Tecnológico de Monterrey
IN 4027: Data Science and Statistical Inference
November, 29th, 2019

**Final Exam**

Name:_____
Student Number:_____

1. Describe a situation in which you would use the Log-rank test instead of a two-sample t test (10 points). Further, describe the hypothesis you would be testing with the Log-rank test (10 points) and describe the steps to carry it out (10 points).

2. Derive the deviance of a Binomial GLM from the general definition of the deviance (30 points)

3. True or false (5 points each, explain why):

a) The James-Stein estimator approaches the MLE estimator as individual averages $p_i$ have a larger variance.

b) $\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2)} dz = \sqrt{2\pi}$

c) The MLE estimator's $\widehat{\boldsymbol{\theta}}$ variance decreases as we increase the dimensionality of parameter vector $\boldsymbol{\theta}$ increases.

d) Ridge regression shrinks all regression parameters towards zero (no effect on response)

e) Bootstrapping requires that we believe that observations are independently drawn from the same distribution.

4. Derive the Kaplan-Meier estimator from the survival and hazard functions (15 points)

5. Assume you have a dataset with 10 observations (label them x1, … x10). Calculate the probability of getting bootstrap sample [x1,x1,x1,x5,x6,x7,x7,x8,x9, x10]. (15 points)