Name: ______________________

ID: ________________________

10 + 13

Stat 421, Test 1, Fall, Oct. 14, 2013; Marzban

ONLY a half-size "cheat sheet" is allowed

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

For questions above horizontal line, do not explain.

For questions below horizontal line,

SHOW answer & work; NO CREDIT for correct answer without explanation

Points

Table 2.7

1. Circle all correct statements.

~1 a) chi-squared tests are always 1-sided c) F-tests are always 1-sided.

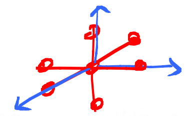
b) chi-squared tests may be 1- or 2-sided d) F-tests may be 1- or 2-sided.

0.5 each with 0.5 penalty

lect + book p.4

~ 1 **2.** Consider 3 factors, each having 3 levels. How many runs are required for a factorial design, and for a one-at-a-time design?

a) $(3 \times 3, 6)$ b) $(3 \times 3, 7)$ c) $(3^3, 6)$ d) $(3^3, 7)$



~ 2 **3.** In the previous problem, what is the answer if there are k factors, each having 3-levels?

$(3^k, (3-1)k + 1)$ take out the middle point, put it back in.

1 point each

2.28 1-sided

4. An article compares two procedures for predicting the shear strength for steel plates. For three plates, the following data are observed. We want to see if there is evidence to support the claim that method A has higher mean shear strength.

Plate	Method A	Method B
1	1.186	1.061
2	1.151	0.992
3	1.322	1.063

a) Write down the appropriate H_0, H_1.

~ 2

$H_0: \mu_A \le \mu_B$

$H_1: \mu_A > \mu_B$

b) Suppose all of the conditions for a t-test are satisfied.

~~~2~~ 1 What kind of t-test is appropriate? Specify as much as you can.

1-sided, welch, ---, paired, --- This is the important one.

2.12 1-sided

~ 2 **5.** The average viscosity of a liquid detergent is supposed to be below 800, otherwise laundary machines can be damaged. Write the appropriate H_0, H_1.

BAD error = $(\mu < 800 \mid \mu > 800)$.

So you better not assume $\mu < 800$

$\Rightarrow \begin{cases} H_0: \mu \ge 800 \\ H_1: \mu < 800 \end{cases}$

lect 7

~~~2~~1 **6.** Suppose we are interested in testing whether elasticity of a certain plastic varies across three temperatures. We have access to six plastic specimen, and so we measure the elasticity of two specimen at each of the three temperatures. What is/are the appropriate types of model(s)?

a) (means model, fixed effects) c) (means model, random effects)

b) (effects model, fixed effects) d) (effects model, random effects)

0.5 each with 0.5 penalty

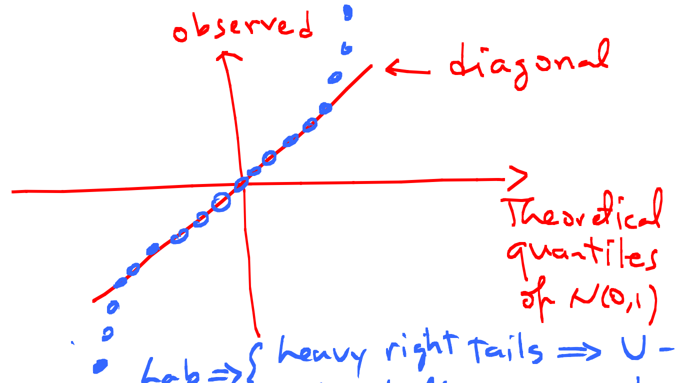
---

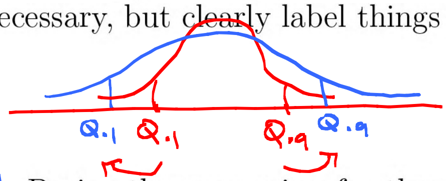
book p.75

~ 1 **7.** Our textbook discusses some of the advantages of a balanced design. One of the reasons is that equal sample size leads to higher power. Does this mean that if we get rid of the "extra" samples in one pop, then we'll have more power? Yes/no, and briefly explain.

No, because if we get rid of the "extras", then the overall sample size goes down, and so, power goes down, too.

~2 Lab 2 **8.** The t-distribution with small df has heavier tails than an analogous normal distribution. Draw an example of a normal probability plot (i.e., qqnorm) where the data have come from a t-distribution with small df. For reference, draw a diagonal line, as well. No explanations necessary, but clearly label things (axes, etc.).





