Let the random variable X represent the number of telephone lines in use by the technical support center of a software manufacturer at noon each day. The probability distribution of X is shown in the table below.

x	0	1	2	3	4	5
$p(x)$	0.35	0.20	0.15	0.15	0.10	0.05

(a) Calculate the expected value (the mean) of X.

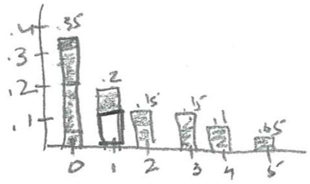
$0 \cdot .35 + 1 \cdot .20 + 2 \cdot .15 + 3 \cdot .15 + 4 \cdot .1 + 5 \cdot .05 = 1.6$

The expected # of lines in use at noon is 1.6.

(b) Using past records, the staff at the technical support center randomly selected 20 days and found that an average of 1.25 telephone lines were in use at noon on those days. The staff proposes to select another random sample of 1,000 days and compute the average number of telephone lines that were in use at noon on those days. How do you expect the average from this new sample to compare to that of the first sample? Justify your response.

larger samples are more likely to be closer to the expected value so I would expect the larger sample to be closer to 1.6 than 1.25 was.

(c) The median of a random variable is defined as any value x such that $P(X \leq x) \geq 0.5$ and $P(X \geq x) \geq 0.5$. For the probability distribution shown in the table above, determine the median of X.



The median would be 1 phone in use.

(d) In a sentence or two, comment on the relationship between the mean and the median relative to the shape of this distribution.

Because the distribution is skewed right I would expect the mean to be higher than the median due to the median being more robust and less affected by skew or outliers.

— And it was.

The manager of a large company that sells pet supplies online wants to increase sales by encouraging repeat purchases. The manager believes that if past customers are offered \$10 off their next purchase, more than 40 percent of them will place an order. To investigate the belief, 90 customers who placed an order in the past year are selected at random. Each of the selected customers is sent an e-mail with a coupon for \$10 off the next purchase if the order is placed within 30 days. Of those who receive the coupon, 38 place an order.

(a) Is there convincing statistical evidence, at the significance level of $\alpha = 0.05$, that the manager's belief is correct? Complete the appropriate inference procedure to support your answer.

Conditions
Random - Says it
Ind - ✓ less than 10% of pop
Normal - $90\left(\frac{38}{40}\right) > 10$
$90\left(\frac{90-38}{90}\right) > 10$
Passes all conditions

$H_0: p = .4$
$H_a: p > .4$
$\alpha = .05$

$\frac{O - E}{SE} = \frac{\frac{38}{90} - .40}{\sqrt{\frac{p(1-p)}{n}}} = .427$

P-Value = .335

We do ~~NOT~~ reject our null that offering a gift card will increase return customers above .4

(b) Based on your conclusion from part (a), which of the two errors, Type I or Type II, could have been made? Interpret the consequence of the error in context.

Type II error could have been made.
This means we could have not rejecte cted a flawed null hypothesis. This means we do not implement the new coupon system, but it would have increased use of store and the store loses money.

A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

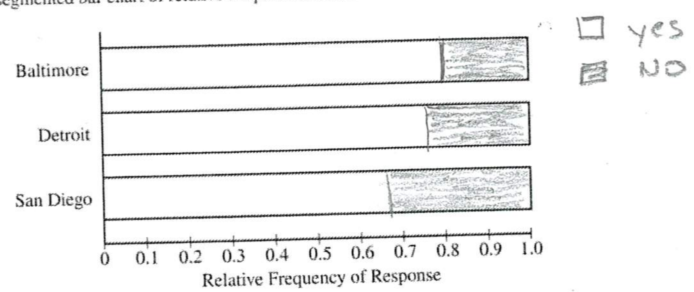
	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847
	.80	.74	.65	

(a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

The researches claim is incorrect, becaus the condition prob in balt is greater than that of the other cities.

(b) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

SAN DIEGO Had the smallest proportion who had a soft drink the previous week at .65.

(c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

(i) Identify the appropriate inference procedure. χ^2 test

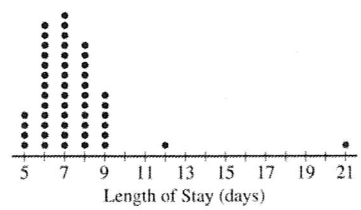
(ii) Identify the hypotheses of the test.

Ho: The city chosen will have no effect on proportin who had a soft drink in the previous week

Ha: The city chosen will have an effect on proportion who had a soft drink in the previous week

1. The length of stay in a hospital after receiving a particular treatment is of interest to the patient, the hospital, and insurance providers. Of particular interest are unusually short or long lengths of stay. A random sample of 50 patients who received the treatment was selected, and the length of stay, in number of days, was recorded for each patient. The results are summarized in the following table and are shown in the dotplot.

Length of stay (days)	5	6	7	8	9	12	21
Number of patients	4	13	14	11	6	1	1



(a) Determine the five-number summary of the distribution of length of stay.

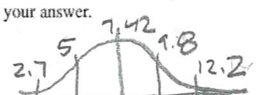
min = 5   Q1 = 6   MED = 7   Q3 = 8   max = 21

(b) Consider two rules for identifying outliers, method A and method B. Let method A represent the $1.5 \times \text{IQR}$ rule, and let method B represent the 2 standard deviations rule.

(i) Using method A, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer. I.Q.R. = 8-6 = 2   1.5 x IQR = 3   fences at 3 & 11

anything outside 3 & 11 are potential outliers so 12 & 21 are outliers.

(ii) The mean length of stay for the sample is 7.42 days with a standard deviation of 2.37 days. Using method B, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.



only 21 is outside 2 σ so there is only 1 outlier

(c) Explain why method A might identify more data points as potential outliers than method B for a distribution that is strongly skewed to the right.

than the median

The mean is less robust and therefore more effected by outliers.

The I.Q.R. method is centered on the median so it is less effected by the skew while 2 σ rule is centered on the mean, making it more susceptable to being pulled at by the skew.