Name: _____

ID: _____

Quiz section or time: _____

Stat/Math 390, Spring, Test 3, June 5, 2015; Marzban

Same deal as test 1, ...

Points

1. We have learned a procedure that can give a sense of the minimum necessary sample size for a study, based on considerations of a CI (for a mean, for example). Which of the following is/are necessary for that minimum-sample-size-estimation procedure?

a) Confidence level b) sample mean c) sample standard deviation d) desired width of the CI

1 **2.** Generally, for which of the following can one NOT build a CI and do hypothesis tests?

a) difference between two pop proportions
b) ratio of two pop proportions
c) difference between two pop variances
d) ratio of two pop variances
e) prediction from a population regression
f) None of the above

1 **3.** Fill in the blanks in the following statement, using the list of phrases appearing below.

In deriving a formula for a CI for a ____, we need the ____ of a ____, and the variance of ____.

a) population parameter b) sample statistic c) sampling distribution

1 **4.** A lower confidence bound for $\mu_1 - \mu_2$ is _____ the upper confidence bound for $\mu_2 - \mu_1$.

a) equal to b) equal to the negative of c) unrelated to

5. Suppose we have tested H0: $\pi_1 - \pi_2 \leq 0.6$ versus H1: $\pi_1 - \pi_2 > 0.6$, and found a p-value less than $\alpha = 0.05$. Which of the following is/are correct?

a) We can be 95% confident that π_1 exceeds π_2.
b) We can be 95% confident that π_1 exceeds π_2 by at least 0.6.
c) There is a 95% chance that π_1 exceeds π_2 by at least 0.6.
d) There is a 5% chance that π_1 does not exceed π_2 by 0.6.

1 **6.** Suppose you have collected data, and performed a hypothesis test, leading to p-value $< \alpha$. But suppose you simply don't like the conclusion. What is the most appropriate action you should take?

a) Switch H0 and H1. b) Collect more data. c) Change α. d) None of the above.

1 **7.** You may not know it, but you have taken 6 engineering, 5 physics, and 8 math courses, and you have numerical grades for each course. Suppose you have a feeling that you are equally good (as measured by gpa) in those three fields. What is the most appropriate test for testing your feeling?

a) z-test b) t-test c) chi-squared d) 1-way ANOVA F-test e) F-test of model utility

8. Based on everything you now know about regression, which of the following probs can NOT be computed, at a given x? In all of the following c is just a constant.

a) $prob(y > c)$, where y is a random y value.
b) $prob(\hat{y}(x) > c)$, where $\hat{y}(x)$ is a random prediction of the mean of y-values at a given x.
c) $prob(\hat{y}(x) > c)$, where $\hat{y}(x)$ is a random prediction for a single case, at a given x.
d) $prob(y(x) > c)$, where $y(x)$ is the true/population prediction at a given x.

1 **9.** A 95% CI for $y(x)$ will cover individual observed values 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 Lect 28 and class discussion

10. In ~~regression~~, if the standard deviation of errors (σ_ϵ) and the standard deviation of the predictions ($\sigma_{\hat{y}}$) are **known** (i.e., not estimated), then the distribution of $\frac{\text{prediction error}}{\sqrt{\sigma_{\hat{y}}^2+\sigma_\epsilon^2}}$ is

a) N(0,1) b) t with df = n-(k+1) c) insufficient information.

Last 30

1 **11.** Which of the following statements is/are generally true?

a) $\alpha + \beta = 1$ c) Statistical significance implies physical significance.

b) β + power = 1 d) Physical significance implies statistical significance.

---

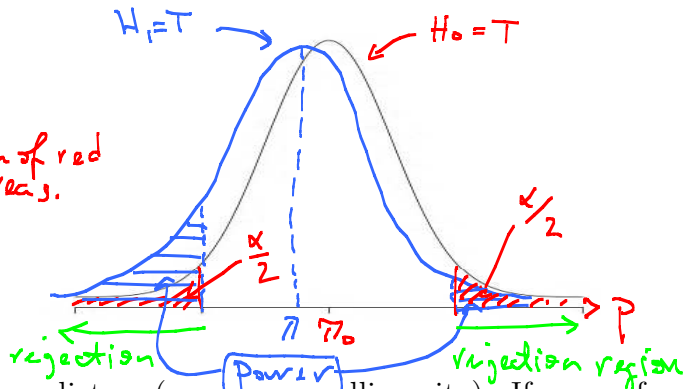
hw and Last

12. We are testing $H_0 : \pi = \pi_0$ vs. $H_1 : \pi \neq \pi_0$ at $\alpha = 0.05$. On the shown normal distribution

~1 a) Label the x-axis p = sample proportion

~2 ~~1~~ b) Shade/label the area(s) corresponding to α. sum of red areas.

~3 ~~2~~ c) Revise the diagram (or make a new one), clearly shading/labeling the area(s) corresponding to power. Sum of blue areas.



Jelly Beans, Chocolate ...

~2 **13.** In a multiple regression problem we have 100 predictors (assume no collinearity). If none of the predictors are actually useful, what will happen if you perform t-tests on each of the hundred β coefficients, at significance level $\alpha = 0.05$?

