Review Test Submission: M2

User	Akhila Juturu
Course	-CSE 544.01 Probability and Statistics for Data Scientists - Spring 2022
Test	M2
Started	5/4/22 8:15 PM
Submitted	5/4/22 9:30 PM
Due Date	5/4/22 9:35 PM
Status	Completed
Attempt Score	250 out of 250 points
Time Elapsed	1 hour, 15 minutes out of 1 hour and 15 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

10 out of 10 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 $c^x e^{-c x}$, 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 of the listed

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

Question 2

10 out of 10 points

Clinical trials for the effectiveness of the keto diet, which claims to reduce weight by 5 pounds, were performed. A random sample of 39 participants lost an average of 6.5 pounds. Given the true variance of the population as 36, the absolute value of the z-statistic for the hypothesis that the keto diet's claim is true is __ . Report the result as a decimal rounded to two digits after the decimal point (ex: 14.466 is rounded and reported as 14.47, 1.103 is rounded and reported as 1.10)

Selected Answer: 1.56

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	1.56	

Question 3

10 out of 10 points

Consider a distribution that takes values 0, 1, and 2 with probabilities x/2, x/2, and (1-x), respectively. Find the MLE of x given data samples {1, 2, 0, 0, 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: 0.80

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity

Exact Match	0.80

Question 4 10 out of 10 points

Consider a distribution that takes value 4 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 $\widehat{se}(\hat{x}_{MME})$. Use {4, 4, 0} 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: 0.27

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.27	

Question 5 10 out of 10 points

Students' grade point average (GPA) can be assumed as a Gaussian distribution. Consider a hypothesis about the mean GPA μ, H_0: $\mu \geq 3.0$, H_1: $\mu < 3.0$. By using an appropriate testing method on a large random sample dataset, we get the observed sample mean 3.20 and the p-value. What does the p-value mean?

Selected Answer: f. The probability of finding a student whose GPA is lower than 3.20, assuming the mean GPA is no lower than 3.0

Answers:

a. The probability of finding a student whose GPA is lower than 3.0

b. The probability of finding a student whose GPA is lower than 3.20

c. The probability that the true mean GPA is lower than 3.0

d. The probability that the true mean GPA is lower than 3.20

e. The probability of finding a student whose GPA is lower than 3.0, assuming the mean GPA is no lower than 3.0

f. The probability of finding a student whose GPA is lower than 3.20, assuming the mean GPA is no lower than 3.0

g. The probability of finding a student whose GPA is lower than 3.20, assuming the mean GPA is lower than 3.0

h. None of above

Question 6 10 out of 10 points

For this question, refer to lecture 15. Consider the courtroom example and use the terminology and notation as in class. A particular jury has overseen a 105 innocent cases and a 100 guilty cases. The jury correctly judged 95 of the 105 innocent cases and 98 of the 100 guilty cases. The remaining cases were judged incorrectly. What is the precision of the jury.

Selected Answer: 98/108

Answers:

None of the listed.

98/100

95/100

98/108

95/97

Question 7 10 out of 10 points

Consider a distribution with two parameters, x and y. The mean of the distribution is x+1 and the variance 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: 0.17

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.17	

Question 8

10 out of 10 points

Given two samples X={0,1,0,0,1,0,1,0}, Y={1,1,1,1,1,0,0,1} sampled i.i.d. from two independent Bernoulli distribution Bern(p1) and Bern(p2) respectively, consider Wald's test to test if p1>=p2, i.e. H0: p1>=p2, H1: p1<p2. Use MLE as your estimator. Calculate the p-value in this problem. Report the result as a decimal rounded to two digits after the decimal point (ex: 14.466 is rounded and reported as 14.47, 1.103 is rounded and reported as 1.10). The entries in the Z score table below represent the area to the left of the corresponding Z score, i.e. , integration from negative infinity to Z of the p.d.f. of the Normal distribution. Z score of current entry is the summation of row head and column head, for example at the entry (0.1, .01), the value is .54380, which means the integration from negative infinity to 0.11 is 0.54380.

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Selected Answer: 0.05

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity

172.30.32.12

Exact Match	0.05

Question 9 10 out of 10 points

A car is being driven on a straight road. The following table gives its position (y) with respect to time (in seconds).

