

STAT303 Secs 509–511

Spring 2001

Exam #1

Form A

Instructor: Julie Hagen Carroll

May 2, 2001

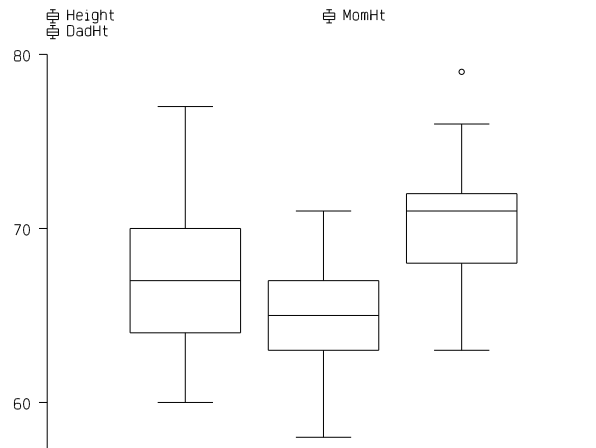
1. **Don't EVEN open this until you are told to do so.**
2. Be sure to MARK your REGISTERED section number and your test form (A, B, C or D) on the scantron!
3. Sign your name where indicated on your scantron and write your (Thursday) section number and computer number beside it. You may keep your exam.
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!

- Which of the following is TRUE?
 - 'Number of roommates' is a discrete numeric variable.
 - 'Test score on this exam' is a continuous numeric variable.
 - 'Number of days in a week' is a discrete numeric variable.
 - 'Telephone number' is a discrete numeric variable.
 - All of the above are true.

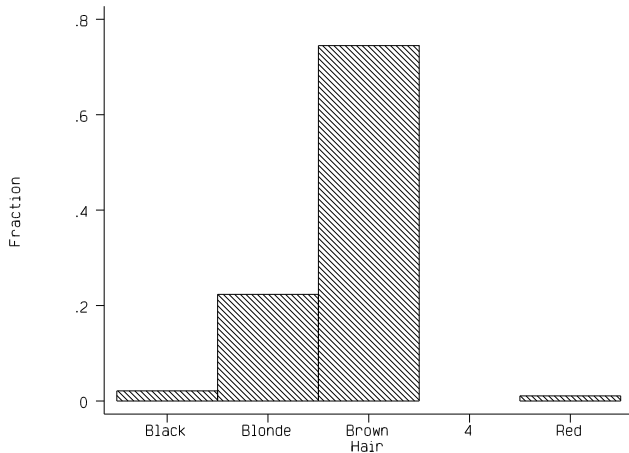
- Which of the following affect the median, \tilde{x} , but not mean, \bar{x} ?
 - the skewedness of the data
 - outliers added to the data
 - scale changes to the data
 - All of the above affect the median but not the mean.
 - None of the above affect the median but not the mean.

x	Freq.	Percent	Cum.
200	3	3.30	3.30
210	1	1.10	4.40
220	6	6.59	10.99
230	6	6.59	17.58
240	9	9.89	27.47
250	10	10.99	38.46
260	15	16.48	54.95
270	17	18.68	73.63
280	23	25.27	98.90
290	1	1.10	100.00
Total	91	100.00	

- Which is the correct list of the 5 Number Summary, if any, for this data?
 - 200,225,245,265,290
 - 200,225,245,265,290
 - 200,230,270,280,290
 - 200,240,260,280,290
 - 200,240,260,270,290



- What can you say about the data represented above? The first is *Height* of a student, the second is *Mom's Height*, and the third is *Dad's Height*.
 - Mom's Height* is the most normal in shape.
 - Mom's are shorter than their children.
 - Mom's are shorter than their husbands (Dad's).
 - All of the above are true.
 - Exactly two of the above are true.
- What is the IQR for *Height*(of a student), still using the boxplot above?
 - 50%
 - 1/2 of the range
 - 10
 - 7
 - 17
- Suppose you know that a distribution is bell-shaped with mean, $\bar{x} = 25$ and standard deviation, $s_x = 5$, what would be the approximate 85th percentile?
 - 29.25
 - 34.25
 - 25.85
 - 30
 - 35



7. What can you say about the shape of the histogram above? 4=Gray (I don't know why it's missing)

- A. It's fairly normal since the peak is in the center.
- B. 'Red' is an outlier.
- C. The data is categorical, so the shape is meaningless.
- D. The data is categorical, but the shape is normal.
- E. Since the data is categorical, this graph is invalid.

