Homework No.1 (MSDS 954:567) Spring 2022

Due date: 2/17/2022 23:59 pm

Problem 1. Given the dataset

ID	RACE	SBP	DBP	HR
001	W	130	80	60
002	B	140	90	70
003	W	120	70	64
004	W	150	90	76
005	B	124	86	72

(NOTE: SBP is systolic blood pressure, DBP is diastolic blood pressure, and HR is heart rate.)

Write an R code to compute the "average" blood pressure. (ABP) defined as a weighted average of the diastolic blood pressure and the systolic blood pressure. Since the heart spends more time in its relaxed state (diastole), the diastolic pressure is weighted two-thirds, and the systolic blood pressure is weighted one-third. Therefore, the average blood pressure could be computed by multiplying the diastolic blood pressure by 2/3, and the systolic blood pressure by 1/3 and adding the two. An equivalent expression would be the diastolic pressure plus one-third of the difference between the systolic and diastolic pressures. Using either definition, add APB to the data set.

Problem 2. Rice (1995, p. 390) gives the following data (Natrella, 1963) on the latent heat of the fusion of ice (cal/gm):

Method A: 79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02

Method B: 80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

(a) Inspect the data graphically in various ways, for example, boxplots, Q-Q plots and histograms.

(b) Assuming normality, test the hypothesis of equal means, both with and without making the assumption of equal variances.

(c) Compare the result with a Wilcoxon/Mann-Whitney nonparametric two sample test.

[*Remark: For (a), solve it both by paper and pencil and also using a computer. For (b) and (c), use a computer for your calculation, but write out the corresponding formulas*]

Problem 3. Given the following data points:

-1.43 -0.95 -0.19 0.02 0.14 0.83 1.35 1.46 2.62

(a) Calculate the sample median, IQR (interquartile range) and MAD (median absolute deviation); What are the (robust) breakdown points of these statistics? (explain your answers)

(b) Draw a histogram, using 3 groups (with equal bandwidths).

[*Remark: Solve this problem by paper and pencil.*]

Problem 4. Consider the model $y_i = \beta_0 + x_i\beta_i + \epsilon_i$, where ϵ_i are iid $N(0, \sigma^2)$ for $i = 1, 2, \ldots, 5$. We have the following data:

X	Y
-1.0	- 1.5
0.0	0.3
1.0	0.9
2.0	2.1
3.0	2.3

(a) Find the Least Squares (LS) and MLE estimates of β_1, σ^2 and variance of $\hat{\beta}_1$.
(b) Find the 95% confidence interval for β_1.

[*Remark: Solve this problem by paper and pencil.*]

Problem 5. Suppose $X_1, \cdots, X_n \overset{iid}{\sim} Uniform[0, \theta]$. We would like to estimate and make inference for θ.
(a) Find the MLE $\widehat{\theta}^{MLE}$ for θ and compute the density of $\widehat{\theta}^{MLE}$.
(b) Let the underlying truth be $\theta = 1$. Suppose we are only able to get $n = 60$ samples. Generate your own observations $x_1, \cdots, x_{60}$. Based on them we can do bootstrap and get the bootstrap MLE $\widehat{\theta^*}$. Bootstrap for $N = 1000$ times and plot the density of $\widehat{\theta^*}$.
Compare the distribution of $\widehat{\theta}^{MLE}$ and $\widehat{\theta^*}$ according to the density plot you obtain. Are they symmetric or asymmetric?
(c)Based on $\widehat{\theta}^{MLE}$ and 1000 $\widehat{\theta^*}$ you get in (b), find $P(\widehat{\theta}^{MLE} = 1)$ and $P(\widehat{\theta^*} = \widehat{\theta}^{MLE})$.
(d) Based on the 1000 $\widehat{\theta^*}$ obtained in (b), build the bootstrap 90% confidence interval for θ. Simulate the data for $N = 1000$ times and get the 90% confidence interval for θ based on the $\widehat{\theta}^{MLE}$. Compare these two intervals.
(e) This example shows that boostrap does poorly. In fact, try to prove $P(\widehat{\theta}^{MLE} = 1) = 0$ and $P(\widehat{\theta^*} = \widehat{\theta}^{MLE}) = 1 - (1 - \frac{1}{n})^n$. You may use these two facts to check whether you have done correctly in (c).

[*Hint: In this example, the Bootstrap Central Limit Theorem does not hold. This is an example that the bootstrap method fails to work.*]

[*Remark: For (a) and (e), solve it by paper and pencil. For (b),(c),(d), use a computer for your calculation, and write down the formula and results.*]