Name: ______________ problem 1–7

ID: ______________

Quiz section or time: ______________

7 + 12.5

Stat/Math 390, Winter, Test 2, February 21, 2014; Marzban

Same deal as test 1, ...

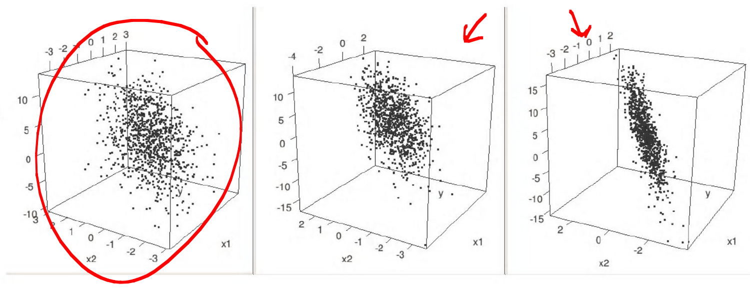
Points

1 Lect 8, 9 **1.** Circle all of the statements that are **generally** true regarding the correlation coefficient, r_{xy} between x and y.

a) The only way r_{xy} can be 0 is if the scatterplot of y vs. x consists of random scatter of points.
a) As the slope of a linear relation in y vs. x increases, so does r_{xy}.
c) For a quadratic regression model for (x, y), the proportion of the variation in y explained by x, is equal to r_{xy}^2.
d) None of the above.

more collinear

1 Lect 13 and Lab **2.** Consider the following three data sets on (x_1, x_2, y). Although it may not be clear in these figs, the relationship between the response y and the predictors (x_1, x_2) is planar in all three sets. Circle the one which is most suitable for a multiple regression fit.



1 Lect 11 and Sample test **3.** Which of the following models/fits is NOT possible using the methods of linear regression learned in this class?

a) $y = \alpha + \beta \log(x)$ b) $x = e^{(y-\alpha)/\beta}$ c) $e^y = \alpha x^\beta$ d) $2^y = \alpha x^\beta$ e) None of the above.

$Y = \alpha + \beta \ln x$ $Y = \ln\alpha + \beta \ln x$ $Y = \log_2 \alpha + \beta \log_2 x$ or $y \ln 2 = \ln \alpha + \beta \ln x$

1 Lect 13 **4.** Which of the following statements is **generally** correct? In the model

a) $y = \alpha + \beta_1 x + \beta_2 x^2$, β_1 measures the average change in y when x changes by one unit.
b) $y = \alpha + \beta_1 x_1 + \beta_2 x_2$, β_1 measures the average change in y when x_1 changes by one unit. if x_2 = Const
c) $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, β_1 measures the average change in y when x_1 changes by one unit, and x_2 is held fixed.
d) None of the above.

1 Lect 17 **5.** Circle all the correct answers. Let A denote the event that it will rain today, and B the event that it will rain tomorrow. Also, suppose $p(A)$ and $p(B)$ are nonzero. Based on this information, and nothing else, then, in general,

rain today does not preclude rain tomorrow.

a) A and A^c are mutually exclusive c) A and B are mutually exclusive.
b) A is independent of A^c d) A is independent of B.

$P(A^c|A) = 0$, not $P(A^c)$

1 5.69, 5.70 **6.** A quantity is known to have an exponential distribution with mean and standard deviation equal to 2 and 2, respectively. The std. dev. of the distribution of the **sum** of three measurements is

a) 3×2 b) $\sqrt{3} \times 2$ c) $2/\sqrt{3}$ d) Depends on the sample.

$T = x_1 + x_2 + x_3 \Rightarrow V[T] = V[x_1 + x_2 + x_3] = 3V[x] \Rightarrow \sigma_T = \sqrt{3}\,\sigma_x = \sqrt{3}\,(2)$. Alternatively,

1 class example **7.** 130 students took test 1 this quarter in this class. We can build the CI for the population mean grade. In one sentence, describe the population in question.

