1.

Part I **Consider a situation in which the data set is divided into two parts according to the values of $x_2$. The categorical variable, $x_{2c}$, is equal to 0 if $x_2$ is less than its sample median and equal to 1 otherwise. For the categorical variable $x_{2c}$, choose "0" as reference group.**

**Fit a model of $y$ on $x_1$ and $x_{2c}$ , i.e.,**

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2c} + \beta_{12} x_{i1} x_{i2c} + e_i \quad , \quad e_i \sim_{iid} N(0, \sigma^2)}$$

(a) Write down the fitted lines of $y$ on $x_1$ for different values of $x_{2c}$.

Answer:

Consider Model II:

$$y_{ij} = \alpha_j + \gamma_j x_{ij} + e_{ij}$$

for j=0,1 and i=1,2,...,10
For each level, it is simple linear regression. Therefore, we have the following results:

$$\begin{aligned}
\hat{\gamma}_0 &= \frac{S^{(0)}_{x_1 y}}{S^{(0)}_{x_1 x_1}} = \frac{199.6}{95.6} = 2.0879 \\
\hat{\alpha}_0 &= \bar{y}^{(0)} - \hat{\gamma}_0 \bar{x}^{(0)} = 11.2 - 2.0879 * 5.2 = 0.3429 \\
\hat{\gamma}_1 &= \frac{S^{(1)}_{x_1 y}}{S^{(1)}_{x_1 x_1}} = \frac{218}{94} = 2.3191 \\
\hat{\alpha}_1 &= \bar{y}^{(1)} - \hat{\gamma}_1 \bar{x}^{(1)} = 26.3 - 2.3191 * 8 = 7.7472
\end{aligned}$$

Therefore, the fitted lines are

$$\begin{aligned}
&\text{for } x_{2c} = 0: \ \hat{y} = 0.3429 + 2.0879 x_1 \\
&\text{for } x_{2c} = 1: \ \hat{y} = 7.7472 + 2.3191 x_1
\end{aligned}$$

(b) Hence or otherwise, find the fitted line for the model of $y$ on $x_1$, $x_{2c}$ and their interaction term.

Answer:

Back to model I,

$$\begin{aligned}
\hat{\beta}_0 &= \hat{\alpha}_0 = 0.3429 \\
\hat{\beta}_1 &= \hat{\gamma}_0 = 2.0879 \\
\hat{\beta}_2 &= \hat{\alpha}_1 - \hat{\alpha}_0 = 7.4043 \\
\hat{\beta}_{12} &= \hat{\gamma}_1 - \hat{\gamma}_0 = 0.2312
\end{aligned}$$

The fitted line is

$$\hat{y} = 0.3429 + 2.0879 * x_1 + 7.4043 x_{2c} + 0.2312 x_1 x_{2c}$$

(c) Find the unbiased estimate of the unknown parameter $\sigma^2$. No need to show that it is unbiased.

Answer:

First calculate the RSS for each level.

$$\mathrm{RSS}_0 = S_{yy}^{(0)} - \hat{\gamma}_0 S_{x_1 y}^{(0)} = 70.8552$$
$$\mathrm{RSS}_1 = S_{yy}^{(1)} - \hat{\gamma}_1 S_{x_1 y}^{(1)} = 22.5236$$

Then the RSS for the dataset is the summation of RSS for each level

$$\mathrm{RSS} = \sum_j \mathrm{RSS}_j = 93.3914$$
$$\hat{\sigma}^2 = \frac{\mathrm{RSS}}{n-p} = \frac{93.3914}{20-4} = 5.8370$$

(d) Test $\beta_{12} = 0$ by t-test at the significance level $\alpha = 0.05$ . State the test statistic, the critical value and your decision rule clearly.

