

Exam 1 - STAT 303 Session 201
Summer 2009

Name:

UIN:

Signature:

1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You may use one one-sided sheet of formulas that you have brought with you. You should have no other printed or written material with you on the exam.
4. You have 60 minutes to work on this exam. There are 16 multiple choice questions and 4 written questions, each worth 5 points.
5. You may use a calculator but not a phone during the exam.
6. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
7. Do not sit directly next to another student.
8. Good Luck!!!

1. Find the mean, median, and standard deviation. (Use three decimal place.)

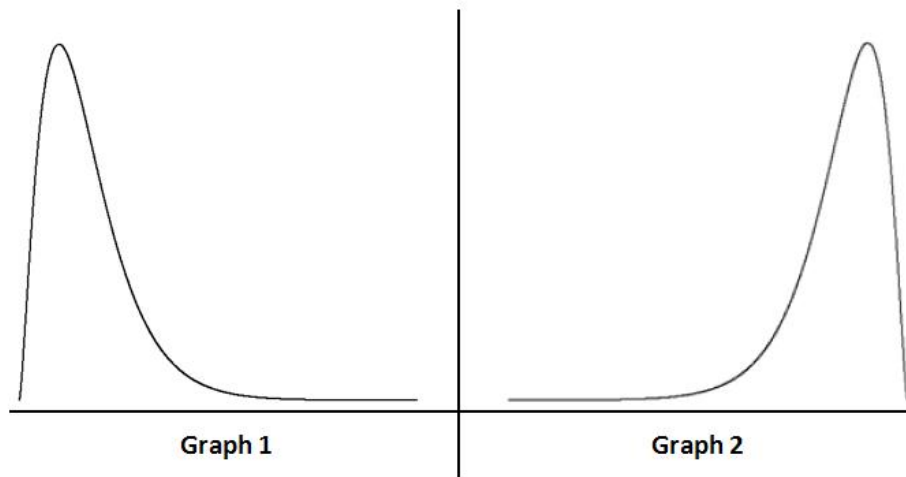
10, 9, 8, 6, 12, 13, 10, 11, 7, 14, 15

- (a) Mean: 10.455, Median: 10, Standard Deviation: 2.876
- (b) Mean: 10.455, Median: 10, Standard Deviation: 8.273
- (c) Mean: 10, Median: 10, Standard Deviation: 2.876
- (d) Mean: 10, Median: 10, Standard Deviation: 8.273

2. Select the answer which is NOT TRUE.

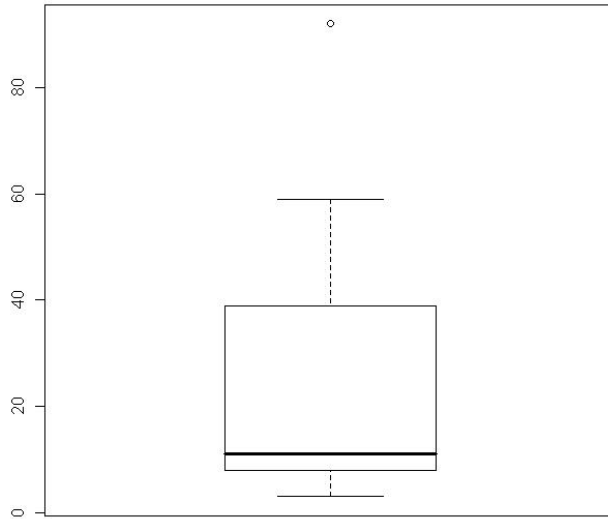
- (a) Histogram is the graphical versions of frequency table
- (b) *Box-plot only shows five number summaries: Min, Q_1 , Mean, Q_3 , Max.*
- (c) Stem-and-leaf plot works best for smaller data sets.
- (d) Normal quantile plot uses for comparing the distribution of the sample to the normal distribution.

3. Match Graph 1 and Graph 2 to right explanation.



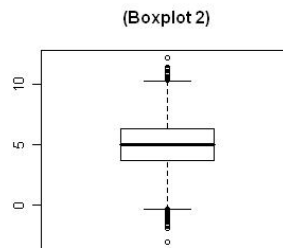
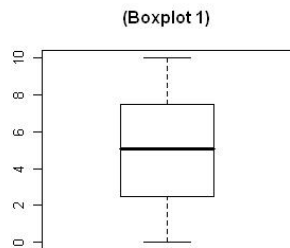
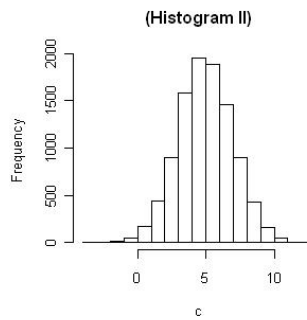
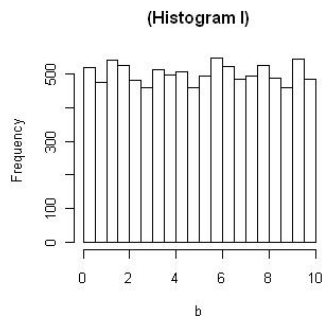
- (a) Graph 1 - Right Skewed, Mean < Median; Graph 2 - Left Skewed, Mean > Median
- (b) Graph 1 - Left Skewed, Mean < Median; Graph 2 - Right Skewed, Mean > Median
- (c) *Graph 1 - Right Skewed, Mean > Median; Graph 2 - Left Skewed, Mean < Median*
- (d) Graph 1 - Left Skewed, Mean > Median; Graph 2 - Right Skewed, Mean < Median

