

Name: _____
Student No.: _____

University of Alberta

Department of Mathematical and Statistical Sciences

Statistics 151 Midterm Examination – Lecture B01 Version A

Date: October 29, 2010

Instructor: Alireza Simchi

Time: 9:00-9:50

Instructions: (READ ALL INSTRUCTIONS CAREFULLY.)

1. This is a closed book exam. You are permitted to use a non-programmable calculator. Please turn off your cellular phones or pagers.
2. The exam has 20 multiple-choice questions. For each multiple-choice question choose the answer that is closest to being correct. Circle one of the letters (a)-(j) corresponding to your chosen answer for each question. All answers will be graded right or wrong (no partial credit) in this part. Each single question is worth 1 point. All numerical answers are rounded.
3. This exam has **4** pages including this cover. Please ensure that you have all pages and write your name and your student ID at the top of each page.
4. The table of the standard normal probabilities and formula sheet are provided in a separate booklet.
5. The exam is graded out of a total of **20** points.

1. In 2002, a report in the *Journal of the American Cancer Institute* indicated that women who work nights have a 60% greater risk of developing breast cancer. Researchers based these findings on the work histories of 763 women with breast cancer and 741 women without the disease. Choose the best option.
 - (a) The conclusion is valid, because it was published in a well known journal.
 - (b) The conclusion is valid, because of large sample sizes.
 - (c) The conclusion is not valid, since the study is too old.
 - (d) The conclusion is not valid, because of possible confounding (or lurking) variables,
 - (e) More information is necessary to decide about the validity of the conclusion.

2. After menopause, some women take supplemental estrogen. There is some concern that if these women also drink alcohol, their estrogen levels will rise too high. Twelve volunteers who were receiving supplemental estrogen were randomly divided into two groups, as were 12 other volunteers not on estrogen. In each case, one group drank an alcoholic beverage, the other a nonalcoholic beverage. An hour after, everyone's estrogen level was checked. Only those on supplemental estrogen who drank alcohol showed a marked increase. Choose the best option.
 - (a) The study is an experiment,
 - (b) The study is an observational study,
 - (c) The study is neither an observational study nor experiment,
 - (d) The data obtained is not sufficient to evaluate the effect of alcohol on the estrogen level.
 - (e) None of the above can be concluded based on the information provided.

3. Refer to Question 2; choose the best option?
 - (a) The results of study can be extended to the population,
 - (b) The result of the study is only valid for twelve volunteers,
 - (c) The result of the study is not valid, since the sample size is small,
 - (d) Twelve samples are enough to decide about the validity of the result.
 - (e) More information is necessary to decide about the validity of the result.

4. The winning long-jump and high-jump distances, in inches, for the summer Olympics from 1912 through 2004 were recorded. The sample correlation between long-jump and high-jump distances was 0.925. What is the sample correlation if data are converted to meters? Use 1 inch = 0.0254 meter.
 - (a) 0.023
 - (b) 0.027
 - (c) 0.5276
 - (d) 0.925
 - (e) 36.42

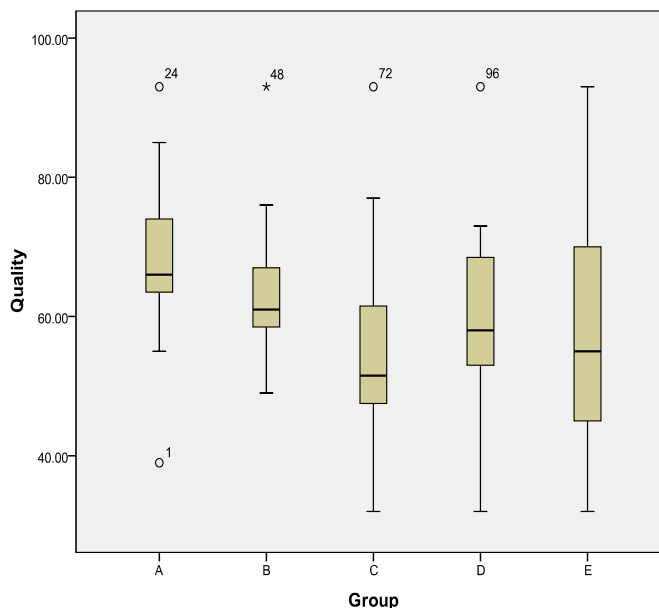
5. A survey was conducted to determine how people rated the quality of programming available on television. Respondents were asked to rate the overall quality from 0 (no quality at all) to 100 (extremely good quality). The stem-and-leaf display of the data is shown below.

3		26
4		03478999
5		0122345
6		1256
7		17
8		
9		3

What measures of centre and variation (or spread) are most appropriate?
 - (a) Mean and Standard Deviation
 - (b) Mean and IQR
 - (c) Median and Standard Deviation
 - (d) Median and IQR
 - (e) Mode and Range

6. Refer to the data in Question 5; what is the median of this data?
 - (a) 50.0
 - (b) 50.5
 - (c) 51.0
 - (d) 51.5
 - (e) 52.0

7. Refer to the data in Question 5; what is the inter-quartile of the data?
 (a) 13.0 (b) 13.5 (c) 14.0 (d) 14.5 (e) 15.0
8. Refer to Question 5; which of the five box-plots labelled A, B, C, D and E displayed below is the box-plot of the data:



- (a) Box-plot A (b) Box-plot B (c) Box-plot C (d) box-plot D (e) Box-plot E
9. Seven students in a graduate program were randomly selected. Their grade point averages (GPA) when they entered the program (X) and their current GPA (Y) are given in the following table..

x	3.5	3.6	3.6	3.7	3.8	3.9	3.9
y	3.6	3.7	3.8	3.7	3.7	3.8	3.9

Find the equation of the line of regression. What is the slope of the regression line of Y on X?

