or to the left of the test statistic under the probability density function of the test statistic, if the alternative is "$>$" or "$<$", respectively; when the alternative is "$\neq$", the p-value is the double of smaller one between the area to the right and the area to the left.) *The calculation of the p-value depends on which parameter we are testing and also depends on the alternative hypothesis*, but the decision rule does not depend on the alternative hypothesis and is always as follows. **If the p-value is less than or equal to α, we reject the null hypothesis at the α significance level**. This decision rule is a general principle, which can be applied to all tests of hypotheses.

In practice, however, without a computer, we are able to calculate the p-value only if the distribution of the test statistic is the standard normal (or other distributions, such as binomial, that we know how to calculate probabilities by pocket calculators); for the t-distribution, the critical values for different degrees of freedom are tabulated as the t-table, but in order to calculate probabilities without a computer, we need one table (similar to the standard normal table) for each value of the degrees of freedom; this is obviously not practically feasible.

After talking about continuous data, we considered binary data. If a population consists of 0's (failures) and 1's (successes) only, then the parameter of interest is the proportion p of successes. In Section 8.4 we already discussed the construction of confidence intervals for p. Now in Section 9.4 we consider a significance test for the population proportion. The null hypothesis is simply

$$\mathrm{H}_0 \colon p = p_0,$$

where p_0 is an explicitly hypothesised known value, and the three different alternative hypotheses are:

$$\mathrm{H}_A \colon p \neq p_0, \qquad \mathrm{H}_A \colon p > p_0, \qquad \text{or} \qquad \mathrm{H}_A \colon p < p_0.$$

As mentioned in the previous example (of testing hypotheses concerning a population mean), if the alternative is $>$ or $<$, the null hypothesis can be $\leq$ or $\geq$, respectively. (Why?)

Suppose the sample proportion is $\hat{p}$. Obvious, we would like to know whether $\hat{p}-p_0$ is too unusual or not. In order to tell whether it is too unusual, we have to know its distribution. This consideration leads to the following test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}.$$

Note that the philosophy of hypothesis testing is that we first assume the null hypothesis is true and calculate the value of the test statistic. If the value of the test statistic is too large or/and too small, then it is unreasonable and so our assumption (i.e. the null hypothesis) is probably wrong. Therefore, in the above formula, the denominator contains $p_0(1-p_0)$ instead of $\hat{p}(1-\hat{p})$, which was used in the formula of confidence interval for p; it is because we assume that $\mathrm{H}_0\colon p = p_0$ is true in this context, while for confidence interval we of course would not assume any value for the parameter that we are estimating. Now, if the alternative is $\neq$, $>$ and $<$, respectively, we reject the null hypothesis at the α significance level whenever $|z| \geq z_{\frac{\alpha}{2}}$, $z \geq z_\alpha$ and $z \leq -z_\alpha$, respectively.

To summarise, we have:

H_0	H_A	Reject H_0 at the α significance level if
$p = p_0$	$p \neq p_0$	$\lvert z\rvert \geq z_{\frac{\alpha}{2}}$
$p = p_0$ or $p \leq p_0$	$p > p_0$	$z \geq z_\alpha$
$p = p_0$ or $p \geq p_0$	$p < p_0$	$z \leq -z_\alpha$

This is the end of Chapter 9. Please read Chapter 9 repeatedly until you understand everything explained in this (and the last two pages of the previous) review and until you are able to work out some numerical exercises in the book. This is very important for the rest of this course.

Cheers,
HENG PENG