2015S Fountain*, Crain
2015S1 2016S1
2015S2 2019S3
2015S3 2018S2 2019S2

2015S4

2018S4 2019S1
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, $i = 1, .., n$, with the restriction that $\beta_1 - \beta_0 = 0$. Find the least-squares estimators of the regression coefficients.
Let $SSE = \sum_{i=1}^n (y_i - \beta_0 - \beta_0 x_i - \beta_2 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_0} = 2\sum_{i=1}^n (y_i - \beta_0 - \beta_0 x_i - \beta_2 x_i^2)(-1 - x_i) \overset{set}{=} 0$; $\hat{\beta}_0 = \frac{\sum_{i=1}^n (1+x_i) y_i - \hat{\beta}_2 \sum_{i=1}^n (1+x_i) x_i^2}{\sum_{i=1}^n (1+x_i)^2}$
$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^n (y_i - \beta_0 - \beta_0 x_i - \beta_2 x_i^2)(-x_i^2) \overset{set}{=} 0$; $\hat{\beta}_0 = \frac{\sum x_i^2 y_i - \hat{\beta}_2 \sum x_i^4}{\sum x_i^2 (1+x_i)}$
$\hat{\beta}_2 = \frac{\sum x_i^2 y_i \sum (1+x_i)^2 - \sum (1+x_i) y_i \sum x_i^2 (1+x_i)}{\sum x_i^4 \sum (1+x_i)^2 - [\sum x_i^2 (1+x_i)]^2}$
$\hat{\beta}_0 = \hat{\beta}_1 = \frac{\sum (1+x_i) y_i \sum x_i^4 - \sum x_i^2 y_i \sum (1+x_i) x_i^2}{\sum x_i^4 \sum (1+x_i)^2 - [\sum x_i^2 (1+x_i)]^2}$

2015F

2015F1

2016S1 [566-HW2-6] [8.3 The One-Quarter Fraction of the 2k Design p.344]
You must design an experiment to test six factors, each having two levels. Your budget will only allow sixteen runs. Furthermore, due to time constraints, only eight runs can be done on a given day, so you will have to conduct the experiment in 2 blocks. You may assume that 4-way and higher interactions are not important. Show all of the following:

- all of your generators (make sure that your resolution is at least III)

2_{IV}^{6-2} ABC=E, BCD=F; I=ABCE=BCDF=ADEF

- the 16 runs to conduct
- the eight runs to be done on each day

4Blk:ABD-ABF:1–;2+-;3-+;4++. $df_{noAE,ABD,ABF}$=1, df_{Blk}=3, df_T=15
2Blk:ABD:1+3,2+4; ABF:1+2,3+4. $df_{noABDorABF}$=1, df_{Blk}=1, df_T=15
ABF- Day1: (1), abce, bcdf, adef, abd, cde, acf, bef
ABF+ Day2: ae, bc, df, abcdef, abf, cef, acd, bde

- the effects to be confounded with blocks

ABF=ACD=BDE=CEF

- the Source and DF columns of the ANOVA table

A-F,AB,AC,AD,AE,AF,BD,BF,ABD,Block=1

Run	A	B	C	D	E	F	ABD	ABF		
(1)	-	-	-	-	-	-	-	-	1	I=ABCE=BCDF=ADEF
ae	+	-	-	-	+	-	+	+	4	A = BCE = DEF = ABCDF
bef	-	+	-	-	+	+	+	-	2	B = ACE = CDF = ABDEF
abf	+	+	-	-	-	+	-	+	3	C = ABE = BDF = ACDEF
cef	-	-	+	-	+	+	-	+	3	D = BCF = AEF = ABCDE
acf	+	-	+	-	-	+	+	-	2	E = ABC = ADF = BCDEF
bc	-	+	+	-	-	-	+	+	4	F = BCD = ADE = ABCEF
abce	+	+	+	-	+	-	-	-	1	AB = CE = ACDF = BDEF
df	-	-	-	+	-	+	+	+	4	AC = BE = ABDF = CDEF
adef	+	-	-	+	+	+	-	-	1	AD = EF = BCDE = ABCF
bde	-	+	-	+	+	-	-	+	3	AF = DE = BCEF = ABCD
abd	+	+	-	+	-	-	+	-	2	BD = CF = ACDE = ABEF
cde	-	-	+	+	+	-	+	-	2	BF = CD = ACEF = ABDE
acd	+	-	+	+	-	-	-	+	3	AE = BC = DF = ABCDEF
bcdf	-	+	+	+	-	+	-	-	1	ABD = BEF = ACF = CDE
abcdef	+	+	+	+	+	+	+	+	4	ABF = BDE = ACD = CEF

2015F2

2017F1 [Example 8.2 The Tool Life Data]
A company is comparing two different methods of processing their raw material. They begin with 8 batches of material, which are randomly assigned to one of the two processes. The quality of the final product is measured on a 50-point scale. There is some concern that the outside temperature on each day might have an effect on the product (and the effect might be different for the two processes), and so it is recorded so that it can be taken into account. The following table shows and ordered pair for each batch, consisting of the quality measurement and the temperature.
Process 1 (45,81)(40,68)(41,77)(41,61)
Process 2 (42,59)(37,62)(41,83)(35,70)

a) Write an appropriate model for the situation described above. Hint: there are 8 observations, and you will need to use at least one indicator variable.

Process 1: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; Process 2: $y_i = \beta_0 + \gamma_0 + (\beta_1 + \gamma_1) x_i + \varepsilon_i$; Let $w_i = \begin{cases} 0 & 1 \le i \le 4 \\ 1 & 5 \le i \le 8 \end{cases}$, overall $y_i = \beta_0 + \beta_1 x_i + w_i \gamma_0 + w_i \gamma_1 x_i + \varepsilon_i$

b) Write the matrix form of the appropriate model. Show the contents and dimensions of all matrices.

$$\begin{bmatrix} 45 \\ 40 \\ 41 \\ 41 \\ 42 \\ 37 \\ 41 \\ 35 \end{bmatrix}_{8\times 1} = \begin{bmatrix} 1 & 81 & 0 & 0 \\ 1 & 68 & 0 & 0 \\ 1 & 77 & 0 & 0 \\ 1 & 61 & 0 & 0 \\ 1 & 59 & 1 & 59 \\ 1 & 62 & 1 & 62 \\ 1 & 83 & 1 & 83 \\ 1 & 70 & 1 & 70 \end{bmatrix}_{8\times 4} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{bmatrix}_{4\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}_{8\times 1}$$

c) Suppose that you wish to test for equality of the two slopes. Write the matrix form of the reduced model. What will be the numerator and denominator degrees of freedom for the additional sum of squares F test?

To test the hypothesis that the two regression lines are identical ($H_0: \gamma_0 = \gamma_1 = 0$),
To test the hypothesis that the two lines have different intercepts and a common slope ($H_0: \gamma_0 = 0$),
$H_0: \gamma_1 = 0$, $y_i = \beta_0 + \beta_1 x_i + w_i \gamma_0 + \varepsilon_i$

$$\begin{bmatrix} 45 \\ 40 \\ 41 \\ 41 \\ 42 \\ 37 \\ 41 \\ 35 \end{bmatrix}_{8\times 1} = \begin{bmatrix} 1 & 81 & 0 \\ 1 & 68 & 0 \\ 1 & 77 & 0 \\ 1 & 61 & 0 \\ 1 & 59 & 1 \\ 1 & 62 & 1 \\ 1 & 83 & 1 \\ 1 & 70 & 1 \end{bmatrix}_{8\times 3} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \end{bmatrix}_{3\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}_{8\times 1}$$

$dfE_{Full} = n - (k+1) = 8 - (3+1) = 4$, $dfE_{Reduced} = n - (k+1) + r = 5$
$F = \frac{(SSE_{Reduced} - SSE_{Full})/r}{SSE_{Full}/dfE_{Full}}$, $df_{nume} = 1$, $df_{deno} = 4$

2015F3

2016S3 2017F2

a) Explain the difference between fixed and random effects in an experimental design. Give an example to illustrate your explanation.

b) Explain the difference between crossed and nested effects in an experimental design. Give an example to illustrate your explanation. Make sure to discuss which terms would be absent from the model, and the resulting effect on sums of squares and degrees of freedom in the ANOVA table.

2015F4

2019S1
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, $i = 1, .., n$ with the additional restrictions that $\beta_1 = 0$, $\beta_2 = 2\beta_0$. Find the least-squares estimators of β_0 and β_1.
Let $SSE = \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2 = \sum_{i=1}^n (y_i - \beta_0 - 2\beta_0 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_0} = 2\sum_{i=1}^n (y_i - \beta_0 - 2\beta_0 x_i^2)(-1 - 2x_i^2) \overset{set}{=} 0$;
$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i (1+2x_i^2)}{\sum_{i=1}^n (1+2x_i^2)^2}$; $\hat{\beta}_2 = \frac{2\sum_{i=1}^n y_i (1+2x_i^2)}{\sum_{i=1}^n (1+2x_i^2)^2}$

2016S

Fountain, Tableman*

2016S1

[2015S1][] 2017SD2 [7.6 Confouding the 2k Factorial Design in Four Blocks]
You must design an experiment to test six factors, each having two levels. Your budget will only allow sixteen runs. Furthermore, due to time constraints, only four runs can be done on a given day, so you will have to conduct the experiment in 4 blocks. You may assume that 4-way and higher interactions are not important. Show all of the following:

- all of your generators (make sure that your resolution is at least III)

2_{IV}^{6-2} ABC=D, ABE=F; I=ABCD=ABEF=CDEF

- the 16 runs to conduct
- the alias structure
- the four runs to be done on each day
- the effects to be confounded with blocks
- the Source and DF columns of the ANOVA table

Run	A	B	C	E	D	F	ACE	ACF			
(1)	-	-	-	-	-	-	-	-	1	I=ABCD=ABEF=CDEF	Day1: (1), abcd, abef, cdef
adf	+	-	-	-	+	+	+	-	2	A = BCD = BEF	Day2: adf, ace, bde, bcf
bdf	-	+	-	-	+	+	-	+	3	B = ACD = AEF	Day3: bdf, acf, bce, ade
ab	+	+	-	-	-	-	+	+	4	C = ABD = DEF	Day4: ab, cd, ef, abcdef
cd	-	-	+	-	+	-	+	+	4	D = ABC = CEF	
acf	+	-	+	-	-	+	-	+	3	E = ABF = CDF	
bcf	-	+	+	-	-	+	+	-	2	F = ABE = CDE	
abcd	+	+	+	-	+	-	-	-	1	AB = CD = EF	Confounded
ef	-	-	-	+	-	+	+	+	4	AC = BD	
ade	+	-	-	+	+	-	-	+	3	AD = BC	A-F, AC,AD,AE,AF,CE,CF=1
bde	-	+	-	+	+	-	+	-	2	AE = BF	Blk=3
abef	+	+	-	+	-	+	-	-	1	AF = BE	
cdef	-	-	+	+	+	+	-	-	1	CE = DF	
ace	+	-	+	+	-	-	+	-	2	CF = DE	
bce	-	+	+	+	-	-	-	+	3	ACE = BDE = BCF = ADF	Confounded
abcdef	+	+	+	+	+	+	+	+	4	ACF = BDF = BCE = ADE	Confounded

2016S2

2017F3 2018S3
The multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \varepsilon_i$ was fit to a data set of 75 observations. The regression SS's (SSR) were partitioned sequentially into the following:
$SSR(X_1) = 108$;
$SSR(X_2|X_1) = 163$;
$SSR(X_3|X_1X_2) = 29$;
$SSR(X_4|X_1X_2X_3) = 41$;
$SSR(X_5|X_1X_2X_3X_4) = 26$
The model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 X_{i3} + \beta_5 X_{i5} + \varepsilon_i$ was also fit to the same data and the following ANOVA was calculated:
Source(df) SS_F $SS_{-3,-4,-5}$ $SS_{1,2}$ $SS_{-2,-4}$ $SS_{1,3,5}$
Regression 367(5) -96(3) 271(2) -153(2) 214(3)
Residual Error 336(69) +96(3) 432(72) +153(2) 489(71)
Total 703(74)
The additional(extral) sum of squares F test (partial F test), $SSE_{reduced} - SSE_{Full}$ is called the extra sum of squares due to j^{th} predictor given that all the other terms are in the model.
$SSR_{Full} - SSR_{Red} = SSE_{Reduced} - SSE_{Full}$
$F = \frac{(SSE_{Red} - SSE_{Full})/(dfE_{Red} - dfE_{Full})}{SSE_{Full}/dfE_{Full}}$
Answer the following from the above information:

(a) Calculate the F-statistic for testing the hypothesis (H_0) that X_3, X_4, and X_5 have no significant effect on the response Y.

