MSCA 31000 - Introduction to Statistical Concepts

Chapter 5

Q11: You win a game if you roll a die and get a 2 or a 5. You play this game 60 times.

a. What is the probability that you win between 5 and 10 times (inclusive)?
b. What is the probability that you will win the game at least 15 times?
c. What is the probability that you will win the game at least 40 times?
d. What is the most likely number of wins.
e. What is the probability of obtaining the number of wins in d?

Answer:

a. This is a binomial distribution with N = 60, x = 5, 6, 7, 8, 9, 10 and probability = 2/6 = 0.33 and summing up

   Plugging into formula for binomial distribution, answer is **0.00371**

b. Plugging N = 60, x = 15 to 60, probability = 0.33 (in R: pbinom(15,60,.33)) and summing up

   Answer is **0.88**

c. Plugging N = 60, x = 40 to 60, probability = 0.33 (in R: pbinom(40,60,.33,lower.tail=F)) and summing up

   Answer is **2.355708e-08**, which is nearly 0.

d. 20

e. pbinom(20,60,.33) - pbinom(19,60,.33) = **0.10856**

Q13: An unfair coin has a probability of coming up heads of 0.65. The coin is flipped 50 times. What is the probability it will come up heads 25 or fewer times? (Give answer to at least 3 decimal places).

Answer: pbinom(25,50,.65) = 0.021

Q15: True/False: You are more likely to get a pattern of HTHHHTHTTH than HHHHHHHHTT when you flip a coin 10 times.

Answer: True.

A = Probability of HTHHHTHTTH (6 heads) = pbinom(6,10,.5) - pbinom(5,10,.5) = **0.2050781**

B = Probability of HHHHHHHHTT (8 heads) = pbinom(8,10,.5) - pbinom(7,10,.5) = **0.04394531**

A > B, hence the statement is True.

Chapter 7

Q12: Use the normal distribution to approximate the binomial distribution and find the probability of getting 15 to 18 heads out of 25 flips. Compare this to what you get when you calculate the probability using the binomial distribution. Write your answers out to four decimal places.

Answer:

a) pbinom(18,25,.5) - pbinom(14,25,.5) = 0.2049

b) Since the value of N * p = 25 * .5 = 12.5 which is greater than 10, we can approximate the normal distribution to a binomial distribution

Here, mean = N * p = 12.5
Std dev = sqrt(N * p * (1 - p)) = 2.5

pnorm(18.5,12.5,2.5) - pnorm(14.5,12.5,2.5) = 0.2037

There is a difference of **0.0012**.

Q14: True/false: In a normal distribution, 11.5% of scores are greater than Z = 1.2.

Answer: True. In a standard normal distribution, pnorm(1.2,0,1,lower.tail=F) = **0.115**.

Q16: True/false: The larger the π, the better the normal distribution approximates the binomial distribution.

Answer: False. Thumb rule: Approximation is good if both Nπ and N(1-π) are both greater than 10.

Q18: True/false: Abraham de Moivre, a consultant to gamblers, discovered the normal distribution when trying to approximate the binomial distribution to make his computations easier.

Answer: True.

Q20: True/false: In the figure below, the red distribution has a larger standard deviation than the blue distribution.

Answer: True.

Q22: For this problem, use the Anger Expression (AE) scores. (a) Compute the mean and standard deviation. (b) Then, compute what the 25th, 50th and 75th percentiles would be if the distribution were normal. (c) Compare the estimates to the actual 25th, 50th, and 75th percentiles.

Answer:
(a) mean: **37**, std dev: **12.94**

b) Computing percentile estimates assuming a normal distribution

c <- seq(.25,.75,.25)
Quantiles for Mean = qnorm(c, mean = 37, sd = 12.94)

25th percentile	50th percentile	75th percentile
28.2721	37	45.7279

c)

Actual percentiles: quantile(angerdata$Anger.Out)

25th percentile	50th percentile	75th percentile
27.00	36.00	44.75

The estimates (in part b) vary from the actual percentiles (estimates are higher) since we have assumed a normal distribution, which is continuous, for estimation whereas the actual quantile values are discrete.

Chapter 9

Q23: (a) How many men were sampled? (b) How many women were sampled?

Answer:

a. # of men sampled = 30
b. # of women sampled = 48

Q24: What is the mean difference between men and women on the Anger-Out scores?

Answer:

- mean difference in sample = **0.7958**

Q25: Suppose in the population, the Anger-Out score for men is two points higher than it is for women. The population variances for men and women are both 20. Assume the Anger-Out scores for both genders are normally distributed. Given this information about the population parameters:

(a) What is the mean of the sampling distribution of the difference between means?
(b) What is the standard error of the difference between means?
(c) What is the probability that you would have gotten this mean difference (see #24) or less in your sample?

Answer:

(a) 2
(b) 1.0408
(c) pnorm(0.7958333, mean=2, sd=1.040833) = **0.12365**