

KINGS UNIVERSITY COLLEGE  
THE UNIVERSITY OF WESTERN ONTARIO  
DEPARTMENT OF STATISTICAL AND ACTUARIAL SCIENCES  
**STATISTICS 1024B MIDTERM TEST**

Saturday, March 1st, 2014, 9:00 - 11:30 AM

<u>Section</u>	<u>Instructor</u>
S001 (Tu/Th)	D. Bellhouse
S200 (Blended)	L. Murray
Kings (570)	S. Kopp
Kings (571)	S. Valluri

INSTRUCTIONS:

- This is a closed book test.
- Use only an HB pencil for the Scantron sheet.
- Fill in NAME, INSTRUCTOR, SIGNATURE and COURSE on the Scantron sheet.
- **Fill in and code your STUDENT NUMBER, SECTION and EXAM CODE on the Scantron sheet.** Your EXAM CODE is 280.
- Leave the ANSWER SHEET NUMBER blank on the Scantron sheet.
- There are 35 multiple choice questions. A formula sheet, Table A and a blank page for rough work is at the end of the paper.
- **Code** your answers on the Scantron sheet.
- Only non-programmable calculators are permitted.
- **NO EXTRA TIME WILL BE GIVEN TO CODE YOUR ANSWERS!!**

**GOOD LUCK!**





- 7) Below you are given two stemplots. Number 1 is the stemplot for the annual average rate of return of stocks in the United States from 1950 to 1997 (to the nearest percent; lowest is  $-26\%$ , highest is  $50\%$ ). Number 2 is the stemplot for the annual average rate of return of US Government Treasury Bills for the same 48-year period (to the nearest decimal point; lowest is  $0.9\%$ , highest is  $14.7\%$ ).

1. Stocks	2. Treasury Bills
-2   6	0   9
-1   500	1   257789
-0   98753	2   4788
0   011456778	3   023556
1   01224667999	4   024449
2   022334479	5   12455699
3   022227	6   149
4   33	7   02577
5   0	8   04
	9   09
	10   68
	11   4
	12   2
	13
	14   7

In describing the two distributions, which would be more informative: the five number summary OR the mean and standard deviation,  $(\bar{x}, s)$ ?

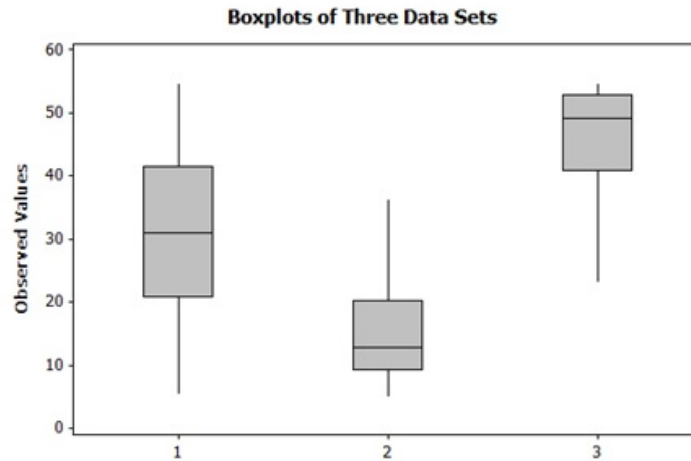
- (A) Five number summary for both stocks and Treasury Bills.  
 (B)  $(\bar{x}, s)$  for both stocks and Treasury Bills.  
 (C) Five number summary for stocks,  $(\bar{x}, s)$  for Treasury Bills  
 (D)  $(\bar{x}, s)$  for stocks, five number summary for Treasury Bills.
- 8) Two sets of data give the following results:

	First Sample	Second Sample
Sample size	10	30
Mean	4	8
Median	3	5
Standard Deviation	3	4

If the two data sets are combined into one large data set of 40 observations, the mean of the combined data set is

- (A) 4                                      (B) 5                                      (C) 6                                      (D) 7

- 9) You are given the following side-by-side boxplots representing the distribution of three different data sets (labelled 1, 2 and 3):



You are also given the following three different descriptions of the distribution of each data set:

$a$  = skewed to the right

$b$  = skewed to the left

$c$  = symmetric

Which of the following is the correct matching?

