THE UNIVERSITY OF WESTERN ONTARIO
DEPARTMENT OF STATISTICAL AND ACTUARIAL SCIENCES
STATISTICS 1024B TEST 1

Friday, February 5th, 2010, 6:00 - 8:00 PM

EXAM CODE 970

INSTRUCTIONS:

- This is a closed book test. Table A from the text and a formula sheet are attached.
- There are 30 multiple choice questions to be answered using the provided Scantron sheet.
- Only non-programmable calculators are permitted.
- Use only an HB pencil for the Scantron sheet.
- Print your name, instructor and course (SS 1024B) on your Scantron sheet, and sign it.
- Fill in STUDENT NUMBER and SECTION on the Scantron sheet.
- Enter 970 as your EXAM CODE on the Scantron sheet.
- Leave the ANSWER SHEET NUMBER blank on the Scantron sheet.
- Code your answers on the Scantron sheet and submit it. You may keep the question sheet.
- **NO EXTRA TIME WILL BE GIVEN TO CODE YOUR ANSWERS!!**

GOOD LUCK!

1) The lengths of needles from white pine trees follow approximately the normal distribution with mean 8.3 cm and standard deviation 1.5 cm. According to the 68-95-99.7 rule, what range of lengths covers the centre 95% of white pine needles?

(A) 5.3 to 11.3 cm

(B) 6.8 to 9.8 cm

(C) 7.9 to 9.4 cm

(D) None of the above

2) A locomotive's "adhesion" is the locomotive's pulling force as a multiple of its weight. This is an important performance measure of a locomotive. A diesel locomotive model has adhesion which varies in actual use according to a Normal distribution with mean 0.37 and standard deviation 0.04. The first quartile for the adhesion distribution is

(A) 0.29 (B) 0.04 (C) 0.25 (D) 0.34

3) Using the standard normal distribution tables, the area under the standard Normal curve corresponding to $-0.5 < Z < 1.2$ is

(A) 0.5764 (B) 0.3085 (C) 0.2815 (D) 0.8849

4) A company produces boxes of laundry detergent labeled "Giant Size 6 Kg." The actual weight of detergent in such a box has a Normal distribution with a mean of 6.05 Kg and a standard deviation of 0.05 Kg. To avoid having dissatisfied customers, the company says a box of detergent is considered underweight if it weighs less than 6 Kg. To avoid losing money, it labels the top 5% (the heaviest 5%) overweight. How heavy does a box have to be for it to be labeled overweight?

(A) 6.01 Kg (B) 6.15 Kg (C) 6.13 Kg (D) 5.97 Kg

5) The time taken for a computer link to be made between the terminal in an executive's office and the computer at a remote factory site has a Normal distribution with a mean of 15 seconds and a standard deviation of 3 seconds. On 90% of the occasions, the computer link is made in less than
(A) 19.39 seconds (B) 15.95 seconds (C) 18.84 seconds (D) 11.16 seconds

6) Which of the following is **not** true of the correlation coefficient r?

(A) If $r = 0$, then there is no relationship between x and y.

(B) Multiplying all data values (x's and y's) by 10 will have no impact on r.

(C) If r is the correlation between x and y, then r is also the correlation between y and x.

(D) $-1 \leq r \leq 1$.

7) When water flows across farm land, some soil is washed away, resulting in erosion. An experiment was conducted to investigate the effect of the rate of water flow (liters per second) on the amount of soil (kilograms) washed away. The data are given in the following table.

Flow rate	.31	.85	1.26	2.47	3.75
Eroded soil	.82	1.95	2.18	3.01	6.07

The association between flow rate and amount of eroded soil is

(A) positive.

(B) negative.

(C) neither positive nor negative.

(D) impossible to determine since both variables are categorical.

8) What is the most important reason that students shop online? The answer may differ by gender of students. The following table summarizes the reasons for shopping online from a sample of university students:

Reason	Male	Female
Save time	29	20
Easy	28	31
More selection	16	34
Low price	11	4
Live far from stores	10	7

What percentage of students shop online because there is more selection?

