201-H02-HR **Final Exam Review** **Fall 2016**

1. For the following descriptions of data, identify the W's (and How, if known), name the variables, specify if the variables are quantitative or categorical, and, for any quantitative variable, identify the units in which it was measured, if provided.

    a. Transport Canada and the Canadian Transportation Agency monitors airlines for safety and customer service. For each flight, carriers must report the type of aircraft, the number of passengers, whether or not the flights departed and arrived on schedule, and any mechanical problems.

    b. At the Tim Hortons Website there is a link called "Nutritional Information" that gives nutrition facts for various categories of food that it sells, such as doughnuts, sandwiches, etc. Under doughnuts, for example, we see a listing of all the doughnut types, and for each, information is provided about the number of calories and the amounts of trans fat, total fat, sodium, sugar, protein, and "% daily value" of iron and calcium.

    c. Scientists at a major pharmaceutical firm conducted an experiment to study the effectiveness of an herbal compound to treat the common cold. They exposed each patient to a cold virus, then gave them either the herbal compound or a sugar solution known to have no effect on colds. Several days later they assessed each patient's condition using a cold severity scale ranging from 0-5. They found no evidence of the benefits of the compound.

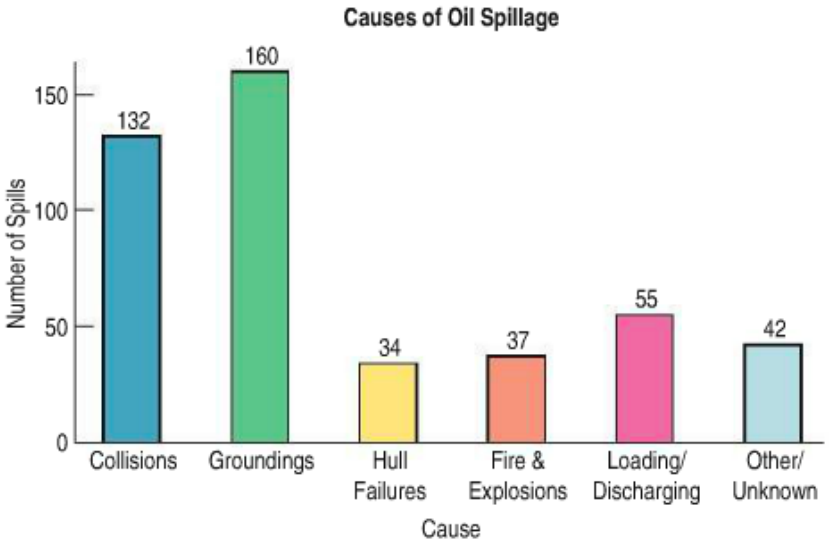
2. Every year, thousands of forest fires burn throughout Canada. In 2010, 6986 forest fires were started. Of those fires, 737 were in Quebec, 931 in Ontario, 1845 in Alberta, and 1673 in BC. Determine the relative frequency distribution of forest fire locations within Canada, and describe it in a sentence or two.

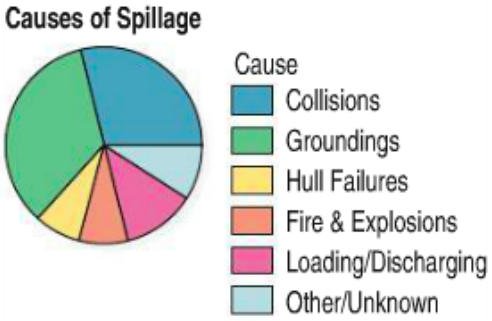
3. The following table shows the number of licenced American drivers (in millions) by age and by sex.

Age	Male Drivers (millions)	Female Drivers (millions)	Total
19 and Under	5.1	4.9	**10.0**
20–24	8.7	8.6	**17.3**
25–29	9.2	9.2	**18.4**
30–34	8.9	8.9	**17.8**
35–39	9.7	9.6	**19.3**
40–44	9.9	9.8	**19.7**
45–49	10.6	10.7	**21.3**
50–54	10.1	10.2	**20.3**
55–59	8.7	8.9	**17.6**
60–64	7.2	7.3	**14.5**
65–69	5.3	5.4	**10.7**
70–74	3.8	4.0	**7.8**
75–79	2.9	3.2	**6.1**
80–84	2.0	2.4	**4.4**
85 and Over	1.4	1.7	**3.1**
Total	**103.5**	**104.8**	**208.3**

a. What percent of total drivers are under 20?
b. What percent of total drivers are male?
c. Write a few sentences comparing the number of male and female licensed drivers in each age group.
d. Do a driver's age and sex appear to be independent? Explain.