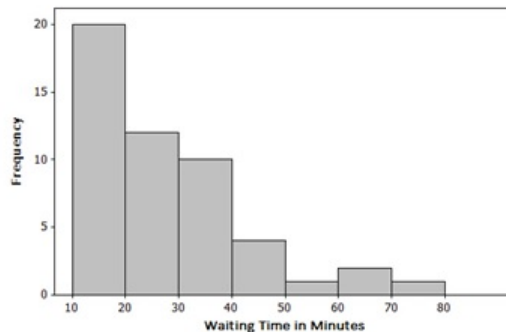
- (A)  $1 = a, 2 = b, 3 = a$
- (B)  $1 = b, 2 = a, 3 = b$
- (C)  $1 = c, 2 = b, 3 = a$
- (D)  $1 = c, 2 = a, 3 = b$

- 10) The stemplot below displays midterm exam scores for the 38 students taking a course. The highest possible test score was 100.

2	8
3	6
4	378
5	1235567889
6	0012334566777
7	11235788
8	9
9	3

Which scores are outliers?

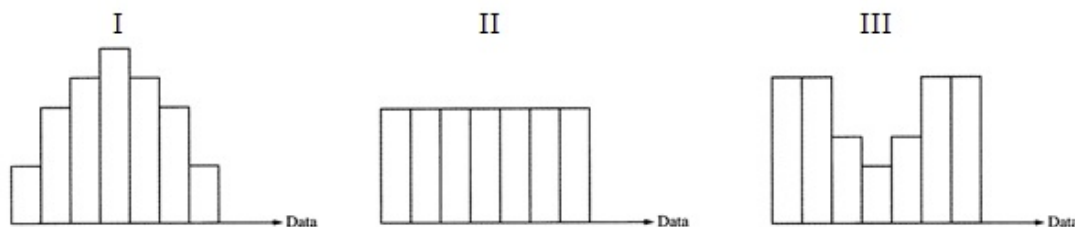
- (A) 28 only
  - (B) 93 only
  - (C) 28, 36 and 93
  - (D) 28 and 93
  - (E) Correct answer is not given by (A), (B), (C) or (D).
- 11) You are given a histogram of the waiting times in minutes that a sample 50 patients experienced waiting to be seen by the specialist after arrival in the specialist’s office.



From the histogram, what can be said about the mean and median of the data.

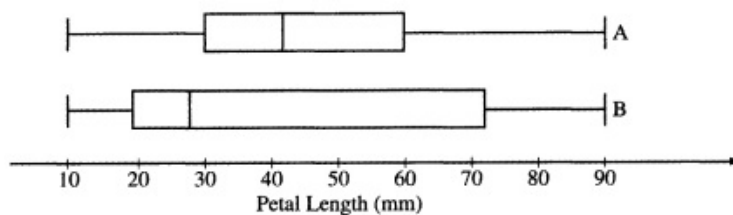
- (A) The median is between 30 and 40 minutes and the mean is less than the median.
- (B) The median is between 20 and 30 minutes and the mean is greater than the median.
- (C) The median is between 30 and 40 minutes and the mean is greater than the median.
- (D) The median is between 20 and 30 minutes and the mean is less than the median.

- 12) The histograms labelled I, II and III represent the distributions of three different data sets, each containing 28 integers, from 1 through 7, inclusive. The horizontal and vertical scales are the same for all graphs. Note that the histograms are all symmetric.



Which graph represents the data set with the largest standard deviation?

- (A) I  
 (B) II  
 (C) III  
 (D) It is impossible to tell from the information given.
- 13) A botanist is studying the petal lengths, measured in millimeters, of two species of lilies. The boxplots above illustrate the distribution of petal lengths from two samples of equal size, one from species A and the other from species B. Based on these boxplots, which of the following is a correct conclusion about the data collected in this study?



Based on these boxplots, which of the following is a correct conclusion about the data collected in this study?

