1. In a random sample of \( n = 30 \) people drawn from a certain population of adults, systolic blood pressure (\( y \), in \( \text{mm Hg} \)) and age (\( x \), in years) are measured with the following results:

\[
\bar{x} = 45.1, \ s_x = 15.3 \quad \bar{y} = 142.5, \ s_y = 22.6 \quad s_{xy} = 227.1 \quad \text{(sample covariance)}
\]

(a) Report the sample correlation coefficient between \( y \) and \( x \).

\[
r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{227.1}{15.3 \times 22.6} = 0.66
\]

(b) Report the least squares estimate of the slope, \( \hat{\beta} \), in a regression of \( y \) on \( x \).

\[
\hat{\beta}_1 = r_{xy} s_y / s_x = 0.66 \times 22.6 / 15.3 = 0.97
\]

(c) Report the least squares estimate of the intercept in a regression of \( y \) on \( x \).

\[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 98.8
\]

(d) Estimate the mean and standard deviation of blood pressures among people in

i. the population who are aged \( x = 30 \)

\[
\hat{\mu}(x = 30) = \hat{\beta}_0 + \hat{\beta}_1(30) = 98.8 + 0.97(30) = 127.9
\]

\[
\hat{\sigma} = \sqrt{MS(E)} = \sqrt{SS[Tot](1 - r^2)/(n - 2)} = \sqrt{22.6^2 \frac{29}{28} (1 - .66^2)} = 17.3
\]

ii. the population who are aged \( x = 40 \)

\[
\hat{\mu}(x = 40) = \hat{\beta}_0 + \hat{\beta}_1(30) = 98.8 + 0.97(40) = 137.6, \ \hat{\sigma} = \sqrt{MS(E)} = 17.3
\]

(e) Report a 95\% confidence interval for the difference in the two means estimated in part (d).

\[
10\hat{\beta}_1 \pm t(.025, n - 2)\sqrt{\text{Var}(10\hat{\beta}_1)}
\]

or

\[
9.7 \pm 2.0510 \sqrt{MS(E)/S_{xx}} \quad \text{or} \quad 9.7 \pm 2.05(2.13)
\]
2. Four multiple linear regression models involving two predictor variables, $x_1$ and $x_2$ are considered for a sample of $n = 15$ observations. The error mean square obtained by fitting each model is given below:

$$M_0 : \mu(x_1, x_2) = \beta_0 \quad MS(E) = 3.652 = s^2_y \quad (df = 14)$$

$$M_1 : \mu(x_1, x_2) = \beta_0 + \beta_1 x_1 \quad MS(E) = 1.138 \quad (df = 13)$$

$$M_2 : \mu(x_1, x_2) = \beta_0 + \beta_2 x_2 \quad MS(E) = 1.343 \quad (df = 13)$$

$$M_3 : \mu(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad MS(E) = 0.392 \quad (df = 12)$$

$$M_4 : \mu(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \quad MS(E) = 0.394 \quad (df = 11)$$

The estimated regression equations under $M_3$ and $M_4$ are

$$M_3 : \hat{\mu}(x_1, x_2) = 2.59 - .38 x_1 + .87 x_2$$

$$M_4 : \hat{\mu}(x_1, x_2) = 2.46 - .19 x_1 + 1.27 x_2 - .61 x_1 x_2$$

(a) Consider the case where $x_2 = 0.3$, report an estimate of the change in the mean response when $x_1$ is increased by 1 unit

i. Under $M_3 \hat{\beta}_1 = -.38$

ii. Under $M_4 \hat{\beta}_1 + 0.3\hat{\beta}_3 = -.19 - .3(.61) = -.37$

(b) Consider $F$-tests for comparing nested models. Report the $F$-ratio for a test of

i. $H_0 : \beta_1 = 0$ in $M_1$

ii. $H_0 : \beta_2 = 0$ when comparing models $M_1$ and $M_3$.

iii. $H_0 : \beta_1 = \beta_2 = 0$ when comparing models $M_0$ and $M_3$.

iv. $H_0 :$ the dependence of the mean on $x_1$ is constant across all levels of $x_2$.

$$F_i = \frac{SS[Tot] - SS(E)}{MS(E)_{M_i}} = \frac{3.652(14) - 1.138(13)}{1.138}$$

$$F_{ii} = \frac{1.138(13) - .392(12)}{.392}$$

$$F_{iii} = \frac{(3.652(14) - .392(12))/2}{.392}$$

$$F_{iv} = \frac{.392(12) - .394(11)}{.394}$$
3. In a simple linear regression of a response variable \( y \) on a predictor variable \( x \), the sample mean and variance of \( y \) are observed to be \( \bar{y} = 10 \) and \( s_y = 2 \), and the observed sample correlation coefficient with \( x \) is \( r_{xy} = -0.5 \).

