Name:
ID:
Quiz section or time:

Points

Stat/Math 390, Winter, Test 3, March 11, 2011; Marzban
Open everything, closed messaging/discussion

1 **1.** You have two R codes, each of which you run on 10 different machines. You are NOT interested in picking the fastest machine. Instead, you are trying to see if there is a difference in the mean run time between the two codes. The appropriate test is

a) 2-sample, 1-sided t-test b) 2-sample, 2-sided t-test

c) 2-sample, 2-sided t-test for paired data d) Chi-sqd test of homogeneity.

2. The inequality of two large-sample proportions can be tested by (multiple answers are possible)

a) z-test, b) chi-squared test c) F-test

1 **3.** For small samples from a Normal population, a 95% -interval covers the population mean about 95% of the time.

a) z- b) t- c) a and b d) neither a nor b.

1 **4.** Of the 91 people who were given a high dose of drug X, 40 developed a severe rash and 5 had mild rash. Of the 96 people who were given a medium dose of the drug, the numbers were 20 and 8. And of the 53 people in the control group, the numbers were 20 and 10. What is the best test of whether the drug (dosage) has an effect on the incidence of rash in the three groups?

a) t-test b) chi-squared test of individual proportions c) chi-squared test of homogeneity d) F-test

1 **5.** In performing an F test of the inequality of several population means, the sample sizes from each population

a) must be equal b) must be unequal c) can be anything (≥ 2)

1 **6.** A weight scale is used to measure the weights of 40 objects, in each of 4 populations. An F-test is performed to test the inequality of means. If later it is found that all of the measurements were consistently off by an additive constant, the "new" p-value will be the one computed earlier.

a) lower than b) equal to c) larger than d) insufficient info provided.

7. The inequality of two large-sample means can be tested by (multiple answers are possible)

a) z-test, b) chi-squared test c) F-test

1 **8.** In any model with parameters to be estimated from some data, when overfitting has occurred, increasing the complexity of the model will cause its R^2 on that data to , and its R^2 on other data will .

a) increase, increase b) increase, decrease c) decrease, decrease d) insufficient info provided.

2 **9.** Do 95% CIs for a response (in regression) cover predictions for individuals more, or less, often than 95% of the time? In words, explain why.

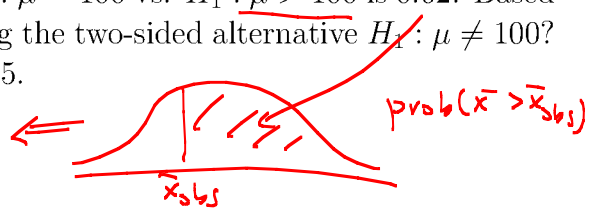
The 95% P.I is designed to cover individual predictions about 95% of the time. But PI is wider than C.I. So, 95% CIs will cover individual predictions less often than 95%

2 Jim

10. Based on a random sample, the p-value of the test $H_0 : \mu = 100$ vs. $H_1 : \mu > 100$ is 0.62. Based on the same random sample, what is the p-value for testing the two-sided alternative $H_1 : \mu \neq 100$? Hint: draw a normal distribution, and note that $0.62 > 0.5$.

p-value = 2x tail area

$= 2(1 - .62) = \boxed{0.76}$



8.42 Jim

11. Consider our study center. It is staffed with the expectation that 40% of its clients are from the business school, 30% from engineering, 20% from social science, and the other 10% from agriculture. A random sample of 120 clients reveals 33, 42, 30, and 15 from the four departments. Does this data suggest that the percentages on which staffing is based are not correct? To that end,

2/3 **a)** State the relevant hypotheses in terms of **clearly-defined quantities**.

π_i = prop. of students from dept i. H_0: $\pi_1 = .4$, $\pi_2 = .3$, $\pi_3 = .2$, $\pi_4 = .1$

H_1: Atleast 1 of the above is wrong.

3 **b)** Compute p-value, and state conclusion **regarding the percentages**, at $\alpha = 0.05$. **BY HAND**

expected counts = .4(120) = 48, 36, 24, 12

obs counts = 33, 42, 30, 15

$\chi^2 = \sum \frac{(exp - obs)^2}{exp} = 4.6875 + 1 + 1.5 + 0.75 = 7.9375$

Table VII, $df = n - 1 = 4 - 1 = 3 \Rightarrow$ $0.045 < p\text{-value} < 0.05$

p-value $< \alpha \Rightarrow$ Reject H_0 in favor of H_1, NOT

I.e. The assumed percentages are consistent with data.

1 **c)** The percentage for which department is least consistent with data? No explanations necessary.

Business school. (Biggest contribution to χ^2) ← Not necessary.

11.18 Carlin

12. A regression equation has been developed based on a sample of size 22: $\hat{y} = 3 + 4x$. Team 1 uses it at $x = 5$ to predict the response, while team 2 uses it at $x = 7$ to make a prediction. When the two teams compare their results, they see that the difference (second - first) between their predictions is 8. Is there evidence from data that the true difference is greater that 6? So,

2/3 **a)** Let y_1 and y_2 denote the true predictions for the two teams. In terms of these, write H_0, H_1.

$H_0 : y_2 - y_1 \leq 6$

$H_1 : y_2 - y_1 > 6$

2/3 **b)** Denote the true regression equation as $y = \alpha + \beta x$. In other words, $y_1 = \alpha + \beta(5)$, and $y_2 = \alpha + \beta(7)$. Using these equations, transform the above hypotheses so that they pertain to β.

$H_0 : \alpha + 7\beta - \alpha - 5\beta \leq 6$,, > 6 $\Rightarrow$ $H_0 : \beta \leq 3$, $H_1 : \beta > 3$

3 **c)** Compute a p-value to test the hypotheses in part b. Suppose $s_e = 1$ and $S_{xx} = 4$. **Show work!**

$t_{obs} = \frac{\hat{\beta} - \beta_0}{s_e / \sqrt{S_{xx}}} = \frac{4 - 3}{1/\sqrt{4}} = 2$

p-value = prob $(t > t_{obs})$ = prob $(t > 2)$ = $\boxed{.030}$

$df = 22 - 2 = 20$

At $\alpha = .05$, There is evidence that $y_2 - y_1 > 6$ ← Not necessary.