4. Data from the International Tanker Owners Pollution Federation Limited give the cause of spillage for 460 oil tanker accidents resulting in spills of more than 700 tons of oil from 1970-2010. Following are some displays.

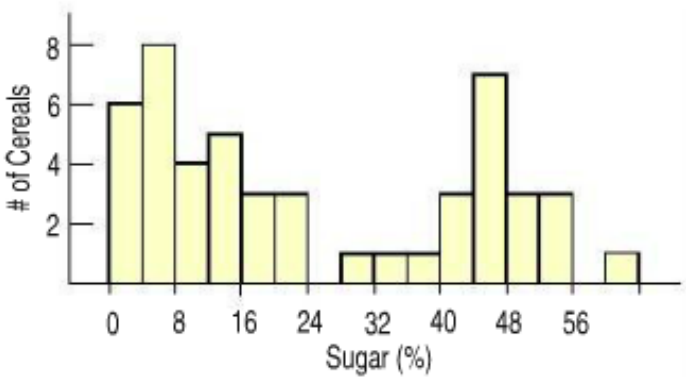




[8]Canada's World Poll's principal sponsor is the Simmons Foundation, and it is also sponsored by the Environics Institute, the Canadian Broadcasting Corporation (CBC), the *Globe and Mail*, and *Le Devoir*.

a. Write a brief report interpreting what the displays show.
b. Is a pie chart an appropriate display for these data? Explain.

5. The histogram displays the sugar content (as a percent of weight) of 49 brands of breakfast cereals.



   a. Describe the modality, skew, and outliers of the data distribution.
   b. What do you think might account for this shape?
   c. Would the mean and standard deviation be appropriate measures of centre and spread, respectively, for this distribution? Explain.

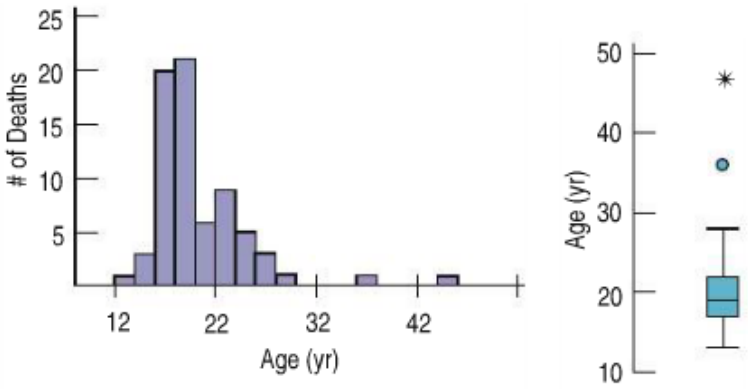
6. A clerk entering salary data into a company spreadsheet accidentally put an extra "0" in the boss's salary, listing it as $2 000 000 instead of $200 000. Explain how this error will affect these summary statistics for the company payroll:
   a. Measures of centre: mean and median
   b. Measures of spread: range, IQR, and standard deviation

7. Below are the annual number of deaths from tornadoes in the United States from 1998 through 2011:

   130, 94, 40, 40, 555, 54, 35, 38, 67, 81, 125, 21, 45, 544

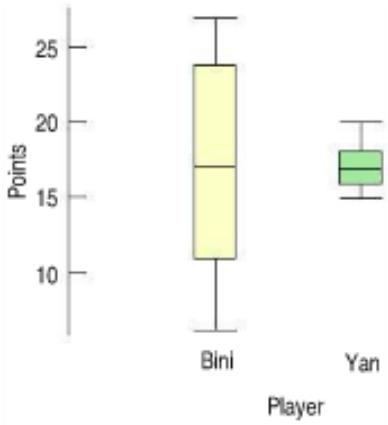
   a. Find the mean, median, first quartile, third quartile, range, standard deviation, and IQR.
   b. Construct a boxplot.

