KING'S UNIVERSITY COLLEGE
THE UNIVERSITY OF WESTERN ONTARIO
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCES

STATISTICS 1024B Practice Midterm (2 hours)

INSTRUCTIONS:

This is a closed-book test.

There are **35** multiple choice questions to be answered using the provided Scantron sheet.

Only non-programmable calculators are permitted.

Use only an HB pencils when filling out the scantron sheet:

***Print* your name, course and section number where requested.**
Carefully code in your ID, and section number
Enter 378 in the CODE box on the Scantron sheet.

Leave the Answer sheet Number box on the scantron sheet blank

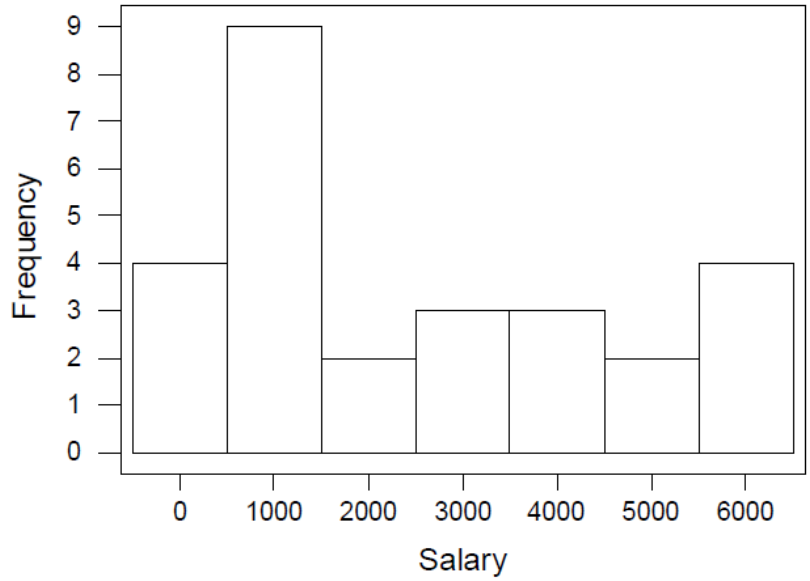
Carefully code your answers with pencil on the Scantron sheet. Also make sure the bubbles for your student ID, section, and exam code are filled out carefully and correctly.

NO EXTRA TIME WILL BE GIVEN TO CODE YOUR ANSWERS!!

GOOD LUCK!

Note:
The formula page, and Table A provided are a separate attachment.

(1) You are given the following histogram. How would you best describe the distribution?



(A) The distribution is skewed to the left

(B) The distribution is skewed to the right

(C) The distribution is symmetric

(D) There are outliers

(E) The distribution is skewed both left and right

(2) In an auction of 24 Wayne Gretzky rookie cards, the prices of the cards are given in the following Stemplot. The leaf's unit is 0.01dollar (e.g. 65 | 0 is $6.50) The prices are in Canadian dollars.

65 | 0 0
66 | 0 5 5 8 9
67 |
68 | 7 9 9
69 | 7 9
70 | 5 5 8 9 9
71 | 0 0 0 9
72 | 3 6
73 | 8

Calculate the median selling price of the Gretzky rookie cards.

