Name: ______________________

ID: ______________________

Quiz section or time: ______________________

Stat/Math 390, Winter, Test 3, March 14, 2014; Marzban

Same deal as test 1, ...

Points

1 **1.** To do a hypothesis test on (or build a CI for) a population parameter, we need to know the
a) sampling distribution of the population parameter, under H_0.
b) sampling distribution of the population parameter, under H_1.
c) sampling distribution of the sample statistic, under H_0.
d) sampling distribution of the sample statistic, under H_1.

1 **2.** A 95% CI for $\pi_1 - \pi_2$ is (-0.1, 0.8). Circle all correct statements. We can be 95% confident that
a) $\pi_1 = \pi_2$ b) $\pi_1 > \pi_2$ c) π_1 cannot exceed π_2 by 0.8 (or more) d) None of the above.

1 8.5 **3.** It is important that a car comes to a complete stop within 50 meters of when the breaks are engaged. Let μ denote the true/mean stopping distance. Data is collected on the stopping distance for 10 cars. Which is the appropriate quantity to compute?
a) lower confidence bound for μ b) upper confidence bound for μ c) 2-sided CI for μ
d) 2-sided CI for the true proportion of cars that stop within 50 meters.

1 **4.** The game of Roulette involves a horizontal spinning disc with 38 slots around it. A ball is thrown onto the disc, and eventually settles into one of the slots. It is suspected that the disc is uneven, and so the ball does not have an equal probability of falling into each of the slots. What is the best test for testing this suspicion?
a) t-test b) chi-squared c) 1-way anova F-test d) F-test of model utility

1 9.12 **5.** In 1-WAY ANOVA for testing k population means, $\mu_i (i = 1, k)$, which of the following is true?
a) The test requires sample means, and sample sizes, but not sample variances.
b) The sample sizes for the k samples must be equal.
c) The null hypothesis is $H_0 : \mu_1 = \mu_{01}, \mu_2 = \mu_{02}, ...,$ where μ_{0i} are the null parameters.
d) None of the above.

1 9.15 Lab **6.** In a study, each of four laboratories is asked to make 100 determinations of the percentage of methyl alcohol in a certain material. We want to test whether there is a difference between the average percentages reported by the laboratories. Which is the most appropriate test?
a) t-test b) chi-squared c) 1-way anova F-test d) F-test of model utility

1 **7.** A 95% PI for $y(x)$ will cover the mean of y ______ than 95% of the time. Hint: think of both the CI and the PI.
a) less often b) equally often c) more often

1 11.53 Lect27 **8.** In a multiple regression model without an interaction term we want to estimate, in a way that conveys information about precision and reliability, the average change in $y(x)$ associated with a 1-unit increase in x_1, when all of the other predictors remain fixed. Which is the appropriate quantity to compute?
a) CI for $\hat{\beta}_1$ b) CI for β_1 c) CI for $y(x)$ d) PI for $y(x)$
e) Not possible, because the absense of interaction implies collinearity.

~2 8.38 **9.** Elevated energy consumption during exercise continues after the workout ends. In a study, energy consumption was measured continuously for 30 minutes for each of 15 subjects both after a weight training exercise and after a treadmill exercise. We want to test whether average consumption after weight training exceeds that for the treadmill exercise by more than 5. a) Write the most appropriate H_0, H_1 in terms of well-defined quantities, and b) write the formula for the corresponding statistic.

μ_2 = true mean consumption after weight traning
μ_1 = " " " treadmill

$H_0: \mu_2 - \mu_1 \le 5$

$H_1: \mu_2 - \mu_1 > 5$

$t = \dfrac{\bar{d} - 5}{s_d/\sqrt{15}}$ where $d = \bar{x}_2 - \bar{x}_1$ [paired data]

~2 example 8.88 **10.** To test whether people's health status depends on the time of year, a study collects the number of admissions into a hospital within 30 days of individual's birthday inclusive (i.e., 30 days before and 30 days after), within 60 days of birthday, and more than 60 days of birthday. Write the appropriate H_0, H_1 in terms of clearly defined quantities. 1 year=365 days. Don't compute p-value.

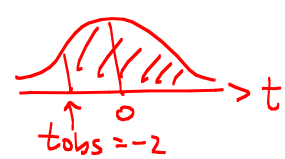
π_1 = prop. of people visiting hospital within 30 days of Birthday
π_2 = --- 60 ---
π_3 = more than 60 ---

30+1+30, 60+1+60, 365−61−121

$H_0: \pi_1 = \dfrac{61}{365}, \ \pi_2 = \dfrac{121}{365}, \ \pi_3 = \dfrac{183}{365}$; H_1 = At least one of These props is wrong.

~2 8.23 **11.** To be safe it is required that true average viscosity of some material exceed 30. Based on a sample of size 9, the sample mean and standard deviation are 28, and 3, respectively. Is the material safe at $\alpha = 0.05$? Clearly state the hypotheses in terms of well-defined quantities, report the p-value, and state the conclusion in "English".

