MATH108 10FA Final A
[answers in web view] Total points: 150

Name:_____
Student ID:_____

> Let the words of my mouth and the meditation of my heart
> Be acceptable in Your sight, O LORD, my Rock and my Redeemer.
> -- Psalm 19:14

- Please show all your work! No partial credit will be given for incorrect answers with no work shown.
- Please draw a box around your final answer.
- You are only permitted to use your own calculator and writing implements. Cell phones should be muted and left in your pocket or bag.
- All relevant tables are attached to the back. You may detach them for your reference.
- Assume $\alpha = 0.05$ everywhere unless indicated otherwise.
- For t-tests on two groups, if the df is not given, you may use the conservative estimate of $df = \min(n_1, n_2) - 1$.

1. Indicate the **level of measurement** for each of the following variables as categorical (G), ordinal (O), discrete (D), or continuous (C). **[6]**

   (a) Location of injury: e.g., knee, lower back, shoulder, chest, etc.

   (b) Number of correct answers on a multiple-choice test

   (c) Number of children in a family, coded as 0, 1, 2, or "at least 3"

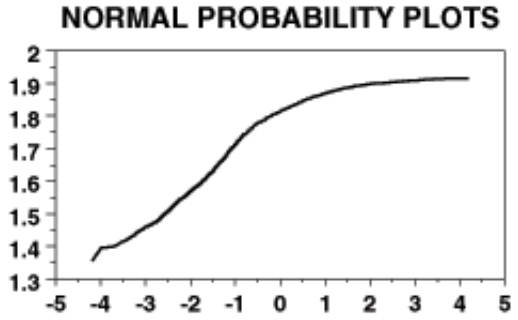
   (d) Blood glucose level (mg/dLi)

   (e) Whether a woman is pregnant or not

   (f) Strength of family bonds, rated as "Very Strong", "Somewhat Strong", "Weak", or "Very Weak"

2. The following is a **normal probability plot** of potassium concentration in a number of geologic samples. The horizontal axis is expected normal scores (*n*-scores), and the vertical axis is observed potassium concentration (this orientation matches the textbook). **How** does the distribution differ from a normal distribution? **Sketch** the distribution, highlighting where it is non-normal. **[4]**

   [Data from USGS Open-File Report 2005-1231.]



3. Suppose that in a study half of the participants are **nurses** and 80% of the participants consider their jobs to be **high-stress**. Consider the probability that a participant in the study is a nurse who considers his/her job to be high-stress.

(a) What is the **minimum** possible value for this probability? Draw a **Venn** diagram illustrating this situation. [3]

(b) What is the **maximum** possible value for this probability? Draw a **Venn** diagram illustrating this situation. [3]

4. In a study of Canadian nurses, say that 70% of the nurses work in **hospitals**, and one quarter of the nurses habitually **smoke**. 20% of all the nurses in the study are smokers who work in hospitals.

   (a) For **each** of the three probabilities given (70%, 25%, 20%), express the probability in **notation** (e.g., P(smoke)) and draw a **Venn diagram**, shading in the relevant region (draw three separate Venn diagrams). [3]

   (b) In this study, what is the chance that a nurse working in a hospital smokes? [3]

   (c) In this study, is working in a hospital **independent** of smoking? Why or why not? [3]

5. A particular FDG-PET (fludeoxyglucose positron-emission tomography) screening test for non-Hodgkin's lymphoma has a 15% false-positive rate (85% specificity) and 90% sensitivity (i.e., 90% of lymphomas are caught by the screening process).