8. Crowd management Strategies monitors accidents at rock concerts. In their database, they list the names and other variables of victims whose deaths were attributed to “crowd crush” at rock concerts. Here are the histograms and boxplot of the victims’ ages for the data from a one-year period:



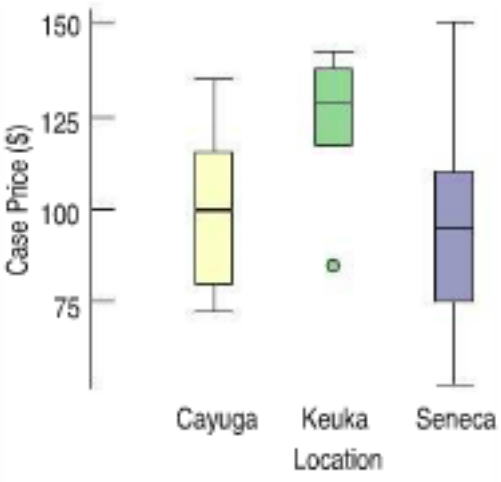
c. What features of the distribution can you see in both the histogram and the boxplot?
d. What features of the distribution can you see in the histogram that you could not see in the boxplot?
e. What summary statistics would you choose to summarize the centre of this distribution? Why?
f. What summary statistic would you choose to summarize the spread of this distribution? Why?

9. Below are boxplots of the points scored during the first 10 games of the season for two players.



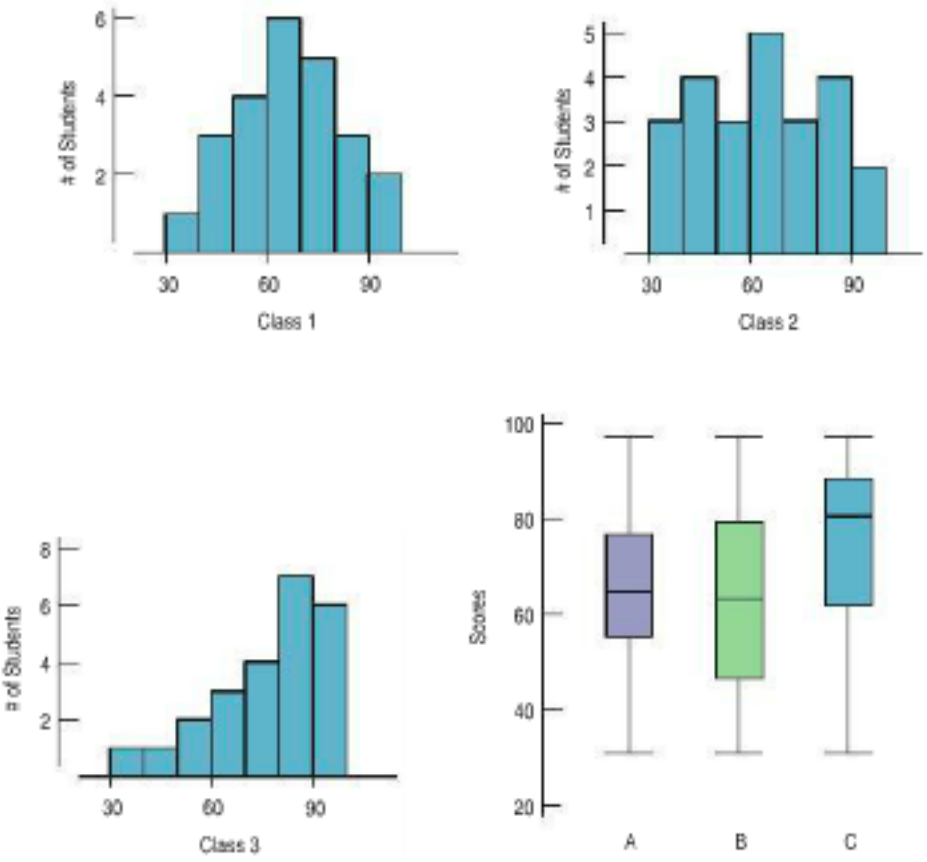
Describe the shapes of the distributions depicted by each of the boxplots separately. Then compare the distributions portrayed by these boxplots.

10. The boxplots display case prices (in dollars) of varieties of wines produced by vineyards along three of the Finger Lakes in New York State.

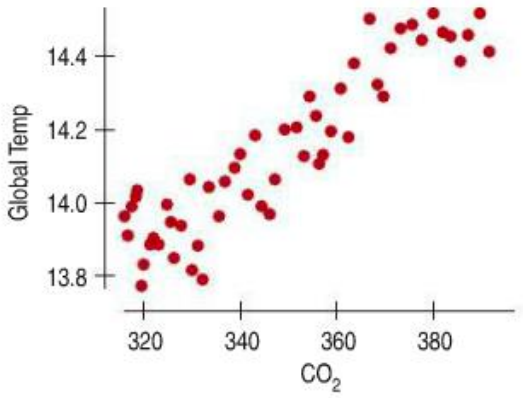


a. Write a few sentences describing these wine prices.
b. Which lake region produces the most expensive wine?
c. Which lake region produces the cheapest wine?
d. In which region are the wines generally more expensive?

