PASS MOCK EXAM

– FOR PRACTICE ONLY –

Course: GEOG 2006 **Facilitator:** Iain Burnett

Dates and locations of mock exam take-up: Monday April 11th, 6-8 pm, ME 3174,
Tuesday April 12th, 12-2 pm, ME 3174

It is **most beneficial** to you to write this mock midterm **UNDER EXAM CONDITIONS**. This means:

- Complete the midterm in ___ hour(s).
- Work on your own.
- Keep your notes and textbook closed.
- Attempt every question.

After the time limit, go back over your work with a different colour or on a separate piece of paper and try to do the questions you are unsure of. Record your ideas in the margins to remind yourself of what you were thinking when you take it up at PASS.

The purpose of this mock exam is to give you practice answering questions in a timed setting and to help you to gauge which aspects of the course content you know well and which are in need of further development and review. Use this mock exam as a ***learning tool*** in preparing for the actual exam.

Please note:

- Come to the PASS session with your mock exam complete. There, you can work with other students to review your work.
- Often, there is not enough time to review the entire exam in the PASS session. Decide which questions you most want to review – the Facilitator may ask students to vote on which questions they want to discuss.
- Facilitators do not bring copies of the mock exam to the session. Please print out and complete the exam before you attend.
- **Facilitators do not produce or distribute an answer key for mock exams.** Facilitators help students to work together to compare and assess the answers they have. If you are not able to attend the PASS session, you can work alone or with others in the class.

Good Luck writing the Mock Midterm!!

Part 1: Multiple Choice

1. Using sample data to make a generalization about a population is best explained as:
a) Descriptive Statistics
b) Inferential Statistics
c) Relational Statistics
d) All of the above

2. Which of the following is an example of a variable at the ratio level of measurement?
a) Temperature (°F)
b) Colours
c) Letter grades for university
d) Length (mm)

3. True or false: An error caused by poor precision can be improved through recalibration.
a) True
b) False

4. Three values were gathered for tree heights at Carleton; 3.63 m, 2.9 m, 3.21 m. You calculate the mean tree height from the three measured. The mean is:
a) 3.25 m
b) 3.2 m
c) 3.247 m
d) 3 m

5. You have a data set with 100 values of mercury levels in fish and want to find the average level across the 100 fish. You create a histogram of your variable and find the data to be heavily skewed to the right. What form of central tendency would be **most representative** for your data set?
a) Mean
b) Median
c) Standard Deviation
d) Mode

6. What is the third quartile for the following data set?
{1, 3, 3, 7, 11, 12, 17, 18, 19, 21, 23, 27, 27, 29, 30}
a) 21
b) 23
c) 27
d) Cannot be determined with this information alone

7. True or false: A p-value of 0.035 on a Shapiro-Wilkes test shows a normally distributed data set.
a) True
b) False

8. The concept that the outcome of one event has no bearing on the probability of another event is best described as:
a) Mutually exclusive
b) Exhaustive
c) Conditional probability
d) Statistical independence

9. What is the probability of getting a value within one standard deviation of the mean on a normal distribution?
a) 50 %
b) 72 %
c) 95%
d) 68%

10. You have a data set with a mean of 13.1, a sample size of 29, a standard deviation of 3.4. What is the unreliability of your sample?
a) 0.63
b) 2. 43
c) 0.26
d) 3.85

11. With regards to the central limit theorem, how does the sample size affect the sampling distribution?
a) As n increases, the distribution becomes less normal
b) As n increases, the distribution becomes more normal
c) As n decreases, the distribution becomes more normal
d) n does not affect the distribution

12. What would be the main difference in moving from a 95% confidence interval to a 99% confidence interval?
a) The interval would be larger
b) The population mean would be larger
c) The population mean would be smaller
d) The interval would be smaller

13. You have a sample mean of 10, a standard deviation of 3.06. What is the probability of getting a value larger than 16?
a) 0.05
b) 0.025
c) 0.01
d) 0.16

14. What does a p-value represent?
a) The probability of rejecting the null hypothesis when it is false
b) The probability of committing a type 2 error
c) The probability of obtaining a more extreme test statistic due to random chance
d) All of the above

