

# Stat 1411 exam 1

Please answer all questions carefully!

1. (4pt) Jason is buying a new computer. He is comparing various models using many different variables. These variables play an important role in making the decision about which computer he will buy. Identify each of the following variables as categorical or quantitative.

- A) Does the computer come with a CD/DVD writer? *categorical*
- B) What is the memory capacity of the computer? *quantitative*
- C) Is the computer a laptop or a desktop model? *categorical*
- D) How much does the computer cost? *quantitative*

2. (7pt) In a statistics class with 136 students, the professor records how much money each student has in their possession during the first class of the semester. The histogram shown below represents the data he collected.



(5pts) A) Give a description of the shape of the histogram. (i.e, symmetric/skewed, mode, outliers)

*skewed right  
unimodal, \$0-\$10  
outlier \$100-\$110*

(2pts) B) Approximately what is the number of students with \$30.00 or more in their possession?

*10*

3. (6pts) During the early part of the 1994 baseball season, many sports fans and baseball players noticed that the number of home runs being hit seemed to be unusually large. Below are separate stemplots for the number of home runs by American League and National League teams based on the team-by-team statistics on home runs hit through Friday, June 3, 1994 (from the *Columbus Dispatch* sports section, Sunday, June 5, 1994).

American League

2	
3	5
4	0 3 9
5	1 4 7   8 8
6	4 8 8
7	5 7

$$M = \frac{57 + 58}{2} = 57.5$$

National League

2	9
3	1
4	2 6 7 8 8
5	3 5 5 5
6	3 3 7
7	

$$M = \frac{48 + 53}{2} = \frac{101}{2} = 50.5$$

A. What is the median for the number of home runs for the American League teams?

57.5

B. Determine whether each of the following statements is true or false.

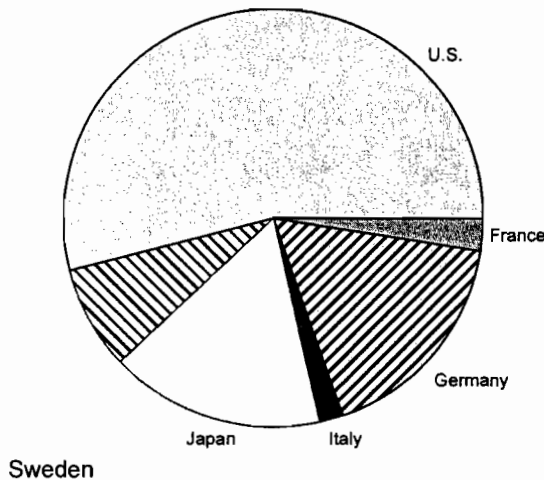
i) The American League plot is reasonably symmetric. *True*

ii) The National League plot is bimodal. *False*

iii) The median number of home runs hit by National League teams for this time period was higher than the median for the American League teams. *False*

iv) The lowest number of home runs hit by *any* team for this time period is 29. *True*

4. (6pts) The Consumers Union measured the gas mileage in miles per gallon of 38 automobiles from the 1978–79 model-year on a special test track. The pie chart below provides information about the country of manufacture of the automobiles used in this study by the Consumers Union.



