(1) Consider a one-way ANOVA model. There are 3 small classes with sizes $n_1 = 10$, $n_2 = 12$ and $n_3 = 15$ students respectively. Let $y_{ij}$ represent the test score of student $j$ in class $i$. Assume $y_{ij}$'s are independent random variables with $y_{ij} \sim N(\theta_i, \sigma^2)$, $j = 1, \ldots, n_i$ and $i = 1, 2, 3$. $\theta_1, \theta_2, \theta_3$ and $\sigma^2$ are unknown parameters. Suppose we want to test $H_0$: $\theta_1 = \theta_2 - 2 = \theta_3 - 5$ (reduced model) versus the full model $H_1$.

(1a) (6 points) Complete the following ANOVA table by filling each empty cell with the correct numerical value of degree of freedom and corresponding mean square.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>D.F.</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>SSR_1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Reduced Model</td>
<td>SSR_0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>SSE_0 - SSE_1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>SSE_1</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SSTO</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: ANOVA table for two nested models

(1b) (6 points) Spell out the following sums of squares in terms of $y_{ij}$'s:

\[
\begin{align*}
SSTO &= \frac{3}{n} \sum_{i=1}^{3} \left( \bar{y}_{i.} - \bar{y}_{..} \right)^2 \\
SSR_1 &= \frac{3}{n} \sum_{i=1}^{3} n_i \left( \bar{y}_{i.} - \bar{y}_{..} \right)^2 \\
SSE_1 &= \frac{3}{n} \sum_{i=1}^{3} \sum_{j=1}^{n_i} \left( y_{ij} - \bar{y}_{i.} \right)^2 \\
SSR_0 &= \frac{3}{n} \sum_{i=1}^{3} \sum_{j=1}^{n_i} \left( y_{ij} - \bar{y}_{i.} - \bar{y}_{..} \right)^2 = \text{constant} \\
SSE_0 &= \frac{3}{n} \sum_{i=1}^{3} \sum_{j=1}^{n_i} \left( y_{ij} - \bar{y}_{i.} \right)^2 \\
\end{align*}
\]

\[
\begin{align*}
\bar{y}_{i.} &= \bar{y}_{1.}, \bar{y}_{2.}, \bar{y}_{3.} \quad i=1,2,3
\end{align*}
\]

Note: Under $H_0$, \( \{ \theta_2 = \theta_1 + 2 \} \) \( \theta_3 = \theta_1 + 5 \) the LSE for the single parameter $\theta_1$ is given by:

\[
\hat{\theta}_1 = \frac{1}{n} \left[ \sum_{i=1}^{3} y_{i.} - \sum_{i=1}^{3} \left( \bar{y}_{i.} - \bar{y}_{..} \right) \right] = \bar{y}_{..} - \frac{2n_1 + 5n_3}{n} = \bar{y}_{..} - \frac{99}{37}
\]

Hence, \( \hat{\theta}_1 = \bar{y}_{1.} \)

\[
\begin{align*}
\hat{\theta}_2 &= \hat{\theta}_1 + 2 \\
\hat{\theta}_3 &= \hat{\theta}_1 + 5
\end{align*}
\]
(2) Check "Yes" or "No". No explanation needed or considered. (2 points for each of the following 4 parts)

(2a) In the simple linear regression model

\[ y_1 = \beta_0 - 3\beta_1 + \epsilon_1, \]
\[ y_2 = \beta_0 + \beta_1 + \epsilon_2, \]
\[ y_3 = \beta_0 + 2\beta_1 + \epsilon_3, \]

\( \epsilon_i, i = 1, 2, 3 \) are i.i.d \( N(0, \sigma^2) \) random variables, and \( \beta_0, \beta_1 \) and \( \sigma \) are unknown parameters. Are the least square estimators \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) uncorrelated?

Yes \( \checkmark \); No \( \big/ \). 

because the coeff. for \( \beta_1 \)
\begin{align*}
&\text{(2b) Recall in Chapter 4, Atkinson proposed the normal/half normal plots for detecting} \\
&\text{influential observations. For example, a simulated envelope consisting of two nearly straight} \\
&\text{lines is generated via percentiles from the standard normal distribution. You did a problem} \\
&\text{in Homework 3 by applying this procedure to the DFFITS measure. Does the same simulated} \\
&\text{envelope method apply to Cook's D measure } D_i = \frac{\hat{e}_i^2}{\hat{\sigma}^2} \frac{1}{(1-h_i)^2} \text{?} \\
&\text{Yes} \big/ \; \text{No} \checkmark.
\end{align*}

(2c) Consider simultaneous confidence intervals (CI's) for \( \beta_i, i = 1, 2, 3 \) in a regression model. Does Scheffe's procedure always yield narrower CI's than Bonferroni's procedure, based on the same data and the same coverage probability?