(A) 26 %

(B) 36 %

(C) 17 %

(D) None of the above

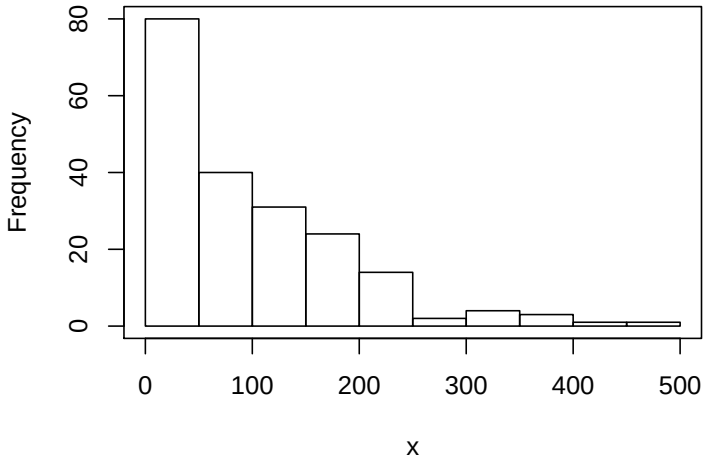
9) Rainwater was collected in water collectors at thirty different sites near an industrial basin and the amount of acidity (pH level) was measured. The following stem-and-leaf display shows the pH values with the leaf unit equal to 0.1.

```
Stems   Leaves
   2       679
   3       237789
   4       1222446899
   5       0556788
   6       0233
```

The acidity amount associated to the second quartile is:

(A) 4.2 (B) 4.5 (C) 4.6 (D) 4.4

10) Look at the following histogram of 200 observations and choose the correct answer.



(A) The distribution is skewed to the left.

(B) The location of the median is in the second bar from the left in the histogram.

(C) The distribution is bimodal with about half in each group.

(D) None of the above.

11) Which of the following statements is usually not true about a Stemplot?

(A) It separates each observation into a stem and a leaf.

(B) It is good for small data sets.

(C) Each leaf contains only a single digit.

(D) It is a graphical display of the distribution for a categorical variable.

12) Consider the following variables which pertain to an observational study on diabetes:

(I)	BTYPE:	blood type (A, B, AB, or O)
(II)	WEIGHT:	subject's weight (kg)
(III)	HYPOG:	hypoglycemia episode experienced by subject (Y or N)
(IV)	DTYPE:	type of diabetes (Type I or Type II)
(V)	BSUGAR:	blood sugar level (mg/dl)
(VI)	VISITL:	length of visit (days)
(VII)	QNUM:	number of questions subject asked physician
(VIII)	EXCER:	number of hours spent on exercise per week
(IX)	IDNUMER:	patient's ID number

The quantitative variables in this list are given by:

(A) (II), (V), (VI), (VII), (VIII).

(B) (II), (V), (VIII) only.

(C) all but (III), (IV), (IX).

(D) None of the above.

13) The stem and leaf plot below summarizes the final year averages of a graduating honors class in the Faculty of Science. Select the correct statement.

6	88
7	24567
8	2346
9	014
10	

(A) Q3 equals 86.5

(B) The range is 23

(C) Q1 equals 73.5

(D) The median is 79.5

14) The linear correlation coefficient between grade point average (GPA) and IQ test score for 78 students in the 7th grade is computed as 0.6337. The summary for the rest of the data set is as follows (we use x_i for the IQ and y_i for the GPA):

$$\sum x_i = 8494.2, \sum y_i = 580.866, s_x = 13.7, s_y = 2.10$$

The equation for the least-squares line for predicting GPA from IQ is

(A) $\hat{y} = 0.0971 - 3.127x$

(B) not determined by the given information.

(C) $\hat{y} = -3.127 + 0.0971x$

(D) $\hat{y} = -3.127 + 4.134x$

15) Consider the 10 data points:

3.06, 4.28, 2.75, 6.39, 4.49, 2.77, 4.73, 5.11, 4.86, 3.54

The upper quartile of this dataset is

(A) 4.86 (B) 4.385 (C) 1.8 (D) 5.11

16) A study of sales of a particular model of car included a wide range of ages of cars. A regression line to model the relationship between the age (in years) and price (in dollars) resulted in the equation: $\hat{y} = 19547 - 2026x$. We plan to buy a car which is 3 years older than the average age of car in this study. What is the expected price compared to the average price?

