DALITE Q6 - Binomial Distribution and Inference for one Proportion. Solutions.

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on October 22, 2018

This DALITE quiz will cover the binomial distribution and inference for a sample proportion. You need to understand the binomial distribution before moving on to the chapter on one sample proportions. This is analgous to learning the normal distribution before inference for a single mean.

Binomial distribution | One sample proportion | Hypothesis testing

1. Binomial Distribution 1

In which of the following would Y not have a Binomial distribution? Provide your justification in the rationale.

a. **Y = Number, out of 60 occupants of 30 randomly chosen cars, wearing seatbelts. (Correct)**
b. Y = Number, out of 60 occupants of 60 randomly chosen cars, wearing seatbelts.
c. Y = Number, out of simple random sample of 100 individuals, that are left-handed.

1.1. Correct rationales.

- 60 occupants from 30 cars would mean more than one occupant per car - whether each occupant is wearing a seatbelt or not isn't independent if more than one comes from the same car.
- There is a chance that if someone in the car is not wearing a seatbelt, the other passenger in the car is not too. This means that it is not independent. #idiotswhodon'twearseatbeltsdrivetogether
- Not independent, idiots who don't wear seatbelts in cars drive together!!! VROOM VROOM
- If there are more than one occupant in a car, they may influence each other to wear or not wear a seat belt.
  - Richard: "hey Timmy, only losers wear seat belts, be cool like me and don't wear your seat belt"
  - Timothy: "oh snap, you're right Ricky, not wearing a seat belt is the fleekest"
  - Richard and Timothy: "we're going to live forever"
- The outcome does not fall into one of two categories as someone can be left-handed, right-handed, or ambidextrous.

1.2. Incorrect rationales.

- Each trial does not have an equal probability of success as there are more right-handed people than left-handed people.
- Because in this case the observations are not independent. If the first volunteer selected is left-handed, the second one is more likely to be right-handed because there are more right-handed individuals than left-handed in the remaining sample.

2. Binomial Distribution 2

In which of the following would Y not have a Binomial distribution? Provide your justification in the rationale.

a. **You want to know what percent of married people believe that mothers of young children should not be employed outside the home. You plan to interview 50 people, and for the sake of convenience you decide to interview both the husband and the wife in 25 married couples. The random variable Y is the number among the 50 persons interviewed who think mothers should not be employed. (Correct)**
b. You observe the sex of the next 50 children born at a local hospital; Y is the number of girls among them.
c. Y = number of occasions, out of a randomly selected sample of 100 occasions during the year, in which you were indoors.

2.1. Correct rationales.

- Husband's opinion may depend on wife's opinion, and vice versa - observations aren't independent.
- One of the conditions for the binomial distribution is that the trials must be independent. In A, they are selecting 25 married couples in order to gather data on 50 people, but in this scenario either the husband or the wife could influence the other, so the trials are not independent.

2.2. Incorrect rationales.

3. Binomial Distribution 3

The U.S. National Center for Health Statistics reports that approximately 12% of emergency department visits result in hospital admissions. Consider 20 randomly selected emergency department visits and assume that visits to emergency departments are independent. What is the approximate probability that at most 2 of the 20 visits would result in hospital admissions?

a. **0.5631 (Correct)**
b. 0.2740
c. 0.1344
d. 0.2891
e. 0.4369

3.1. Correct rationales.

- We have to callculate the probability at 0,1, and 2 and add them up: For 0: (1)(1)(0.88)ˆ20 For 1: 20(0.12)(0.88)ˆ19 For 2: $190(0.12)^{2(0.88)}18$ Adding these all up equals 0.5631
- P(X< or=2) = P(X=2)+P(X=1)+P(X=0) = 0.5631
- pbinom(2, 20, 0.12, lower.tail = TRUE)

3.2. Incorrect rationales.

- 1- pbinom(2, size = 20, prob = 0.12)

4. Binomial DIstribution 4

The U.S. National Center for Health Statistics reports that approximately 12% of emergency department visits result in hospital admissions. Consider 20 randomly selected emergency department visits and assume that visits to emergency departments are independent. How many hospital admissions do we expect, on average, in a random sample of 20 emergency department visits?

a. 2
b. **2.4 (Correct)**
c. 12
d. 24
e. 1.4533

4.1. Correct rationales.

- mean of a binomial distribution = np = 20*0.12 = 2.4
- μ = np = (20)(0.12) = 2.4 hospital admissions

4.2. Incorrect rationales.

- 0.12 x 20 = 2.4 But we can't in a random sample get .4, so we choose 2.