Yes \( \checkmark \); No \( \big/ \).

(2d) A common sense experience tells us the older a used car, the lower sale value it tends to have. To model this causal relationship between response \( y \) (used car value) and covariate \( x \) (age of the car), is it appropriate to start with a simple linear regression \( y = \beta_0 + \beta_1 x + \epsilon \)?

Yes \( \checkmark \); No \( \big/ \). 

Use \( \log y \) as the response instead of \( y \).

(3) The director of admissions in a small college administered a newly designed entrance test to 20 students selected at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (\( y \)) can be predicted from the entrance test score (\( x \)). A simple linear regression model was fitted to the data as follows:

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i; \quad \epsilon_i \text{ i.i.d } \sim N(0, \sigma^2), \quad i = 1, \ldots, n. \]

The edited R output and other summaries are given below.
Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -1.6996  | 0.7268     | -2.338  | 0.0311 * |
| Entrance       | 0.6399   | 0.1440     | 4.331   | 1.60e-05 *** |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response: GPA

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrance</td>
<td>1</td>
<td>6.4337</td>
<td>6.4337</td>
<td>33.998</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>3.4063</td>
<td>0.1892</td>
<td>0.1892</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

\[ \sum_{i=1}^{n} y_i = 50, \quad \sum_{i=1}^{n} x_i = 100, \quad \sum_{i=1}^{n} x_i y_i = 257.66, \quad \sum_{i=1}^{n} y_i^2 = 134.84, \quad \sum_{i=1}^{n} x_i^2 = 509.12 \]

(3a) (5 points) Use Bonferroni and Scheffé methods to obtain 90% simultaneous confidence intervals for \( \beta_0 \) and \( \beta_1 \). Which method do you prefer? Why?

For \( B^- \), \( G = t_{n-2, 1-\alpha/2} = t_{18, 1-0.05} = 2.101 \)

For \( S^- \), \( G = \sqrt{S_{S-}^2/n-2; 1-\alpha} = \sqrt{2 \cdot S_{S}, 18} = 2.291 \)

Also, \( SE(\hat{\beta}_0) = 0.7268 \), \( SE(\hat{\beta}_1) = 0.1440 \). Hence, the simultaneous CI's:

For \( B^- \), \[ \beta_0 : -1.6996 \pm 2.101 \cdot 0.7268 = (-3.2267, -0.1726) \]
\[ \beta_1 : 0.8399 \pm 2.101 \cdot 0.1440 = (0.5374, 1.1424) \]

For \( S^- \), \[ \beta_0 : -1.6996 \pm 2.291 \cdot 0.7268 = (-3.3647, -0.0345) \]
\[ \beta_1 : 0.8399 \pm 2.291 \cdot 0.1440 = (0.51, 1.1698) \]

(3b) (5 points) Calculate a 90% prediction interval for the GPA at the end of the freshman year for a student with the entrance test score 5.0.

Using \( \bar{y} = 77.8, \bar{x} = (2, 33) \) with \( K = 1 \), \( G = t_{18, 0.95} = 1.734 \), 
\[ s = \sqrt{0.1892} = 0.435 \], and \( \bar{x} = 5.0 \) implies
\[ \sqrt{\frac{1}{n} + \frac{(x-x)^2}{\sum(x_i-x)^2}} + 1 = \sqrt{\frac{1}{18} + 1} = 1.0247 \]. Hence the 90% prediction interval is given by \[ 77.8 \pm t_{18, 0.95} \cdot s \cdot \sqrt{\frac{1}{n} + 1} \]
\[ = 2.5 \pm 1.734 \cdot 0.435 \cdot 1.0247 = (1.727, 3.273) \]
(3c) (5 points) Suppose we want to test \( H_0 : y_i = -1 + x_i + \epsilon_i \) (reduced model) versus \( H_1 : y_i = \beta_0 + \beta_1 x_i + \epsilon_i \) (full model). Derive a F-test procedure, calculate the observed test statistic and state your conclusion using the significance level \( \alpha = 0.1 \).

\[
\text{With } q = 2 \quad \text{and} \quad n-p = 18, \quad F = \frac{(SSE_0 - SSE_1)/(n-p)}{SSE_1/(n-p)} = 119.51
\]

Hence reject \( H_0 \).

