(5) No b = 0
Some b = 0.5
Only b = 1

(6) No c = 0
Some c = 0.5
Only c = 1

(7) a, b, d, ab, ad, bd = 0
bc, cd = 0.5
c = 1

Name: ______________________

ID: ______________________

Quiz section or time: _____9+11_____

Stat/Math 390, Winter, Test 2, Feb. 17, 2012; Marzban

CLOSED everything. ONLY a half-size "cheat sheet" is allowed. No magnifying lens. Check front page and back page.

Multiple-choice: mark answers on these pages. DO NOT EXPLAIN.

The rest: SHOW answer & WORK on these pages; NO CREDIT FOR CORRECT ANSWER WITHOUT EXPLANATION

Points

1 2.34 **1.** x is known to have a normal distribution with parameters μ and σ. Which one of the following is sufficient to determine their values? Its
a) 50^{th} percentile b) 68^{th} percentile c) 1^{st} and 5^{th} percentiles d) None of the above
Takes 2 eqns to solve for 2 unknowns.

1 2.51, Sample Test 1, and Lab 2 **2.** For which of the following distributions can the idea of a qq plot be used to test the distribution?
a) Normal b) Uniform c) Exponential d) All of the above.

1 5.45 **3.** For which one of the following one canNOT compute a sampling distribution?
a) sample standard deviation b) sample maximum c) population mean d) none of the above
parameter

1 3.16 **4.** Suppose the r between x and y (both positive) is very low. Assuming there are no clusters or outliers, the r for which of the following may be large? Between
a) x and y^2 b) x^2 and y c) x and $\log(y)$ d) $\log(x)$ and y e) All of the above

1 3.* **5.** The relationship between x and y is known to be generally linear, with a finite, nonzero slope. Then, the r between them (multiple answers possible, with penalty)
a) is ± 1 b) depends on the scatter between x and y c) depends on the slope

1 3.36c **6.** To predict height (y) in meters, from weight (x_1) in pounds, and arm-length (x_2) in meters, data is collected. Let r_{uv} denote the correlation coefficient between u and v. It is found that the $r_{yx_1} = 0.7$ $r_{yx_2} = 0.8$, and $r_{x_1x_2} = 0.2$. The OLS-based regression model $y = 0.1 + 0.3x_1 + 0.4x_2 + 0.5x_1x_2$ is found adequate (i.e., $R^2 = 0.9$). Then (multiple answers possible, with penalty)
a) On the average, height increases by 0.3 m, for every pound increase in weight. Interaction
b) On the average, height increases by 0.7 m, for every pound increase in weight. No interpretation for r.
c) On the average, 90% of the variability in height can be accounted for by weight and arm-length.
d) The large R^2 implies that overfitting has probably occurred. Overfitting ⇒ large R^2, but not vice versa.
e) None of the above.

1 7.86+7.50b **7.** The 95% lower confidence bound for a population proportion is 0.2. Then, with 95% confidence, one can say that the true proportion (multiple answers are possible, with penalty)
a) is lower than 0.2 b) exceeds 0.1 c) exceeds 0.2 d) exceeds 0.3
Not with 95% Confidence

1 7-11 **8.** A 95% 2-sided CI for μ is computed to be (10,12). Then which one of the following is correct?
a) 95% of sample means will be ~~between 10 and 12~~. within $\mu \pm 1.96\, \sigma/\sqrt{n}$ not $\bar{x}$.
b) We can be 95% confident that the true mean is less than 12. (10,12) = 2 sided, not 1-sided.
c) 95% of the CIs will ~~be (10,12)~~. cover μ.
d) We can be 95% confident that (10,12) is the ~~true CI~~. There is no such thing; CI = statistic.
e) None of the above.