$H_0: \beta_3 = \beta_4 = \beta_5 = 0; r = 3; SST = 703;$
$SSR_{Full} = \sum_{i=1}^{5} SSR_{X_i} = 367; SSE_{Full} = SST - SSR_{Full} = 703 - 367 = 336;$ $dfE_{Full} = n - (k+1) = 75 - (5+1) = 69;$
$SSR_{Red} = \sum_{i=1}^{2} SSR_{X_i} = 271; SSE_{Red} = SST - SSR_{Red} = 703 - 271 = 432; dfE_{Red} = n - (k+1) + r = 69 + 3 = 72$
$F = \frac{(432-336)/(72-69)}{336/69} = 6.571429; F_{p,3,69} = 6.571429; F_{0.05,3,50} = 2.79, F_{0.05,3,100} = 2.70; \therefore p < 0.05$, reject H_0 at 0.05 level of significance

(b) Calculate R^2 for the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 X_{i2} + \varepsilon_i$

$SSR = \sum_{i=1}^{2} SSR_{X_i} = 271$
$R^2 = \frac{SSR}{SST} = \frac{271}{703} = 0.3855$

(c) Describe the meaning or interpretation of the statistic R^2 calculated in part (b).

R^2 is the coefficient of determination, is the proportion of variation explained by the regressor x. Values of R^2 that are close to 1 imply that most of the variability in y is explained by the regression model

(d) Calculate the R^2_{adj} for the model in part (b).

$R^2_{adj} = 1 - \frac{SSE/dfE}{SST/dfT} = 1 - \frac{432/72}{703/74} = 0.3684211$

(e) Calculate the F-statistic for testing $H_0: \beta_2 = \beta_4 = 0$.

$SSE_{Red} = 489; r = 2; dfE_{Red} = n - (k+1) + r = 71,$
$F = \frac{(489-336)/(71-69)}{336/69} = 15.70982, F_{p,2,69} = 15.70982; F_{0.05,2,50} = 3.18, F_{0.05,2,100} = 3.09; \therefore p < 0.05$, reject H_0 at 0.05 level of significance

2016S3

2017F2

a) Explain the difference between fixed and random effects in an experimental design. Give an example to illustrate your explanation.
b) Explain the difference between crossed and nested effects in an experimental design. Give an example to illustrate your explanation. Make sure to discuss which terms would be absent from the model, and the resulting effect on sums of squares and degrees of freedom in the ANOVA table.

2016S4

2019S1

Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, i = 1, .., n$ with the additional restrictions that $\beta_1 = 1$, $\beta_2 = \beta_0$. Find the least-squares estimators of the coefficients.
Let $SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}(y_i - \beta_0 - x_i - \beta_0 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^{n}(y_i - x_i - \beta_0 - \beta_0 x_i^2)(-1 - x_i^2) \overset{set}{=} 0; \hat{\beta}_0 = \hat{\beta}_2 = \frac{\sum_{i=1}^{n}(1+x_i^2)(y_i - x_i)}{\sum_{i=1}^{n}(1+x_i^2)^2}$

2016S5

In the multiple regression model with $p-1$ independent variables X_j, let the $n \times p$ matrix $\mathbf{X}$ denote the design matrix which contains the column of 1's to fit the intercept term and has full rank. Let $\mathbf{H}$ denote the hat matrix. Let h_{ii} denote the i_{th} diagonal element of $\mathbf{H}$. Prove that $0 \leq h_{ii} \leq 1$.
$\mathbf{H}$ is an idempotent matrix and symmetric.
$\mathbf{H^2} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H}$

$$h_{ii} = h_i' h_i = \begin{bmatrix} h_{1i} & \cdots & h_{ni} \end{bmatrix} \begin{bmatrix} h_{1i} \\ \vdots \\ h_{ni} \end{bmatrix} = \sum h_{ii}^2 = h_{ii}^2 + \sum_{j \neq j} h_{ij}^2 \geq 0$$
$$h_{ii} - h_{ii}^2 = h_{ii}(1 - h_{ii}) = \sum_{j \neq j} h_{ij}^2 \geq 0 \implies h_{ij} \leq 1$$

Residual: $e = (I - H)Y$
$$Cov(\mathbf{e}) =$$
$$Var(e_i) = \sigma^2(1 - h_{ii}) \geq 0 \implies h_{ij} \leq 1$$

2016F

Jong Sung Kim*, Brad Crain

2016F1

The data for this question consist of 12 measurements on each of 2 quantitative regressor variables x_1 and x_2 and on a dependent variable y. The data are displayed below:
Obs x1 x2 y m $\bar{y}_i$ $(y_{ij} - \bar{y}_i)^2$
1 3 1 5 1 5 0
2 3 2 5 2 4 1
3 3 2 3 2 4 1
4 7 1 23 3 23 0
5 7 3 19 4 19 0
6 12 1 75 5 78 9
7 12 1 81 5 78 9
8 12 2 67 6 67 0
9 12 3 51 7 49 4
10 12 3 47 7 49 4
11 19 2 135 8 135 0
12 19 3 121 9 121 0

```
x1 <- c(3,3,3,7,7,12,12,12,12,12,19,19)
x2 <- c(1,2,2,1,3,1,1,2,3,3,2,3)
y <- c(5,5,3,23,19,75,81,67,51,47,135,121)
table <- data.frame(y,x1,x2)
model <- lm(y~x1+x2,table)
model2 <- lm(y~x2+x1,table)
vcov(model) # Covariance of Estimates
```

```
##             (Intercept)         x1          x2
## (Intercept)   54.812118 -1.5107739 -16.5305567
## x1            -1.510774  0.2483464  -0.4966928
## x2           -16.530557 -0.4966928  10.7694378
```

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = table)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0814  -4.0788   0.7909   5.9059  10.1021
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.4465     7.4035  -1.411   0.1919
## x1            8.1556     0.4983  16.365 5.28e-08 ***
## x2           -9.5612     3.2817  -2.914   0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.844 on 9 degrees of freedom
## Multiple R-squared:  0.968,  Adjusted R-squared:  0.9609
## F-statistic: 136.2 on 2 and 9 DF,  p-value: 1.869e-07
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## x1         1 20646.9 20646.9 263.9987 5.625e-08 ***
## x2         1   663.9   663.9   8.4885   0.01721 *
## Residuals  9   703.9    78.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## x2         1   364.5   364.5   4.6606   0.05918 .
## x1         1 20946.3 20946.3 267.8266 5.282e-08 ***
## Residuals  9   703.9    78.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model to be fit to the data is $Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \varepsilon_i$. What follows is partially incomplete SAS output. Although the output is incomplete, there is enough information given that you can answer the questions that follow with a minimal amount of calculation. Note that Type I SS is the same as Seq SS.
Information I: model: $y = x_1 x_2$;
Analysis of Variance
Source df SS MS
Model 2 21292.2369 10646.1185
Error 9 703.87576 78.2084
Total 11 21996
Root MSE 8.84355 R-Square 0.9680
Dependent Mean 52.66667 Adj R-Sq 0.9609
Coeff Var 16.79156
Parameter Estimates
Variable DF ParameterEstimate Type I SS
Intercept 1 -10.44655 33285
x1 1 8.15560 20647
x2 1 -9.56119 663.87250
Information II: model:$y = x_2 x_1$;
Analysis of Variance
Source SumofSquares
Model
Error 703.87576
Total
Root MSE 8.84355 R-Square 0.9680
Dependent Mean 52.66667 Adj R-Sq 0.9609
Coeff Var 16.79156
Parameter Estimates
Variable DF ParameterEstimate Type I SS
Intercept 1 -10.44655 33285
x2 1 -9.56119 364.50000
x1 1 8.15560 20946

(a) Do a hypothesis test, at the .01 level of significance, of $H_0: \beta_1 = \beta_2 = 0$ vs. $H1$: At least one of β_1 or $\beta_2 \neq 0$.

$dfR = 2, dfT = 12 - 1, dfE = 11 - 9 = 2;$
$SSE = 703.87576$, $MSE = \frac{703.87576}{9} = 8.84355^2 = 78.2084$; $SSR = 364.5 + 20946 = 20647 + 663.8725$; $MSR = 21310.87/2 = 10655.43$
$F = \frac{MSR}{MSE} = \frac{10655.43}{78.2084} = \frac{(21996.1125 - 703.87576)/2}{78.2084} = 136.2441 > F(0.01, 2, 9) = 8.02$

(b) What is the value of R^2? $SST = 21310.87 + 703.87576 = 22014.75$ $R^2 = \frac{21310.87}{22014.75} = 0.9680,$

(c) Do the following two hypothesis tests, each at the .05 level of significance:

i. $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

$F = \frac{20946/1}{78.20842} = 267.8228 > F(0.05, 1, 9) = 5.12$

ii. $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 \neq 0$

$F = \frac{663.8725/(10-9)}{78.20842} = 8.488504 > F(0.05, 1, 9) = 5.12$

(d) Obtain a 99% confidence interval for β_1.

$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-k-1} se(\hat{\beta}_1)$, $se(\hat{\beta}_1) = \sqrt{MSE \cdot C_{22}}$; $8.1556 \pm t(0.005, 9) 8.84355 * \sqrt{0.2483463917}, 8.1556 \pm 3.249835541 \times 4.407127; (-6.166838, 22.47804)$

```
c(8.1556-qt(0.995,9)*8.84355*sqrt(0.2483463917),8.1556+qt(0.995,9)*8.84
```

```
## [1] -6.166838 22.478038
```

(e) Give an unbiased estimate of the variance of $\hat{\beta}_1 - \hat{\beta}_2$.

$Var(\hat{\beta}_1 - \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2) = MSE(C_{22} + C_{33} - 2C_{23}) = 78.20842[0.2483 + 10.7694 - 2(-0.49669)] = 939.3676$

(f) Obtain MS(Pure Error). Hint: Pure Error can be found exactly the same way as we did for simple linear regression model. That is, group according to different combinations of levels from X_1 and X_2. First compute SS(Pure Error), and then divide it by degrees of freedom.

$SS_{PE} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 28$; $df_{PE} = n - m = 12 - 9 = 3$; $MS_{PE} = 28/3$

(g) Perform a test for lack-of-fit at the .05 level of significance. Note: If you are unable to answer part (f), use MS(Pure Error) = 7.5. This is not the correct answer to (f), but if you use it in this part of the problem, you will receive full credit on this part, provided your answer is otherwise correct. [4.5]

$SSE = SS_{LOF} + SS_{PE}$, $\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m} n_i (y_{ij} - \hat{y}_i)^2$
H_0: There is no lack of fit, the model is appropriate; H_1: There is a lack of fit, the model is not appropriate;
$SS_{LOF} = SSE - SS_{PE} = 703.87576 - 28 = 675.8758$
$df_{LOF} = dfE - df_{PE} = m - (k+1) = 6$
$F = \frac{SS_{LOF}/df_{LOF}}{MS_{PE}} = \frac{675.87576/6}{28/3} = 12.06921$
$F(0.05, 6, 3) = 8.94$. Reject H_0 at the .05 level of significance.

