Review Test Submission: M2

User	Venkata Subba Narasa Bharath Meadam
Course	-CSE 544.01 Probability and Statistics for Data Scientists - Spring 2020
Test	M2
Started	5/6/20 2:31 PM
Submitted	5/6/20 3:21 PM
Due Date	5/6/20 3:30 PM
Status	Completed
Attempt Score	44 out of 100 points
Time Elapsed	50 minutes out of 50 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

0 out of 4 points

Consider a distribution that takes value 2 with probability x and 0 with probability (1-x). Find the MME of x and report a numerical estimate of its standard error. That is, report $\hat{se}(\hat{x}_{MME})$. Use {2, 0, 2} as the sample data. Report your final answer as a number with one digit before the decimal and rounded to two digits after the decimal (ex: 0.033 is reported as 0.03 and -1.607 is reported as -1.61).

Selected Answer: 0.00

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.27	

Question 2

4 out of 4 points

Consider a specific version of the one-tailed Z-test which is defined with H_0: u = u_0 vs H_1: u > u_0. Assume the data is Normally distributed with true mean u and true standard deviation sigma. For confidence level a, what is the Pr(accept H_1 | H_0 true) of this version of the Z-test?

Selected Answer: $1-\Phi(z_a)$

Answers:

$\Phi(z_a)$

$2\Phi(z_a)$

$2\Phi(z_{a/2})$

$1-2\Phi(z_a)$

$1-\Phi(z_a)$

$1-2\Phi(z_{a/2})$

None of the listed

$1-\Phi(z_{a/2})$

$\Phi(z_{a/2})$

Question 3

4 out of 4 points

Refer to the Pearson Correlation test from class and use the thresholds from class to decide the correlation between X={0, 4} and Y={-1, 1}.

Selected Answer: Positive linear correlation

Answers:

Positive linear correlation

No correlation

Negative linear correlation

None of the listed

Question 4

4 out of 4 points

Consider a probability distribution X that takes values 4 and 1 with probabilities c and (1-c), respectively. Given some i.i.d. data samples $\{X_1, X_2, ..., X_n\}$ distributed as X, with sample mean X_b, find the MLE of c. You can ignore the 2nd derivative condition when finding the MLE.

Selected Answer: $(X_b - 1)/3$

Answers:

$X_b/3$

$(X_b - 1)/3$

None of the stated

X_b

$X_b/4$

Question 5

4 out of 4 points

Use K-S test to compare X={0.3, 0.9, 0.5} with Unif(0, 1) and report the statistic. However, for the statistic, report the minimum distance between the two distributions from among all values in your table. Compute the differences to the left and right of data points as usual, as in class, but then find the min value from among these and report that. Report the result as a decimal rounded to two digits after the decimal point and with exactly one digit before the decimal (ex: 1.466 is rounded and reported as 1.47 and 0.103 is rounded and reported as 0.10).

Selected Answer: 0.03

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.03	

Question 6

0 out of 4 points

Consider a distribution that takes values -1, 0, and 1 with probabilities x/3, x/3, and (3-2x)/3, respectively. Find the MLE of x given data samples {-1, 0, 1, 1, 1}. You do not have to worry about the 2nd derivative when computing MLE. Hint: for this question, it may make sense to make use of data samples while computing MLE. Report your answer with exactly one digit before the decimal and rounded to two decimal places (ex: 0.066 is reported as 0.07, 1.033 is reported as 1.03, and 1.1 is reported as 1.10).

Selected Answer: 0.50

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.60	

Question 7

0 out of 4 points

Let $D=\{X_1, X_2, ..., X_n\}$ be i.i.d. distributed with a distribution having p.m.f. p(x) such that p(x) is proportional to $c^x e^{-c}$, where c is the unknown parameter. Let $S = X_1 + X_2 + ... + X_n$. Let the prior of c be the H(a, b) distribution which has p.d.f. f(x) proportional to $x^{(a-1)}$

$e^{-b x}$. Use Bayesian inference to compute the posterior of c given data D. You will find that the posterior looks like the H distribution (after ignoring any constants). What are the parameters of this posterior H distribution?

