Matt Isaac

Rates & Proportions

Homework 1

Due: 9/8/17

1.2

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

* 1. Political party affiliation (Democrat, Republican, unaffiliated) – **nominal**
  2. Highest degree obtained (none, high school, bachelor’s, master’s, doctorate) – **ordinal.**
  3. Patient condition (good, fair, serious, critical) – **ordinal.**
  4. Hospital location (London, Boston, Madison, Rochester, Toronto) – **nominal.**
  5. Favorite beverage (beer, juice, milk, soft drink, wine, other) – **nominal.**
  6. How often feel depressed (never, occasionally, often, always) – **ordinal.**

1.3

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

1. *Specify the distribution of the student’s number of correct answers on the exam:*

Let X count the number of correct answers on the exam. Then X~Binomial(100, 0.25).

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

We can approximate the binomial distribution with a normal distribution, using the mean and variance as the parameters of the normal distribution.

E(X) = np = 100 \* 0.25 = 25

Var(X) = np(1-p) = 100 \* 25 (1 – 0.25) = 18.75

So, X~N(25, 18.75). To answer the question, I will find the probability of the student getting a score of fifty or more.

P(X ≥ 50 ) = 1 – P(X ≤ 50) = 1 – P(Z ≤ (50 – 25)/) = 1 – P(Z ≤ 5.773) = 1 – 1 = 0

There is essentially a 0% chance that this student will score 50 or higher on the exam. I would be shocked and amazed if this ever occurred.

1.6

*Genotypes AA, Aa, and aa occur with probabilities (p1, p2, and p3). For n = 3 indppendnet observations the observed frequencies are (n1, n2, n3).*

1. *Explain how you can determine n3 from knowing n1 and n2. Thus, the multinomial distribution of (n1, n2, n3) is actually two-dimensional.*

As far as we’ve been told, AA, Aa, and aa are the only three possible genotypes. In other words, everyone will fall into one of those three categories. Thus, p1 + p2 + p3 = 1. If p1 and p2 are fixed, p3 is also fixed to equal 1 – (p1 + p2).

1. *Show* the set of all *possible observations, (n1, n2, n3) with n = 3.*

(AA, AA, AA), (AA, AA, Aa), (AA, Aa, AA), (Aa, AA, AA), (AA, Aa, Aa), (Aa, AA, Aa), (Aa, Aa, AA), (Aa, Aa, Aa), (AA, AA, aa), (AA, aa, AA), (aa, AA, AA), (AA, aa, aa), (aa, AA, aa), (aa, aa, AA), (aa, aa, aa), (Aa, Aa, aa), (Aa, aa, Aa), (aa, Aa, Aa), (Aa, aa, aa), (aa, Aa, aa), (aa, aa, Aa), (AA, Aa, aa), (AA, aa, Aa), (aa, AA, Aa), (aa, Aa, AA), (Aa, AA, aa), (Aa, aa, AA).

Or

(AA, AA, AA)

(AA, AA, Aa)

(AA, AA, aa)

(AA, Aa, Aa)

(AA, Aa, aa)

(AA, aa, aa)

(Aa, aa, aa)

(Aa, Aa, Aa)

(aa, aa, aa)

1. *Suppose (π1, π2, π3) = (0.25, 0.50, 0.25). Find the multinomial probability that (n1, n2, n3) = (1, 2, 0).*

P(1, 2, 0) = π1n1 π2n2 π3n3  = 0.251 0.502 .250 = .1875

The P(n1, n2, n3) = (1, 2, 0) = .1875

1. *Refer to (c). What probability distribution does n1 alone have? Specify the values of the sample size index and parameter for that distribution.*

n1 will have a binomial distribution. Each observation will either be n1 or not n1. Each trial is independent and identical.

n1~Binomial(0.25, 3).

1.7

*Russian Roulette problem…*

1. *Find the probability of playing Russian Roulette six times and never having the bullet fire.*

P(Playing 6 times with no bullet fired) =

1. *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 (5/6) y-1 (1/6) for y = 1,2,3,…*

Playing 6 times, we had for the bullet never firing. Let us assume that the bullet fires on the seventh game. The probability of this happening would be . We multiply the probabilities of each sub event together to find the probability of the overall event happening. In our case here, y = 7. So .

This will hold for any value of y, where y = 1, 2, 3,…

1.8

*When the 2000 General Social Survey asked subjects whether they would be willing to accept cuts in their standard of living to protect the environment, 344 of 1170 subjects said “yes.”*

1. *Estimate the population proportion who would say “yes.”*

An unbiased estimate for p is

1. *Conduct a significance test to determine whether a majority or minority of the population would say “yes.” Report and interpret the p-value.*

H0 : p ≥ 0.5

HA : p < 0.5

z =

P(Z < -15.83) = 0.

Our p-value = 0 < 0.05. There is essentially no chance that we would have gotten the value of that we did if p is greater than or equal to 0.5. If we surveyed the entire population, a minority would say ‘yes’.

1. *Construct and interpret a 99% confidence interval for the population proportion who would say ‘yes’.*

A 99% confidence interval for p =

Our confidence interval is (0.256, 0.324). We are 99% confident that our interval covers the true proportion of the population that would say ‘yes’.

2.17

*Refer to table 2.3. Find the P-value for testing that the incidence of heart attacks is independent of aspirin intaking using (a) X2, (b) G2. Interpret results.*

1. *X2 =*

2.18

2.21