Comments on normal theory tests (27 April 2020)

We have seen in the last days a number of tests, all assuming a normal distribution of the population sampled. Sometimes this may be an unwarranted assumption, and we have seen that a permutation test can help when testing difference of means. Al alternative is to use Tchebiychev's inequality, which only requires the existence of the first two moments of the distribution sampled.

Thebychev's inequality: when everything else fails

Slides **Testing $H_0 : m = m_0$ with no normality (I)-(II)** show the use of Thebychev's inequality. If in the presence of normality we know that

$$\text{Prob}\left\{\left|\frac{\overline{X}_n - m}{\sigma/\sqrt{n}}\right| \leq z_{\alpha/2}\right\} = 1 - \alpha \tag{1}$$

where $z\alpha/2$ is the quantile leaving probability $\alpha/2$ to its right in the $N(0,1)$ distribution. If σ is not known, we have a similar expression with the quantile taken from the Student's t distribution.

If normality cannot be assumed, we have from Tchebichev's inequality:

$$\text{Prob}\left\{\left|\overline{X}_n - m\right| < k\frac{\sigma}{\sqrt{n}}\right\} \geq 1 - \frac{1}{k^2} \tag{2}$$

If we want the right hand side to have probability at least $1 - \alpha$, it suffices to replace k by $1/\sqrt{alpha}$ throughout:

$$\text{Prob}\left\{\left|\overline{X}_n - m\right| < \frac{\sigma}{\sqrt{n\alpha}}\right\} \geq 1 - \alpha \tag{3}$$

This is similar to (**??**) and enables us to either test hypothesis or construct confidence intervals for m. These intervals will be wider than those based on normal theory: when $\alpha = 0.05$ for instance, $z_{\alpha/2} = 1.96$ while $1/\sqrt{\alpha} = 4.4721$, over twice as large.

Using normal tests as an approximation: proportions

It is still the case that even for distributions that are far from the normal, the tests studied are quite good approximations. One important case is the estimation of a proportion (slides **The case of a proportion (I)-(III)**. Here the mean $\overline{X}_n = (X_1 + \ldots + X_n)/n$ is the binomial frequency and although the individual X_i are binary we know that for large enough n (which in this context is $np > 18$), $\overline{X}_n$ is close to normal and we can expect the normal theory approximation to work quite well. This is indeed the case and

$$\frac{\overline{X}_n - p}{\sigma/\sqrt{n}} \approx N(0,1) \tag{4}$$

We can replace s by the upper bound 0.25 ($\sqrt{p(1-p)}$ in the most unfavourable event, when $p = 1 - p = 0.5$). Slides **The case of a proportion (II)-(III)** show a worked example in which we have estimated s, as the bound would be too conservative.

Using normal tests as an approximation: other cases

Slides **Testing differences of means** and **Testing differences of proportions** both show uses of the normal theory as an approximation. Notice that we use the $N(0,1)$ for the distribution of the test statistic, whether or not the variance is known or estimated. For the sample sizes needed for these approximations to work out, we assume that our estimations of the variances are quite close to the true variances.

In the case of the difference of two means, for small samples we required equal variances. If the sample is large enough the the estimated variances can be taken as the true variances, we still use the normal approximation, whether or not variances are equal.

Paired comparisons

One of the most common misuses of the two-sample t test is the situation where there *seems* to be two populations, but in fact what we have is single population with a bi-variate response. In this last case, the two-sample t test is usually quite powerless, and will fail to reject the null hypothesis when in fact would be indicated.

The example presented in slides **Paired comparisons (I)-(III)** should make the matter clear. We are comparing weights of the first and second children of a sample of mothers. We must resist the temptation of considering two independent samples (first born and second born children), for the weights **cannot** be assumed independent: each pair of brothers come from the same mother and for this fact their weights at birth should be regarded as related.

The simple idea that the paired comparisons method exploits is: if weights of first and second babies have equal mean, their differences should have mean zero. We can thus test this later hypothesis.

Slides **Paired comparisons (IV)-(V)** illustrate the computations (and different conclusions) using the two methods: only the paired comparisons method would be correct in this setting.

We have checked already the computations made by the function `t.test` in the standard two sample test. In the case at hand, taking the differences of weights for brothers, we would have:

```
> First  <- c(3.80, 2.40, 2.750, 1.800)
> Second <- c(4.150, 2.755, 2.900, 1.990)
> Dif <- First - Second
> n   <- length(Dif)
> Xb  <- mean(Dif)
> s2  <- sum( (Dif - Xb)^2 ) / n
> t   <- (Xb - 0) / ( sqrt(s2/(n-1)) )
> t

[1] -4.899484

> Xb

[1] -0.26125
```

This checks out all right with the results in the slide **Paired comparisons (V)** which we reproduce here:

```
> t.test(x=First, y=Second, paired=TRUE)

        Paired t-test

data:  First and Second
t = -4.8995, df = 3, p-value = 0.01627
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.43094421 -0.09155579
sample estimates:
mean of the differences
               -0.26125
```

All we have to do to switch from an ordinary to a paired t-test is to add the argument `paired=TRUE`.

Notice that this rejects at the $\alpha = 0.05$ level the null hypothesis, which the ordinary (an here incorrect) two sample t-test did not. The reason is clear: we have a difference in the means of -0.26125 Kg., the second babies being heavier. This was not significant in the two sample t-test, for there was wide variation on children weight from mother to mother. When we take the differences, the "mother effect" disappears and -0.26125 is compared to a much smaller variance and becomes significant.