Selected Answer: [None Given]

Answers:
- None of the listed
- (n+a-1, 1+b)
- (S+a-1, 1+b)
- (n+a, 1+b)
- (S+a, 1+b)

Question 8

4 out of 4 points

Consider the simple linear regression fit as discussed in class. However, this time assume that the ground truth error is always 0. Find the β_0 term given sample data {(2, 6), (1, 9)}, where each sample pair refers to (Y_i, X_i). Report the final answer with exactly one digit before the decimal and rounded to two decimal places (ex: 0.066 is reported as 0.07, -1.033 is reported as -1.03, and 1 is reported as 1.00).

Selected Answer: 4.00

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	4.00	

Question 9

0 out of 4 points

You are given data samples {2, 2, -2, -2} which you know to be Normally distributed with some unknown mean but variance 3. Compute the t-statistic for this data assuming a simple 2-tailed t-test for checking whether the true mean is 0.2.

Selected Answer: None of the listed

Answers:
- -0.1
- -0.2
- -0.067
- -0.4
- None of the listed
- -0.133

Question 10

0 out of 4 points

Consider a distribution with two parameters, x and y. The mean of the distribution is x+1 and the second moment is y/x. Report the MME for y given sample data {2, 0, 2, 1}. Report your answer with exactly one digit before the decimal and rounded to two decimal places (ex: 0.066 is reported as 0.07, 1.033 is reported as 1.03, and 1.1 is reported as 1.10).

Selected Answer: 0.20

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.56	

Question 11

0 out of 4 points

Consider the paired t-test to test whether H_0: $\mu_X = \mu_Y$ vs H_1: $\mu_X \neq \mu_Y$, and for this question assume that the test is applicable. Let X={1, 3, 2} and Y={0, 2, 2}. Let the critical threshold or the right-hand-size of the t-test (the $t_{n-1,alpha/2}$ value) be denoted as c. For what values of c will paired t-test NOT result in a rejection for the stated X and Y?

Selected Answer: None of the listed

Answers:
- 4.75

4.5

None of the listed

5.25

5

Question 12

4 out of 4 points

In class, we covered the simple Z-test. We also saw in class how to go from the simple t-test to the paired t-test. Apply the same approach to the simple Z-test to get the paired Z-test for checking the null hypothesis that the two population means are equal. Consider the two-tailed version of this test (so alternate hypothesis is that the two population means are unequal).

Let the two sample sets be X and Y, and let the set of differences (X-Y) be denoted as set D, as in class. Assume that the test is applicable, and so samples in D are Normally distributed with known standard deviation of 2. For alpha value of 0.05, which of the following cases will NOT be rejected by the paired Z-test as designed above?

Selected Answer: X={1, 3, 2, 0} and Y={1, 0, 3, 0}

Answers:

X={3, 0, 1, 2} and Y={5, 3, 1, 7}

X={4, 4, 3, 3} and Y={2, 1, 2, 1}

X={1, 3, 2, 0} and Y={1, 0, 3, 0}

None of the listed

Question 13

0 out of 4 points

Let $\{X_1, X_2, ..., X_n\}$ be i.i.d. samples distributed as Exponential(1/b). Use MLE of b as your estimator. Let H_0: b <= b_0 vs H_1: b > b_0. Using Wald's test, for what value of b_0 can you reject the null hypothesis given {1, 1, 1, 3} as your sample data. Use alpha value of 0.05. Hint: A4, Q2 can help.

Selected Answer: 0.75

Answers:

None of the listed

1.5

0.75

2

2.8

0.2

Question 14

0 out of 4 points

Consider the Bernoulli-Beta conjugate prior example from class. Specifically, you are given D=$\{X_1, X_2, ..., X_n\}$ which are i.i.d. distributed as Bern(p), where p is unknown. Using Unif(0, 1) as your prior for p, you conduct Bayesian estimation after each round using the posterior at the end of the previous round as your new prior and using the new data seen on that day. You are also given that for the Beta(a, b) distribution, the mean is a/(a+b).

