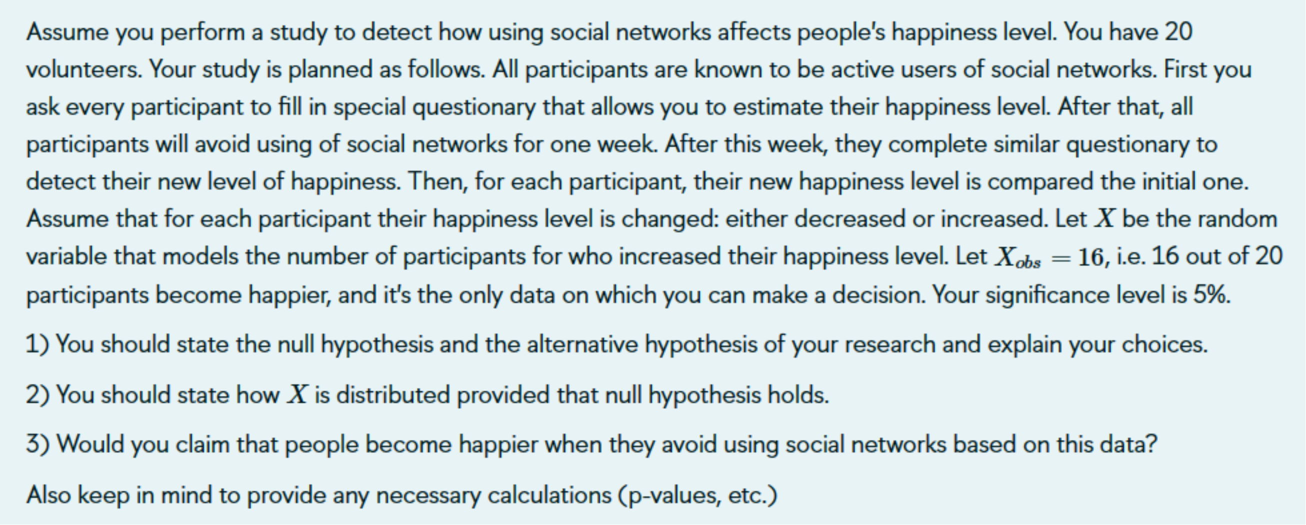
Question 1



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 special questionary 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 similar questionary to detect their new level of happiness. Then, for each participant, their new happiness level is compared the initial one. Assume that for each participant their happiness level is changed: either decreased or increased. Let X 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.

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

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

Also keep in mind to provide any necessary calculations (p-values, etc.)

1) The hypotheses are based on the goal of the study, which is to detect how using social networks affects people’s happiness level.

The null hypothesis H_0 is that **there is no** significant difference in the happiness level of participants before and after avoiding social networks for one week. In other words, the use of social networks has no effect on the happiness level of participants, so the null hypothesis represents the assumption that there is no relationship between the two variables (use of social networks and happiness level).

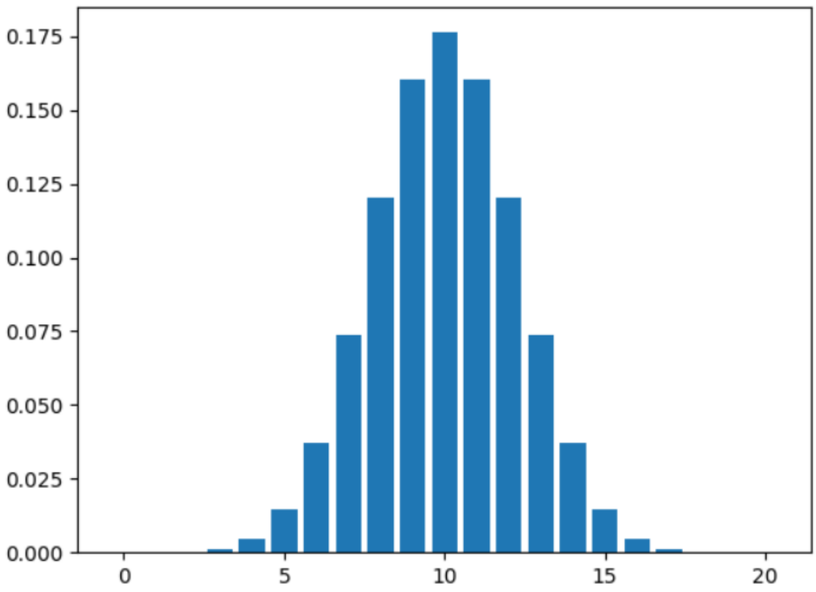
The alternative hypothesis H_1 is that **there is** a significant difference in the happiness level of participants before and after avoiding social networks for one week. This could mean that avoiding social networks has an effect on the happiness level of participants, so the alternative hypothesis represents the assumption that there is certain relationship between the two variables (use of social networks and happiness level).

