Problem 1 /15

Use R for all quantitative calculations. Include code and output in all cases.

This problem refers to the `carseats_sample` data set, available on Moodle. This set includes information about pricing and marketing of a certain brand of carseat at a sample of 75 stores in 2014.

(a) Is the `US` variable quantitative or qualitative? What is the level of measurement? Briefly explain your answers.

The US variable is a qualitative/catagorical nominal variable. Although the variable has been assigned 1 or 0, what it is describing is descriptive with no particular order.

(b) Compute the five-number summary and inter-quartile range of competitor prices at the stores in this set.

quantile(carseats_sample $CompPrice)

0%	25%	50%	75%	100%
77.0	115.5	127.0	136.5	161.0
min	Q_1	m	Q_3	max

$IQR = Q_3 - Q_1 = 21$ $1.5(IQR) = 1.5(21) = 31.5$

$Q_1 - 1.5(IQR) < x < Q_3 + 1.5(IQR)$

$Q_1 - 1.5(IQR) = 115.5 - 31.5 = 84.0$

$Q_3 + 1.5(IQR) = 136.5 + 31.5 = 168.0$

$84.0 < x < 168.0$

(c) Determine the 60^{th} percentile of carseat prices (not competitor prices) in this data set.

quantile(carseats_sample$Price, 0.6)

60%

122.4

(d) Compute a level 95% confidence interval for the mean price of carseats (not competitor price) assuming that the population standard deviation is \$25. Briefly explain your choice of method. Identify the point estimate, margin or error, and endpoints of the interval.

$\mu = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ z^* & σ are used since pop. sd is given

$\bar{x}$ = point estimate = mean(carseats_sample$Price) = 115.5

$z^* \frac{\sigma}{\sqrt{n}}$ = margin of error = 1.960 * 25/sqrt(75) = 5.66

1.960 comes from CI of 95%

Endpoints: $115.5 - 5.66 < x < 115.5 + 5.66 = 109.8 < x < 121.2$

t.test() not used since z_* can be used & t.test uses t_*

Problem 2 /15

Use R for all quantitative calculations. Include code and output in all cases.

This problem refers to the `jumping` data set, available on Moodle. This set gives the heights of seven children (in cm) and the horizontal distances they jumped (in cm) in a simple experiment. The data is also displayed in the following table.

Height	134	134	136	135	130	138	132
Distance	132	106	125	150	105	130	114

(a) In just a sentence or two, explain the circumstances under which it would be appropriate to calculate the correlation coefficient of these two variables. No R code is needed for this part.

If the relationship is linear then it is appropriate to calculate the correlation coefficient

(b) Assume the conditions in part (b) are met and determine the correlation of these two variables.

cor(jumping$height, jumping$distance) = 0.590

(c) Find the equation of the least-squares regression line. Use `height` as the explanatory variable.

lm(distance ~ height, data = jumping)

Coefficients:

(Intercept)	height
-365.587	3.643

distance = 3.643(height) - 365.587

(d) What is the predicted jump distance of a child with height 150cm? If this calculation isn't appropriate, briefly explain why.

distance $= (3.643 * 150) - 365.587 = 180.863$

It is not appropriate to estimate the jump distance of a child with a height of 150 cm because 150cm is outside the range of observed heights so it would be an extrapolation that could give a probamatic value.

(e) What is the residual of the child with height 135 cm? Briefly interpret this number in ordinary human language.

residual = observed - fitted $= 150 - 126.218 = 23.782$

observed $= 150$

fitted $= (3.643 * 135) - 365.587 = 126.218$

The child with a height of 135cm jumped 23.782 cm farther than would be expected for a child of that height.

Problem 3 /15

Use R for all quantitative calculations. Include code and output in all cases.

A hospital emergency room classifies incoming patients as either high, medium, or low priority, hopefully with equal proportions. As part of an internal audit, a random sample of 140 patients is collected. The results are as follows.

- 49 patients were classified as high priority
- 63 patients were classified as medium priority
- 28 patients were classified as low priority

Is the hospital in alignment with its own standard? Test at significance level $\alpha = .05$. Use **both** of the methods covered in class and follow all of the best practices we have established. Make sure your process is clear!

H_0 = patients are classified in equal proportions

H	M	L

ratio <- c(1/3 | 1/3 | 1/3)

H_a = patients are not classified in equal proportions

	H	M	L
observed <- c(	49	63	28)
expected <- c(	46.7	46.7	46.7) = ratio * 140

χ^2 = sum((observed - expected)^2 / expected) = 13.3

df = 3 - 1 = 2

p = 1 - pchisq(χ^2, df) = 0.001

---

chisq.test(observed, p=ratio)

χ^2 = 13.3 df = 2 p-value = 0.001

$p < \alpha$ therefore the data does not support that the patient priority is evenly distributed and the hospital is in alignment with its standards.