

Practice Exam 1

Math 283

1. Distribution of number of computers X per family is given below (assume that $X \leq 3$)

x_i	0	1	2	3
p_i	0.25	0.5	0.15	?

- What is the proportion of families that have 3 computers?
- What is the proportion of families that have at least one computer?
- If two families were picked independently, what is the probability that both have exactly one computer?

2. In specimens of soil, the average pH equals 4.5 and standard deviation is 0.5. Assume Normal distribution.

- What is the probability that pH will be between 4.2 and 5.2?
- Find 17th percentile of pH values.
- Find an interval of pH values that would contain 95% of all values.

3. For a group of students, their scores for English course (X) and for Calculus course (Y) were recorded.

Student #	English(X)	Calculus.(Y)	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	3	1	0	0
2	3	4	0	1	0
3	3.5	2.5	0.25	0.25	
4	2	3.5	1	0.25	-0.5
5	2.5	2		1	0.5
Σ	15	15		2.5	-0.25

- Fill in the blanks
- Compute the standard deviation of X
- Compute correlation coefficient r .
- What can you say about the relationship between English and Calculus grades?

4. A Southern California water district measured daily evaporation for reporting to the local farmers. The farmers could then determine optimum watering regimen for their crops. One week's measurements are recorded below (in centimeters):

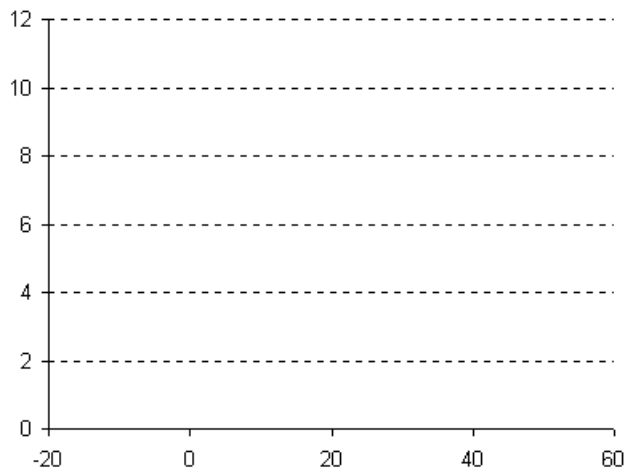
5.2, 4.0, 3.6, 4.5, 2.5, 4.8

- Find Q_1 (first quartile) and Q_3 (third quartile)
- Make a boxplot of the data.

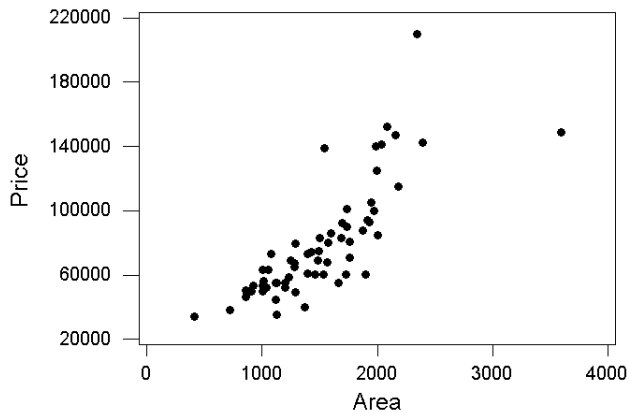
5. A stemplot for a sample of yearly returns (in %) on US stocks is given below.

```
-1 | 850
-0 | 742
 0 | 124589
 1 | 03458
 2 | 158
 3 | 29
 4 |
 5 | 3
```

- a) Find the median of the sample.
- b) Estimate the 80th percentile
- c) Describe the shape of stock returns distribution.
- d) based on the shape, which is true?
 - i. mean < median
 - ii. mean is very close to median
 - iii. mean > median
- e) Make a histogram of the returns, using class intervals -20-0, 0-20, etc.



6. For the regression of House Price (Y, \$) versus Area (X, sq. ft) the following results were obtained: $r = 0.796$, $\bar{x} = 1507$, $\bar{y} = 78,802$, $s_x = 509$, $s_y = 35,020$



- Compute the slope of the least squares (regression) line
- Compute the intercept and write down the regression equation:

Price = _____ + _____ * Area

- Predict the price of a house with Area = 1000 ft.
- What influence does the outlier (house with Area = 3600) have on the regression line?
- Compute the residual for this outlier.

7. Multiple choice.

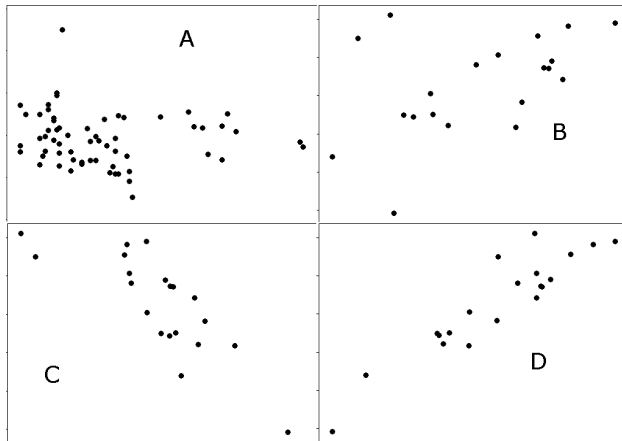
- Correlation between $X = \text{age (months)}$ and $Y = \text{height (inches)}$ was computed. If we change inches to cm (1 inch \approx 2.5 cm), how will the correlation coefficient change?
 - increase
 - decrease
 - stay the same
 - Cannot be determined from information given
- For the above situation, how will the slope b of the regression line change?
 - increase
 - decrease
 - stay the same
 - Cannot be determined from information given
- Which of these measures is resistant to outliers?
 - Range
 - IQR (interquartile range)
 - standard deviation
 - variance
 - none of the above

8. Quick ones.

- A study had shown strong correlation between number of movie theaters and number of fires in US cities. What did the study miss?
- What is a major difference between observational study and experiment?

c) match the correlation numbers to plots

- i. -0.8 ii. -0.05 iii. 0.48 iv. 0.92



d) In order to test how accurate the fiber labels on "Whole grain" products are, we randomly collected 25 such products and measured the fiber content in each.

- i. identify the population
- ii. identify the sample
- iii. identify a possible variable of interest. Is it categorical or numerical?

e) In order to investigate how the price of a law book depends on its length, a group of students searched `amazon.com` and picked the first 30 books in the list. Discuss the quality of their sample.