- (a) 0.04 (b) 0.14 (c) 0.24 (d) 0.34 (e) 0.44
 (f) 1.05 (g) 1.10 (h) 2.05 (i) 2.10 (j) 120.56
10. Refer to question 9; a new graduate student with entrance GPA of 3.8 was selected. According to the regression model, her residual for predicting her current GPA is +0.02. What is her current GPA?
- (a) 3.58 (b) 3.60 (c) 3.65 (d) 3.70 (e) 3.75
 (f) 3.80 (g) 3.85 (h) 3.90 (i) 3.95 (j) 4.00
11. It is difficult to determine a person's body fat percentage (Y) accurately without immersing him or her in water. Researchers hoping to find ways to make a good estimate immersed 20 male subjects, then measured their waists, and recorded their weights (X). Summary statistics for X and Y are given in the following.

$$\bar{x} = 188.6, \quad s_x = 26.66, \quad \bar{y} = 19.75, \quad s_y = 9.56, \quad r = 0.697$$

What is the y-intercept of the regression line of Y on X?

- (a) 1.39 (b) 7.39 (c) 17.39 (d) 27.39 (e) 346.84
 (f) -1.39 (g) -7.39 (h) -17.39 (i) -27.39 (j) -346.84
12. Refer to question 11; what percentage of variation in Y is explained by the regression model of Y on X?
- (a) 8% (b) 10% (c) 28% (d) 30% (e) 48%
 (f) 50% (g) 68% (h) 70% (i) 88% (j) 90%

13. Employment data at a large company reveal that 56% of the workers are married, 44% are college graduates, and 20% are college graduates and married. What is probability that a randomly selected worker is neither married nor a college graduate?

(a) 0.0 (b) 0.1 (c) 0.2 (d) 0.3 (e) 0.4
(f) 0.5 (g) 0.6 (h) 0.7 (i) 0.8 (j) 1.0

14. Suppose Sarah is flying from Regina to Vancouver with a connection in Edmonton. The probability that her first flight leaves on time is 0.77. If the first flight is on time, the probability that her luggage will make the connection flight in Edmonton is 0.92. But if the first flight is delayed, the probability that the luggage will make it is only 0.63. What is probability that her luggage arrives in Vancouver with her?

(a) 0.0616 (b) 0.0851 (c) 0.1467 (d) 0.1627 (e) 0.1889
(f) 0.2300 (g) 0.2355 (h) 0.7084 (i) 0.7700 (j) 0.8533

15. Refer to question 14; suppose that her luggage arrived in Vancouver with her, what is the probability that the first flight was delayed?

(a) 0.07 (b) 0.17 (c) 0.27 (d) 0.37 (e) 0.47
(f) 0.57 (g) 0.67 (h) 0.77 (i) 0.87 (j) 0.97

16. The random variable X has the following probability distribution.

Value of X	-2	-1	0	1	2
Probability	0.1	0.2	0.2	0.4	?

What is the mean of X or $\mu_X = E(X)$?

(a) -0.4 (b) -0.3 (c) -0.2 (d) -0.1 (e) 0.0
(f) 0.4 (g) 0.3 (h) 0.2 (i) 0.1 (j) 1.0

17. The length of human pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days. What percent of pregnancies last less than 270 days?

(a) 33.00 (b) 36.69 (c) 38.97 (d) 40.13 (e) 42.86
(f) 57.14 (g) 59.87 (h) 61.03 (i) 63.31 (j) 67.00

18. Refer to question 17; how long do the longest 20% of pregnancies (approximately) last?

(a) 250 (b) 253 (c) 258 (d) 263 (e) 266
(f) 269 (g) 274 (h) 279 (i) 282 (j) 286

19. An insurance company found that in the entire population of homeowners, the mean loss from fire is $\mu = 250$ and the standard deviation of the loss is $\sigma = 1000$. The distribution of losses is strongly right-skewed; many policies have \$0 loss, but a few have large losses. If the company sells 10000 policies, what is the approximate probability that the average loss will less than 256.4?

(a) 0.0000 (b) 0.1314 (c) 0.2611 (d) 0.3632 (e) 0.4168
(f) 0.5832 (g) 0.6368 (h) 0.7389 (i) 0.8686 (j) 1.0000

20. Suppose 5% of the people who make airline reservations with Air Canada will not appear for the flight. For 500 randomly selected people with Air Canada reservations, what is the probability that less than 6% do not appear for their flight?

(a) 0.0307 (b) 0.1515 (c) 0.2514 (d) 0.3669 (e) 0.4960
(f) 0.5040 (g) 0.6331 (h) 0.7486 (i) 0.8485 (j) 0.9693