Assume you see one coin flip each day, and the coin flip result alternates between failure and success depending on whether the day is odd or even. Thus, on day 1, you see {0} as your data. On day 2, you see a new {1} as your data, so your full data up to day 2 is {0, 1}. Likewise you see a new {0} on day 3 so your full data on day 3 is {0, 1, 0}. Likewise, on day 4, you see a new {1}, so your full data on day 4 is {0, 1, 0, 1}. As in class, assume you compute a new posterior at the end of each day using the posterior from the end of the previous day as your prior and using the new data seen on that day as your input data; you then take the new posterior mean as your estimate of p. Thus, at the end of day 1, you use {0} as your dataset and compute the posterior mean. At the end of day 2, using the posterior at the end of day 1 as your new prior, you compute the new posterior using the result of coin flip seen on day 2, i.e., {1}, as the input data. At the end of how many days will your estimate of p first be greater than or equal to 0.5?

Selected Answer: 1

Answers:

2

4

3

None of the listed

1

Question 15

4 out of 4 points

Consider a case where 25 females voted for party A and 35 females voted for party B. In the same election, 25 males voted for party A and 15 males voted for party B. You are asked to find the chi-squared statistic, Q, for this question. However, in the definition of the Q statistic in class, we have a sum of terms, where each term has a squared term in the numerator. That is, each term in the summation is of the form N^2/D. For this question, we redefine the Q statistic to not have the square term in the numerator but instead have the absolute value of the numerator, thus, we replace each N^2/D term with $|N|/D$. Report this redefined Q statistic value for this question. Report the value with exactly one digit before the decimal and rounded to 2 digits after the decimal (ex: 1.466 is reported as 1.47 and 0.103 is reported as 0.10).

Selected Answer: 0.83

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.83	

Question 16

0 out of 4 points

Consider a set of i.i.d. Bernoulli RVs D={X_1, X_2, ..., X_n} with true but unknown Bernoulli parameter p. Consider the second element, X_2. Since it is a Bernoulli RV, we know that Var(X_2) = p(1-p). Since p is unknown, we can estimate Var(X_2) as X_b(1-X_b), where X_b is the sample mean of D. However, we can also estimate Var(X_2) using the sample variance, $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - X_b)^2$. It turns out that both estimates are equal. Why are the two estimates of Var(X_2) equal?

Selected Answer: Because sample mean is a consistent estimator.

Answers:

- Because the data is distributed as Bernoulli.
- Because both estimators are consistent.
- Because sample mean is a consistent estimator.
- Because sample variance is a consistent estimator.
- None of the listed.
- Because the data samples are i.i.d.

Question 17

4 out of 4 points

We defined the correlation metric in class. Using the same notation, consider a variant of that metric called squared correlation defined as $\frac{E[(X^2 - E[X^2])(Y^2 - E[Y^2])]}{\sigma_X^2 \sigma_Y^2}$. After some algebra, the squared correlation simplifies to which of the below expressions?

Selected Answer: ($E[X^2Y^2]$ - $E[X^2]$ $E[Y^2]$) / (Var(X) Var(Y))

Answers:

- Var(XY) / (Var(X) Var(Y))
- ($E[X^2Y^2]$ - $E[X^2]$ $E[Y^2]$) / (Var(X) Var(Y))
- ($E[X^2Y^2]$ - $E^2[X]$ $E^2[Y]$) / (Var(X) Var(Y))
- 0
- None of the listed

Question 18

0 out of 4 points

In Bayesian inference, when computing the posterior distribution, consider the case where the prior is Unif(0, 1). Now, after computing the posterior, if you use the MAP (Maximum a posteriori) estimate of the parameter, we can say that this estimate will always (even if if the number of data samples is not large) be the same as

Selected Answer: Bayesian mean of the parameter

Answers:

MME of the parameter

None of the listed

Bayesian mean of the parameter

MLE of the parameter

Question 19

0 out of 4 points

Consider Q5 from A5 (one-sided unpaired T-test) but this time let H_0: $u_1 < u_2$ and H_1: $u_1 >= u_2$. Also, assume that n and m are small. With these changes to the question, what will be the p-value?

