Exercise: simple linear regression: body weight and plasma volume. Example data contain the body weight (kg) and plasma volume (literes) for eight healthy men.

1 Estimating model coefficients

weight [kg]	plasma [l]	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$	x^2
58.00	2.75						
70.00	2.86						
74.00	3.37						
63.50	2.76						
62.00	2.62						
70.50	3.49						
71.00	3.05						
66.00	3.12						

1. Calculate:

$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i =$

$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i =$

2. Fill in columns 3rd to 6th (leave the last 2 columns for now)

3. Calculate $\hat{\beta}_1$:

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} =$

4. Calculate $\hat{\beta}_0$:
$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} =$

5. Write equation for the best-fitting straight line:

2 Accuracy of the coefficient estimates

1. Fill in the remainig columns in the table above

2. Calculate s $s = \sqrt{[\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2-\overline{\beta_1}\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-2}]} =$

3. Calculate $s.e(\hat{\beta}_0) = s * \sqrt{[\frac{1}{n} + \frac{x_i^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}]} =$

4. Calculate $s.e(\hat{\beta}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}} =$

5. Have a look at Figure 3.3 in *An Introduction to Statistical Learning* and answer questions
   - What do 10 light blue lines represent on the plot (right)?
   - What is an `unbiased estimator`?
   - Have we underestimated or overestimated β_1?

3 Hypothesis testing

Is there an association between body weight and plasma volume?

1. Write down the null hypothesis and alternative hypothesis 2. Calculate t-statistics for $\hat{\beta}_1$

$t = \frac{\hat{\beta}_1 - 0}{s.e.(\hat{\beta}_1)} =$

3. Use t distribution table containign critical values of the t distribution, to check if whether the p-value for our calculated t-statistics is lower than 5% threshold? Is it lower than 1% threshold?.

4. Can we reject the null hypothesis? Is there an association between body weight and plasma volume.