Weekly Review 9

It is important but not so straightforward for beginners to distinguish two independent samples from *paired-sample*, which can be described as follows. We have paired-population, e.g. pre- and post-treatment, left- and right-arms, husband and wives, twin brothers, test and exam scores, etc. For paired-data we are still interested in the difference of the means but the difference of the means is the same as the mean of differences for paired-data. Thus, the *paired-sample t-test* is the same as the t-test for one sample of pairwise differences. So, what we have to do is to take all pairwise differences. These differences form one sample and we are interested in the population mean μ_{diff}, which is just equal to $\mu_1 - \mu_2$. Tests for

$$\begin{aligned}
&\mathrm{H}_0\colon \mu_1 - \mu_2 = \delta_0,\\
&\mathrm{H}_A\colon \mu_1 - \mu_2 \neq \delta_0, \qquad \mathrm{H}_A\colon \mu_1 - \mu_2 > \delta_0 \qquad \text{or} \qquad \mathrm{H}_A\colon \mu_1 - \mu_2 < \delta_0
\end{aligned}$$

in the paired-sample case is equivalent to tests for

$$\begin{aligned}
&\mathrm{H}_0\colon \mu_{\text{diff}} = \delta_0,\\
&\mathrm{H}_A\colon \mu_{\text{diff}} \neq \delta_0, \qquad \mathrm{H}_A\colon \mu_{\text{diff}} > \delta_0 \qquad \text{or} \qquad \mathrm{H}_A\colon \mu_{\text{diff}} < \delta_0.
\end{aligned}$$

Thus, we just have to apply the test we learnt in Section 9.3 to the sample of the differences $\{d_1, d_2, \ldots, d_n\}$, where each d_i denotes the difference of the i^{th} pair. Read Section 10.4. Note that in practice we seldom know the population variance of the differences. Even if we know the two variances of the two original populations, we cannot derive the population variance of their pairwise differences, and so for paired-data it is unlikely that we can use z-test.

After discussing $\mu_1 = \mu_2$, we talk about the null hypothesis $p_1 = p_2$ in Section 10.5. To test the equality of two unknown population proportions:

$$\mathrm{H}_0\colon p_1 = p_2$$

against one of the three possible alternative hypotheses:

$$\mathrm{H}_A\colon p_1 \neq p_2, \qquad \mathrm{H}_A\colon p_1 > p_2, \qquad \text{or} \qquad \mathrm{H}_A\colon p_1 < p_2,$$

we consider the test statistic

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

where $\hat{p}$ is the sample proportion obtained by combining two samples into one. If in the first sample we have x_1 successes out of n_1 trials and in the second sample we have x_2 successes out of n_2 trials. Then

$$\hat{p}_1 = \frac{x_1}{n_1}, \qquad \hat{p}_2 = \frac{x_2}{n_2}, \qquad \text{and} \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

We can do so because if the null hypothesis is true, the two populations have the same proportion and so we have to estimate only one unknown proportion. When n_1 and n_2 are large, this z has approximately the standard normal distribution and so the decision rule is the same as other tests having a standard normal test statistic, i.e.

H_0	H_A	Reject H_0 at the α significance level if
$p_1 = p_2$	$p_1 \neq p_2$	$\lvert z \rvert \geq z_{\frac{\alpha}{2}}$
$p_1 = p_2$ or $p_1 \leq p_2$	$p_1 > p_2$	$z \geq z_\alpha$
$p_1 = p_2$ or $p_1 \geq p_2$	$p_1 < p_2$	$z \leq -z_\alpha$

To conclude Chapter 10, we should also mention the confidence intervals for the difference between two population means and between two population proportions from two independent samples. For $\mu_1 - \mu_2$, a $100(1-\alpha)\%$ confidence interval is

$$\overline{x}_1 - \overline{x}_2 \pm \begin{cases} z_{\alpha/2} \cdot \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} & \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are known,} \\[2ex] t_{\alpha/2,\, n_1+n_2-2} \cdot s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} & \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are unknown but equal,} \\[2ex] t_{\alpha/2,\, \nu} \cdot \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} & \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are unknown and unequal,} \end{cases}$$

and for $p_1 - p_2$, it is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Obviously we would not assume $p_1 = p_2$ in the construction of the confidence interval for $p_1 - p_2$ (why not?), and so two different estimators, $\hat{p}_1$ and $\hat{p}_2$, are needed to estimate the two unknown proportions p_1 and p_2, respectively.

What I would like you to be able to do is to write down the formulae of test statistics and confidence intervals by using statistical arguments, rather than by retrieving them from your memory. There are too many formulae and some of them are very similar. If you can obtain each of them by statistical argument, then they are easy to distinguish; if you remember them by brute force, then they are easy to be confused. I hope you have noticed that I spent most of my efforts not on how to use the formulae (which will be nothing more than routine work; what you need is not more examples shown by me but more practice by yourselves) but on how these formulae can be obtained by using statistical arguments based on what we have learnt (what you need is not more practice but a thorough understanding of what we have learnt).