(a) Report the estimated mean of the response at the average of the predictor variable, \( x = \bar{x} \).

\[
\hat{\mu}(x = \bar{x}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} = \bar{y} = 10
\]

(b) Report the estimated mean of the response when \( x \) is one sample standard deviation above its the mean, \( x = \bar{x} + s_x \).

\[
\hat{\mu}(x = \bar{x} + s_x) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} + \hat{\beta}_1 s_x = \bar{y} + \hat{\beta}_1 s_x = \bar{y} + r_{xy} \frac{s_y}{s_x} s_x = \bar{y} + r_{xy} s_y = 10 - 0.5(2) = 9
\]

4. A food scientist studies the way the texture of chicken depends on the way it is cooked when using a microwave oven. He fits a model in which the response variable, \( \text{stress} \), depends on the \( \text{power} \) settings and duration (\( \text{ctime} \)) of the microwave, along with an elasticity measurement called \( \text{gprime} \). A multiple linear regression function of the form

\[
\mu(\text{power,ctime,gprime}) = \beta_0 + \beta_1 \text{power} + \beta_2 \text{ctime} + \beta_3 \text{gprime}
\]

was fit using PROC GLM from the SAS software, with output given on the next page.

(a) How many columns does the design matrix \( X \) have? \( \dim(X) = (27 \times 4) \), so \( X \) has \( p + 1 = 4 \) columns.

(b) How many rows does the design matrix \( X \) have? \( n = 27 \) columns

(c) One element of the design matrix has been hidden by “AAAAAA”. Report this value. \( AAA = -1.97 \) by symmetry of \( (X'X)^{-1} \)

(d) Report a \( t \)-statistic for a test that the partial slope for \( \text{power} \) is 0 \( (H_0 : \beta_1 = 0) \).

\[
t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{-1.46}{\hat{MS}(E)(X'X)^{-1}1} = \frac{-1.46}{.038(5.15)} = \frac{-1.46}{.44} = -3.3
\]

(e) Report these extra regression sums of squares:

i. \( R(\beta_2 | \beta_0, \beta_1) \) .053, from output

ii. \( R(\beta_2 | \beta_0, \beta_1, \beta_3) \) .00159, from output

iii. \( R(\beta_1 | \beta_0, \beta_2, \beta_3) \)

\[
R(\beta_1 | \beta_0, \beta_2, \beta_3) = F_{(\beta_1=0)} MS(E) = t^2_{(\beta_1=0)} MS(E) = (-3.3)^2 .038 = 0.41
\]
/* begin SAS code */
proc glm data=microwave;
   model stress=power ctime gprime/solution inverse;
run;
/* end SAS code */

The SAS System
The GLM Procedure

X'X Inverse Matrix

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>power</th>
<th>ctime</th>
<th>gprime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.2019129125</td>
<td>AAAAAAAAAAAAA</td>
<td>-0.010542606</td>
<td>0.0275943835</td>
</tr>
<tr>
<td>power</td>
<td>-1.965134572</td>
<td>5.150361951</td>
<td>0.0118771473</td>
<td>-0.104536448</td>
</tr>
<tr>
<td>ctime</td>
<td>-0.010542606</td>
<td>0.1119771473</td>
<td>0.0092844166</td>
<td>-0.000330082</td>
</tr>
<tr>
<td>gprime</td>
<td>0.0275943835</td>
<td>-0.104536448</td>
<td>-0.000330082</td>
<td>0.00290521</td>
</tr>
</tbody>
</table>

Sum of Source DF Squares Mean Square F Value Pr > F
Model 3 4.43544056 1.47848019 39.25 <.0001
Error 23 0.86631891 0.03766604
Corrected Total 26 5.30175947

R-Square Coeff Var Root MSE strain Mean
0.836598 11.24161 0.194077 1.726420

Source DF Type I SS Mean Square F Value Pr > F
power 1 4.14880015 4.14880015 110.15 <.0001
ctime 1 0.05335556 0.05335556 1.42 0.2461
gprime 1 0.23328485 0.23328485 6.19 0.0205

Source DF Type III SS Mean Square F Value Pr > F
power 1 ------------------ (hidden) -------------------
ctime 1 0.00158671 0.00158671 0.04 0.8392
gprime 1 0.23328485 0.23328485 6.19 0.0205

Standard Parameter Estimate Error t Value Pr > |t|
Intercept 3.268360570 0.21277053 15.36 <.0001
power -1.463716111 --------- (hidden) ---------
c-time -0.000671778 0.00327305 -0.21 0.8392
gprime -0.026033468 0.01046077 -2.49 0.0205