Name: ________________________
ID: ____________________________

Stat 421, Test 1, Fall, Oct. 19, 2016; Marzban

9 +17

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

By clustering.

2 **1.** Which distributions are involved in testing

a) a single mean	z, t	d) two variances	F
b) a single variance	χ^2	e) three or more means	F
c) two means	z, t, F	e) a contrast	t (or z)

2014/Q3

1 **2.** In comparing data from two populations, comparative boxplots are useful only when
a) the samples in each population are normal
b) the samples in each population have the same variance
c) the two samples are independent (e.g., not paired)
d) none of the above.

Note: If paired, x_i may be larger than y_i, for each i, in which case μ_x and μ_y are diff. But boxplots may still show big overlap.

2015/Q4

1 **3.** For which of the following one canNOT have a sampling distribution? Circle all right answers.
a) sample variance b) sample minimum c) sample maximum d) population variance.

not a r.v.

hw.F, lect3

1 **4.** For which of the following distributions one canNOT compute a mean and variance? Circle all appropriate answers.
a) Normal b) Exponential c) Chi-squared d) F e) none of the above.

lect3

1 **5.** The expressions $E[\bar{y}] = \mu_y$, $V[\bar{y}] = \sigma_y^2/n$ (where μ_y and σ_y^2 are population mean and variance, respectively) are NOT true when the population is
a) Normal b) Exponential c) Chi-squared d) F e) none of the above.

2,23, hw-X

1 **6.** Suppose we are looking at qqplots of Y for different levels of X, separately. Then, one can use transformations to make them
a) more straight b) more parallel c) both a and b d) neither a nor b.

more normal   more equal variance

Lect8

1 **7.** "Snooping" (i.e., using contrasts to test specific hypotheses) leads to
a) Lower power b) Increased Type I error c) higher contrast d) none of the above.

1 **8.** Joe is expected to manufacture a concrete whose mean breaking point μ is at least $13KN/m^2$. Although larger values of μ are desirable, increasing μ is an expensive process. But smaller values of μ can have disastrous consequences. Joe should set up the problem so that $\alpha = prob($ choose $\mu > 13$ $|$ $\mu < 13$ $)$. Fill in the blanks with statements regarding μ.

← Technically $prob(\mu > 13 | \cdots)$ makes no sense.

2013/Q8

~ 2 **9.** Let $f_X(t)$ and $f_Y(t)$ denote the pdf of X and $Y = X^2$, respectively. Since $E_X[X^2] = E_Y[Y]$, it follows that $\int t^2 f_X(t)dt = \int t f_Y(t)dt$. Is the above statement correct (Yes/No); if not, why not?

Yes, it is correct.

Nst: It may be confusing to see $\int t^2 \, f_X(t)\, dt = \int t \, f_Y(t)\, dt$, when both sides are supposed to be the same mean. But then you need to note that the $f(t)$ are different on the two sides.

~~~2~~ 1.5 hw-T **10.** In a problem with 1 factor with a levels, 1 response, and n replications, explain why the df of SSE is $n * a - a$.

$$SS = \sum_i^a \sum_j^n (Y_{ij} - \bar{Y}_{i.})^2$$

There are an terms in that sum. But $\sum_j^n (Y_{ij} - Y_{i.}) = 0$, for each $i = 1, \cdots a$, which is $\underline{a}$ different equations / constraints. Hence $df = an - a$.

~ 2 **11.** Let η_p denote the p^{th} quantile of $N(\mu, \sigma)$, and λ_p denote the p^{th} quantile of $N(0, 1)$. Show that $\eta_p = \sigma\ \lambda_p + \mu$. Hint: Start with the definition of the p^{th} quantile of $N(\mu, \sigma)$.

$$\frac{p}{100} = \int_{-\infty}^{\eta_p} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{N(\mu,\sigma)} dx = \int_{-\infty}^{\frac{\eta_p - \mu}{\sigma}} \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} e^{-\frac{1}{2}z^2} \cancel{\sigma}\, dz = \int_{-\infty}^{\frac{\eta_p - \mu}{\sigma}} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}_{N(0,1)} dz$$

$$z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{1}{\sigma} dx$$

$$\therefore \lambda_p = \frac{\eta_p - \mu}{\sigma}$$

$$\therefore \underline{\eta_p = \mu + \sigma \lambda_p}$$

~ 2 **12.** Show that if a contrast vector is multiplied by a positive nonzero constant, k, i.e., $c_i \to kc_i$, then the p-value of the test of $H_0 : \Gamma = 0,\ H_1 : \Gamma \neq 0$ is unaffected

If $c_i \to kc_i$, then $\overline{c^2} = \frac{1}{n}\sum_i c_i^2 \longrightarrow \frac{1}{n}\sum_i k^2 c_i = k^2 \overline{c^2}$

$$\Gamma = \sum_i c_i \mu_i \longrightarrow k \sum c_i \mu_i = k\Gamma$$

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

$$t = \frac{\hat{\Gamma} - \Gamma}{\sqrt{\underbrace{MSE}_{\text{no } \vec{c}} \cdot \overline{c^2}}} \longrightarrow \frac{\cancel{k}(\hat{\Gamma} - \Gamma)}{\sqrt{MSE \cdot \cancel{k^2}\, \overline{c^2}}} = t \Rightarrow \text{p-value}(t \,\square\, t_{obs}) = \text{unaffected}$$

