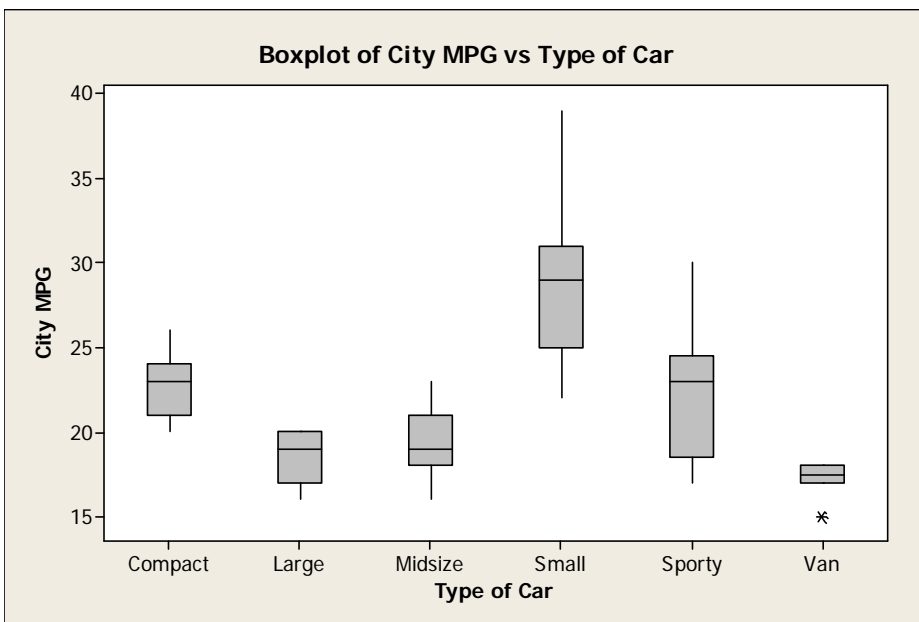


Answer the following questions NEATLY. Show all necessary work directly on the exam. Scratch paper will be discarded unread. 1 point each part unless otherwise marked. Total points: 41

1) In order to determine which kind of data display (for example: histogram versus bar graph) is appropriate for a given variable, one should consider which of the following: (Choose one.)

- A. The range of the data
- B. The median and the mean of the data
- C. Whether the relevant variable is quantitative or categorical

2) The boxplots show city MPG for 6 different types of cars. For each question below, choose one from Compact, Large, Midsize, Small, Sporty, or Van.



a) Which type of car has the lowest MPG?

van

b) Which type of car has the highest first quartile for MPG?

small

c) Which type of car has the have the smallest IQR, and what is it? (2 points)

van, ≈ 2

d) Which type of car generally has the worst MPG?

van

3) You are given the 5 Number Summary for a dataset:  $\{-3, 7, 8, 12, 23\}$ . Which of the following is **true**? (Choose one.)

- A. Fewer of the values were below 8 than were above 8.
- B. More of the values were above 8 than were between 7 and 12.
- C. The same number of the values were above 7 as were below 12.
- D. There must be a mistake in the 5 Number Summary because the values can't be negative.

4) Below is a summary of kind of airbags and class of vehicle for 90 vehicles.

Rows: airbags Columns: class

	Compact	Large	Midsize	Small	Sporty	Van	All
driver	9	7	11	4	7	3	41
none	5	0	4	15	3	6	33
passenger	2	4	7	0	3	0	16
All	16	11	22	19	13	9	90

