

STAT303 Secs 508–510

Fall 2000

Exam #1

Form A

Instructor: Julie Hagen Carroll

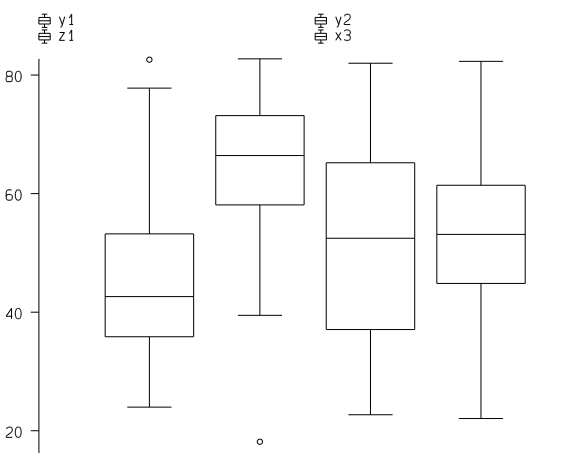
October 25, 2000

1. **Don't EVEN open this until you are told to do so.**
2. Be sure to MARK your REGISTERED section number and your test form (A, B, C or D) on the scantron!
3. Sign your name where indicated on your scantron and write your (Thursday) section number and computer number beside it. You may keep your exam.
4. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly** on the scantron. Multiple marks will be counted wrong.
5. You will have 60 minutes to finish this exam.
6. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
7. This exam is worth 100 points, and will constitute 20% of your final grade.
8. Good luck!

1. Which of the following statements is/are true?
 - A. The value of the median will change if you add 10 to the maximum in a dataset.
 - B. The IQR covers 50% of the range (spread) of the data.
 - C. A z-score tells us how many standard deviations an observation is from its mean.
 - D. All of the above are true statements.
 - E. Exactly two of the above are true statements (excluding D.).

2. Which of the following is NOT affected by outliers?
 - A. the mean, \bar{x}
 - B. the z-scores
 - C. the correlation coefficient
 - D. Exactly two of the above are not affected.
 - E. All of the ARE affected by outliers.

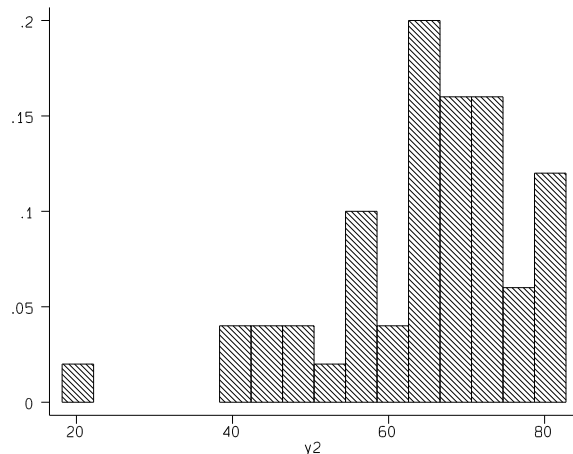
3. Suppose you know that a distribution is bell-shaped, with mean, $\bar{x} = 30$ and standard deviation, $s_x = 4$. Approximately what percent of the data will fall between 22 and 34?
 - A. about 95%
 - B. between 50 and 68%
 - C. about 80%
 - D. between 95 and 100%
 - E. about 75%



4. Numbering the boxplots 1 through 4 from left to right, which distribution is approximately uniform?
 - A. You can't tell by looking at a boxplot.
 - B. 3 only since it's the most spread out.
 - C. 4 only since it's the most symmetric.
 - D. Both 3 and 4 are uniform.
 - E. None of the boxplots are uniform.

5. Which boxplot has an Interquartile Range, IQR, of about 20?
 - A. 1
 - B. 2
 - C. 3
 - D. 4
 - E. 1,2 and 4 are all the same.

6. When can we correctly apply the Empirical Rule?
 - A. The Empirical Rule can be applied to any of the datasets.
 - B. Only if the data is bell-shaped.
 - C. Only if the data is symmetric.
 - D. All of the above are true.
 - E. Exactly two of the above are true.



7. Which of the following are 'better' statistics for the distribution above?
 - A. The median and the range since the data is skewed.
 - B. The median and the IQR since the data is skewed.
 - C. The mean and the standard deviation since the data is skewed.
 - D. The mean and the IQR since the data is skewed.
 - E. The median and standard deviation are always the best statistics to use.

8. Suppose you know that a distribution is bell-shaped with mean, $\bar{x} = 25$ and standard deviation, $s_x = 2$, what would be the approximate 98th percentile?

- A. 0.98
- B. 98% of 25 = 24.5
- C. about 27
- D. about 29
- E. about 31

9. If we know that a distribution is skewed to the right, which of the following is true?

- A. The mean is smaller than the median.
- B. Outliers would most likely be to the left.
- C. The histogram would be shorter on the left.
- D. All of the above are true.
- E. None of the above are true.