2016F2

Data were collected on each of two quantitative regressor variables X_1 and X_2, a dichotomous categorical variable which we shall call "group", and a dependent variable Y. The data are displayed below:
Obs y x1 x2 group
1 14 3.54 17 1
2 23 4.87 15 1
3 31 2.89 13 1
4 28 3.03 10 1
5 25 4.10 21 2
6 35 3.55 18 2
7 18 5.17 18 2
8 23 3.99 16 2
9 40 2.02 13 2
10 34 2.44 13 2
The model to be fit to the data is $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 Z_i + \beta_5 X_{1i} Z_i + \beta_6 X_{2i} Z_i + \beta_7 X_{2i}^2 Z_i + \varepsilon_i$, where $Z_i = 1$, if case i is in group 1, and $Z_i = 0$, otherwise.

(a) What are the first and last rows of the X-matrix (assuming that the data are entered in the same order in which they are displayed above)?

First row: $1, X_{1i}, X_{2i}, X_{2i}^2, 1, X_{1i}, X_{2i}, X_{2i}^2, 1, 3.54, 17, 289, 1, 3.54, 17, 289$
Last row: $1, X_{1i}, X_{2i}, X_{2i}^2, 0, 0, 0, 0, 1, 3.54, 17, 289, 0, 0, 0, 0$

(b) For each of the following objectives, give the appropriate null hypothesis.

i. It is desired to know whether the slope coefficient on x_1 is the same for both groups.$\beta_5 = 0$
ii. It is desired to know whether the entire regression models for the two groups are identical.$\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$
iii. It is desired to know whether a quadratic term in x_2 is needed by both groups.$\beta_3 = \beta_7 = 0$
iv. It is desired to know whether the slope coefficients on x_1 and x_2 for the first group are equal.$\beta_1 + \beta_5 = \beta_2 + \beta_6$

2016F3

The following is part of the SAS output from a simple linear regression model: $y_i = \beta_0 + \beta x_i + \varepsilon_i$, where $i = 1, .., 13$, and y_i and x_i are the ith punter's average punting distance and right leg strength, respectively. Each punter punted 10 times and the average distance was measured. In addition, measure of right leg strength (lb lifted) was taken via a weight lifting test.

```r
table_2016f3 <- matrix(c(170,162.50,140,144.00,180,147.50,160,163.50,170,192.00,150,171.75,170,162.00,110,104.83,120,105.67,130,117.58,120,140.
      nrow=13,ncol =2,byrow = T)
colnames(table_2016f3) <- c("rleg","distance")
model_2016f3 <- lm(distance~rleg,data.frame(table_2016f3))
summary(model_2016f3)
```

```
##
## Call:
## lm(formula = distance ~ rleg, data = data.frame(table_2016f3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.888  -9.371   2.719   8.889  23.639
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.9070    31.3704   0.475  0.64395
## rleg          0.9027     0.2101   4.296  0.00126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.58 on 11 degrees of freedom
## Multiple R-squared:  0.6266, Adjusted R-squared:  0.5926
## F-statistic: 18.46 on 1 and 11 DF,  p-value: 0.001264
```

```r
anova(model_2016f3)
```

```
## Analysis of Variance Table
##
## Response: distance
##           Df Sum Sq Mean Sq F value   Pr(>F)
## rleg       1 5076.9  5076.9  18.457 0.001264 **
## Residuals 11 3025.7   275.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
vcov(model_2016f3)
```

```
##              (Intercept)        rleg
## (Intercept)  984.103717 -6.51994181
## rleg          -6.519942  0.04414544
```

```r
model_2016f3$fitted.values
```

```
##        1        2        3        4        5        6        7        8
## 168.3611 141.2810 177.3879 159.3344 168.3611 150.3077 168.3611 114.2
##        9       10       11       12       13
## 123.2276 132.2543 123.2276 141.2810 159.3344
```

```r
model_2016f3$residuals
```

```
##          1          2          3          4          5          6
##  -5.861136   2.719012 -29.887852   4.165580  23.638864  21.442296  -
##          8          9         10         11         12         13
##  -9.370840 -17.557556 -14.674272  17.022444   8.889012   5.835580
```

Dependent Variable: distance X'X Inverse, Parameter Estimates, and SSE
Variable Intercept rleg distance
Intercept 3.5777777778 -0.023703704 14.906962963
rleg -0.023703704 0.0001604938 0.9026716049
distance 14.906962963 0.9026716049 3025.6604973
Root MSE 16.58493 R-Square 0.6266
Dependent Mean 148.22462 Adj R-Sq 0.5926
Coeff Var 11.18906
Output Statistics

```
 Dep Var  Predicted Std Error    95% CL           95% CL
```

Obs distance Value MeanPredict Mean Predict Residual
1 162.5000
2 144.0000 141.2810 4.8755 130.5501 152.0119 103.2332 179.3288 2.7190
3 147.5000 177.3879 8.1998 159.3402 195.4355 136.6668 218.1089 -29.8879
4 163.5000 159.3344 5.2769 147.7201 170.9488 121.0281 197.6408 4.1656
5 192.0000
6 171.7500 150.3077 4.6253 140.1274 160.4880 112.4115 188.2039 21.4423
7 162.0000
8 104.8300 114.2008 9.1584 94.0433 134.3583 72.5018 155.8999 -9.3708
9 105.6700 123.2276 7.4170 106.9028 139.5523 83.2403 163.2148 -17.5576
10 117.5800 132.2543 5.9141 119.2374 145.2712 93.4996 171.0089 -14.6743
11 140.2500 123.2276 7.4170 106.9028 139.5523 83.2403 163.2148 17.0224
12 150.1700 141.2810 4.8755 130.5501 152.0119 103.2332 179.3288 8.8890
13 165.1700 159.3344 5.2769 147.7201 170.9488 121.0281 197.6408 5.8356

```r
x <- c(170,140,180,160,170,150,170,110,120,130,120,140,160)
y <- c(162.50,144.00,147.50,163.50,192.00,171.75,162.00,104.83,105.67,1
bar_x <- mean(x)
sum((x-mean(x))^2)
```

```
## [1] 6230.769
```

```r
sum(x^2)-sum(x)^2/13
```

```
## [1] 6230.769
```

```r
sum(x^2)-mean(x)^2*13
```

```
## [1] 6230.769
```

```r
sd(x)^2
```

```
## [1] 519.2308
```

```r
S_xx <- var(x)*(13-1)
hat_y <- 14.90696+0.90267*170
qt(0.025,11,lower.tail = F)
```

```
## [1] 2.200985
```

```r
se_y_mean <- sqrt(275.06005*(1/13+(170-bar_x)^2/S_xx))
se_y_new <- sqrt(275.06005*(1+1/13+(170-bar_x)^2/S_xx))
```

(a) Find the three residual values at x = 170.

$\hat{y} = 14.90696 + 0.90267 * 170 = 168.3609$;$162.5 - 168.3609 = -5.8609$,$192 - 168.3609 = 23.6391$,$162 - 168.3609 = -6.3609$

(b) Compute a 95% confidence interval for the mean response at x = 170. Hint:Compute the variance of the estimate of the mean response at x = 170.

$\bar{x} = 147.6923$,$S_{XX} = \sum_{i=1}^{13} (x_i - \bar{x})^2 = 6230.769$
$se(y_0) = \sqrt{MSE(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}})} = \sqrt{275.06005(\frac{1}{13} + \frac{(170 - 147.6923)^2}{6230.769})} = 6.567093$
$\hat{y} \pm t_{n-2, 0.025} se(y_0) = 168.3609 \pm 2.200985 * 6.567093$, $(153.9068, 182.815)$

(c) Compute a 95% prediction interval on a new response observation at x = 170. Hint: You can use part of the expression in (b).

$se(y_0) = \sqrt{MSE(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}})} = 17.83779$
$168.3609 \pm 2.200985 * 17.83779$, $(129.1002, 207.6216)$

2016F4

Prior to 1985, Meily Lin had observed that some colors of birthday balloons seem to be harder to inflate than others. She ran this experiment to determine whether balloons of different colors are similar in terms of the time taken for inflation to a diameter of 7 inches. Four colors were selected from a single manufacturer. An assistant blew up the balloons and the experimenter recorded the times (to the nearest 1/10 second) with a stop watch. This experiment was replicated 4 times, and the data including the order are displayed in the SAS code below, where color 1 = pink, 2 = yellow, 3 = orange, and 4 = blue.

```
baloon <- matrix(c(1,1,22.4,2,3,24.6,3,1,20.3,4,4,19.8,5,3,24.3,6,2,22.2,7,2,28.5,8,2,25.7,9,3,20.2,10,1,19.6,11,2,28.8,12,4,24.0,13,4,17.1,14,
                 nrow = 32,ncol = 3,byrow = T)
colnames(baloon) <- c("runorder", "color", "inftime")
cbind(baloon[1:8,],baloon[9:16,],baloon[17:24,],baloon[25:32,])
```

a. Why or why not do we need to record the run order in the model?

Blocking can systematically eliminate the nuisance's effect on the statistical comparisons among treatments. The four times as blocks can test the robustness of the process variable to conditions which not easily control.

b. What kind of model would be appropriate for the above experiment?

BIBD, Balanced Incomplete Block Design. Every treatment is not present in every block twice. Incomplete means each block doesn't contain two pairs of treatments. Each block contain a unique combination of treatments. Any pairs of treatments occur together the same number of times as any other pair.
a=4 treatments occurs r=8 times in the design. b=4 block contains k=8 treatment.

c. Read the following . If we had assumed that there is an equal slope linear relationship between the inflation time and the run order for each color, how can we test the assumption? How would you adjust the following program? Why?

```
<Program I>
proc glm data=balloon;
class color; /* color: 1 = pink, 2 = yellow, 3 = orange, 4 = blue */
model inftime = color runorder;
estimate 'pink vs. orange' color -1 0 1 0;
lsmeans color/pdiff;
run;
```

d. Based on the following output from , one can apply a Bonferroni multiple comparison test with level .05. Which are significantly different and which are not?

The GLM Procedure; Least Squares Means
color inftimeLSMEAN LSMEANNumber
1 18.3341795 1
2 22.3883782 2
3 22.0882820 3
4 18.2141603 4
Least Squares Means for effect color Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: inftime
i/j 1 2 3 4
1 0.0043 0.0076 0.9271
2 0.0043 0.8195 0.0034
3 0.0076 0.8195 0.0060
4 0.9271 0.0034 0.0060

2016F5

[BIBD]
Consider an experiment to compare 7 treatments in block of size 5. Taking all possible combinations of five treatments from seven gives a balanced incomplete block design with each treatment level occurring 15 times. Hint: Figure out p, t, k, r, λ and their relationships.
$a = 7, k = 5, r = 15$, $ar = bk$, replications of each pair $\lambda = \frac{(k-1)}{a-1} r = \frac{k(k-1)}{a(a-1)} b = 10$

a. How many blocks does the design have?

$b = \frac{ar}{k} = 21$

b. Show that the number of times each treatment level occurs must be a multiple of five for a balanced incomplete block design with 7 treatments and blocks of size 5 to exist.

$r = \frac{bk}{a} = \frac{5}{7} b$

c. Show that the smallest balanced incomplete block design has 15 observations per treatment.

$\lambda = \frac{(k-1)}{a-1} r = \frac{2}{3} r \in \mathbf{N}^+$
r is a multiple of 3 and 5 (in b.), $r = 15$ is the smallest number of observations per treatment for a BIBD with $a = 7, k = 5$.

2017S

Brad Crain, Jong Sung Kim*

2017SR1

2018S1
A company is comparing two different methods of processing their raw material. They begin with 8 batches of material, which are randomly assigned to one of the two processes. The quality of the finalproduct is measured on a 50-point scale. There is some concern that the outside temperature on each day might have an effect on the product (and the effect might be different for the two processes), and so it is recorded so that it can be taken into account. The following table shows and orderedpair for each batch, consisting of the quality measurement and the temperature.
Process 1 (45,81)(40,68) (41,77)(41,61)
Process 2 (42,59)(37,62) (41,83)(35,70)

a) Write an appropriate model for the situation described above. Hint: there are 8 observations, and you will need to use at least one indicator variable.
b) Write the matrix form of the appropriate model. Show the contents and dimensions of all matrices.
c) Suppose that you wish to test for equality of the two slopes. Write the matrix form of the reduced model. What will be the numerator and denominator degrees of freedom for the additional sum of squares F test?