1 7.16 7.* **9.** Which of the following is/are true, if the population is Normal, and sample size is large? The sampling distribution of the sample (multiple answers possible, with penalty)
a) mean is Normal b) proportion is ~~Binomial~~ Normal. c) standard deviation is Normal.

b = 0
a, c, ab, bc, abc = 0.5, ac = 1

2,3 3.37c or lect 13 (R^2)

10. In a regression problem, the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ is employed. The data contains 15 cases, the sample standard deviation of the y is 21, and the typical deviation of the data from the fitted surface is 7. Compute R^2 (in terms of numbers, not symbols.)

$$R^2 = \frac{SS_{expl}}{SST} = 1 - \frac{SSE}{SST} \Longrightarrow R^2 = 1 - \frac{[n-(k+1)]\, s_e^2}{(n-1)\, s_y^2}$$

$$s_e^2 = \frac{SSE}{n-(k+1)} \qquad = 1 - \left(\frac{15-6}{15-1}\right)\left(\frac{7}{21}\right)^2$$

$$s_y^2 = \frac{1}{n-1} SST \qquad = 1 - \frac{9}{14}\cdot\frac{1}{9} = \boxed{\frac{13}{14}}$$

2,3 hw J hw L **11.** In a hw you took random samples of size 10 and showed that if the sample mean of the 10 numbers is used as a prediction for each of the 10 numbers, then SSE is minimum. Here, you will prove that result, quite generally. Suppose you are given $y_i, i = 1, 2, ..., n$, and use a constant α as a prediction for each of the y_i. **a)** Write down the expression for SSE in terms of y_i and α. **b)** Then differentiate that expression with respect to α, and find the critical value of α which makes the derivative of SSE zero. Show work.

a) $SSE = \sum_{i=1}^{n} (y_i - \alpha)^2$

b) $$\frac{d}{d\alpha} SSE = \sum_i \frac{d}{d\alpha}(y_i - \alpha)^2 = -\sum_i 2(y_i - \alpha) = -2\Big(\sum_i^n y_i - \underbrace{\sum_i^n \alpha}_{n\alpha}\Big)$$

$$= -2n\left(\frac{1}{n}\sum_i y_i - \alpha\right) = -2n(\bar{y} - \alpha)$$

$$= 0 \Longrightarrow \boxed{\hat{\alpha} = \bar{y}}$$

2,3 5.5 LC **12.** Suppose the lifetime (x) of batteries has a Normal distribution with mean of 8 and std dev of 1. Let T denote the total lifetime of four batteries in a randomly selected package. Find the numerical value of T_0 for which prob$(T \geq T_0) = 0.95$. Show work!

$$0.95 = prob(T \geq T_0) = prob(n\bar{x} \geq T_0) = prob\left(\bar{x} \geq \frac{T_0}{n}\right) = prob\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq \frac{\frac{T_0}{n}-\mu}{\sigma/\sqrt{n}}\right)$$

$$0.95 = prob\Big(z \geq \underbrace{\frac{\frac{T_0}{n}-\mu}{\sigma/\sqrt{n}}}\Big)$$

Table I $\Rightarrow$ -1.645 $\rightarrow$ $\frac{T_0}{n} - \mu = -1.645\frac{\sigma}{\sqrt{n}} \Rightarrow T_0 = n\left(\mu - 1.645\frac{\sigma}{\sqrt{n}}\right)$

$$= 4\left(8 - 1.645\frac{1}{\sqrt{4}}\right)$$

$$= 4(7.18) = \boxed{28.7}$$

2,3 p.300 **13.** Starting from a self-evident fact, derive the formula for a 91% upper confidence bound for a population mean. Make sure the z^* in your answer has a numerical value. Show work.

Self-evident fact: prob($z >$?) = .91 $\quad -1.34 \quad$ (Table I)

$$\therefore prob\left(\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} > -1.34\right) = .91$$

$$\therefore \quad \bar{x} - \mu_x > -1.34\frac{\sigma_x}{\sqrt{n}} \Longrightarrow \mu_x < \bar{x} + 1.34\frac{\sigma_x}{\sqrt{n}}$$

$\therefore$ $\boxed{91\% \text{ upper bound: } \bar{x} + 1.34\frac{\sigma_x}{\sqrt{n}}}$