

Assignment #2 SOLUTIONS

1)

A distributor of Canada Dry was interested in examining the relationship between the number of ads (x) for their ginger ale product during prime time on a local television station and the number of sales per week (y) in 1000s of cases. She compiled the figures for 20 weeks and computed the following summary information:

$$n = 20, \sum x_i = 92, \sum y_i = 177, \sum x_i y_i = 884, s_x = 1.3917, \text{ and } s_y = 2.9069.$$

1a.) Find the correlation coefficient for the number of ads during prime time and weekly sales.

(2 points)

ANS:

$$\text{Since } s_{xy} = \frac{884 - \frac{(92)(177)}{20}}{19} = 3.6737, \text{ then } r = \frac{3.6737}{(1.3917)(2.9069)} = 0.908.$$

1b) Interpret the correlation coefficient as to whether the number of ads is associated with sales. Justify your answer.

(1 point)

ANS:

There is a strong positive linear relationship between the ads and sales.

1c) Find the best-fitting line relating the number of ads during prime time and weekly sales.

(2 points)

ANS:

$$\text{Since } b = 0.908 \left(\frac{2.9069}{1.3917} \right) = 1.8966, \text{ and } a = \frac{177}{20} - (1.8966) \left(\frac{92}{20} \right) = 0.1256, \text{ then } \hat{y} = 0.1256 + 1.8966x.$$

1d) If the Canada Dry distributor ran 21 TV ads per week for her product, what would you predict her sales to be?

(1 point)

ANS:

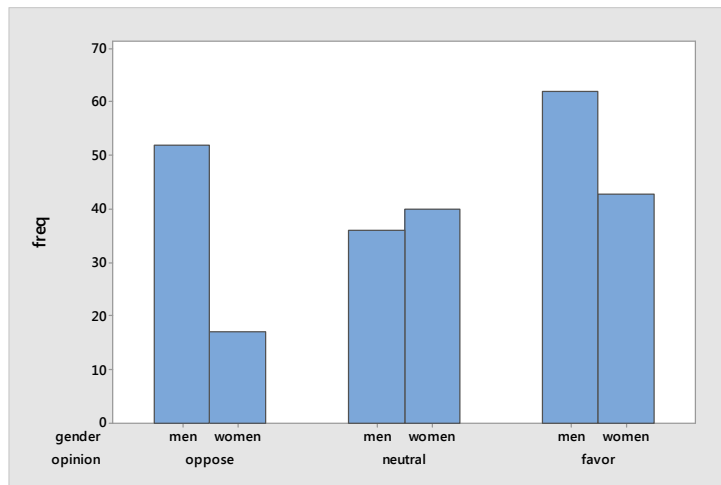
$$\hat{y} = 0.1256 + 1.8966(21) = 39.9542 \text{ thousand cases}$$

2)

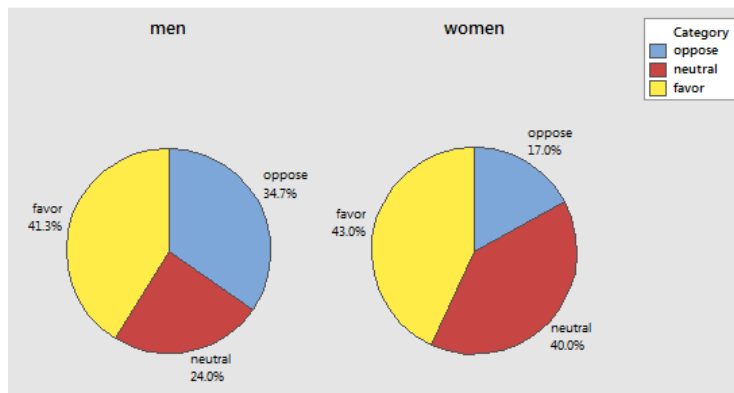
Concordia students were asked to indicate whether they favored or opposed closing down the Tim Horton's in the LB building for a Starbucks. They were categorized into three groups according to their answers, as shown below:

	Oppose	Neutral	Favor
Men	52	36	62
Women	17	40	43

2a.) Create a side-by-side bar chart to describe these data.
(1 point)



2b) Create two pie charts (one for men and one for women) to describe these data.
(1 point)



2c) Describe the different pros/cons of the bar and pie charts in their abilities to depict the differences and similarities of the responses of men and women.

