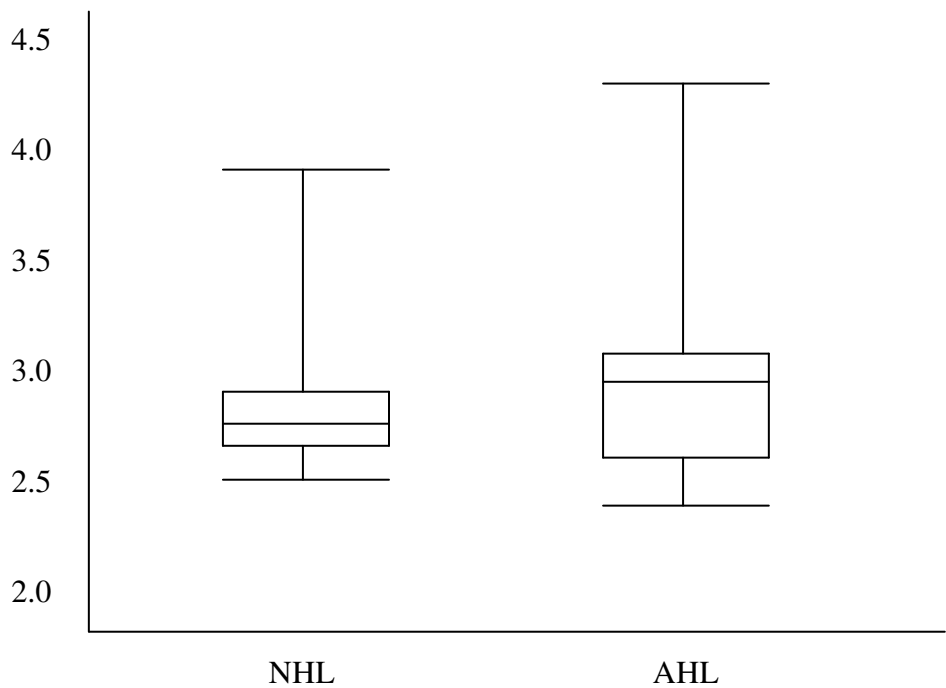


**STAT 2220 – Sample Midterm Test 1B**

**Part A**

The first **two** questions (**1** and **2**) refer to the following:

The average number of goals per game for the 30 National Hockey League teams and the 29 American Hockey League teams for the 2009/10 season are displayed in the side-by-side basic (quantile) boxplots shown below:



Some summary statistics for the two distributions are shown below:

	Min	Q1	med	Q3	max	mean	std. dev.
NHL	2.51	2.65	2.74	2.90	3.88	2.85	0.29
AHL	2.41	2.60	2.93	3.05	4.28	2.92	0.38

1. Which of the following statements is/are true?

I. The interquartile range for the AHL is greater than that for the NHL.  
II. The distribution of goals per game for NHL teams is skewed to the right.  
III. About 7 AHL teams scored more than 3.05 goals per game.  
  
(A) I only      (B) I and II only      (C) I and III only      (D) II and III only      (E) I, II and III
2. The NHL consists of 24 teams from the U. S. and 6 Canadian teams. The mean number of goals per game for Canadian teams was 2.74. What was the mean number of goals for the U. S. teams?  
  
(A) 2.88              (B) 2.90              (C) 2.92              (D) 2.94              (E) 2.96
3. The GPAs for a sample of 15 students at a university are ordered and are shown below:

1.98      2.23      2.40      2.63      2.89      3.01      3.15      3.22  
3.37      3.41      3.68      3.96      4.08      4.19      4.30

What is the interquartile range of GPAs for this sample?

- (A) 1.21              (B) 1.06              (C) 1.28              (D) 1.55              (E) 1.33

4. The frequency distribution shown below displays the daily saturated fat intakes (in grams) for a sample of people.

Saturated Fat Intake	Frequency
10 – 20	18
20 – 30	19
30 – 40	20
40 – 50	8
50 – 60	4
60 – 70	3
70 – 80	1

The distribution of saturated fat intake is:

