Practice Problems

1. In the following questions, please choose the correct answer. There is only one correct answer in each question. If you select multiple answers, you will be given no credit.

   (a) Assume you have computed a *single-factor ANOVA F-value* and obtained a p-value, $p = 0.001$. Your conclusion is that

   (i) all the factor level means are equal.
   (ii) all the factor level means are statistically different.
   (iii) there is at least one factor level mean significantly different that the other.
   (iv) none of the above.

   (b) You just fitted a 2-way ANOVA model with interactions and you obtained the following ANOVA table (only partial information shown here):

Source	SS	DF	F-Ratio	p-value
Factor A	176.584	1	18.915	$<$0.0001
Factor B	242.570	1	25.984	$<$0.0001
Interaction of A*B	1.870	1	0.200	0.660

   The interaction term is **not** statistically significant and hence you remove it from the model. When you remove it, you expect that

   (i) The SS values of the main effects will increase, even slightly.
   (ii) The SS values of the main effects will decrease, even slightly.
   (iii) The SS values of the main effects will remain the same.
   (iv) We do not have enough information to conclude.

   (c) In simultaneous testing, the *family confidence interval* for a mean/difference/contrast is

   (i) always wider than the individual (i.e. non-family) confidence intervals.
   (ii) always narrower than the individual confidence intervals.
   (iii) always the same as the individual confidence intervals.
   (iv) none of the above.

   (d) When performing a *simultaneous test*, then

   (i) Tukey, Scheffe, and Bonferroni tests, all give the same conclusion.
   (ii) Tukey always performs better than Bonferroni and Scheffe.
   (iii) which test performs better depends on the hypotheses we test.
   (iv) none of the above.

   (e) In an ANOVA with $n = 1$, you cannot test for the interaction terms because

   (i) interactions are not present.
   (ii) you do not have any degrees of freedom left to test for the interaction.
   (iii) interactions are not statistically significant.
   (iv) none of the above.

2. An economist compiled data on productivity improvement for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (*low*, *moderate*, *high*). We had 9 firms with low expenditures, 12 firms with moderate expenditures and 6 firms with high expenditures. The productivity improvement is measured on a scale from 0 to 15. *Assume that the ANOVA model is appropriate.*

   (a) Write down the *cell means model* and draw (by hand) a representation of it. Make sure you explain the notation you are using.

   (b) Using the appropriate Table, estimate the effects α_1, α_2, α_3 that satisfy the sum constraint. (A point estimate is sufficient in this question). Interpret the meaning of the $\hat{\alpha}_1$ estimator in the context of the problem.

   (c) Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Control the α risk at 0.05. State the hypotheses, decision rule and conclusion.

   (d) Estimate the mean productivity improvement for firms with high research and development expenditures levels; use a 95% confidence interval. Use $|t_{24}(0.05/2)| = 2.064$.

   (e) Obtain a 95% confidence interval for $D = \mu_2 - \mu_1$. *Interpret* your interval estimate.

   (f) In *Problem 1: Table 4* you can see the Tukey 95% confidence intervals for all the pairwise comparisons. *Interpret the results.*

   (g) Estimate the following contrast with a 95% confidence interval:

$$L = \frac{\mu_1 + \mu_2}{2} - \mu_3$$

The **R** output for this problem is summarized below in pages 3 and 4. In your answers, please **specify which table(s) you are using**.

```r
means.improve <- lm(productivity~expenditure-1)
summary(means.improve)
```

```
##
## Call:
## lm(formula = productivity ~ expenditure - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43333 -0.50556  0.02222  0.53333  1.32222
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## expenditurehigh       9.2000     0.3266   28.17   <2e-16 ***
## expenditurelow        6.8778     0.2667   25.79   <2e-16 ***
## expendituremoderate   8.1333     0.2310   35.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8001 on 24 degrees of freedom
## Multiple R-squared:  0.9912, Adjusted R-squared:  0.9901
## F-statistic: 899.6 on 3 and 24 DF,  p-value: < 2.2e-16
```

