

STAT*2080 F15
Midterm 1
Version 1
Oct. 16, 2015

This first page is to satisfy the U of G's privacy policy. Please fill out your first and last names on this page, but NOT your ID # or any other information.

Last Name:

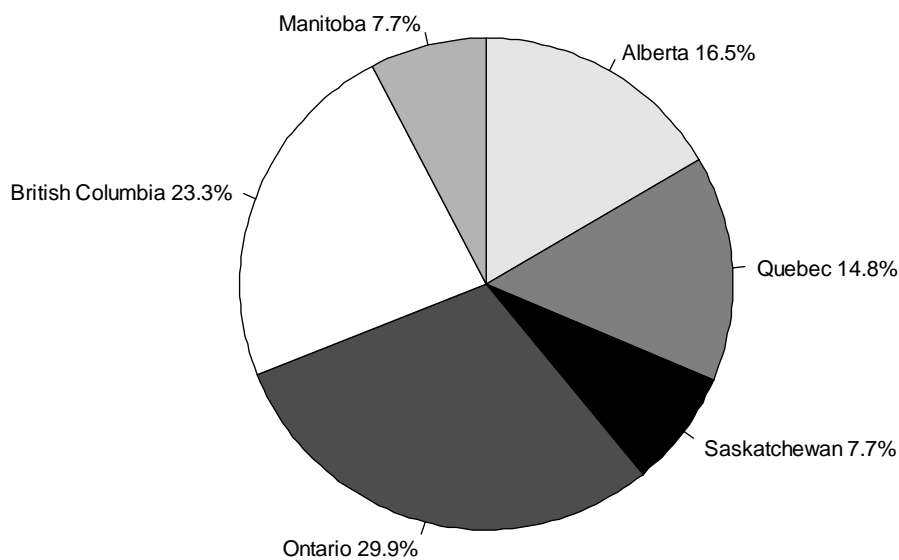
First Name:

On the next page, please fill out your name and ID number.

1. AC Nielsen is conducting a province-wide survey for Rogers in which several variables of interest are recorded. Which one of the following variables is categorical?
- (A) Number of people, both adults and children, living in the household.
 - (B) Most watched TV show.
 - (C) Hours of TV watched per day.
 - (D) Total household income, before taxes, in 2014.
 - (E) Number of televisions in the home.

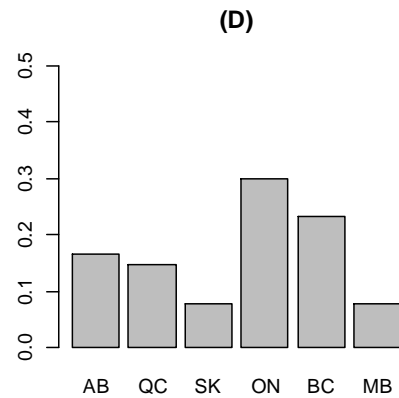
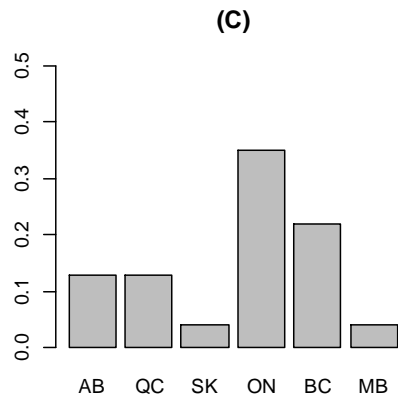
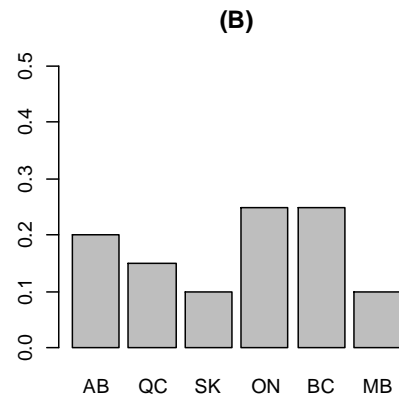
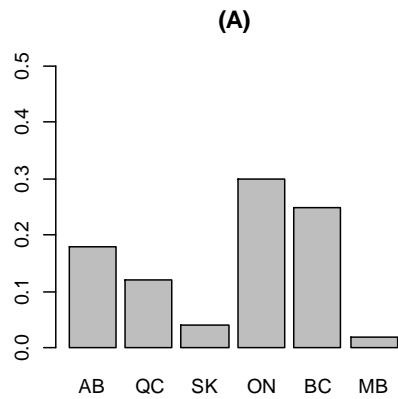
The next two questions refer to the following information. Statistics Canada tracks the number of movements made by airlines at airports across the country. The pie chart below provides information on the number of movements made at some of the major airports in each province, throughout July 2015:

Number of Movements by Airport



2. Note that there were 41,416 movements recorded at the airport in Ontario. How many movements were recorded in the Alberta airport? Choose the option closest to your answer.
- (A) 6,833
 - (B) 16,534
 - (C) 22,855
 - (D) 34,582
 - (E) Not enough information is provided to answer the question.

