Supplementary exercises 10.17 and 10.26 of IPS7e

Data: Measurement of heart rate (HR) and oxygen uptake (VO2) for one individual under different exercise conditions. Both variables are response variables, but the interest is in predicting VO2 from HR because the heart rate is much easier to measure. 19 measurements were taken.

Model: The statistical model of interest is a linear regression model:

$$\text{VO2}_i = \beta_0 + \beta_1 \text{HR}_i + \varepsilon_i,$$

where the errors $\varepsilon_1, \ldots, \varepsilon_{19}$ are assumed i.i.d. from $\text{N}(0, \sigma)$.

Exercise 10.17
We start with Minitab listings and a fitted line plot.

Regression Analysis: VO2 versus HR

The regression equation is
VO2 = - 2.804 + 0.03865 HR

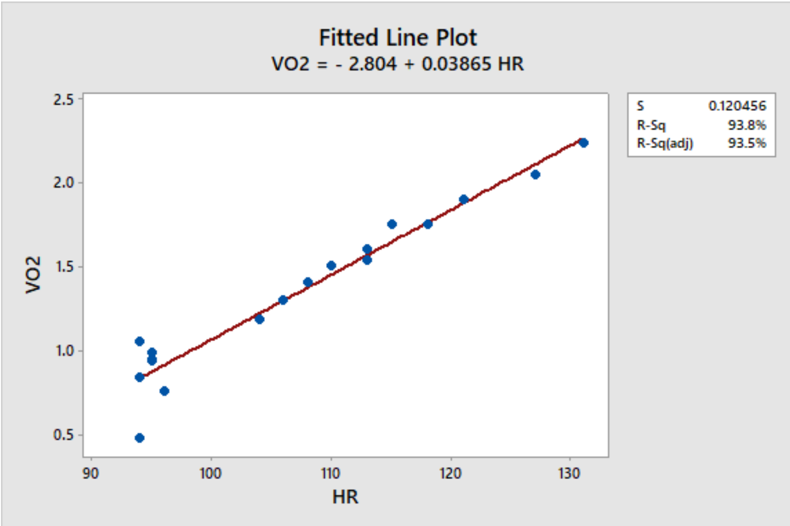
Model Summary

S	R-sq	R-sq(adj)
0.120456	93.85%	93.48%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3.76185	3.76185	259.27	0.000
Error	17	0.24666	0.01451		
Total	18	4.00852			

Fitted Line: VO2 versus HR



(a) All points are pretty close to the line, and none of them seem unusual or outlying. There is a small cluster of points around HR = 95, but this is not a problem with model assumptions and could in this case correspond to the experimentor having deliberately taken several measurements in this range, by setting the exercise conditions the subject was exposed to. There also appears to be a wider spread of the points around the line in this cluster, but this is difficult to assess visually because more points will naturally scatter more than a single point. In summary, no obvious problems with any of the points can be identified.

(b) The estimated regression line is given in the Minitab listing: $\text{VO2} = -2.804 + 0.03865 \cdot \text{HR}$.

(c) The Minitab listing also gives an ANOVA table, which can be used for testing the hypothesis $\text{H}_0 : \beta_1 = 0$, as follows:

$$F = 259.27, \quad \text{df} = (1, 17), \quad P < 0.0005.$$

There is very strong evidence against H_0. The data show, beyond any reasonable doubt, that a (linear) relation between HR and VO2 exists. We could also have tested H_0 by a t-test: the value is $t = 16.10$ (in the listing below), with the same P-value and conclusion.

(d) For confidence and prediction intervals we need to rerun the model in the Regression menu, followed by the Predict menu.

Regression Analysis: VO2 versus HR

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3.76185	3.76185	259.27	0.000
HR	1	3.76185	3.76185	259.27	0.000
Error	17	0.24666	0.01451		
Lack-of-Fit	11	0.07331	0.00666	0.23	0.983
Pure Error	6	0.17335	0.02889		
Total	18	4.00852			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.120456	93.85%	93.48%	92.13%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-2.804	0.258	-10.86	0.000	
HR	0.03865	0.00240	16.10	0.000	1.00

Prediction for VO2

Regression Equation

VO2 = -2.804 + 0.03865 HR

Settings

Variable	Setting
HR	96

Prediction

Fit	SE Fit	95% CI	95% PI
0.906248	0.0382219	(0.825607, 0.986889)	(0.639621, 1.17288)

Settings

Variable	Setting
HR	115

Prediction

Fit	SE Fit	95% CI	95% PI
1.64064	0.0336520	(1.56964, 1.71164)	(1.37677, 1.90451)

Comments:
We will use prediction intervals because we want to assess the plausible range for a new observation, instead of getting a sense of how precisely the line is estimated at these points. From the listing we read off the estimates and intervals for those two values.

$$\begin{aligned} \text{HR} = 96: &\quad \hat{y} = 0.906, \quad 95\% \text{ prediction interval}: \ (0.640, 1.173), \\ \text{HR} = 115: &\quad \hat{y} = 1.641, \quad 95\% \text{ prediction interval}: \ (1.377, 1.905). \end{aligned}$$

(e) It depends on how accurately the researchers want to know VO2. The regression equation predicts only the subject's mean VO2 for a given heart rate, and the prediction intervals show that there is considerable variation about the line so that the actual VO2 is quite variable around that estimate.

Exercise 10.26

(a) The ANOVA table was already shown in the listing above.

(b) The null hypothesis $H_0 : \beta_1 = 0$, that is, a horizontal line, or no linear relation between heart rate and oxygen uptake.

(c) Already above: $F(1, 17)$ and $P < 0.0005$.

(d) From our previous calculation of $t = 16.10$, we get $t^2 = 259.21$, which apart from round-off error equals the $F = 259.27$ of the ANOVA table.

(e) $R^2 = 93.8\%$, a high percentage of the variance is explained.