

SGA - Basic Statistics - Isupov Ilya

April 2024

Telegram: [@Jeremix](#)

1 What makes you happy?

Assume you perform a study to detect how using social networks affects people's happiness level. You have 20 volunteers. Your study is planned as follows. All participants are known to be active users of social networks. First, you ask every participant to fill in a special questionnaire that allows you to estimate their happiness level. After that, all participants will avoid using of social networks for one week. After this week, they complete a similar questionnaire to detect their new level of happiness. Then, for each participant, their new happiness level is compared to the initial one. Assume that for each participant, their happiness level is changed: either decreased or increased. Let X_{obs} be the random variable that models the number of participants for who increased their happiness level. Let $X_{obs}=16$, i.e. 16 out of 20 participants become happier, and it's the only data on which you can make a decision. Your significance level is 5%.

1. You should state the null hypothesis and the alternative hypothesis of your research and explain your choices.

We know that each participant's happiness level has changed. Let's assume that avoiding the use of social networks doesn't affect the level of happiness in any way, in this case, we will assume for the null hypothesis that half of the participants had an increased level of happiness, and the other half decreased.

$$H_0 : p = \frac{1}{2}$$

We are interested in determining whether the absence of social networks increases the level of happiness, so we define an alternative hypothesis as follows:

$$H_1 : p > \frac{1}{2}$$

We have the following parameters:

- number of participants:

$$n = 20$$

- X - random variable that models the number of participants for who increased their happiness level.

$$X_{obs} = 16$$

- significance level:

$$\alpha = 5\% = 0.05$$

2. You should state how X is distributed provided that null hypothesis holds.

X is distributed as a binomial random variable, because X has two outcome - happy and unhappy.

$$X \sim \text{Bin}(n = 20, p = \frac{1}{2})$$

3. Would you claim that people become happier when they avoid using social networks based on this data?

Let's check this using p-value and determine the probability that 16 or more people will become happier without using social networks:

$$\begin{aligned} p - \text{value} &= P(X \geq 16|H_0) = \\ &= P(X = 16|H_0) + P(X = 17|H_0) + P(X = 18|H_0) + P(X = 19|H_0) + P(X = 20|H_0) \\ P(X = 16|H_0) &= C_{16}^{20} \cdot \left(\frac{1}{2}\right)^{20} = \frac{20!}{16! \cdot 4!} \cdot \left(\frac{1}{2}\right)^{20} = \frac{4845}{1048576} \\ P(X = 17|H_0) &= \frac{1140}{1048576} \\ P(X = 18|H_0) &= \frac{190}{1048576} \\ P(X = 19|H_0) &= \frac{20}{1048576} \\ P(X = 20|H_0) &= \frac{1}{1048576} \\ p - \text{value} &= P(X \geq 16|H_0) = \frac{4845 + 1140 + 190 + 20 + 1}{1048576} = \frac{6196}{1048576} \approx 0.006 \\ p - \text{value} &\approx 0.006 < \alpha = 0.05 \end{aligned}$$

Thus, the p-value is less than the significance level, which means we must reject our null hypothesis in favor of alternative hypothesis. The result confirms that **people become happier when they avoid using social networks.**

Let's check this p-value using Python, scipy stats and binomial distribution:

```
p_value = 1 - stats.binom.cdf(k=15, n=20, p=0.5)
round(p_value, 3)
✓ 0.0s
0.006
```