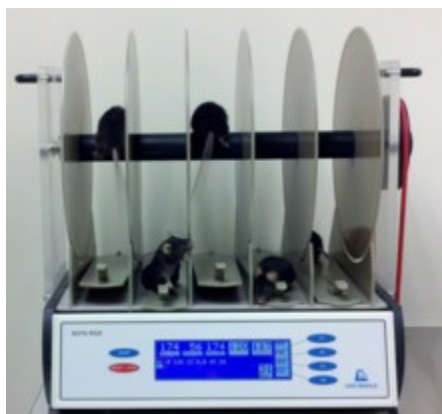# Hypothesis testing

Example 1.

That is widely known that benzodiazepine tranquilizers possess negative effect on muscle tone. So the search for some new tranquilizers free from such effect is actual and needed. In one of the experiments the effect of newly synthesized indole derivative on muscle tone in mouses (rodents) has been compared with reference tranquilizer diazepam using the rotarod performance test [Tsubanova N.A., Shtrygol S.Y. The study of antidepressive and anxyolythic effect of spirocyclic oxindole derivative. Bulletin of pharmacy. - 2011. - №1. - P. 77-80.]. Results of the experiment can be orgaized into the contingency table:

	Diazepam	Studied compound
Fell down	8	4
Didn't fell	0	7

Our task is to say whether such results are enough to claim that the studied compound has less negative effect on muscle tone than diazepam.

Solution

We have two groups in study: the first consists of mouses which were given diazepam, and the second — mouses which were given studied compound. And we have one variable for study — that is the status of mouse: whether mouse fell down or not. The variable is nominal, since it is expressed with words, not numbers. The nominal data has no location and scale parameters, it has only distribution itself. So we will using some test for distributions. The distribution of mouse status in diazepam group is 100% felt down, 0% didn't fell. The distribution of mouse status in compound group is $4/(7+4)=36.4\%$ felt down, $7/(7+4)=73.6\%$ didn't fell. The null hypothesis H_0 is "the distributions of mouse status in both groups (diazepam and compound) are the same", the alternative hypothesis H_1 is "the distributions of mouse status in both groups are not the same". The difference in proportions is evident, however, we should know whether it is statistically significant. (Here and further we adopt common level of statistical significance $\alpha=0.05$.). For nominal data we have as options chi-squared test, Fisher's exact test and McNemar test. The data is not dependent (different mouses come to two studied groups, not the same), so McNemar test does not suit. The data has two low counts: zero and four, thus chi-squared test also is not appropriate. So our choice is for Fisher's exact test. Let us do it with R!

First of all — data input

>mouse=cbind(c(8,0),c(4,7))

Then the test itself

>fisher.test(mouse)

You should search in R output for p-value and compare given value with 0.05. If p-value is greater than 0.05, then we have no enough evidence to reject the null hypothesis. If it is less than 0.05, then we reject null hypothesis as highly unlike, and take alternative hypothesis as true.

Note to your copybooks a) a type of studied variable b) null and alternative hypothesis c) used test d) used commands e) resulting p-value f) conclusion

Task 1. Preeclampsia is a disorder of pregnancy characterized by high blood pressure and large amounts of protein in the urine. In the Uzhgorod perinatal center women were studied for the relations between preeclampsia and other complications (N. Loya, 2014, yet unpublished results). The preterm births in women with preeclampsia and healthy pregnant women were as follows:

	Preeclampsia women	Healthy women
Normal term birth	91	32
Preterm birth	33	3

You need to find percents for preterm birth in preeclampsia and healthy women and to test the statistical significance of difference between studied groups.

ToDo: Note to your copybooks a) a type of studied variable b) percents of preterm birth c) null and alternative hypothesis d) used test and reasoning for it e) used commands f) resulting p-value g) conclusion

Example 2.

Let us examine classical Student's sleep data [Student The probable error of the mean // Biometrika. -1908. - 6. - P. 20.]. Two hypnotic drugs were evaluated. The data is attached to a workspace after calling

>data(sleep)

Now, when calling its name, we can see the data itself

>sleep

The increase in sleeping time is in the first variable extra, then variable group denotes the drug number, and the ID variable corresponds to patient ID. Give some attention to the fact, that ID goes from 1 to 10 and then repeats again from 1 to 10. That is becouse each patient took both drugs in consequent days. So we have the first group, which takes drug 1, and we have the second group, which takes drug 2. However, the patients in both groups are the same. Such groups are called dependent and require proper methods for analysis.

