1. Consider the 928 bivariate measurements of height ($y$) and midparent height ($x$).

(a) Use SAS to compute $\bar{y}$. Obtain a 95% confidence interval for the population mean height, or the expected height of a randomly sampled person from the population, $\mu = E(Y)$.

(b) Report the following summary statistics:
   i. $\bar{y} = 1/n \sum y_i = 1/n(y_1 + y_2 + \cdots + y_n)$
   ii. $\bar{x} = 1/n \sum x_i = 1/n(x_1 + x_2 + \cdots + x_n)$
   iii. $\sum(x_i - \bar{x})(y_i - \bar{y})$
   iv. $\sum(x_i - \bar{x})y_i$
   v. $\sum x_i(y_i - \bar{y})$
   vi. $s_{xx} = \sum(x_i - \bar{x})^2$
   vii. $s_x^2 = \frac{1}{n-1} \sum(x_i - \bar{x})^2$
   viii. $s_{yy} = \sum(y_i - \bar{y})^2$
   ix. $s_y^2 = \frac{1}{n-1} \sum(y_i - \bar{y})^2$
   x. $r$

(c) Express the slope from the least squares regression line as a function of $r$, $s_x$ and $s_y$ above.

(d) Obtain a 95% confidence interval for the population mean height among males with average midparent height ($x = \bar{x}$). Compare with part (a).

(e) Obtain a 95% confidence interval for the population mean height among males with midparent height ($x = \bar{x} + s_x$).
2. These are $n = 15$ bivariate measurements on striped ground crickets in
the file (“crickets.dat”). Treat chirp frequency (“CPS”) as the response
variable, and temperature (Temp) as the predictor.

(a) Obtain a scatterplot of these msmts (sketched or use software).

(b) Specify the simple linear regression model for these data. Identify all
parameters in the model, providing the interpretation of each.

(c) Explain how the interpretation (and the estimate) of the slope pa-
parameter changes if temperature is expressed in Celcious.

(d) Estimate mean chirp frequency among crickets when the temp. is
$T = 80^\circ F$. Estimate the standard deviation among chirp frequency
measurements made at this fixed temperature.

(e) Estimate the mean chirp frequency among crickets when $T = 105^\circ F$.
Is there a problem with this estimate?

(f) Report the sum of squared deviations between the fitted values and
the average chirp frequency, $\bar{y}$?

(g) What proportion of variance in chirp frequencies is explained by the
linear regression model?

3. Do the exercises on page 15 of the lecture notes:

(a) Examine the butterfat and temperature data plotted on the next
page. Is there evidence of linear association? The sample correlation
coefficient is $r = -0.45$ based on randomly sampled days. Carry out
an appropriate test. Obtain a 95% confidence interval for the popula-
tion correlation coefficient describing the linear association between
% butterfat and temperature.

(b) Suppose that two variables $X$ and $Y$ have correlation $\rho = 0.6$. What
is the probability that a random sample of $n = 30$ observations from
this bivariate population will yield a sample correlation coefficient of
0.7 or higher?

Hint for (a):

```
proc corr data=butterfat FISHER;
   var butterfat temp;
run;
```

4. Rao 11.3bc. In the regression equation:

$$E(Y|a, b, c, d, f) = \beta_0 + \beta_1 a + \beta_2 b + \beta_3 c + \beta_4 d + \beta_5 f$$

(b) Interpret $\beta_1$ and $\beta_2$

(c) Interpret the quantity $\beta_1 + \beta_2 + \beta_3$
5. 11.11 abdeh (p. 494-495, Rao text needed, data available on website as “ex11.8.dat”, SAS code also available as “rao-ex11-8.sas”)

---

The SAS System 1

Model: MODEL1

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>4.40039</td>
<td>2.20020</td>
<td>34.56</td>
<td>0.0012</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>0.31836</td>
<td>0.06367</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>7</td>
<td>4.71875</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Standard

| Variable  | DF | Estimate  | Error  | t Value | Pr > |t| |
|-----------|----|-----------|--------|---------|------|---|
| Intercept | 1  | -0.86114  | 0.50012| -1.72   | 0.1457|
| x1        | 1  | 0.80007   | 0.09658| 8.28    | 0.0004|
| x2        | 1  | -0.07878  | 0.03715| -2.12   | 0.0874|

Model: MODEL2

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>4.40325</td>
<td>1.46775</td>
<td>18.61</td>
<td>0.0082</td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>0.31550</td>
<td>0.07887</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>7</td>
<td>4.71875</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Standard

| Variable  | DF | Estimate  | Error  | t Value | Pr > |t| |
|-----------|----|-----------|--------|---------|------|---|
| Intercept | 1  | -0.59775  | 1.49078| -0.40   | 0.7089|
| x1        | 1  | 0.74355   | 0.31561| 2.36    | 0.0780|
| x2        | 1  | -0.12610  | 0.25186| -0.50   | 0.6429|
| x1x2      | 1  | 0.00998   | 0.05238| 0.19    | 0.8582|

(a) Argue that while \( \beta_0 \) is the same in both models, the meanings of \( \beta_1 \) and \( \beta_2 \) are different.

(b) Determine the least squares prediction equation using model 1.

(d) Test \( H_0 : \beta_1 = \beta_2 = 0 \) at level \( \alpha = 0.05 \) in model 1. Give a brief conclusion.

(e) Determine the least squares prediction equation using model 2.

(h) Obtain estimates of the error variance based on the two models. Do the estimates lead you to prefer one model over the other?

(i) Test the hypothesis that the linear dependence of \( Y \) on \( x_1 \) does not depend on \( x_2 \), using level of significance \( \alpha = .05 \).