Please print in pen:
Waterloo Student ID Number:

WatIAM/Quest Login Userid:

UNIVERSITY OF WATERLOO

Examination
Test
Winter 2020
STAT 341

Times: Friday 2020-02-07 at 09:30 to 10:20
Duration: 50 minutes
Exam ID: 4449268
Sections: STAT 341 LEC 001
Instructors: Nathaniel Stevens

Special Materials

Candidates may bring only the listed aids.
· Calculator - Pink Tie

Instructions:

- You have 50 minutes to complete this test.
- This test consists of 6 questions and 8 pages (including this cover page).
- Page 8 contains additional space for rough work. DO NOT use this page for anything that you would like to have marked.
- Numeric answers should be rounded to four decimal places (unless the answer is exact to fewer than four decimal places).
- Incorrect answers may receive partial credit if your work is shown. An incorrect answer with no work shown will receive 0 points.

Question	Points
Q1	4
Q2	6
Q3	4
Q4	3
Q5	5
Q6	8
Total	**30**

- Please identify yourself by signing here: ______________________________

Please initial:

1. **[4 points]** Consider the population $\mathcal{P} = \{y_1, \ldots, y_N\}$ and the attribute $a(\mathcal{P})$.

   (a) [1 point] What does it mean for $a(\mathcal{P})$ to be **location-scale equivariant**?

   (b) [3 points] Suppose that the attribute $a(\mathcal{P})$ is location equivariant **and** scale equivariant. Show that $a(\mathcal{P})$ is therefore location-scale equivariant.

Please initial:

2. **[6 points]** This question concerns sensitivity analysis.

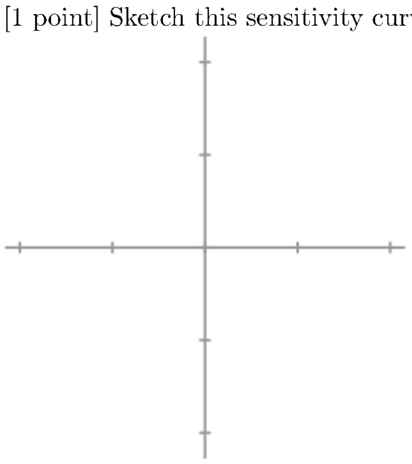
(a) [1 points] Given a population $\mathcal{P} = \{y_1, \ldots, y_{N-1}\}$ and an attribute $a(\mathcal{P})$, define the sensitivity curve for the attribute.

(b) [5 points] Consider the population above and assume N is odd (i.e., $N = 2m+1$). The order statistics in this case are $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(N-1)}$ and the median can be written as

$$a(\mathcal{P}) = a(y_1, \ldots, y_{N-1}) = \frac{y_{(m)} + y_{(m+1)}}{2}$$

i. [3 points] Derive the median's sensitivity curve.

ii. [1 point] Sketch this sensitivity curve.



iii. [1 point] From this curve, what do you conclude about the median's resistance to outliers in y?

3. [**4 points**] Consider the population $\mathcal{P} = \{y_1, \ldots, y_N\}$, and the following two attributes that measure its spread. The first attribute is the standard deviation defined as

$$a_1(\mathcal{P}) = \sqrt{\frac{\sum_{u \in \mathcal{P}} (y_u - \overline{y})^2}{N}}$$

and the second attribute is the interdecile range (the difference between the first and ninth deciles)

$$a_2(\mathcal{P}) = Q_y(0.9) - Q_y(0.1)$$

where $Q_y(p)$ is the $100 \times p\%$ quantile of the population.

(a) [1 point] Define the breakdown point for a population attribute.

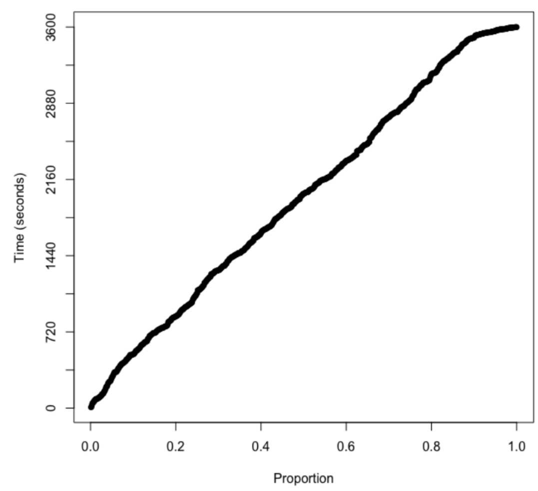
(b) [2 points] Determine the breakdown point for both attributes $a_1(\mathcal{P})$ and $a_2(\mathcal{P})$. Provide rationale for your answer.

