12+15

Name: ______________________

ID: ________________________

Quiz section time: ________________________

Stat/Math 390, Spring, Final Exam- June 10, 2008; Marzban
Open everything, closed messaging/discussion
Check pages 1, 2, and 3. SHOW WORK !!!

Points

3.42a

2 **1.** The following joint mass function is given. Is x independent of y?

		y=0	y=1	y=2	
x	1	0.125	0.125	0.25	0.5
	2	0.125	0.125	0.25	0.5
		0.25	0.25	0.5	$h(y)$

$g(x)$

a) Yes. $f(x,y) = g(x)\,h(y)$.
b) Cannot tell because it is necessary to compute ρ.
c) Cannot tell because the joint distribution is not given.
d) No.

7.501

2 **2.** In regression, the estimate of the regression coefficient β is more precise if
a) the errors have a smaller variance.
b) the x values have a wider variance.
c) a and b.
d) None of the above.

$V[\hat\beta] = \frac{\sigma^2}{S_{xx}}$, $S_{xx} = (n-1)S_x^2$

7.52b

2 **3.** Data is collected on the concentration of harmful exhaust gasses both inside and outside a sample of cars in a factory. The manager of the factory asks you "For the next car that comes off the assembly line, between what values would you expect the difference in concentrations to lie?" Which of the following 2-sided intervals should you compute?
a) A CI for the difference, for unpaired data.
b) A PI for the difference, for unpaired data.
c) A CI for the difference, for paired data.
d) A PI for the difference, for paired data.

8.16

2 **4.** To obtain information on corrosion-resistance properties of a certain type of steel conduit, a sample of specimens are buried in soil for an extended period. The maximum penetration (in mils) is then measured for each specimen, yielding a sample mean penetration of 52.7 and a sample std dev of 4.8. The conduits were manufactured with the specification that the true average penetration be at most 50 mils. What should be the hypotheses if we want to see whether the sample data indicate that specifications have not been met?
a) $H_0 : \mu = 50 \quad H_1 : \mu \neq 50$
b) $H_0 : \mu \leq 50 \quad H_1 : \mu > 50$
c) $H_0 : \mu \geq 50 \quad H_1 : \mu < 50$
d) $H_0 : \mu \leq 50 \quad H_1 : \mu \neq 50$

Without data, I conclude H_0, i.e. specs have been met. Data tries to reject that.

8.46

2 **5.** Three different design configurations are being considered for a particular component. There are four possible failure modes for the component. An engineer obtains data on the number of failures in each mode for each of the configurations, and wants to see if the configurations appear to have an effect on type of failure. What is the best test?
a) A chi-squared test based on 1 population with multiple categories.
b) A chi-squared test based on multiple populations each with multiple (6) categories.
c) A 1-way ANOVA F-test.
d) An F-test of model utility in regression.

		failure 1	2	3	4
Config	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.

11.5

2 **6.** Assuming x=temperature and y=oxygen diffusivity are related by a simple linear regression model, we use the equation to compute a point estimate for mean diffusivity at a given temperature, T. How does this point estimate compare to a point prediction of the diffusivity value that would result from making one more observation at temperature T? It is
a) same.
b) larger.
c) smaller.
d) unrelated.

4 8.79 **7.** Two types of coating - type A and B - are being examined regarding their effect on the lifetime of a screw. The type A coating is applied to a sample of 7 screws, and the type B coating is applied to another sample of 7 screws. Use the accompanying data to decide at significance level 0.01 whether there is strong evidence for concluding that true average lifetime for the type B coating is more than four times that of the type A coating. Specifically, test the parameter $\theta = 4\mu_A - \mu_B$, by **1)** writing the hypotheses for it, **2)** computing a p-value, and **3)** stating your final conclusion in words referring to the two types of coating. Hint: It is known that $\hat{\theta} = 4\overline{x_A} - \overline{x_B}$ has a normal distribution with mean θ; but you find its variance yourself.

	Sample mean	Sample std. dev
Type A	12.4	7.8
Type B	53.0	38.3

$H_0: \theta \geq 0$

$H_1: \theta < 0$

$Z = \dfrac{\hat{\theta} - \theta}{\sqrt{V[\hat{\theta}]}} \in N(0,1)$

$V[\hat{\theta}] = 16\, V[\bar{x}_A] + V[\bar{x}_B] = 16 \dfrac{s_A^2}{7} + \dfrac{s_B^2}{7} = 348.6186$

$t = \dfrac{[4(12.4) - 53] - 0}{\sqrt{348.6186}} = \dfrac{-3.4}{18.67} = -0.18$

or Z

p-value $= prob(t < -0.18) = 0.4286$, df = Welch $\simeq 10$

There is insufficient evidence (at any reasonable α) to conclude that the true avg. lifetime for B is more than 4 times that of A.