- (A) skewed to the left and so the mean is greater than the median.  
 (B) skewed to the right and so the mean is greater than the median.  
 (C) skewed to the left and so the median is greater than the mean.  
 (D) skewed to the right and so the median is greater than the mean.  
 (E) approximately symmetric and so the mean and median are approximately equal.
5. A father drops his three children off at the movie theater and gives them each \$25 to spend on a movie and concession snacks. After each of the children buys a movie ticket for \$9, what is the standard deviation of the amounts of money the children have left?
- (A) \$3.00            (B) \$2.12            (C) \$4.00            (D) \$2.83            (E) \$0.00
6. We want to calculate the correlation  $r$  between two variables  $X$  and  $Y$ . Which of the following conditions are necessary for  $r$  to be a meaningful measure of association?
- I.  $X$  and  $Y$  are both quantitative variables.  
 II. The relationship between  $X$  and  $Y$  is linear.  
 III.  $X$  is an explanatory variable and  $Y$  is a response variable.
- (A) I only  
 (B) II only  
 (C) I and II only  
 (D) I and III only  
 (E) I, II and III
7. Three friends took the same university course. Each of them scored 25% lower on the final exam than they did on the midterm test (i.e. each student's final mark was only 75% of their midterm mark). What is the correlation between midterm score and final score for these three friends?
- (A) 0.25            (B)  $-1$             (C) 0.75            (D) 1            (E)  $-0.25$
8. Can a student's midterm score be used to predict his final exam score? The midterm scores  $X$  (out of 50) and the final exam scores  $Y$  (out of 100) are measured on a sample of students in a large class. The least-square regression line is calculated to be  $\hat{y} = 8 + 1.8x$ . What would be the equation of the least-squares regression line if we had instead recorded the student's midterm score as a percentage (i.e., out of 100)?
- (A)  $\hat{y} = 8 + 0.9x$   
 (B)  $\hat{y} = 16 + 3.6x$   
 (C)  $\hat{y} = 8 + 1.8x$   
 (D)  $\hat{y} = 16 + 0.9x$   
 (E)  $\hat{y} = 8 + 3.6x$

9. A local research firm often uses newspaper advertisements to invite individuals to volunteer for various experiments the firm is planning to conduct. One such study was to determine if a newly formulated cream had any beneficial effect on individuals suffering with psoriasis (a skin condition). Psoriasis sufferers were invited to participate in a study with volunteers to be given either the new cream or a placebo in the form of a cream with similar appearance and feel as the new cream. What conclusions can be drawn from such studies involving volunteers?
- (A) Useful information can be gained from studies such as this, provided that the individuals who volunteer are randomly assigned to either receive the experimental treatment or the placebo.
  - (B) Because there is bias associated with the selection of subjects, nothing can be concluded.
  - (C) In such a study there is great potential for the presence of confounding or lurking variables so little information can be gained.
  - (D) Little if anything can be concluded because this is an observational study.
  - (E) Because there is no control over the selection of subjects, little will be learned from such studies.
10. Will a fluoride mouthwash used after brushing reduce cavities? Twenty sets of twins were used To investigate this question. One member of each set of twins was randomly assigned to use the mouthwash after each brushing, while the other did not. After 6 months, the difference in the number of cavities of those using the mouthwash was compared with the number of cavities of those who did not use the mouthwash. This is an example of:
- (A) a double blind experiment.
  - (B) a single blind experiment.
  - (C) an experiment that uses a placebo.
  - (D) a matched pairs experiment.
  - (E) an observational study.
11. A simple random sample of size  $n$  is the only type of sample that guarantees that:
- (A) every individual in the population has a known chance of being selected into the sample.
  - (B) every individual in the population has an equal chance of being selected into the sample.
  - (C) every group of  $n$  individuals has an equal chance of being selected into the sample.
  - (D) results will not be biased.
  - (E) all of the above
12. A stratified random sample is conducted by:
- (A) grouping the subjects using some criteria, then randomly selecting subjects from each group to interview.
  - (B) randomly selecting subjects to interview, then grouping the subjects using some criteria.
  - (C) grouping the subjects using some criteria, then randomly selecting entire groups to interview.
  - (D) randomly selecting subjects to interview, then randomly selecting groups to interview.
  - (E) randomly grouping the subjects, then randomly selecting the subjects from each group to interview.



