Exam I
Test A

STAT 201 - 502

October 6, 2005

Do not open test until instructed to do so.

You should only have a calculator, a scantron, 1 page of written notes, and something with which to write.

Read all question very carefully and make sure you have answered all questions fully before turning in your test.

On Written Section, ALWAYS SHOW YOUR WORK!! You will be graded for how you arrived at an answer not just for the answer. If you don't have room in the space provided use scratch paper.

GOOD LUCK!!!

Name: UIN:

Part 1: Multiple Choice

1. Which of the following variables are quantitative?

   (A) Gender (male or female)

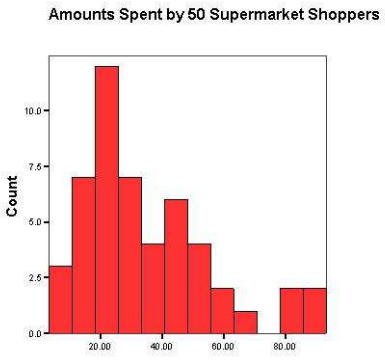
   (B) Race (Asian, Black, White, Hispanic, or other)

   (C) Weight (kilograms)

   (D) Zip Code (5-digit code)

   (E) None of the above

2. Consider the following histogram:



   Which of the following are true?

   (A) mean = median; symmetric

   (B) mean $\leq$ median; left-skewed

   (C) mean $\geq$ median; right-skewed

   (D) There is no mean

   (E) There is no median

3. I want to determine the most common car color amongst students at Texas A&M University. I randomly ask 100 students what color their car is. The best way to graphically display this data is with a

   (A) time plot

   (B) stemplot

   (C) bar graph

   (D) histogram

   (E) box plot

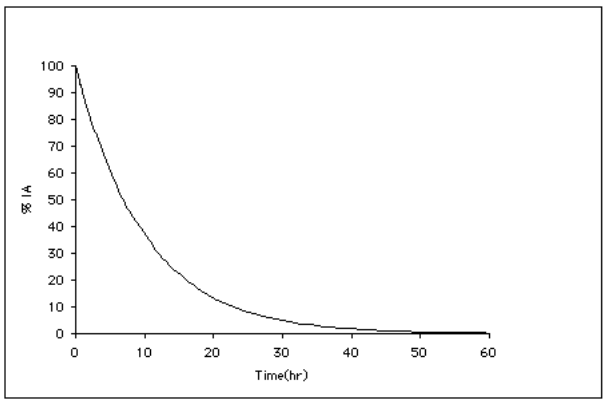
4. Which measure of distribution is the most resistant/robust to outliers?

   (A) median
   (B) standard deviation
   (C) mean
   (D) correlation
   (E) all of the above

5. The standard deviation of the three numbers 3, 8, 7 is

   (A) 6.39
   (B) 7
   (C) 2.65
   (D) 1.41
   (E) 5

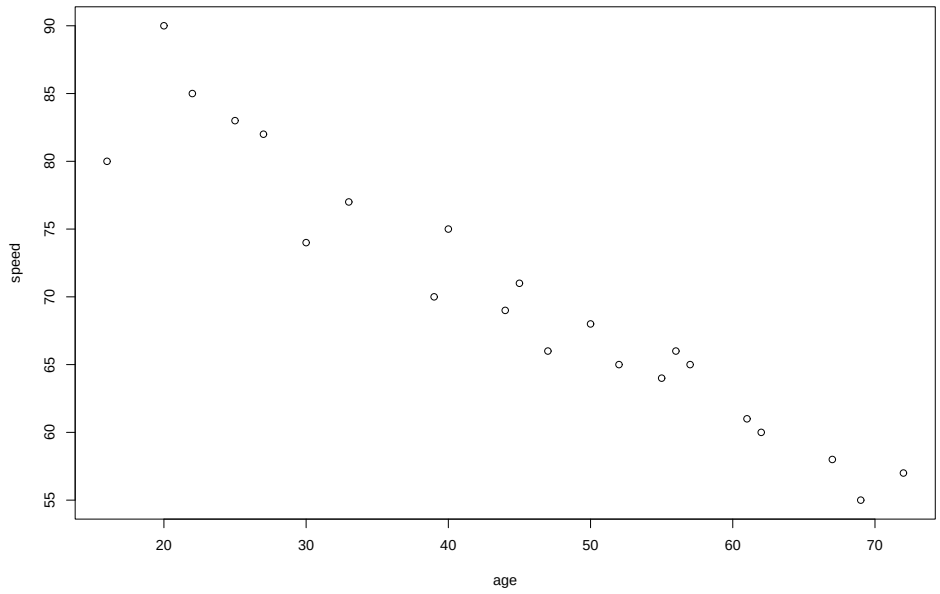
6. Consider the density curve:



   Which of the following statements is true:

   (A) This density curve is symmetric.
   (B) This density curve is right-skewed.
   (C) This density curve is left-skewed.
   (D) This density curve is normal.
   (E) The area under this density curve is 1.

