Name: ______________________ ID: ________________ Quiz section or time: _________

Stat/Math 390, Winter, Test 3, Mar. 11, 2016; Marzban

AS BEFORE

9 + 15

Points

sect 22

1. In testing H0: $\mu \le \mu_0$ vs. H1: $\mu > \mu_0$, circle all TRUE statements. The p-value is given by the following area under the sampling distribution of $\bar{x}$:
a) Area to the right of μ_0.
b) Area to the right of the observed sample mean if it's to the left of μ_0.
c) Area to the right of the observed sample mean if it's to the right of μ_0.
d) Area to the left of the observed sample mean if it's to the left of μ_0.
e) Area to the left of the observed sample mean if it's to the right of μ_0.

7.86, sp.12 (test2)

2. A 95% lower confidence bound for a population mean is 13. Then, which are correct?
a) The true mean is above 13, with 95% confidence.
b) 95% of computed lower bounds will be lower than the true mean.
c) There is a 95% probability that the true mean is greater than 13.
d) 95% of sample means will be greater than ~~13~~

13 μ $\bar{x}$

not $\bar{x}$ $\to \mu - z^* \frac{\sigma}{\sqrt{n}}$ .95 μ $\bar{x}$

8.36

1 **3.** We want to see if an update to a software makes it faster. So, we time the software on a computer, and then time the updated software on the same computer. Then we repeat this experiment on 9 more computers. The most appropriate test is
a) 1-sample t-test b) 2-sample indep. t-test c) 2-sample paired t-test d) 1-way ANOVA F-test

1 **4.** Which of the following is the correct interpretation of a p-value?
a) It is the probability that the null hypothesis is correct.
b) It is the probability that the alternative hypothesis is correct.
c) It measures the evidence from data in favor of the null hypothesis.
d) It measures the evidence from data in favor of the alternative hypothesis.

1 **5.** Which of the following statements is TRUE? In testing $H_0 : \mu \ge \mu_0$; $H_1 : \mu < \mu_0$, if in fact
a) $\mu \ge \mu_0$, then surely p-value $< \alpha$
b) $\mu \ge \mu_0$, then surely p-value $> \alpha$
c) $\mu < \mu_0$, then surely p-value $< \alpha$
d) $\mu < \mu_0$, then surely p-value $> \alpha$
e) All of the above
f) None of the above

7.58

1 **6.** Suppose you compute the 95% upper confidence bound for a population mean. Also compute the 95% lower confidence bound for the population mean. The interval defined by these two numbers covers the true mean what percentage of the time?

CI x

a) less than 95% b) 95% c) more than 95%

7. When performing a 1-way ANOVA F-test, which of the following is/are TRUE?
a) The number of observations in each population must be equal. n_i
b) The larger the F statistic, the more evidence there is against H_0. p-value = pr(F > F_obs)
c) p-value is the sum of the areas (left & right of F_{obs}), because of "At least, ..." appearing in H_1.
d) For 2 populations, the F-test is equivalent to a 2-sample, 2-sided t-test. F is generalization of t.

w.15 test 2

8. In regression, circle all of the quantities that can be computed.
a) CI for α b) CI for β_i c) CI for $y(x)$ d) PI for α e) PI for β_i f) PI for $y(x)$

w.15 test 2

1 **9.** Suppose we have k brands of phones, and we want to see if $\mu_1 < 1, \mu_2 < 1, \cdots \mu_k < 1$, where μ_i is the mean lifetime of brand i. The most appropriate test is
a) t-test b) chi-squared c) 1-way ANOVA F-test d) F-test of model utility e) None of the above

$\bar{x} = \frac{\sum x_i}{n}$, $s_x^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$, $s_e^2 = \frac{\sum(y_i-\hat{y}_i)^2}{n-(k+1)}$, $r = \frac{1}{n-1}\sum(\ )(\ )$ || $s_{\hat{\beta}} = \frac{s_e}{\sqrt{S_{xx}}} = \frac{s_e}{\sqrt{n-1}\,s_x} \to 0$

df of numerator   df ...

$s_{\hat{y}} = s_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}} \to 0$

~1.5 6/4

10. Consider the following **sample** statistics. What happens to each as the sample size (n) increases? Write one of the following possible answers (a,b,c) in the space provided.

a) Generally decreases b) Generally remains constant c) Generally increases

mean ($\bar{x}$):	b	standard deviation of errors (s_e):	b
variance (s^2):	b	standard deviation of predictions ($s_{\hat{y}}$):	a
correlation coefficient (r):	b	standard deviation of a regression slope ($s_{\hat{\beta}}$):	a

~2 97.9

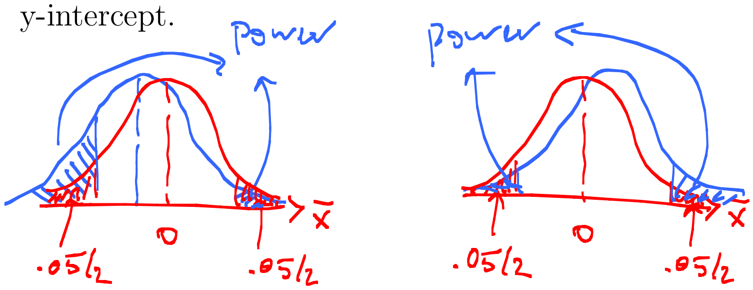
11. The number (or proportion) of times a given letter in the English language occurs in any English text is known (not shown). Consider the following sample counts from a single document.