2017SR2

2019S1
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, $i = 1, .., n$ with the additional restrictions that $\beta_1 = 1$ and $\beta_2 = \beta_0/2$. Find the least-squares estimators of β_0, β_1, β_2.
Let

$$SSE = \sum_{i=1}^{n} (y_i - \beta_0 - x_i - \frac{\beta_0}{2} x_i^2)^2$$
$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - x_i - \frac{\hat{\beta}_0}{2} x_i^2)(1 + \frac{x_i^2}{2}) \overset{set}{=} 0$$
$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} (y_i - x_i)(1 + \frac{x_i^2}{2})}{\sum_{i=1}^{n} (1 + \frac{x_i^2}{2})^2}, \hat{\beta}_1 = 1, \hat{\beta}_2 = \frac{\sum_{i=1}^{n} (y_i - x_i)(1 + \frac{x_i^2}{2})}{2 \sum_{i=1}^{n} (1 + \frac{x_i^2}{2})^2}$$

2017SD1

[Latin Square]
Given a educational material evaluation experiment where there are three possible blocking factors [R,C,G], each with six levels $[R_{1..6}; C_{1..6}; G_{1..6}]$:

1. Write out the model equation of the Latin Square design if the blocking factors R and C are used, and G is disregarded.

$y_{ijk} = \mu + \tau_i + \alpha_j + \beta_k + \varepsilon_{ijk}$, $i, j, k = 1, .., 6$; where μ overall mean
τ_i is effect of i^{th} treatment; α_j is effect of j^{th} block of factor R; β_k effect of k^{th} block of factor C;
ε_{ijkl} is random error when i^{th} treatment is applied at j^{th} block of factor R and k^{th} block of factor C; y_{ijkl} is response ;
Assumptions: $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$. Further assumptions would be based on whether the treatment and blocking factors are random or fixed.

2. Explain why all three blocking factors can not be used simultaneously without a modification

The Latin-Squre design can only use 2 blocking factors, as we distribute the levels of the treatment factor on a table with rows of one blocking factor (each row is fore one block) and columns of the other blocking factor (each column is one block)

3. What is the modification required?

You can test three blocking factors by turning the Latin-square design into a Graeco-Latin square design, which allows to add a Greek letter to each entry in the table, where each Greek letter stands for a block of the factor G.

4. If the Relative Efficiency for the modified experiment was calculated to be 2.3, how many observations of heterogeneous experimental units in a CRD would be expected to obtain the same variance for the treatment mean as one replicate of the modified experiment.

$\frac{(df_{E(LS)}+1)(df_{E(CRD)}+3)MS_{CRD}}{(df_{E(LS)}+3)(df_{E(CRD)}+1)MS_{LS}} = 2.3$
$df_{E(LS)} = (p-1)(p-2) = 20$,$(df_{E(GS)} = (p-1)(p-3) = 15$, $df_{E(CRD)} = a(n-1) = 7.2, 32.5$

2017SD2

[8.3 The One-Quarter Fraction of the 2k Design p.344] [7.7 table 7.9]
Given a Blocked 2^{6-2} design with Factors [A,B,C,D,E,F],Generators E=ABC, F=BCD and Defining Contrasts AB, CD

1. How many blocks are included in this design?

4

2. What is the Defining Relationship in this design?

generating relations I=ABCE=BCDF=ADEF

3. What is the Resolution of this Design?

IV

4. List the aliases of AE

AE=BC=ABCDEF=DF

5. Show the effect on two-way interactions that include A, if you augment by **folding** on A [8.5.2]

I=-ABCE=BCDF=-ADEF
AB=-CE=ACDF=-BDEF
AC=-BE=ABDF=-CDEF
AD=-BCDE=ABCF=-EF
AE=-BC=ABCDEF=-DF
AF=-BCEF=ABCD=-DE

6. List the aliases of the defining contrasts [including the generalized interaction]?

AB=CE=ACDF=BDEF
CD=ABDE=BF=ACEF

2017F

Robert Fountain*, Daniel Taylor-Rodriguez

2017F1

2018S1 2019S3
A company has developed two specialized training workshops for their employees. Each of the 12 employees is randomly assigned to one of the two workshops. The company would like to develop a model that could be used to predict each employee's performance score (a number from 0 to 100) based on their attendance at the workshops. The previous year's performance score is also available for use as a predictor. The following table shows and ordered pair for each employee, consisting of the current performance score and the previous year's score.
WorkshopA (70,58)(70,62) (68,60)(72,65) (72,66)(72,62)
WorkshopB (75,60)(74,62) (72,60)(71,60) (73,61)(73,65)

a) Write an appropriate model for the situation described above, allowing for different slopes and different intercepts for the two workshops. Hint: there are 12 observations, and you will need to use at least one indicator variable.
b) Write the matrix form of the appropriate model. Show the contents and dimensions of all matrices.
c) Suppose that you wish to test for equality of the two slopes. Write the matrix form of the reduced model. What will be the numerator and denominator degrees of freedom for the additional sum of squares F test?

2017F2

a) Explain the difference between fixed and random effects in an experimental design. Give an example to illustrate your explanation.

https://stats.stackexchange.com/questions/4700/what-is-the-difference-between-fixed-effect-random-effect-and-mixed-effect-mode

- Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts ai and fixed slope b corresponds to parallel lines for different individuals i, or the model yit=ai+bt. Kreft and De Leeuw (1998) thus distinguish between fixed and random coefficients.
- Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. Searle, Casella, and McCulloch (1992, Section 1.4) explore this distinction in depth.
- "When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random." (Green and Tukey, 1960)
- "If an effect is assumed to be a realized value of a random variable, it is called a random effect." (LaMotte, 1983)
- Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage ("linear unbiased prediction" in the terminology of Robinson, 1991). This definition is standard in the multilevel modeling literature (see, for example, Snijders and Bosker, 1999, Section 4.2) and in econometrics.

b) Explain the difference between crossed and nested effects in an experimental design. Give an example to illustrate your explanation. Make sure to discuss which terms would be absent from the model, and the resulting effect on sums of squares and degrees of freedom in the ANOVA table.

https://www.theanalysisfactor.com/the-difference-between-crossed-and-nested-factors/
Two factors are crossed when every category of one factor co-occurs in the design with every category of the other factor. In other words, there is at least one observation in every combination of categories for the two factors.
A factor is nested within another factor when each category of the first factor co-occurs with only one category of the other. In other words, an observation has to be within one category of Factor 2 in order to have a specific category of Factor 1. All combinations of categories are not represented.
If two factors are crossed, you can calculate an interaction. If they are nested, you cannot because you do not have every combination of one factor along with every combination of the other.

2017F3

2018S3 2016S2
The multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \varepsilon_i$ was fit to a data set of 75 observations. The regression SS's (SSR) were partitioned sequentially into the following:
$SSR(X_1) = 108$
$SSR(X_2|X_1) = 163$
$SSR(X_3|X_1X_2) = 29$
$SSR(X_4|X_1X_2X_3) = 41$
$SSR(X_5|X_1X_2X_3X_4) = 26$
The model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 X_{i3} + \beta_5 X_{i5} + \varepsilon_i$ was also fit to the same data and the following ANOVA was calculated:
Source SS
Regression 214
Residual Error 489
Total 703
Answer the following from the above information:

(a) Calculate the F-statistic for testing the hypothesis (H_0) that X_3, X_4, and X_5 have no significant effect on the response Y.
(b) Calculate R^2 for the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 X_{i2}\varepsilon_i$
(c) Calculate the R^2_{adj} for the model in part (b).
(d) Calculate the F-statistic for testing $H_0 : \beta_2 = \beta_4 = 0$.

2017F4

2019S1
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, i = 1,..,n$ with the additional restrictions that $\beta_0 = 1$, $\beta_1 - \beta_2 = 0$. Find the least-squares estimators of the regression coefficients.
Let $SSE = \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (y_i - 1 - \beta_1 x_i - \beta_1 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^n (y_i - 1 - \beta_1 x_i - \beta_1 x_i^2)(-x_i - x_i^2) \overset{set}{=} 0; \hat{\beta}_1 = \hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i + x_i^2)(y_i - 1)}{\sum_{i=1}^n (x_i + x_i^2)^2}$

2018S

Robert Fountain*, Daniel Taylor-Rodriguez

2018S1

2019S3
A company has developed three specialized training workshops for their employees. Each of the 12 employees is randomly assigned to one of the three workshops. The company would like to develop a model that could be used to predict each employee's performance score (a number from 0 to 100) based on their attendance at the workshops. The previous year's performance score is also available for use as a predictor. The following table shows and ordered pair for each employee, consisting of the current performance score and the previous year's score.
WorkshopA (70,58)(70,62) (68,60)(72,65)
WorkshopB (75,60)(74,62) (72,60)(71,60)
WorkshopC (72,66)(72,62) (73,61)(73,65)

a) Write an appropriate model for the situation described above, allowing for different slopes and different intercepts for the three workshops. Hint: there are 12 observations, and you will need to use at least one indicator variable.

Let $w_{1i} = \begin{cases} 0 & 1 \le i \le 4 \\ 1 & 5 \le i \le 8 \\ 0 & 9 \le i \le 12 \end{cases}$, $w_{2i} = \begin{cases} 0 & 1 \le i \le 4 \\ 0 & 5 \le i \le 8 \\ 1 & 9 \le i \le 12 \end{cases}$
overall $y_i = \beta_0 + \beta_1 x_i + w_{1i}(\gamma_0 + \gamma_1 x_i) + w_{2i}(\delta_0 + \delta_1 x_i) + \varepsilon_i$
WorkshopA: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, 1 \le i \le 4;$
WorkshopB: $y_i = \beta_0 + \gamma_0 + (\beta_1 + \gamma_1)x_i + \varepsilon_i, 5 \le i \le 8;$
WorkshopC: $y_i = \beta_0 + \delta_0 + (\beta_1 + \delta_1)x_i + \varepsilon_i, 9 \le i \le 12;$

b) Write the matrix form of the appropriate model. Show the contents and dimensions of all matrices.

$$\begin{bmatrix} 70\\70\\68\\72\\75\\74\\72\\71\\72\\72\\73\\73 \end{bmatrix}_{12\times1} = \begin{bmatrix} 1&58&0&0&0&0\\1&62&0&0&0&0\\1&60&0&0&0&0\\1&65&0&0&0&0\\1&60&1&60&0&0\\1&62&1&62&0&0\\1&60&1&60&0&0\\1&60&1&60&0&0\\1&66&0&0&1&66\\1&62&0&0&1&62\\1&61&0&0&1&61\\1&65&0&0&1&65 \end{bmatrix}_{12\times6} \times \begin{bmatrix} \beta_0\\\beta_1\\\gamma_0\\\gamma_1\\\delta_0\\\delta_1 \end{bmatrix}_{6\times1} + \begin{bmatrix} \varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\\\varepsilon_8\\\varepsilon_9\\\varepsilon_{10}\\\varepsilon_{11}\\\varepsilon_{12} \end{bmatrix}_{12\times1}$$

c) Suppose that you wish to test for equality of the three slopes. Write the matrix form of the reduced model. What will be the numerator and denominator degrees of freedom for the additional sum of squares F test?