2) If the null hypothesis holds, then the number of participants who increased their happiness level after avoiding social networks for one week would follow a binomial distribution. This is because each participant has two possible outcomes (increased happiness level or not), the trials are independent, and the probability of success (increased happiness level) is constant for each trial.

In this case, the binomial random variable X representing the number of participants who increased their happiness level would have parameters n = 20 (the number of participants) and
p = 0.5 (the probability of success under the null hypothesis that there is no difference in happiness level before and after avoiding social networks).

The code below allows to plot the distribution.

```python
from scipy.stats import binom
import matplotlib.pyplot as plt
n = 20
p = 0.5
X = binom(n, p)
x = list(range(n + 1))
plt.bar(x, X.pmf(x));
```



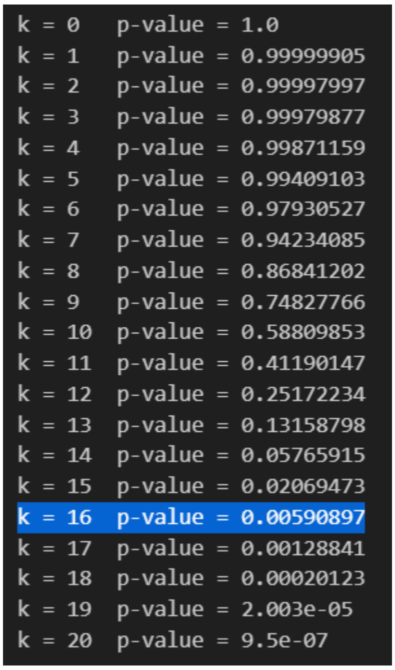
3) To determine whether to claim that people become happier when they avoid using social networks based on this data, we need to calculate the p-value for the observed data and compare it to the chosen significance level of 5%.

The p-value represents the probability of observing a result as extreme or more extreme than the observed data, assuming that the null hypothesis is true. If the p-value is less than or equal to the chosen significance level, then we would reject the null hypothesis and conclude that there is evidence to suggest that avoiding social networks has an effect on happiness level. Otherwise, if the p-value is greater than the significance level, we would fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that avoiding social networks has an effect on happiness level.

In this case, the observed data is that 16, i.e. X_{obs} (or k) = 16, out of 20 participants increased their happiness level after avoiding social networks for one week. Since this is a one-sided test (we are testing whether avoiding social networks increases happiness level), the p-value would be calculated as the probability of observing 16 or more successes (increased happiness level) in 20 trials (participants), assuming that the probability of success is 0.5 under the null hypothesis.

The code snippet below performs p-value calculations for different values of X_{obs} (k).

```python
from scipy.stats import binom
n = 20
p = 0.5
for k in range(n + 1):
    p_value = binom.sf(k - 1, n, p)
    print(f"k = {k} \tp-value = {round(p_value, 8)}")
```



```
k = 0   p-value = 1.0
k = 1   p-value = 0.99999905
k = 2   p-value = 0.99997997
k = 3   p-value = 0.99979877
k = 4   p-value = 0.99871159
k = 5   p-value = 0.99409103
k = 6   p-value = 0.97930527
k = 7   p-value = 0.94234085
k = 8   p-value = 0.86841202
k = 9   p-value = 0.74827766
k = 10  p-value = 0.58809853
k = 11  p-value = 0.41190147
k = 12  p-value = 0.25172234
k = 13  p-value = 0.13158798
k = 14  p-value = 0.05765915
k = 15  p-value = 0.02069473
k = 16  p-value = 0.00590897
k = 17  p-value = 0.00128841
k = 18  p-value = 0.00020123
k = 19  p-value = 2.003e-05
k = 20  p-value = 9.5e-07
```

From the calculation results above we can see that for k = 16 p-value amounts to 0.0059 which is much lower than the significance level of 0.05 (5%). Basing on the given data this enables us to reject the null hypothesis and claim that people become happier when they avoid using social networks.

In the context of this study, a Type I error would occur if we concluded that avoiding social networks has an effect on people’s happiness level when, in reality, it does not. A Type II error would occur if we concluded that avoiding social networks does not have an effect on people’s happiness level when, in reality, it does.

The probability of making a Type I error is controlled by the significance level of the test. A lower significance level means that the test is more stringent and less likely to make a Type I error. The probability of making a Type II error depends on several factors, including the sample size, the true effect size, and the significance level of the test. Increasing the sample size or using a less stringent significance level can reduce the probability of making a Type II error.