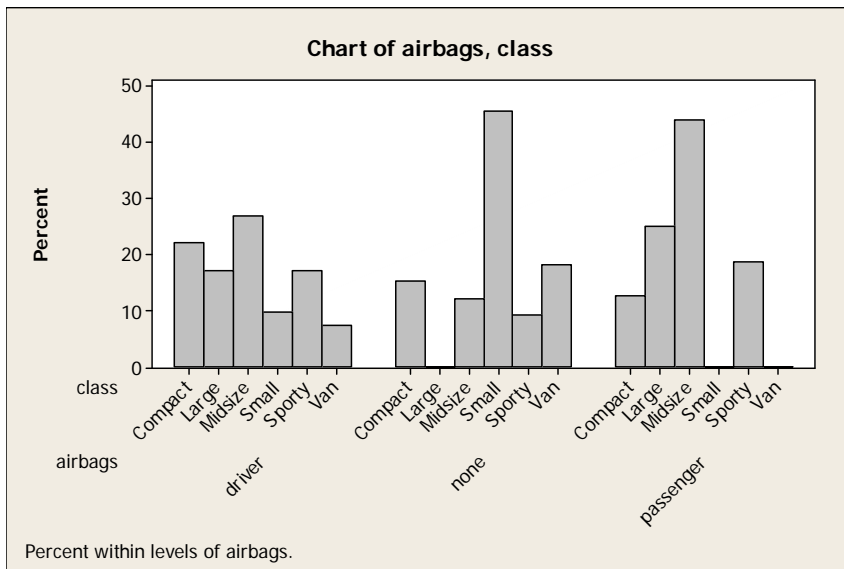
Cell Contents: Count

a) What percent of vehicles are in the compact class?  $\frac{16}{90} = 18\%$

b) What percent of vehicles are large with driver-side airbags?  $\frac{7}{90} = 8\%$  or  $\frac{11}{90} = 12\%$

c) What percent of small cars have no airbags?  $\frac{15}{19} = 79\%$

d) Below are side-by-side bar charts for class of vehicle, separated by type of airbags. Based on this chart, is type of airbags independent of class of vehicle? Explain. (2 points)



No.  
The distribution of class depends on the type of airbags.

5) A student was studying the relationship between how much money students spend on food and on entertainment per week. Based on a sample of 270, he calculated a correlation coefficient of 0.013 for these two variables. Which of the following is an appropriate interpretation?

- A. To get a higher correlation he needs to take a larger sample
- B. The correlation indicates that there is some type of linear relationship
- C. There is no linear relationship, but there may be a nonlinear relationship
- D. If he switches the response and predictor variables, he might get a better correlation.

6) A student gathered data on a large number of cars, from new cars to those that were up to 10 years old. Using the data on car ages (in years) and car prices (in US dollars), he found a linear relationship and produced the following regression model: Predicted Price =  $15620 - 1440 * \text{Age}$ .  $r = -0.93$ .

a) In the context of the problem, provide of  $-1440$ .

For each year a car ages, the predicted price decreases by \$1440.

b) In the context of the problem, provide an interpretation of 15620.

A new car costs \$15,620.

c) Find  $R^2$  and provide an interpretation of  $R^2$  in the context of this problem. (2 points)

$R^2 = .86$   
86% of the variation in car prices is explained by the linear relationship

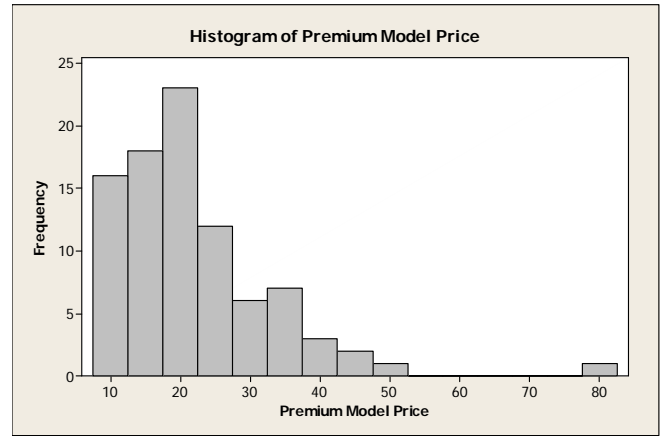
d) In the context of this problem, what would it mean to have a negative residual?

The car sold for less than predicted. with the car's age.

a) When can we make a prediction based on this model? (Choose one).