$V[\bar{x}] = \frac{\sigma_x^2}{n}$
$\parallel$
$V[\frac{T}{n}]$
$\frac{1}{n^2} V[T] = \frac{\sigma_x^2}{n}$
$V[T] = n\sigma_x^2, \ldots$

The population is all of the students who have ever taken 390 (and ever will), etc.

e.g. with Marzban, ...

example in lect 10

~ 2 **8.** In a regression problem involving 2 predictors, the errors are found to be -2, +1, 0, -1, 2. The deviation of the predictions from $\overline{y}$ are +1, -1, 0, -1, +1.

a) What is the value of s_e, AND its meaning?

$$s_e^2 = \frac{SSE}{n-(k+1)} = \frac{4+1+0+1+4}{5-(2+1)} = \frac{10}{2} = 5 \implies \underline{s_e = \sqrt{5}} \simeq 2.3$$

(3)

meaning: Typical deviation (error) of data from the fit (prediction) is about 2.3.

~ 2 **b)** What is the value of R^2, AND its interpretation?

$$R^2 = \frac{SS_{expl}}{SST} = \frac{\sum_i (\hat{y}_i - \overline{y})^2}{SS_{expl} + SSE} = \frac{1+1+0+1+1}{\cdots + 10} = \frac{4}{14} = \underline{\frac{2}{7}} \sim 0.3$$

part a.

Interpret: About 30% of the variability in y is explained by (or attributed to) x_1, x_2.

3.48 and Lect 9 (on r)

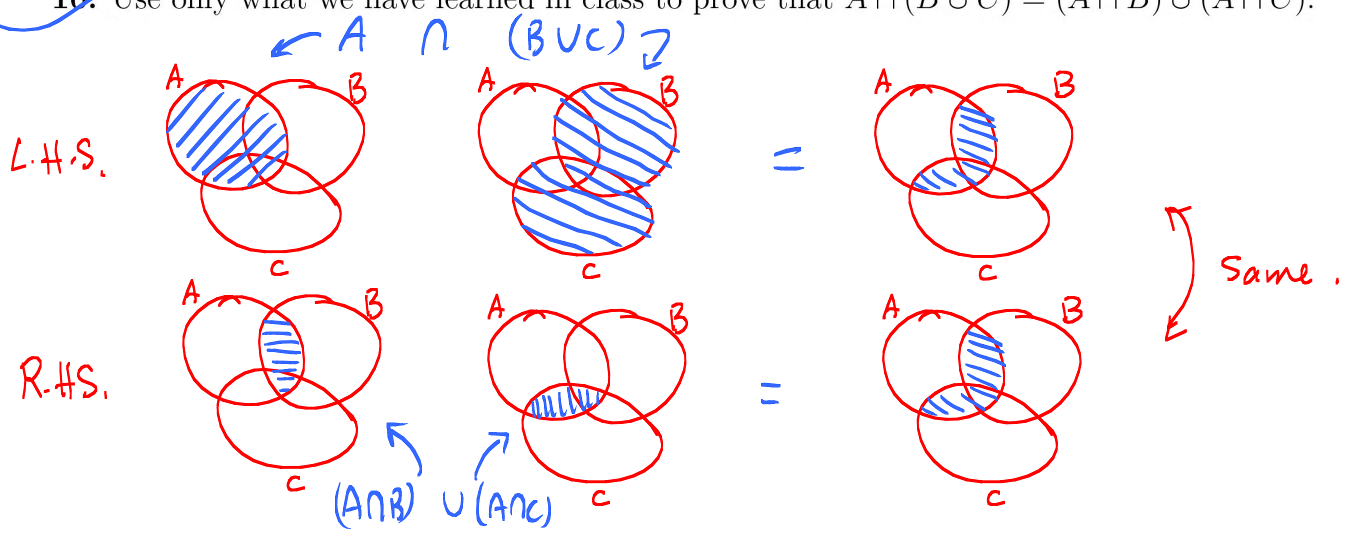
~ 2 **9.** Let $\hat{\beta}$ denote the OLS estimate of the slope parameter in a simple linear regression fit to (x, y) data. If all the x_i data are scaled to $x_i' = 2x_i$, and the y_i data are scaled to $y_i' = 3y_i$, show that the OLS estimate of the slope parameter in $y' = \alpha' + \beta' x$ is given by $(3/2)\hat{\beta}$.

