Solutions to Pollard sheet 10

The regression equation is **distance = 21.0 + 0.196 height + 0.191 weight**

S = 3.943 R-Sq = 80.5%

Source	DF	SS	MS	F	P
Regression	2	578.82	289.41	18.62	0.001
Residual Error	9	139.91	15.55		
Total	11	718.73			

Source	DF	Seq SS
height	1	558.06
weight	1	20.75

Note that $1-R^2$ = Residual Error / Total = 139.91/718.73

i. The regression equation is **distance = 12.1 + 0.597 height**

S = 4.008 R-Sq = 77.6%

Source	DF	SS	MS	F	P
Regression	1	558.06	558.06	34.73	0.000
Residual Error	10	160.67	16.07		
Total	11	718.73			

Note that $1-R^2$ = Residual Error / Total = 160.67/718.73

ii. The regression equation is **weight = - 46.6 + 2.10 height**

S = 7.549 R-Sq = 92.4%

Source	DF	SS	MS	F	P
Regression	1	6899.2	6899.2	121.06	0.000
Residual Error	10	569.9	57.0		
Total	11	7469.1			

iii. The regression equation is **f = - 0.00 + 0.191 g**

S = 3.740 R-Sq = 12.9%

Source	DF	SS	MS	F	P
Regression	1	20.75	20.75	1.48	0.251
Residual Error	10	139.91	13.99		
Total	11	160.67			

Note that $1-R^2$ = Residual Error / Total = 139.91/160.67

iv. The regression line in (iii) must pass through the point with coordinates $\overline{g}$ and $\overline{f}$. Both means are zero, because the sum of the residuals from a regression (with an intercept term) is zero. The least squares line in (iii) must have a zero intercept.

v. Look at them, or perhaps plot them against each other and observe the straight line.

vi. Parts (i), (ii), and (iii) give

```
distance = 12.1 + 0.597 height + f
weight = - 46.6 + 2.10 height + g
f =  0.191 g + e
```

Thus

```
distance = 12.1 + 0.597 height + (0.191 g + e)
         = 12.1 + 0.597 height + 0.191 (weight + 46.6 - 2.10 height) + e
         = (12.1 + 0.191 × 46.6)
            + (0.597 - 0.191 × 2.10 )height
            + 0.191 weight + e
```

Compare with the fit with both predictors,

```
distance = 21.0 + 0.196 height + 0.191 weight + e
```

Note that

```
12.1 + 0.191 × 46.6 = 21.0
0.597 - 0.191 × 2.10 = 0.196
```

vii. From the regression with two predictors,

$1-R^2$ = Residual Error / Total = 139.91/718.73

From (i)

$1-R^2$ = Residual Error / Total = 160.67/718.73

From (iii)

$1-R^2$ = Residual Error / Total = 139.91/160.67

Clearly, to get the $1-R^2$ for the two predictors we take the product of the $1-R^2$ for parts (i) and (iii). That makes sense because the Residual Error for part (i) becomes the Total (sum of squares) for part (iii); the column of residuals **f** plays the role of the y's for part (iii).

viii. The decomposition of the regression sum of squares

```
Source       DF      Seq SS
height        1      558.06
weight        1       20.75
```

for the two predictors corresponds to the fact that the Regression (sum of squares) in part (i) is the component of the variability in distance "accounted for" by the height, leaving the residual sum of squares `160.67,` and that the Regression (sum of squares) in part (iii) is the component of this `160.67` that is subsequently "accounted for" by the weight.