- A. For a car 6 years old, but not one 20 years old.
- B. We do not need to make a prediction for a car 6 years old because we have data for that age.
- C. For cars 6 years old and 20 years old.
- D. We cannot make a prediction because  $r$  is negative.

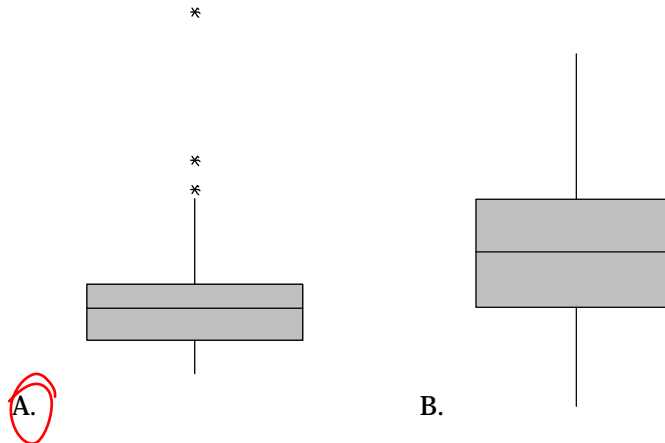
7) Provided is a histogram of costs (in \$1,000) of a premium version of 89 car models in 1993.



a) Based on the histogram, what can be said about car prices? (Choose one.)

- A. The majority of premium models cost low, although a few are priced high.
- B. The premium models of cars in America cost less than they do in Europe
- C. If you remove the cars with the highest costs, the results will show an approximately normal distribution.
- D. More people buy cheaper cars than expensive ones.

b) Which of the following boxplots represents the same data as the histogram shown above? (Choose one.)



- A.
- B.

c) Which estimates of the mean and median are most plausible? (Choose one.)

- A. Mean = 47    Median = 45
- B. Mean = 21    Median = 37
- C. Mean = 13    Median = 26
- D. Mean = 22    Median = 21

d) You are later told that the value of 80 is a mistake and should have been 50. Which of the following would be true? (Choose one.)

- A. The value of the median decreases; the value of the mean stays the same.
- B. The values of the median and mean both decrease.
- C. The value of the median stays the same; the value of the mean decreases.
- D. The values of the median and the mean both stay the same.

- e) For the histogram given on the previous page, which is the best measure of spread? (*Choose one.*)
- A. The range of approximately 70 is best because it captures the effect of the outlier.
  - B. The standard deviation is best because the data is unimodal and the standard deviation provides an approximate distance to the center.
  - C. The IQR is best because the data is skew and the IQR measures the distance between the first and third quartiles.
  - D. All of the above.

8) IQ scores are normally distributed with a mean of 100 and a standard deviation of 20.

- a) If these scores are converted to z scores, which of the following statements will be correct? (*Choose one.*)
- A. The mean will equal 5, and the standard deviation will be 20.
  - B. The mean will equal 100 because the distribution doesn't change.
  - C. The mean will equal 0, but the standard deviation cannot be determined.
  - D. Both the mean and the median score will be 0.

b) Stephy scored 117 on the IQ test. What percent of students performed lower than her? (*2 points*)

Rough sketch of the corresponding diagram:

Percent:



80.23%

c) On the IQ test, 9% of subjects scored higher than Jess. What was her score? (*2 points*)

Rough sketch of the corresponding diagram:

Score:



127

9) Data Analysis.

The data are of 1993 new car data taken from *Consumer Reports*.

