## Scope

This document shows the solution for the data set in the table below. Your data set will almost certainly be different than this one, but the methods illustrated here are still valid for your data set. To refresh your memory, we begin with the general formulas involved. If you want to skip the formulas, click here.

## Formulas

There are $n$ bivariate observations, $(X_i, Y_i),\ i = 1, \ldots, n$.

### Means

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

### Standard Deviations

$$S_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}, \quad S_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

### Pearson Correlation

$$r = \frac{1}{n-1}\sum_{i=1}^{n} X_i' Y_i',$$

where $X_i'$ and $Y_i'$ are the standardized variates

$$X_i' = \frac{X_i - \overline{X}}{S_X} \text{ and } Y_i' = \frac{Y_i - \overline{Y}}{S_Y}$$

### Least Squares Estimators

$$\hat{\beta}_1 = r\frac{S_Y}{S_X}, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

### Residuals

$$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i),\ i = 1, \ldots, n$$

### Sums of Squares and Mean Squares

SSTO= $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

SSE=$\sum_{i=1}^{n} e_i^2$, MSE=SSE/$(n-2)$

SSR=SSTO−SSE, MSR=SSR

**Standard Errors of Least Squares Estimators**

Define $\hat{\sigma} = \sqrt{\text{MSE}}$. Then the standard errors are

$$\begin{aligned} \hat{\sigma}(\hat{\beta}_0) &= \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}} \\ \hat{\sigma}(\hat{\beta}_1) &= \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}} \end{aligned}$$

**Confidence Interval for $\hat{\beta}_1$**

A level $L$ interval is

$$(\hat{\beta}_1 - \hat{\sigma}(\hat{\beta}_1)t_{n-2,(1+L)/2}, \hat{\beta}_1 + \hat{\sigma}(\hat{\beta}_1)t_{n-2,(1+L)/2})$$

## Data

There are $n = 6$ bivariate observations given in the table below, which also includes their standardized values, the means and standard deviations of both variables and their Pearson correlation.

| Height ($X$) | Pulse ($Y$) | $X'$ | $Y'$ | $X'Y'$ |
|---|---|---|---|---|
| 163 | 68 | −0.5740 | −1.9022 | 1.0918 |
| 165 | 78 | −0.3714 | −0.0312 | 0.0116 |
| 160 | 83 | −0.8779 | 0.9043 | −0.7939 |
| 178 | 79 | 0.9454 | 0.1559 | 0.1474 |
| 184 | 82 | 1.5531 | 0.7172 | 1.1139 |
| 162 | 79 | −0.6753 | 0.1559 | −0.1053 |
| $\overline{X} = 168.\overline{66}$ | $\overline{Y} = 78.\overline{16}$ | | | $r = 0.2931$ |
| $S_X = 9.87$ | $S_Y = 5.34$ | | | |

## SAS Code

The following SAS code will produce all the output needed to answer questions a.)-f.) (and more). The data step reads the data into the SAS data set *pulse*. The SAS procedure *proc reg* performs the regression and outputs results. The corr option produces a correlation matrix. The clb option produces confidence intervals for the betas. The alpha option specifies level 1-alpha (here, $1 - .05 = .95$) intervals.

```
data pulse;
 input height pulse;
datalines;
163 68
165 78
160 83
178 79
184 82
162 79
;
run;
proc reg data=pulse corr;
 model pulse=height/clb alpha=.05;
run;
```

## Solutions

Below, we show how the SAS output can be used to answer questions a.)-f.).

a.) The Pearson correlation, $r = 0.2931$ can be read from the correlation matrix in the SAS output:

Correlation

| Variable | height | pulse |
|---|---|---|
| height | 1.0000 | 0.2931 |
| pulse | 0.2931 | 1.0000 |

b.) The least squares estimates, $\hat{\beta}_0 = 51.40150$ and $\hat{\beta}_1 = 0.15869$ are found in the Parameter Estimates table in the SAS output:

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | 95% Confidence Limits | |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 51.40150 | 43.71324 | 1.18 | 0.3049 | -69.96590 | 172.76891 |
| height | 1 | 0.15869 | 0.25880 | 0.61 | 0.5729 | -0.55986 | 0.87723 |

c.) and d.) The error sum of squares, SSE$= 130.56156$ can be found in the Analysis of Variance table in the SAS output, as can the mean squared error, MSE$= 32.64039$:

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 1 | 12.27177 | 12.27177 | 0.38 | 0.5729 |
| Error | 4 | 130.56156 | 32.64039 | | |
| Corrected Total | 5 | 142.83333 | | | |

e.) The estimated standard errors, $\hat{\sigma}(\hat{\beta}_0) = 43.71324$ and $\hat{\sigma}(\hat{\beta}_1) = 0.25880$ can be found in the Parameter Estimates table shown above.

f.) The 95% confidence interval for $\beta_1$, $(-0.55986, 0.87723)$ can also be found in the Parameter Estimates table shown above.