

Notes about my answers are marked by •

General things about my grading:

- If I wrote math error, you had the right idea/equation, but made a mistake in the calculation. I deducted one point for these.
- The exam had many places where results from early parts were used in subsequent parts. I deducted points for the mistake when (and only when) they occurred. I did not deduct subsequent points because your answer(s) in later parts didn't match mine. If you did the right things in subsequent parts, I gave you full credit.
- These solutions indicate some common mistakes. If you don't understand a comment on your exam or something in these solutions, please come and see me (or call/e-mail if you're offcampus).
- If I misadded points, or you don't understand why I deducted points, please see / call / e-mail me.

1. Making components - design based inference

(a) experimental study. Treatments are randomly assigned to components (the e.u.'s).

(b) $p = 15/1000 = 0.015$.

The observed ratio = $14.65/7.34 = 1.996$. The p-value comes from the randomization distribution. How many values are ≥ 1.996 ? I count 14. So $p = (N+1)/(R+1) = 15/1000$.

(c) $T = (1.996 - 1.5)/0.0946 = 5.24$

- I didn't ask you for a p-value because the d.f. here are not trivial. The usual 2-sample formula doesn't apply because the s.e. is not $s_p\sqrt{1/n_1 + 1/n_2}$.

(d) Not reasonable. The randomization distribution is not symmetric.

- Various folks said that the randomization distribution was not centered at 0. That's not a problem. The rand. distribution looks centered at 1.5. If you calculated the rand. distribution of ratio-1.5, that would be centered at 0. The T distribution is centered at 0 because the T formula includes a -1.5.

- Saying 'check assumptions' got no credit because the usual assumptions are for appropriate tests of differences, not ratios. You have to think about the randomization distribution.

2. Litter size in pigs

(a) The mean of the experimental group is known more precisely because it has the smaller s.e.

- You are looking for the group with the smallest s.e. Since you do not assume equal variances, calculate the s.e. for each group using its own s.d. I get control = 0.54 and expt = 0.47. Because you are not pooling variances, larger sample size is not sufficient. Because the sample sizes are unequal, larger s.d. is not sufficient.

- (b) $T = 0.48$, $0.6 < p < 0.8$. No evidence of a difference between the two means.
- I decided to pool the sd's, which gave: $s_p = \sqrt{(19 * 2.4^2 + 9 * 1.5^2)/28} = \sqrt{4.63} = 2.15$
 $T = (10.3 - 10.7)/(2.15 * \sqrt{(1/20 + 1/10)}) = 0.4/0.833 = 0.48$. This has 28 d.f. T is close to 0, so p is large. The tables allow you to bracket it.
 - You could decide to not pool. That's fine. The T statistic is similar. The problem is the d.f. You got full credit if you carried through the Satterthwaite approximation. You lost 2 points if you used the pooled 28 d.f.
- (c) Various methods are acceptable for each assumption. The most commonly named methods were:
- | | |
|---------------------|---|
| Assumption | Method |
| normality of errors | QQ plot of residuals |
| independence | evaluate the design, compare ou to eu |
| equal variance | Levene's test, plot residual vs predicted value |
- (d) Use a block design, with blocks of 3 sows with the same mother. In each block, 2 sows get the control treatment and one sow gets the experimental treatment.
- A number of folks proposed a paired design: 15 blocks, with one sow receiving the control and the other sow receiving the experimental treatment. This is a fine design; the only difficulty is that the problem said to keep the same sample sizes (i.e. 20 control and 10 experimental). The paired design gives you 15 control and 15 experimental sows.
- (e) There are 7 significant markers at $FDR = 0.05$. They are 427, 157, 10, 495, 86, 259, and 121.
- The Benjamini and Hochberg FDR adjustment depends on $p * 500/i$ where i is the index in the sorted p-values. The furthest down the list with adjusted values < 0.05 is marker 121.
 - Although marker 259 has an adjusted value > 0.05 , it is included because there is a subsequent marker with adjusted value < 0.05 .
- (f) There are 3 significant markers if you use a Bonferroni correlation. They are 427, 157, and 10.
- The Bonferroni correction for 500 tests is to compare the p-value to $\alpha/500$. That is all p-values < 0.0001 . Equivalently, you can compare $p * 500$ to 0.05.

3. Sterilizing peas

- (a) e.u. = batch of canned peas, because a batch is randomly assigned to a treatment
o.u. = batch of canned peas, because there is one observation from each batch

- (b) Filled in ANOVA table:

Source	df	SS	MS	F
Days	5	11.68	2.34	2.41
Treatments	4	38.25	9.56	9.87
Error	20	19.36	0.968	
Total	29	69.29		

- The d.f. should be straightforward. You didn't need to compute any SS from the means,

once you have the d.f. The error SS is MSerror * DFerror. Then, remember that the SS sum to the SStotal, so SSstr = SStotal - SSdays - SSstr - SSerror. MStrt and the F statistic were straightforward.

