

STATISTICS 368
SAMPLE FINAL EXAM

Note: You should attempt these questions only after reviewing the entire course. If you have difficulty on any of them, then you should take this as an indication of an area in which more review is required. SOLUTIONS will not be posted. You are most welcome to see me to check your solutions, etc.

Time: 3 hours.

1. The following short questions are designed to test general knowledge; they do not pertain to any particular designs.
 - (a) Suppose that the population of women's heights is normally distributed, with a variance of 1 cm. I am curious as to whether or not the mean height is 165 cm. To test the hypothesis that it is 165 cm. I observe that a certain (randomly selected) woman has a height of 166 cm. On the basis of this, how do I compute the p -value associated with my hypothesis?
 - (b) What are the three most important techniques of experimental design?
 - (c) Explain why, in a fixed effects model which contains an overall mean and various treatment effects, we can always assume that the sum of the treatment effects is zero.
2. In a one way analysis of variance the effects model is

$$E[Y_{ij}] = \mu + \tau_i, \quad (j = 1, \dots, n, \quad i = 1, \dots, a); \quad \sum_i \tau_i = 0.$$

- (a) The sum of squares function for this model is

$$S(\mu, \boldsymbol{\tau}) = \sum_{i,j} (Y_{ij} - E[Y_{ij}])^2 = \sum_{i,j} (Y_{ij} - \mu - \tau_i)^2.$$

Write down the least squares estimates of μ and τ_i .

- (b) Prove your answer in (a) by carrying out the decomposition of $S(\mu, \boldsymbol{\tau})$ into a sum of three non-negative terms, one of which does not involve $(\mu, \boldsymbol{\tau})$ and the other two of which equal 0 when evaluated at the estimates you wrote down in (a).

3. In order to investigate the relative effectiveness of three insecticides, a sample of 30 fields was compiled. One of the insecticides was applied in each of ten fields, and one hour later the abundance of insects was measured.
- (a) What is the name of the design being used here?
 - (b) Describe the nature of the randomization which should be used. Are all 30 observations made in a completely random order, or is some other mechanism used?
 - (c) Write down the (fixed) effects model describing the abundance. Remember to include any relevant constraints on the parameters.
 - (d) Using the terms which you introduced in (c), write down expressions for
 - i. the mean abundance when insecticide i is used,
 - ii. the effect (on the abundance) of insecticide i , over and above any effect which is common to all insecticides.
 - (e) Using the terms which you introduced in (c), write down expressions for the estimates of
 - i. the mean abundance when insecticide i is used,
 - ii. the effect (on the abundance) of insecticide i , over and above any effect which is common to all insecticides.
4. (a) Write down a 3×3 Graeco-Latin square.
- (b) Suppose that you wish to investigate 3 fertilizer types, by applying them to crops in 3 regions of land. Each region is only large enough for two crops, so that a Balanced Incomplete Block design is called for. Write down an appropriate design. What is λ for your design? Give both the meaning of λ , and its numerical value.
5. Consider a 2^2 factorial design with factors A and B, for which the effects model is that the expected value of the k^{th} observation with A at level i ($i = 1$ for low, $i = 2$ for high) and B at level j ($j = 1$ for low, $j = 2$ for high) is

$$E[y_{ijk}] = \mu + A_i + B_j + (AB)_{ij}.$$

- (a) In terms of y_{ijk} and appropriate averages, give expressions for
 - i. the estimate of the main effect of the high level of A,
 - ii. the estimate of $(AB)_{ij}$.

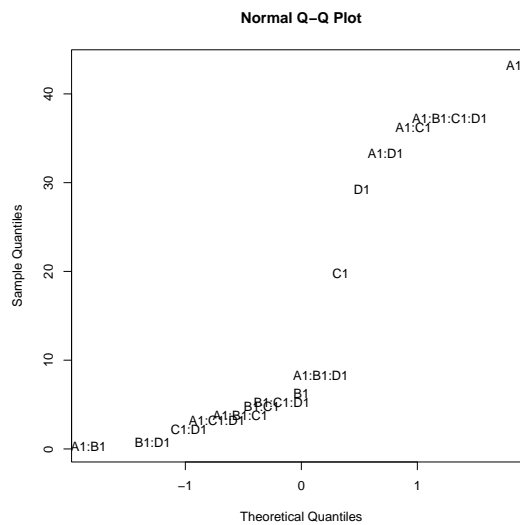
(b) The interaction effect is defined to be

$$AB = \frac{[(AB)_{22} - (AB)_{12}] - [(AB)_{21} - (AB)_{11}]}{2}.$$

Use your answer in (a) to show that this is estimated by

$$\widehat{AB} = \frac{\bar{y}_{22\cdot} - \bar{y}_{12\cdot} - \bar{y}_{21\cdot} + \bar{y}_{11\cdot}}{2}.$$

6. Consider the following half-normal plot, which resulted from an analysis of a 2^4 factorial design with factors A, B, C and D.