$H_0 : \gamma_1 = \delta_1 = 0, y_i = \beta_0 + \beta_1 x_i + w_{1i}\gamma_0 + w_{2i}\delta_0 + \varepsilon_i$

$$\begin{bmatrix} 70\\70\\68\\72\\75\\74\\72\\71\\72\\72\\73\\73 \end{bmatrix}_{12\times1} = \begin{bmatrix} 1&58&0&0\\1&62&0&0\\1&60&0&0\\1&65&0&0\\1&60&1&0\\1&62&1&0\\1&60&1&0\\1&60&1&0\\1&66&0&1\\1&62&0&1\\1&61&0&1\\1&65&0&1 \end{bmatrix}_{12\times4} \times \begin{bmatrix} \beta_0\\\beta_1\\\gamma_0\\\delta_0 \end{bmatrix}_{4\times1} + \begin{bmatrix} \varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\\\varepsilon_8\\\varepsilon_9\\\varepsilon_{10}\\\varepsilon_{11}\\\varepsilon_{12} \end{bmatrix}_{12\times1}$$

$$\begin{bmatrix} 0&0&0&1&0&-1\\0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} \beta_0\\\beta_1\\\gamma_0\\\gamma_1\\\delta_0\\\delta_1 \end{bmatrix} = \begin{bmatrix} 0\\0 \end{bmatrix}, Rank(T) = 2$$

$dfE_{Full} = n - (k+1) = 12 - (5+1) = 6, dfE_{Reduced} = n - (k+1) + r = 8$
$F = \frac{(SSE_{Reduced} - SSE_{Full})/(dfE_{Reduced} - dfE_{Full})}{SSE_{Full}/dfE_{Full}}, df_{nume} = 10 - 8 = 2, df_{deno} = 6$

2018S2

[2015S3][] 2019S2
A company that produces textiles is trying to determine if the final quality is determined by production site (A), machine operator (B), and thread type (C). The company operates 3 production sites, and all 3 will participate in the experiment. At each of the 3 sites, 5 machine operators will be randomly selected. The company uses two different types of thread. Each of the operators will produce 3 samples of cloth using each type of thread, yielding a total of $3 \times 5 \times 2 \times 3 = 90$ observations.
Source SS df MS F pval<0.05
A 17 2 8.5 3.70 *
B 25 4 6.2 2.70 *
C 4 1 4.0 1.74
AB 32 8 4.0 1.74
AC 5 2 2.5 1.09
BC 12 4 3.0 1.30
ABC 12 8 1.5 0.65
Error 138 60 2.3
Total 245 89

a) State which effects are fixed at which effects are random.
b) State which effects are nested within others and which effects are crossed.
c) Create an abbreviated ANOVA table that has two columns: one column that lists the effects in the model (including the appropriate main effects, interactions, and nested effects), and a second column that gives the degrees of freedom for each item in the first column.

R(F)F
Site (A): τ_i a=3, Fixed;
Operator (B): $\beta_{j(i)}$ Nested in A, b=5, Random;
Thread Type (C): γ_k Crossed with B, c=2, Fixed;
Replications: n=3, Random
Model: $y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{kj(i)} + \varepsilon_{(ijk)l}, i_{1:3}, j_{1:5}, k_{1:2}, l_{1:3}$
$\sum_{i=1}^a \tau_i = 0, \sum_{k=1}^c \gamma_k = 0, \sum^a (\tau\gamma)_{ik} = 0, \sum^c (\tau\gamma)_{ik} = 0, \varepsilon_{(ijk)l} \sim N(0, \sigma^2)$
$\beta_{j(i)} \sim N(0, \sigma^2_\beta), (\beta\gamma)_{kj(i)} \sim N(0, \frac{c-1}{c}\sigma^2_{\beta\gamma}), \sum^c (\beta\gamma)_{kj(i)} = 0$
Source SS df MS F
A 17 2 8.5 8.5/4.75

B(A) 57 12 4.75 5.8/2.3
C 4 1 4.0 4.0/2.0
AC 5 2 2.5 2.5/2.0
CB(A) 24 12 2.0 2.0/2.3
Error 138 60 2.3

term	i(f)	j(r)	k(f)	l(r)	df	EMS	F
τ_if	0	b	c	n	a-1	$\sigma^2 + cn\sigma^2_\beta + \frac{bcn}{a-1}\sum_{i=1}^a \tau_i^2$	$\frac{A}{B(A)}$
$\beta_{j(i)}$r	1	1	c	n	a(b-1)	$\sigma^2 + cn\sigma^2_\beta$	$\frac{B(A)}{E}$
$(\gamma)_k$f	a	b	0	n	c-1	$\sigma^2 + n\sigma^2_{\gamma\beta} + \frac{abn}{c-1}\sum_{k=1}^c \gamma_k^2$	$\frac{C}{CB(A)}$
$(\tau\gamma)_{ik}$f	0	b	0	n	(a-1)(c-1)	$\sigma^2 + n\sigma^2_{\gamma\beta} + \frac{bn}{(a-1)(c-1)}\sum^a\sum^c(\tau\gamma)^2_{ik}$	$\frac{AC}{CB(A)}$
$(\gamma\beta)_{kj(i)}$r	1	1	0	n	a(b-1)(c-1)	$\sigma^2 + n\sigma^2_{\gamma\beta}$	$\frac{CB(A)}{E}$
$\varepsilon_{(ijk)l}$	1	1	1	1	abc(n-1)	σ^2	
Total					abcn-1		

2018S3

2017F3 2016S2
The multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \varepsilon_i$ was fit to a data set of 75 observations. The regression SS's (SSR) were partitioned sequentially into the following:
$SSR(X_1) = 108$
$SSR(X_2|X_1) = 163$
$SSR(X_3|X_1X_2) = 29$
$SSR(X_4|X_1X_2X_3) = 41$
$SSR(X_5|X_1X_2X_3X_4) = 26$
The model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 X_{i3} + \beta_5 X_{i5} + \varepsilon_i$ was also fit to the same data and the following ANOVA was calculated:
Source SS
Regression 214
Residual Error 489
Total 703
Answer the following from the above information:

(a) Calculate the F-statistic for testing the hypothesis (H_0) that X_3, X_4, and X_5 have no significant effect on the response Y.

(b) Calculate R^2 for the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 X_{i2}\varepsilon_i$

(c) Calculate the R^2_{adj} for the model in part (b).

(d) Calculate the F-statistic for testing $H_0 : \beta_2 = \beta_4 = 0$.

2018S4

2019S1
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, $i = 1, .., n$ with the additional restrictions that $\beta_1 = 0$, $\beta_0 = 2\beta_2$. Find the least-squares estimators of β_0, β_1, and β_2.
Let $SSE = \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2 = \sum_{i=1}^n (y_i - 2\beta_2 - \beta_2 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^n (y_i - 2\beta_2 - \beta_2 x_i^2)(-2 - x_i^2) \overset{set}{=} 0$; $\hat\beta_2 = \frac{\sum_{i=1}^n (2+x_i^2) y_i}{\sum_{i=1}^n (2+x_i^2)^2}$ $\hat\beta_0 = \frac{2\sum_{i=1}^n (2+x_i^2) y_i}{\sum_{i=1}^n (2+x_i^2)^2}$

2018F

Robert Fountain*, Daniel Taylor-Rodriguez

2018F1

The weights (y_i, kilograms) and corresponding heights (x_i, centimeters) of 10 randomlysampled adolescents (i= 1,..,10) are recorded, and the following summary statistics are computed:
$\sum_{i=1}^{10}(x_i - \bar{x})^2 = 472$,$\sum_{i=1}^{10}(y_i - \bar{y})^2 = 731$, $\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 274$
You will perform a simple linear regression of weight on height, under the usual assumption of independent, identically distributed, normal errors.

a) Compute the least squares estimates for the intercept and slope parameters.

$\hat\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{274}{472} = 0.5805085$;
$\hat\beta_0 = \bar{y} - \hat\beta_1\bar{x} = \bar{y} - 0.5805085\bar{x}$

b) Compute the usual unbiased estimate of the error variance.

$\hat\sigma^2 = \frac{SSE}{n-2} = \frac{1}{8}(S_{yy} - \frac{S^2_{xy}}{S_{xx}}) = \frac{1}{8}(731 - \frac{274^2}{472}) = 71.49258$

c) Compute unbiased estimates of the variances of the least squares estimates in part (a).

$Var(\hat\beta_1) = \frac{\hat\sigma^2}{S_{xx}} = \frac{71.49258}{472} = 0.1514673$
$Var(\hat\beta_1) = \hat\sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}) = 71.49258(\frac{1}{10} + \frac{\bar{x}^2}{472})$

d) Perform a two-sided test for whether or not height and weight are related (assuming the simple linear regression model holds). State the null and alternative hypotheses, and use $\alpha = 0.05$.

$H_0 : \hat\beta_1 = 0; H_1 : \hat\beta_1 \neq 0$
$t_0 = \frac{\hat\beta_1 - 0}{\sqrt{Var(\hat\beta_1)}} = \frac{0.5805085}{\sqrt{0.1514673}} = 1.491589 < t_{\frac{0.05}{2}, n-2} = 2.31$
Fail to reject H_0 at 0.05 level of significance.

e) Compute 95% simultaneous two-sided confidence intervals for the intercept and slope parameters, using the Bonferroni method.

$\hat\beta_1 \pm t_{\frac{0.05}{2p}, n-2} se(\hat\beta_1) = 0.5805085 \pm 2.75\sqrt{0.1514673}$, $(-0.4905, 1.6515)$

2018F2

[565-HW1]
City planners are evaluating the effectiveness of a new "intelligent" traffic control system in reducing the amount of time motorists must spend on city streets. A total of 24 simulations are run: 4 simulations for each of the 6 combinations of control system (old or new) and traffic intensity (light, moderate, or heavy). All simulations use different random seeds, the combinations are run in a completely random order, and the median travel time (minutes) is recorded for each simulation. For each combination, the following table gives the average and sample standard deviation of the median travel times from the 4 simulations assigned that combination:

```
                    Old System            New System
```

Sample light Moderate Heavy light Moderate Heavy
Mean 13 14 15 5 8 17
Standard Deviation 1 2.5 3.5 2.5 2 3.5

a) Write a (univariate) linear model equation of the usual full form for data from this experiment, with median travel time as the response. Explain each term and specify any conditions it satisfies. What crucial assumption are you making about the error variances?

$y_{ijkl} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$, $i = 1, 2$; $j = 1, 2, 3$; $k = 1, 2, 3, 4$; $l = 1, 2$, $[a = 2, b = 3, n = 4]$ where μ overall mean
τ_i is fixed main effect of i^{th} level of Factor A; β_j is fixed main effect of j^{th} level of Factor B;
$(\tau\beta)_{ij}$ is fixed interaction effect of i^{th} level of Factor A and j^{th} level of Factor B;
ε_{ijkl} is random error for the k^{th} replicate EU when i^{th} level of Factor A and j^{th} level of Factor B are applied; y_{ijkl} is response for the;
Assumptions: $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$ (constant variance, zero mean, independent); $\sum_i^2 \tau_i = 0; \sum_j^3 \beta_j = 0; \sum_i^2 (\tau\beta)_{ij} = 0; \sum_j^3 (\tau\beta)_{ij} = 0$

b) Produce an ANOVA table with all appropriate sources of variation, including the (corrected) total. Include sums of squares, degrees of freedom, and appropriate mean squares.