(c) [1 point] Given your answers in part (b), which attribute is more robust (i.e., resistent to outliers) and why?

Please initial:

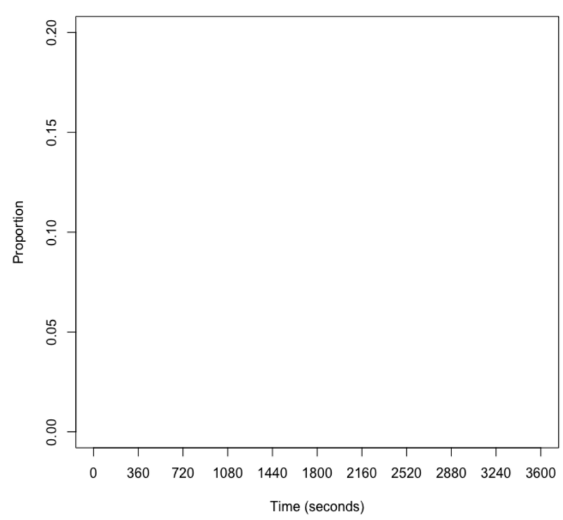
4. **[3 points]** Wayne Gretzky "The Great One" is a Canadian former professional ice hockey player. He played 20 seasons in the National Hockey League (NHL) and he is considered to be the greatest hockey player ever. Below is a quantile plot of the $N = 894$ goals he scored during his time in the NHL. In particular, we examine the times (in seconds) at which the goals occurred in a sixty-minute game.

(a) [1 point] Using this plot provide an estimate of the interquartile range.



(b) [1 points] Using your answer from (a) together with the Freedman-Diaconis Rule, explain why when creating a histogram of these data, 10 bins would be a sensible choice.

(c) [1 point] Using the axes below, construct a histogram for these data using 10 bins. (Hint: Consider concentration boxes and the connection between quantile plots and histograms.)



Please initial:

5. **[5 points]** This question concerns the implicitly defined attribute $\boldsymbol{\theta} \in \mathbb{R}^k$ in population $\mathcal{P}$.

(a) [1 point] Provide the objective function-based definition of $\widehat{\boldsymbol{\theta}}$. Define any notation that you use.

(b) [1 point] Provide the system of equations-based definition of $\widehat{\boldsymbol{\theta}}$. Define any notation that you use.

(c) [3 points] In point form, describe the batch-stochastic gradient descent algorithm. Define any notation that you use.

6. **[8 points]** Determine whether the following statements are True or False. In each case circle the correct answer.

   (a) [1 point] The ratio of two scale equivariant attributes is a scale invariant attribute.

   i. True
   ii. False

   (b) [1 point] A relative-frequency histogram (where the height of each bar reflects the *proportion* of data lying in the bin) is replication invariant.

   i. True
   ii. False

   (c) [1 point] When evaluating a population attribute, the notions of *influence* and *sensitivity* are the same.

   i. True
   ii. False

   (d) [1 point] Consider the population $\mathcal{P} = \{y_1, \ldots, y_N\}$. Suppose that a histogram of this data is left-skewed (i.e., negatively skewed). If we wanted to use a power transformation to make the histogram more symmetric, we should use a power that is less than 1.

   i. True
   ii. False

   (e) [1 point] Compared to batch-sequential gradient descent, batch-stochastic gradient descent is less senstive to outliers in the data.

   i. True
   ii. False

   (f) [1 point] Consider the objective function $\rho(\theta) = 2\theta^2 - 5\theta + 3$. Stochastic gradient descent would be a useful technique for determining

   $$\underset{\theta \in \mathbb{R}}{\operatorname{argmin}}\, \rho(\theta)$$

   i. True
   ii. False

   (g) [1 point] In objective function-minimization problems, the Newton-Raphson algorithm can be viewed as a form of gradient descent.

   i. True
   ii. False

   (h) [1 point] Iteratively reweighted least squares is a root-finding algorithm.

   i. True
   ii. False

This space is left for rough work

Please initial: