Stat/Math 390, Spring, Test 2, May 14, 2021; Marzban

READ the following paragraph!!

8+6+6=20

BY SUBMITTING THIS TEST, I HEREBY PLEDGE ON MY HONOR THAT I HAVE TAKEN IT IN PERSON AND WITHOUT ASSISTANCE FROM ANY OTHER PERSON. I ACKNOWLEDGE THAT THE PENALTY FOR VIOLATING ACADEMIC INTEGRITY IS MOST SEVERE.

- This is an online test given during the Corona Virus outbreak.
- For questions that have only 1 correct response, the options will appear as circles.
- For questions that may have multiple response, the options will appear as squares.
- It is open book/web/hw/solutions/past_tests/calculator/etc, but closed collaboration.
- The questions are presented to students in random order.
- You have the option of changing your answer to an already answered question.
- The list of questions and the remaining time for the whole test appear at the bottom of each page.
- Questions 1-8 are worth 1 point and do not require much calculation or writing.
- Questions 9-11 are worth 2 points each, and require a bit more work.
- Questions 12-13 are worth about 2 or 3 points and require varying levels of calculation, all of which are to be done on paper (or Tablet), and saved/scanned/photographed, and uploaded to canvas before 3:20. For these non-multiple-choice question, SHOW WORK. NO CREDIT FOR CORRECT ANSWER WITHOUT EXPLANATION/WORK. To save time, you may upload a single file that contains the solutions to all of the file-upload questions. But make sure you upload that file for all file-upload questions.

hw-lect 9-2 → $\mu_x = \frac{1}{\lambda}$, $\sigma_x = \frac{1}{\lambda}$

1 **1.** For the exponential distribution, as the mean increases, the width of the distribution
a) becomes narrower b) remains constant c) becomes wider d) depends on sample size.

lect 9

1 **2.** Even though, in practice one doesn't take multiple samples from a population, suppose someone actualy did take 3 samples of size n from a given population, and got the following sample means; -1.2, 0, 1.2. Then, the distribution mean is ← need the dist. p(x) or f(x).
a) 0 b) 0 only if the distribution is symmetric
c) 0 only if the distribution is normal d) None of the above.

3.14

1 **3.** They say that there exists a correlation between the price of a violin and who made it. For example, violins made by Stradivari are believed to be more expensive than those made by Guarneri, followed by Rugierri, and followed by those made by Amati. To quntify such claims, one can compute
a) Scatterplot of violin prices vs. violin makers. b) Pearson's correlation coefficient, r.
c) The R^2 of a regression model of violin makers and violin prices. d) None of the above.

categorical.

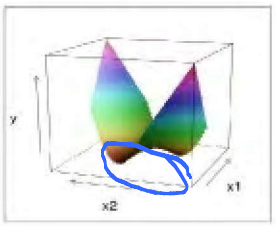
lect 10, autumn 19 → $R^2 = (r)^2$

1 **4.** In a regression problem $y = \alpha + \beta x$, which of the following is/are true if x and y are switched?
a) $r \to r$ b) $R^2 \to R^2$ c) $\hat{\beta} \to 1/\hat{\beta}$ d) None of the above.

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \to \frac{S_{xy}}{S_{yy}} \neq 1/\hat{\beta}$ $\hat{\beta} \neq \beta$!

lect 14, winter 17

1 **5.** Suppose in a problem involving two predictors x_1, x_2, and one response y, the data (not shown) closely reside on the surface shown in this figure; Which of the following conditions is/are likely to exist?



a) interaction ← saddle
b) collinearity ← look at the x_1-x_2 plain

hw-lect 17-2

1 **6.** Consider a sample of size n from a Bernoulli distribution with parameter π. Then (select the correct statement(s).
a) The sample mean is equal to the sample proportion (of 1's). $\bar{x} = p$
b) The sample mean is equal to π.
c) The expected value of the sample mean is equal to π. $E[\bar{x}] = E[p] = \pi$
d) The expected value of the sample mean is equal to the sample proportion (of 1's).

Important practical Moral: Even in the simplest situation where you expect/believe a (linear) plain to be adequate for your problem, when you're choosing what predictors to include in your study,do NOT pick the ones with high corr. with y!

1 **7.** Consider the CI for the population mean μ_x. If we were to increase the confidence level from 95% to 99%, then (select the true statement(s))

a) There is a higher chance that the observed CI will cover μ_x. ✗

b) We can be more confident that the observed CI covers μ_x.

c) A random CI will cover μ_x with higher probability.

} 2 interpretations of CI.

d) There is a 99% probability that a random $\bar{x}$ will fall in the observed CI. (see hw-lect16-2)

1 **8.** Suppose the CI for some (strictly positive) population parameter θ is given by (L, R). Then, the CI for $\frac{1}{\theta}$ is given by $L<\theta<R \Rightarrow \frac{1}{R}<\frac{1}{\theta}<\frac{1}{L}$

a) (R, L) b) $(\frac{1}{R}, \frac{1}{L})$ c) $(\frac{1}{L}, \frac{1}{R})$ d) none of the above.

hw-lect9-4, spring19

2 **9.** On test1 we found the n^{th} percentile of $Exp(\lambda)$. If we make a qq-plot of a very large sample that comes from such a distribution, and if we place the percentiles of $Exp(\lambda = 1)$ on the x-axis, the y-intercept and the slope, respectively, of the qq-plot will be $u(\lambda) = -\frac{1}{\lambda}\log(1+\frac{n}{100})$ } $\Rightarrow u(\lambda) = \frac{1}{\lambda}u(1)$

$u(1) = -\log(1+\frac{n}{100})$ }

a) $(0, 1)$ b) $(0, \frac{1}{\lambda})$ c) $(\frac{1}{\lambda}, \frac{1}{\lambda})$ d) $(\frac{1}{\lambda}, 0)$

Spring 2017

2 **10.** Select the true statement(s). Suppose we have selected two independent/uncorrelated predictors x_1, x_2, for the prediction of a response y. Also, suppose there are no clusters and outliers. We then notice that one of the predictors is very highly (and linearly) correlated with the response, then (Hint: this is a 3d visualization problem.)

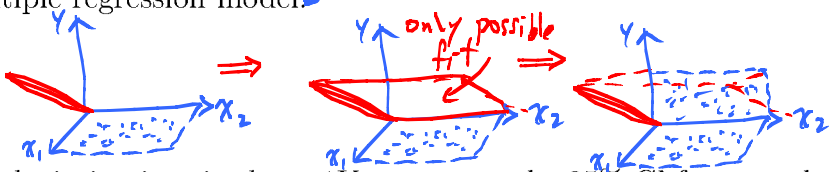
see lect.

a) that's ideal for developing a multiple regression model.

b) the other predictor is useless.

c) there is collinearity.

d) there is interaction.



lect17

2 **11.** In a problem where the typical deviation in x is about 1Kg, we want the 85% CI for μ_x to have a half-width of 0.2Kg. Then, the sample size should be about

a) 7 b) 10 c) 52 d) 96

$CI: \bar{x} \pm z^* \frac{\sigma_x}{\sqrt{n}}$

$\Rightarrow \bar{x} \pm 1.44 \frac{1}{\sqrt{n}} \Rightarrow \frac{1}{\sqrt{n}} = .2 \Rightarrow n = \left(\frac{1.44 \times 1}{.2}\right)^2 = 52$

hw-lect13-2, hw-lect14-1

~ 3 **12.** Consider the multiple regression model $y = \beta_1 x_1 + \beta_2 x_1 x_2 + \epsilon$. Find the Normal equations of regression (i.e., the equations that must be set to zero in order to minimize the SSE), and write them in "bar" notation. DO NOT solve them.

0.5+0.5 $SSE = \sum_i \epsilon_i^2 = \sum_i (y_i - \beta_1 x_{1i} - \beta_2 x_{1i} x_{2i})^2$

bar notation 0.5+0.5 ±0.25

deriv. 0.5+0.5

$\frac{\partial}{\partial \beta_1}: \sum_i (y_i - \beta_1 x_{1i} - \beta_2 x_{1i} x_{2i}) x_{1i} \Rightarrow \overline{x_1 y} - \hat{\beta}_1 \overline{x_1^2} + \hat{\beta}_2 \overline{x_1^2 x_2} = 0$

$\frac{\partial}{\partial \beta_2}: \sum_i (y_i - \beta_1 x_{1i} - \beta_2 x_{1i} x_{2i}) x_{1i} x_{2i} \Rightarrow \overline{x_1 x_2 y_2} - \hat{\beta}_1 \overline{x_1^2 x_2} - \hat{\beta}_2 \overline{x_1^2 x_2^2} = 0$

hw-lect16-3

~ 3 **13.** Suppose we have a distribution whose mean and variance are α and $(1-\alpha)^2$, respectively. I'm not going to give you the name of the distribution because you may panic! But assume that for samples of size n taken from this distribution, the sampling distribution of $\bar{x}$ is Normal. Find the CI (at arbitrary confidence level) of the α parameter. Do NOT worry about which is the lower end and which is the upper end.

0.5+0.5

$\mu_x = \alpha, \sigma_x = 1-\alpha \Rightarrow \bar{x} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}}) = N(\alpha, \frac{1-\alpha}{\sqrt{n}}) \Rightarrow z = \frac{\bar{x}-\alpha}{(1-\alpha)/\sqrt{n}} \sim N(0,1)$

$pr(-z^* < z < z^*) = \text{Conf. level} \Rightarrow$ Solve for α in $\frac{\bar{x}-\alpha}{(1-\alpha)/\sqrt{n}} = \pm z^*$ 0.5+0.5 ±0.25

Our TA (Qiliang) has shown that this answer assumes a number of things about $\bar{x}, \alpha, n$, etc. But we graded the method!

$\therefore \bar{x} - \alpha = \pm \frac{z^*}{\sqrt{n}}(1-\alpha) \Rightarrow \bar{x} \mp \frac{z^*}{\sqrt{n}} = (1 \mp \frac{z^*}{\sqrt{n}})\alpha \Rightarrow \alpha = \frac{\bar{x} \mp z^*/\sqrt{n}}{1 \mp z^*/\sqrt{n}}$

$\frac{\bar{x} - z^*/\sqrt{n}}{1 + z^*/\sqrt{n}} < \alpha < \frac{\bar{x} + z^*/\sqrt{n}}{1 - z^*/\sqrt{n}}$. Solve for α: 0.5+0.5