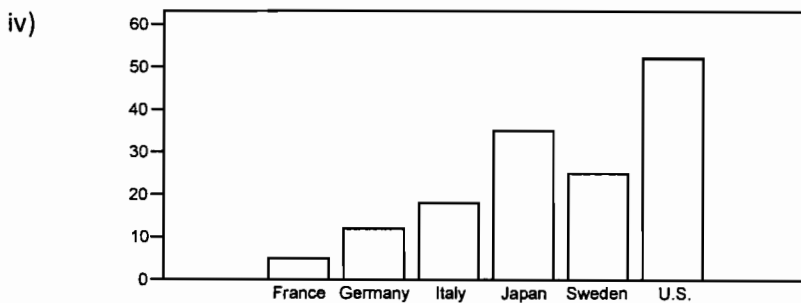
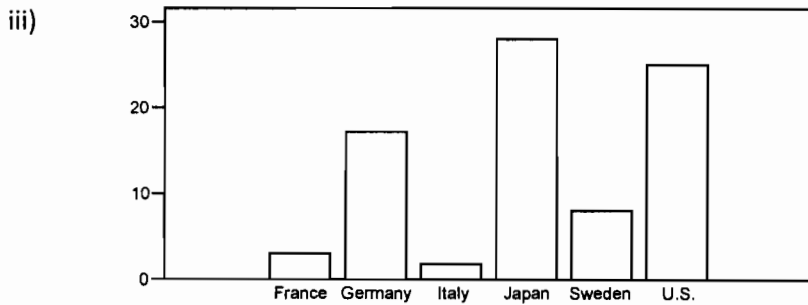
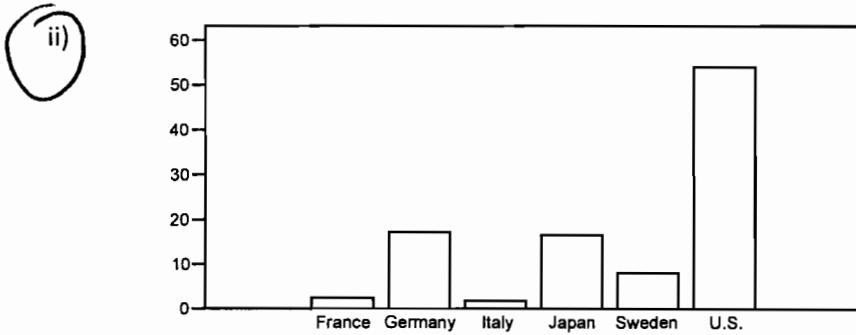
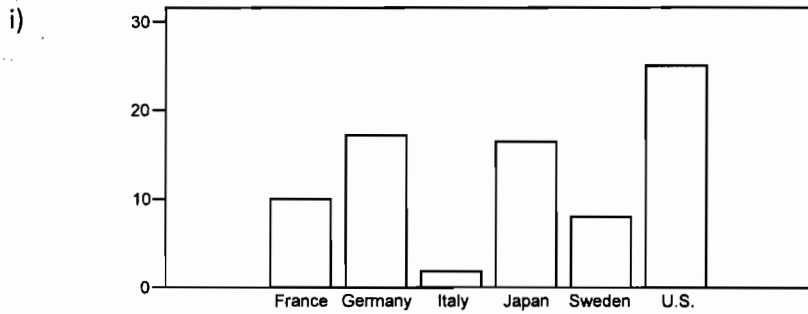
A. What conclusion can we draw based on this pie chart?

i) Italian cars get significantly lower gas mileage than cars of other countries. This is because their slice of the pie is the smallest of the chart.

ii) More than half of the cars in the study were from the United States.

iii) Swedish cars get gas mileages that are between those of Japanese and U.S. cars.

B. Which of the following bar graphs represents the same data as the pie chart in the previous problem?



5. (4pt) The average salary of all female workers is \$35,000. The average salary of all male workers is \$41,000. What must be true about the average salary of all workers?

The avg salary is b/w \$35,000 and \$41,000

6. (5pt) As part of a large ongoing study on the treatment of women with breast cancer, the treatment time (in months) of eight patients whose lymph nodes were cancer free is recorded. Treatment time is defined as the time from the moment the cancer treatment starts until the patient is declared cancer free. Give the 5 number summary.

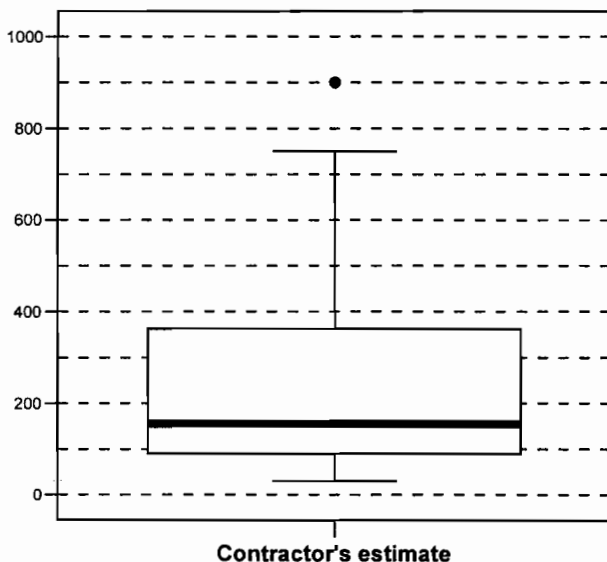
9.47 8.60 19.33 16.33 8.50 9.40 17.67 9.30

8.50 8.60 | 9.30 9.40 | 9.47 16.33 | 17.67 19.33  
 8.95 9.435 17  
 Q1 M Q3

5 # summary

8.50 8.95 9.435 17 19.33

7. (5pts) The Michigan Department of Transportation (M-DOT) is working on a major project: 80% of the highways in Michigan need to be repaved. To speed completion of this project, many contractors will be working for M-DOT. Contractors are currently bidding on the next part of the project. To help make a decision about which contractor to hire, M-DOT collects many variables besides just the estimated cost. One of those variables is the contractor's estimate of the number of work days required to finish the job. Twenty contractors have bid on the next job. The boxplot below represents their estimates of the number of work days required.