(A) It cannot be determined.

(B) Lower by \$6078.

(C) You will pay \$13422.

(D) Higher by \$6078.

17) X and Y are two categorical variables. The best way to determine if there is a relationship between these variables is to:

(A) draw a scatterplot of the X and Y values.

(B) fit a regression line to the data for prediction purposes.

(C) make a two-way table of the X and Y values.

(D) calculate the correlation between X and Y.

18) Find the mean of the 5 observations:

13.1, 14.1, 10.4, 12.2, 12.2

(A) 10.4 (B) 12.4 (C) 12.25 (D) 12.2

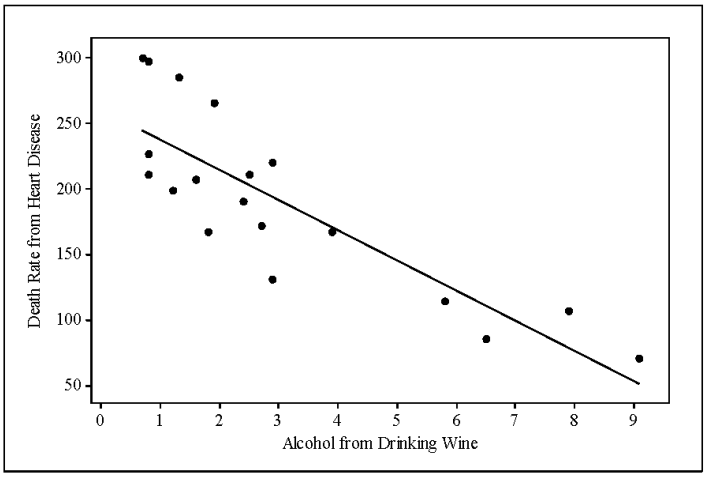
19) Approximately what percent of the observations in a distribution lie between the minimum and the third quartile?

(A) 100% (B) 50% (C) 25% (D) 75%

20) A study measured the heights in meters of 13 rose bushes. The mean was 1.43 and the standard deviation was 0.31. The correct units for the standard deviation are:

(A) meters (B) square meters (C) no units (D) inverse meters

21) The following is a scatterplot of the liters of alcohol from drinking wine per person and the death rates from heart disease per 100,000 people for each of 19 countries. The least-squares regression line has been drawn in on the plot.



Based on the least-squares regression line we would predict that in a country where, per person, 7 liters of alcohol from wine is consumed, the death rate from heart disease per 100,000 people would be about

(A) 260

(B) 50

(C) can not be predicted since causality has not been established

(D) 100

22) The penetrability of soil was measured at five locations, giving values

3.10, 2.86, 2.68, 3.32, 3.04

The mean of this sample is 3.00. What is the standard deviation?

(A) 0.059 (B) 0.243 (C) 0.217 (D) 0.053

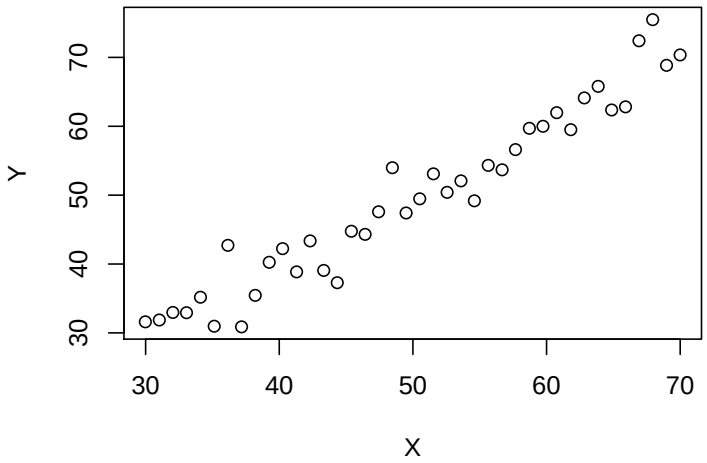
23) In a study of blood types, 300 people were selected at random from a population. Their blood type (O, A, B, or AB) was determined, as well as their Rh status (positive or negative). The results are shown in the following table:

	O	A	B	AB
Rh positive	82	89	54	19
Rh negative	13	27	7	9

In the marginal distribution for blood type, what is the percentage of individuals with type O blood?

