1. (20 pts – 1 pt each)

a. Match the symbols/notation with the definitions on the right. Put the letter of the definition in the space next to the symbol on the left.

_____ $\Pr(X \in A) = ?$

_____ μ

_____ r

_____ S_{XY}

_____ $f_X(x)$

_____ $\cup$

_____ $\bar{X}$

_____ σ_X^2

_____ ρ

_____ S_X

_____ $F_X(x)$

_____ $\cap$

a. set intersection; "AND"

b. set union; "OR"

c. population mean; expected value; 1st moment

d. sample mean

e. population variance, 2nd moment

f. sample standard deviation

g. population correlation coefficient

h. sample correlation coefficient

i. general probability statement

j. sample covariance

k. p.d.f. of X (probability distribution function)

l. c.d.f. of X (cumulative pdf)

b. Match the symbol/notation with its defining equation on the right.

_____ $E(X)$

_____ r

_____ $\bar{X}$

_____ S_{XY}

_____ σ_X^2

_____ $\hat{X}$

_____ S_X^2

_____ $F_X(x)$

a. $\Sigma x_i p_i ; \int_{-\infty}^{\infty} x f(x) dx$

b. $\dfrac{1}{n-1} \cdot \dfrac{1}{S_X S_Y} \Sigma(X - \bar{X})(Y - \bar{Y})$

c. $\dfrac{1}{n} \sum_{i=1}^{n} X_i$

d. $\dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$

e. $E(X - \mu_X)^2$

f. $(\prod_{i=1}^{n} X_i)^{1/n}$

g. $\dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

h. $\Pr(X \le x)$

1.   A study of time spent on housework found that men who are employed spend an average of 8.2 hours per week doing housework. Assume that the amount of time spent on housework per week by all employed men in the United States is normally distributed with a mean of 8.2 hours and a standard deviation of 2.1 hours.

a. What percent of men spend more than 9 hours per week on housework?

b. The 10% of employed men who spend the most time on housework spend more than how many hours per week on housework?

2.   A study was conducted to assess the relationship between number of absences and final grade in a statistics class. The regression equation was found to be $\hat{y} = 102 - 3.6x$. Number of absences explained 89% of the variation in final grade.

a. What is the response variable, number of absences or final grade?

b. What is the numerical value of the correlation between number of absences and final grade?

c. If the standard deviation S_y of final grades is 1.1, then what is S_x?

3.   Age at diagnosis for each of 20 patients under treatment for meningitis are given below

17 18 18 18 18 18 19 19 20 20 22 22 23 23 25 25 25 25 26 39

The following descriptive statistics were obtained: $\overline{X} = 22$, $s = 4.99$

a. Calculate the median and the IQR for the data

b. Are there any outliers in the data set? Explain.

c. Construct a boxplot for the data. **Write a one sentence interpretation of your graph.**

15   20   25   30   35   40

4.   A population of taxpayers is divided into five income levels and a simple random sample is selected from each one for an audit. This is an example of a _____ sample.

   a. Stratified
   b. Systematic
   c. Biased
   d. Simple random

5.   It is usual in finance to describe the returns from investing in a single stock by regressing the stock's returns on the returns from the stock market as a whole. The monthly percent total return (y) on Philip Morris common stock and the monthly return (x) on the Standard and Poor's 500-stock index were analyzed for the period between July 1990 and May 1997. Here are the results:

$\bar{x} = 1.304$ $s_x = 3.392$ $r = 0.525$

$\bar{y} = 1.878$ $s_y = 7.554$

   a. Calculate the equation of the regression line from this information.

   b. What percent of the variation in Philip Morris stock is explained by the linear relationship with the market as a whole?

   c. Explain what the slope of the line tells us about how Philip Morris stock responds to changes in the market.

   d. Suppose this model cost you $1,000, and there were another model which cost $5,000 but which had an r^2 of 65%. How might you decide if the more expensive prediction model be worthwhile?

6. A lurking variable is

a. The true cause of any response.
b. Any variable that produces a large residual.
c. A variable that is not among the variables studied but that affects the response variable.
d. The true variable that is explained by the explanatory variable.

7. A group of college students believe that herb tea has remarkable restorative powers. To test this belief, they make weekly visits to a local nursing home, visiting with the residents and serving them herb tea. The nursing staff reports that after several months many of the residents are more cheerful and healthy. The students conclude that they were correct, herb tea has restorative powers. Do you agree? Why or why not?