Answer:

To test $\beta_{12}$=0 is equivalent to test

$$\gamma_1 - \gamma_0 = 0$$

From part(b), we have got the point estimate for $\beta_{12}$

$$\hat{\beta_{12}} = \hat{\gamma}_1 - \hat{\gamma}_0 = 0.2312$$

For s.e. of $\hat{\beta}_{12}$,

$$Var(\hat{\beta}_{12}) = Var(\hat{\gamma}_1 - \hat{\gamma}_0) = Var(\hat{\gamma}_1) + Var(\hat{\gamma}_0) = \hat{\sigma}^2 (\frac{1}{S_{x_1x_1}^{(1)}} + \frac{1}{S_{x_1x_1}^{(0)}}) = 5.8370 * (\frac{1}{95.6} + \frac{1}{94}) = 0.1232$$

s.e. of $\hat{\beta}_{12} = \sqrt{0.1232} = 0.3510$

Then we can apply t test

$$t_{obs} = \frac{\hat{\beta}_{12}}{\sqrt{Var(\hat{\beta}_{12})}} = \frac{0.2312}{\sqrt{0.1232}} = 0.6588$$

Since $|t_{obs}| < t_{0.025,16}$=2.12, we do not reject $H_0$.

(e) Based on the model in (b) (i.e., $\beta_{12} \neq 0$), test whether the difference of $E(y)$ between $x_{2c} = 1$ and $x_{2c} = 0$ at $x_1 = 4$, is significant at $\alpha = 0.05$. State the test statistic, the critical value and your decision rule clearly.

Answer:

$$E(y_i)|_{x_1=4} = \beta_0 + 4\beta_1 + \beta_2 x_{2c} + 4\beta_{12} x_{2c}$$

Therefore, the difference between $x_{2c}=0$ and $x_{2c}=1$ is

$$\beta_2 + 4\beta_{12}$$

Therefore, the null hypothesis is

$$\begin{aligned}
&\beta_2 + 4\beta_{12} = \alpha_1 - \alpha_0 + 4\gamma_1 - 4\gamma_0 = 0\\
\Leftrightarrow &\alpha_1' + (4-\bar{x}^{(1)})\gamma_1 - \alpha_0' - (4-\bar{x}^{(0)})\gamma_0 = 0\\
&\alpha_1' - 4\gamma_1 - \alpha_0' + 1.2\gamma_0 = 0
\end{aligned}$$

where $\alpha_1'$ and $\alpha_0'$ correspond to the parameters for centered model when $x_{2c}=1$ and $x_{2c}=0$ respectively, and t test can be applied.

$$\begin{aligned}
\hat{\alpha}_1 - \hat{\alpha}_0 + 4\hat{\gamma}_1 - 4\hat{\gamma}_0 &= 7.4043 + 4 * 0.2312 = 8.3291\\
Var(\hat{\alpha}_1' - 4\hat{\gamma}_1 - \hat{\alpha}_0' + 1.2\hat{\gamma}_0) &= Var(\hat{\alpha}_1') + 16Var(\hat{\gamma}_1) + Var(\hat{\alpha}_0') + 1.44Var(\hat{\gamma}_0)\\
&= \hat{\sigma}^2(\frac{1}{n} + \frac{16}{S_{x_1x_1}^{(1)}} + \frac{1}{n} + \frac{1.44}{S_{x_1x_1}^{(0)}})\\
&= 5.8370 * (\frac{1}{10} + \frac{16}{94} + \frac{1}{10} + \frac{1.44}{95.6}) = 2.2489\\
\text{s.e. of } \hat{\alpha}_1 - \hat{\alpha}_0 + 4\hat{\gamma}_1 - 4\hat{\gamma}_0 &= \sqrt{2.2489} = 1.4996\\
t_{obs} &= \frac{\hat{\alpha}_1 - \hat{\alpha}_0 + 4\hat{\gamma}_1 - 4\hat{\gamma}_0}{\sqrt{Var(\hat{\alpha}_1' - 4\hat{\gamma}_1 - \hat{\alpha}_0' + 1.2\hat{\gamma}_0)}} = \frac{8.3291}{\sqrt{2.2489}} = 5.5541
\end{aligned}$$

Since $t_{obs}$ =5.5541 $> t_{0.025,16} = 2.12$, reject $H_0$.

Part II **Consider a situation in which the data set is divided into four parts according to the values of $x_1$ and $x_2$. The categorical variable, $x_{1c}$, is equal to 0 if $x_1$ is less than its sample median and equal to 1 otherwise. The categorical variable, $x_{2c}$, is equal to 0 if $x_2$ is less than its sample median and equal to 1 otherwise. For both categorical variables, $x_{1c}$ & $x_{2c}$, choose "0" as reference group**.