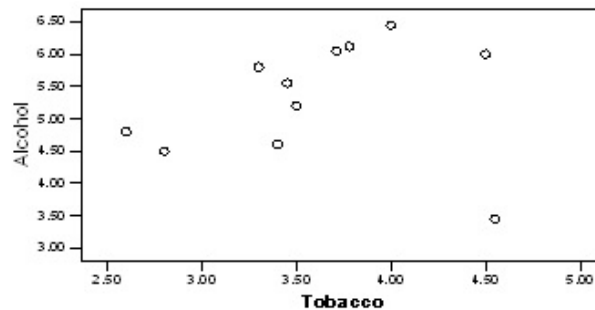
- (A) The range for species B is greater than the range for species A.  
 (B) The interquartile ranges are the same for both samples.  
 (C) There are more petal lengths that are greater than 70 mm for species A than there are for species B.  
 (D) There are more petal lengths that are less than 30 mm for species B than there are for species A.

- 14) Which of the following is a true statement about a normal density curve?
- (A) The mean and the median differ in value.
  - (B) It has a peak centered above its mean.
  - (C) About 5% of the curve lies to the right of the mean plus two standard deviations.
  - (D) Each of (A), (B) and (C) are true.
- 15) A study recently concluded that the amount of digital information Canadians consume per day follows a normal distribution with mean  $\mu = 34.5$  gigabytes and standard deviation  $s = 10.2$  gigabytes. Using the 68-95-99.7 rule, approximately what percentage of Canadians consume between 24.3 and 54.9 gigabytes of information per day?
- (A) 81.5%                      (B) 84.0%                      (C) 79.0%                      (D) 83.7%
- 16) The marks achieved by Mikki in Mathematics, English and French, together with the mean and the standard deviation for each subject, are given in the following table:

Subject	Mark	Mean	Standard Deviation
Mathematics	60	52	10
English	70	70	5
French	65	60	3

- If marks in each class are normally distributed, which of the following statements is correct in terms of Mikki's relative standing in each of her classes?
- (A) Mikki's best subject was English, followed by Mathematics and then French.
  - (B) Mikki's best subject was Mathematics, followed by French and then English.
  - (C) Mikki's best subject was French, followed by Mathematics and then English.
  - (D) Mikki's best subject was English, followed by French and then Mathematics.
- 17) Birthweights at a local hospital have a normal distribution with a mean of 3120 grams and a standard deviation of 425 grams. The proportion of infants with birthweights above 3545 grams is
- (A) 0.1587                      (B) 0.3409                      (C) 0.5000                      (D) 0.8413

- 18) Suppose that pulse rates of people in a certain population are normally distributed with a mean of 69.4 beats per minute. If 10% of people have pulse rates of more than 79 beats per minute, then the standard deviation of pulse rate in this population is (to one decimal place):
- (A) 4.9                      (B) 4.1                      (C) 9.6                      (D) 7.5
- 19) The weights of a population of adult male gray whales are approximately normally distributed with a mean weight of 18,000 kilograms and a standard deviation of 4,000 kilograms. The weights of a population of adult male humpback whales are approximately normally distributed with a mean weight of 30,000 kilograms and a standard deviation of 6,000 kilograms. A certain adult male gray whale weighs 24,000 kilograms. This whale would have the same standardized weight ( $z$ -score) as an adult male humpback whale whose weight, in kilograms, is equal to which of the following?
- (A) 24,000                      (B) 36,000                      (C) 21,000                      (D) 39,000
- 20) A study on alcohol and tobacco was conducted and the results are shown in the following scatterplot. An outlier can be found on the bottom right side of the plot.

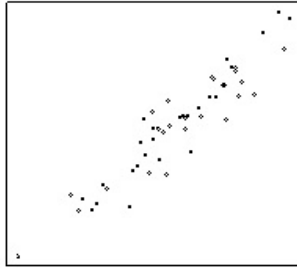


If the outlier was removed, the value of correlation  $r$  would

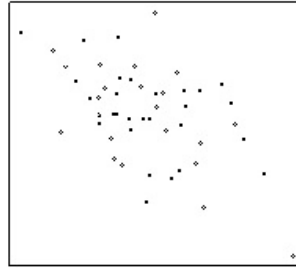
- (A) become closer to  $-1$ .  
(B) become closer to  $1$ .  
(C) become closer to  $0$ .  
(D) be impossible to determine.

21) Match the four plots with the most likely values of their correlations.