8. A card is to be selected from an ordinary deck of 52 cards. Suppose that a casino will pay $10 if you select an ace. If you fail to select an ace, you are required to pay the casino $1. Construct a probability distribution for **the amount of money the casino wins**.

9. An electronics store sells a particular brand of computer notebook. Let X be the number of computer notebooks sold in a day. The probability distribution for X is below

X	0	1	2	3	4
Probability	0.1	0.45	0.25	?	0.15

a. What is the probability that three computer notebooks are sold in a day?

b. Find the probability that no more than two computer notebooks are sold in a day

c. What is the probability that **at least three** computer notebooks are sold in a day?

d. Find the mean for the above distribution.

10. Event A has probability 0.3. Event B has probability 0.5. Find P(A or B) if:

    a. A and B are disjoint.


    b. A and B are independent.

11. Suppose that for a certain Caribbean Island in any 3-year period the probability of a major hurricane is 0.25 and the probability of an earthquake is 0.14. If the two events are independent, what is the probability of both an earthquake and a major hurricane in any 3-year period?

    a. 0
    b. 0.39
    c. 0.035
    d. 0.355
    e. The probability cannot be calculated from the information given

12.   After keeping track of his heating expenses for several winters, a homeowner believes he can predict the monthly cost from the average daily Fahrenheit temperature using the following equation:
$$\hat{y} = 133 - 2.13x$$

The residual plot for the data is shown below:

Residuals Versus avg temp
(response is cost)
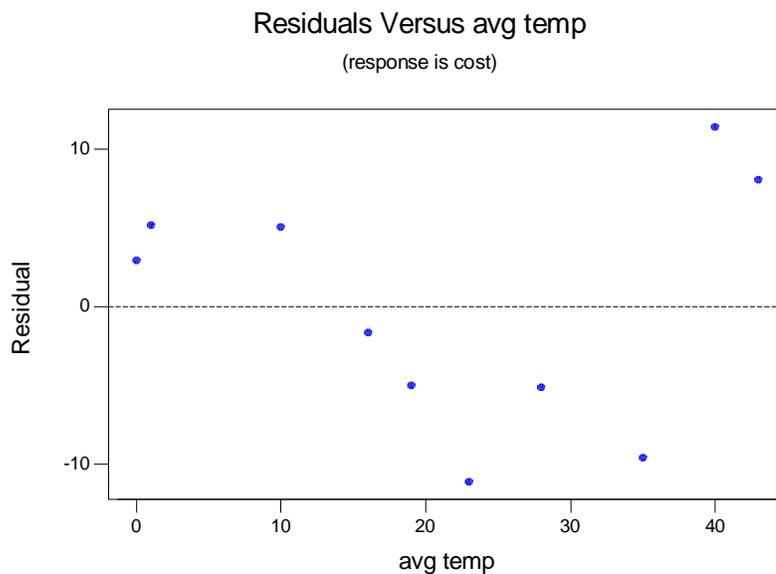
During months where the temperature stays around forty, would you expect your cost predictions based on this model to be accurate, too low, or too high? Use the residual plot to justify your answer.

Use the residual plot to describe how well the regression line describes the data.

13.   A high school Latin teacher wished to demonstrate the favorable effect of studying Latin on mastery of English. For all seniors she obtained scores on a standard English-proficiency examination. In that school, the average SAT verbal score for students studying Latin is 532 and for those not studying Latin it is only 489. The Latin teacher concluded that "the study of Latin greatly improves one's command of English"

a. Is this an observational study or a randomized experiment? Explain.

b. What are the explanatory and response variables?

c. List a possible confounding variable, and explain how it would impact the study.

d. Was the teacher correct in her conclusion? Why or why not?

14.   Find the (sample) correlation coefficient for the data set below:

x	y
1	3
1	5
2	3
3	5
3	4

a. (1 point) What is the name for this very large number (you must spell this correctly): 10^{100}

b. (3 points) Suppose you have a cube, *X* miles on a side, which you fill with 2-inch golf balls. How large (i.e., what dimensions in miles) would the cube have to be to hold 10^{99} golf balls?
HINT: 1 mile = 5,280'

c. (1 point) How many of these cubes in part b would we have to have to have 10^{100} golf balls?