(c) $Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$

- Another acceptable model is $Y_{ij} = \mu_j + \beta_i + \varepsilon_{ij}$

(d) The efficiency is 1.24. Intermediate calculations are: there are I=6 blocks and J=5 treatments.

$$\sigma_{CR}^2 = \frac{(I - 1) MS_{blocks} + I(J - 1) MS_{error}}{IJ - 1} = \frac{5 \cdot 2.34 + 6 \cdot 4 \cdot 0.968}{29} = 1.20$$

$$Eff = \frac{\sigma_{CR}^2}{\sigma_{RCB}^2} = \frac{1.20}{0.968} = 1.24$$

(e) i. H0: all treatments have the same mean, $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E$. Or, all treatment effects (τ_i) are zero.

ii. The appropriate F distribution is the $F_{4,20}$ distribution. (numerator d.f. = 4, denominator d.f. = 20). 9.87 is larger than the largest tabulated value, so $p < 0.001$.

iii. Strong evidence that at least one mean is different, or strong evidence that not all means are equal.

- If you reported $0.01 < p < 0.025$, you were comparing the observed F statistic to the $F_{20,4}$ distribution. That would be appropriate for 21 treatments with error d.f. = 4.

- “Treatments have different means” is wrong, because it may not be true. Most treatments may have the same mean. “Reject H0” is not a very clear conclusion.

- If you made mistake(s) earlier, I adjusted the expected answer and did not deduct points twice.

(f) Tukey (or Tukey-Kramer) because you’re interested in the family of all pair-wise comparisons.

(g) i. Contrast coefficients:

	A	B	C	D	E
C-E	0	0	1	0	-1
(A+B+D)/3 - (C+D)/2	1/3	1/3	-1/2	1/3	-1/2

- A few folks proposed 1/2 and -1/2 for the C-E coefficients. This estimates 1/2 of the difference between C and E. You get the same test, but the question very specifically asked for the coefficients to estimate the contrast.

ii. Yes, they are orthogonal. $\sum b_i c_i / n_i = \sum b_i c_i / n = (0 + 0 + 1/2 + 0 - 1/2) / 10 = 0$.

(h) (-3.0, -0.7) The estimate is the difference in means: $2.95 - 4.817 = -1.87$ with s.e. = 0.57. You need the 0.975 quantile of a 20 df t distribution = 2.086. So the ci is $(-1.87 - 2.086 \cdot 0.57, -1.87 + 2.086 \cdot 0.57) = (-3.04, -0.68)$.

- Since the se is 0.57 and $se/3 = 0.19$, I rounded my answer to the 0.1 place.
- You can compute the estimate and s.e. by hand or get them from the SAS output.
- Quite a few folks used the wrong d.f. Remember, the s.e. is computed using the MSE, so you look at the d.f. for that MSE.

(i) You need to find the SS between A, B, and D. That represents SS for two orthogonal contrasts (e.g. A-B, and (A+B)/2 - D). The ANOVA table gives you SS treatments (4 d.f.). The SAS output gives you SS for two additional contrasts: C-E and (A+B+D)/3 - (C+E)/2. These two are orthogonal to the two desired contrasts, so you can get the desired SS by $SS_{\text{treatments}} - SS_{\text{ce}} - SS_{\text{abcd}} = 38.25 - 10.453 - 27.38 = 0.42$. This has 2 d.f. Hence, the F statistic is $F = (0.42 / 2) / 0.968 = 0.22$, which is compared to a $F_{2,20}$ distribution. The observed F is tiny, so $p > 0.5$.

- This was the hardest problem on the exam. I was glad to see quite a few folks get it.
- Saying you needed the SS among the A, B, and D treatments got you partial credit. You're on the right track, but you could compute that from the available information.
- Proposing 3 pairwise tests did not get you any credit, because this is three tests of three hypotheses. Remember, we didn't test equality of the 5 means in the original problem by doing 10 pairwise tests.

(j) Stick with the current design because the efficiency is larger than 1.

- The proposed design for the new study is a CRD. Since the efficiency of blocking is much larger than 1, you should stick with the previous design.

(k) $n = 97$ warehouses.

- Yes, I wouldn't ever want to use this large number. However, design computations can be very useful to identify unreasonable expectations.

Details: variance of the specific contrast $(A+B/2 - D) = 1.5 \sigma^2/n$. Hence, the general formula is

$$\delta = (t_{1-\alpha/2} + t_{1-\beta})\sqrt{1.5\sigma^2/n}$$

which leads to

$$n \geq 1.5(t_{1-\alpha/2} + t_{1-\beta})^2\sigma^2/\delta^2.$$

Using z quantiles to get started gives $n > 96.1$. The experiment with 97 warehouses and 3 treatments will have $(2)(96) = 192$ error d.f., so the appropriate t quantiles are essentially the z quantiles, so no need to iterate.

- A number of folks used the 'difference between two means' approach, which has a 2 instead of the 1.5. That gives $n > 128$, I believe.
- A number of folks forgot to check the degrees of freedom. The computations using z quantiles are only appropriate if the computed n produces a design with a large error d.f.
- If I wrote 'one more round' on your paper, that is because you didn't repeat the calculation for a second iteration. Remember, the first calculation is based on a guess for the d.f. If you said that repeating was not necessary because the df are large (and you started with something like z quantiles, that's fine.