11. Three statistics classes all took the same test. Histograms and boxplots of the scores for each class are shown below. Match each class with the corresponding boxplot.



12. The Earth's climate is getting warmer due to the increase in atmospheric levels of carbon dioxide. Below is a scatterplot showing the mean annual CO_2 concentration in the atmosphere, and the mean annual air temperature, from 1959 to 2011.



A regression predicting *Temperature* from CO_2 produces the following output table:

Dependent variable is Global Temperature
R-squared = 84.0%

Variable	**Coefficient**
Intercept	11.0276
CO_2	0.0089

a. Describe the direction, form, and strength of this scatterplot.

b. Comment on each of the Correlation Conditions for this scatterplot.

c. What is the correlation between CO_2 and *Temperature*?

d. What is the correlation between CO_2 and *Temperature*?

e. Explain the meaning of R-squared in this context.

f. Give the regression equation.

g. What is the meaning of the slope in this equation?

h. What is the meaning of the y-intercept of this equation?

13. You find that a scatterplot, which meets the Correlation Conditions, and has the following attributes:

$$\bar{x} = 20$$
$$\bar{y} = 50$$
$$\sum (x - \bar{x})^2 = 100$$
$$\sum (y - \bar{y})^2 = 400$$
$$\sum (x - \bar{x})(y - \bar{y}) = -170$$

    a. Calculate and interpret the Correlation Coefficient.

    b. Construct the Linear Model and use it to approximate the value of the response variable when *x* = 25.

    c. Calculate and interpret the Coefficient of Determination.

14. In 2007, 51% of all emigrants from Canada were male, 82% were at least 18 years old, and 42% were males of at least 18 years of age.
    a. Make a Venn Diagram of this data.
    b. Find the probability that a randomly selected person who emigrated from Canada in 2007 was:
        i. Male and under 18 years of age
        ii. Either male or at least 18 years of age
        iii. Female and under 18 years of age.

15. You draw a card at random from a standard deck of 52 cards. Find each of the following conditional probabilities:
    a. The card is a heart, given that it is red.
    b. The card is red, given that it is a heart.
    c. The card is an ace, given that it is red.
    d. The card is a queen, given that it is a face card.

16. A private high school report contains these statistics:
    70% of incoming Grade 9 students attend public primary schools.
    75% of the students who come from public primary school go on to university after completing high school.
    90 % of the students who did not come from public primary schools go on to university after completing high school.

    a. Is there any evidence that a private-school student's chances of going to university may depend upon what kind of primary school the students attended? Explain.

    b. What percent of this private high school's students go on to university?

17. Calculate each of the following probabilities.

    a. Calculate the probability that a randomly-generated 5-digit number will have no 3s. For this question, a 5-digit number cannot start with a 0.

    b. Six men and five women are eligible to receive 4 identical randomly-allocated prizes. If only one prize can be allocated per person, what is the probability that two men and two women will win?

18. In a survey of 113 and 137 women, 61 men reported that they like prefer Star Wars to Star Trek and 55 women reported that they do not prefer Star Wars.

    a. Construct and complete a contingency table that displays the numbers of men and women and whether they prefer Star Wars or Star Trek. Be sure to have a row and a column for the totals.

    b. Compute the probability that a person surveyed is female given that the person prefers Star Wars.

    c. Are gender and preferring Star Wars dependent? Justify your answer by comparing two relevant probabilities.

19. Consider the random variable, *X*, whose possible values, and their corresponding probabilities, are listed below.

X	100	200	300	400
P(X=x)	0.5	0.1	0.2	0.2

Compute the expected value, variance, and standard deviation of X.

20. A carnival game offers a $100 cash prize for anyone who can break a balloon by throwing a dart at it. It costs $5 to play, and you're willing to spend up to $20 trying to win that cash prize. You estimate that you have about a 10% chance of hitting the balloon on any throw.
    a. Create a probability model for the possible outcomes what you play this carnival game.
    b. Find the expected number of darts you'll throw.
    c. Find your expected winnings.

