STA4504 Chapter 1 Homework

Edgard Maboudou

1. In the following examples, identify the response variable and the explanatory variables.

   a. Attitude toward gun control (favor, oppose), Gender (female, male), Mothers education (high school, college).

   b. Heart disease (yes, no), Blood pressure, Cholesterol level.

   c. Race (white, nonwhite), Religion (Catholic, Jewish, Protestant), Vote for president (Democrat, Republican, Other), Annual income.

   d. Marital status (married, single, divorced, widowed), Quality of life (excellent, good, fair, poor).

   Solution:

   a. The response variable, attitude toward gun control, depends on the explanatory variables, gender and mother's education.

   b. The response variable, heart disease, depends on the explanatory variables blood pressure and cholesterol level.

   c. The response variable, vote for president, depends on the explanatory variables, race, religion, and annual income.

   d. The response variable, quality of life, depends on the explanatory variable marital status.

   □

2. Which scale of measurement is most appropriate for the following variables – nominal, or ordinal?

   a. Political party affiliation (Democrat, Republican, unaffiliated).

   b. Highest degree obtained (none, high school, bachelors, masters, doctorate).

   c. Patient condition (good, fair, serious, critical).

   d. Hospital location (London, Boston, Madison, Rochester, Toronto).

   e. Favorite beverage (beer, juice, milk, soft drink, wine, other).

   f. How often feel depressed (never, occasionally, often, always).

   Solution:

   a. The order of valuables is irrelevant, so they are **nominal** variables.

   b. The variables have ordered scales (by level), so they are **ordinal** variables.

   c. The variables have ordered scales (by intensity), so they are **ordinal** variables.

   d. The order of valuables is irrelevant, so they are **nominal** variables.

   e. The order of valuables is irrelevant, so they are **nominal** variables.

   f. The variables have ordered scales (by frequency), so they are **ordinal** variables.

   □

3. Each of 100 multiple-choice questions on an exam has four possible answers but one correct response. For each question, a student randomly selects one response as the answer.

   a. Specify the distribution of the students number of correct answers on the exam.

b. Based on the mean and standard deviation of that distribution, would it be surprising if the student made at least 50 correct responses? Explain your reasoning.

Solution:

a. It is a **binomial distribution** because the categorical data result from 100 independent and identical trials have two possible outcomes (correct or incorrect) for each.

b. The binomial distribution for $n = 100$ trials with parameter $\pi = \frac{1}{4}$ has mean and standard deviation

$$E = n\pi = 25 \quad \text{and} \quad \sigma = \sqrt{n\pi(1-\pi)} = \sqrt{100 \cdot \frac{1}{4} \cdot \frac{3}{4}} \approx 4.33.$$

Since making 50 correct responses is more than 3 standard deviation from the mean, it is very surprising. In fact, the probability of 50 correct response equals

$$P(50) = \binom{100}{50} \times .25^{50} \times .75^{50} \approx 0.$$

□

6. Genotypes AA, Aa, and aa occur with probabilities (π_1, π_2, π_3). For $n = 3$ independent observations, the observed frequencies are (n_1, n_2, n_3).

   a. Explain how you can determine n_3 from knowing n_1 and n_2. Thus, the multinomial distribution of (n_1, n_2, n_3) is actually two-dimensional.

   b. Show the set of all possible observations, (n_1, n_2, n_3) with $n = 3$.

   c. Suppose $(\pi_1, \pi_2, \pi_3) = (0.25, 0.50, 0.25)$. Find the multinomial probability that $(n_1, n_2, n_3) = (1, 2, 0)$.

   d. Refer to (c). What probability distribution does n_1 alone have? Specify the values of the sample size index and parameter for that distribution.

Solution:

a. In multinomial distribution, $n = n_1 + n_2 + n_3$. Thus, knowing n_1 and n_2 gives n_3 since we are given n.

b. The set of all possible observations of(n_1, n_2, n_3) with $n = 3$ is:

$$S = \{(3,0,0), (0,3,0), (0,0,3), (2,1,0), (2,0,1),$$
$$(1,2,0), (1,0,2), (0,1,2), (0,2,1), (1,1,1)\}.$$

c. By the multinomial probability function,

$$P((n_1, n_2, n_3) = (1,2,0) = \frac{3!}{1!2!0!} \cdot 0.25^1 \cdot 0.50^2 \cdot 0.25^0 = 0.1875.$$

d. From (c), we know n_1 alone has binomial distribution because a person either has genotype AA or does not have genotype AA. The the sample size index $n = 3$ and parameter $\pi = .25$.

