

## Stat-201 – Surrey

NAME \_\_\_\_\_

Student # \_\_\_\_\_

Please do not open the exam until you are instructed to do so.

No aids are permitted except for standard scientific calculators without Wi-Fi, Bluetooth or cellular capabilities.

No cell phones, PDAs, media players, or other electronic items are permitted.

1 letter size (or A4) page of notes is permitted

Normal,  $\chi^2$ , F and t-distribution tables appear at the end of the exam package. Feel free to remove them

Multiple choice out of 32 points \_\_\_\_\_

Question 1 out of 4 points \_\_\_\_\_ Question 2 out of 11 points \_\_\_\_\_

Question 3 out of 10 points \_\_\_\_\_ Question 4 out of 12 points \_\_\_\_\_

Question 5 out of 12 points \_\_\_\_\_ Question 6 out of 15 points \_\_\_\_\_

Question 7 out of 11 points \_\_\_\_\_ Question 8 out of 17 points \_\_\_\_\_

Question 9 out of 18 points \_\_\_\_\_ Question 10 out of 8 points \_\_\_\_\_

Question 11 out of 5 points \_\_\_\_\_ Question 12 out of 3 points \_\_\_\_\_

Total of 158 possible points

**Multiple choice questions. Chose the best answer (2points each, total 32 points)**

1. A public opinion poll in Ohio wants to determine whether registered voters in the state approve of a measure to ban smoking in all public areas. The researchers select a simple random sample of 50 registered voters from each county in the state and ask whether they approve or disapprove of the measure. This is an example of
    - a. a systematic county sample.
    - b. a stratified sample.
    - c. a multistage sample.
    - d. a simple random sample.
  2. A health club is interested in finding out which of two brands of aerobic exercise equipment provides a more vigorous workout. They purchase 10 machines of each type and for five days between 9 and 12 AM they measure the average pulse rate of each person who is working out on one of these 20 machines. This is an example of
    - a. an experiment.
    - b. an observational study, not an experiment.
    - c. a paired data experiment.
    - d. a stratified experiment.
  3. Sickle-cell disease is a painful disorder of the red blood cells that in the United States affects mostly blacks. To investigate whether the drug hydroxyurea can reduce the pain associated with sickle-cell disease, a study by NIH gave the drug to 150 sickle-cell sufferers and the placebo to another 150. Neither doctors nor patients were told who received the drug. The number of episodes of pain reported by each subject was recorded. This is an example of
    - a. an observational study.
    - b. an experiment, but not a double blind experiment.
    - c. a double blind experiment.
    - d. a paired data experiment.
  4. A stratified random sample corresponds to which of the following experimental designs?
    - a. a block design
    - b. a double blind experiment
    - c. an experiment with a placebo
    - d. a confounded, nonrandomized study
  5. A news release for a diet products company reports: "There's good news for the 65 million Americans currently on a diet." Its own study showed that people who lose weight can keep it off. The sample was 20 graduates of the company's program who endorsed the program in commercials. The results of the sample are probably
    - a. biased, overstating the effectiveness of the diet.
    - b. biased, understating the effectiveness of the diet
    - c. unbiased since the people in the sample are nationally recognized individuals.
    - d. unbiased but they could be more accurate. A larger sample size should be used.
-

---

**These next 2 questions use this information:**

A researcher is interested in investigating the relationship between sugar consumption and weight gain for high school students. Fifteen volunteers were randomly assigned to one of two groups. The first group contained six volunteers who were put on a low sugar diet. The second group consisting of the remaining nine volunteers was put on a diet with sugar constituting approximately 15% of their diet. After eight weeks, the change in weight was recorded for each of the volunteers.

6. The response is
- a. the percentage of sugar in the diet.
  - b. the eight-week time period.
  - c. the assignment to groups.
  - d. the change in weight.
7. This is an example of
- a. a randomized observational study.
  - b. a matched pairs experiment.
  - c. an experiment, but not a double blind experiment.
  - d. a randomized comparative experiment.

---

**These next 2 questions use this information**

One hundred volunteers who suffer from agoraphobia are available for a study. Fifty are selected at random and are given the drug imipramine, which is believed to be effective in treating agoraphobia. The other 50 are given a placebo. A psychiatrist evaluates the symptoms of all volunteers after two months to determine if there has been substantial improvement in the severity of the symptoms.