Define $\mu_{ij}$ to be the population mean of $y$ when $x_{1c}=i, x_{2c}=j$ for $i,j=0,1$.

(a) Formulate a suitable model using the population means of these four parts, $\mu_{ij}$. Then, estimate all unknown parameters in the model.

Answer:

$$y_{ijk} = \mu_{ij} + e_{ijk} \quad i=0,1 \quad j=0,1 \quad k=1,2,...,5$$

$$\begin{aligned}
\hat{\mu}_{00} &= \bar{y}_{00} = \frac{26}{5} = 5.2\\
\hat{\mu}_{01} &= \bar{y}_{01} = \frac{96}{5} = 19.2\\
\hat{\mu}_{10} &= \bar{y}_{10} = \frac{86}{5} = 17.2\\
\hat{\mu}_{11} &= \bar{y}_{11} = \frac{167}{5} = 33.4
\end{aligned}$$

(b) Test the population means of these four parts are equal at $\alpha = 0.05$. Write down the test statistic, critical value and your conclusion clearly.

Answer:

$$
\begin{aligned}
TSS &= \sum (y_{ijk} - \bar{y}_{..})^2 = S_{yy} = 2155.75 \\
RSS &= \sum_i \sum_j S_{yy}^{(i,j)} = 151.6 \\
\hat{\sigma}^2 &= \tfrac{RSS}{n-p'} = \tfrac{151.6}{20-4} = 9.475 \\
RegSS &= TSS - RSS = 2004.15 \\
F &= \frac{RegSS/(p'-1)}{RSS/(n-p')} = \frac{2004.15/(4-1)}{151.6/(20-4)} = 70.5066
\end{aligned}
$$

Critical value is $F_{0.05,3,16} = 3.24 < F_{obs} = 70.5066$

So reject $H_0$.

(c) Find orthogonal single-degree-of-freedom contrasts such that the sum of their contrast S.S. is equal to Reg.S.S. in (b). Hence or otherwise, find the significant contrast(s) at the level of $\alpha = 0.05$ to explain the inequality of population means of $y$ for different combinations of $x_{1c}$ & $x_{2c}$.

Answers:

Three orthogonal contrasts are

$$
\begin{aligned}
\omega_1 &= \mu_{11} - \mu_{00} \\
\omega_2 &= \mu_{01} - \mu_{10} \\
\omega_3 &= \mu_{11} - \mu_{01} - \mu_{10} + \mu_{00}
\end{aligned}
$$

It is easy to see that $\omega_1$ $\omega_2$ $\omega_3$ are orthogonal. Since the degree of freedom of RegSS is 3, we have $RegSS = SSW_1 + SSW_2 + SSW_3$.

$$
\begin{aligned}
SSW_1 &= \frac{(\hat{\mu}_{11} - \hat{\mu}_{00})^2}{\frac{1}{n^{(1,1)}} + \frac{1}{n^{(0,0)}}} = \frac{(33.4 - 5.2)^2}{\frac{1}{5} + \frac{1}{5}} = 1988.1 \\
SSW_2 &= \frac{(\hat{\mu}_{01} - \hat{\mu}_{10})^2}{\frac{1}{n^{(0,1)}} + \frac{1}{n^{(1,0)}}} = \frac{(19.2 - 17.2)^2}{\frac{1}{5} + \frac{1}{5}} = 10 \\
SSW_3 &= \frac{(\mu_{11} - \mu_{01} - \mu_{10} + \mu_{00})^2}{\frac{1}{n^{(0,1)}} + \frac{1}{n^{(1,0)}} + \frac{1}{n^{(1,1)}} + \frac{1}{n^{(0,0)}}} = \frac{(33.4 - 19.2 - 17.2 + 5.2)^2}{\frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5}} = 6.05
\end{aligned}
$$

$$
\begin{aligned}
f_1 &= \frac{SSW_1}{\hat{\sigma}^2} = \frac{1988.1}{9.475} = 209.8259 \ > \ F_{0.05,1,16} = 4.49 \\
f_2 &= \frac{SSW_2}{\hat{\sigma}^2} = \frac{10}{9.475} = 1.0554 \ < \ F_{0.05,1,16} \\
f_3 &= \frac{SSW_3}{\hat{\sigma}^2} = \frac{6.05}{9.475} = 0.6385 \ < \ F_{0.05,1,16}
\end{aligned}
$$

Therefore $\omega_1$ is the significant contrast.