T	0	1	2	3	6	10
y	0	1	4	5	9	12

Use a simple EWMA to estimate position y at T = 4 seconds. Set $\alpha = 0.9$, $\hat{y}_0 = y_0$, where $y0$ is the position at T = 0 seconds. Report the result as a decimal rounded to two digits after the decimal point (ex: 14.466 is rounded and reported as 14.47, 1.1 is rounded and reported as 1.10).

Selected Answer: 4.87

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	4.87	

Question 10 10 out of 10 points

Given the following data, when using multiple linear regression to fit the equation $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_0$, we need to use the formula we learned from lectures, $\vec{\beta} = (X^T X)^{-1} X^T Y$.

What is the size of X^T in this problem? Refer to notation in class.

Y	X1	X2	X3
1	3	1	1
4	6	5	1
3	4	2	1
7	8	9	1

Selected Answer: b. 4*4

Answers:
a. 4*3
b. 4*4
c. 3*4
d. 5*4

Question 11 10 out of 10 points

Consider the Z test to test whether H0: mu<=1, H1: mu>1 for a large data {X1,X2,...,Xn} i.i.d. sampled from an unknown distribution with unknown mean mu and known deviation sigma. Assume the test is applicable. The test statistic is Z=-2 Which of the dark shaded areas in the following graphs can represent the p-value in this hypothesis test.

fig1

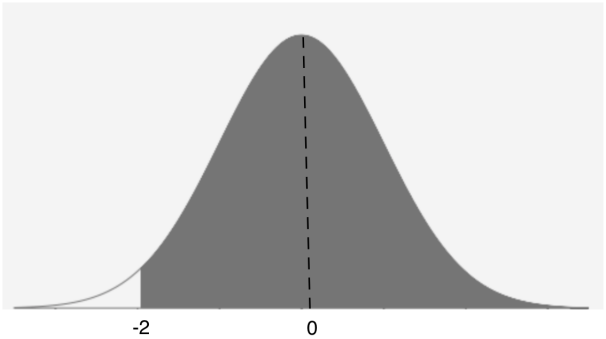


fig2

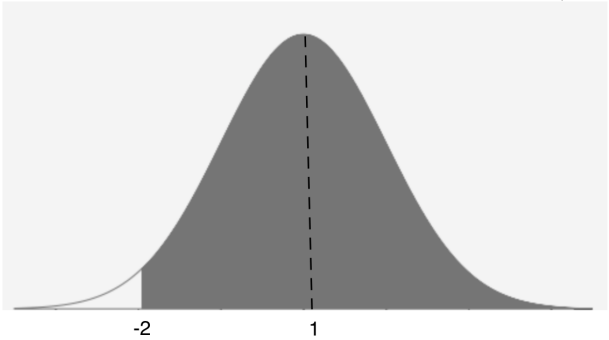
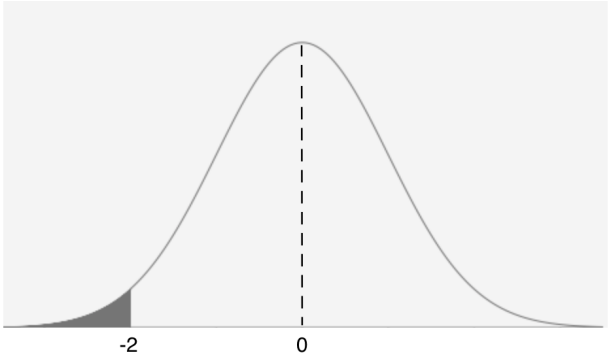
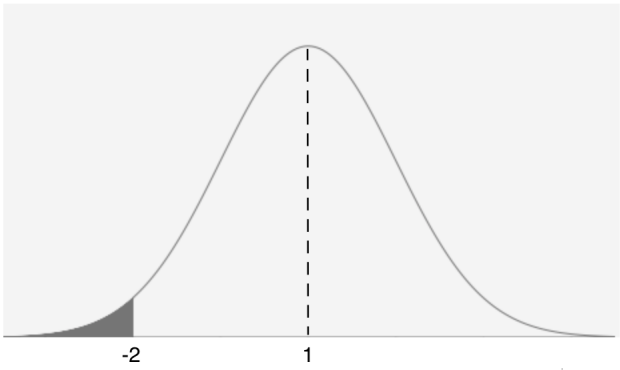


fig3

fig4





Selected Answer: a. fig1

Answers: a. fig1
b. fig2
c. fig3
d. fig4

Question 12

10 out of 10 points

Use K-S test to report the K-S statistic (max difference) when comparing X={0.9, 0.5, 0.1} with Unif(0, 1). 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: 0.23

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.23	

Question 13

10 out of 10 points

Given the set of (Y_i, X_i) samples: {(1, 2), (2, 4), (1, 3), (4, 6)} and the equation Y = X - 2 as the line of best fit (regression fit), the MAPE statistic averaged over these data points is __. Report the result as a decimal rounded to two digits after the decimal point (ex: 14.466 is rounded and reported as 14.47, 1.103 is rounded and reported as 1.10).

