MATH108 10FA Midterm ch1-6 B
[answers in web view] Total points: 70

Name:_____
Student ID:_____

> "For You light my lamp; the LORD my God illumines my darkness.
> For by You I can run upon a troop; and by my God I can leap over a wall.
> As for God, His way is blameless; the word of the LORD is tried;
> He is a shield to all who take refuge in Him."
> -- Psalm 18:28-30

- 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.
- Calculators are permitted, but no notes, text, laptops, PDAs, or electronic dictionaries. Cell phones should be muted and left in your pocket or bag.
- Tables 3 and 4 are attached to the back. You may detach them for your reference.

1. Say that in a certain town 60% of the people vote Conservative, and the probability that a townsperson drives a large truck is 4/7. If you pick a random townsperson, there is a 27% chance that the townsperson is a Conservative driving a large truck.

   (a) What fraction of Conservatives in this town drive large trucks? **[3]**

   (b) In this town, is voting Conservative **independent** of driving a large truck? Why or why not? Interpret what this means in the context of the townspeople.**[3]**

2. Classify each of the following statements (which may or may not be true) as either **(D)escriptive** or **(I)nferential**: **[4]**

   (a) About 90% of students in our MATH108 class are nursing students.

   (b) 30.1% of male registered nurses (RNs) are under the age of 40.

   (c) 19.2% of BC residents in 2004 were clinically obese.

   (d) Of the children and adolescents studied in the 2004 Canadian Community Health Survey, 8% were clinically obese.

3. Mark each of the following **variables** as categorical (G), ordinal (O), discrete (D), or continuous (C): **[5]**

   (a) Hemoglobin count, in g/dL

   (b) Age, divided into categories "0-10", "11-18", "19-25", "26-39", "40-59", "60 and up"

   (c) How many children are in a family

   (d) Satisfaction with current family doctor, rated as "Very Dissatisfied", "Dissatisfied", "Satisfied", or "Very Satisfied"

   (e) Whether a student passes a course or not

4. The list below shows the total serum bilirubin level (mg/dL) in a newborn baby, measured at various times over a one-month period.

3.7, 4.4, 5.8, 3.1, 5.2, 5.6, 3.3, 2.5, 5.1

(a) Construct a relative frequency **histogram**, classifying the data by bins of width 1 mg/dL.**[4]**

(b) Find the sample **mean**. Show your work. **[2]**

(c) Find the sample **standard deviation**. Show your work. **[4]**

(d) Draw a **boxplot** for the data. Show your work. **[4]**

(e) Find a 95% **confidence interval** for the average total serum bilirubin level for this baby over the one-month period. **[4]**

(f) **Interpret** this confidence interval in words and in the context of the study. **[2]**

(g) We feel this confidence interval is still too broad. In order to estimate the average total serum bilirubin level to within ±0.4 mg/dL with 90% confidence, **how many** measurements would we need to average? (Assume the underlying true bilirubin level does not change.) **[4]**

(h) If we increase the sample size to n=40 measurements, would we expect the distribution of our sample to be more **normal**? Why or why not? [2]

5. An assay (test) measuring blood lactate during exercise is imprecise, returning values normally distributed around the true lactic acid concentration, with a standard deviation of 3 mmol/L. Blood lactate values of over 20 mmol/L are considered "high".

   (a) What is the probability that this assay returns a result **within** ±2.7 mmol/L of the true value? [2]

   (b) If the true concentration of blood lactate in a patient is 18.2 mmol/L (not "high"), what is the probability that the assay could still return a value that is considered "**high**"? [2]

   (c) Since the assay is so imprecise, we run it **five** times and average the results. Using the same true concentration of 18.2 mmol/L, what is the probability that the average of five runs is still "high"? [4]

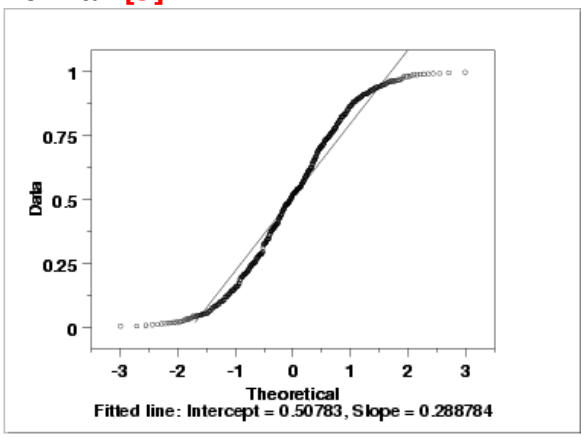
6. A particular screening test for colon cancer has a 17% false-positive rate (i.e., 83% **specificity**) and a 5% false-negative rate (i.e., 95% **sensitivity**).

   (a) Suppose the test is applied to a group of patients, 40% of whom are known to have colon cancer. Draw an **event tree** for the outcomes of the test. Label the tree with probabilities for each branch. Also calculate the probabilities of each final outcome (leaf of the tree). [5]

   (b) What is the probability that a random patient from this group will **test negative** for colon cancer using this screening test? [2]

   (c) What is the probability that a patient from this group who tested negative **actually** does not have the disease? [2]

7. The diagram below is a **normal probability plot** for an artificially-generated dataset. (Note that the n-scores are along the horizontal axis, and the y-scores from the data are along the vertical axis; this matches the textbook.) How does the distribution differ from a normal distribution? Sketch the distribution, highlighting where and how it deviates from the normal. **[5]**



[NIST Engineering Statistics Handbook, §1.4.2.2.2]

8. Suppose you wish to study whether male nurses are more likely than female nurses to quit within the first five years of working as a nurse.

   (a) What is the **population** in question? **[1]**

   (b) List the **variables** which need to be measured. **[1]**

   (c) For each variable, indicate its **level** of measurement and whether it is a **predictor** (independent variable) or **outcome** (dependent variable). **[2]**

   (d) Discuss how you might do the **sampling** process. What are some steps you could take to ensure a random sampling? Address both criteria for random sampling. How would you measure the variables? Would the measurement process bias your sampling? **[3]**