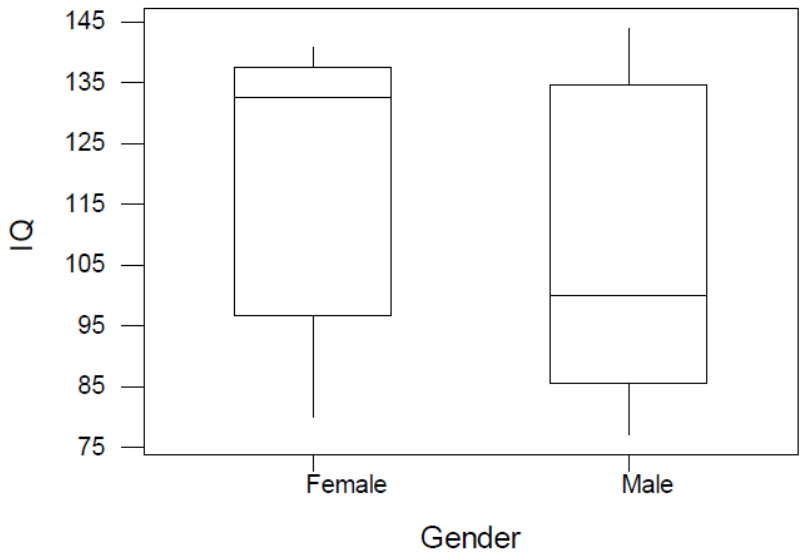
(A) $ 6.99 (B) $70.2 (C) $7.02 (D) $7.05

(3) You are given the following sample data: -2 , -2 , -1, 0, 1, 4

Calculate the sample standard deviation.

(A) 2.08 (B) 2.28 (C) 4.33 (D) 5.2

You are given the following side-by-side boxplots that compare IQ and gender. Please answer the next <u>two questions</u> using the boxplots.



(4) What is the difference in means between males and females?

(A) 36 (B) 42 (C) 50 (D) it cannot be determined with the given graphs

(5) Which of the following statements is/are <u>False</u>?

(i) According to the boxplots the IQR for women is lower than it is for men.

(ii) According to the boxplots women have a higher median IQ than men.

(iii) There is no way to determine which gender has the higher median IQ.

(iv) The distributions of IQ are skewed for both men and women.

(A) (i) (B) (ii) (C) (iii) (D) (iv) (E) none of them are false

(6) A study of 6 heart patients was made. Here are their ages in years.

31 49 37 53 51 43

What is the first quartile?

(A) 37 (B) 51 (C) 46 (D) 52

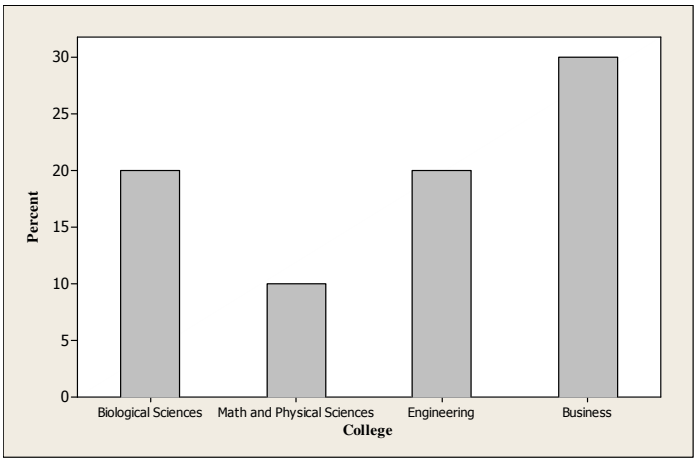
(7) Consider the following variables:

(1) Women's favorite TV program
(2) Salaries of football players
(3) Number of pets owned by families
(4) Total claims paid by 15 auto insurance companies in 2002
(5) Colour of cars
(6) Marital status of people
(7) Length of frog's jump
(8) Spring break locations favored by college students
(9) Ethnic background

The categorical variables in this list are given by:

(A) (1), (3), (4), (8), and (9).
(B) (2), (4), and (7).
(C) (1), (5), (6), (8), and (9).
(D) (2), (5), and (6).

(8) A large university is divided into six colleges, with most students graduating from four of these colleges. The following bar chart gives the distribution of the percent graduating from the four most popular colleges in 2003.



Which of the following is a correct statement?

(A) The ordering of the colleges is wrong.
(B) The bar graph is skewed to the right.
(C) The bar graph is skewed to the left.
(D) It would be correct to make a pie chart for this data if you added an "Other" category
(E) None of the above statements are true

(9) The actual ages of the trucks (in months) to be sold at an auction are as follows :

20 19 50 18 45 20 19 47 19 46 48 45 19 19

Which one of the following is the 5-number summary for this data?