3. Which one of the following bar graphs represents the same data as the pie chart on the number of movements?



(E) None of the above.

4. Which one of the following statements is **TRUE**?
- (A) The interquartile range is a resistant measure of spread.
 - (B) Both the sample mean, \bar{x} , and sample variance, s^2 , are in the same units as the original observations used to compute them.
 - (C) The sample mean is a good measure of central tendency for symmetric and for skewed distributions.
 - (D) All of the above statements are true.
 - (E) None of the above statements is true.

The next two questions refer to the following information. The housing prices of 15 three-bedroom homes with two or more bathrooms currently for sale in Guelph are shown in the table below.

Price (\$K)	200	300	400	500	550	600	750
Count	2	1	2	2	4	3	1

5. What are the mean and median housing prices (\$K), respectively, for the 15 prices listed above?
- (A) 500 and 525
 - (B) 500 and 550
 - (C) 483.3 and 525
 - (D) 483.3 and 550
 - (E) None of the above.
6. Suppose a recording error occurred and one of the \$600K homes should have been \$950K. Using the data, with one 600 replaced by 950, the first and third quartiles are approximately \$400K and \$600K, respectively. How many suspected outliers are in the corrected data?
- (A) 0
 - (B) 1
 - (C) 2
 - (D) 3
 - (E) None of the above.
7. How many of the following statements regarding randomization are FALSE?
- (i) Randomly sampling units from the population ensures your sample does not contain any outliers.
 - (ii) Randomly assigning treatment in an experiment helps to average out the effects of other important factors on the response.
 - (iii) Randomization eliminates the need to conduct blind studies.
 - (iv) Statistics based on random samples always have low variance, regardless of sample size.
- (A) 0
 - (B) 1
 - (C) 2
 - (D) 3
 - (E) 4

The next two questions refer to the following information. The lifetime of a statistical calculator is known to approximately follow a normal distribution with $\mu = 1800$ hours and $\sigma = 120$ hours.

8. What proportion of calculators has a lifetime of at least 2050 hours?
- (A) 0.0188
 - (B) 0.2917
 - (C) 0.9812
 - (D) 0.0163
 - (E) None of the above.
9. What percentage of calculators will last between 1750 and 2050 hours? Choose the option closest to your answer.
- (A) 41.67
 - (B) 33.72
 - (C) 66.72
 - (D) 64.40
 - (E) 98.12
10. Suppose the length of a contemporary Hollywood blockbuster movie is well approximated by a normal distribution with $\mu = 130$ minutes and that the percentage of movies that are less than 95 minutes long is 0.8%. What is the approximate standard deviation for the length of a randomly selected contemporary Hollywood blockbuster? Choose the option closest to your answer.
- (A) $\sigma = 10$
 - (B) $\sigma = 15$
 - (C) $\sigma = 25$
 - (D) $\sigma = 150$
 - (E) Not within ± 5 of any of the above.

The next three questions refer to the following information. The following table categorizes 522 Canadians aged 25 to 64 by highest level of education attained and employment status from 2012.

Employed	Highest Level of Education Attained					Total
	Primary	Jr High	Sr High	College	University	
Yes	7	26	145	105	121	404
No	9	16	43	25	25	118
Total	16	42	188	130	146	522

11. What proportion of employed adults attained a college or university education? Choose the option closest to your answer.
- (A) 0.424
 - (B) 0.433
 - (C) 0.528
 - (D) 0.559
 - (E) None of the above.
12. What proportion of adults had highest attained education of senior high school or lower and was unemployed? Choose the option closest to your answer.
- (A) 0.048
 - (B) 0.130
 - (C) 0.364
 - (D) 0.471
 - (E) None of the above.
13. Among those whose highest attained level of education was senior high school, what proportion was unemployed? Choose the option closest to your answer.
- (A) 0.229
 - (B) 0.364
 - (C) 0.564
 - (D) 0.771
 - (E) None of the above.

The next four questions refer to the following information. Researchers measured the pace taken to run a four-mile route (minutes/mile) and the number of calories burned during the run. Summary statistics for these data are given in the following table.