~~~2~~ 1.5 Lect4 **13.** Let $Y \sim N(\mu, \sigma)$. Then, $z = \frac{\bar{y}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$. Suppose **we know the value of μ.** Starting from a "self-evident fact," construct the expression for the CI of σ. This is a "strange" (not hard) problem, but try anyway.

$$pr(-z^* < z < +z^*) = 1 - \alpha$$

$$pr\left(-z^* < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < z^*\right)$$

$$pr\left(-\frac{1}{z^*} < \frac{\sigma/\sqrt{n}}{\bar{Y} - \mu} < \frac{1}{z^*}\right)$$

It will be interesting to check the power of this test!

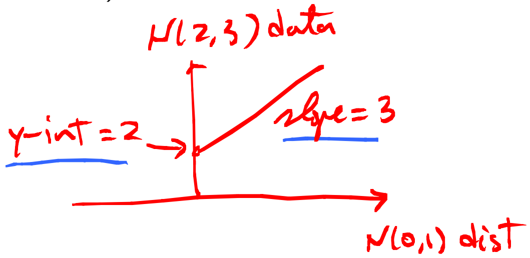
$$pr\left(-\frac{1}{z^*}(\bar{Y}-\mu)\cdot\sqrt{n} < \sigma < \frac{1}{z^*}(\bar{Y}-\mu)\sqrt{n}\right) \Rightarrow \text{C.I. for } \sigma : \pm \frac{(\bar{Y}-\mu)\sqrt{n}}{z^*}$$

Note: There is no s in this formula even though it's a C.I. for σ. We don't need s to estimate σ, because how far $\bar{y}$ is from μ is determined by σ. So, $\bar{y} - \mu$ carries the info we need to infer σ.

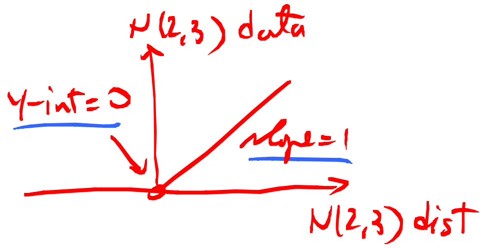
~ ~~3~~ 2.5

14. Draw the qqplot corresponding to the following situations; if the qqplot is a straight line, SPECIFY the (y-intercept, slope); if a qqplot is NOT possible, EXPLAIN why not.
Data are from N($\mu = 2, \sigma = 3$), and the x-axis of the qqplot shows quantiles of

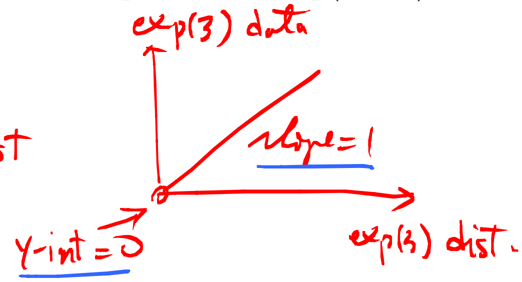
a) the standard normal:



b) N($\mu = 2, \sigma = 3$):