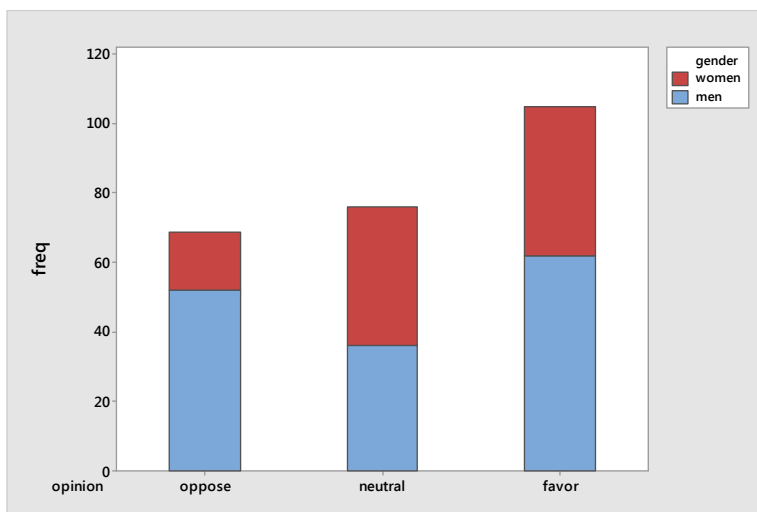
(1 point)

ANS:

The differences in the proportion of men and women in the three groups is best graphically portrayed by the pie charts, since the unequal number of men and women tend to confuse the interpretation of the bar charts. However, the bar charts are useful in retaining the actual frequencies in each group, which are lost in the pie chart.

2d) Create a stacked bar chart to describe these data.

(1 point)



3) The distribution of final exam scores in an inferential statistics class is normal, with standard deviation 6. Find the mean score if 93.32% of the students received scores no higher than 90.

(2 points)

ANS:

Let the normal random variable x be the final exam score. Since $P(x \leq 90) = 0.9332$, then P

$$\left(z \leq \frac{90 - \mu}{6} \right) = 0.9332. \text{ Hence, } (90 - \mu)/6 = 1.5, \text{ which implies that } \mu = 81.$$

4) Concordia University decides to check travel expenses to see if they are correctly reported. An auditor selects 200 expense reports at random to audit. What is the probability more than 20% of the sampled reports will be incorrect when, in fact, only 10% of the university's expense reports are improperly documented?

(2 points)

ANS:

Let the random variable x denote the number of travel reports improperly documented; x is binomial with $n = 200$ and $p = 0.10$. Then, using the normal approximation to the binomial with

$\mu = np = 20$, and $\sigma = \sqrt{n \cdot p \cdot (1 - p)} = 4.243$, we have $P(x > 40) \approx P\left(\frac{39.5 - 20}{4.243}\right) = P(z > 4.83) = 0$.

5)

The time necessary to complete an examination has a mean of 45 minutes and a standard deviation of 4 minutes. The average time necessary for 64 randomly selected students is computed.

5a) What is the mean and standard deviation of the sampling distribution of \bar{x} ?

(1 point)

ANS:

The mean: $\mu_{\bar{x}} = \mu = 45$ minutes.

The standard deviation is $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 0.5$ minutes.

5b) Can we say that the sampling distribution of \bar{x} is approximately normally distributed? Why or why not?

(1 point)

ANS:

Yes, because $n = 64 > 30$; our rule-of-thumb minimum value for large samples, therefore the CLT can be applied to conclude that the sampling distribution of \bar{x} is approximately normally distributed.

6)

NBC believed that 30% of Canadian TV households would be tuned in to the NHL All-Star game. A random sample of 500 TV households is selected.

6a) Describe the sampling distribution of the sample proportion.

(1 point)

ANS

Since $np = 150$ and $n(1 - p) = 350$ are both greater than 5, the sampling distribution is

approximately normal with a mean $\mu_{\hat{p}} = p = 0.30$ and a standard deviation $\sigma_{\hat{p}} = \sqrt{pq/n} = 0.02049$.

6b) Find the probability that the sample proportion watching the game will be between 0.25 and 0.35.

(2 points)

ANS:

$$P(0.25 \leq \hat{p} \leq 0.35) = P(-2.44 \leq z \leq 2.44) = 2(0.4927) = 0.9854$$

7) The distribution of IQ scores for CEGEP graduates is normally distributed with $\mu = 104$ and $\sigma = 16$. What percentage of the IQ scores would be between 97 and 126?

(2 points)

ANS:

$P(97 \leq x \leq 126) = P(-0.44 \leq z \leq 1.38) = 0.5862$. Thus, approximately 58.6% of the IQ scores would lie between 97 and 126.

8)

The number of household members, x , and the amount spent on groceries per week, y , rounded to the nearest dollar, are measured for eight households in Montreal. The data are shown below:

x	5	2	2	1	4	3	5	3
y	140	100	75	40	125	70	130	65

8a) Find the best-fitting regression line for these data.

(3 points: 2 points off if work not shown)

$$\hat{y} = 26.6 + 21.30x$$

8b) What would you estimate a household of seven to spend on groceries per week? Should you use the fitted line to estimate this amount? Why or why not?

(1 point)

When $x = 7$, the estimated value of y is $\hat{y} = 175.67$. However, it is risky to try to estimate the value of y for a value of x outside the experimental region, that is, the range of x values for which you have collected data.