   (a) Suppose the screening test is applied to 200 patients, of which 80 have non-Hodgkin's lymphoma. Draw an **event tree** for the outcomes of the test, and **label** the tree with probabilities for each branch of the tree. [4]

(b) On average, how many people in this group will **test positive** for non-Hodgkin's lymphoma? **[3]**

(c) If a patient tests positive using this test, what is the probability that the patient **really** has non-Hodgkin's lymphoma? **[2]**

6. The average number of hours of exercise per week was measured for a number of urban dwellers and rural dwellers. A 95% **confidence interval** for the difference of means (urban - rural) is (-0.27, 1.23). Based on this information, indicate whether each of the following statements is "True" or "False". (Please write the entire word, "True" or "False".) **[6]**

   (a) Urban dwellers exercise an average of between 0.27 hrs less and 1.23 hrs more per week than rural dwellers.

   (b) 95% of urban dwellers exercise between 0.27 hrs less and 1.23 hrs more per week than rural dwellers.

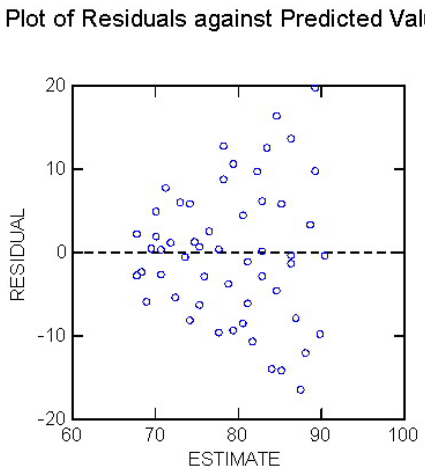
   (c) We are 95% certain that urban dwellers exercise between 0.27 hrs less and 1.23 hrs more per week than rural dwellers.

   (d) With 95% confidence, the difference in hrs/week of exercise between urban and rural dwellers in this study is between -0.27 and 1.23.

   (e) At a 5% level of significance, this study is unable to find a difference in amount of exercise between urban and rural dwellers.

   (f) There is no difference in the amount of exercise for urban and rural dwellers.

7. Below is a **residual plot** for a linear regression model relating blood pressure to age (data from UNC SOCI709 course). From this plot, is there evidence to indicate that any of the **assumptions** of regression may have been violated? Sketch a possible **scatterplot** of blood pressure versus age that would reflect this residual plot. *(Hint: generally, blood pressure increases with age.)* **[5]**



8. In a study of BC nurses, an analysis was run to determine whether which nursing **school** the nurse graduated from had an impact on **salary**.

   (a) What are the **variables** which need to be measured for each baby? For each variable, indicate its level of measurement and whether it is a predictor or outcome variable. [2]

   (b) What is the appropriate parametric statistical **test** to run? [1]

   (c) State the null and alternate **hypotheses**, both in words and in appropriate notation. [3]

   (d) The data were collected and an appropriate analysis run, obtaining a *p*-value of 0.07. State the **conclusion** of the analysis, and interpret it in the context of the original research question. [2]

   (e) The ***p*-value** was 0.07. What does this number '0.07' mean, in the context of the research question? 0.07 of what? [3]

9. For BC nurses, is being **married** independent of working **over 60 hours/week**? The number of participants in each category is listed in the table below.

	Married	**Not Married**
≤ 60hrs	150	80
> 60hrs	90	80

   (a) What is the **population** of interest? [1]

   (b) Name the **variable(s)** which need to be measured, indicate their levels of measurement, and whether each is a predictor or outcome variable. [2]

   (c) State the null and alternate **hypotheses**, both in words and in appropriate notation. Which statistical test(s) would be appropriate? [3]

(d) Run the appropriate test and bracket a ***p*-value**. [4]

(e) State the **conclusion** from this test, and interpret it in the context of the original research question. [2]

10. Does blood vitamin **B12** level (pg/mL) have an impact on **depressive** symptoms (Beck Depression Inventory (BDI-II), on a scale from 0-63 points)?