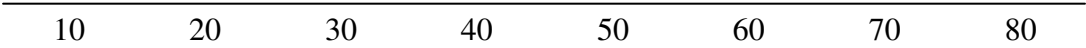
**Part B**

1. The Winnipeg Symphony Orchestra is performing one night at the Centennial Concert Hall. A random sample of 22 audience members is selected. Their ages are ordered and are shown below:

19   28   33   36   42   45   48   50   52   52   53  
55   58   59   60   61   61   65   68   70   72   79

The five-number summary for this data set is:   **19   45   54   61   79**

- (a) Use the scale below to construct a modified (outlier) boxplot for this data set. Show any necessary calculations.



What is the shape of the distribution of ages?

- (b) The same evening, Elton John is performing at the MTS Center. A random sample of 17 audience members is selected. Their ages are recorded and ordered and are shown below:

17   18   23   25   29   30   33   37   37   41   42   46   49   50   54   54   62

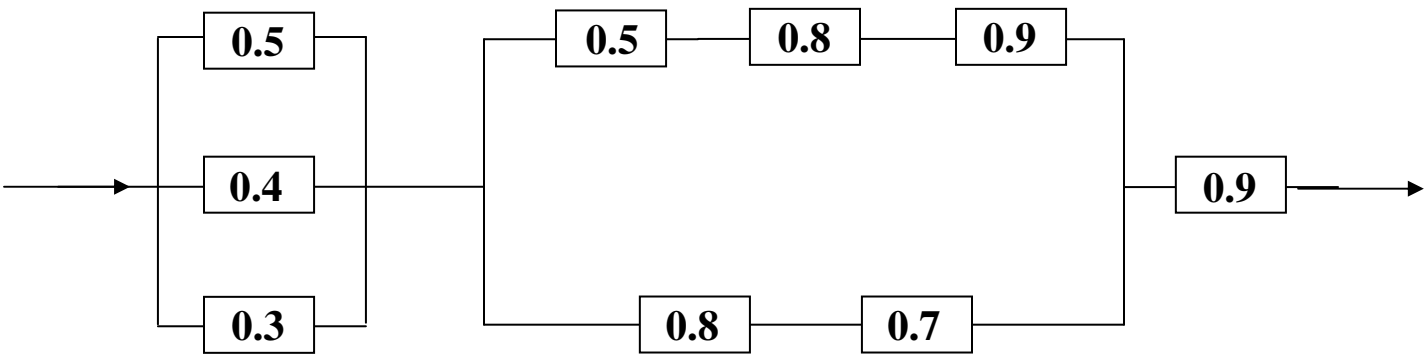
Create a back-to-back stemplot of the ages of audience members at the symphony and the Elton John concert. Would it be more appropriate to use the five-number summary or the mean and standard deviation to summarize the distribution of ages at the Elton John concert? Explain.

2. The owner of an ice cream truck would like to know if the temperature affects the amount of ice cream he sells. The temperature (in °C) and ice cream sales (in \$) for a sample of six days from the past summer are recorded. The sample means and standard deviations are shown below.

	mean	std. dev.
Temperature	24.00	4.86
Sales	310.00	123.98

A scatterplot of the data reveals that a linear relationship between Temperature and Sales is an appropriate assumption. The equation of the least squares regression line is calculated to be  $\hat{y} = -231.92 + 22.58x$ .

- Find the value of the correlation between Temperature and Sales.
  - Interpret the slope of the least squares regression line in this example.
  - One day, the temperature was 27°C and the ice cream sales were \$350. Find the value of the residual for this day. What does the sign of the residual tell us?
  - What percentage of the variation in Sales can be accounted for by its regression on Temperature?
3. An automobile manufacturer would like to examine the effect of speed and type of tire on the stopping distance of its vehicles in winter driving conditions. Vehicles will be outfitted with either regular tires or snow tires, and will drive at a speed of either 40, 60 or 80 km/h. The stopping distances will then be compared. Each combination of factor levels will be tested on three vehicles. The manufacturer believes that the effect of speed and type of tire may differ for its different models of car, so the experiment is conducted separately for each model of car they make. Identify each of the following in this experiment: (0.5 marks each)
- Factor(s):
  - Treatment(s):
  - Response Variable(s):
  - Blocking Variable(s):
4. (a) Calculate the reliability of the following system, where the value in each box represents the reliability of that component.



- (b) A simple system consists of three components placed in parallel. Each component has the same reliability  $r$ . What must be the value of  $r$  in order for the reliability of the entire system to be at least 0.95?