15. You want to test to see if your sample tree height of 23.1 feet is significantly greater than the world average tree height of 21.95 feet. You have a sample size of 25 and a standard deviation of 4.56. What test do you use to test your hypothesis?
a) A paired t-test
b) A Z-test
c) A student's t-test
d) A Welch's t-test

16. You run a Pearson's Correlation test and you get an r equal to 0. What is the shape of the relationship between the variables?
a) Linear and positive
b) Horizontal
c) Linear and negative
d) No way of knowing without looking at a scatterplot

17. You calculate the covariance between two variables and get an incredibly low (negative) value. What direction is the relationship between the variables?
a) Positive
b) Negative
c) Vertical
d) Not enough information to know

18. When you have a negative residual in a linear regression model, it means the model is:
a) Over predicting the value
b) Under predicting the value
c) Not significant
d) Not in the right position and it needs to be moved up or down.

19. Which value shows what is explained by the regression line in a bivariate regression?
a) SSE
b) SSR
c) SSX
d) SSY

20. You are testing to see if there is a relationship between amount of snowfall and elevation. You get a coefficient of determination of 0.67 and a p-value of 0.03. You plot the residuals with the elevation (independent variable) and find a linear relationship. This means:
a) Your residuals are statistically independent and your linear regression is likely accurate
b) The coefficient of determination is over estimating your response variable
c) The correlation coefficient is much lower than you expected
d) The errors are statistically dependent and further analysis should be conducted.

Part 2: Applied Questions

1. Calculate the mean, sum of squares, variance, and standard deviation for the following data set; 4.2, 2.1, 7.8, 5.4, 3.3, 6.01.

2. You have a data set with 36 values of CO_2 emissions from cars. The set has a standard deviation of 3.5, and a mean of 31.4 ppm. The data has a normal distribution.
a) Calculate the standard score associated with a value of 34 ppm.

b) You want to know what amount of CO_2 emissions 25% of the values are below. You know that $P(< z) = 0.25$ where $z = -0.67$. At what value does 25% of the data lie below?

3. You took two samples sets of ant lengths; one from Ottawa and one from Toronto. Each had a sample size of 51 ants. You run a Shapiro-Wilk test on both sets and receive p-values of 0.04 and 0.47 respectively. You are looking to see if the average length of ants from Ottawa is different than that of Toronto.
a) Write the null and alternate hypothesis for the test you will conduct, as well as the test you will use.

b) You get a test statistic of 1.25 and a p-value of 0.072 on your difference test, what does this mean, and what can you conclude about your hypothesis?

c) How would your answer have changed if the Shapiro-Wilk test produced both p-values as 0.04? As 0.47?

4. You want to test how related global temperature and amount of snowfall are in British Colombia.
a) Describe the steps you would take to make this decision (preliminary steps, hypothesis, and any tests you would conduct). Assume both variables have a normal distribution.

b) Assume you receive a correlation coefficient of -0.36. What does this imply about the relationship between these variables?

c) How would your answer change if the variables were not normally distributed?

5. You wish to produce a linear regression model for two variables. As you did above, state the steps required to produce a linear regression. Proceed with the steps necessary, the hypothesis, the assumptions that would need to be followed, and any post analysis that needs to be done.

b) The linear regression produced provides the following output:

```
Residuals:
   Min     1Q Median     3Q    Max
 -5337  -2381  -1188   1672  14630

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     40382.3     1943.6  20.777  < 2e-16 ***
data$UnemployM   -932.1      295.2  -3.157  0.00278 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3807 on 47 degrees of freedom
Multiple R-squared:  0.175,       Adjusted R-squared:  0.1574
F-statistic: 9.967 on 1 and 47 DF,  p-value: 0.002782
```

Interpret the results and report on the coefficient of determination and the slope with respect to their p-values. What does the coefficient of determination tell you about the linear regression produced?

Part 3: Theory

1. Describe the three types of probability and give an example of each.

2. What are the three measures of central tendency and in which circumstances would they be most useful?

3. What are the different types of “two sample difference of means tests”. Explain when you would use them and how they differ. Include the non-parametric version.

4. Define skewness and kurtosis, and explain how they are related to a data sets' distribution.

5. Discuss the differences between the correlation coefficient and the coefficient of determination. Discuss their similarities.

6. What is the difference between accuracy and precision?

7. What are the differences between a t-test and a wilcoxon rank sum test? Why would you use one rather than the other?