(A) 18 19 20 46 50

(B) 18 19 20.5 46.5 50

(C) 18 19 20 46.5 50

(D) 18.5 19 20 46.5 50

(E) 18 19 20 47 50

(10) In drawing a histogram, which one of the following suggestion(s) should be followed?

(A) Leave large gaps between bars for clarity.

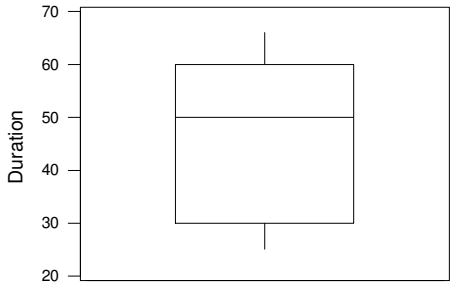
(B) The heights of bars should equal the class frequency or the relative class frequency.

(C) Generally, bars should be square so that both the height and width equal the class count.

(D) The x-axis can be marked either by the frequency or classes.

(E) More than one of the above suggestions should be followed.

(11) The boxplot below describes the duration time in hours for a particular application of an appliance bulb. All 100 data values have been rounded to the nearest hour for the analysis.



At least how far above the median would a rounded data value have to be in order for it to be considered an outlier? (Use the convention for deciding outliers adopted in this course.)

(A) 62 (B) 46 (C) 50 (D) 53 (E) 56

(12) In the 2006 Olympics, a French skier skied the slalom race in 88.45 seconds and this time was about one standard deviation faster than the mean. Assuming the race times are normally distributed, about how many of the 34 skiers finishing the event would you expect had a faster race time than the French skier?

(A) 5 (B) 2 (C) 10 (D) 29 (E) can't be determined

(13) Which of the following is a correct statement about Normal Distributions?

(A) The standard normal distribution only takes on values between -3 and +3.

(B) The width of central 68% of the normal distribution is one standard deviation.

(C) The standard normal random variable has no units.

(D) The standard deviations of all normal distributions are the same.

(E) None of the above statements are correct

(14) The marks on a statistics test are normally distributed with a mean of 62 and a standard deviation of 15. The instructor assigns a grade of B+ or higher to the top 30% of the students in the class. What is the minimum mark required to get a grade of B+ or higher?

(A) 68.7 (B) 71.5 (C) 73.2 (D) 74.6 (E) 69.8

(15) The lengths x of nails in a large shipment received by a carpenter are approximately normally distributed with mean 2 inches and standard deviation 0.1 inches. The carpenter can only use nails that are between 1.75 inches and 2.25 inches in length. What proportion of the shipment of nails will the carpenter reject?

(A) 0.9876 (B) 0.9938 (C) 0.9500 (D) 0.0062 (E) 0.0124

(16) Let z be a standard normal random variable. What proportion of its values fall between $z = -1.0$ and $z = 0.5$?

(A) 0.6826 (B) 0.5328 (C) 0.6915 (D) 0.4672 (E) 0.3085

(17) Which of the cases below has the highest relative score?

(A) A score of 2.7 on a test for which the mean is 3.2 and the standard deviation is 1.1

(B) A score of 37 on a test for which the mean is 35 and the standard deviation is 12

(C) A score of 1080 on a test for which the mean is 921 and the standard deviation is 87

(D) A score of 65 on a test for which the mean is 37 and the standard deviation is 15

(18) You are given that SAT scores are normally distributed with $\mu=500$ and $\sigma=100$. *Using the 68-95-99.7% rule*, what proportion of SAT scores falls between 600 and 700?