4. Select proper explanation for following boxplot.



- (a) The shape of distribution is left skewed.
- (b) The maximum is near 60 and the minimum is near 0.
- (c) Inter-quantile Range (IQR) is approximately 40.
- (d) The criterion for outlier is $Q_1 - IQR$ and $Q_3 + IQR$
- (e) *None of the above.*

5. Work with a group to match the histograms to the appropriate boxplots and appropriate summary statistics.



	Mean	Median	St. Dev.	Q_1	Q_3
Summary 1	4.992	5.033	2.890	2.471	7.492
Summary 2	4.963	4.945	1.980	3.636	6.306

- (a) Histogram I - Boxplot 2 - Summary 1
- (b) Histogram II - Boxplot 2 - Summary 1
- (c) *Histogram I - Boxplot 1 - Summary 1*
- (d) Histogram II - Boxplot 1 - Summary 2

[Question 6 - 7] Raw scores on behavioral tests are often transformed for easier comparison. A test of reading ability has mean 58 and standard deviation 15 when given to third graders. Fourth graders have mean score 66 and standard deviation 10 on the same test. Fifth graders have mean score 71 and standard deviation 5 on the same test. To provide separate “norms” for each grade, we want scores in each grade to have mean 50 and standard deviation 15. David, a third-grade student, scores 73 on the test. Nancy, a fourth-grade student, scores 76 on the test. And Peter, a fifth-grade student, scores 81 on the test.

6. Using the transformed scores, who scores higher within his or her grade?
 - (a) David
 - (b) Nancy
 - (c) *Peter*
 - (d) Their scores are the same.

7. What is the Davis’s transformed score?
 - (a) 50
 - (b) 55
 - (c) 60
 - (d) *65*
 - (e) 70

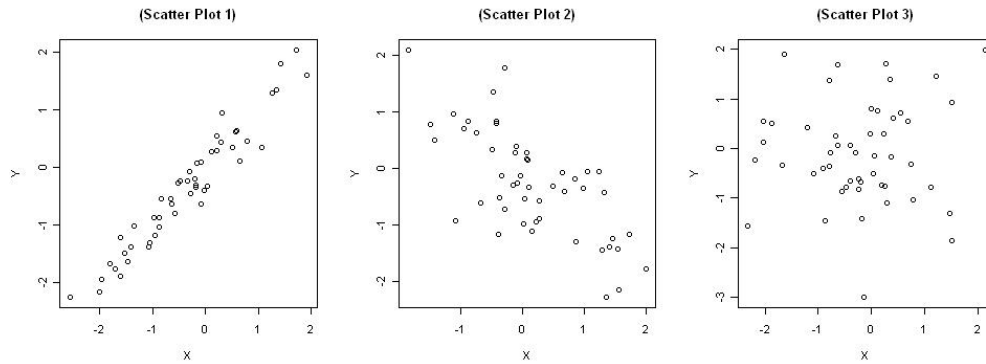
8. Select the answer which is NOT TRUE?
 - (a) *When we find some outliers in our sample, we should delete these outliers*
 - (b) An observation is *influential* for a statistical calculation if removing it would markedly change the result of the calculation.
 - (c) Sometimes outliers can reveal some important features of the data, not be just mistakes!

- (d) Outliers have an influence on mean and standard deviation.
9. If the grades on an exam are normally distributed with a mean of 70 and a variance of 9, what grade do you approximately have to make to be in the top 15% and 2.5% of the class?
- (a) Top 15%: 79, Top 2.5%: 88
 (b) *Top 15%: 73, Top 2.5%: 76*
 (c) Top 15%: 76, Top 2.5%: 82
 (d) Top 15%: 76, Top 2.5%: 79
 (e) We cannot calculate the points using above information.
10. Select the answer which is NOT TRUE?
- (a) Correlation coefficients measures the direction and strength of the linear relationship between two numeric variables.
 (b) *If r is 0, there is no relationship between two variables.*
 (c) If explanatory and response are switched, r remains the same.
 (d) Scale changes do not affect r .
 (e) Correlation coefficients must be between -1 and 1, or $(-1 \leq r \leq 1)$.
11. Suppose we measure the weight of everyone on a football team and obtain the statistics for a team report. Then, suppose we found out the scale was 10 lbs. under so we need to add 20 lbs. to every weight. Further, suppose we found out that we are supposed to report the weights and statistics in kilograms, not lbs (Remember, 1 lb = 0.45 kilograms). Choose the WRONG linear transformation for each statistics.