So, first of all let us find average increase in sleeping time in both groups. Calling mean() functions for several groups at a time can be done with tapply() function:

>tapply(sleep$extra,sleep$group,mean)

We can see the difference, however, is it statistically significant or not? To test this we should choose proper test. First of all, let us think about extra variable. It has ratio type, since it has a natural zero and both addition and multiplication of data makes sense. Our null hypothesis H_0 is "There is no difference between sleep time increase in studied groups", our alternative hypothesis H_1 is "There is some difference between sleep time increase in studied groups". We are interested in comparing average values, so one of the location tests should be chosen. Among location tests there are parametric and non-parametric methods. Parametric methods are more powerful, but require the distribution of residuals (deviations from the group means) to be close to normal. To calculate residuals for each groups separately we use function scale() within tapply(). To combine the results together and downgrade them from complex data type to simple vector we use unlist() function.

>res=unlist(tapply(sleep$extra,sleep$group,scale))

Now, there are two possible ways to check whether the distribution of residuals is close to normal. The first one is to use special test for checking normality. The most powerful among such tests is Shapiro-Wilks test:

>shapiro.test(res)

The resulting p-value is greater than 0.05, that means that we cannot reject the null hypothesis that the distribution of residuals is normal. Thus we can decide to choose parametric methods.

Another way is more preferred by experience biostatisticians — that is a visual inspection of quantile-quantile normal plot:

>qqnorm(res)

and adding reference line

>qqline(res)

The line denotes ideal normal distribution. The greater distance between points and the line means the less similarity of current distribution with normal one.

In our case points are close enough to make a choice in favour of parametric methods.

The parametric method among location tests for dependent data is paired t-test.
>t.test(sleep$extra~sleep$group, paired=T)
In the output we have p-value and also mean of the differences and 95% confidence interval for that mean.
ToDo: Note to your copybooks a) a type of studied variable b) average increase in sleep time in both groups c) null and alternative hypothesis d) used test and reasoning for it e) used commands f) mean of the differences and 95% confidence interval for that mean g) resulting p-value h) conclusion

Example 3.
Let us continue with Student's sleep data and try to downgrade extra sleep time to the nominal variable (medical researchers like that, for example, they always try to downgrade age to nominal age intervals). Let our intervals will be: the first from -3 to 0 (i.e. decrease of sleeping time), the second — from 0 to 3 (increase in sleeping time) and the third — from 3 to 6 (great increase in sleeping time). To do this, we will use cut() function with proper breaks parameters
>extra.g=cut(sleep$extra,breaks=c(-2,0,3,6))
To see entire result by each drug, we may type
>table(extra.g, sleep$group)
However, remember that our data is dependent, so each patient has been used to test both drugs. So we should preserve information about his interval for extra both when taking the first and the second drug. To do this, we call tabulate the first drug data by the second drug data (to divide data into two samples we are using tapply()):
>tab=table(tapply(extra.g,sleep$group, unlist))
>tab
In the resulting table we can see, that there was one patient, that has descreased slleping time when taking both drugs. Three patients have decreased sleep when taking the first drug, but increased sleep when taking the second. One patient has decreased sleep when taking the first drug, but great increase when taking the second and so on. The sum of the whole table is 10 — exactly the number of our patients. To analyze such table the test for nominal dependent data should be utilized. That is McNemar test:
>mcnemar.test(tab)
ToDo: Note to your copybooks a) a type of studied variable b) null and alternative hypothesis c) used test and reasoning for it d) used commands and their output e) resulting p-value f) conclusion

Task 2.
The patients that had stroke were followed-up in Uzhgorod regional hospital for a several years. Some of them have got later another stroke accident. Find whether the probability of accident is related to high level of systolic blood pressure (measured after the first stroke episode). The data can be downloaded from stat.org.ua using command
>bp=read.csv("http://stat.org.ua/data/systolic.csv")
ToDo: Note to your copybooks a) a type of studied variable b) average systolic blood pressure in both groups (having the second stroke accident or not) c) null and alternative hypothesis d) used test and reasoning for it e) used commands f) resulting p-value g) conclusion