The variables are:

manufacturer:	The car's manufacturer
model:	The model of the car
class:	The type of the car as defined in the Consumer Reports article
basic:	Price for a basic version of this model (in \$1,000)
midrange:	Average of Min and Max prices (in \$1,000)
premium:	Price for a premium version (in \$1,000)
city:	City MPG (miles per gallon by EPA rating)
highway:	Highway MPG
airbags:	Air Bags standard (none, driver only, driver & passenger)
drivetrain:	Drive train type (rear wheel drive, front wheel drive, all wheel drive)
cylinders:	Number of cylinders
engine:	Engine size (liters)
horsepower:	Horsepower (maximum)
RPM:	RPM (revs per minute at maximum horsepower)
RevPerMile:	Engine revolutions per mile (in highest gear)
manual:	Manual transmission available
fuel tank:	Fuel tank capacity (gallons)
passenger:	Passenger capacity (persons)
length:	Length (inches)
wheelbase:	Wheelbase (inches)
width:	Width (inches)
turn:	U-turn space (feet)
weight:	Weight (pounds)
domestic:	Domestic (non-U.S. manufacturer, U.S. manufacturer)

Data can be found at <http://www.canyons.edu/faculty/morrowa/140/examsandquizzes/>.

This prompt can be found at <http://www.canyons.edu/faculty/morrowa/140/examsandquizzes/> (FORM A)

*Use Word to answer the following questions. Print your solutions when you are ready.*

*Your final write-ups should include ONLY the graphs/statistics that are relevant.*

*Suggested Discussion Points:*

- *Describe the distribution/relationship as applicable*
- *If it is appropriate, fit an appropriate model (Normal model, linear model).*

*Provide evidence (appropriate graphs and statistics) for all of your findings.*

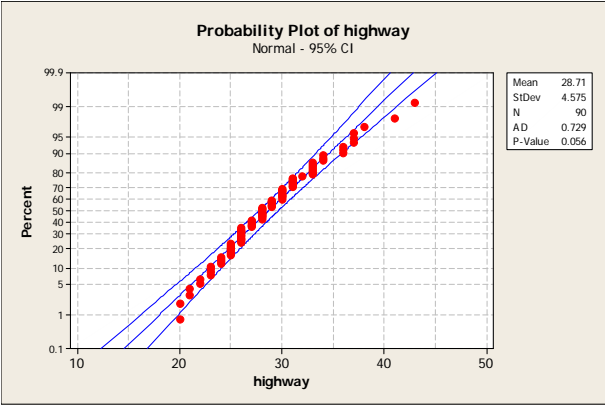
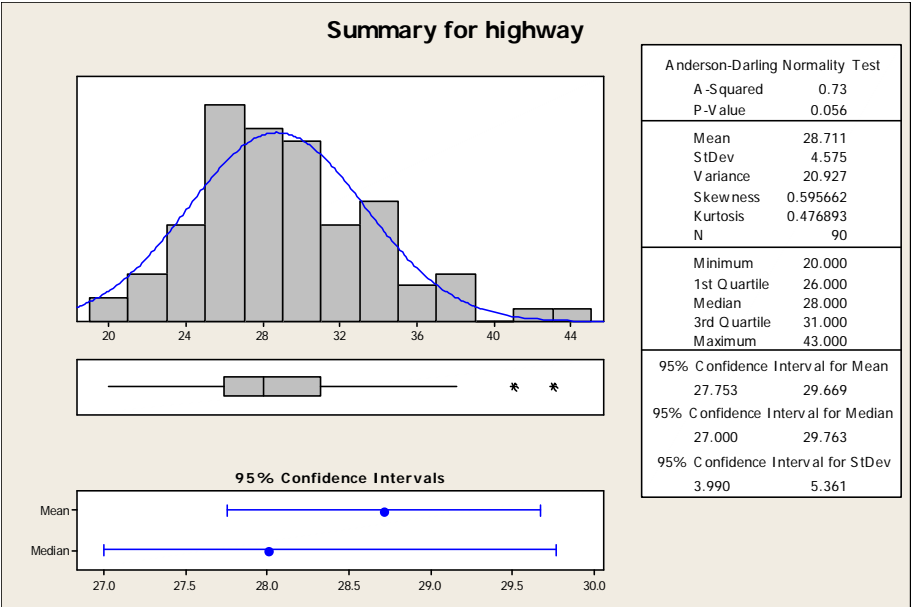
*3 points each*

**GENERAL NOTES ON SOLUTIONS... IN ADDITION TO THE POINTS BELOW, CORRECT SOLUTIONS WILL NOT INCLUDE INCORRECT CONCLUSIONS.**

a) One variable: Analyze/summarize the highway MPG of the cars in the study. Variable: **highway**.