Problem 1: Table 1

```r
contrasts(expenditure) = contr.sum(3)
anova.model.improve = lm(productivity~expenditure)
summary(anova.model.improve)
```

```
##
## Call:
## lm(formula = productivity ~ expenditure)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43333 -0.50556  0.02222  0.53333  1.32222
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.0704     0.1603  50.359  < 2e-16 ***
## expenditure1   1.1296     0.2475   4.565 0.000126 ***
## expenditure2  -1.1926     0.2222  -5.366 1.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8001 on 24 degrees of freedom
## Multiple R-squared:  0.5671, Adjusted R-squared:  0.531
## F-statistic: 15.72 on 2 and 24 DF,  p-value: 4.331e-05
```

Problem 1: Table 2

```
anova(anova.model.improve)
```

```
## Analysis of Variance Table
##
## Response: productivity
##             Df Sum Sq Mean Sq F value    Pr(>F)
## expenditure  2 20.125 10.0626   15.72 4.331e-05 ***
## Residuals   24 15.362  0.6401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 1: Table 3

```
TukeyHSD(aov(productivity~expenditure))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = productivity ~ expenditure)
##
## $expenditure
##                    diff        lwr         upr     p adj
## low-high      -2.322222 -3.3752471 -1.26919735 0.0000335
## moderate-high -1.066667 -2.0656538 -0.06767956 0.0347870
## moderate-low   1.255556  0.3745317  2.13657937 0.0043755
```

Problem 1: Table 4

3. An economist compiled data on productivity improvement for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (*low*, *high*). We had 9 firms with low expenditures, and 6 firms with high expenditures. The economist had also information on annual productivity improvement in the prior year that they decided to use as a continuous predictor. The productivity improvement (response) is measured on a scale from 0 to 15.

   (a) Based on the **R** output below, state

      (i) the regression line for the firms with low expenditures for research and development.

      (ii) the regression line for the firms with high expenditures for research and development.

   (b) Test whether or not the interaction term is statistically significant. State the hypotheses, decision rule and conclusion.

   (c) Does the productivity improvement vary according to the expenditure level? Justify your answer.

   The **R** output for this problem is summarized below in pages 6 and 7. In your answers, please **specify which table(s) you are using**.

```r
productivity.full = lm(productivity~past.improvement + expenditure + past.im
provement:expenditure)
summary(productivity.full)
```

```
##
## Call:
## lm(formula = productivity ~ past.improvement + expenditure +
##     past.improvement:expenditure)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32417 -0.12215 -0.03904  0.09955  0.55556
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -6.8697     2.0438  -3.361  0.00635 **
## past.improvement               1.3410     0.1704   7.872 7.61e-06 ***
## expenditurelow                 7.1479     2.1730   3.289  0.00721 **
## past.improvement:expenditurelow  -0.4086     0.1994  -2.049  0.06511 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2426 on 11 degrees of freedom
## Multiple R-squared:  0.9773, Adjusted R-squared:  0.9711
## F-statistic: 157.6 on 3 and 11 DF,  p-value: 2.564e-09
```

Problem 2: Table 5

```r
anova(productivity.full)
```

```
## Analysis of Variance Table
##
## Response: productivity
##                              Df  Sum Sq Mean Sq  F value    Pr(>F)
## past.improvement              1 25.3352 25.3352 430.4241 3.606e-10 ***
## expenditure                   1  2.2396  2.2396  38.0487 7.023e-05 ***
## past.improvement:expenditure  1  0.2471  0.2471   4.1976   0.06511 .
## Residuals                    11  0.6475  0.0589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 2: Table 6

```r
productivity.reduced = lm(productivity~past.improvement + expenditure)
summary(productivity.reduced)
```

```
##
## Call:
## lm(formula = productivity ~ past.improvement + expenditure)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44811 -0.15059  0.00333  0.12405  0.46475
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.29702    1.19958  -2.748  0.01765 *
## past.improvement  1.04287    0.09967  10.463 2.19e-07 ***
## expenditurelow    2.79362    0.50968   5.481  0.00014 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.273 on 12 degrees of freedom
## Multiple R-squared:  0.9686, Adjusted R-squared:  0.9633
## F-statistic:   185 on 2 and 12 DF,  p-value: 9.624e-10
```

Problem 2: Table 7

4. In a study for the effectiveness of different rust inhibitors, *four* brands (A, B, C, D) were tested, each on a different set of 5 units. The basic results were as follows:

i	Inhibitor	n_i	$\bar{Y}_{i.}$
1	A	5	43
2	B	5	89
3	C	5	67
4	D	5	40

The higher the mean, the more effective is the rust inhibitor.