(d) Re-write the model in (a) as a model with the interaction term between $x_{1c}$ and $x_{2c}$ and then estimate the unknown parameters in the model.

Answer:

$$y_i = \beta_0 + \beta_1 x_{i1c} + \beta_2 x_{i2c} + \beta_{12} x_{i1c} x_{i2c} + e_i$$

$$\begin{aligned}
\hat{\beta}_0 &= \hat{\mu}_{00} = 5.2 \\
\hat{\beta}_1 &= \hat{\mu}_{10} - \hat{\beta}_0 = 12 \\
\hat{\beta}_2 &= \hat{\mu}_{01} - \hat{\beta}_0 = 14 \\
\hat{\beta}_{12} &= \hat{\mu}_{11} - \hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2 = 2.2
\end{aligned}$$

(e) Write down the hypothesis that there is no interaction between $x_{1c}$ and $x_{2c}$ in terms of $\mu_{ij}$. Hence or otherwise, perform the test at the significance level of 0.05. Write down the test statistic, critical value and your conclusion clearly.

Answer:

$H_0$: $\mu_{00} - \mu_{01} - \mu_{10} + \mu_{11} = 0$

$$\begin{aligned}
\text{s.e. of } \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10} + \hat{\mu}_{11} &= \sqrt{\hat{\sigma}^2 (\sum_{i,j} \frac{1}{n^{(i,j)}})} = \sqrt{9.475 * (\frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5})} = \sqrt{7.58} = 2.7532 \\
t &= \frac{\hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10} + \hat{\mu}_{11}}{\sqrt{\hat{\sigma}^2 (\sum_{i,j} \frac{1}{n^{(i,j)}})}} = \frac{2.2}{2.7532} = 0.7991
\end{aligned}$$

Critical Value is $t_{0.025,16} = 2.12 > t_{obs}$

So do not reject $H_0$.

(f) Assume that there is no interaction between $x_{1c}$ and $x_{2c}$.

i. Find the unbiased estimate of the unknown parameter $\sigma^2$. No need to show that it is unbiased.

Answer:

$$X_c = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ \vdots & \vdots & \vdots \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & -\frac{1}{2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{2} & \frac{1}{2} \\ \vdots & \vdots & \vdots \\ 1 & -\frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$X_c^T X_c = \begin{pmatrix} \text{n} & 0 & 0 \\ 0 & \frac{n}{4} & 0 \\ 0 & 0 & \frac{n}{4} \end{pmatrix} = \begin{pmatrix} 20 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

$$X_c^T Y = \begin{pmatrix} 375 \\ 131/2 \\ 151/2 \end{pmatrix}$$

$$\hat{\beta}' = \begin{pmatrix} \hat{\beta}'_1 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 18.75 \\ 13.1 \\ 15.1 \end{pmatrix}$$

$$\text{RSS} = Y^T Y - \hat{\beta}'^T X_c^T Y = 157.65$$

$$\hat{\sigma}^2 = \frac{\text{RSS}}{n - p'} = \frac{157.65}{20 - 3} = 9.2735$$

ii. Find the sum of squares for the main effect of "$x_{1c}$". Hence or otherwise, test whether the population means of $y$ for $x_{1c} = 1$ and $x_{1c} = 0$ are different at $\alpha = 0.05$.

iii. Find the sum of squares for the main effect of "$x_{2c}$". Hence or otherwise, test whether the population means of $y$ for $x_{2c} = 1$ and $x_{2c} = 0$ are different at $\alpha = 0.05$.

