Name: ______________________
ID: ________________________

Stat 421, Test 1, Fall, Nov 13, 2014; Marzban

21

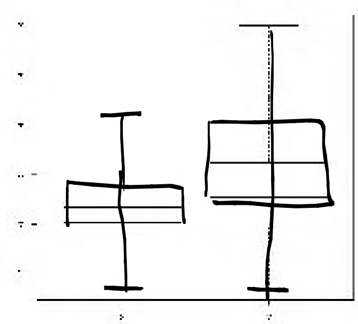
ONLY a half-size "cheat sheet" is allowed

Multiple choice: Circle all the correct answers; there is wrong-answer penalty

For rest, SHOW answer & work; NO CREDIT for correct answer without explanation

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.** Consider the data on x and y in the adjacent diagram.
Circle all of the reasonably correct conclusions.
a) x and y have equal sample medians
b) x and y have equal population medians
c) x and y have different sample medians
d) x and y have different population medians

1 **3.** Ten model X cars are selected at random, and the CO_2 level is measured both inside and outside the cars. Denote the measurements x_i and y_i, respectively, with $i = 1, 2, ..., 10$. We are interested in whether or not there is a difference between the mean CO_2 level inside and outside model X cars. Which of the following is/are useful to examine?

Paired data

a) comparative boxplots of the x_i and y_i
b) histograms of the x_i and y_i
c) boxplot of $x_i - y_i$.
d) boxplot of x_i/y_i (assuming $y_i \neq 0$).

1 hw, Q **4.** Suppose there is a difference between two population means. If data are in fact paired, then an unpaired test will generally yield a ____ p-value than that of a paired test.
a) Higher b) Lower c) zero d) Cannot tell in general.

1 **5.** Suppose we have to decide whether or not to approve a medicine. Suppose you have decided that it's less dangerous to deny a medicine that actually works, and it's more dangerous to approve a medicine that does not work. What are the appropriate hypotheses?
a) H_0 : medicine works, H_1 : medicine does not work.
b) H_0 : medicine does not work, H_1 : medicine works.
c) Based on this information, one cannot decide H_0, H_1.

~~~2~~ 1.5 Lect 7 **6.** A swarm of bees flies around a honey comb. Suppose their x, y, and z coordinates are standard normally distributed about the honey comb. What is the radius of the sphere which includes about half of the bees? Explain.

$r^2 = x^2 + y^2 + z^2$ where $x, y, z \sim N(0,1) \Rightarrow r^2 \sim \chi^2_3$

$0.5 = pr(r < r_c) = pr(r^2 < r_c^2) = 1 - pr(r^2 > r_c^2)$

$0.5 = pr(r^2 > r_c^2) \Rightarrow r_c^2 = 2.37$ (Table III)

$\therefore \boxed{r_c = \sqrt{2.37}}$
~~~

~~2~~ 1

7. Before agreeing to purchase a large order of polyethylene sheaths, a company wants to see conclusive evidence that the population standard deviation of sheath thickness is less than 0.05 mm. Write the appropriate hypothesis pair in terms of **well-defined parameters**.

$H_0: \sigma \geq .05$

$H_1: \sigma < .05$

σ = pop. std. dev. of sheath Thickness.

hw, C, D

8. The rv X is known to have the pdf $f_X(x) = 1$, for $0 < x < 1$, and zero for all other x values.

~2 **a)** Find the expected value of X^2.

2

$$E_X[X^2] = \underline{\underline{\int_{-\infty}^{\infty} x^2 f_X(x)\,dx}} = \int_0^1 x^2\, 1\, dx = \frac{1}{3} x^3 \Big|_0^1 = \frac{1}{3}$$

~2 **b)** In homeworks we have seen how one can *transform* from X to Z, for example $Z = (X-\mu)/\sigma$. Now, suppose the transformation is $Z = X^2$, and we already know that the pdf of Z is $f_Z(z) = \frac{1}{2\sqrt{z}}$ for $0 < z < 1$, and zero elsewhere. Compute the expected value of Z.

$$E_Z[Z] = \underline{\underline{\int_{-\infty}^{\infty} z\, f_Z(z)\, dz}} = \int_0^1 z \frac{1}{2\sqrt{z}}\, dz = \frac{1}{2}\int_0^1 \sqrt{z}\, dz = \frac{1}{\cancel{2}} \frac{z^{3/2}}{3/\cancel{2}} \Big|_0^1 = \frac{1}{3}$$

2.32

~~2~~ 2.5

9. Two popular medications are being compared on the basis of the average absorption rate by the body. Tablet 1 is claimed to be absorbed at least twice as fast as Tablet 2. Assume that σ_1^2 and σ_2^2 are known. Using what we know about the sampling distribution of a sample mean, write the expression for the statistic which has a standard normal distribution. Confirm that its expected value and variance are correct. μ_i = mean absorption rate for Tablet i.