Mean	IQR	Variance	Max
200 lbs.	70 lbs.	1600 lbs.	400 lbs.

- (a) Mean = $(200 + 20) \times 0.45$
 (b) *IQR = $(70 + 20) \times 0.45$*
 (c) Variance = 1600×0.45^2
 (d) Max = $(400 + 20) \times 0.45$
 (e) Every above linear transformation is proper.

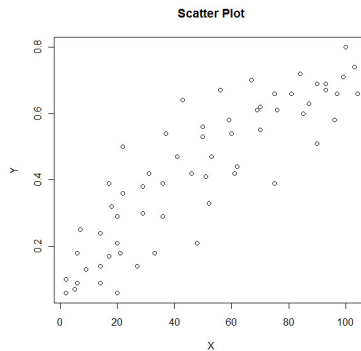
12. Match the scatter plot to the proper correlation coefficients. **(Write your answer)**



$$r = -.70, \quad r = .05, \quad r = .97$$

- (a) Scatter Plot 1: $r = .97$
- (b) Scatter Plot 2: $r = -.70$
- (c) Scatter Plot 3: $r = .05$

13. Using the linear regression result which displays below, select the statement which is NOT TRUE.



$$Y = 0.136 + 0.006 \times X \quad R^2 = 0.7588$$

- (a) The explanatory variable x and the response variable y are positively related.
- (b) When $X = 0$, the Y will be 0.136.
- (c) The correlation coefficient between X and Y is 0.871
- (d) Using linear regression, you can explain 75.88% of variation in the data.
- (e) *When X increases by 1, Y will be increased by 0.136*

[Question 14 - 15] Write down your answer with following frequency table.

x	Frequency	Rel. Freq.	Cum. Rel. Freq.
0	5	.10	.10
1	9	.18	(d)
2	7	(b)	.42
3	4	.08	.50
4	10	.20	.70
5	(a)	(c)	(e)
6	4	.08	.94
7	3	.06	1.00
Total	50	1.00	1.00

14. Write down (a), (b), (c), (d), and (e). **(Write your answer)**

$$a = 8, b = .14, c = .16, d = .28, e = .86$$

15. Write down Min, Q_1 , Median, Q_3 , Max. **(Write your answer)**

$$\text{Min} = 0, Q_1 = 1, \text{Median} = 3.5, Q_3 = 5, \text{Max} = 7$$

[Question 16 - 17] Write down your answer with following two-way table.

Status	Alaska Airlines	America West
On time	718	5534
Delayed	74	532
Total	792	6066

16. Calculate the following conditional probability.

(a) How likely are the delayed airplanes is Alaska airline if airplane is delayed.

$$\frac{74}{74+532} = .122$$

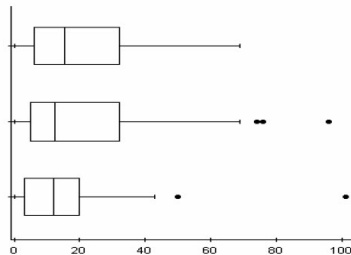
- (b) Calculate the probability of delayed case if your airline is American West.

$$\frac{532}{5534+532} = .088$$

17. However, when we divide above table by the departure city, the interpretation will be different. Select the statement which is NOT TRUE.

	Los Angeles			Phoenix		
Airlines	On time	Delayed	Total	On time	Delayed	Total
Alaska Airlines	497	62	559	221	12	233
America West	694	117	811	4840	415	5255

- (a) This association is called *Simpson's Paradox*.
 (b) *In actual, America West would be preferable.*
 (c) *The conclusion of two tables are contradictory.*
18. Which of the following is/are true for the data represented above? There are 50 points in each.



- (a) Although the range of the bottom boxplot is larger than the middle one, the middle one has a larger standard deviation. (Remember the definition of standard deviation).
 (b) The middle boxplot has the largest mean.
 (c) Only the top boxplot has no outliers.
 (d) *All of the above are true.*
 (e) We need summary numbers to determine any of the above.
19. Suppose we weigh 50 people. 49 of them are between 180 and 280 lbs. One weighs only 150 lbs. Which of the following is true for this sample?
- (a) *The sample mean, \bar{X} , is NOT the best measure of center for this data.*
 (b) The sample mean, \bar{X} , will be larger than the sample median.

- (c) The data cannot be normally distributed since there is an outlier.
- (d) All of the above are true for this data.
- (e) Only two of the above are true for this data (excluding D).

20. Select the statements which are TRUE.

- (a) We can use of a regression line for prediction far outside the range of values of the explanatory variable x that you used to obtain the line.
- (b) The further an (x, y) data point is from the mean, the less it affects the sample statistics.
- (c) Influential points should be deleted since they change the linear relationship between x and y .
- (d) *The size of a sample, n , affects the spread of the sampling distribution of the statistic.*
- (e) None of the above are TRUE.

21. Exam Summary

$Q_1 = 57.25$, Median = 67.5, $Q_3 = 72.5$, Max = 92.5, Mean = 64.5