□

7. In his autobiography *A Sort of Life*, British author Graham Greene described a period of severe mental depression during which he played Russian Roulette. This "game" consists of putting a bullet in one of the six chambers of a pistol, spinning the chambers to select one at random, and then firing the pistol once at one's head.

a. Greene played this game six times, and was lucky that none of them resulted in a bullet firing. Find the probability of this outcome.

b. Suppose one kept playing this game until the bullet fires. Let Y denote the number of the game on which the bullet fires. Argue that the probability of the outcome y equals $\left(\frac{5}{6}\right)^{y-1}\left(\frac{1}{6}\right)$, for $y = 1, 2, 3, \ldots$. (This is called the geometric distribution.)

Solution:

a. Since the probability of having a bullet firing is $\frac{1}{6}$ each time. As the trials are independent, the multiplicative rule is used. So, the probability of none of the six trials resulted in a bullet firing is $p = \left(1 - \frac{1}{6}\right)^6 \approx 0.335$.

b. Since the game will conclude with the first success, it has to fail $y - 1$ times in order to have $Y = y$, the bullet fires on the yth game. Thus, as the trials are independent of each other, $P(Y = y) = \left(\frac{5}{6}\right)^{y-1}\left(\frac{1}{6}\right)$, for $y = 1, 2, 3, \ldots$.

□

1.9 A sample of women suffering from excessive menstrual bleeding have been taking an analgesic designed to diminish the effects.Anewanalgesic is claimed to provide greater relief. After trying the new analgesic, 40 women reported greater relief with the standard analgesic, and 60 reported greater relief with the new one.

a. Test the hypothesis that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic. Report and interpret the P-value for the two-sided alternative. (Hint: Express the hypotheses in terms of a single parameter. A test to compare matched-pairs responses in terms of which is better is called a sign test.)

b. Construct and interpret a 95% confidence interval for the probability of greater relief with the new analgesic.

Solution:

a. Let π denote the proportion of women reported greater relief with the standard analgesic. We test $H_0 : \pi = 0.5$ against the two-sided alternative hypothesis, $H_a : \pi \neq 0.5$. The sample proportion of "yes" responses was $p = \frac{60}{100} = 0.6$. For a sample of size $n = 100$, $n\pi_0 = 50 > 5$ and $n(1 - \pi_0) = 50 > 5$ so n is large enough. The null standard error of p equals $\sqrt{\frac{0.5\times0.5}{100}} = 0.05$. The test statistic is

$$z = \frac{0.6 - 0.5}{0.05} = 2.$$

Assume $\alpha = 0.05$. Since $z > z_{\alpha/2} = z_{0.025} = 1.96$, we rejected H_0.
Conclusion: There is no sufficient evidence that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic.
The two-sided P-value is the probability that the absolute value of a standard normal variate exceeds 2, which is $P = 0.046$. There is strong evidence that, $\pi > 0.5$.

b. A large sample $100(1 - \alpha)\%$ confidence interval for π has the formula

$$p \pm z_{\alpha/2} \cdot SE, \quad \text{with } SE = \sqrt{\frac{p(1-p)}{n}}.$$

For 95% confidence, $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96$. With $p = 0.6$ and $n = 100$, the 95% confidence interval equals

$$0.6 \pm 0.096 = (0.504, 0.696).$$

We can be 95% confident that the proportion of women reported greater relief with the new one is between 0.504 and 0.696.

□

1.10 Refer to the previous exercise. The researchers wanted a sufficiently large sample to be able to estimate the probability of preferring the new analgesic to within 0.08, with confidence 0.95. If the true probability is 0.75, how large a sample is needed to achieve this accuracy? (Hint: For how large an n does a 95% confidence interval have margin of error equal to about 0.08?)

Solution: Since we required that

$$1.96 \cdot \sigma_p = 0.08 \text{ and } \sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75 \times 0.25}{n}},$$

we can solve for $n = 0.75 \times 0.25 \times \left(\frac{1.96}{0.08}\right)^2$. So a sample of 113 is needed to achieve this accuracy.

□