$H_0: \mu_1 \geq 2\mu_2 \qquad H_1: \mu_1 < 2\mu_2$

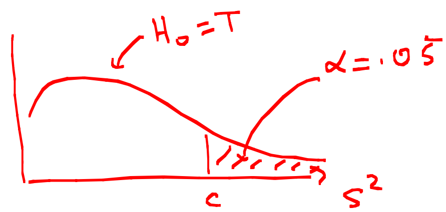
$$\left.\begin{aligned} E[\bar{X}_1 - 2\bar{X}_2] &= E[\bar{X}_1] - 2E[\bar{X}_2] = \mu_1 - 2\mu_2 \\ V[\bar{X}_1 - 2\bar{X}_2] &= V[\bar{X}_1] + 4V[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + 4\frac{\sigma_2^2}{n_2} \end{aligned}\right\} \Rightarrow z = \frac{(\bar{X}_1 - 2\bar{X}_2) - (\mu_1 - 2\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + 4\frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

2.5 hw-power

~~3~~ hw-J

10. We are testing the hypotheses $\underline{H_0: \sigma^2 \leq 9}$ versus $\underline{H_1: \sigma^2 > 9}$, at $\alpha = 0.05$. Based on a sample of size $n = 16$, what is the (approximate) power of the (most appropriate) test if the true variance is in fact 36?

$$0.05 = pr(s^2 > c) = pr\left(\underbrace{\frac{(n-1)s^2}{\sigma^2}}_{\chi^2_{16-1}} > \underbrace{\frac{(16-1)c}{9}}_{\Rightarrow 25 \text{ (Table III)}}\right)$$



$$\frac{\cancel{15}^{5} c}{\cancel{9}_3} = \cancel{25}^{5} \Rightarrow \boxed{c = 15}$$

$$\text{power} = pr(s^2 > 15 \mid \sigma^2 = 36)$$

$$= pr\left(\frac{(n-1)s^2}{\sigma^2} > \frac{(16-1)15}{36}\right)$$

$$= pr\left(\chi^2 > \frac{\cancel{15}^{5} \cdot \cancel{15}^{5}}{\cancel{6}_2 \cdot \cancel{6}_2}\right) = pr\left(\chi^2 > \frac{25}{4}\right) = pr(\chi^2 > 6.2) \simeq \boxed{.975}$$

(Table III)

~ 2 Sample version of I, II Lect 7 p. 4 **11.** In a paired design, data have been collected on two variables x and y. Let $d_i = x_i - y_i$, with $i = 1, 2, ..., n$ denote the differences. Express the sample variance of the differences, s_d^2, in terms of the two sample variances s_x^2 and s_y^2, and the Pearson correlation coefficient $r = \frac{1}{n-1}\sum_i(\frac{x_i-\overline{x}}{s_x})(\frac{y_i-\overline{y}}{s_y})$, where $\overline{x}$ and $\overline{y}$ are the sample means.

$$S_d^2 = \frac{1}{n-1}\sum_i \left[(x_i - y_i) - (\bar{x} - \bar{y}) \right]^2$$

$$= \frac{1}{n-1}\sum_i \left[(x_i - \bar{x}) - (y_i - \bar{y}) \right]^2$$

$$= \frac{1}{n-1}\sum_i (x_i-\bar{x})^2 + \frac{1}{n-1}\sum_i (y_i - \bar{y})^2 - 2\,\boxed{\frac{1}{n-1}\sum_i (x_i-\bar{x})(y_i-\bar{y}) \cdot \frac{S_x S_y}{S_x S_y}}$$

r

$$S_d^2 = S_x^2 + S_y^2 - 2 r S_x S_y$$

~ 2 qq **12.** A Normal qq-plot involves plotting the data versus quantiles of the standard normal distribution. It turns out one can make a qq plot to check whether some data come from a distribution other than Normal. For example, one can plot the data versus the theoretical quantiles of an exponential distribution. A straight line would suggest that the data do come from an exponential. Compute the p^{th} quantile of the exponential distribution with parameter λ.

$$\int_0^q \lambda e^{-\lambda x}\,dx = p \Rightarrow \lambda \frac{e^{-\lambda x}}{-\lambda}\Big|_0^q = p \Rightarrow 1 - e^{-\lambda q} = p$$

$$\therefore e^{-\lambda q} = 1-p \Rightarrow -\lambda q = \ln(1-p) \Rightarrow q = -\frac{1}{\lambda}\ln(1-p)$$

These q values would go on the x-axis of the qq plot.

2.5 ~ ~~2~~ Lect 6 hw-4 **13.** Equal size samples ($n_1 = n_2 = 31$) are taken from two populations, and a 2-sided confidence interval (CI) for σ_1^2/σ_2^2 is desired. However, it is also required that the upper limit of the CI is four times larger than the lower limit. What must the significance level α be?

C.I. for $\frac{\sigma_1^2}{\sigma_2^2}$: $\underbrace{\frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}}_{A} < \frac{\sigma_1^2}{\sigma_2^2} < \underbrace{\frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1}}_{B}$

$B = 4A$, $n_1 = n_2 = 31$

$$\therefore F_{\frac{\alpha}{2},30,30} = 4\, F_{1-\frac{\alpha}{2},30,30} = 4 \cdot \frac{1}{F_{\frac{\alpha}{2},30,30}}$$

$$\therefore \left(F_{\frac{\alpha}{2},30,30}\right)^2 = 4 \Rightarrow F_{\frac{\alpha}{2},30,30} = 2 \Rightarrow \frac{\alpha}{2} = .025$$

F Table

$$\therefore \boxed{\alpha = .05}$$