(A) 0.135 (B) 0.270 (C) 0.160 (D) 0.340 (E) 0.115

(19) You are given that $\bar{x} = 4.14$, $s_x = 2.14$, $\bar{y} = 2.73$, $s_y = 0.43$ and the least squares regression line is $\hat{y} = 1.9 - 0.2x$. Based on this information, what is the correlation between *x* and *y*?

(A) 0.995 (B) -0.995 (C) -0.07 (D) 0.04 (E) -0.04

(20) Which of the following is a correct statement about scatterplots?

(A) A scatterplot would be a good tool to use in examining the color of eyes of statistics students by gender.

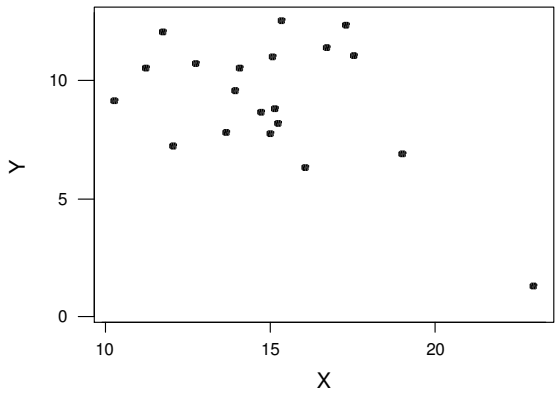
(B) Strength of association and outliers provide a complete description of the overall pattern in a scatterplot.

(C) As interest rates decrease, the prices of preferred stocks rise. A scatterplot would show that these two variables are positively associated.

(D) A scatterplot that has a curved pattern with all of the points following the curved pattern very closely is said to be weak because we have a preference for a linear relationship.

(E) None of the above statements are correct.

(21) The correlation *r* between the variables X and Y shown in the scatterplot is – 0.48.



If the outlier, the point at the bottom right of the scatterplot, is removed, the correlation will

(A) remain the same.

(B) change, but the direction (increase or decrease) of the change is uncertain.

(C) increase towards 0.

(D) decrease towards –1.

(22) With regards to correlation, which of the following statements is <u>false</u>?

(i) The correlation must lie between -1 and +1.

(ii) When there is a clear negative association, the correlation will be less than zero.

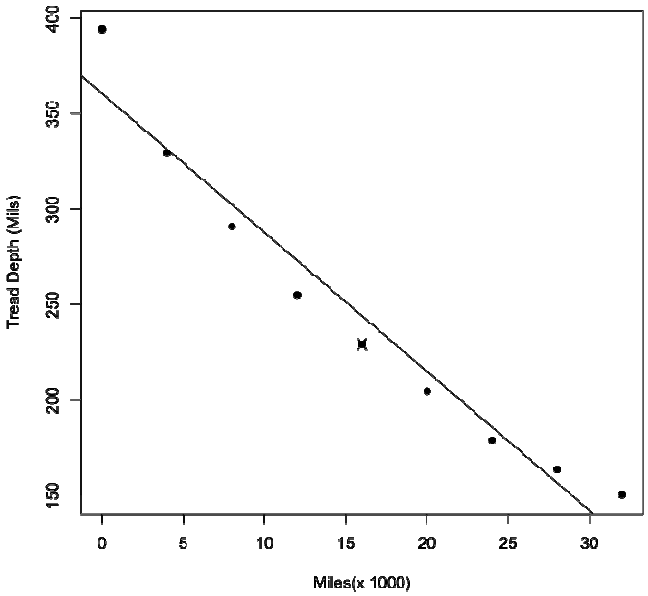
(iii) The correlation does not change when we switch explanatory and response variables.

(iv) Correlation is a resistant measure.

(A) (i) (B) (ii) (C) (iii) (D) (iv) (E) none of them are false

Use the following information to answer the next <u>three questions</u>

In a study, nine tires of a particular brand were driven on a track under identical conditions. Each tire was driven a particular controlled distance (measured in thousands of miles), and afterward the tread depth was measured. Tread depth is measured in "mils." The least-squares regression line was computed, and added to a scatterplot of these data.