What do you judge to be the most significant effects?

7. Suppose that one replicate of a 2^4 factorial experiment must be run in 4 blocks. The experimenter chooses two effects ABC and ACD to confound with blocks.
- (a) What are the two defining contrasts?
- (b) Derive and write down the treatment combinations which are to appear in each of the four blocks.

8. Three replicates of a 2^3 factorial experiment are run, each in two blocks. In the first replicate, the three factor interaction is confounded with blocks. In the second replicate, it is AB which is confounded. In the third it is AC. After running the experiment, an analysis on R was performed, starting with the commands

```
h <- lm(y ~ Rep + Block%in%Rep + (A + B + C)^3)
anova(h)
```

- (a) Fill in the degrees of freedom in the following table:

Rep	
A	
B	
C	
Rep:Block	
A:B	
A:C	
B:C	
A:B:C	
Residuals	

- (b) If one wishes to calculate an estimate of AB which is not confounded in any way, then how many of the observations can be used in the calculation?
9. An experimenter wished to carry out a 2^4 factorial design, but has only enough resources for 8 runs. She decides to run the principal half of the design.
- (a) Write down the eight combinations of factors which are to be run.
- (b) How is the AB interaction estimated in this design?
- (c) Is AB aliased? If so, with what?
- (d) Suppose that the experimenter acquired more resources, and ran the complementary fraction. Can she now compute an unbiased estimate of CD? If so, how?

10. Consider the mixed effects model

$$\begin{aligned} y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \\ i &= 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n, \end{aligned}$$

in which the τ_i are fixed effects and β_j and $(\tau\beta)_{ij}$ are random. The expected mean squares for this model are

$$\begin{aligned} E[MS_A] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1}, \\ E[MS_B] &= \sigma_\varepsilon^2 + an\sigma_\beta^2, \\ E[MS_{AB}] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2, \\ E[MS_E] &= \sigma_\varepsilon^2. \end{aligned}$$

Suppose that the following ANOVA was obtained from R.

	Df	Sum Sq	Mean Sq	F value
A	2	42.30	21.15	21.36
B	3	12.62	4.21	4.25
A:B	6	27.05	4.51	4.55
Residuals	12	11.88	0.99	

There are three levels of Factor A. Write down the formula for simultaneous 95% confidence intervals on all three differences $\tau_i - \tau_{i'}$ ($i < i'$). Be explicit - give the numerical value of anything which can be gotten from this output, and explain how any other required numbers would be obtained.

11. An agricultural experiment, to test the effectiveness of certain crop stimulants on certain crops, is carried out as follows. Three 100 hectare plots of land are selected and randomly labelled 1,2,3. In plot 1, wheat is grown. In plot 2, rye is grown and in plot 3, canola is grown. Then each plot is divided into four 25 hectare fields and the four possible stimulants are applied, one to a field, in random order. This entire experiment is then replicated 4 times.

- What is the name of the design being used here?
- Viewing replicates as a random factor, clearly write down the effects model which you would use to analyze these data. Be sure to describe any constraints on, or distributions of, the terms in your model.

12. An experiment was carried out to test the effectiveness of three drugs. The experimenters originally thought of using a completely randomized design, treating the drug types as the levels, and fitting the one way model $E[y_{ij}] = \mu + \tau_i$. Here y_{ij} is the response of the j^{th} subject ($j = 1, 2, 3, 4$) to the i^{th} drug type ($i = 1, 2, 3$). It was then thought that much of the variation might be due to the covariate $x = \text{'Age of the subject'}$, and so the experimenters measured x along with y and fit the model

$$y_{ij} = \mu + \tau_i + \beta X_{ij} + \varepsilon_{ij},$$

with $X_{ij} = (x_{ij} - \bar{x}_{..})$. The following ANOVA was obtained.

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
drug	2	1.17	0.58	0.0329	0.9678
X	1	1097.48	1097.48	61.8192	4.946e-05
Residuals	8	142.02	17.75		

In order to test the hypothesis that the drugs are equally effective, the experimenters then dropped the τ_i from the model and re-ran the ANOVA, obtaining

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	1078.15	1078.15	66.342	1.006e-05
Residuals	10	162.51	16.25		

- What is the null hypothesis to be tested?
- What is the value of F_0 , the F-statistic used to test this hypothesis?
- Do you think the drugs are equally effective? Why or why not?