7. A study was done to look at the association between age and average driving speed in the US.



The above scatterplot suggests

(A) there is a positive association between age and average driving speed; as age increases average driving speed decreases

(B) there is a negative association between age and average driving speed; as age increases average driving speed decreases

(C) there is a negative association between age and average driving speed; as age increases average driving speed increases

(D) there is a positive association between age and average driving speed; as age increases average driving speed increases

(E) None of the above describe the scatterplot

8. A lurking variable

(A) is the true variable which is explained by the explanatory variable

(B) hides in the corner before kidnapping experimental units

(C) is any variable that produces a large residual

(D) is a variable which is not explicitly measured but which affects the response variable

(E) is true cause of a response

9. John's parents recorded his height and weight at various ages up to 66 months. Below is a record of the height results:

Age (months)	36 48 54 60 66
Height (inches)	35 38 41 43 45

Referring to the data above, John's parents decide to use a least-squares regression line of John's height on age to predict his height at 21 years (252 months), we conclude that:

(A) John's height should be about half his age.

(B) They first need to convert the age units from months to years.

(C) Such a prediction could be misleading since it involves extrapolation.

(D) The parents will get a fairly accurate estimate of his height at 21 years since all of the data are clearly correlated.

(E) All of the above

10. The S&P500 stock index is the average of the price of 500 stocks. There is a moderately strong correlation between how much this index changes in January and how much it changes during the entire year. Which of the following would be true of the correlation if all 500 individual stocks were used as the data?

(A) The correlation would be lower; correlation is lower for average data than for raw data

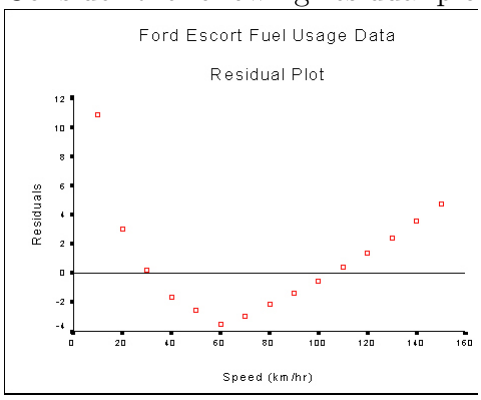
(B) The correlation would be lower; correlation is higher for average data than for raw data

(C) The correlation would be higher; correlation is higher for average data than for raw data

(D) The correlation would be higher; correlation is lower for average data than for raw data

(E) There would be no change in the correlation

11. Consider the following residual plot:



Which is the best suggestion for optimizing the fitted model?

(A) The linear regression model is a good fit; no further changes needed

(B) A variable transformation should be used on the response variable

(C) A variable transformation should be used on the explanatory variable

(D) A higher-order model should be fit

(E) A time series model should be fit

Use the following paragraph for the next 2 questions:

The Insurance Institute for Highway Safety published data on the cost of total damage suffered by compact automobiles in a series of controlled low-speed collisions. The total damage cost for a sample of 9 cars, in hundreds of dollars, is provided below:

10 6 8 10 4 3.5 7.5 8 6

12. What is the mean?

(A) 7.5
(B) 7
(C) 6
(D) 6.5
(E) 8

13. What is the median?

(A) 8
(B) 6.5
(C) 6
(D) 7
(E) 7.5

Part 2: Written Section

1. The dataset "Televisions, Physicians, and Life Expectancy" has data on the number of people per television set and the number of people per physician for 32 countries.

Country	People per TV	People per Physician
Argentina	4.0	370
Bangladesh	315.0	6166
Brazil	4.0	684
Canada	1.7	449
China	8.0	643
Colombia	5.6	1551
Egypt	15.0	616
Ethiopia	503.0	36660
France	2.6	403
Germany	2.6	346
India	44.0	2471
Indonesia	24.0	7427
Iran	23.0	2992
Italy	3.8	233
Japan	1.8	609
Kenya	96.0	7615
NorthKorea	90.0	370
SouthKorea	4.9	1066
Mexico	6.6	600
Morocco	21.0	4873
Myanmar	592.0	3485
Pakistan	73.0	2364
Peru	14.0	1016
Philippines	8.8	1062
Poland	3.9	480
Romania	6.0	559
Russia	3.2	259
SouthAfrica	11.0	1340
Spain	2.6	275
Sudan	23.0	12550
Taiwan	3.2	965
Thailand	11.0	4883
Turkey	5.0	1189
Ukraine	3.0	226
UnitedKingdom	3.0	611
UnitedStates	1.3	404
Venezuela	5.6	576
Vietnam	29.0	3096

mean for people per TV = $\bar{TV} = 38.88$
mean for people per physician = $\bar{PHYS} = 2925.90$
standard deviation for people per TV = $s_{TV} = 106.66$
standard deviation for people per physician = $s_{PHYS} = 6219.80$
correlation = $r = 0.0735$

(A) Find the equation for the least-squares regression line. Make sure to use the variable names in the equation to identify the explanatory variable(TV) and response variable(PHYS).

(B) Interpret the slope in words.

(C) Find the residuals for Vietnam. On a graph, where would this observed point lie in relation to the line (above, below, or on the line)?

2. A market researcher employs a large number of typists to enter data into a computer database. The time it takes for a new typist to learn the computer system is known to have a normal distribution with a mean of 90 minutes and a standard deviation of 18 minutes. A candidate is automatically hired if she learns the computer system in less than 100 minutes. A cut-off time is set at the slowest 10% of the learning distribution. Anyone slower than this cut-off time is definitely not hired.

   (A) What proportion of the candidates will be automatically hired?

   (B) What is the cut-off time that the market research company uses?

(C) What is the probability that a new higher takes more than 45 minutes but less than 120 minutes to learn the computer system?