Chapter 11 introduces some more complicated tests. The first one is to test how good a hypothesised distribution can describe the given data, which is known as the *goodness-of-fit test*. It can be considered as a null hypothesis concerning several proportions for a population containing not only 0's and 1's but also, say, 2's, 3's and 4's (representing e.g. grade A, grade B, grade C, grade D and grade F) so that we have not only a proportion of successes (which consequently gives us a proportion of failures) but more than two proportions such that the sum of all proportions is still equal to one. That is, such a problem is a generalisation of the hypothesis concerning a hypothesised value of one proportion $p = p_0$ to the hypothesis concerning hypothesised values for several proportions $p_1 = p_{10}$, $p_2 = p_{20}$, ..., $p_k = p_{k0}$, where $p_1 + \cdots + p_k = 1$. (What does p_{i0} stands for? Compare it with the notation μ_0, p_0 and δ_0.) However, in the discussion below, we use English words to describe the true proportions and p_i to denote the i^{th} hypothesised proportion, so that we can omit the second one in the double subscript '$_{i0}$' to avoid messy expressions. The price to pay is some inconsistence in the use of notations.

The idea of the test statistic is to compare the observed values and the expected values under the null hypothesis. The observed values are e.g. the number of grade A, the number of grade B, etc., observed in the given sample; the expected values are values obtained by assuming the null hypothesis. More precisely, each expected value is simply the product of the total and the corresponding probability calculated by using the distribution specified in the null hypothesis.

Suppose the distribution specified in the null hypothesis has k possible outcomes with probabilities p_1, ..., p_k. Then $E_i = np_i$, where n is the total number of observations. Denote by O_i the observed frequency of the i^{th} outcome. Then the larger the difference $O_i - E_i$, the more likely the null hypothesis is wrong. To combine the information in all k outcomes, we first have to get rid of the sign of the differences and so we consider $\sum_{i=1}^{k}(O_i - E_i)^2$. However, we do not know the distribution of this sum. Mathematicians worked out the large-sample distribution (i.e. an approximation when n is large) of the sum of the ratios:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum \frac{(O - E)^2}{E},$$

which is called the χ^2-distribution, with degrees of freedom $k - 1$. We lose one degree of freedom because there is one constraint: $\sum_i (O_i - E_i) = 0$, which actually is a consequence of the fact (or, in some sense, the constraint) that $\sum_i p_i = 1$. This large-sample distribution provides a good approximation when n is so large that $E_i \geq 5$ for all i (I did not spell out this requirement in class).

The story now becomes: we test

$$\begin{aligned} \mathrm{H}_0 &: \text{the distributioin is } \{p_i\} \\ \mathrm{H}_A &: \text{the distribution is not } \{p_i\}. \end{aligned}$$

(As I said above, to avoid making the expression too messy, I did not give any symbols to the true distribution and did not add a subscript '$_0$' to the symbols of the hypothesised distribution. Of course it is perhaps mathematically more proper to express the two hypotheses in terms of only mathematical symbols but then the details would be quite messy or scary.) We reject H_0 only when what we observed is very different from what we expected, i.e. reject H_0 at the α significance level whenever $\chi^2 \geq \chi^2_\alpha$. This is a one-sided test even we have a two-sided alternative hypothesis because we take the square of each difference and hence a very negative value becomes a very large positive value.

Sometimes, the distribution specified in the null hypothesis is just a distribution without specifying the value(s) of the parameter(s), e.g. the null hypothesis says that the data follow the binomial or the Poisson. Then we simply estimate the parameter(s) and then calculate the probabilities and expected values according to the estimated parameter(s). If the χ^2-statistic, now measuring the difference between the observed values and the *estimated* expected values, is large, we can believe that other values for the parameters (leading to different estimated expected values) would give us even larger χ^2 value. Thus, if the χ^2-statistic calculated by using the estimated expected values is already larger than the critical value, we would reject H_0. However, because we will lose degrees of freedom when we estimate parameters, the degrees of freedom of the χ^2-distribution in such a situation is $k - m - 1$, where m is the number of parameters estimated. (This extension has not been mentioned in the textbook but is important. The choice of the symbols is not important; different textbooks may use different symbols for our k and m.)

Note that we do not have time to illustrate everything by numerical examples in class. We only discussed the idea behind the technical details of the test. We will do one or two exercises in the example class and assignments for each problem, but this is definitely not sufficient. You must read the textbook to gain a better understanding and then do some exercises in the textbook to gain a real mastery of the above and other forthcoming testing procedures.

The discussion on the goodness-of-fit test in this review should be read again after you have gone through the details of the examples in Chapter 11. You may not understand thoroughly my summary unless you understand the examples. So, please read the book carefully and repeatedly until you really understand everything.

Next week we will see the generalisation of the hypothesis of the equality of two proportions $p_1 = p_2$ in two populations to the equality of two distributions of two populations, and then introduce a different scenario where the χ^2-distribution is applicable, namely testing hypotheses concerning σ^2 and the construction of the confidence interval for σ^2. Then we will discuss a problem that we mentioned above when we discussed the two-sample t-test, namely, the problem of testing whether two variances are equal or not.

Cheers,
Heng Peng