Lab $\Rightarrow$ { heavy right tails $\Rightarrow$ U-shape
" left " $\Rightarrow$ dome-shape

~~~2~~ 2.5 lect 3 p.3,6 **9.** Derive the expression for the expected value of y^2 if y is normal with parameters μ, σ. Note: $\int_{-\infty}^{\infty} x^2 \exp^{-x^2/2} = \sqrt{2\pi}$

$z=(x-\mu)/\sigma$

$$E[y^2]=\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \overset{\downarrow}{=} \int_{-\infty}^{\infty} (\mu+\sigma z)^2 \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} e^{-\frac{1}{2}z^2} \cdot \cancel{\sigma}\, dz$$

$$= \mu^2 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}_{=1 \quad \int f(z)dz} + \sigma^2 \underbrace{\int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}_{=1 \text{ (Hint)}} + 2\mu \underbrace{\int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}_{=0 \text{ (odd)}}$$

$$= \boxed{\mu^2+\sigma^2}$$

~2 lect 3, p.6 Lect 4, p.1 **10.** Consider a sample of size n. We have shown that if the individual elements of the sample are drawn independently, then the variance of the sample mean is σ^2/n, where σ is the standard deviation of the population. Now, suppose the individual samples are not independent, and that the covariance is given by a constant K. Derive the expression for the variance of the sample mean.

$$V[\bar{y}] = V\left[\frac{1}{n}\sum_i^n y_i\right] = \frac{1}{n^2} V\left[\sum_i y_i\right] = \frac{1}{n^2}\left(\sum_i^n \underbrace{V[y_i]}_{\sigma^2\ \forall i} + 2\sum_{i<j} \underbrace{Cov[y_i, y_j]}_{K\ \forall i,j}\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 \cdot n + 2K\frac{n(n-1)}{2}\right)$$

$$= \frac{\sigma^2}{n} + \frac{n-1}{n}K$$

[$= K + \frac{1}{n}(\sigma^2 - K) \longrightarrow K$ as $n \longrightarrow \infty$
$\therefore$ $V[\bar{y}]$ is dominated by K (ie. by dependence)]

2.21 **11.** a) Suppose we are testing $H_0 : \mu_1 \leq \mu_2$ versus $H1 : \mu_1 > \mu_2$. We have evidence to indicate the two population standard deviations are equal. From a sample with $n_1 = 9, n_2 = 16$, we have estimated a common standard deviation of $\frac{24}{5}$. At $\alpha = 0.05$, what is the power of the test if the
~2 true difference $\mu_1 - \mu_2$ is 0.790?

$\alpha = p(t > t_c)$ H_1 where $t = \frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\quad df = n_1+n_2-2 \Longrightarrow t_c = 1.714$

$\therefore$ critical value: $(\bar{y}_1-\bar{y}_2)_c = t_c s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$$\text{power} = p\left(\bar{y}_1-\bar{y}_2 > (\bar{y}_1-\bar{y}_2)_c \mid \mu_1-\mu_2 = 0.790\right) = p\left(t > \frac{t_c s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} - .790}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

$$= p\left(t > t_c - \frac{.790}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right) = p\left(t > t_c - \frac{.790}{2\,\cancel{\frac{24}{5}}\sqrt{\cancel{\frac{25}{9(16)}}}}\right) = p\left(t > t_c - \frac{.790}{2}\right)$$

$$= p(t > t_c - .395) = p(t > 1.714 - .395) = p(t > 1.319) \overset{df=23}{=} \boxed{0.10}$$
~~~

~~2~~ ~1

b) Now, suppose we have good evidence to believe that the two populations have unequal standard deviations. Meanwhile, the sample standard deviations are both $\frac{24}{5}$. What is the power now? If you're short on time, just point out the places where the calculation in part a would be different.

This time, must use $df = \text{Welch} = \dfrac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2\left(\frac{1}{n_1-1}\right)+\left(\frac{s_2^2}{n_2}\right)^2\left(\frac{1}{n_2-1}\right)}$

But when $s_1^2 = s_2^2$, it drops out:

$$\therefore df = \frac{(n_1+n_2)^2\cdot(n_1-1)(n_2-1)}{(n_1-1)n_1^2+(n_2-1)n_2^2} = \frac{(25)^2\cdot \cancel{8}\cdot\cancel{15}^{5}}{\cancel{8}\cdot\cancel{9}_{3}\cdot 9+\cancel{15}_{5}\cdot 16\cdot\cancel{16}_{2}} = \frac{(25)^2\cdot 5}{27+160} = \frac{(25)^2\cdot 5}{187} = 16.7$$

$t_c = 1.335$

$$\text{power} = p\left(t > t_c - \frac{.790}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\right) = p(t > 1.335 - .395) = p(t > 0.94) \approx \boxed{.18}$$

$\Longrightarrow$ In short, df and the std. error $\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)$ change.

~~3~~ ~2.5

lect 7 **12.** Do the anova decomposition of SST (i.e., derive $SS_{between}$ and SS_{within} for unbalanced designs. The latter should be written in terms of the (conditional) variance of the response for each level of the treatment, and all of the limits of the sums should be clearly indicated.

$+\bar{Y}_{i.} - \bar{Y}_{i.}$

$$SST = \sum_i^a \sum_j^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_i^a \sum_j^{n_i} \left[(Y_{ij}-\bar{Y}_{i.}) + (\bar{Y}_{i.}-\bar{Y}_{..})\right]^2$$

$$= \underbrace{\sum_i^a \sum_j^{n_i} (Y_{ij}-\bar{Y}_{i.})^2}_{(n_i-1)s_i^2} + \sum_i^a \underbrace{\sum_j^{n_i}}_{n_i} (\bar{Y}_{i.}-\bar{Y}_{..})^2 + 2\sum_i^a \underbrace{\sum_j^{n_i} (Y_{ij}-\bar{Y}_{i.})}_{\sum_j^{n_i} Y_{ij} - \bar{Y}_{i.}\sum_j^{n_i} 1}(\bar{Y}_{i.}-\bar{Y}_{..})$$

$$n_i\bar{Y}_{i.} - n_i\bar{Y}_{i.} = 0$$

$$= \sum_i^a (n_i-1)s_i^2 + \sum_i^a n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

$$= SS_{within} + SS_{between}$$