    (a) What is the **population** of interest? [1]

    (b) Name the **variable(s)** which need to be measured, indicate their levels of measurement, and whether each is a predictor or outcome variable. [2]

    (c) What is the appropriate parametric statistical **test** to run? [1]

    (d) State the null and alternate **hypotheses**, both in words and in notation. [2]

    (e) A study with 60 participants results in the following data:
    $SS_X = 1{,}350{,}000$, $SS_Y = 9{,}000$, $SS_{XY} = -70{,}000$.
    Find the **slope** of the best-fit line, indicate its **units**, and **interpret** the slope in light of the model for vitamin B12 and depression. (Keep at least 4 significant figures in the slope.) [3]

    (f) The **average** vitamin B12 level in the study was 500 pg/mL, and the average BDI score in the study was 45 points. Find the **equation** of the best-fit line, and interpret the **intercept** of the line in light of the model. [3]

(g) Find the **correlation** between vitamin B12 level and depressive level in this study. [2]

(h) What **fraction** of the variability in depressive levels in this study is explained by the linear relationship with vitamin B12 levels? [2]

(i) Describe the **distribution** of BDI depressive levels predicted by the linear model when vitamin B12 levels are at 600 pg/mL. [4]

(j) Answer the original research question: bracket a p-value, state your conclusion, and interpret it in light of the original research question. [4]

11. Does **income** level (low, middle, high) have an impact on **caloric intake** (calories per day)?

(a) Name the **variable(s)** which need to be measured, indicate their levels of measurement, and whether each is a predictor or outcome variable. [2]

(b) What is the appropriate parametric statistical **test** to run? [1]

(c) State the null and alternate **hypotheses**, both in words and in notation. [2]

(d) Data for this experiment are given below. **Run** an appropriate test and bracket the *p*-value. [5]

Low-income:	1200	1800	2400
Middle-income:	1200	1350	
High-income:	2100	2200	

(e) State the **conclusion** from this test, and interpret it in the context of the original research question. [2]

(f) What are the **assumptions** of the statistical test you performed? Is there evidence to suggest that any of these assumptions have been violated in this dataset? [3]

12. Human beta-endorphin (HBE) is a hormone secreted by the pituitary gland under conditions of stress (like exams!). Suppose we wish to determine whether blood concentration of **HBE** (pg/mL) is higher for a group of men **jogging** as compared to a group of men **resting**.

(a) Name the **variable(s)** which need to be measured, indicate their levels of measurement, and whether each is a predictor or outcome variable. [2]

(b) State the null and alternate **hypotheses**, both in words and in appropriate notation. Which statistical test(s) would be appropriate? [3]

(c) Data for this experiment are given below. Sketch **boxplots** for the data, on a common axis (number line). [4]

								Mean:	**SD:**
Jogging:	60	58	62	49	51	58	54	56	4.7958
Resting:	41	37	51	60	28	35		42	11.6276

(d) Run an appropriate **parametric** test and bracket the *p*-value. [5]

(e) State the **conclusion** from this test, and interpret it in the context of the original

research question. [2]

(f) Using the same data, perform an appropriate **non-parametric** test and bracket the *p*-value. [4]

(g) State the **conclusion** from this test, and interpret it in the context of the original research question. [2]

(h) Which test do you think is more **appropriate** for this data, the parametric or the non-parametric test? Why? [2]

13. Does HBE concentration in men increase after they exercise? Data from a study of 6 men are below.

							Mean	SD
Before:	42	58	38	50	49	48	47.5	6.921
After:	47	57	44	53	49	53	50.5	4.722

(a) Name the **variable(s)** which need to be measured, indicate their levels of measurement, and whether each is a predictor or outcome variable. [2]

(b) State the null and alternate **hypotheses**, both in words and in appropriate notation. Which statistical test(s) would be appropriate? [3]

(c) Run an appropriate **parametric** test and bracket the *p*-value. [5]

(d) State the **conclusion** from this test, and interpret it in the context of the original research question. [2]

(e) Using the same data, perform an appropriate **non-parametric** test and bracket the *p*-value. **[3]**

(f) State the **conclusion** from this test, and interpret it in the context of the original research question. **[2]**

(g) Which test do you think is more **appropriate** for this data, the parametric or the non-parametric test? Why? **[2]**