5. Binomial Distribution 5

The U.S. National Center for Health Statistics reports that approximately 12% of emergency department visits result in hospital admissions. Consider 20 randomly selected emergency department visits and assume that visits to emergency departments are independent. If Y is the number of emergency department visits that result in hospital admissions in random samples of 20 visits, what is approximately the standard deviation of Y?

a. **1.45 (Correct)**
b. 2.11
c. 4.46

5.1. Correct rationales.

- The standard deviation of U can be solved using the equation sigma = sqrt(np(1-p)) = sqrt(20.*12*(1-0.12)) = 1.45

5.2. Incorrect rationales.

6. Binomial Distribution 6

According to the 2015 U.S. census update, approximately 13% of Americans are black. Let Y be the number of blacks in a random sample of 1500 Americans. What is the probability that the sample will contain 200 or more blacks?

a. pbinom(q = 200, size = 1500, prob = 0.13)
b. 1-dbinom(x = 200, size = 1500, prob = 0.13)
c. dbinom(x = 200, size = 1500, prob = 0.13)
d. `pnorm(200, mean = 195, sd = 13.025, lower.tail = FALSE)` **(Correct)**
e. `1 - pbinom(q = 200, size = 1500, prob = 0.13, lower.tail = FALSE)`

6.1. Correct rationales.

- a) is not correct because it is looking at all values below 200 b) is not correct because it is looking at the probability of getting anything but 200 c) is not correct because it is looking at getting exactly 200 e) is not correct because it is looking at all values below **201** d) since sample size is large, we can use normal approximation for binomial distribution
- In this case, we can use the normal approximation because n is large enough, n*p = 195 > 10, n(1-p) = 1305 > 10. Therefore, we can use pnorm function with q=200, lower.tail false. A doesn't specify a lower.tail=FALSE, so R gives lower.tail = TRUE by default. dbinom are wrong.

6.2. Incorrect rationales.

- because you are calculating the sample that will contain 200 or more, lower.tail is false.
- E) would be right if it didn't have the 1 - before the code.
- Enough ppl not to do a t-test

7. One Sample Proportion 1

A CDC report on secondhand smoke at home gives the following 95% confidence interval for the percent of California households that are free of secondhand smoke: (90.8, 92.2). The correct interpretation for this confidence interval is that we can be 95% confident that

a. the proportion of households free of secondhand smoke in another sample of California households would be between 0.908 and 0.922
b. the population mean number of households in California that are free of secondhand smoke is between 90.8 and 92.2
c. **the true proportion of all California households that are free of secondhand smoke is between 0.908 and 0.922 (Correct)**

7.1. Correct rationales.

- The answer is C because the confidence interval looks at where the true population proportion will fall 95% of the time.
- A is not right because the CI is only for that trial, not for all trials. B is not right because they are not talking about a mean value but a proportion.

7.2. Incorrect rationales.

8. One Sample Proportion 2

A CDC report on secondhand smoke at home gives the following 95% confidence interval for the percent of California households that are free of secondhand smoke: (90.8, 92.2). What is the margin of error for this interval?

a. **0.007 (Correct)**
b. 0.014
c. 0.028

8.1. Correct rationales.

- The margin of error is half the width of the confidence interval.
- The margin of error is half of the confidence interval. 92.2 - 90.8 / 2 = 0.007

8.2. Incorrect rationales.

9. One Sample Proportion 3

How many observations must be recorded to estimate a population with unknown proportion p to within +/- 0.02 with 95% confidence?

a. n=25
b. n=1225
c. **n=2401 (Correct)**
d. n=2350
e. n=1691

9.1. Correct rationales.

- We would have to guess p in this case because it is not given. It is best to assume p is 0.5 when it is unknown because the margin of error is largest when p is 0.5 so by doing so we are giving a more conservation estimation for how many people we will need. We then use the formula n=(z/m)^2(p)(1-p) where p=0.5, z=1.96, m=0.02.

9.2. Incorrect rationales.

- 1691 = (1.645/0.02)^2 * (0.5 * (1-0.5))
- Using the formula: n = (z*/m)(p)(1-p) = (1.96/0.02)(0.5)(1-0.5) = 24.5 = 25. Here, test is conservative and p is set to 0.5.