(a) (*10 points*)Assume that an ANOVA model is appropriate. Write down the ANOVA model as a cell means model <u>and</u> as a factor effects model. Explain the notation you use (in terms of the experiment), draw a representation of your model and state the model assumptions.

(b) (*5 points*) We fit an ANOVA model and the results are summarized below:

Source of Variation	SS	df	MS
Between Inhibitors	7,893.75		
Error	72.00		
Total	7,965.75		

Fill in the blanks in the table.

(c) (*5 points*) Test whether the four rust inhibitors differ in effectiveness, using a level of significance 0.05. State the hypothesis, test statistic and conclusion. Interpret your results in the context of the problem. (The critical F value here is 0.56).

(d) (*5 points*) Brands B and C are nationally advertised. Brands A and D are private brands. Estimate the contrast

$$L_1 = \frac{\mu_2 + \mu_3}{2} - \frac{\mu_1 + \mu_4}{2}$$

with a 95% confidence interval. What does your confidence interval indicate about the comparative effectiveness of the two groups of brands?

(e) (*5 points*) Consider the following contrasts:

$$L_1 = \frac{\mu_2 + \mu_3}{2} - \frac{\mu_1 + \mu_4}{2}, \text{ and } L_2 = \mu_1 - \mu_4$$

Obtain simultaneous confidence intervals for the above contrasts with the Bonferrroni procedure using a 90% family confidence coefficient. *Interpret your results.*

5. A company studied the effect of two factors, brazing material and honeycomb structure material, on the bonding of the material. The observed variable was the number of cells (out of 512 cells in a panel) that are not completely bonded to the core. Twenty four panels of each of the treatment combinations were studied and the results are summarized below:

	Brazing Material (Factor B)				
Structure Material (Factor A)	GE81	AG-MN	Coast Metal 53	Electrolytic Copper	Row Total
17-7 PH	88 ($\bar{Y}_{11\cdot}$)	108 ($\bar{Y}_{12\cdot}$)	13 ($\bar{Y}_{13\cdot}$)	23 ($\bar{Y}_{14\cdot}$)	58 ($\bar{Y}_{1\cdot\cdot}$)
Titanium	63 ($\bar{Y}_{21\cdot}$)	46 ($\bar{Y}_{22\cdot}$)	40 ($\bar{Y}_{23\cdot}$)	29 ($\bar{Y}_{24\cdot}$)	44.5 ($\bar{Y}_{2\cdot\cdot}$)
Column Total	75.5 ($\bar{Y}_{\cdot 1\cdot}$)	77 ($\bar{Y}_{\cdot 2\cdot}$)	26.5 ($\bar{Y}_{\cdot 3\cdot}$)	26($\bar{Y}_{\cdot 4\cdot}$)	51.25($\bar{Y}_{\cdot\cdot\cdot}$)

(a) Based on the information given above, prepare an estimated interaction plot. Does your graph suggest that any interaction effects are present? Explain.

(b) Assume that the ANOVA model assumptions are satisfied. You fit an ANOVA model and the results are summarized below:

Source of Variation	SS	df	MS
Between Treatments	182,868		26,124
Factor A	8,748		8,748
Factor B	120,060		40,020
Interactions	54,060		18,020
Error	19,872		108
Total	202,740	191	

Fill in the missing degrees of freedom.

(c) Test whether or not the two factors interact. Using $\alpha = 0.05$, the critical value F is 2.66. State the alternatives, decision rule and conclusion. Do your results confirm your graphic analysis?

(d) Estimate μ_{23} with a 95% confidence interval. Interpret your interval estimate.