The equation of the least-squares regression line is: Tread Depth = 360.64 – 11.39x (thousands of miles) .
Also, $r^2 = 0.953$.

(23) The least-squares line would predict that the tread depth of a tire driven 16 thousand miles is

(A) 260.5 mils. (B) 178.4 mils. (C) 360.64 mils. (D) –181,879 mils.

(24) Which of the following statements is <u>true</u>?

(A) About 95.3% of the variation in tread depth is explained by the regression on miles.

(B) According to the least-squares regression line, the tread depth of a new tire (driven 0 miles) is predicted to be 360.64 mils.

(C) According to the least-squares regression line, we would predict a decrease in groove depth of 11.39 mils for each 1000 miles driven on a tire.

(D) All of the above statements are true.

(25) Which of the following is <u>true</u>?

(A) The correlation between tread Depth and miles is –0.976.

(B) 95.3% of the data points fall on the least-squares regression line.

(C) The correlation between tread depth and miles is 0.976.

(D) At lease one of observations must be highly influential.

Use the following information to answer the next <u>two questions</u>.

Applicants looking for a job at a restaurant chain may apply to be a server or kitchen worker. The table below summarizes the numbers of male and female applicants hired for the jobs they applied for.

	Server			Kitchen worker	
	Male	Female		Male	Female
Not hired	80	120	Not hired	30	15
Hired	20	50	Hired	80	25

(26) The proportion of female applicants for a job as kitchen worker that were hired is

(A) 0.100 (B) 0.375 (C) 0.400 (D) 0.357 (E) 0.625

(27) From these data we may conclude

(A) Men are advantaged in applying for kitchen jobs, while women are advantaged in applying to be servers.

(B) Overall, the proportion of male applicants hired is greater than that for women overall.

(C) The restaurant hires more kitchen workers than servers.

(D) All of the above statements are true

(28) In a recent round of lay-offs in a company, the percent of employees 50 and over who were laid off was much higher that the percent under 50 who were laid off. However, when the data were analyzed separately in each job category, the percent of employees 50 and over who were laid off was lower than the percent of employees under 50 laid off in each job category. This reversal of direction of the association between age and being laid off, when job category is taken into account, is called

(A) Simpson's paradox

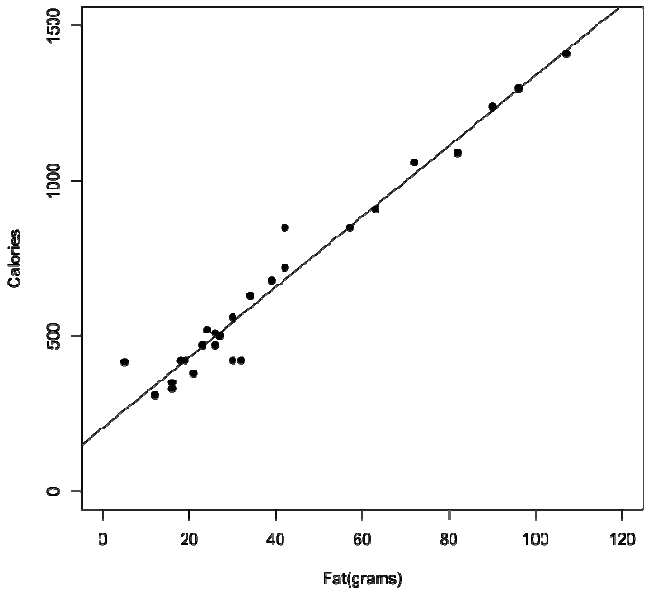
(B) Least-squares regression

(C) Negative association

(D) A residual plot

Use the following information to answer the next <u>two questions</u>

In many fast food restaurants, there's a strong correlation between a menu item's fat content (measured in grams) and its calorie content. We want to investigate this relationship. Using all of the food menu items at a well-known fast food restaurant, the fat content and calorie contents were measured. We decide to fit the least-squares regression line to the data, with fat content (x) as the explanatory variable and (y) as the response variable. A scatterplot of the data (with regression line included), and a summary of the data are provided. One of the menu items is a hamburger with 107 grams of fat and 1410 calories.



