

**Instructions:** This is a closed-notes, closed-book exam. No calculator or electronic device of any kind may be used. Use nothing but a pen or pencil. Please write your name and answers on blank paper. Please do NOT write your answers on the pages with the questions. For questions that require extensive numerical calculations that you should not be expected to do without a calculator, simply set up the calculation and leave it at that. For example,  $(3.45 - 1.67)/\sqrt{2.34}$  would be an acceptable answer. On the other hand, some quantities that are very difficult to compute one way may be relatively easy to compute another way. Part of this exam tests your ability to figure out the easiest way to compute things, based on the information provided and the relationships between various quantities. In general, I expect you to be able to invert diagonal matrices of any dimensions and non-diagonal  $2 \times 2$  matrices. If you find yourself trying to deal with something more complicated, there is probably a better way to solve the problem.

- Each of three labs independently conducted a large experiment to estimate the effect of a new drug relative to an existing control treatment. Let  $\delta$  represent the mean response for the new drug minus the mean response for the existing control treatment. The estimate of  $\delta$  and the standard error of the estimate of  $\delta$  obtained by each lab are provided in Table 1 below.

**Table 1**

Lab	Estimate of $\delta$	Standard Error of the Estimate of $\delta$
1	5.5	2.0
2	6.7	3.0
3	1.4	1.0

For  $i = 1, 2, 3$ , let  $y_i$  be the estimate of  $\delta$  obtained by lab  $i$ , and suppose

$$y_i = \delta + \varepsilon_i,$$

where  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are independent and normally distributed with zero mean. Furthermore, suppose  $\text{Var}(\varepsilon_i) = \sigma_i^2$  for all  $i = 1, 2, 3$ .

- Based on the information provided, determine estimates of  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$ .

For all remaining parts of this problem, assume the true values of  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$  are known and equal to the estimates you provided in part (a).

- Find the BLUE of  $\delta$ .
- Find the standard error of the BLUE of  $\delta$ .
- Now suppose, for  $i = 1, 2, 3$ ,

$$y_i = \delta + u_i + \varepsilon_i,$$

where  $\delta$  and  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are defined as before;  $u_1, u_2, u_3 \stackrel{iid}{\sim} N(0, \tau^2)$ ; and  $u_1, u_2$ , and  $u_3$  are independent of  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$ . Provide an expression for an unbiased estimator of  $\tau^2$ , and prove that your estimator is unbiased. More than one answer is possible. Full credit will be given for any unbiased estimator accompanied by a correct proof. Thus, you may wish to report a relatively simple unbiased estimator even if it does not make full use of all the available data.

2. An Internet retailer sells merchandise to customers who visit the retailer's website. In order to purchase merchandise, a customer ultimately must visit the retailer's checkout webpage. This checkout webpage has several fields that a customer must complete (e.g., mailing address, credit card information, shipping option selection, etc.) and can include additional features like minimum purchase amounts or minimum purchase amounts to qualify for free or reduced-cost shipping. The retailer is considering four specific setups (labeled S1, S2, S3, and S4) for their checkout webpage that vary in field arrangement and/or minimum purchase requirements. The retailer is also considering three different choices for the background image (BI1, BI2, or BI3) that is shown to customers upon checkout. A total of 100,000 customers were randomly assigned to the four checkout webpage setups, with 25,000 customers for each setup. Whenever one of these 100,000 customers visited the checkout webpage, he or she was always presented with the same checkout webpage setup that was randomly assigned to the customer at the beginning of the study. However, the background image on the checkout webpage was varied so that each customer saw the three background images in random order over the course of his or her first three visits to the checkout webpage. The dollar amount spent by the customer on each of his or her first three visits to the checkout webpage was recorded. Example data for 3 of the 100,000 customers is displayed in Table 2 below.

**Table 2**

Customer	Visit	Setup	Background Image	Dollars Spent
1	1	S3	B2	0.00
1	2	S3	B1	23.18
1	3	S3	B3	42.95
2	1	S1	B3	59.99
2	2	S1	B2	0.00
2	3	S1	B1	13.75
3	1	S4	B1	32.55
3	2	S4	B3	114.00
3	3	S4	B2	98.99

The retailer would like to know which combination of checkout webpage setup and background image is best for business in the sense that it encourages customers to spend more dollars on each visit to the checkout webpage. For the sake of this problem, assume the dollar amount spent by each customer on each visit follows a normal distribution.