term	i(f)	j(f)	k(r)	df	SS
A τ_if	0	b	n	a-1	$bn\sum^a(\bar{y}_{i..} - \bar{y}_{...})^2; \frac{\sum^a y^2_{i..}}{bn} - \frac{y^2_{...}}{abn}; \bar{y}_{1..} = 14; \bar{y}_{2..} = 12$ $3*4*[(14-12)^2 + (10-12)^2] = 96$
B β_{ij}f	a	0	n	b-1	$an\sum^b(\bar{y}_{.j.} - \bar{y}_{...})^2; \frac{\sum^b y^2_{.j.}}{an} - \frac{y^2_{...}}{abn}; \bar{y}_{.1.} = 9; \bar{y}_{.2.} = 11; \bar{y}_{.3.} = 16$ $2*4*[(9-12)^2 + (11-12)^2 + (16-12)^2] = 208$
AB $(\tau\beta)_{ij}$f	0	0	n	(a-1)(b-1)	$n\sum^a\sum^b(y_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2; n\sum\sum y^2_{ij.} - \frac{1}{abn}y^2_{...} - SS_A - S$ $4*[(13-14-9+12)^2 + 1^2 + (-3)^2 + (-2)^2 + (-1)^2 + 3^2]$
Eε_{ijk}r	1	1	1	ab(n-1)	$SST - \sum SS; (n-1)\sum^a\sum^b S^2_{ij}$;126
Total				abn-1	$\bar{y}_{...} = 12; \sum\sum\sum(y_{ijk} - \bar{y}_{...})^2; \sum\sum\sum y^2_{ijk} - \frac{y^2_{...}}{abn}$;542

```
bar_y... <- (13+14+15+5+8+17)/6; bar_y1.. <- (13+14+15)/3; bar_y2.. <-
hat_y11. <- 13-14-9+12; hat_y12. <- 14-14-11+12; hat_y13. <- 15-14-16+1
SS_a <- 3*4*((bar_y1..-bar_y...)^2+(bar_y2..-bar_y...)^2)
SS_b <- 2*4*((bar_y.1.-bar_y...)^2+(bar_y.2.-bar_y...)^2+(bar_y.3.-bar_
SS_ab <- 4*(hat_y11.^2+hat_y12.^2+hat_y13.^2+hat_y21.^2+hat_y22.^2+hat_
SSE <- (4-1)*(1^2+2.5^2+3.5^2+2.5^2+2^2+3.5^2)
SS_a/1;SS_b/2;SS_ab/2;SSE/18; SS_a+SS_b+SS_ab+SSE
## [1] 96
## [1] 104
## [1] 56
## [1] 7
## [1] 542
```

```
sd <- c(1,2.5,3.5,2.5,2,3.5)
yij.bar <- c(13,14,15,5,8,17)
(yi..bar <- c(mean(yij.bar[1:3]),mean(yij.bar[4:6])))
## [1] 14 10
(y.j.bar <- c(mean(yij.bar[c(1,4)]),mean(yij.bar[c(2,5)]),mean(yij.bar[
## [1]  9 11 16
(y...bar <- mean(yij.bar))
## [1] 12
(ssa <- 3*4*sum((yi..bar-y...bar)^2))
## [1] 96
(ssb <- 2*4*sum((y.j.bar-y...bar)^2))
## [1] 208
(ssab <- 4*sum((yij.bar-
            rep(yi..bar,1,each=3)-
            rep(y.j.bar,2,each=1)+
            y...bar)^2))
## [1] 112
(sse <- (4-1)*sum(sd^2))
## [1] 126
(sst <- ssa+ssb+ssab+sse)
## [1] 542
ssa/1; ssb/2; ssab/2; sse/18
## [1] 96
## [1] 104
## [1] 56
## [1] 7
```

c) Test whether your model in part (a) may be reduced to a model in which the effects of system and traffic intensity are purely additive. Remember to state the null and alternative hypotheses. Use $\alpha = 0.05$.

```
(F0 <- ssab/2*18/sse)
## [1] 8
pf(8,2,18,lower.tail = F)
## [1] 0.003266949
qf(0.05,2,18,lower.tail = F)
## [1] 3.554557
```

$H_0: (\tau\beta)_{ij} = 0 \forall i,j$; $F_{p,2,18} \frac{MS_{AB}}{MSE} = \frac{56}{7} = 8$; $F_{0.05,2,8} = 3.55$. There is enough evidence to reject H_0. The model may not be reduced, as the interaction effects is significant at 5% significance level.

d) Form a two-sided 95% confidence interval for the difference in median travel time between the new system and the old system under moderate traffic conditions.

```
yij.bar[2]-yij.bar[5]-qt(0.025,18,lower.tail = F)*sqrt(2*sse/18/4)
## [1] 2.069535
yij.bar[2]-yij.bar[5]+qt(0.025,18,lower.tail = F)*sqrt(2*sse/18/4)
## [1] 9.930465
```

$\bar{y}_{12.} - \bar{y}_{22.} \pm t_{\frac{\alpha}{2},18}\sqrt{\frac{2MSE}{n}} = 14 - 8 \pm 2.1\sqrt{\frac{2*7}{4}} = 6 \pm 3.9287$;[2.0713, 9.9287]

2018F3

[13.2]
Consider the linear mixed model $y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$, $i = 1,..,a$ $j = 1,..,b$ $k = 1,..,n$, $\sum_{i=1}^{a} \alpha_i = 0$, $\beta_{ij} \sim N(0, \sigma_\beta^2)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
with all β_{ij}'s and ε_{ij}'s independent, where $a \geq 2$, $b \geq 2$, and $n \geq 2$. The parameters μ, α_i, σ_β^2, and σ_ε^2 are assumed to be unknown. Adopt the following notation: $\bar{y}_{ij.} = \frac{1}{n}\sum_{k=1}^{n} y_{ijk}$, $\bar{y}_{i..} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$, $\bar{y}_{...} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$
$Cor(y_{111}, y_{112}) = \frac{Cov(y_{111}, y_{112})}{se(y_{111})se(y_{112})} = \frac{MSE \cdot C_{12}}{\sqrt{MSE \cdot C_{11} MSE \cdot C_{12}}}$

a) In terms of the parameters, find the correlations between (i) y_{111} and y_{112}, (ii) y_{111} and y_{121}, and (iii) y_{111} and y_{211}.

$Cov(y_{111}, y_{112}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{11} + \varepsilon_{112}) = Var(\beta_{11}) + Cov(\varepsilon_{111}, \varepsilon_{112}) = \sigma_\beta^2$
$Var(y_{111}) = \sigma_\beta^2 + \sigma_\varepsilon^2 = Var(y_{112})$; $Cor(y_{111}, y_{112}) = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_\varepsilon^2}$
$Cov(y_{111}, y_{121}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{12} + \varepsilon_{121}) = Cov(\beta_{11}, \beta_{12}) + Cov(\varepsilon_{111}, \varepsilon_{121}) = 0$
$Cov(y_{111}, y_{211}) = Cov(\beta_{11} + \varepsilon_{111}, \beta_{21} + \varepsilon_{211}) = Cov(\beta_{11}, \beta_{21}) + Cov(\varepsilon_{111}, \varepsilon_{211}) = 0$
$Cor(y_{111}, y_{121}) = Cor(y_{111}, y_{211}) = 0$

b) For any given value of i, specify the **joint** distribution of $\bar{y}_{i1.}, ..\bar{y}_{ib.}$. [3.4.3]

$\bar{y}_{ij.}$ is a linear combination of $\mu, \alpha_i, \beta_{ij}, \varepsilon_{ijk}$
A linear combination of normal distributed random variables and constants are normal distributed.
$E[\bar{y}_{ij.}] = E[\frac{1}{n}\sum_k^n y_{ijk}] = E[\mu + \alpha_i + \beta_{ij} + \bar{\varepsilon}_{ij.}] = \mu + \alpha_i, \forall i_{1:a}; j_{1:b}$
$Var[\bar{y}_{ij.}] = Var[\mu + \alpha_i + \beta_{ij} + \bar{\varepsilon}_{ij.}] = \sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2$
$f(\bar{y}_{i1.}, ..\bar{y}_{ib.}) = \prod_j^b f(\bar{y}_{ij.}) = (2\pi(\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2))^{-\frac{b}{2}} \exp[\frac{-1}{2(\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2)}\sum_j^b (\bar{y}_{ij.} - \mu - \alpha_i)^2]$

c) In terms of the data, write a formula for the usual unbiased estimator of $\alpha_1 - \alpha_2$. What is the exact distribution of this estimator?

$E[\bar{y}_{1..} - \bar{y}_{2..}] = E[\alpha_1 - \alpha_2 + \bar{\beta}_{1.} - \bar{\beta}_{2.} + \bar{\varepsilon}_{1..} - \bar{\varepsilon}_{2..}] = \alpha_1 - \alpha_2$
$V[] = Var[\bar{\beta}_{1.} - \bar{\beta}_{2.} + \bar{\varepsilon}_{1..} - \bar{\varepsilon}_{2..}] = V[\bar{\beta}_{1.}] + V[\bar{\beta}_{2.}] + V[\bar{\varepsilon}_{1..}] + V[\bar{\varepsilon}_{2..}] = \frac{2}{b}\sigma_\beta^2 + \frac{2}{bn}\sigma_\varepsilon^2$
$\hat{\alpha}_1 - \hat{\alpha}_2$ is a combination of normal distributed r.v. $\sim N(\alpha_1 - \alpha_2, \frac{2}{b}\sigma_\beta^2 + \frac{2}{bn}\sigma_\varepsilon^2)$

d) Show that $E[\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..})^2] = a(b-1)(\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2)$ Justify all important steps. (Hint: Your answer to part (b) might be useful.)

$\bar{y}_{ij.} - \bar{y}_{i..} = \mu + \alpha_i + \beta_{ij} + \bar{\varepsilon}_{ij.} - (\mu + \alpha_i + \bar{\beta}_{i.} + \bar{\varepsilon}_{i..}) = \beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}$
$E[\bar{y}_{ij.} - \bar{y}_{i..}] = E[\beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = 0$
$Cov(\beta_{ij}, \bar{\beta}_{i.}) = \frac{1}{b}Cov(\beta_{ij}, \sum_j^b \beta_{ij}) = \frac{1}{b}[1 \cdot \sigma_\beta^2 + (b-1) \cdot 0]$
$Cov(\bar{\varepsilon}_{ij.}, \bar{\varepsilon}_{i..}) = Cov(\frac{1}{n}\sum_k^n \varepsilon_{ijk}, \frac{1}{bn}\sum_j^b\sum_k^n \varepsilon_{ijk}) = \frac{1}{bn^2}\sum_k^n Cov(\varepsilon_{ijk}, \sum_j^b \varepsilon_{ijk}) = \frac{1}{bn}\sigma_\varepsilon^2$
$Var[\bar{y}_{ij.} - \bar{y}_{i..}] = V[\beta_{ij} - \bar{\beta}_{i.} + \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = V[\beta_{ij} - \bar{\beta}_{i.}] + V[\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..}] = V[\beta_{ij}] + V[\bar{\beta}_{i.}] - 2Cov(\beta_{ij}, \bar{\beta}_{i.}) + V[\bar{\varepsilon}_{ij.}] + V[\bar{\varepsilon}_{i..}] - 2Cov(\bar{\varepsilon}_{ij.}, \bar{\varepsilon}_{i..}) = \sigma_\beta^2 + \frac{1}{b}\sigma_\beta^2 - \frac{2}{b}\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2 + \frac{1}{bn}\sigma_\varepsilon^2 - \frac{2}{bn}\sigma_\varepsilon^2 = \frac{b-1}{b}(\sigma_\varepsilon^2 + \frac{1}{n}\sigma_\varepsilon^2)$
$E[\sum_i^a\sum_j^b(\bar{y}_{ij.} - \bar{y}_{i..})^2] = \sum_i^a\sum_j^b(V[\bar{y}_{ij.} - \bar{y}_{i..}] + E[\bar{y}_{ij.} - \bar{y}_{i..}]^2)$ $\sum_i^a\sum_j^b[\frac{b-1}{b}(\sigma_\varepsilon^2 + \frac{1}{n}\sigma_\varepsilon^2) + 0] = a(b-1)(\sigma_\beta^2 + \frac{1}{n}\sigma_\varepsilon^2)$

e) In terms of the data, write a formula for the usual unbiased (ANOVA) estimate of σ_β^2. (Define all new notation, if you use any)

term	i(f)	j(r)	k(r)	df	EMS	F
α_if	0	b	n	a-1	$\frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i + n\sigma_\beta^2 + \sigma_\varepsilon^2$	$\frac{MS_A}{MS_{AB}}$
β_{ij}r	0	1	n	a(b-1)	$n\sigma_\beta^2 + \sigma_\varepsilon^2$	$\frac{MS_{AB}}{MS_E}$
ε_{ijk}r	1	1	1	ab(n-1)	σ_ε^2	
Total				abn-1		