A. What is (approximately) the interquartile range, based on the boxplot?

i) 140 days

ii) 270 days

iii) 360 days

iv) 760 days

B. Determine whether each of the following statements is true or false.

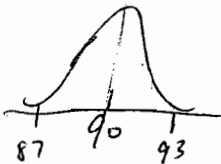
i) The median number of days is approximately 180. *True*

ii) The minimum number of days is approximately 40. *False*

iii) The maximum number of days is approximately 750. *True*

iv) Twenty-five percent of contractors estimated the number of days to be more than 100. *False*

8.(5pt) Items produced by a manufacturing process are supposed to weigh 90 grams. However, there is variability in the items produced and they do not all weigh exactly 90 grams. The distribution of weights can be approximated by a normal distribution with mean 90 grams and a standard deviation of 1 gram. What percentage of the items will either weigh less than 87 grams or more than 93 grams?



Can use 68-95-99.7 Rule  
87 is 3 std dev below 90  
93 is 3 std dev above 90  
99.7% b/w 87 and 93 so

**0.3%** below 87 or above 93 grams

9. (5pts) The temperature at any random location in a kiln used for manufacturing bricks is normally distributed with a mean of 1000°F and a standard deviation of 50°F. If bricks are fired at a temperature above 1125°F, they will crack and must be discarded. If the bricks are placed randomly throughout the kiln, what is the percentage of bricks that crack during the firing process?  $X = \text{firing temp}$

$$X > 1125 \quad z = \frac{1125 - 1000}{50} = 2.5$$

$$z > 2.5 = 1 - \text{amt } z < 2.5 = 1 - 0.9938 = 0.0062 \times 100\% = \boxed{0.62\%}$$

10. (5pt) The scores on a university examination are normally distributed with a mean of 62 and a standard deviation of 11. If the bottom 5% of students will fail the course, what is the lowest mark that a student can have and still be awarded a passing grade?

$z < z$  is 0.05 from table  $z = -1.645$

$$11(-1.645) = \frac{(x - 62)}{11} \cdot 11$$

$$-18.095 = x - 62$$

$$+62 \quad +62$$

$$x = 43.905 \approx \boxed{44}$$

11. (5pts) Chocolate bars produced by a certain machine are labeled with 8.0 oz. The distribution of the actual weights of these chocolate bars is normal with a mean of 8.1 oz. and a standard deviation of 0.1 oz. What proportion of chocolate bars weighs between 8.2 and 8.3 oz?  $x = \text{choc bar weight}$

$$8.2 < x < 8.3$$

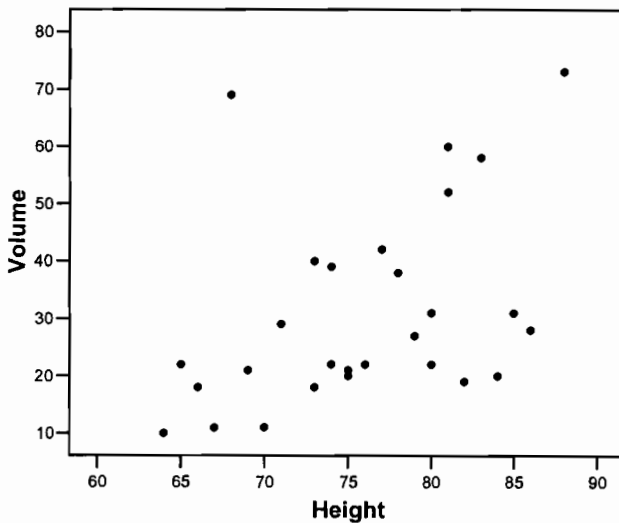
$$x = 8.2 \quad z = \frac{8.2 - 8.1}{0.1} = 1$$

$$x = 8.3, \quad z = \frac{8.3 - 8.1}{0.1} = 2$$

$$1 < z < 2$$

$$p \sim p \quad z < 2 - p \sim p \quad z < 1 = 0.9772 - 0.8413 = \boxed{0.1359}$$

12. (6pts) The height (in feet) and volume of usable lumber (in cubic feet) of 32 cherry trees are measured by a researcher. The goal is to determine if the volume of usable lumber can be estimated from the height of a tree. The results are plotted below.