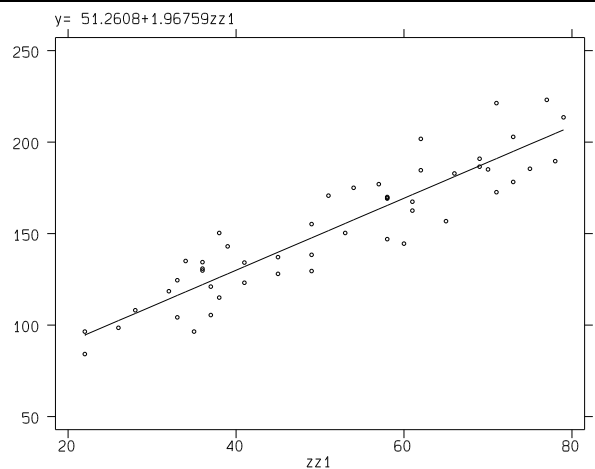
Percentiles	Smallest	Largest
1%	22	22
5%	26	22
10%	32.5	26
25%	37	28
50%	52	
75%	65	75
90%	73	77
95%	77	78
99%	79	79

10. How many outliers are there in the output above?

- A. at least 4
- B. 3
- C. 2
- D. 1
- E. none

11. Which best describes the distribution in the problem above?

- A. normal
- B. uniform
- C. skewed left
- D. skewed right
- E. symmetric



12. Which of the following is the best approximation of the linear equation for this data?

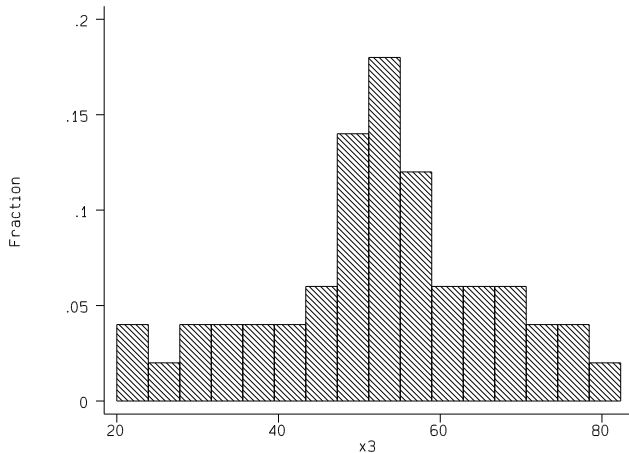
- A. $y = 90 + 2 * x$
- B. $y = 50 + 2 * x$
- C. $y = 90 + 0.5 * x$
- D. $y = 50 + 0.5 * x$
- E. $y = 150 + 50 * x$

13. If the point (100,60) was added to the previous scatterplot, which of the following would be true?

- A. It would be an outlier, so the correlation would decrease, but the slope would be about the same.
- B. It would be an outlier, so the correlation would increase, but the slope would be about the same.
- C. It would be an influential point, so the slope and intercept would change.
- D. It would be an outlier, so the correlation and the slope would increase.
- E. It would be an influential point, so the correlation would change, but the slope would be about the same.

14. If we know the data is bell-shaped, which of the following would be true?

- A. The z-score for the median is 0.
- B. The z-score for the minimum is between -3 and -4.
- C. The length of the IQR is less than 2.
- D. All of the above are true.
- E. Exactly two of the above are true.



15. Which of the following best describes the distribution above?

- A. The mean is approximately 50 and the standard deviation is about 10.
- B. The mean is approximately 50 and the standard deviation is about 20.
- C. The median is approximately 50 and the interquartile range is about 10.
- D. Exactly two of the above are true.
- E. You cannot tell the mean and standard deviation from a histogram.

Driver	Gender		Total
	F	M	
aggress	37	31	68
passve	14	16	30
timid	1	1	2
Total	52	48	100

Pearson chi2(2) = 0.5036 Pr = 0.777 (p-value)

16. Are 'type of Driver'(aggressive, passive or timid) and 'Gender' statistically related?

- A. Yes, they are statistically related because the p-value is large.
- B. No, they are not statistically related because the p-value is large.
- C. Yes, males are more aggressive than females.
- D. No, only 1 male and 1 female are timid drivers.
- E. We would need the correlation coefficient to tell the strength of the relationship.

17. Using the previous table, how likely are you to be an aggressive driver if you are female?

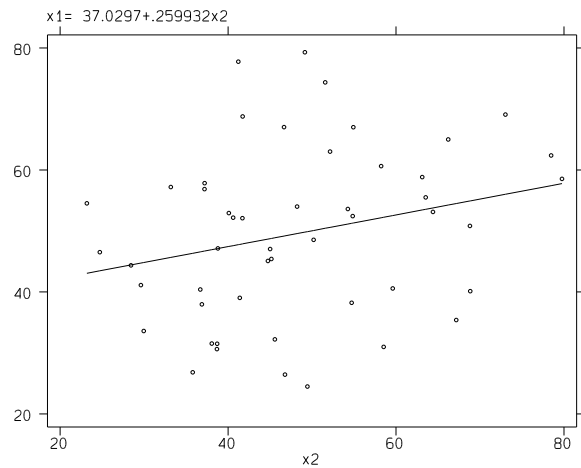
- A. 37/52
- B. 37/68
- C. 68/100
- D. 52/68
- E. 37/100

18. How likely are you to be a passive male driver?

- A. 16/48
- B. 16/30
- C. 30/100
- D. 16/100
- E. 48/100

19. Which of the following is NOT true?

- A. Social Security Number is a discrete numeric variable.
- B. Classification is a an ordinal variable.
- C. Height is a continuous numeric variable.
- D. Number of cars that just ran the red light is a discrete numeric variable.
- E. All of the above are true.



20. How would you describe the correlation of the data above?

- A. Strongly negative
- B. Moderately negative
- C. Weak
- D. Moderately positive
- E. Strongly positive

1C,2E,3C,4B,5A,6B,7B,8D,9E,10E,11B,12B,13C,14D,15A,16B,17A,18D,19A,20C