3

4 9.14 **8.** An article in a scientific journal reports on an experiment in which 3 groups of rats consisting of 5 rats per group were put on diets with differing amounts of carbohydrates. At the conclusion of the study, the DNA content of the liver of each rat was determined. The following table shows the results of the experiment. The paper also reports that the total sum of squares for the experimental data was SST=850. Complete the ANOVA table for this study, and show your work, below.

Group	Source of Carb	Sample mean of DNA content
1	Starch	10
2	Sucrose	20
3	Fructose	15

$k = 3 \quad n_i = 5 \quad \Rightarrow n = 15$

$\bar{\bar{y}} = \sum_i^3 \dfrac{n_i}{n} \bar{y}_i = \dfrac{5}{15}(10 + 20 + 15) = 15$

Source of Variation	df	SS	MS	F	p-value
Between group	2	250	125	2.5	>0.1
Within group	12	600	50		
Total Variation	14	850			

k−1 250/2

n−k n−1 600/12 125/50

Table VIII.

$SS_{between} = \sum^k n_i (\bar{y}_i - \bar{\bar{y}})^2 = 5(10-15)^2 + 5(20-15)^2 + 5(15-15)^2 = 250$

df=7

$(t < \dfrac{3.44 - 3.5}{0.0229}) = prob(t < -2.6) = \boxed{.011}$

3 bt as hyp. test (instead of C.I.)

~~4~~ 11.20 **9.** The following is included in a printout from a regression analysis on x=temperature and y=diffusivity. Does the data provide sufficient evidence for concluding that the true mean diffusivity at $x = 1.5$ is less than 3.5? Specifically, **1)** State the hypotheses, **2)** compute a p-value, and **3)** state the conclusion "in words" at an alpha level of 0.01 .

$n = 9, \overline{x} = 1.4, var(x) = 0.075, y = -2.095 + 3.693\,x, s_e = 0.064$

$\hat{y}_{obs} = -2.095 + 3.693(1.5) = 3.44$

H_0: true mean of y at $x = 1.5 \geq 3.5$
H_1: " " " " " < 3.5

$s_{\hat{y}} = s_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = 0.064\sqrt{\frac{1}{9} + \frac{(1.5-1.4)^2}{8(.075)}} = 0.0229$, $S_{xx} = (n-1)s_x^2$

if $H_0 = T$

$p\text{-value} = pr(\hat{y} < \hat{y}_{obs}) = pr\left(\frac{\hat{y} - Y(x)}{s_{\hat{y}}} < \frac{\hat{y}_{obs} - Y(x)}{s_{\hat{y}}}\right) = pr\left(t < \frac{3.44-3.5}{.0229}\right) = pr(t < -2.6) = .011$

$df = 9-2 = 7$

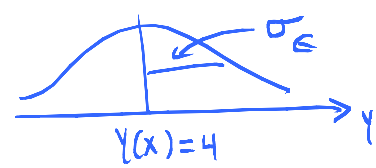
At $\alpha = .05$, There is evidence that the true mean of y at $x = 1.5$ is < 3.5

2 11.25 **10.** Let y = total daily travel time, x_1 = distance traveled, and x_2 = the number of deliveries made by some truck. For a company, the multiple regression model $y = 0.05\,x_1 + 0.5\,x_2 + \epsilon$ is found to be adequate. A truck has to make 3 deliveries, spanning a total of 50 miles; assuming $\sigma_\epsilon = 1.0$ hour, what is the probability that travel time is at most 6 hours? I.e., What is the probability that the truck completes the job in less than 6 hours?

True mean of y at x $= .05(50) + .5(3) = 4$

$prob(y < 6\text{ hrs}) = prob\left(\frac{y - Y(x)}{\sigma_\epsilon} < \frac{6 - Y(x)}{\sigma_\epsilon}\right)$

$= pr\left(z < \frac{6-4}{1}\right) = pr(z < 2) = 0.9772$



3 11.36 **11.** The following printout is produced from a regression analysis to predict energy content from burning objects with different concentrations of plastics, paper, garbage, and water. WITHOUT any calculation, answer the following questions AND refer to the specific entry/entries in the printout.

Predictor	Coef	StDev	T	p
Constant	2244.9	177.9	12.62	0.00
plastics	28.925	2.82	10.24	0.00
paper	7.64	4.06	1.88	0.03
garbage	4.297	10.48	0.41	0.34
Changed water	-37.35	1.83	-20.36	0.00

because $\beta_{garbage}$ may be zero.

s=31.48, R-sq=0.964, R-sq-adj=0.958 .

Analysis of Variance:

Source	DF	SS	MS	F	p
Regression	4	664931	166233	167.71	0.000
Error	25	24779	991		
Total	29	689710			

a) Does the model have any utility? Yes, because At least 1 β_i is non zero.

b) With all the variables included in the model, does garbage provide useful information? No

c) Under what conditions can you conclude that an increase in water content is expected to lead to a decrease in energy content? When water is not collinear with any of the other predictors.