Name: ______________________
ID: ______________________

6.5 + 11.5

Stat 421, Test 1, Fall, Oct. 18, 2017; Marzban

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 lect2 **1.** Suppose you see a table like this in a paper, where the dots represent the grades of different students. The implication is that Gender is one of the nuisance factors. Fill the blank space.

	Major				
	physics	math	statistics	CS	Gardening
Boy	...	...	...	...	...
Girl	...	...	...	...	...
Average	.	.	.	.	.

suggest that we don't care about gender.

everywhere **2.** Which of the following statements is/are FALSE?

1.5

a) If y_i are iid, then $Cov[y_i, \overline{y}] = 0$ — in $\overline{y}$, there is one term that's y_i. Then $Cov[y_i, \overline{y}] = V[y_i]$

b) $E_Y[s^2] = \sigma_Y^2$ only if $Y \sim N$. — check the proof.

c) $V_Y[\overline{y}] = \sigma_Y^2/n$ only if $Y \sim N$.

d) $\sum_i (y_i - \mu_Y) = 0$

e) $E[\frac{y_i - \overline{y}}{s}] = 0$

f) Because the constrast Γ is defined in terms of population params, a p-value test does not exist.

check the proof.

$\Sigma(y_i - \mu_Y) = n(\overline{y} - \mu_y) \neq 0$

s = r.v.

X lect2 **3.** We have observed the following data, and intend to perform a randomization test of the effect of Tip on Depth. In the table provided, write one of the random trials that should NOT be considered.

Coupon	1	2	3	4
Tip	A	B	A	B
Depth	13	11	12	10

Did not grade this.

Coupon	1	2	3	4
Tip	A	B	A	B
Depth	12	11	13	10

1 2014/5 **4.** 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, it does not matter how we set-up H_0, H_1.

Bad err = (Approve | does not work)
(H_1 | H_0)

1 hw-lect3-2 **5.** Let $x \sim N(\mu, \sigma)$. Then, the most likely value of sample maximum is

a) $-\infty$ b) 0 c) 1 d) $+\infty$ e) none of the above

1 lect6 **6.** We have collected data and computed $\overline{y}_{obs}$. Suppose we know that $\sigma_y = 1$. Then, the observed 95% CI for μ_y is $\overline{y}_{obs} \pm 1.96/\sqrt{n}$. Which of the following statements is/are TRUE:

a) There is 95% probability that a random $\overline{y}$ is within $\mu_y \pm 1.96/\sqrt{n}$ — defn. of CI.

b) There is 95% probability that a random $\overline{y}$ is within $\overline{y}_{obs} \pm 1.96/\sqrt{n}$ — $pr(\overline{y}_{obs} - \cdots < \overline{y} < \overline{y}_{obs} + \cdots)$

c) There is 95% probability that the observed $\overline{y}$ is within $\mu_y \pm 1.96/\sqrt{n}$

d) There is 95% probability that the observed $\overline{y}$ is within $\overline{y}_{obs} \pm 1.96/\sqrt{n}$

e) The rejection region is the region outside of $\overline{y}_{obs} \pm 1.96/\sqrt{n}$ — $Y_{critical}$

$pr\left(-1.96 < \frac{\overline{y} - \overline{y}_{obs}}{\sigma_y/\sqrt{n}} < 1.96\right)$

we don't know the dist. of this guy!

1 hw-lect6-3 **7.** Let F denote $(s_2/s_1)^2$. The CI for $(\sigma_2/\sigma_1)^2$ is given by

a) $[F_{obs} F_{1-\alpha/2,n1-1,n2-1}, F_{obs} F_{\alpha/2,n1-1,n2-1}]$

b) $[-F_{obs} F_{1-\alpha/2,n1-1,n2-1}, F_{obs} F_{\alpha/2,n1-1,n2-1}]$

c) $[F_{obs} - F_{1-\alpha/2,n1-1,n2-1}, F_{obs} + F_{\alpha/2,n1-1,n2-1}]$

d) None of the above.

Only a couple of student's got this right. That is not good, because this is a stat 220 question. Now that you can derive CI's, you should know how to interp. them.

~~2~~ 1.5 (lect 5, done differently)

8. Suppose we are given the following facts (always assuming iid):
1) If $y_i \sim N(0,1)$, then $z_i = \frac{y_i - \mu_y}{\sigma_y} \sim N(0,1)$, and $\frac{\bar{y}-\mu_y}{\sigma_y/\sqrt{n}} \sim N(0,1)$
2) If $z_i \sim N(0,1)$, then $\sum_i^n z_i^2 \sim \chi^2_n$.
3) If $x_1 \sim \chi^2_{df_1}$, $x_2 \sim \chi^2_{df_2}$, then $x_1 - x_2 \sim \chi^2_{df1-df_2}$.
From these facts, show that $\sum_i^n (\frac{y_i - \bar{y}}{\sigma_y})^2 \sim \chi^2_{n-1}$. Hint: Use an "ANOVA decomposition," but you have to decide whether to decompose $\sum_i^n (y_i - \bar{y})^2$ or $\sum_i^n (y_i - \mu_y)^2$