Selected Answer:

$$1-\Phi\left(\frac{\overline{X}-\overline{Y}-(u_1-u_2)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}\right)$$

Answers:

$$1-\Phi\left(\frac{\overline{X}-\overline{Y}-(u_1-u_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\right)$$

$$\Phi\left(\frac{\overline{X}-\overline{Y}-(u_1-u_2)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}\right)$$

None of the listed

$$\Phi\left(\frac{\overline{X}-\overline{Y}-(u_1-u_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\right)$$

$$1-\Phi\left(\frac{\overline{X}-\overline{Y}-(u_1-u_2)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}\right)$$

Question 20

0 out of 4 points

Consider the simple linear regression but with the predicted fit being $\widehat{Y}=(\widehat{\beta})^2 X$, and given $\widehat{\beta} \neq 0$. Note that there is no intercept term here. Solve for the OLS estimate of $\widehat{\beta}$ in this case (do not worry about the second derivative condition) and report the value for sample data of {(1, 3), (2, 4)}, where each sample pair refers to (Y_i, X_i). Report the final answer with exactly one digit before the decimal and rounded to two decimal places (ex: 0.066 is reported as 0.07, 1.033 is reported as 1.03, and 1.1 is reported as 1.10).

Selected Answer: 0.20

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.66	

Question 21

4 out of 4 points

Perform a 2-sample KS test and find the max difference statistic for X={8, 2} and Y={2, 2, 3, 1}.

Selected Answer: 0.50

Answers:
- 0.50
- 0.25
- 0.00
- None of the listed
- 0.75

Question 22

0 out of 4 points

Consider the Wald's test for checking whether a given coin in unbiased or not given an i.i.d. sample of data $\{X_1, X_2, ..., X_n\}$. However, for the $\hat{\theta}$ estimator in the definition of Wald's statistic, use X_1 (first element) as your estimator; even though this estimator is not consistent or AN, assume the test is applicable with this estimator. Given sample data {1, 0, 0}, report the absolute value of the Wald's statistic using the given estimator.

Selected Answer: None of the listed

Answers:
- $\frac{3\sqrt{3}}{2\sqrt{2}}$
- None of the listed
- $\frac{3}{2}$
- $\frac{3}{2\sqrt{2}}$
- 1

Question 23

0 out of 4 points

Consider the Normal(a, b^2) distribution with mean a and variance b^2. Define M as the second moment of the distribution. Find the MLE of M for data {-1, 1, 1}. Report your final answer with exactly one digit before the decimal and rounded to two digits after the decimal (ex: 1.003 is reported as 1.00 and 0.109 is reported as 0.11).

Selected Answer: 0.94

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	1.00	

Question 24

4 out of 4 points

Let X={1, 9} and Y={2}. Use Permutation test with the statistic as absolute difference in means. The p-value, rounded to two decimal places and including one digit before the decimal (ex: .466 is rounded and reported as 0.47 and 1.133 is rounded and reported as 1.13) for the given data is [FILL IN BELOW].

Selected Answer: 0.67

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.67	

Question 25

4 out of 4 points

For this question, refer to lecture 15, slides 6-8. Consider the sick patient example and use the terminology and notation as in class. In a clinical trial of a new disease detection test, there were a 100 healthy patients and a 100 sick patients. The test correctly identified 98 out of the 100 healthy patients as healthy. The test also correctly identified 99 of the 100 sick patients as sick. The remaining patients were incorrectly classified. What is the Type II error of the test.

Selected Answer:	None of the listed.
Answers:	2/100
	2/101
	1/99
	None of the listed.
	1/101

Wednesday, May 6, 2020 7:57:40 PM EDT

← OK