Stats I

Midterm Exam

Barboza-Salerno

Due:

1. A survey is distributed and the fraction of the population, *p*, that supports Issue 1 establishing a state constitutional right to "make and carry out one's own reproductive decisions” was calculated.

a. A new poll in Ohio shows that the probability is .5. Assume that p = .5. Use the central limit theorem to estimate the probability that in a new survey of n = 25 people, at least 14 people support Issue 1.

b. With *p* being unknown, let be the fraction of the survey respondents who support Issue 1. What is the smallest sample size*, n*, in order to have 95% confidence that is within .01 of the true value of *p?*

2. Let *A =* the event that an individual is diagnosed with PTSD and *B* = the event that an individual is diagnosed with major depressive disorder (MDD).

a. Suppose the probability that an individual is diagnosed with neither PTSD nor MDD is 3/8. What is the probability that at least one of the events occurs?

b. Suppose that P(A) = .5, = .2, What is P(B)?

3. The success rate for an intervention was documented across multiple sites. At each of 60 centers, researchers tested the intervention on 12 participants and recorded the number of successes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *x* | 0 | 1 | 2 | 3 | 4 | 5 |
| Counts | 4 | 15 | 17 | 10 | 8 | 6 |

The following table gives counts from 60 centers, e.g., x = 2 occurred in 17 of the 60 centers.

1. Compute the average number of successes over the 60 centers.
2. Conduct a goodness of fit test that tests the assumption that the probability of success is the same at each center using a significance of .05.

4. In a trial of a new experimental cancer drug, 821 patients were treated with 1 mg doses of the drug. 30 experienced nausea within 24 hours of taking the drug. In 725 persons who did not take the drug, 9 experienced nausea in this same time frame. Under an assumption that the new cancer drug does not cause nausea and that 9/725 is the probability that a randomly selected person experiences nausea within a 24-hour period, what is the probability that out of 821 people, the observed number of people (i.e., 30) -- or more -- experience nausea in a given day? Does this seem unusual? What reasonable conclusion might one make about the new experimental drug?

5. Suppose we have 49 data points with sample mean 6.25 and variance 12. The null hypothesis, the data are drawn from a normal distribution, versus the data is not.

1. Test for significance at the .05 level and use R to compute the exact p-value.
2. Draw a picture showing the null hypothesis, the rejection region and the area used to compute the p-value

6. In my work on domestic violence, I noticed that when data are disaggregated across space and time, that victimization is more common immediately following holidays (e.g., New Years) compared to the period immediately preceding them. Suppose 20 domestic violence incidents falling within one week of holidays are randomly selected. Under the assumption that there is nothing special about holidays, and there are no issues suggesting that victimization events should be uniformly spread throughout these two weeks -- find the probability that more than 14 victimization incidents occur in the week after Thanksgiving. What likely conclusion can one make upon seeing 15 out of the 20 victimization incidents selected occurred in the week after Thanksgiving?

7. In a large population of graduate students, 35% have experienced feelings of anxiety. To study this further, you take a random sample of 10 students from OSU. Find the probability that

* 1. Exactly 2 students have experienced anxiety
  2. At most 7 students have experienced anxiety
  3. The 10th person selected experienced anxiety
  4. Calculate the standard deviation of the number of students in the sample who have experienced anxiety
  5. If you took a larger sample of students, say 150, how many would be expected to experience anxiety?
  6. In a sample of 150 randomly selected students, 30% experienced anxiety. Test the hypothesis that the sample mean is significantly different from the population mean.

8. Make a statistical argument (i.e., compute the likelihood) that the selection process in the Ahmaud Arbery case was inherently unfair after reading this article: <https://www.npr.org/2021/11/05/1052435205/ahmaud-arbery-jury>. Choose a way to illustrate the results visually by selecting a plot and graphing it (you can use SPSS, JASP or R).

9. I have been studying child maltreatment fatalities recently. I collect information about survivorship (*B*)of 1430 children who were substantiated for child abuse or neglect (*A*). The exposure is that the child was substantiated for abuse or neglect and the outcome is whether they subsequently died (i.e., did not surviv). Using the table below, answer the following questions:

|  |  |  |
| --- | --- | --- |
| Table of child deaths following substantiated maltreatment | | |
|  | Died | Did not die |
| Sub. Maltreatment | 40 | 746 |
| Not substantiated | 12 | 632 |

1. What is the risk that a child who was substantiated for maltreatment dies?
2. What is the risk that a child who was not substantiated for maltreatment dies?
3. Describe the relative risk of dying for the children who were exposed?
4. Compute and interpret the odds ratio of dying (versus not dying) among those who were substantiated for maltreatment (versus those who were not).
5. Take the reciprocal of (d) and interpret it (i.e., 1/answer in d)

10. (Central Limit Theorem) Let be independent and identically distributed (i.i.d.), each with expected value , and variance of the sum . Approximate , using the central limit theorem.

Conceptual Questions

1. T/F: the following random variable distributed as a binomial random variable? The number of questions correct if one randomly guesses on a quiz of 20 multiple choice questions where each question has 4 possible answers. (binomial, normal, chi-square, F, neither)
2. T/F: the following random variable distributed as a binomial random variable? The number of people with blue eyes in a group of 10 people drawn from a room of 30 people without replacement.
3. T/F: the following random variable distributed as a binomial random variable? The number of each of 3 different mental health disorders present in a group of 100 children.
4. T/F: The odds of passing statistics for those who study are 4.55 times higher compared to those who do not study, and the associated confidence interval is [3.45, 8.45]. The odds ratio is statistically significant.
5. T/F: The “sampling distribution of the mean” refers to the distribution of all possible means regardless of sample size.
6. T/F: The uncertainty around a point estimate can be reduced by decreasing the level of confidence from 95% to 90%.
7. T/F: The uncertainty around a point estimate can be reduced by increasing the sample size.
8. T/F: Two events that are independent have the same chance of occurring.
9. T/F: If two events are independent, knowing the outcome of one changes the probability of the other occurring.
10. T/F: Event A is "a randomly selected student will graduate in four years from a PhD program". Event B is "a randomly selected student who does not graduate in four years will leave academia and get a high paying job". Events A & B are not mutually exclusive events, but they are dependent.