Answer:

$$\text{SSA} = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = 858.05$$

$$\text{SSB} = an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = 1140.05$$

$$F_1 = \frac{\text{SSA}}{\hat{\sigma}^2} = \frac{858.05}{9.2735} = 92.5271 \ > \ F_{0.05,1,17} = 4.45$$

$$F_2 = \frac{\text{SSB}}{\hat{\sigma}^2} = \frac{1140.05}{9.2735} = 122.9363 \ > \ F_{0.05,1,17} = 4.45$$

Therefore, both null hypotheses are rejected.

2. (a) Note that we can rewrite the data as follows
$y = (34, 32.7, 32, 33.2, 29.8, 26.7, 28.7, 30.1, 29.8, 29, 28.1, 28.8, 27.3, 29.7, 28.3)$,
polymer$=x_1 = (0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1)$,
temper$=x_2 = (0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1)$,
And the number for four gropus are

$$n_{0,0} = 4,\ n_{0,1} = 3,\ n_{1,0} = 3,\ n_{1,1} = 5.$$

Thus, in each group, we have

$$\begin{aligned}
n_{0,0} &= 4,\ T_1 = \sum y_i = 131.9,\ \bar{y}_{0,0} = 32.975,\ s_{yy}^{0,0} = 2.1275;\\
n_{0,1} &= 3,\ T_2 = \sum y_i = 85.2,\ \bar{y}_{0,1} = 28.4,\ s_{yy}^{0,1} = 4.94;\\
n_{1,0} &= 3,\ T_3 = \sum y_i = 88.9,\ \bar{y}_{1,0} = 29.63333,\ s_{yy}^{1,0} = 0.64667;\\
n_{1,1} &= 5,\ T_1 = \sum y_i = 142.2,\ \bar{y}_{1,1} = 28.44,\ s_{yy}^{1,1} = 3.152.
\end{aligned}$$

So we have

$$\begin{aligned}
RSS &= \sum_{i,j} s_{yy}^{i,j} = 2.1275 + 4.94 + 0.64667 + 3.152 = 10.86617,\\
\hat{\sigma}^2 &= \frac{RSS}{15-4} = 0.9878333.
\end{aligned}$$

(b) Consider Model II:

$$\begin{aligned}
y_{ijk} &= \mu_{ij} + e_{ijk}, \text{for } i,j = 0,1; and \quad k = 1,2,3,4\ for\ i = j = 0; \quad k = 1,2,3\ for\ i = 0, j = 1;\\
&\quad k = 1,2,3\ for\ i = 1, j = 0; \quad k = 1,2,3,4,5\ for\ i = j = 1.
\end{aligned}$$

The estimates are

$$\begin{aligned}
\hat{\mu}_{00} &= \bar{y}_{0,0} = 32.975\\
\hat{\mu}_{01} &= \bar{y}_{0,1} = 28.4,\\
\hat{\mu}_{10} &= \bar{y}_{1,0} = 29.63333\\
\hat{\mu}_{11} &= \bar{y}_{1,1} = 28.44
\end{aligned}$$

$$\beta_{12} = 0 \Leftrightarrow \mu_{11} + \mu_{00} - \mu_{01} - \mu_{10} = 0$$

Here, t test can be applied.

$$\begin{aligned}
\hat{\mu}_{11} + \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10} &= 32.975 + 28.44 - 28.4 - 29.63333 = 3.381667\\
Var(\hat{\mu}_{ij}) &= \hat{\sigma}^2 \frac{1}{n_{(i,j)}}\\
t_{obs} &= \frac{\hat{\mu}_{11} + \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10}}{\sqrt{Var(\hat{\mu}_{11} + \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10})}}\\
&= \frac{\hat{\mu}_{11} + \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10}}{\sqrt{\hat{\sigma}^2 \sum_{ij} \frac{1}{n_{i,j}}}} = \frac{3.381667}{\sqrt{1.103081}} = 3.219787 > t_{0.025,11} = 2.201
\end{aligned}$$

Therefore, reject H0, i.e., the interaction between polymer and temperature is significant.