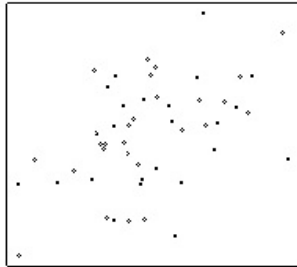
Plot A



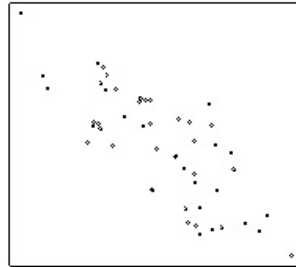
Plot B



Plot C



Plot D



- (A) Plot A: 0.45, Plot B:  $-0.8$ , Plot C: 0.9, Plot D:  $-0.5$
- (B) Plot A:  $-0.9$ , Plot B: 0.5, Plot C:  $-0.45$ , Plot D: 0.8
- (C) Plot A: 0.9, Plot B:  $-0.8$ , Plot C: 0.45, Plot D:  $-0.5$
- (D) Plot A: 0.9, Plot B:  $-0.5$ , Plot C: 0.45, Plot D:  $-0.8$

22) A researcher states that bone density in women is negatively associated with age. This means that

- (A) above-average values of age tend to accompany above-average values of bone density.
- (B) above-average values of age tend to accompany below-average values of bone density.
- (C) below-average values of age tend to accompany below-average values of bone density.
- (D) older women aren't any more likely than younger women to have below-average bone density.



Use the following information for the next TWO questions (Q27 and Q28):

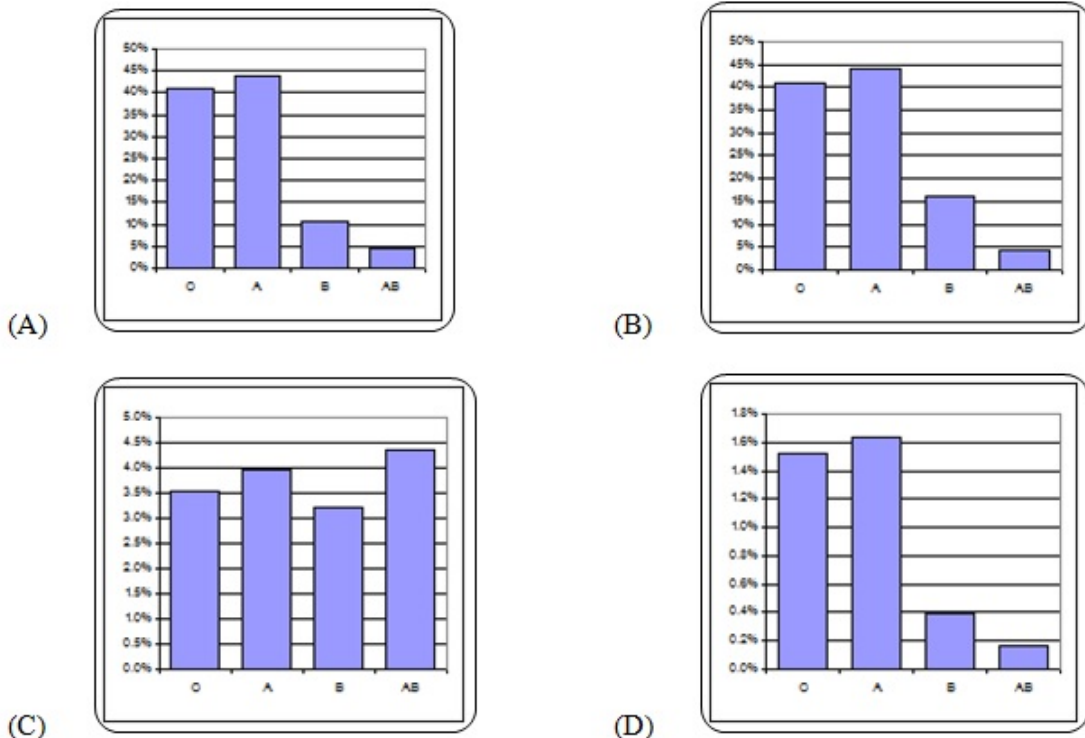
The following table gives the breakdown of blood type and ethnic background in the state of Hawaii.