- (a) Name the experimental units in this experiment.
- (b) Consider the model you would fit to these data and create a corresponding ANOVA table with columns labeled *Source* and *Degrees of Freedom*. You are not required to write down your model; you are only required to provide the entries in the *Source* and *Degrees of Freedom* columns of the ANOVA table corresponding to your model.

3. A large company that owns a chain of restaurants was interested in determining which of four renovation styles would be most profitable. Two restaurants with similar profit histories were chosen in each of six US cities. Within each city, two of the four renovation styles (labeled RS1, RS2, RS3, and RS4) were randomly assigned to the two restaurants according to the design in Table 3 below.

**Table 3**

City	Restaurant 1	Restaurant 2
1	RS1	RS2
2	RS1	RS3
3	RS1	RS4
4	RS2	RS3
5	RS2	RS4
6	RS3	RS4

Each of the 12 restaurants in the study was renovated according to its assigned renovation style. Restaurant profit in the year following renovation was recorded for each of the 12 restaurants. Let  $y_{ij}$  denote the profit of the restaurant renovated according to renovation style  $RS_i$  in city  $j$ . Note that  $y_{ij}$  is available for only some combinations of  $i$  and  $j$ . For those combinations of  $i$  and  $j$  occurring in the data, suppose

$$y_{ij} = \mu_i + u_j + e_{ij},$$

where  $\mu_i$  is an unknown constant for  $i = 1, 2, 3, 4$ ;  $u_1, \dots, u_6 \stackrel{iid}{\sim} N(0, \sigma_u^2)$ ; and the  $e_{ij}$  terms are distributed as  $N(0, \sigma_e^2)$  and independent of each other and the  $u_j$  terms. For  $i = 1, 2, 3, 4$ , let  $\bar{y}_i$  be the average of responses whose first subscript is  $i$ ; in other words,  $\bar{y}_i$  is the average profit of restaurants remodeled according to renovation style  $RS_i$ . Consider the following partially completed ANOVA table that is constructed from sequential (Type I) sums of squares.

Source	DF	MS	EMS
Renovation Style  1	3	27.7	
City  1, Renovation Style	5	7.4	$\sigma_e^2 + 1.6\sigma_u^2$
Error	3	0.9	$\sigma_e^2$
Corrected Total	11		

- Compute the value of an unbiased estimator of  $\sigma_u^2$ .
- Derive  $E(\bar{y}_1 - \bar{y}_2)$ .
- Derive  $\text{Var}(\bar{y}_1 - \bar{y}_2)$ .
- Provide an estimate of  $\text{Var}(\bar{y}_1 - \bar{y}_2)$ .
- Determine the degrees of freedom associated with the estimate provided in part (d).

4. A square field was partitioned into four rows. Two types of fertilizer (labeled F1 and F2) were randomly assigned to the rows such that each fertilizer type was applied to two of the four rows. The same field was partitioned into four columns. Two types of pesticide (labeled P1 and P2) were randomly assigned to the columns such that each pesticide type was applied to two of the four columns. At the end of the growing season, the yield was measured for each plot formed by the intersection of a row with a column. The experimental layout with observed yields is provided in Table 4 below. Yields are presented as whole numbers to make calculations easier, but please assume that yields are normally distributed.

**Table 4**

		Column 1	Column 2	Column 3	Column 4
		P1	P2	P2	P1
Row 1	F1	9	2	7	10
Row 2	F2	5	1	3	3
Row 3	F1	12	6	7	11
Row 4	F2	10	3	7	8

A researcher (who has never taken a class like STAT510) generates the following R code and output.

```
> y=c(9,2,7,10,5,1,3,3,12,6,7,11,10,3,7,8)
> f=factor(c(1,1,1,1,2,2,2,2,1,1,1,1,2,2,2,2))
> p=factor(c(1,2,2,1,1,2,2,1,1,2,2,1,1,2,2,1))
> o=lm(y~f+p+f:p)
> anova(o)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
f           1     36   36.000   6.1714 0.028734 *
p           1     64   64.000  10.9714 0.006198 **
f:p         1      4    4.000   0.6857 0.423789
Residuals 12     70    5.833
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Use the data in Table 4 and/or the R code and output to complete the following parts.

- Compute the value of the  $F$  statistic you would use to test for a fertilizer-type main effect.
- Give the degrees of freedom associated with the test for a fertilizer-type main effect.
- Compute the value of the  $F$  statistic you would use to test for fertilizer-type  $\times$  pesticide-type interaction.
- Give the degrees of freedom associated with the test for fertilizer-type  $\times$  pesticide-type interaction.