More details: \( SSE_1 = 3.4063 \)

and \( SSE_0 = \sum_{i=1}^{20} \left[ y_i - (-1 + x_i) \right] = \sum_{i=1}^{20} y_i^2 + 20 + \sum_{i=1}^{20} x_i^2 + 2 \sum_{i=1}^{20} y_i \beta_0 - 2 \sum_{i=1}^{20} x_i \beta_1 \)

\[
= 134.84 + 20 + 509.12 + 100 - 200 - 2 \cdot 257.66 = 48.64
\]

(3d) (5 points) Consider two pairs of parameter values under the alternative \( H_1: (\beta_0, \beta_1) = (-0.5, 1.2) \) and \( (\beta_0, \beta_1) = (-0.8, 1.5) \). Which pair would give a higher power for the test you derived in (3c)? Why?

\( \sigma^2 \delta^2 = (h-h')^T \left[ C (X^T X)^{-1} C^T \right]^{-1} (h-h') \)

In this case, \( C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) hence

\[
\sigma^2 \delta^2 = (h-h')^T (X^T X) (h-h') \quad \text{where} \quad X^T X = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = \begin{pmatrix} 20 & 100 \\ 100 & 509.12 \end{pmatrix}
\]

For the alternative pair \( (\beta_0, \beta_1) = (-0.5, 1.2) \):

\[
\sigma^2 \delta^2 = (0.5, 0.2) \begin{pmatrix} 20 & 100 \\ 100 & 509.12 \end{pmatrix} (0.5, 0.2) = 45.36
\]

For \( (\beta_0, \beta_1) = (-0.8, 1.5) \):

\[
\sigma^2 \delta^2 = (0.2, 0.15) \begin{pmatrix} 20 & 100 \\ 100 & 509.12 \end{pmatrix} (0.2, 0.15) = 148.08
\]

Hence the pair \( (-0.8, 1.5) \) will give a higher power.
(4) (10 points) Assume a linear model $Y = X\beta + \epsilon$, where the error vector $\epsilon$ has iid $N(0, \sigma^2)$ components. Let $\alpha$ be a fixed $p \times 1$ vector of the same dimension as $\beta$, and $L$ be an $n \times 1$ vector of constants, such that $LY$ is an unbiased estimator of $\alpha^T \beta$. Show that $LY$ is the minimum variance linear unbiased estimator of $\alpha^T \beta$ if and only if $D^T Y$ and $LY$ are statistically independent for every $n \times 1$ constant vector $D$ which satisfies $E(D^T Y) = 0$ for all $\beta$.

Note: Please prove the sufficiency and necessity respectively by stating clearly the assumption and conclusion in each direction.

Sufficiency $\implies$ : Let $c^T Y$ be any unbiased est. of $\alpha^T \beta$, then $E[(c-L)^T Y] = 0$, hence $(c-L)^T Y$ & $LY$ are independent RV's. Therefore,

$$\text{Var}(c^T Y) = \text{Var}[(c-L)^T Y + LY] = \text{Var}[(c-L)^T Y] + \text{Var}(LY) \geq \text{Var}(LY).$$

Necessity $\implies$ : Suppose $LY$ is BLUE, and $E(D^T Y) = 0$.

For an arbitrary constant $a > 0$, both $(L+aD)^T Y$ & $(L-aD)^T Y$ are unbiased est. of $\alpha^T \beta$.

$$\text{Var}[(L+aD)^T Y] = \text{Var}(LY) + a^2 \text{Var}(D^T Y) + 2a \text{Cov}(LY, D^T Y) \geq \text{Var}(LY) \implies a \text{Var}(D^T Y) + 2 \text{Cov}(LY, D^T Y) \geq 0 \quad \forall \ a > 0 \quad \text{(\#)}$$

By the same token,

$$\text{Var}[(L-aD)^T Y] \geq \text{Var}(LY) \implies a \text{Var}(D^T Y) - 2 \text{Cov}(LY, D^T Y) \geq 0 \quad \forall \ a > 0 \quad \text{...(\#\#)}$$

Let $a \neq 0$ in (\#) & (\#\#) we have

$$\text{Cov}(LY, D^T Y) \geq 0 \quad \text{Hence \ Cov}(LY, D^T Y) = 0$$

$$\text{Cov}(LY, D^T Y) \leq 0 \quad \text{which implies}$$

$LY$ & $D^T Y$ are indep because they are both normal RV's.