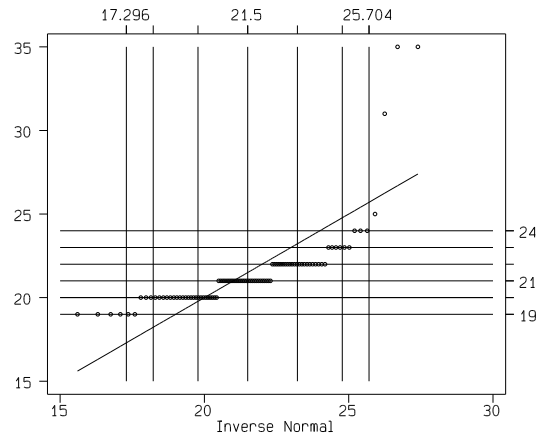
8. Which of the following is the *Fundamental Rule for Using Data for Inference*?

- A. Sample data can never be used to make inferences about a population. (You must look at the whole population.)
- B. Sample data can be used to make inferences about a much larger group if the data can be considered to be representative with regard to the question(s) of interest.
- C. Any data can be used to make inferences about a much larger group if the data came from a random sample.
- D. Sample data can be used to make inferences about any larger group as long as the data came from a random sample.
- E. Data is only useful if it came from a simple random sample.

9. What can we say about z-scores?

- A. The mean of any set of z-scores is always 0 whether the distribution is normal or not.
- B. We can determine the percentiles of any distribution by looking at the z-scores.
- C. If we convert to z-scores, Q_1 and Q_3 always fall within -1 and $+1$.
- D. All of the above are true statements.
- E. None of the above are true statements.

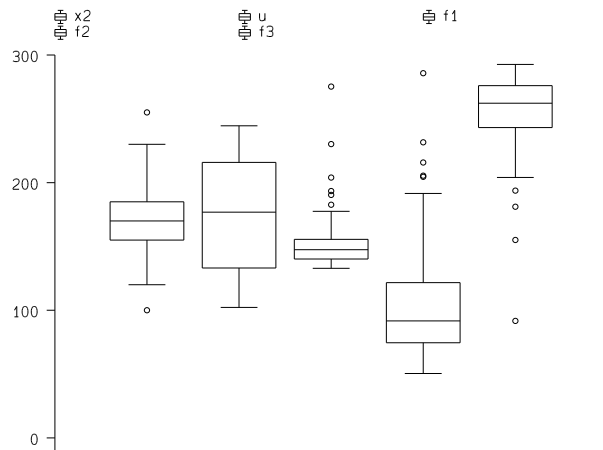
(Grid lines are 5, 10, 25, 50, 75, 90, and 95 percentiles)



10. Assuming the data above has a mean, $\bar{x} = 20$ and standard deviation, $s = 3$, what can we say about the shape of the distribution?

- A. Since this is a normal quantile plot, we can use the Empirical Rule and say about 95% of the observations will fall between 17 and 23.
- B. Since this is a normal quantile plot, we can use the Empirical Rule and say about 95% of the observations will fall between 14 and 26.
- C. Since this is a normal quantile plot, we can use the Empirical Rule and say about 95% of the observations will fall between 11 and 29.
- D. The points don't fit the line, so we can't say anything about the shape.
- E. The quantile plot shows the data is skewed to the right, so we can't use the Empirical Rule.

