Instructions

Upload your solutions (handwritten or typed) via Canvas Assignments on or before 10/12/2021.

Grading scheme

a	b	c	d	e	f	g	h	Total
2	2	2	3	3	2	3	3	20

---

a. State carefully the probability model involved, make sure to define all your variables and explain which parameters are known and which unknown.

Solution. Let X denote the number of people who present mild to moderate side effects from a random sample of 25 subjects. We assume that the responses of the subjects are independent, so they follow a binomial distribution $X \sim \text{Binom}(n, p)$ with $n = 25$ and p *unknown.* □

b. Which type of test should they perform? A two-sided test, a left-sided test, or a right-sided test? State the null and alternative hypotheses.

Solution. The researchers want to know if $p < 0.30$, therefore the correct test is a one-sided left (or lower) tail test. The correct hypotheses are: $H_0 : p \geq 0.3$ and $H_1 : p < 0.3$. □

c. Define the test statistic, justify your choice. Now give the distribution of it under the null hypothesis.

Solution. The sample size $n = 25$ is a borderline case. Some people consider it large enough to perform a normal approximation, while other more conservative people might not deem it large enough.
You have two acceptable options: compute the exact probabilties of the binomial distribution using the binomail table; or, use a normal approximation including a continuity correction.
However: *why approximate something when you can easily compute its exact value?*
Since the sample size is very small, if you do not use a continuity correction your answers will be a poor approximation! Unaceptable in this situation.

Option 1. For a binomial test the test statistic is the random variable X itself. Under H_0 the distribution of X is Binom(20, 0.3).

Option 2. If we choose to use a normal approximation, we should still use X as the test statistic. Under the null hypothesis, and because of the Central Limit Theorem, we have that $X \approx N(np_0, np_0(1 - p_0)) = \mathcal{N}(7.5,\ 5, 25)$. □

d. Define the rejection region using a critical value.

Solution.
Option 1. We search for the largest integer c such that $\mathbf{P}(X \leq c \,|\, H_0) \leq 5\%$. Looking at the table we choose $c = 3$. Thus we reject H_0 if $X \leq 3$.

Option 2. We search for the a value of c such that $\mathbf{P}(X \le c \,|\, H_0) = 5\%$.

$$\begin{aligned} 5\% &= \mathbf{P}(X \le c \,|\, H_0) \\ &= \mathbf{P}(X \le c + 0.5 \,|\, H_0) && \text{continuity correction} \\ &= \mathbf{P}\left(\frac{X - 7.5}{\sqrt{5.25}} \le \frac{(c+0.5) - 7.5}{\sqrt{5.25}} \,\middle|\, H_0\right) \\ 5\% &\approx \mathbf{P}\left(Z \le \frac{(c+0.5) - 7.5}{\sqrt{5.25}}\right) && \text{CLT, null hypothesis} \end{aligned}$$

We know that $\mathbf{P}(Z \le -1.645) = 5\%$, so we conclude that $\frac{(c+0.5)-7.5}{\sqrt{5.25}} = -1.645$. Solving this equation for c gives: $c = 3.23$. Therefore, we reject if $X \le 3$. □

e. What is the P-value of the result?

Solution.
Option 1. The observed value of the test statistic is $x = 5$. Using the binomial table, we get that the P-value is $\mathbf{P}(X \le 5 \,|\, H_0) = 0.193$.

Option 2. The P-value is $\mathbf{P}(X \le 5 \,|\, H_0)$. Using a continuity correction we compute this to be

$$\mathbf{P}\left(\frac{X - 7.5}{\sqrt{5.25}} \le \frac{(5+0.5) - 7.5}{\sqrt{5.25}} \,\middle|\, H_0\right) = \mathbf{P}(Z \le -0.873) = 0.191.$$

□

f. Explain in words the conclusion that can be drawn from this clinical trial.

Solution. There is not enough evidence from the data provided to conclude (at a 5% level of significance) that the proportion of people who would present mild to moderate symptoms is smaller than 30%. □

g. Compute the power of the test, assuming the true proportion is $p = 0.2$. What could the researchers do to increase the power of the test?

Solution.
Option 1. Using the binomial table, we get that Power(0.2) $= \mathbf{P}(X \le 3 \,|\, p = 0.2) = 0.234$. To increase the power the sample size needs to be increased.

Option 2. If the true value is $p = 0.2$, then $X \approx \mathcal{N}(5, 4)$. We have Power(0.2) $= \mathbf{P}(X \le 3 \,|\, p = 0.2)$. Using a continuity correction we compute this to be

$$\mathbf{P}\left(\frac{X - 5}{\sqrt{4}} \le \frac{(3+0.5) - 5}{\sqrt{4}} \,\middle|\, p = 0.2\right) = \mathbf{P}(Z \le -0.750) = 0.227.$$

To increase the power the sample size needs to be increased. □

h. After the trial is concluded, the researchers noticed that 5 out of the 25 subjects involved are members of the same family. The researchers have different opinions about how to react to this fact.

- Researcher 1 thinks there is no reason to revise the results in **f.**
- Researcher 2 thinks that the 5 family members should be dropped from the trial, and the data should be re-analyzed using only the remaining 20 subjects.
- Researcher 3 thinks that the integrity of the trial is compromised and they need to start all over again.

For each of the 3 opinions above, judge whether the statement is valid from a statistical perspective. What would you advise the researchers to do (and why)?

Solution.

- Researcher 1 is wrong. Because the response to disease and medication usually has a genetic component, it is not possible to regard these 5 participants as a random sample of the total population! In an extreme case, suppose the five people who presented symptoms are precisely these 5 family members (or suppose none of the family members presented symptoms). Do you still trust the test? You can simply not accept these results!
- Perhaps Researcher 2 has a good point. Without the 5 family members it *might* be possible to consider the remaining 20 participants as a random sample. On the other hand, hand picking which participants to include and which to exclude might introduce a bias in the study. An important question to ask is: how was it possible that 5 family members were "randomly chosen"? This suggests that whatever method was used to recruit participants is not very reliable! This needs to be checked. You can also argue here that the power of the test is rather low as it is, a smaller sample size will decrease the power even more.
- Researcher 3 is the most reasonable one. The test should be re-designed in such a way that the selection of participants resembles a random sample.

Because of the points above, I would agree with Researcher 3.

□