Selected Answer: 25.00

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	25.00	

Question 14

10 out of 10 points

Consider the simple linear regression but with the predicted fit being $\hat{Y} = (\hat{\beta})^2 X$, and given $\hat{\beta} \neq 0$. Note that there is no intercept term here. Solve for the OLS estimate of $\hat{\beta}$ 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: 0.66

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.66	

Question 15

10 out of 10 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? Use the uncorrected sample standard deviation (i.e., with n in the denominator and not (n-1)).

Selected Answer: 2.5

Answers:
- 2
- None of the listed
- 1.75
- 1.25
- 2.5

Question 16

10 out of 10 points

Let X={3, 5} and Y={1}. Use Permutation test with the statistic as absolute difference in 1st element of each set. For example, for the given X and Y, statistic would be |3 - 1| = 2. The p-value, rounded to two decimal places and including 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) for the given data is [FILL IN BELOW].

Selected Answer: 0.33

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	0.33	

Question 17

10 out of 10 points

Refer to the Pearson Correlation test from class and use the thresholds from class to decide the correlation between X={0, 2} and Y={3, 1}.

Selected Answer: Negative linear correlation

Answers:
- Positive linear correlation
- Negative linear correlation
- No correlation
- None of the listed

Question 18

10 out of 10 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 it is always a success. Thus, on day 1, you see {1} as your data. On day 2, you see a new {1} as your data, so your full data up to day 2 is {1, 1}. Likewise you see a new {1} on day 3 so your full data on day 3 is {1, 1, 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 {1} 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 strictly greater than 0.8?

Selected Answer: 4

Answers:
- None of the listed
- 3
- 4
- 2
- 1

Question 19

10 out of 10 points

Consider a case where 20 females voted for party A and 30 females voted for party B. In the same election, 40 males voted for party A and 10 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: 1.67

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
Exact Match	1.67	

Question 20

10 out of 10 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: $b \neq b_0$ (i.e., H_1 is b not equal to b_0). Using Wald's test, for what value of b_0 can you NOT reject the null hypothesis given {1, 2, 4, 5} as your sample data. Assume $z_{\alpha/2} = 1.96$. Hint: A4, Q2 can help.

Selected Answer: 4.5

Answers:
- -1.5
- None of the listed
- 4.5
- 6
- 0

Question 21

10 out of 10 points

Which of the following is NOT true.

Selected Answer: The paired t-test requires that both sets of samples be Normally distributed.

Answers:
- Wald's test for 2 populations requires that both estimators be Asymptotically Normal.
- The Z-test is unrealistic because true standard deviation is rarely known.
- The paired t-test requires that both sets of samples be Normally distributed.
- The Permutation test has no assumptions.
- The K-S test has no assumptions.

Question 22

10 out of 10 points

Perform a 2-sample KS test and find the max difference statistic for X={8, 2} and Y={1, 6, 8, 8}.

Selected Answer: 0.25

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

Question 23

10 out of 10 points

You are given data samples {2, 2, -2, -2} which you know to be Normally distributed with some unknown mean but variance 1. Compute the t-statistic for this data assuming a t-test for checking whether the true mean is 0.2. Use the uncorrected sample standard deviation (i.e., with n in the denominator and not (n-1)).

Selected Answer: -0.2

Answers:
-0.4
None of the listed
-0.8
-0.1
-0.05
✓ -0.2

Question 24 10 out of 10 points

Consider the Unif(a, b) distribution. Define H as (a+b+ab)/4. Find the MLE of H for data {7, 3, 2, 5, 8}. 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: ✓ 6.50

Correct Answer:

Evaluation Method	Correct Answer	Case Sensitivity
✓ *Exact Match*	6.50	

Question 25 10 out of 10 points

Consider a probability distribution X that takes values 10 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: ✓ $(X_b - 1)/9$

Answers:
None of the stated
$X_b/9$
✓ $(X_b - 1)/9$
X_b
$X_b/10$

Saturday, May 14, 2022 3:59:02 PM EDT

← OK