About 5 of the β's will turn-up as significant (e.g. $\beta \neq 0$) even though in reality none are significant (e.g. all $\beta = 0$)

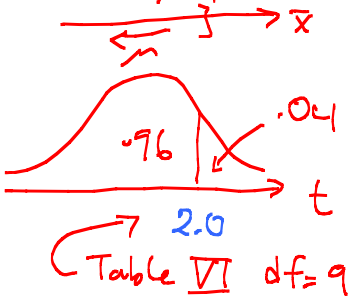
hw-A6 7.18

14. The response time (in milliseconds) was measured in a small sample of 10 pilots, and the sample mean and standard deviation of response time were found to be 100, and 21 respectively. For these pilots a long response time is problematic. Suppose you are a General in the airforce and are facing the decision to hire or not hire these pilots. Assuming the population of response time is normal, at a confidence level of 96% (NOT 95%),

~3 ~~2.5~~ a) compute the most relevant confidence interval for response time; assume $\sqrt{10} \sim 3$.

Because a slow response time is a "BAD" Thing, The most appropriate CI is a upper confidence bound: $\bar{x} + t^* s/\sqrt{n}$

$$\bar{x} + t^* \frac{s}{\sqrt{n}} = 100 + 2.0 \frac{21}{\sqrt{10}} \simeq 100 + 2.0 \frac{21}{3} = 100 + 14 = 114$$



~2 b) provide TWO interpretations of the answer in part a.

1) We can be 96% confident That The true mean response-time is lower Than 114

2) 96% of upper confidence bounds computed from random samples will be larger than The true mean response time

2') Equivalently, There is a 96% prob. That a random sample will yield an upper Conf. bound That is higher Than The true mean.

Lots to do

15. To test whether an electrical wire manufacturing process is functioning correctly, the resistivity of 100 wire segments is measured at the beginning and the end of the wire. These measurements are denoted x_i and y_i, respectively, with $i = 1, ..., 100$. We have decided to compute a 95% 2-sided CI for the difference in mean resistivity $\mu_x - \mu_y$ between the two ends. Suppose we find that for each and every wire segment, there is a difference of exactly 1.0 (Ohm) in the resistivity between the two ends, i.e., $x_i - y_i = 1.0$.

~ 3 a) Compute an appropriate CI. Hint: see part b.

This suggests that the appropriate CI is for paired data. Then

C.I. for $\mu_x - \mu_y$: $\bar{d} \pm t^* \frac{s_d}{\sqrt{n}} = \boxed{1}$

Here, $d_i = x_i - y_i = 1 \Rightarrow \bar{d} = 1$ and $\boxed{s_d = 0}$ (d_i does not vary; it's always 1)

~ 2 ~~2~~ 1 b) Explain why the answer in part a makes sense.

The Conf. "Interval" is 1, ie. has no width. This makes sense because d_i is 1 for all i; so we can be certain that $\mu_x - \mu_y = 1.0$.

8.44 **16.** An information retrieval system has 4 storage locations. Information has been stored with the *a priori* expectation that the long-run proportion of requests for location i is given by $\pi_i = (2.5 - |i - 2.5|)/6$. A sample of 30 retrieval requests gave the following frequencies for locations $i = 1, ..., 4$, respectively: 4, 10, 11, and 5. We want to test if this data are consistent with the *a priori* expectations.

~ 2 a) Set-up the hypotheses, in terms of the parameters defined in the problem.

H_0: $\pi_1 = \frac{1}{6}$, $\pi_2 = \frac{2}{6}$, $\pi_3 = \frac{2}{6}$, $\pi_4 = \frac{1}{6}$

H_1: At least one of these assignments is wrong.

~ 2 b) Write the expected count in each of the 4 locations, under H_0: $\frac{1}{6}(30)$, $\frac{2}{6}(30)$, $\frac{2}{6}(30)$, $\frac{1}{6}(30)$ → 5, 10, 10, 5

~ 1 c) Write the observed count in each of the 4 locations: 4, 10, 11, 5

~ 2 ~~2~~ 1 d) Write the most appropriate test statistic (just name it; **do not** compute it): Chi squared df=3

Lots of places

~ 3 ~~3~~ 3.5 **17.** ~~Let us assume~~ that the true regression fit for a certain problem is $y = 1 + 5x + \epsilon$. We do not know the true standard deviation of errors, but we do have an estimate of 12, based on a sample of size 10. We also have $S_{xx} = 12$. What is the probability that the average change in y, with respect to 1 unit change in x, will exceed 6?

"β"

$$? = pr(\hat{\beta} > 6) = pr\left(\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} > \frac{6 - \beta}{s_{\hat{\beta}}}\right) = pr\left(t > \frac{6 - \beta}{s_e/\sqrt{S_{xx}}}\right)$$

$$= pr\left(t > \frac{6 - 5}{12/\sqrt{12}}\right) = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}} = \frac{1}{3.4} \approx .3$$

$$= pr(t > 0.3) = \boxed{0.386}$$

Table VI, df = 10 - 1 = 9