The following information was also calculated from the sample:
$r = 0.979$ $\bar{x} = 40.35$ grams $\bar{y} = 662.88$ calories $s_x = 27.99$ grams $s_y = 324.90$ calories

(29) The intercept of the least-squares regression line is;

(A) 204.50 (B) 662.88 (C) 150.38 (D) 201.49 (E) 243.54

(30) Refer to the example data point (107 grams, 1410 calories). What, approximately, is the residual corresponding to this observation?

(A) 10 calories (B) 10 grams (C) 0 calories (D) –10 grams (E) -10 calories

(31) A high school marketing class creates two videos advertising an expensive BMW sport car. They test the videos by asking 200 fellow students to view both videos (in random order), and indicate which video makes them more likely to buy the BMW sports car. BMW should be reluctant to agree that the video favoured in this study will sell more of their sports car because

(A) the study used a matched pairs design instead of a completely randomized design
(B) results from students may not be the same as from older and richer customers who might buy a higher end sports car
(C) not enough students were asked
(D) there was no control group
(E) male responses should have been considered only

(32) You are testing a new medication for the relief of depression. You are going to give the new medication to subjects suffering from depression, and after a month see if their symptoms have lessened. You have sixteen subjects available (numbered 1 to 16 for simplicity). Four of the subjects are to be given the new medication and four of the others, a placebo. Selections are made using a random number table beginning with line:

81507 27102 56027 55892 33063 41842 81868 71035 09001 43367 49497

The first four subjects randomly selected will receive the new medication. These are subjects numbered

(A) 1, 2,6 and 10.
(B) 2, 3, 8 and 9.
(C) 1, 2, 6 and 7.
(D) 2, 6, 9 and 10.
(E) 8,15,7 and 2.

(33) A recreational boating retailer buys several different types of kayak paddles from a manufacturer in China. Recently, some customers have complained that the paddles have broken and the boating retailer is worried that this may be due to manufacturing defects. To assess the quality of the manufacturer's production, the retailer sends a sample of 8 paddles from the last shipment to a laboratory for study. The Kayak paddles vary tremendously in price, from under $30 to over $300. In order to save money on the study, the retailer chose paddles costing under $100 for the samples sent to the laboratory. This is an example of:

(A) Nonresponse.
(B) Undercoverage.
(C) A simple random sample.
(D) A stratified random sample.
(E) multistage sampling

(34) Suppose the volunteers were first divided into groups of men and women. Then each group of men or women was divided into two groups according to his or her age (40 and over, or less than 40). Therefore, each volunteers will be belong to one of 4 groups, and then for each group, half of them were randomly assigned to the new drug and the remaining people in the group received the placebo. This would be an example of

(A) a block design.
(B) stratified sampling.
(C) confounding. The effects of gender will be mixed up with the effects of the drugs.
(D) systematic sampling.
(E) a matched pairs design

(35) In an experiment with 4 treatments and 100 male and 100 female subjects, which of the following would be a <u>block design</u>?

(A) Use a table of random digits to randomly assign each subject to one of the 4 treatments.
(B) Use a table of random digits to randomly assign 50 subjects to each of the 4 treatments.
(C) Systematically assign the first 50 subjects to the first treatment, the second 50 subjects to the second treatment, the third 50 subjects to the third treatment, and the fourth 50 subjects to the fourth treatment.
(D) Use a table of random digits to randomly assign 25 men and 25 women to each of the 4 treatments.

<u>Answer Key:</u>

1	6	10	16	21	26	31
BCBDC	ACDAB	EACEE	BDABE	CDBDA	EDAAE	BDBAD