$$\sum_i^n (y_i - \overset{+\bar{y}-\bar{y}}{\mu_y})^2 = \sum_i^n (y_i - \bar{y})^2 + \sum_i^n (\bar{y} - \mu_y)^2 - 2(\bar{y}-\mu_y)\overbrace{\sum_i (y_i - \bar{y})}^{0}$$

$$\underbrace{\sum_i^n \left(\frac{y_i - \mu_y}{\sigma_y}\right)^2}_{\sim \chi^2_n \ (2)} = \underbrace{\sum_i^n \left(\frac{y_i - \bar{y}}{\sigma_y}\right)^2}_{(3)} + \underbrace{\left(\frac{\bar{y}-\mu_y}{\sigma_y/\sqrt{n}}\right)^2}_{\sim \chi^2_1 \ (1)}$$

$$\therefore \sim \chi^2_{n-1}$$

If you try $+\mu_y - \mu_y$

$$\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \mu_y)^2 + \sum_i (\bar{y}-\mu_y)^2 - 2(\bar{y}-\mu_y)\underbrace{\sum_i (y_i - \mu_y)}_{n(\bar{y}-\mu_y)}$$

$$= \sum_i (y_i - \mu_y)^2 - n(\bar{y}-\mu_y)^2$$

same

(hw, lect 8-1)

9. For the model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $i = 1 \cdots a, j = 1 \cdots n$, with $\epsilon_{ij} \sim N(0, \sigma_\epsilon)$, we built CIs for $\mu, \mu_i = \mu + \tau_i$, and σ^2_ϵ. But we forgot the CI for τ_i. Let's start it here. Estimate τ_i with $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$.

~2 a) Find $E[\hat{\tau}_i]$ and $V[\hat{\tau}_i]$, in terms of a, n, τ_i, and σ^2_ϵ. Use (w/o proof) the identity $V[\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}] = \frac{a-1}{an}\sigma^2_\epsilon$.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \Rightarrow \bar{y}_{i.} = \mu + \tau_i + \bar{\epsilon}_{i.} \Rightarrow \bar{y}_{..} = \mu + \bar{\tau}_. + \bar{\epsilon}_{..}$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} = \cancel{\mu} + \tau_i + \bar{\epsilon}_{i.} - \cancel{\mu} - \bar{\tau}_. - \bar{\epsilon}_{..} = \tau_i - \bar{\tau}_. + \bar{\epsilon}_{i.} - \bar{\epsilon}_{..}$$

$$E[\hat{\tau}_i] = (\tau_i - \bar{\tau}_.) + E[\bar{\epsilon}_{i.}] - E[\bar{\epsilon}_{..}]$$

$$E[\bar{\epsilon}_{i.}] = \frac{1}{n}\sum_j \underbrace{E[\epsilon_{ij}]}_{0} \qquad E[\bar{\epsilon}_{..}] = \frac{1}{an}\sum_{jk} \underbrace{E[\epsilon_{jk}]}_{0}$$

$$V[\hat{\tau}_i] = V[\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}] = \frac{a-1}{an}\sigma^2_\epsilon$$

We graded this part with d) below.

~2 b) Write the appropriate quantity that has a N(0,1) distribution. You may set $\bar{\tau}_. = 0$ hereafter.

We are trying to build a CI for τ. Then $z = \frac{\hat{\tau}_i - E[\hat{\tau}_i]}{\sqrt{V[\hat{\tau}_i]}} = \frac{\hat{\tau}_i - \tau_i + \overset{0}{\bar{\tau}_.}}{\sqrt{\frac{a-1}{an}\sigma^2_\epsilon}}$

~~2~~ (correction) c) Write the appropriate quantity that has a t distribution with $df = an - a$.

$$t = (\hat{\tau}_i - \tau_i)/\sqrt{\frac{a-1}{an} MSE} \qquad (\sigma^2_\epsilon \to MSE)$$

~~2~~ 1 d) Now, prove the identity used above. Assume we already know $V[\bar{\epsilon}_{i.}] = \sigma^2_\epsilon/n$, and $V[\bar{\epsilon}_{..}] = \sigma^2_\epsilon/(an)$.

$$V[\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}] = \underbrace{V[\bar{\epsilon}_{i.}]}_{\frac{\sigma^2_\epsilon}{n}} + \underbrace{V[\bar{\epsilon}_{..}]}_{\frac{\sigma^2_\epsilon}{an}} - 2\underbrace{Cov[\bar{\epsilon}_{i.}, \bar{\epsilon}_{..}]}_{\frac{1}{a}\sum_j^a Cov[\bar{\epsilon}_{i.}, \bar{\epsilon}_{j.}]}$$

$$\underbrace{\frac{1}{a}\sum_j^a Cov[\bar{\epsilon}_{i.}, \bar{\epsilon}_{j.}]}_{\text{only } j=i \text{ survives}} = V[\bar{\epsilon}_{i.}] = \frac{\sigma^2_\epsilon}{n}$$

$$= \frac{\sigma^2_\epsilon}{n}\left(1 + \frac{1}{a} - \frac{2}{a}\right) = \frac{\sigma^2_\epsilon}{n}\left(\frac{a-1}{a}\right)$$

To finish the problem: C.I. for τ_i: $\hat{\tau}_i \pm t_{critical}\sqrt{\frac{a-1}{an} MSE}$.

