SGA - Basic Statistics - Isupov Ilya

April 2024

Telegram: @Jeremix

1 How to test a drug?

You perform a clinical trial to study new drug. You have 20 volunteers with some disease. You randomly split all the volunteers into two groups (10 volunteers in each): the treatment group and the control group. Volunteers in the treatment group receive the new drug, volunteers in the control group receive placebo (pills that looks like a drug but do not have active substance). You conclude that new drug is effective if people who take the drug will recover faster (on average) than people in the control group. If your drug is effective, you will invest in its production, otherwise you will look for another drug. Assume that you obtained the following data (disease duration in days).
Control group = [6, 7, 7, 5, 7, 8, 8, 7, 7, 7]
Treatment group = [7, 6, 6, 5, 5, 6, 7, 5, 5, 8]

1. Describe this problem in terms of statistical hypothesis testing framework. How would you model your data in terms of random variables?
We have two groups of people, one took the drug, the other a placebo. We are interested to know if the drug works or not. If the drug works, then the recovery time of people taking the drug will be less, if not, then the recovery time will be about the same.
Let's use random variables - $\mathbf{X}$ and $\mathbf{Y}$:
- $\mathbf{X}$ is the recovery time of the treatment group,
- $\mathbf{Y}$ is the recovery time of the control group.
Suppose that the random variables $\mathbf{X}$ and $\mathbf{Y}$ are distributed normally and independently.

2. State the null hypothesis and the alternative. Will your alternative be one-sided or two-sided? Why?
Let's assume for the null hypothesis that the average recovery time of the two groups will not differ:

$$H_0 : \mu_x = \mu_y$$

Let's take as an alternative hypothesis that the average recovery time of a group taking drugs will be less than the recovery time of a group of people taking a placebo:

$$H_1 : \mu_x < \mu_y$$

We have adopted a one-sided alternative hypothesis, since in this case we want to invest in drug development. In which case do we want to invest in it? Only if it works. We are not interested in the case when patients will recover longer when taking the drug than without it. Therefore, it is worth considering a one-sided alternative.

3. What kind of statistical test will you use? Why this test?

In this case, we have 20 different people, 10 in each group. People are not related to each other in any way, they are just random people. Thus, we have two independent random variables - X and Y. In this case, **Two-sample t-test** should be used.

But if we were testing our drug on one group of people - first we would look at their recovery time without taking the medicine, and then at the recovery time when taking the medicine, then in this case we should consider a Paired t-test.

Let's assume that we plan to invest a lot of money in this project and want to reduce the likelihood of a first-type error, so we'll assume a significance level of 1 percent:

$$\alpha = 1\% = 0.01$$

4. Use this test (apply Python if necessary and provide your code).

```
from scipy.stats import ttest_ind
import numpy as np
```
✓ 0.0s

```
X = [7, 6, 6, 5, 5, 6, 7, 5, 5, 8]  # Treatment group
Y = [6, 7, 7, 5, 7, 8, 8, 7, 7, 7]  # Control group
alpha = 0.01
```
✓ 0.0s

```
_, p_value = ttest_ind(
                       X, Y,
                       alternative='less',
                       equal_var=(np.var(X, ddof=1) == np.var(Y, ddof=1))
                       )
p_value
```
✓ 0.0s

```
0.026452854031349084
```

```
if p_value < alpha:
    print(f'P-value = {round(p_value, 3)} < alpha = {alpha}')
    print('Reject the null hypothesis.')
else:
    print(f'P-value = {round(p_value, 3)} > alpha = {alpha}')
    print('Fail to reject the null hypothesis.')
```
✓ 0.0s

```
P-value = 0.026 > alpha = 0.01
Fail to reject the null hypothesis.
```

5. Analyse the results and provide a conclusion in mathematical and real-life terms. Would you invest into production of this drug?

We got a p-value = 0.026 bigger than the accepted significance level of alpha = 0.01, which means that we **don't reject the null hypothesis**. Therefore I would not invest in the development of the drug, but I would conduct additional experiments with a large sample size to determine the effect of the drug on people during the course of the disease, because the planned amount of investment is significantly more than the cost of conducting one experiment, even with a larger sample. So it's better to check the results a couple more times.