11. If we know that a distribution is normal with mean, $\bar{x} = 100$ and standard deviation, $s_x = 10$, what would mostly be the minimum value?
- A. 0
 - B. 10
 - C. 70
 - D. 80
 - E. 90
12. Which of the 4 stages in statistics are most important to you, a user of statistics?
- A. The Collection phase since that's where we get the information.
 - B. The Organization phase since that's where we figure out what's what.
 - C. The Analyzation phase since that's where we determine statistical significance.
 - D. The Interpretation phase since that's where we make our decisions.
 - E. Each is important to anybody who reads the newspaper.

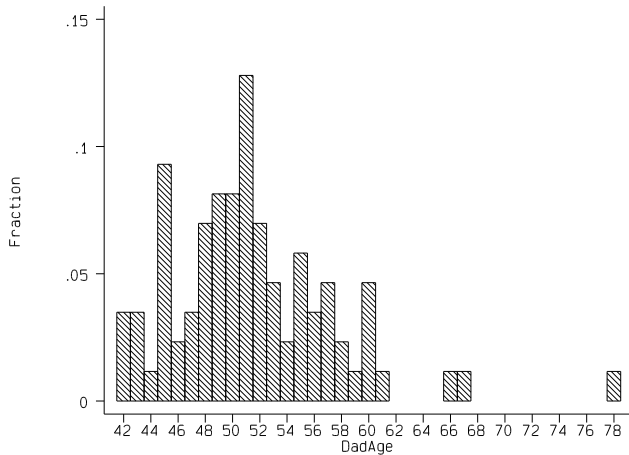


13. Numbering the boxplots above 1 to 5, which one most likely has a **mean** greater than 200?
- A. No information about the mean can be determined from a boxplot.
 - B. 5
 - C. 4
 - D. 3
 - E. 2

14. What is the best definition of *percentiles*?
- A. A value in a dataset that occurs that percent of the time.
 - B. A value in a dataset that has that percent of the observations less than or equal to it's value.
 - C. A value in a dataset that has that percent of the observations equal to it's value.
 - D. A value in a dataset that has that percent as it's value.
 - E. None of the above are correct.

	sec			
Age	509	510	511	Total
19	1	2	3	6
20	11	2	13	26
21	8	7	12	27
22	8	8	6	22
23	3	2	1	6
24	2	1	0	3
25	1	0	0	1
31	0	0	1	1
35	0	2	0	2
Total	34	24	36	94

15. According to the information above,
- A. the median age for section 510 is 22.
 - B. the most likely (typical) student is 20 and in section 511.
 - C. if you're in section 509, you're most likely 20.
 - D. All of the above are correct statements.
 - E. None of the above are correct statements.
16. How likely (what's the probability) are you to be in section 511 if you're 22 years old?
- A. 22/94
 - B. 22/36
 - C. 6/22
 - D. 6/36
 - E. 6/94



17. If the 78 year old dad was dropped for the data above, what would happen?
- A. The mean and the standard deviation would decrease.
 - B. The mean and the standard deviation would increase.
 - C. The mean would decrease, but the standard deviation would increase.
 - D. The mean and the median would decrease.
 - E. The median and the IQR would decrease.
18. Which of the following is/are true?
- A. A sample survey does not provide any useful information since it is not randomized.
 - B. Experiments are the best type of study since they control the response variable.
 - C. Case control studies must be used for any medical study since we can't randomize who gets any type of treatment.
 - D. All of the above are true.
 - E. None of the above are true.

Percentiles	Smallest	Largest
1%	80	80
5%	135	128
10%	145	130
25%	160	135
50%	172.5	
75%	185	230
90%	210	245
95%	230	260
99%	265	265

19. How many outliers are there in the data above?
- A. 1
 - B. exactly 4
 - C. exactly 5
 - D. at least 4
 - E. at least 5
20. What is the shape of the distribution in the previous problem?
- A. uniform since \tilde{x} is centered between Q_1 and Q_3
 - B. normal since \tilde{x} is centered between Q_1 and Q_3
 - C. uniform since \tilde{x} is centered between Q_1 and Q_3 and the minimum and maximum
 - D. normal since \tilde{x} is centered between Q_1 and Q_3 and the minimum and maximum
 - E. normal since it's symmetric but Q_1 and Q_3 are further from the minimum and maximum than they are from \tilde{x}

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