$H_0: \mu \le 30$ μ = true mean viscosity.
$H_1: \mu > 30$



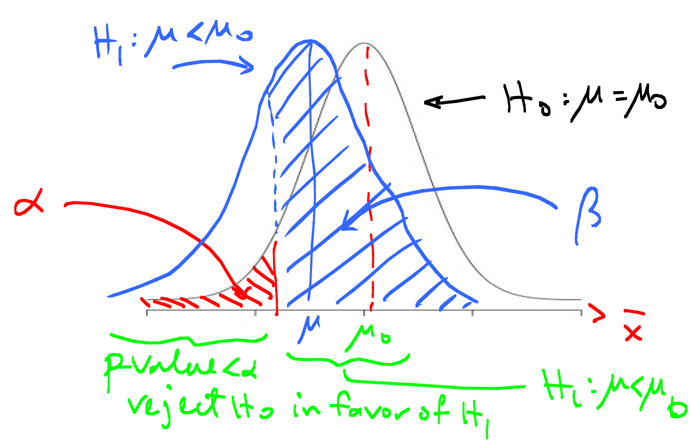
$t_{obs} = \dfrac{28 - 30}{3/\sqrt{9}} = \dfrac{-2}{1} = -2$ $df = n - 1 = 8$

p-value = prob($t > t_{obs}$) = prob($t > -2$) = $1 - .04 = .96 > \alpha$

∴ There is no evidence (at $\alpha = .05$) That The material is safe.

~2 hw-AK **12.** We are testing $H_0: \mu \ge \mu_0$ vs. $H_1: \mu < \mu_0$, at $\alpha = 0.05$. On the shown normal distribution
a) Label the x-axis $\bar{x}$ (or z or t)
b) Shade the appropriate area corresponding to α
c) Revise the diagram (or make a new one), clearly shading the area corresponding to β.

with $H_1: \mu < \mu_0$
p-value = prob($\bar{x} < \bar{x}_{obs}$)
∴ α = left-area



11.53 Lec 27

13. The following is the output from a regression analysis of a sample with $\bar{x} = 2$.

~ 2 a) Predict the mean value of $y(x)$ when $x = 4$, in a way that conveys reliability and precision, at 95% confidence level. HINT: recall what we did in class when we didn't have S_{xx}. Use $t^* = 2.3$, and $\sqrt{1.1} = 1.05$.

Predictor	Coef	StDev	T	p
Constant	0.6	0.2	3.0	0.017
x	0.2	1.0	0.2	0.846

Analysis of Variance:

Source	DF	SS	MS	F	P
Regression	1	39.4	39.4	9.85	0.986
Error	8	32.0	4.0		
Total	9	64.4			

$n-2$ SSE $\frac{SSE}{n-2} = s_e^2 \Rightarrow \boxed{s_e = 2}$

$n-1 \Rightarrow \boxed{n=10}$

Hint $\rightarrow s_{\hat\beta} = \frac{s_e}{\sqrt{S_{xx}}} \Rightarrow 1.0 = \frac{2}{\sqrt{S_{xx}}}$

$\therefore \boxed{S_{xx} = 4}$

C.I. for mean $y(x)$:

$\hat{y}(x) \pm t^* s_e \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

$= \hat\alpha + \hat\beta x$

$[0.6 + (0.2)4] \pm (2.3)(2)\sqrt{\frac{1}{10} + \frac{(4-2)^2}{4}}$

$1.4 \pm (2.3)(2)\underbrace{\sqrt{\frac{1}{10}+1}}_{\sqrt{1.1} = 1.05}$

$\boxed{1.4 \pm (2.3)\boxed{(2)(1.05)}}$

~ 2 b) What is the probability that the estimation error will exceed 2.10?

$\text{prob}(\text{est. err.} > 2.10) = \text{prob}\left(\frac{\text{est. err.}}{s_{\text{est. err}}} > \frac{2.10}{s_{\hat{y}}}\right)$

standardize

$= \text{prob}\left(t > \frac{2.1}{2.1}\right) = \text{prob}(t > 1) \overset{df=n-2=8}{\doteq} \boxed{.173}$

~ 2 c) What is the probability that the prediction error will exceed 2.53? Hint: $\sqrt{2.1^2 + 2.0^2} = 2.53$

$\text{prob}(\text{pred. err.} > 2.53) = \text{prob}\left(\frac{\text{pred. err}}{s_{\text{pred. err}}} > \frac{2.53}{\sqrt{s_{\hat{y}}^2 + s_e^2}}\right)$

$= \text{prob}\left(t > \frac{2.53}{\sqrt{2.1^2 + 2^2}}\right) = \text{prob}(t > 1) \overset{df=n-2=8}{\doteq} \boxed{.173}$