$\hat{\sigma}_\beta^2 = \frac{1}{n}(MS_{B(A)} - MS_E) = \frac{\sum_i^a\sum_j^b(\bar{y}_{ij.} - \bar{y}_{i..})^2}{a(b-1)} - \frac{\sum_i^a\sum_j^b\sum_k^n(y_{ijk} - \bar{y}_{ij.})^2}{abn(n-1)}$

2018F4

Consider a **randomized complete block design** with 12 blocks and a single treatment factor having 3 levels. Let Y_{ij} denote the response measured for an experimental unit in block j that receives treatment i for $i = 1,2,3$ and $j = 1,..,12$. Suppose there is also a covariate whose value X_{ij} is measured for each experimental unit.
The following four models are fit to the data (using least squares), with the resulting residual (error) sums of squares as specified:
Model 1: $Y_{ij} = \mu + \gamma_j + \varepsilon_{ij}$ $SS(Res) = 660$; Model 2: $Y_{ij} = \mu + \alpha_i + \gamma_j + \varepsilon_{ij}$ $SS(Res) = 550$; Model 3: $Y_{ij} = \mu + \alpha_i + \gamma_j + \beta x_{ij} + \varepsilon_{ij}$ $SS(Res) = 300$; Model 4: $Y_{ij} = \mu + \gamma_j + \beta x_{ij} + \varepsilon_{ij}$ $SS(Res) = 420$
The treatment effects are $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ and the block effects are $\alpha = (\gamma_1, \gamma_2, .., \gamma_{12})'$. The corrected total sum of squares is 820.

	df	SS_F	$-\alpha$	$SS_{\beta\gamma}$	$-\beta$	$SS_{\alpha\gamma}$	$-\alpha$	SS_γ
α	2	110	-110	0		110	-110	0
γ	11	160		160		160		160
β		250	-10	240	-250	0		0
E	22	300	+120	420	+250	550	+110	660
T	35	820						

a) Find the sequential sums of squares for γ_j, α_i, and β, in that order.

$SS_\gamma = 160$, $SS_\alpha = 120$, and $SS_\beta = 240$

b) Form an ANOVA table for the randomized complete block design without the covariate X_{ij}, that is, based on Model 2. The table should include all appropriate sources of variation (including the corrected total), with degrees of freedom, sums of squares, and mean squares where appropriate. Then test whether or not there is any treatment effect based on this model. Use $\alpha = 0.05$.
SS DF MS F P

R 110 2 55 2.2 >.05
B 160 11 14.545 0.5818 >.05
E 550 22 25
T 820 35
$F_{0.05,2,22} = 3.44$

c) Test whether there is any treatment effect, after accounting for both blocking and the covariate. Use $\alpha = 0.05$.
SS DF MS F P

R 110 2 55 3.849 <.05
B 160 11 14.545 1.018 >.05
RB 250 1
E 300 21 14.29
T 820 35

d) Suppose the (possibly incorrect) model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ is fit to the data. Compute the residual sum of squares for this model.

$SSE_\alpha = 820 - (420 - 300) = 710$

2019S

Robert Fountain*, Daniel Taylor-Rodriguez

2019S1

2018S4
Assume the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, $i = 1,..,n$, with the restriction that $\beta_0 = 0$. Find the least-squares estimators of the regression coefficients.
Let $SSE = \sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_2 x_i^2)^2$
$\frac{\partial SSE}{\partial \beta_1} = 2\sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_2 x_i^2)(-x_i) \overset{set}{=} 0$; $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \hat{\beta}_2 \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^2}$
$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_2 x_i^2)(-x_i^2) \overset{set}{=} 0$; $\sum_{i=1}^{n} x_i^2 y_i = \hat{\beta}_1 \sum_{i=1}^{n} x_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^4 = \frac{\sum_{i=1}^{n} x_i y_i - \hat{\beta}_2 \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} x_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^4$
$\hat{\beta}_2\left[\sum_{i=1}^{n} x_i^4 - \frac{(\sum_{i=1}^{n} x_i^3)^2}{\sum_{i=1}^{n} x_i^2}\right] = \sum_{i=1}^{n} x_i^2 y_i - \frac{\sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^2}$
$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i^2 y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^4 \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i^3)^2}$
$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} - \frac{\sum_{i=1}^{n} x_i^3[\sum_{i=1}^{n} x_i^2 y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^3]}{\sum_{i=1}^{n} x_i^2[\sum_{i=1}^{n} x_i^4 \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i^3)^2]}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i^4 - \sum_{i=1}^{n} x_i^2 y_i \sum_{i=1}^{n} x_i^3}{\sum_{i=1}^{n} x_i^4 \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i^3)^2}$$

2019S2

[2015S3][] 2018S2[566-HW2-1] [566-HW5-2]
A manufacturer wishes to know if operator (A), material (B), and heat (C) affect the outcome of the product being produced. All of the machines being used operate at the same four heat settings. Five operators are randomly chosen. There are fifteen varieties of material that can be used, and three of these are assigned to each of the five operators. For each operator, that means that there are 12 combinations of heat and material that can be used. The operator produces two replications of each of these, for a total of $5 \times 3 \times 4 \times 2 = 120$ observations. Incorrectly treating all main effects as fixed and all combinations of factors as crossed, the statistician produces the ANOVA table shown below.
Source SS df MS F pval<0.05
A 34 4 8.5 3.70 *
B 12 2 6.0 2.61
C 24 3 8.0 3.48 *
AB 32 8 4.0 1.74
AC 30 12 2.5 1.09
BC 18 6 3.0 1.30
ABC 36 24 1.5 0.65
Error 138 60 2.3
Total 324 119
Produce the corrected ANOVA table, including the expected mean squares and the correct F statistics. If an exact F test is not available, construct an approximate F statistic (for which you need not compute the degrees of freedom).
F(R)F
Operator (A): a=5 Random;
Material (B): Nested in A, b=3, Fixed;
Heat (C): Crossed with B, c=4, Fixed
Replications: n=2, Random
Model: $y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{kj(i)} + \varepsilon_{(ijk)l}$, $i = 1,2,3,4,5$, $j = 1,2,3$, $k = 1,2,3,4$, $l = 1,2$
$\sum_j^b \beta_{j(i)}$=$\sum_k^c \gamma_k$=$\sum_k^c (\tau\gamma)_{ik}$=$\sum_j^b (\beta\gamma)_{(i)jk}$=$\sum_k^c (\beta\gamma)_{(i)jk}$=0;
E[τ_i=$\beta_{(i)j}$=$(\tau\gamma)_{ik}$=$(\beta\gamma)_{(i)jk}$]=0; V[]=σ_τ^2; $\frac{b-1}{b}\sigma_\beta^2$; $\frac{c-1}{c}\sigma_{\tau\gamma}^2$; $\frac{(b-1)(c-1)}{bc}\sigma_{\beta\gamma}^2$
Source SS df MS F
A 34 4 8.5 8.5/2.3=3.696
B+AB 44 10 4.4 4.4/2.3=1.913
C 24 3 8.0 8/2.5=3.2
AC 30 12 2.5 2.5/2.3=1.087
BC+ABC 54 30 1.8 1.8/2.3=0.7826

Error 138 60 2.3
From Fountain's note, in the j column, treat j as fixed. in the EMS column, treat j(i) and jk(i) as random.

term	i(r)	j(f)	k(f)	l(r)	df	EMS	F
τ_ir	1	b	c	n	a-1	$\sigma^2 + bcn\sigma^2_\tau$	$\frac{A}{E}$
$\beta_{j(i)}$r	1	0	c	n	a(b-1)	$\sigma^2 + cn\sigma^2_\beta$	$\frac{B(A)}{E}$
$(\gamma)_k$f	a	b	0	n	c-1	$\sigma^2 + bn\sigma^2_{\tau\gamma} + \frac{abn\sum^c \gamma_k^2}{c-1}$	$\frac{C}{AC}$
$(\tau\gamma)_{ik}$r	1	b	0	n	(a-1)(c-1)	$\sigma^2 + bn\sigma^2_{\tau\gamma}$	$\frac{AC}{E}$
$(\gamma\beta)_{kj(i)}$r	1	0	0	n	a(b-1)(c-1)	$\sigma^2 + n\sigma^2_{\beta\gamma}$	$\frac{CB(A)}{E}$
$\varepsilon_{(ijk)l}$	1	1	1	1	abc(n-1)	σ^2	

2019S3

[2015S2][]
A company has developed two possible manufacturing processes. They will produce 5 items with each process, in a completely random order. Then they will measure the quality (Y) of each item on a 100-point scale. They suspect that the relative humidity during production might affect the outcome, so they also record it (X). They would like to develop a model that could be used to predict the outcome quality based on the process, with the relative humidity as a covariate. The following table shows and ordered pair for each item, consisting of the quality score and the humidity measurement.
Process A (70,38)(70,55)(68,40)(72,45)(72,36)
Process B (75,30)(74,42)(72,30)(71,30)(73,41)

a) Write an appropriate model for the situation described above, allowing for different slopes and different intercepts for the two processes. Hint: there are 10 observations, and you will need to use at least one indicator variable.

Process A: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; Process B: $y_i = \beta_0 + \gamma_0 + (\beta_1 + \gamma_1)x_i + \varepsilon_i$; Let $w_i = \begin{cases} 0 & 1 \le i \le 5 \\ 1 & 6 \le i \le 10 \end{cases}$, overall $y_i = \beta_0 + \beta_1 x_i + w_i\gamma_0 + w_i\gamma_1 x_i + \varepsilon_i$

b) Write the matrix form of the appropriate model. Show the contents and dimensions of all matrices.

$$\begin{bmatrix}70\\70\\68\\72\\72\\75\\74\\72\\71\\73\end{bmatrix}_{10\times1} = \begin{bmatrix}1&38&0&0\\1&55&0&0\\1&40&0&0\\1&45&0&0\\1&36&0&0\\1&30&1&30\\1&42&1&42\\1&30&1&30\\1&30&1&30\\1&41&1&41\end{bmatrix}_{10\times4} \begin{bmatrix}\beta_0\\\beta_1\\\gamma_0\\\gamma_1\end{bmatrix}_{4\times1} + \begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\\\varepsilon_8\\\varepsilon_9\\\varepsilon_{10}\end{bmatrix}_{10\times1}$$

c) Suppose that you wish to test for equality of the two slopes. Write the matrix form of the reduced model. What will be the numerator and denominator degrees of freedom for the additional sum of squares F test?