FYI

~~2~~ 1.5 hw/lect 3-3 similar

10. It is easy to show that the pdf of the sample standard deviation s is given by $f_s(t) = 2(\frac{n-1}{\sigma^2})t f_W(\frac{n-1}{\sigma^2}t^2)$, where f_W is the pdf of the chi-squared random variable with $df = n-1$. Show that $E_s[s] = \frac{\sigma}{\sqrt{n-1}} E_W[\sqrt{W}]$.

$$E_s[s] = \int_0^\infty t\, f_s(t)\,dt = \int_0^\infty t\, \underbrace{2\left(\tfrac{n-1}{\sigma^2}\right) t\, f_W\left(\tfrac{n-1}{\sigma^2}t^2\right) dt}_{du}$$

$u = \frac{n-1}{\sigma^2} t^2$

$du = 2\left(\frac{n-1}{\sigma^2}\right) t\, dt$

$$= \int_0^\infty \sqrt{\frac{\sigma^2}{n-1}} \sqrt{u}\, f_W(u)\,du = \sqrt{\frac{\sigma^2}{n-1}} \underbrace{\int_0^\infty \sqrt{u}\, f_W(u)\,du}_{E_W[\sqrt{W}]} = \frac{\sigma}{\sqrt{n-1}} \underbrace{E_W[\sqrt{W}]}_{2\frac{\Gamma(1/2)}{\Gamma(\frac{n-1}{2})}} \neq \sigma$$

biased

FYI

~~2~~ 1 2016/11, 2014/12

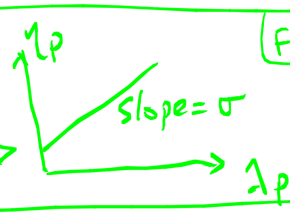
11. It is easy to show that $\eta_p = \mu + \lambda_p \sigma$, where η_p and λ_p are the p^{th} quantile (or percentile) of $N(\mu, \sigma)$ and $N(0,1)$, respectively. It is this identify that allows us to look at a normal qqplot and identify a linear pattern as indication of normality, and to identify the y-intercept and slope of the linear pattern as the μ and σ, respectively, of the Nomal distribution from which the data must have come. It's also easy to show that the p^{th} quantile of an exponential distribution with parameter λ is $(-1/\lambda)\log(1-p)$. Suppose we make a plot of the quantiles of some data versus quantiles of an exponential distribution with $\lambda = 1$, and find a linear pattern. How are the intercept AND the slope of that line related to the λ parameter of the exponential distribution from which our data must have come?

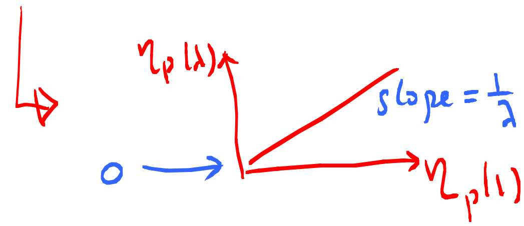
($\eta_p = \eta_p(\mu,\sigma)$, $\lambda_p = \eta_p(0,1)$)

$$\left.\begin{aligned} \eta_p(\lambda) &= -\tfrac{1}{\lambda}\log(1-p) \\ \eta_p(\lambda{=}1) &= -1\log(1-p) \end{aligned}\right\} \Rightarrow \frac{\eta_p(\lambda)}{\eta_p(1)} = \frac{1}{\lambda} \Rightarrow \eta_p(\lambda) = \underbrace{\tfrac{1}{\lambda}}_{\text{slope}}\eta_p(1) + \underbrace{0}_{\text{intercept}}$$

FYI

$\eta_p = \mu + \lambda_p \sigma$





~~2~~ 1.5 2016/12

12. How does the CI for a contrast Γ change if the contrast vector is multiplied by a positive nonzero constant, k, i.e., $c_i \to kc_i$?

C.I. for Γ : $\hat{\Gamma} \pm t_{crit}\sqrt{MSE \frac{1}{n}\sum_i^a c_i^2} \longrightarrow k\left[\hat{\Gamma} \pm t_{crit}\sqrt{MSE\frac{1}{n}\sum_i^a c_i^2}\right]$

$\hat{\Gamma} = \sum c_i \bar{Y}_{i.} \rightarrow k\sum c_i \bar{Y}_{i.}$

MSE: No c_i

$\sum_i (kc_i)^2 = k^2 \sum c_i^2$

I.e. $[CI] \longrightarrow k[CI]$.

FYI

The p-value does not change, But The C.I. does. For example, instead of estimating $\mu_1 - \mu_2$, we will be estimating $k(\mu_1 - \mu_2)$. So, This makes sense.

This document was created with Win2PDF available at http://www.win2pdf.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.