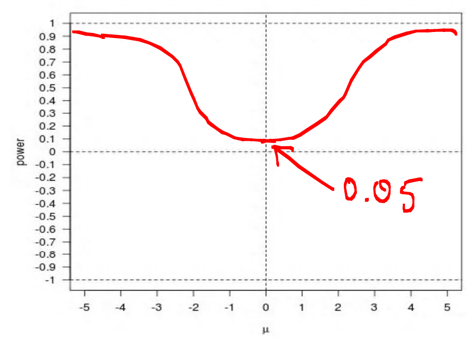
Letter	E	T	A	O	Other
Count	130	90	80	80	620

We want to know if this data suggest that the document is in English. Write the name/type of the most appropriate test: Chi-squared (with df = 5−1)

~~2~~ 3 Sp.15

12. Consider a 2-sided test of a population mean, i.e., $H_0 : \mu = 0$, $H_1 : \mu \neq 0$. Suppose n is large, $\sigma = 1, \alpha = 0.05$. On the adjacent diagram, plot the resulting power curve. The curve can be qualitative, but show the numerical value of the y-intercept.





7.86 sp.10

13. A survey taken in 2002 shows that 10 individuals in a sample of 100 adults are obese. A 1998 census (i.e., the whole **population**) concluded that 4% of the adults are obese. You want to decide with 95% confidence if the 2002 true percentage (π) is more than twice the 1998 percentage?

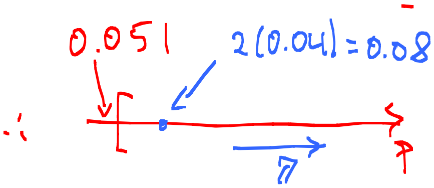
~~3~~ 2.5 a) Compute an appropriate interval (you may use the simplest formula you have, with $t^* = 1.645$)

$p = 10/100 = 0.1$

The appropriate interval is (a 1-sided) lower confidence bound for π:

$$p - t^*\sqrt{\frac{p(1-p)}{n}} = 0.1 - 1.645\sqrt{\frac{.1(.9)}{100}} = 0.1 - 1.645\,\frac{1}{10}\sqrt{0.09}$$

$$= 0.1 - 1.645\,\frac{1}{10}\,\frac{3}{10} \simeq 0.1 - \frac{4.9}{100} \simeq 0.051 \qquad \sqrt{\frac{9}{100}} = \frac{3}{10}$$



~1 b) What is the conclusion in English.

We are 95% confident that $\pi > 0.051$.
But twice 0.04 is in that interval.
So, there is no evidence (at 95% conf. level) that the 2002 proportion π is larger than twice the 1998 proportion.

~~3~~ 2.5 CI version of sp10

14. It is known that the quantity $X^2 = (n-1)s^2/\sigma^2$ has a chi-squared distribution with $df = n-1$. According to chi-squared tables, the 0.025 quantile and the 0.975 quantile of the chi-squared distribution with $df = 20$ are approximately 3.0 and 9.0, respectively. In other words, $pr(3.0 < X^2 < 9.0) = 0.95$, for $df = 20$. **Starting** from what we have been calling a "self-evident fact," use all of this information to compute a 2-sided, 95% CI for σ^2; suppose in a sample of size 21 the sample variance is 27.0. df = 20.

$pr(3 < x^2 < 9) = 0.95$ ← Self-evident fact

$3 < \frac{(n-1)s^2}{\sigma^2} < 9$

$\frac{1}{3} > \frac{\sigma^2}{(n-1)s^2} > \frac{1}{9}$

$\boxed{\frac{(n-1)s^2}{3} > \sigma^2 > \frac{(n-1)s^2}{9}}$

C.I

For $n = 21$, $s^2 = 27$, $df = n-1 = 20$

$\frac{20(\cancel{27})^{9}}{\cancel{3}} > \sigma^2 > \frac{20(\cancel{27})^{3}}{\cancel{9}}$

$\boxed{180 > \sigma^2 > 60}$

~~3~~ 2.5 w14, Summer 12

15. In a simple regression fit to a sample of size 16, we have $\bar{x} = 10, S_{xx} = 16/127, s_e = 4$. If we make a prediction of y at $x = 11$, what is the probability that a prediction from a sample fit will exceed the observed value of y by 4.0? In other words, what's the probability that a prediction will be greater than $(y^* + 4)$?

$? = pr(\hat{y} > y^* + 4) = pr(\hat{y} - y^* > 4) = pr\left(\underbrace{\frac{\hat{y} - y^*}{S_{pred.err}}}_{t} > \frac{4}{S_{pred.err.}}\right)$

$= pr\left(t > \frac{4}{\sqrt{S_e^2 + S_{\hat{y}}^2}}\right)$

$= pr\left(t > \frac{4}{S_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}}\right) = pr\left(t > \frac{\cancel{4}}{\cancel{4}\sqrt{1 + \frac{1}{16} + \frac{(11-10)^2\, 127}{16}}}\right)$

$= pr\left(t > \frac{1}{\sqrt{1 + \frac{1}{16} + \frac{127}{16}}}\right) = pr\left(t > \frac{1}{\sqrt{1 + \frac{128}{16}}}\right) = pr\left(t > \frac{1}{\sqrt{1+8}}\right)$

$= pr\left(t > \frac{1}{3}\right) \approx \boxed{0.36}$ (between 0.347 and 0.384)

$\frac{1}{3} \approx 0.33$; $df = n - 2 = 14$; $t = 0.3$; $t = 0.4$