(A) 86 % (B) 32 % (C) 27 % (D) 34 %

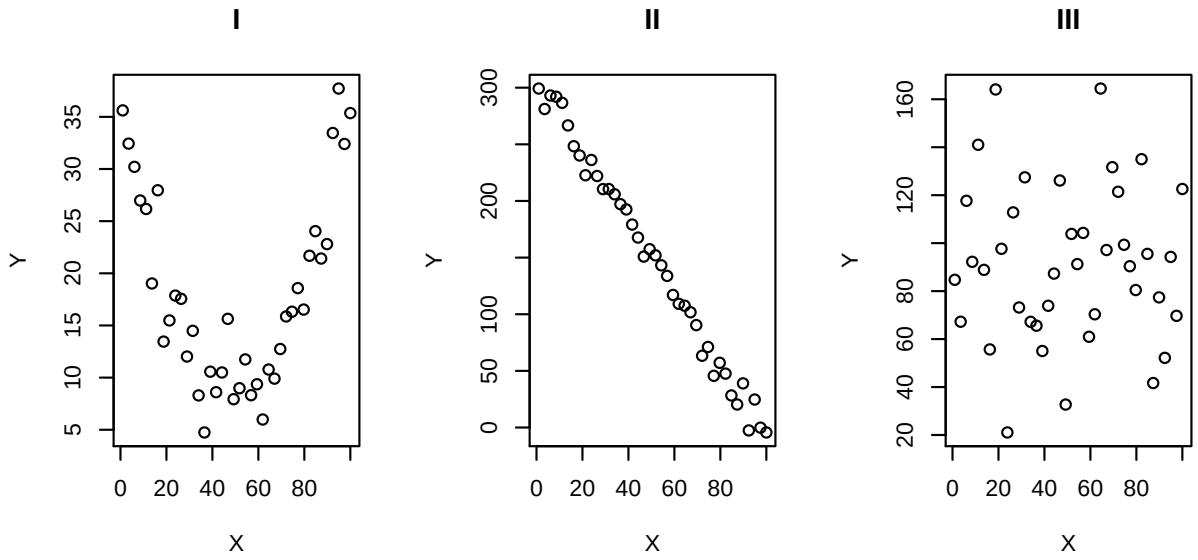
24) The plot below shows the results of 40 measurements of standard weights using two scales. On the X axis the measurements are shown using a professional digital electronic scale, and on the Y axis the same items were weighed on a standard bathroom scale. All measurements are in Kg.



The correlation between the two measurements is approximately

(A) 0.97 (B) -0.97 (C) 0.09 (D) -0.09

25) Consider the three plots below.

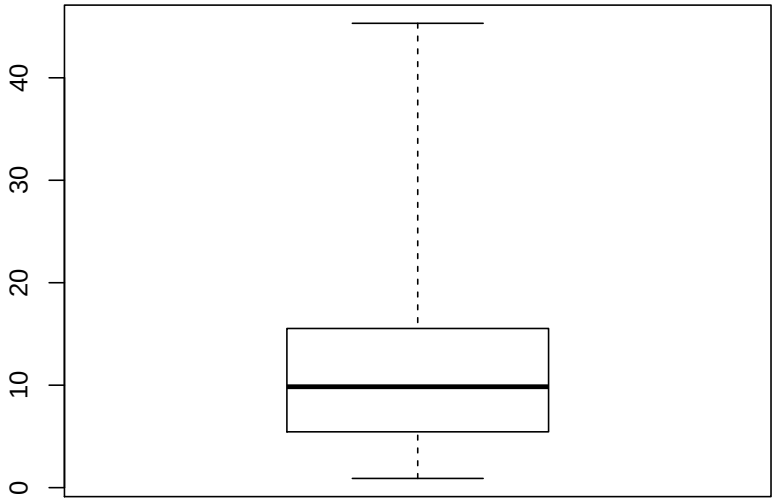


Which of these scatterplots **do not** show a strong relationship between X and Y?
(A) II only (B) III only (C) I and III (D) I only

26) The interquartile range (IQR) for the standard normal distribution is

(A) -0.67
(B) cannot be determined with the information given
(C) 1
(D) 1.35

27) The amount of money (in dollars) each student had in their pocket during a class was found and the distribution is summarized in the boxplot shown below.