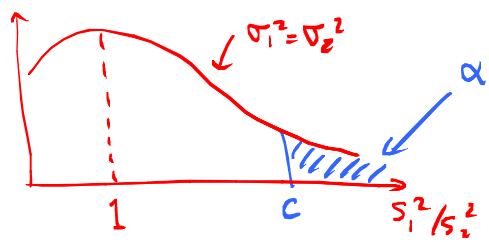


c) Data are from exp($\lambda = 3$) and the x-axis shows quantiles of exp($\lambda = 3$):



F-version of χ^2 in 2013

~ 3 **15.** Consider two independent samples, each of size 9, taken from two normal populations. We want to perform a test of $H_1 : \sigma_1^2 > \sigma_2^2$.
a) At $\alpha = 0.001$, find the numerical value of the rejection region for s_1^2/s_2^2 . Hint: draw a figure with the axes appropriately labeled.
b) What is the power if in fact $\sigma_1^2 = 2\,\sigma_2^2$?



$pr\left(\frac{s_1^2}{s_2^2} > c \,\middle|\, \underbrace{\sigma_1^2 = \sigma_2^2}_{H_0 = T}\right) = \alpha$

$pr\left(\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} > c\right) = \alpha$

$pr(F > c) = \alpha \Rightarrow \underline{c = 12.05}$

$power = pr\left(\frac{s_1^2}{s_2^2} > c \,\middle|\, \underbrace{\sigma_1^2 = 2\sigma_2^2}_{H_1 = T}\right)$

$= pr\left(\frac{s_1^2}{s_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} > c\cdot\frac{\sigma_2^2}{\sigma_1^2} \,\middle|\, \frac{\sigma_2^2}{\sigma_1^2} = \frac{1}{2}\right)$

$= pr\left(\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} > 12.05\cdot\frac{1}{2}\right)$

$= pr(F > 6.025) = .01$

F-Table with df=(9-1, 9-1)

hw-w

~ ~~3~~ 2.5 **16.** For the model $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, with $i = 1 \cdots a$, and $j = 1 \cdots n$, show that $E[SS_c] = \sigma_\epsilon^2 + \frac{\Gamma^2}{\cancel{\sigma_\epsilon^2} c^2}$, where SS_c is a contrast sum-of-squares. You may simply use the following, without derivation: $Cov[y_{i.}, y_{j.}] = Cov[\epsilon_{i.}, \epsilon_{j.}] = 0$ for $i \neq j$.

$E[SS_c] = E\left[\frac{\hat{\Gamma}^2}{c^2}\right] = \frac{1}{c^2} E[\hat{\Gamma}^2] = \frac{1}{c^2}\left(V[\hat{\Gamma}] + E^2(\hat{\Gamma})\right)$

$= \frac{1}{c^2}\left(V[\sum c_i \bar{Y}_{i.}] + E^2(\sum c_i \bar{Y}_{i.})\right)$

$= \frac{1}{c^2}\left(\sum_i c_i^2 \underbrace{V[\bar{Y}_{i.}]}_{\sigma_\epsilon^2/n} + 2\sum_{i<j} c_i c_j \underbrace{Cov[\bar{Y}_{i.}, \bar{Y}_{j.}]}_{0} + \underbrace{\left(\sum c_i \underbrace{E[\bar{Y}_{i.}]}_{\mu_i}\right)}_{\Gamma}^2\right)$

$= \frac{1}{c^2}\left(c^2\sigma_\epsilon^2 + \Gamma^2\right)$

$= \sigma_\epsilon^2 + \frac{\Gamma^2}{c^2}$

Note: If $H_0: \Gamma = 0$ is true, Then $E[SS_c/\sigma_\epsilon^2] = 1$, consistent with $SS_c/\sigma_\epsilon^2 \sim \chi^2_{1=df}$