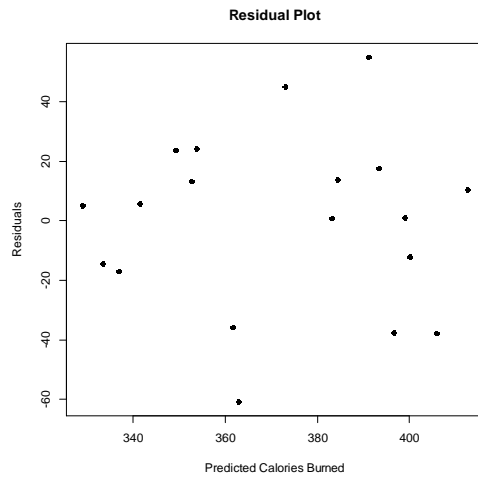
Variable	Mean	Standard Deviation
Pace (x)	8.44	0.395
Calories burned	371.58	39.740
Ln(Calories burned)	5.91	0.108

The fitted lines from two different regressions are given below.

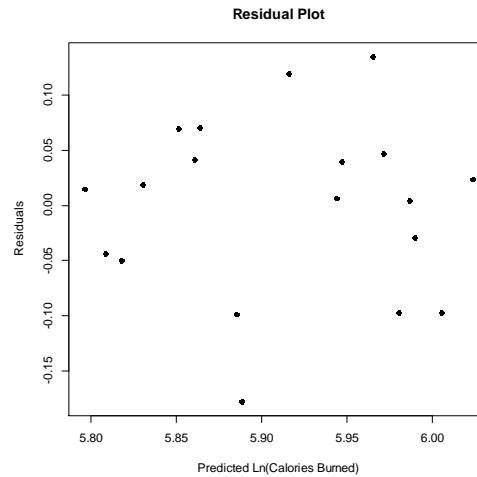
calories burned on pace: $\hat{y} = b_0 - 67.76 \times \text{PACE}$
ln(calories burned) on pace: $\hat{y} = 7.47 - 0.184 \times \text{PACE}$

14. What is the intercept, b_0 , for the regression of **calories burned on pace**? Choose the option closest to your answer.
- (A) $b_0 = -200.31$
(B) $b_0 = 8.44$
(C) $b_0 = 943.47$
(D) $b_0 = 2693.18$
(E) None of the above.
15. One of the runs was run at a pace of 8.7 minutes per mile and the runner was observed to have burned 378 calories. What is the residual associated with this observation for the regression of **ln(calories burned) on pace**? (Note that $\ln(378) = 5.93$). Choose the option closest to your answer.
- (A) 0.06
(B) 6.38
(C) -2.32
(D) -0.06
(E) Cannot be determined from the information given.
16. What is the correlation between **ln(calories) and pace**? Choose the option closest to your answer.
- (A) $r = 0.7$
(B) $r = 0.5$
(C) $r = -0.1$
(D) $r = -0.2$
(E) $r = -0.7$

17. Consider the residual plots shown below.



(i) Calories burned on Pace



(ii) Ln(Calories Burned) on Pace

How many of the following statements are TRUE?

- (i) The residuals appear to be more or less random in both residual plots.
- (ii) The regression of calories burned on pace shows evidence of low variance.
- (iii) The residual plot for the regression of ln(calories burned) on pace appears to have a few suspected outliers.
- (iv) A log transformation of calories burned improved the fit of the regression.

- (A) 0
- (B) 1
- (C) 2
- (D) 3
- (E) 4

18. Which one of the following statements regarding a statistic is FALSE?

- (A) The value of a statistic typically varies from sample to sample.
- (B) A statistic is often used to estimate a characteristic of a population.
- (C) Statistics based on random samples are less likely to be affected by lurking variables compared to statistics based on non-random samples.
- (D) A statistic is a characteristic of a sample.
- (E) The value of a statistic always increases as the sample size increases.

The next two questions refer to the following information. Fifty police officers were recruited into a study and randomly assigned to attend one of two training programs. The control program provided technical training on how to better fight crime and on scientific techniques. The new training program provided the technical training along with training on: (i) public communication and mediation; and (ii) stress management, team building and leadership skills. Pre-study and post-study surveys completed by random community members were used to assess improvement in public's perception score of policemen. The police officers and community members were unaware of which training program the officers took.

19. Which of the following best describes the type of study?

- (A) An observational study.
- (B) A randomized block design.
- (C) A stratified survey.
- (D) A matched pairs design.
- (E) A double-blind study.

20. How many of the following explain why pre-study and post-study surveys were conducted?

- (i) To adjust for natural differences in the skills of the policemen at baseline.
- (ii) If the response rate was too low for the post-study survey then the pre-study survey results could be used instead to compare training program effectiveness.
- (iii) To facilitate calculation of score improvement via difference in perception scores (pre – post).
- (iv) The distribution of the survey scores was not symmetric.

- (A) 0
- (B) 1
- (C) 2
- (D) 3
- (E) 4