Blood Type	Ethnic Background				Totals
	Hawaiian	Hawaiian-White	Hawaiian-Chinese	White	
O	1,903	4,469	2,206	53,759	62,337
A	2,490	4,671	2,368	50,008	59,537
B	178	606	568	16,252	17,604
AB	99	236	243	5,001	5,579
Totals	4,670	9,982	5,385	125,000	145,057

27) What percentage of people have blood type B, conditioned on the fact that they are white?

- (A) 92.3%                      (B) 13.0%                      (C) 12.1%                      (D) 11.2%

28) Which of the following is the correct bar graph for representing the distribution of blood type for the Hawaiian-Chinese?



- (A) Graph A                      (B) Graph B                      (C) Graph C                      (D) Graph D

- 29) One hundred people were interviewed and classified according to their attitude toward small cars and their personality type. The results are shown in the table below.

Attitude	Personality Type		Total
	Type A	Type B	
Positive	25	12	37
Neutral	11	9	20
Negative	24	19	43
Total	60	40	100

Which of the following statements is (are) true?

- (A) The proportion that has a positive attitude toward small cars is higher among people with a type B personality type than among people with a type A personality type.
  - (B) Of the three attitude groups, the group with the neutral attitude has the highest proportion of type B personality types.
  - (C) Of the three groups, the group with the negative attitude has the highest proportion of type A personality types.
  - (D) For each personality type, more than half of the 100 respondents have a neutral attitude toward small cars.
- 30) Which of the following is not a characteristic of stratified random sampling?
- (A) The strata are based on facts known before the sample is selected.
  - (B) Every possible subset of the population of the desired sample size has an equal chance of being selected.
  - (C) Random sampling is part of the sampling procedure.
  - (D) The population is divided into groups of units that are similar on some characteristic.

- 31)** An amateur gardener decides to compare two varieties of tomatoes for the current growing season to see if the yield of one is better than the other. He divides his garden into twelve plots and puts the plants of one variety into six of the plots nearest the edge of the garden and the plants of the other variety into the remaining plots in the middle of the garden. The average yield per plant of the first variety was 11.3 pounds per plant while the average yield of the second variety was 14.5. This is an example of
- (A) a multistage design, since two years were involved.
  - (B) an experiment.
  - (C) an observational study, not an experiment.
  - (D) the elimination of all confounding variables by design, since the gardener used the same part of the garden in both years.
- 32)** Automobile brake pads are either metallic or nonmetallic. An experiment is to be conducted to determine whether the stopping distance is the same for both types of brake pads. In previous studies, it was determined that car size (small, medium, large) is associated with stopping distance, but car type (sedan, wagon, coupe) is not associated with stopping distance. The experiment would be best done
- (A) by blocking on car type.
  - (B) by blocking on car size.
  - (C) by blocking on brake pad type.
  - (D) by blocking on stopping distance.
- 33)** An experiment is said to be biased if
- (A) it allows racial or gender differences in the subjects taking part in the experiment.
  - (B) the researcher has speculated in advance as to what the outcome might be.
  - (C) it is conducted in such a way as to systematically favour certain outcomes.
  - (D) it allows individuals with strong opinions to take part in the experiment.

- 34)** A group of medical researchers studied the effects of two different diets: LOW CARB and LOW FAT. The researchers randomly assigned each of 132 severely obese patients to one of the two diets (66 patients to each diet) and monitored their results for 6 months. The researchers themselves cautioned against interpreting these results as a general endorsement for a low-carb diet for anyone wanting to lose weight. Why was this caution likely made?
- (A) This experiment only used severely-obese subjects.
  - (B) This experiment did not preserve subject confidentiality because we know how much weight was lost.
  - (C) This is an observational study, and we know it is difficult to generalize the results.
  - (D) This experiment did not use blinding.
- 35)** A population consists of 52 individuals labelled 00 to 51. Using the following list of random digits, a sample of five of these individual is to be chosen by simple random sampling.

55890 07436 14863 35119 71035 09001 43367 13258 54580

The sample consists of the units labelled

- (A) 07 09 14 35 43
- (B) 00 14 33 43 51
- (C) 00 14 33 36 51
- (D) 07 11 14 35 43

Use this page for rough work

# FORMULA SHEET

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$z = \frac{x - \mu}{\sigma}$$

$$\hat{y} = a + bx$$

$$b = r \frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