8. This study would be double blind if
- a. neither drug had any identifying marks on it.
  - b. Neither the volunteers nor the psychiatrist were allowed to see each other during the session during which the psychiatrist evaluated the severity of the symptoms.
  - c. neither the volunteers nor the psychiatrist knew which subjects had received the placebo.
  - d. all of the above.
9. Suppose the volunteers were first divided into men and women, and then half of the men were randomly assigned to the new drug and half of the women were assigned to the new drug. The remaining volunteers received the placebo. This would be an example of
- a. replication.
  - b. confounding. The effects of gender will be mixed up with the effects of the drugs.
  - c. a block design.
  - d. a matched pairs design.
-

---

**These next 2 questions use this information**

Researchers wish to determine if a new experimental medication will reduce the symptoms of allergy sufferers without the side effect of drowsiness. To investigate this question, the researchers give the new medication to 50 adult volunteers who suffer from allergies. Forty-four of these volunteers report a significant reduction in their allergy symptoms without any drowsiness.

10. This study could be improved by
- a. including people who do not suffer from allergies in the study in order to represent a more diverse population.
  - b. repeating the study with only the 44 volunteers who reported a significant reduction in their allergy symptoms without any drowsiness, and giving them a higher dosage this time.
  - c. using a control group.
  - d. all of the above.
11. The experimental units are
- a. the researchers.
  - b. the 50 adult volunteers.
  - c. the 44 volunteers who reported a significant reduction in their allergy symptoms without any drowsiness.
  - d. the 6 volunteers who did not report a significant reduction in their allergy symptoms without any drowsiness.
- 

12. A national survey interviewed 3,800 people age 18 and older nationwide by telephone. One question asked was about the annual income of the person. Of those surveyed, the average annual income was found to be \$42,010. If the survey had interviewed only 1,000 people, which of the following would be true?
- a. By the law of large numbers, the average income would again be \$42,010.
  - b. By the law of large numbers, the average income in the sample of 1,000 people would have to be farther from the true mean income of all people age 18 and older than the average found in the sample of 3,800 people.
  - c. The average income computed from the sample of 1,000 people would be more accurate, because smaller samples tend to be more homogeneous than larger samples.
  - d. None of the above.

13. The number of undergraduates at the University of Michigan is approximately 25,000, while the number at the Ohio State University is approximately 40,000. At both schools, a simple random sample of about 3% of the undergraduates is taken. We conclude

- a. the sample from the University of Michigan has less sampling variability than that from the Ohio State University.
- b. the sample from the University of Michigan has more sampling variability than that from the Ohio State University.
- c. the sample from the University of Michigan has almost the same sampling variability as that from the Ohio State University.
- d. it is impossible to make any statements about the sampling variability of the two samples, since, the students surveyed were different.

14. The sampling distribution of a statistic is

- a. the probability that we obtain the statistic in repeated random samples.
- b. the mechanism that determines whether randomization was effective.
- c. the distribution of values taken by a statistic in all possible samples of the same size from the same population.
- d. the extent to which the sample results differ systematically from the truth.

15. The number of column inches of classified advertisements appearing on Mondays in a certain daily newspaper has mean 320 inches and standard deviation 30 inches. Suppose that the results for 100 consecutive Mondays can be regarded as a random sample and let  $\bar{x}$  denote the mean number of column inches of classified advertisements in the sample. Assuming a sample of 100 is sufficiently large, the random variable  $\bar{x}$  has

- a. a distribution that is exactly normal by the central limit theorem.
- b. a distribution that is approximately normal by the central limit theorem.
- c. a mean of 3.2 inches by the law of large numbers.
- d. a standard deviation of 3.2 inches by the law of large numbers.

16 We wish to estimate the mean price,  $\mu$ , of all hotel rooms in Las Vegas. The Convention Bureau of Las Vegas did this in 1999 and used a sample of  $n = 112$  rooms. In order to get a better estimate of  $\mu$  than the 1999 survey, we should

- a. Take a larger sample because the sample mean will be closer to  $\mu$ .
- b. Take a smaller sample since we will be less likely to get outliers.
- c. Take a different sample of the same size since it does not matter what  $n$  is.
- d. Take a new sample today and combine this data with the data from 1999.