(c) i. Note that the interaction term is insignificant. Then the model becomes

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i\\
&= \beta_0' + \beta_1 x_{i1c} + \beta_2 x_{i2c} + e_i\\
&= X_c \beta',
\end{aligned}$$

where $\quad \beta_0' = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2;\ x_{i1c} = x_{i1} - \bar{x}_1;\ x_{i2c} = x_{i2} - \bar{x}_2,$

and

$$\mathbf{X_c} = \begin{pmatrix} 1 & -\frac{8}{15} & -\frac{8}{15} \\ \vdots & \vdots & \vdots \\ 1 & -\frac{8}{15} & -\frac{8}{15} \\ \hdashline 1 & -\frac{8}{15} & \frac{7}{15} \\ \vdots & \vdots & \vdots \\ 1 & -\frac{8}{15} & \frac{7}{15} \\ \hdashline 1 & \frac{7}{15} & -\frac{8}{15} \\ \vdots & \vdots & \vdots \\ 1 & \frac{7}{15} & -\frac{8}{15} \\ \hdashline 1 & \frac{7}{15} & \frac{7}{15} \\ \vdots & \vdots & \vdots \\ 1 & \frac{7}{15} & \frac{7}{15} \end{pmatrix} \quad \mathbf{X_c}^T\mathbf{X_c} = \begin{pmatrix} 15 & 0 & 0 \\ 0 & \frac{56}{15} & \frac{11}{15} \\ 0 & \frac{11}{15} & \frac{56}{15} \end{pmatrix}$$

$$\mathbf{X_c}^T\mathbf{Y} = \begin{pmatrix} 448.2 \\ \text{-7.94} \\ \text{-11.64} \end{pmatrix}$$

$$\Rightarrow \hat{\beta}' = (\mathbf{X_c}^T\mathbf{X_c})^{-1}\mathbf{X_c}^T\mathbf{Y} = \begin{pmatrix} 29.88 \\ \text{-1.575124} \\ \text{-2.808458} \end{pmatrix}$$

$$\Rightarrow \hat{\beta} = \begin{pmatrix} \text{29.88-(-1.575124) * } \frac{8}{15}\text{-(-2.808458)* } \frac{8}{15} \text{ =32.21791} \\ \text{-1.575124} \\ \text{-2.808458} \end{pmatrix}$$

Thus, we have

$$\begin{aligned} ResS.S. &= \mathbf{Y}^T\mathbf{Y} - \hat{\beta'}^T\mathbf{X_c}^T\mathbf{Y} = 21.10706 \\ \tilde{\sigma^2} &= \frac{Rse.S.S}{15-3} = 1.758922 \end{aligned}$$

So estimation of $\sigma^2$ is 1.758922.

ii. Denote the factor polymer as A, and factor temperature as B, since no interaction between A and B, thus main effect of A can be written as

$$SS(A) = (C\hat{\beta})^T(C(X^TX)^{-1}C^T)^{-1}(C\hat{\beta}),$$

where $C = (0, 1, 0)$. It is easy to show

$$(\mathbf{X_c}^T\mathbf{X_c})^{-1} = \begin{pmatrix} \frac{1}{15} & 0 & 0 \\ 0 & \frac{56}{201} & \text{-}\frac{11}{201} \\ 0 & \text{-}\frac{11}{201} & \frac{56}{201} \end{pmatrix}.$$

Note the diagonal element of $(X^TX)^{-1}$ and $(X_c^TX_c)^{-1}$ are the same, so we have

$$SS(A) = (-1.575124)^2 * (\frac{56}{201})^{-1} = 8.905078.$$

Thus value of test statistic is

$$f_1 = \frac{SS(A)/1}{\tilde{\sigma^2}} = 5.062804,$$

note that critical value is $F_{0.05,1,12} = 4.75$, thus $f_{obs} > F_{0.05,1,12}$, hence regect $H_0$.

iii. Similarly, main effect of B can be written as

$$SS(B) = (C\hat{\beta})^T(C(X^TX)^{-1}C^T)^{-1}(C\hat{\beta}),$$

where $C = (0, 0, 1)$. And

$$SS(B) = (-2.808458)^2 * (\frac{56}{201})^{-1} = 28.31026.$$

Thus value of test statistic is

$$f_2 = \frac{SS(B)/1}{\tilde{\sigma^2}} = 16.095233,$$

note that critical value is $F_{0.05,1,12} = 4.75$, thus $f_{obs} > F_{0.05,1,12}$, hence regect $H_0$.