- A. What is the response variable, and what is the explanatory variable in this study?

response: volume

explanatory: height

- B. Describe the scatterplot.

moderate linear w/ positive association

one obvious outlier

13. (5pts) A college newspaper interviews a psychologist about a proposed system for rating the teaching ability of faculty members. The psychologist says, "The evidence indicates that the correlation between a faculty member's research productivity and teaching rating is close to zero." What would be a correct interpretation of this statement?

- A. Good researchers tend to be poor teachers and vice versa.
- B. Good teachers tend to be poor researchers and vice versa.

- C. Good researchers are just as likely to be good teachers as they are bad teachers. Likewise for poor researchers.
- D. Good research and good teaching go hand in hand.

14. (8pts) A student wonders if people of similar heights tend to date each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man whom each woman dates. Here are the data (heights in inches):

Women	64	65	65	66	66	70
Men	68	68	69	70	72	74

Determine whether each of the following statements is true or false.

- A. If we had measured the heights of the men and women in centimeters (1 inch  $\approx$  2.5 cm), the correlation coefficient would have been 2.5 times larger. *False*
- B. There is a strong negative association between the heights of men and women because the women are always smaller than the men they date. *False*
- C. There is a positive association between the heights of men and women. *True*
- D. Any height above 70 inches must be considered an outlier. *False*

15.(8pts)

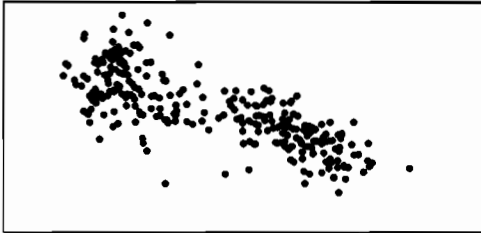
Determine whether each of the following statements regarding the correlation coefficient is true or false.

- A) The correlation coefficient equals the proportion of times that two variables lie on a straight line. *False*
- B) The correlation coefficient will be +1.0 if all the data points lie on a perfectly horizontal straight line. *False*
- C) The correlation coefficient measures the strength of any relationship that may be present between two variables. *False*
- D) The correlation coefficient is a unitless number and must always lie between -1.0 and +1.0, inclusive. *True*

16.(8pts)

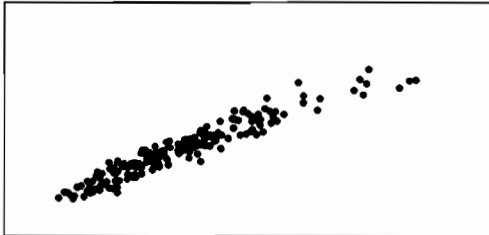
Match the four graphs labeled A, B, C, and D, with the following four possible values of the correlation coefficient: -0.9, -0.7, 0.4, 0.95. Assume all four graphs are made on the same scale.

A)



$$r = -0.7$$

B)



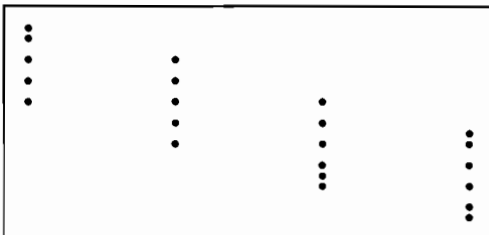
$$r = 0.95$$

C)



$$r = 0.4$$

D)



$$r = -0.9$$

17. (8pts) Given the following table, find the correlation.

$x$	$y$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	
2	5	-1.6	-1.2	2.56	1.44	3.6864	1.92
3	6	-0.6	-0.2	0.36	0.04	0.0144	0.12
4	5	0.4	-1.2	0.16	1.44	0.2304	-0.48
4	8	0.4	1.8	0.16	3.24	0.5184	0.72
5	7	1.4	0.8	1.96	0.64	1.2544	1.12
ms: 18	31	0	0	5.2	6.8	5.704	3.4

$$\bar{x} = \frac{18}{5} = 3.6 \quad \bar{y} = \frac{31}{5} = 6.2$$

$$S_x^2 = \frac{5.2}{4} = 1.3, \quad S_x = \sqrt{1.3} = 1.1402$$

$$S_y^2 = \frac{6.8}{4} = 1.7, \quad S_y = \sqrt{1.7} = 1.3038$$

$$r = \frac{1}{4} \cdot \frac{1}{(1.1402)(1.3038)} (5.704) = \boxed{0.9592}$$

mistake in  $(x_i - \bar{x})(y_i - \bar{y})$   
if recalculated values,  
will have

$$r = \frac{1}{4} \frac{1}{(1.1402)(1.3038)} (3.4) = \boxed{0.5718}$$