**A CORRECT SOLUTION WILL INCLUDE:**

- **BOXPLOT/HISTOGRAM**
- **SHAPE: SLIGHT SKEW, UNIMODAL, TWO OUTLIERS (BOXPLOT)**
- **CENTER: MEDIAN (28)**
- **SPREAD: IQR (5)**
- **DATA APPEARS SLIGHTLY SKEW—NORMAL MODEL IS NOT APPROPRIATE AS SHOWN IN PROBABILITY PLOT**



b) Two variables: Analyze/summarize the relationship between the drive train and the type of airbags of the cars in the study. Variables: **drivetrain, airbags**.

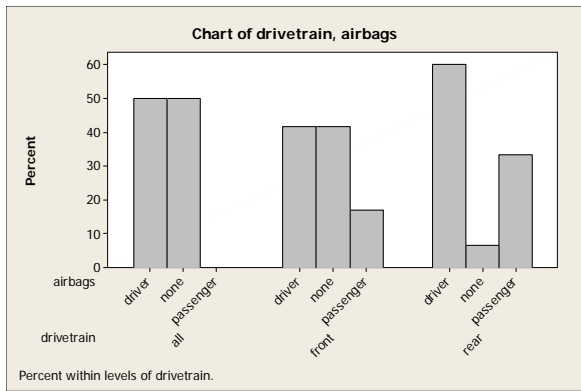
**A CORRECT SOLUTION WILL INCLUDE:**

- **SIDE-BY-SIDE BAR CHARTS**
- **CONTINGENCY TABLE**
- **THE DRIVE TRAIN AND TYPE OF AIRBAGS OF A CAR APPEAR TO BE DEPENDENT.**

Rows: drivetrain    Columns: airbags

	driver	none	passenger	All
all	5	5	0	10
front	27	27	11	65
rear	9	1	5	15
All	41	33	16	90

Cell Contents:            Count

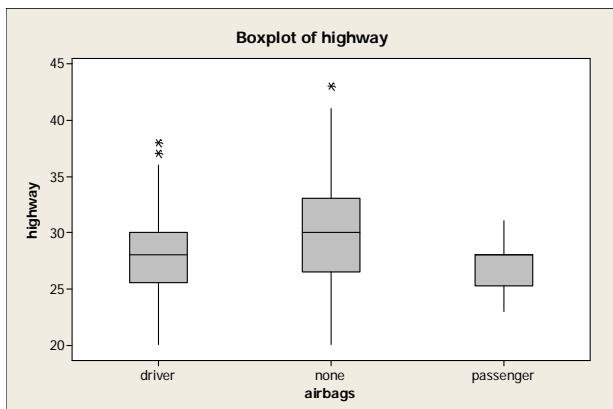
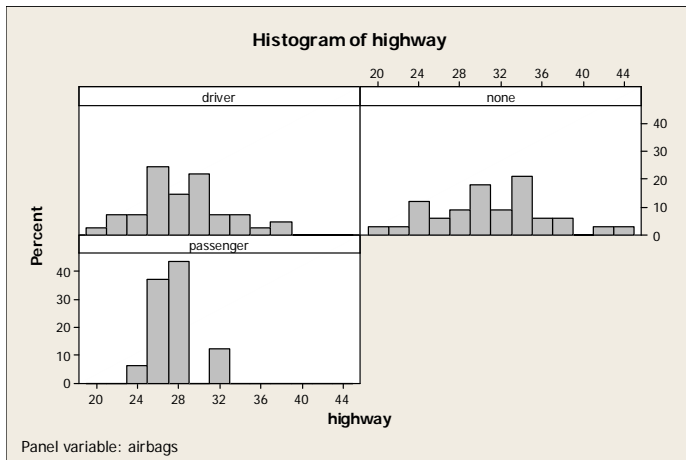


c) Two variables: Analyze/summarize the relationship between highway MPG and the type of airbags of the cars in the study. Variables: **highway**, **airbags**.