End of multiple choice

Name: \_\_\_\_\_

6

1a. Draw a boxplot showing what data would look like if it came from a normally distributed population. (2pts)

1b. Draw a boxplot showing what data would look like if it came from a distribution that was skewed to the right with two large outliers. (2pts)

Name: \_\_\_\_\_

7

2) A researcher examined records of medical transportation from an accident site to the hospital and the rate of survival. The counts are given in the table below.

Outcome	Helicopter	Road
Victim died	64	260
Victim survived	136	840
Total	200	1100

2a) What are the proportions of survival for each of the 2 transportation methods? From this table is it better for survival to be transported by helicopter or by road? (4pts)

Looking closer, the researcher splits the data into 2 tables and finds the following:

	<u>Serious Accidents</u>			<u>Less Serious Accidents</u>	
Outcome	Helicopter	Road		Helicopter	Road
Victim died	48	60		16	200
Victim survived	52	40		84	800
Total	100	100		100	1000

2b) What are the proportion surviving for each type of transportation method in the serious and less serious accident categories? (4pts)

(continued on next page)

Name: \_\_\_\_\_

8

(continued from last page)

2c) From this table is it better for survival to be transported by helicopter or by road?(1pt)

2d) What is this an example of? (Use a maximum of 10 words for your answer, there is no need for a complete sentence.)(2pts)

Name: \_\_\_\_\_

9

3) An intelligence test gives results which are approximately normally distributed with a mean of 100 and standard deviation of 15 for Canadian adults.

3a) What is the probability that a randomly chosen Canadian adult has a score of 105 or higher?(3pts)

3b) What are the mean and standard deviation of the average score of a SRS of 60 people?(2pts)

3c) A SRS of 400 people gives a mean intelligence test score of 104. Make a 90% confidence interval for the mean.(3pts)

3d) Which (if any) of your answers to part a, b and/or c would change if the population distribution was somewhat skewed instead of normal? (2pts)

Name: \_\_\_\_\_

10

4a ) A 90% confidence interval for the mean of normally distributed variable  $X$  with known variance is from 82 to 102. Find a 95% confidence interval for the mean.(5pts)

4b ) A measurement of lung quality in healthy non-smoking 20 year olds living in Surrey is approximately normally distributed with  $\sigma^2 = 16$ . How many subjects will we need to measure to get a 95% confidence interval with a margin of error of 0.5 units (4pts)

Name: \_\_\_\_\_

11

4c) A simple random sample of statistics exam scores gives a 95% confidence interval for the mean of 68 to 84. Can you safely reject the null hypothesis

$H_0: \mu=60$ ,  $H_a: \mu \neq 60$  at the 5% significance level? Why or why not? (3pts)

Name: \_\_\_\_\_

12

5) A group is interested in knowing if there is a difference between any of the 3 bridge designs proposed for the twinning of the Port Mann bridge. The group builds 10 models of each of the 3 bridge designs and tests their strength. The measured strength variable is approximately normally distributed. The resulting data is partly summarized in the table below.

5a) State the hypothesis being tested. (2pts)

5b) Fill in the blanks in the table (8pts)

Source:	df:	SS:	MS:	F	p-value
BridgeModel	___	7433.9	_____	___	_____
Error	___	_____	_____		
Total	___	20013.4			

5c) What do you conclude? (2pts)

Name: \_\_\_\_\_

13

6) I have a quarter and a loonie that I would like to toss and am interested in whether heads (H) or tails (T) lands face up on.

6a) List the sample space for a single toss of these two coins.(2pts)

6b) Draw a Venn diagram for these events(4pts):

6c) Assuming that the coins are fair, that is, for any of these two coins  $P(H)=P(T)=.5$ . Find the probability of both coins landing heads up:  $P(HH)$ . (2pts)

(continued on next page)

Name: \_\_\_\_\_

14

(continued from last page)

6d) What is the probability of one H and one T? (2pts)

6e) If I toss both coins and tell you that at least one of the coins landed heads up, what is the probability that both are heads? (5pts)

Name: \_\_\_\_\_

15

7) A candy coated chocolate candy called w&w's comes in 6 colours. The colours are thought to be found in candy packages with this distribution:

Pink: 1/10	Blue: 2/10
Purple: 1/10	Red: 3/10
Brown: 1/10	Yellow: ?