**Notice:** For unbalanced design,

$$SS.total \neq SSE + SS(A) + SS(B) + SS(AB), \quad (*)$$

since the cross-product terms are not equal to 0. You can obtain

$$SS(AB) = (\hat{\mu}_{11} + \hat{\mu}_{00} - \hat{\mu}_{01} - \hat{\mu}_{10})(\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{1}{n_{ij}})^{-1} = 10.2409,$$

from results of part (b). Thus, it is easy to verify conclusion of $(*)$.

(d) In this case, the point test is $ptest = \hat{\mu}_{0,0} - \hat{\mu}_{0,1} = 4.575$ and $Var(\hat{\mu}_{0,0} - \hat{\mu}_{0,1}) = \hat{\sigma}^2(\frac{1}{4} + \frac{1}{3}) = 0.5762361$, thus test statistic is (since there is interaction, so estimaiton of $\sigma^2$ should be 0.9878333),

$$t_{obs} = \frac{ptest}{\sqrt{Var(\hat{\mu}_{0,0} - \hat{\mu}_{0,1})}} = 6.026857,$$

note that critical value is $t_{0.025,11} = 2.201$, thus $t_{obs} > t_{0.025,11}$, reject $H_0$.

3. In addition to ten men (details are given in Q2 of Assignment 2), eleven women were also studied during a maximal exercise treadmill test. Based on the observations from women, we obtain the following table of parameter estimates and standard error.

| Variable | Parameter Estimate $\hat{\beta}_i$ | Standard Error |
|---|---|---|
| Intercept | -51.9625 | 26.5128 |
| $x_1$ | -0.4168 | 0.2014 |
| $x_2$ | 0.4415 | 0.09881 |
| $x_3$ | 0.3629 | 0.1599 |

The $\hat{\sigma}^2$ for women is equal to 16.7327.

(e) Assuming that the population variances of $y$ for men and women are equal, estimate the common variance.

Solution:
$RSS_2 = \sigma_2^2(n_2 - 4) = 16.7327 \times 7 = 117.1289$

$$\hat{\sigma}^2 = \frac{RSS_1 + RSS_2}{n_1 + n_2 - p'} = \frac{55.9687 + 117.1289}{10 + 11 - 8} = \mathbf{13.3152}$$

(f) Test whether the regression coefficients of $y$ on $x_2$ for men and women are equal at the 5% significance level. Write down your estimate, standard error, test statistic, critical value and your conclusion clearly.

Solution:
$H_0$: $\beta_{21} = \beta_{22}$

$$Var(\hat{\beta}_{21}) = \sigma^2\frac{Var(\hat{\beta}_1)}{\sigma_1^2},\ Var(\hat{\beta}_{22}) = \sigma^2\frac{Var(\hat{\beta}_2)}{\sigma_2^2}$$

$$\Rightarrow \quad Var(\hat{\beta}_{21} - \hat{\beta}_{22}) = \sigma^2(\frac{Var(\hat{\beta}_1)}{\sigma_1^2} + \frac{Var(\hat{\beta}_1)}{\sigma_2^2})$$

$$s.e.(\hat{\beta}_{21} - \hat{\beta}_{22}) = \hat{\sigma}\sqrt{\frac{s.e.(\hat{\beta}_{12})^2}{\hat{\sigma}_1^2} + \frac{s.e.(\hat{\beta}_{22})^2}{\hat{\sigma}_2^2}}$$

$$|t| = \left| \frac{0.2895 - 0.4413}{\hat{\sigma}\sqrt{\frac{0.0781^2}{\hat{\sigma}_1^2} + \frac{0.09881^2}{\hat{\sigma}_2^2}}} \right| = \left| \frac{0.1518}{\sqrt{0.01648}} \right| = \left| \frac{0.1518}{0.1284} \right| = \mathbf{1.1822} < 2.160 = t_{0.025,13}$$

$\Rightarrow$ can't reject $H_0$