A CORRECT SOLUTION WILL INCLUDE:

- SIDE-BY-SIDE BOXPLOTS/HISTOGRAMS
- SHAPE: EACH HISTOGRAM APPEARS ROUGHLY SYMMETRIC, UNIMODAL. PASSENGER CARS HAVE A GAP. DRIVER AND NONE HAVE OUTLIERS ON THE HIGH END.
- CENTER: CARS WITHOUT AIRBAGS HAVE THE HIGHEST MEAN/MEDIAN HIGHWAY MPG.
- SPREAD: CARS WITH PASSENGER AIRBAGS HAVE THE LOWEST STANDARD DEVIATION AND IQR. CARS WITHOUT AIRBAGS HAVE THE HIGHEST SPREAD.

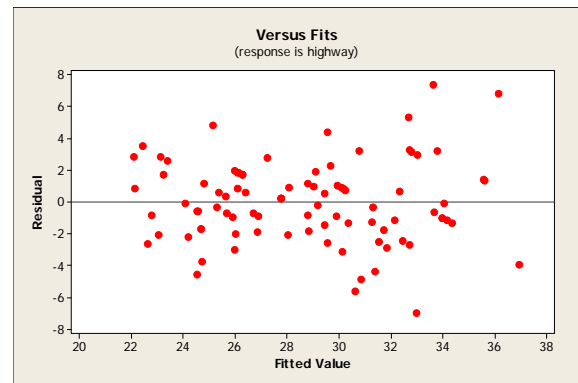
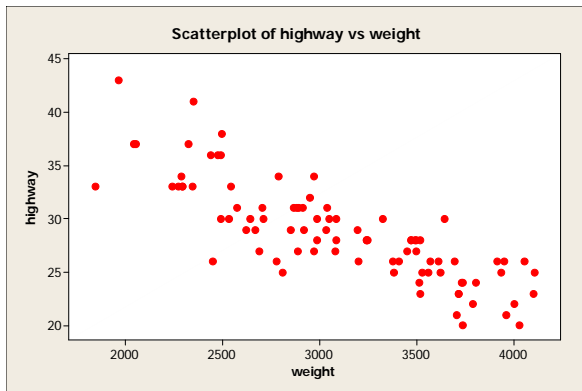
Variable	airbags	Mean	StDev	Median	IQR
highway	driver	28.073	4.156	28.000	4.500
	none	30.273	5.490	30.000	6.500
	passenger	27.125	2.156	28.000	2.750



- d) Two variables: Analyze/summarize the relationship between weight and highway MPG of the cars in the study. Variables: **weight**, **highway**.

**A CORRECT SOLUTION WILL INCLUDE:**

- A SCATTERPLOT WITH HIGHWAY ON Y AND WEIGHT ON X
- DESCRIPTION OF SCATTERPLOT: MODERATE POSITIVE LINEAR RELATIONSHIP,
- REPORT  $r$  (SEE BELOW).  $r$  IS MODERATE, INDICATING THAT THE LINEAR RELATIONSHIP IS MODERATE.
- FIT LINEAR REGRESSION OF HIGHWAY ON WEIGHT (SEE BELOW)
- ANALYZE LINEAR FIT USING BOTH  $R^2$  AND RESIDUAL PLOT
  - $R^2$  IS HIGH. THE LINEAR MODEL EXPLAINS 68.1% OF THE VARIATION IN HIGHWAY
  - OVERALL, THE RESIDUAL PLOT LOOKS GOOD



Pearson correlation of highway and weight = -0.825

The regression equation is  
highway = 49.0 - 0.00656 weight

Predictor	Coef	SE Coef	T	P
Constant	49.020	1.508	32.51	0.000
weight	-0.0065550	0.0004786	-13.70	0.000

S = 2.59976    R-Sq = 68.1%    R-Sq(adj) = 67.7%