7a) What must be the probability of a randomly selected candy being yellow in this distribution? (2pts)

7b) I take a SRS of 200 candies and record the counts of each colour in the table below. The null hypothesis that I'm interested in testing is that the true population distribution is the same as the one given above. State the alternative hypothesis. (1point)

7c) After taking a SRS of candies I observe the counts of colours in the table below. What are the expected counts? (6pts)

	Observed	Expected
Pink:	24	_____
Purple:	28	_____
Red:	55	_____
Blue:	64	_____
Brown:	24	_____
Yellow:	5	_____
Total count:	200	_____

(continued on next page)

Name: \_\_\_\_\_

16

(continued from last page)

7d) If the value of the test statistic  $X^2 = 42$  (note that this value is given for the convenience of the exam and is not the value that you would compute using the numbers above) what is your conclusion and p-value?(4pts)

Name: \_\_\_\_\_

17

8) A new fraud detection machine called "the Greco" is being tested for detecting gambling fraud at the Banks casino in Las Vegas. In a new casino game version of dominos a player wins with probability of .4. If a player wins more than 6 out of 7 plays the "Greco" machine considers the win suspicious.

8a) Let  $X$  be the number of wins out of 7 attempts for the dominos game. What is the distribution of  $X$ ? (3pts)

8b) In a particular night 10,000 players try the game for 7 attempts each. Assuming that these players are playing fairly, about how many times would you expect the machine to falsely accuse people of cheating? (3pts)

(continued on next page)

Name: \_\_\_\_\_

18

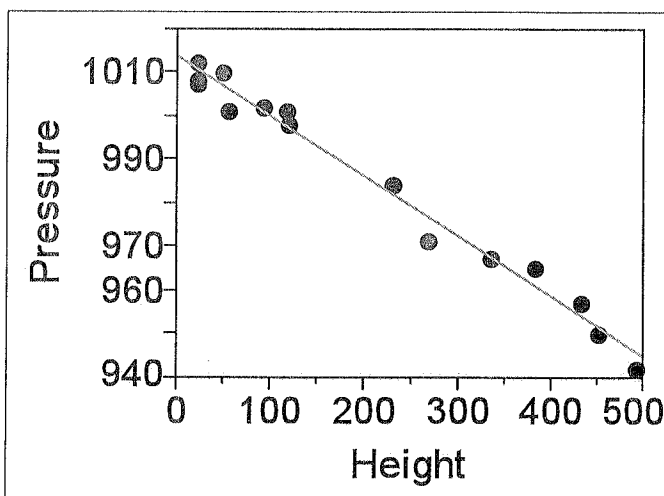
(continued from last page)

8c) One player named Danny Ocean plays the game often and is observed to have won 55 out of 80 tries. Let  $p$  be the proportion of times that Danny wins this game. We are interested in knowing if his proportion of wins is somehow larger than it should be. State the hypothesis in statistical symbols. (2pts)

8d) Can you conclude that Danny's winnings are suspicious (test the hypothesis from part c)? Don't forget to check conditions for your test. (9pts)

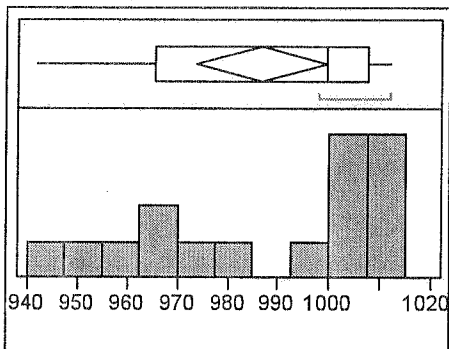
9) The US Environmental Protection Agency (EPA) is monitoring the pressure above the city of Springfield. They are interested in how it changes with as they move from the low altitude (near the houses) to the higher altitude regions usually reserved for birds, planes and Superman. They find the results below. The questions are on the next pages. If it's helpful you can remove this page.