From this boxplot we can say

(A) the first and second quartile are both larger than the median

(B) the distribution is skewed towards low values

(C) the distribution is skewed towards both high values and low values

(D) the distribution is skewed towards high values

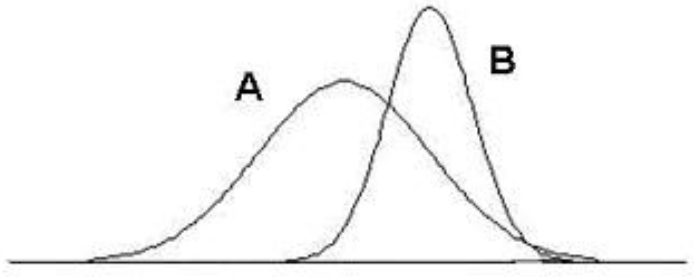
28) The table below summarizes the soft drink preferences, by age group, obtained from a recent survey taken at a London Knights Hockey game:

	Cola	Root Beer	Lemon Lime
Under 21 years of age	40	25	20
Between 21 and 40	35	20	30
Over 40 years of age	20	30	35

Of those surveyed, what proportion of people were under 21 years of age and drank root beer?

(A) 0.098 (B) 0.333 (C) 0.294 (D) None of the above

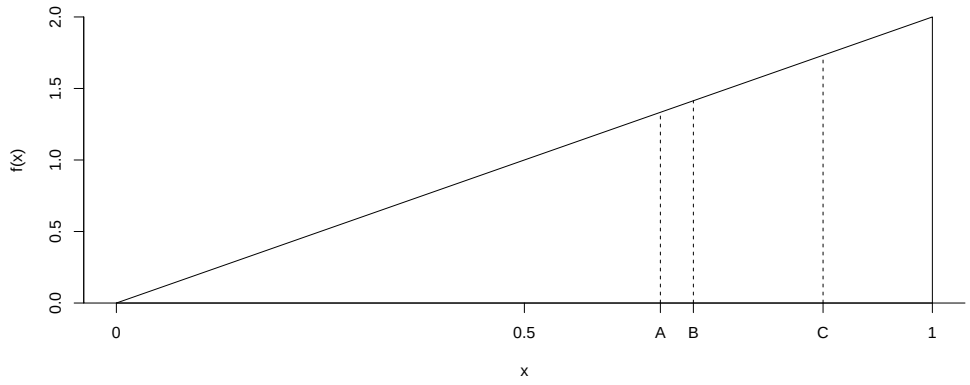
29) Two normal density curves are given in the graph below.



Suppose that curve A represents the density curve of a normal distribution with mean μ_A and standard deviation σ_A. Let μ_B and σ_B be the mean and standard deviation, respectively, from density curve B. Which of the following statements is true?

(A) μ_A is less than μ_B and σ_A is less than σ_B.
(B) μ_A is less than μ_B and σ_A is greater than σ_B.
(C) μ_A is greater than μ_B and σ_A is greater than σ_B.
(D) μ_A is greater than μ_B and σ_A is less than σ_B.

30) The plot of a probability density function is shown below.



The points marked at A, B and C could represent respectively:

(A) the mean, the median and Q_3
(B) Q_1, the median and the mean
(C) the median, the mean and Q_3
(D) Q_1, the median and Q_3

Use this page for rough work.

FORMULA SHEET

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

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

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

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

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

Answer key for version 970:

1	6	11	16	21	26
ADACC	AAABB	DADCA	BCBDA	DBBAB	DDABA