$$\hat{\beta}' = \frac{\overline{x'y'} - \overline{x'}\,\overline{y'}}{\overline{x'^2} - \overline{x'}^2} = \frac{\frac{1}{n}\sum_i x_i' y_i' - \cdots}{\frac{1}{n}\sum_i (x_i')^2 - \cdots}$$

$$= \frac{(2)(3)\frac{1}{n}\sum_i x_i y_i - 2\overline{x}\,3\overline{y}}{2^2\frac{1}{n}\sum_i x_i^2 - 2^2\overline{x}^2} = \frac{2(3)}{2^2}\hat{\beta} = \frac{3}{2}\hat{\beta}.$$

See hw-Y

~ 2 1.5 **10.** Use only what we have learned in class to prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.



572, 573

~ 3 **11.** The USA has about 100 Small colleges, 200 Medium-size colleges, and 900 Large colleges. Among Small colleges 1/6 of the students are Boys, while the proportion of Boys in Medium-sized and Large colleges are 1/6 and 5/6, respectively. If a random student is selected from the USA and found to be a Boy, what is the probability that he came from a Large college? CLEARLY define the events of interest, and EXPLAIN every step. Do NOT assume Boys and Girls are equally likely.

Let S = The event that student is from Small College.
M = ..., L = ...
B = The event that student is Boy.

5/6   900/(100+200+900)   Way 1 (or 2) for calc. probs.

$$? = P(L \mid B) \overset{\text{Bayes}}{=} \frac{P(B \mid L)\,P(L)}{P(B)}$$

$$P(B) \xrightarrow{\text{Trick}} P(B \mid L)P(L) + P(B \mid M)P(M) + P(B \mid S)P(S)$$

$$\frac{5}{6}\cdot\frac{900}{1200} + \frac{1}{6}\cdot\frac{200}{1200} + \frac{1}{6}\cdot\frac{100}{1200}$$

$$= \frac{\frac{5}{6}\cdot\frac{9}{12}}{\frac{5}{6}\cdot\frac{9}{12} + \frac{1}{6}\cdot\frac{2}{12} + \frac{1}{6}\cdot\frac{1}{12}} = \frac{45}{45+2+1} = \frac{45}{48} = \boxed{\frac{15}{16}}.$$

hw-F, 7.7, 7.17

~ 3 **12.** A sample of size 25 has been taken from a normal population, and the sample mean and standard deviation are found to be 0.1 and 1.0 respectively. What is the confidence level at which we can conclude that there is evidence that μ exceeds zero? Hints: 1) think about a lower confidence bound, 2) find the z^*, and then 3) find the confidence level.

Lower conf. bound for μ_x : $\bar{x} - z^* \frac{s}{\sqrt{n}}$, where z^* s.t.

$\text{Prob}(z < z^*) = \text{Conf. level}.$

$P\left(\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} < z^*\right) = \text{Conf. level}$
$\vdots$
$\mu_x > \bar{x} - z^* \frac{\sigma_x}{\sqrt{n}}$

$\therefore\ \bar{x} - z^* \frac{s}{\sqrt{n}} > 0$

μ_x 0 $\bar{x} - z^* \frac{s}{\sqrt{n}}$

$$\therefore\ z^* < \frac{\bar{x}\sqrt{n}}{s} = \frac{0.1(5)}{1.0} = \frac{1}{2}$$

$$\therefore\ \text{Conf. level} = \text{prob}\left(z < \tfrac{1}{2}\right) = \boxed{0.6915}$$ Table I.