21. At a certain university, 6% of all students come from outside Canada. Incoming students are assigned at random to freshman dorms, where they live in residential clusters of 40 freshmen sharing a common lounge area. How many international students would you expect to find in a typical cluster, on average? With what standard deviation?

22.
    a. List the three properties that define every set of Bernoulli trials.

    b. If the population modeled by a set of trials is finite, then the trials cannot, technically, be Bernoulli trials if we draw without replacement. We still draw from finite populations for Bernoulli trials provided that they meet the *10% Condition*. Explain what this condition is and which of the 3 properties in part a) is violated when using a finite population.

    c. Why do we sometimes approximate Binom *(n,p)* with a Normal Model? Explain what extra condition must we check in order to make this approximation.

23. Failures of O-rings on the space shuttle are fairly rare, but often disastrous, events. If we are testing O-rings, suppose that the probability of a failure of any one O-ring is 0.01. Let *X* be the number of failures in the next 10 O-rings tested.
    a. What model will you use to model X?
    b. What is the mean number of failures in the next 10 O-rings?
    c. What is the probability that these is exactly one failure in the next 10 O-rings?
    d. What is the probability that these is at least one failure in the next 10 O-rings?

24. Suppose that 70% of the women who suspect they may be pregnant and purchase an in-home pregnancy test are actually pregnant. Further suppose that the test is 98% accurate. What's the probability that a woman whose test indicates that she is pregnant actually is?

25. A television news reporter says that a proposed constitutional amendment is likely to win approval in an upcoming referendum because a poll of 1505 likely voters indicated that 52% would vote in favour. The reporter goes on to say that the margin of error for this poll was 3.0%.

    a. Calculate the confidence interval
    b. What confidence level did the pollster use?
    c. Explain why the poll is actually inconclusive.

26. A province's environmental ministry worries that many cars may be violating clean air emissions standards. The agency hopes to check a sample of vehicles in order to estimate that percentage with a margin of error of 3% and 90% confidence. To gauge the size of the problem, the agency first picks 60 cars and finds nine with faulty emissions systems. How many should be sampled for a full investigation?

27. During the 2013 National Hockey League playoffs, the home team won 59 of the 86 games. Is this strong evidence of a home-ice advantage in professional hockey? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

28. A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this is proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion with a 5% significant level. Be sure that appropriate assumptions and conditions are satisfied before you proceed.

29. Give an example of a situation where you would prefer a Type I Error.

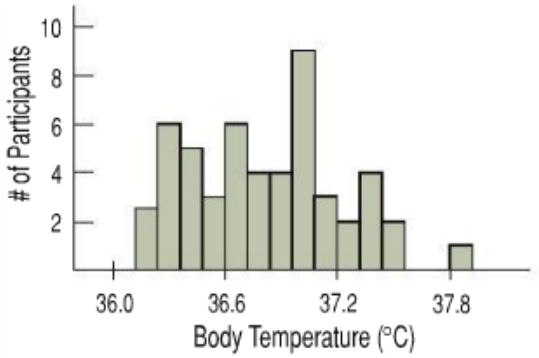
30. Give an example of a situation where you would prefer a Type II Error.

31. A local member of parliament fears he has a problem with women voters. His campaign staff plans to run a poll to assess the situation. They'll randomly sample 300 men and 300 women, asking if they have a favourable impression of the candidate. Obviously, the staff can't know this, but suppose the candidate has a positive image with 59% of males and 53% of females.

    a. What kind of sampling design is his staff planning to use?
    g. Test an appropriate hypothesis to show whether the MP has a problem women. For an alternative hypothesis, claim that the two proportions are different. Be sure to properly state the null hypothesis, use a 5% significance level, and draw some conclusions. Do not forget to pool the data when it is called for in your solution.

32. A researcher measured the body temperatures of that randomly selected group of adults. The data he collected are summarized below. We wish to estimate the average (or "normal") temperature among the adult population.

Summary	
Count	52
Mean	36.83°C
Median	36.78°C
MidRange	37.00°C
StdDev	0.38
Range	1.55
IntQRange	0.58



a. Are the necessary conditions for a *t*-interval satisfied? Explain.

b. Assuming that the distribution does satisfy all the required conditions, construct a 98% confidence interval for the mean from the given that the sample mean sample standard deviation from the table. Write a concluding sentence interpreting this confidence interval.

c. 37 degrees C is commonly assumed to be "normal". Do these data suggest otherwise? Explain.