$H_0: \gamma_1 = 0, y_i = \beta_0 + \beta_1 x_i + w_i\gamma_0 + \varepsilon_i$

$$\begin{bmatrix}70\\70\\68\\72\\72\\75\\74\\72\\71\\73\end{bmatrix}_{10\times1} = \begin{bmatrix}1&38&0\\1&55&0\\1&40&0\\1&45&0\\1&36&0\\1&30&1\\1&42&1\\1&30&1\\1&30&1\\1&41&1\end{bmatrix}_{10\times3} \times \begin{bmatrix}\beta_0\\\beta_1\\\gamma_0\end{bmatrix}_{3\times1} + \begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\\\varepsilon_8\\\varepsilon_9\\\varepsilon_{10}\end{bmatrix}_{10\times1}$$

$$\begin{bmatrix}0&0&0&1\end{bmatrix}_{10\times4}\begin{bmatrix}\beta_0\\\beta_1\\\gamma_0\\\gamma_1\end{bmatrix}_{4\times1} = 0, r = 1$$

$dfE_{Full} = n - (k+1) = 10 - (3+1) = 6, dfE_{Reduced} = n - (k+1) + r = 7$
$F = \frac{(SSE_{Reduced} - SSE_{Full})/(dfE_{Reduced} - dfE_{Full})}{SSE_{Full}/dfE_{Full}}, df_{nume} = 7 - 6 = 1, df_{deno} = 6$

2019F

2019F1

```
xx <- matrix(c(47, 6500, 397.69, 6500, 1501754, 71104.33, 397.69,
              71104.33, 7587.418),nrow = 3,ncol = 3)
(ixx <- solve(xx))
##              [,1]          [,2]          [,3]
## [1,]  0.0589966941 -1.958369e-04 -1.257021e-03
## [2,] -0.0001958369  1.847088e-06 -7.045029e-06
## [3,] -0.0012570207 -7.045029e-06  2.637045e-04
xy <- c(178.97, 22080.38, 1291.301)
yy <- 819.0221
yHy <- 699.3965
n <- 47 ; p <- 3
(beta <- ixx%*%xy)
##              [,1]
## [1,]  4.611292369
## [2,] -0.003361779
## [3,] -0.040003966
(sse <- yy-t(beta)%*%xy) ; (sse <- yy-yHy)
##          [,1]
## [1,] 119.6256
## [1] 119.6256
(mse <- sse/(n-p))
## [1] 2.718764
(var.beta <- mse*ixx)
##              [,1]          [,2]          [,3]
## [1,]  0.1603980666 -5.324343e-04 -3.417542e-03
## [2,] -0.0005324343  5.021796e-06 -1.915377e-05
## [3,] -0.0034175421 -1.915377e-05  7.169503e-04
(sst <- yy-xy[1]^2/n)
## [1] 137.5272
(ssr <- t(beta)%*%xy-xy[1]^2/n); (ssr <- yHy-xy[1]^2/n)
##          [,1]
## [1,] 17.90156
## [1] 17.90159
(r.sq <- ssr/sst)
## [1] 0.1301676
(f0 <- (ssr)/2/mse)
## [1] 3.292229
pf(f0,2,(n-p),lower.tail = F)
## [1] 0.04651387
(t0 <- abs(beta[2]/sqrt(mse*ixx[2,2])))
## [1] 1.500167
pt(t0,(n-p),lower.tail = F)
## [1] 0.07035658
(t025 <- qt(0.975,(n-p)))
## [1] 2.015368
x0 <- c(1,300,5)
(y0 <- x0%*%beta)
##          [,1]
## [1,] 3.402739
(var.y0 <- mse*x0%*%ixx%*%x0)
##           [,1]
## [1,] 0.2191862
c(y0-t025*sqrt(var.y0),y0+t025*sqrt(var.y0))
## [1] 2.459198 4.346280
```

2019F2

```
sd <- c(3.6,3.1,2.3,4.8,1.9,1.2,5.5)
(sse <- (3-1)*sum(sd^2))
## [1] 172.4
rep <- 3
y...bar <- 10.7
yi..bar <- c(10.5,10.3,11.5)
yij.bar <- c(11.3,9.7,7.7,14.7,8.5,3.8,19.2)
(ssa <- 3*(2*(yi..bar[1]-y...bar)^2+3*(yi..bar[2]-y...bar)^2+2*(yi..bar
## [1] 5.52
(ssb <- 3*(sum((yij.bar[1:2]-yi..bar[1])^2)+sum((yij.bar[3:5]-yi..bar[2
## [1] 447.66
(sst <- ssa+ssb+sse)
## [1] 625.58
(msa <- ssa/2) ; (msb <- ssb/4); (mse <- sse/14)
## [1] 2.76
## [1] 111.915
## [1] 12.31429
pf(msa/mse,2,14,lower.tail=F)
## [1] 0.8020241
pf(msb/mse,4,14,lower.tail=F)
## [1] 0.0007775859
(t_a.bonferroni <- qt(0.05/2/choose(3,2),14,lower.tail = F))
## [1] 2.717755
(t_b.bonferroni <- qt(0.05/2/choose(7,2),14,lower.tail = F))
## [1] 3.699229
(min.diff_a <- t_a.bonferroni*sqrt(mse/(1/6+1/9)))
## [1] 18.09532
(min.diff_b <- t_b.bonferroni*sqrt(2*mse/3))
## [1] 10.59913
(t_a.tukey <- qtukey(0.95,nmeans = 3, df =14))
## [1] 3.701394
(t_b.tukey <- qtukey(0.95,nmeans = 7, df =14))
## [1] 4.828954
(min.diff_a <- t_a.tukey*sqrt(mse/2*(1/6+1/9)))
## [1] 4.840649
(min.diff_b <- t_b.tukey*sqrt(mse/3))
## [1] 9.783564
sort(yij.bar)
## [1]  3.8  7.7  8.5  9.7 11.3 14.7 19.2
```

2019F3

2020S

2020S1

H_0:$\hat\beta_0 = 273.15\hat\beta_1$; $\frac{\hat\beta_0 - 273.15\hat\beta_1}{\sqrt{mse(\frac{1}{n} + \frac{\bar x^2}{S_{xx}})}} = 66.81681$ Wrong

$S_{xx} = \sum x^2 - n\bar x^2 = \sum x^2 - 20 * 10.5^2 = 665; \sum x^2 = 2870$
$X'X = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} = \begin{bmatrix} 20 & 210 \\ 210 & 2870 \end{bmatrix}$
$(X'X)^{-1} = \frac{1}{nS_{xx}} \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix} = \frac{1}{20*665}\begin{bmatrix} 2870 & -210 \\ -210 & 20 \end{bmatrix}$
$se(\hat\beta_0 - 273.15\hat\beta_1) = \sqrt{\frac{MSE}{nS_{xx}}[1, -273.15](X'X)^{-1}[1, -273.15]^T}$
$\sqrt{\frac{MSE}{nS_{xx}}(\sum x^2 + 273.15^2 n + 2 \times 273.15 n\bar x)} = \sqrt{\frac{0.09035889}{20*665}1609811} = 3.307098$
$\frac{\hat\beta_0 - 273.15\hat\beta_1}{se(\hat\beta_0 - 273.15\hat\beta_1)} = \frac{9.3301}{3.307098} = 2.821234 > \Delta$
Bonferronii:$\Delta = t_{\frac{\alpha}{2p}, n-p} = 2.445006$
Scheffe: $\Delta = \sqrt{2F_{\alpha,p,n-p}} = 2.666292$

- Elliptical Joint Conf reg:

$\frac{(\hat\beta - \beta)'(\hat\beta - \beta)}{\sigma^2 (X'X)^{-1}} \sim \chi^2_p$;
P($\frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{p \cdot MSE}$<$F_{\alpha,p,n-p}$)=1-$\alpha$
$n(\hat\beta_0 - \beta_0)^2 + 2\sum x(\hat\beta_0 - \beta_0)(\hat\beta_1 - \beta_1) + \sum x^2(\hat\beta_1 - \beta_1)^2 \le F_{0.05,p,n-p} pMSE$
$20(98.377 - 273.15\beta_1)^2 + 420(98.377 - 273.15\beta_1)(0.326 - \beta_1) + 2870(0.326 - \beta_1)^2 \le 2 * F * 0.09$=0.6423717
There is no real numerical solution. Reject null hypothesis.

```
rm(list=ls())
(xbar <- 10.5); (ybar <- 101.8); (n <- 20)
## [1] 10.5
## [1] 101.8
```

```r
## [1] 20
(sxx <- 665)
## [1] 665
(syy <- 72.3)
## [1] 72.3
(sxy <-  216.79)
## [1] 216.79
(b1 <- sxy/sxx)
## [1] 0.326
(b0<- ybar - b1*xbar)
## [1] 98.377
(var.e <- (syy-sxy^2/sxx)/(n-2))
## [1] 0.09035889
(var.b1 <- var.e/sxx)
## [1] 0.000135878
(var.b0 <- var.e*(1/n+xbar^2/sxx))
## [1] 0.0194985
pt((b0-273.15*b1)/(sqrt(var.b0)),(n-2),lower.tail=F)
## [1] 2.522397e-23
```

```r
##
xx <- matrix(c(20,210,210,2870),2,2)
ixx <- solve(xx)
a <- c(1,-273.15)
se.b0.b1 <- sqrt(var.e*a%*%ixx%*%a)
##
sx <- sxx+n*xbar^2
se.b0.b1 <- sqrt(var.e/n/sxx*(sx+n*273.15^2+2*273.15*n*xbar))
pt((b0-273.15*b1)/se.b0.b1,(n-2),lower.tail=F)
## [1] 0.005654872
##
##
qt(0.025/2,(n-2),lower.tail=F) # Bonferroni
## [1] 2.445006
(f <- qf(0.05,2,18,lower.tail = F))
## [1] 3.554557
sqrt(2*f)# Scheffe
## [1] 2.666292
y <- function(x){
  20*(98.377-273.15*x)^2+420*(98.377-273.15*x)*(0.326-x)+2870*(0.326-x)^2
}
optim(0.326,fn=y,method="BFGS" )$value < 2*f*var.e
## [1] FALSE
```

2020S2

```r
x<- rep(1:4,1,each=3)
y<- c(17, 20, 23,29, 21, 25,31, 29, 30,45, 43, 47)
(xbar <- mean(x)); (ybar <- mean(y))
## [1] 2.5
## [1] 30
(sxx <- sum((x-xbar)^2))
## [1] 15
(syy <- sum((y-ybar)^2))
## [1] 1110
(sxy <-  sum((x-xbar)*(y-ybar)))
## [1] 120
(b1 <- sxy/sxx)
## [1] 8
(b0<- ybar - b1*xbar)
## [1] 10
(yhat <- b0 + b1*x)
##  [1] 18 18 18 26 26 26 34 34 34 42 42 42
(sst <- sum((y-ybar)^2))
## [1] 1110
(sse <- sum((y-yhat)^2))
## [1] 150
(ssr <- sum((yhat-ybar)^2))
## [1] 960
ssr/sst
## [1] 0.8648649
(var.e <- sse/(12-2)); (var.e <- (syy-sxy^2/sxx)/(12-2))
## [1] 15
## [1] 15
yi.bar <- c(mean(y[1:3]),mean(y[4:6]),mean(y[7:9]),mean(y[10:12]))
(ssa <- 3*sum((yi.bar-ybar)^2))
## [1] 1050
(sse <-  sum((y-rep(yi.bar,1,each=3))^2))
## [1] 60
# library(pastecs); stat.desc(x); stat.desc(y)
fit1 <- lm(y~x)
fit2 <- aov(y~as.factor(x))
summary(fit1)
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5.00  -3.25   0.00   3.00   5.00
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.000      2.739   3.651  0.00445 **
## x              8.000      1.000   8.000 1.18e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.873 on 10 degrees of freedom
## Multiple R-squared:  0.8649, Adjusted R-squared:  0.8514
## F-statistic:    64 on 1 and 10 DF,  p-value: 1.177e-05
anova(fit1)
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x          1    960     960      64 1.177e-05 ***
## Residuals 10    150      15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(fit2)
## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(x)  3   1050   350.0  46.667 2.055e-05 ***
## Residuals     8     60     7.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(fit1,fit2)
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ as.factor(x)
##   Res.Df RSS Df Sum of Sq  F Pr(>F)
## 1     10 150
## 2      8  60  2        90  6 0.0256 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
olsrr::ols_pure_error_anova(fit1)
## Lack of Fit F Test
## ---------------
## Response :   y
## Predictor:   x
##
##                       Analysis of Variance Table
## ----------------------------------------------------------------------
##                 DF    Sum Sq    Mean Sq    F Value       Pr(>F)
## ----------------------------------------------------------------------
## x                1    960.00     960.00     128.00    5.072821e-07
## Residual        10    150.00      15.00
##  Lack of fit     2     90.00      45.00       6.00          0.0256
##  Pure Error      8     60.00       7.50
## ----------------------------------------------------------------------
```

2020S3

$E[y_{ijk}] = \mu + \tau_i; V[y_{ijk}] = \sigma^2 + \sigma_\beta^2$

$Cov[y_{ijk}, y_{ijk'}] = \sigma_\beta^2; Cov[y_{ijk}, y_{ij'k'}] = 0$

SSA=4000; SSB(A)=2400; SSE=2250; SST=8650

dfA=2; dfB(A)=12; dfE=45; dfT=59

MSA=2000; MSB(A)=200; MSE=50

Fa=10; Fb(a)=4