More Review Problems Solutions

1. A particularly common question in the study of wildlife behavior involves observing contests between "residents" of a particular area and "intruders." In each contest, the "residents" either win or lose the encounter (assuming there are no ties). Observers might record several variables, some of which are listed below. Determine which variables are categorical and which ones are quantitative.

   (a) The duration of the contest (in seconds).

   (b) The number of animals involved in the contest.

   (c) Whether the "residents" win or lose.

   (d) The total number of contests won by the "residents."

   Answers:

   (a) Quantitative
   (b) Quantitative
   (c) Categorical
   (d) Quantitative

   $\square$

2. Of the quantitative variables identified in Problem 1, which ones are continuous and which ones are discrete?

   Answers:

   (a) Continuous
   (b) Discrete
   (c)
   (d) Discrete

   $\square$

3. The asking prices (in thousands of dollars) for a sample of 13 houses currently on the market in Neighborville are listed below. For convenience, the data have been ordered.

   175 199 205 234 259 275 299 304 317 345 355 384 549

   (a) What is the five-number summary?

(b) Use the $1.5 \times \text{IQR}$ rule to determine if there are any outliers present. What is/are the value(s) of the outlier(s)?

(c) Sketch a boxplot of the data.

Answers:

(a) The five–number summary is

Min.	Q1	Median	Q3	Max.
175	234	299	345	549

(b) The inter–quartile range is $\text{IQR} = Q_3 - Q_1 = 111$; thus, $1.5 \times \text{IQR} = 166.5$. Going this distance above the third quartile we get 511.5, which is below the maximum value. Thus, we can consider 549 to be an outlier.
Going $1.5 \times \text{IQR}$ below the first quartile, we get 67.5, which is below all the values, so there are no outliers at the bottom of the distribution.

(c) The boxplot is shown in Figure 1.

Figure 1: Box Plot of House Prices

□

4. A study was conducted in a large population of adults concerning eyeglasses for correcting reading vision. Based on an examination by a qualified professional

the individuals were judged as to whether or not they needed to wear glasses for reading. In addition it was determined whether or not they were currently using glasses for reading. Table 1 provides the proportions found in the study:

Need Grasses? \ Used Glasses?	Yes	No
Yes	0.42	0.18
No	0.04	0.36

Table 1: Proportions for Problem 4

(a) If a single adult is selected at random from this large population, what is the probability that the adult is judged to need eyeglasses for reading?

(b) What is the probability that the selected adult is judged to need eyeglasses but does not use them for reading?

(c) Suppose two adults are selected from the population independently, and at random. What is the probability that both were judged to need eyeglasses and neither was using them for reading?

Answers: Table 2 shows the marginal distributions of the variables X, whether the person uses glass or not, and the variable Y, whether the person has been judged to need glasses or not.

Need Grasses? \ Used Glasses?	Yes	No	p_Y
Yes	0.42	0.18	0.60
No	0.04	0.36	0.40
p_X	0.46	0. 54	1

Table 2: Two-Way Table for Problem 4

(a) The probability that the adult is judged to need eyeglasses for reading is $p_Y(\text{“Yes”}) = 0.60$ or 60%.

(b) The probability that the selected adult is judged to need eyeglasses but does not use them for reading is the joint probability

$$P(X = \text{“No”},\ Y = \text{“Yes”}) = 0.18.$$

(c) Since the events are independent and the probability of each event is 0.18, by the previous part, it follows that the probability of the joint events is the product $(0.18)\cdot(0.18) = 0.0324$.

□

5. A company is being criticized because only 3 of 16 people in executive-level positions are female. The company claims that although the number is lower than it might wish, it is not surprising given the fact that only 40% of their employees are women. Suppose we can consider the 16 people in executive-level positions as a simple random sample of all employees. What is the probability of observing 3 or fewer women in the group of 16 executives?

   ***Solution**:* Assume that the selection of the 16 executives is done at random, and let X denote the number of women in the selected group of 16. Then, we may assume that X follows a binomial distribution with parameters $n = 16$ and $p = 0.40$. We then want to compute

$$P(X \leqslant 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3),$$

   where

$$P(X = k) = \binom{16}{k}(0.4)^k(0.6)^{16-k} \quad \text{for } k = 0, 1, 2, \ldots, 16.$$

   We then obtain that

$$\begin{aligned} P(X \leqslant 3) &= (0.6)^{16} + 16(0.4)(0.6)^{15} + 120(0.4)^2(0.6)^{14} + 560(0.4)^3(0.6)^{13} \\ &\approx 0.00028 + 0.00301 + 0.01505 + 0.04681 \\ &\approx 0.065, \end{aligned}$$

   or about 6.5%. □

6. The scores in the second midterm of this course had a mean of 85.35 and a standard deviation of about 8.16. If the scores were normally distributed then we would expect to see the following distribution of scores:

$$10 \quad 7 \quad 3 \quad 7 \quad 10$$

   in a class of 37 in the ranges

   score $\leqslant$ 80.45,
   80.45 $<$ score $\leqslant$ 84.53,
   84.53 $<$ score $\leqslant$ 86.18,
   86.18 $<$ score $\leqslant$ 90.25,
   score $>$ 90.25,

   respectively. The actual distribution was

$$11 \quad 5 \quad 3 \quad 6 \quad 12$$

Perform a goodness of fit test to determine if the distribution of scores is close to normal.

***Solution**:* The Chi–Square distance in this case turns out to be

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \approx 1.21$$

We can now run simulations to generate Chi–Squared distances under the assumption that that the exam scores come from a normal distribution with mean 85.35 and standard deviation 8.16 to obtain an array, `ChiSqr`, of Chi–Squared distances whose frequency histogram is shown in Figure 2. The p–value can then be estimated as usual

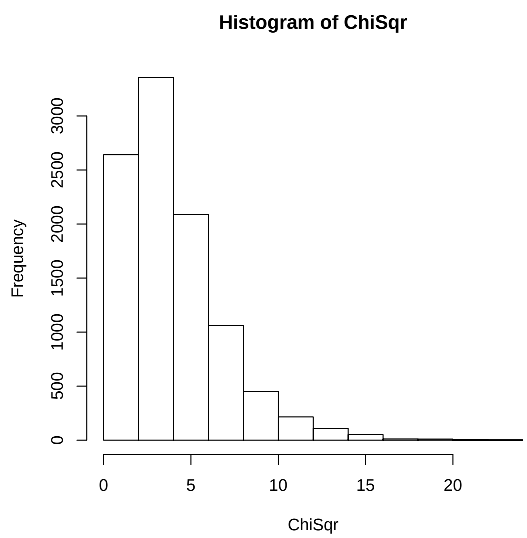

Figure 2: Histogram of `ChiSqr`

with the `pHat()` function

$$p\text{–value} \approx \texttt{pHat(ChiSqr,Xsqr)} \approx 0.8851,$$

which is a very high p–value. We therefore cannot reject the null hypothesis that the scores come from a normal distribution. $\square$