**Bivariate Fit of Pressure By Height**



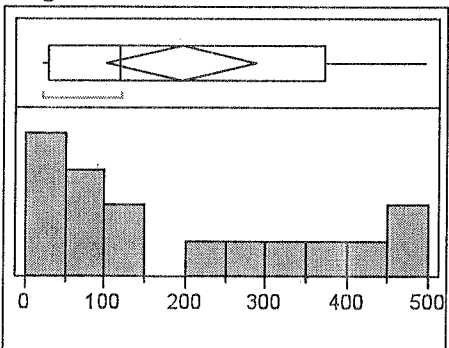
RSquare	0.98
RSquare Adj	0.98
Root Mean Square Error	3.28
Mean of Response	986.7
Observations (or Sum Wgts)	16

**Pressure**



Mean	986.7
Std Dev	24.2
Std Err Mean	6.055
upper 95% Mean	999.59
lower 95% Mean	973.78
N	16

**Height**



Mean	195.4
Std Dev	174.4
Std Err Mean	43.59
upper 95% Mean	288.34
lower 95% Mean	102.53
N	16

(continued on next page)

Name: \_\_\_\_\_

20

(Continued from last page)

9a) Find the slope ' $a$ ' and intercept ' $b$ ' for the regression line:  $Pressure = a + b Height$   
(4pts)

9b) What proportion of the variation in the pressure is explained by the variation in height? (1point)

9c) Test if the correlation is significantly less than zero. Although it's not exactly true, for this question let's say that  $\sum (x - \bar{x})^2 = 100260$ . Also, to avoid potentially building

on mistakes from earlier, for this question we will say that the regression line equation is  $Pressure = 800 - 0.1 Height$ . Note that these numbers are not what you would get if you used the actual data from the plot but please use them anyways. Be sure to state your hypothesis and conclusion. (10pts)

(continued on next page)

Name: \_\_\_\_\_

21

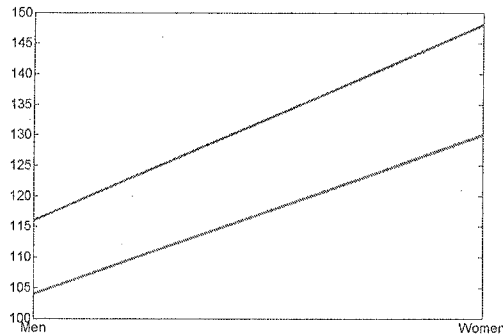
(Continued from last page)

9d) Using the regression equation:  $Pressure = 800 - 0.1 Height$ , and the plots and JMP output from earlier, what is the expected pressure at a height of 1000m (just give a point estimate, there is no need for a confidence interval)? Is this a reasonable value? (3points)

Name: \_\_\_\_\_

22

10) A study of road rage risk factors compared commuters whose average daily drive is at least 30 minutes with a control group who mostly use transit but drive on occasion. The study examined the participants heart rate while being subjected to annoying behaviours for 200 men and 200 women in each of the treatment groups. A portion of the 2-way ANOVA results from JMP are given below.



green line is control group, blue line is commuter group

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	215256.09	71752.0	296.3455
Error	796	192729.83	242.1	
C. Total	799	407985.92		
				Prob > F
				<.0001

#### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	104.49	0.55014	226.29	0.0000
Group[Commuter]	11.51	0.55014	26.38	<.0001
Gender[Female]	25.50	0.55014	13.64	<.0001
Group[Commuter]*Gender[Female]	17.69	0.55014	2.72	0.0066

What are the expected values (group means) for each of the four factor combinations?(8pts)

Name: \_\_\_\_\_

23

11) The number of SFU Surrey students who ask for more information about taking a minor in statistics in a given term follows a Poisson distribution with mean 25. Let  $X$  be the number of students inquiring in one year. What is the probability that at least 100 students ask for this information in one year if there are 3 terms per year? Be clear about checking conditions. (5pts)

Name: \_\_\_\_\_

24

12) Researchers want to examine the distribution of salaries in Canadian urban centers. Should they use the mean and standard deviation or the median and quartiles to summarize the population? Why? (3pts) (keep your answer short)