# Notes on Applied Linear Regression

Tianbai Xiao

August 2022

## 1 Linear Regression Model

The general multiple linear regression model with response $Y$ and explanatory variables $X_1, \ldots, X_p$ have the form [1]

$$\begin{aligned} &\mathrm{E}(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \\ &\mathrm{Var}(Y|X) = \sigma^2 \end{aligned} \tag{1}$$

The symbol $X$ in $(Y|X)$ means that we are conditioning on all the terms on the right side of the equation. Both the $\beta$s and $\sigma^2$ are unknown parameters that we need to estimate. Eq.(1) is a linear function of the parameters $\beta$s, which is why this is called linear regression. When $p = 1$, $X$ has only one element, and we get the simple regression problem. When $p = 2$, the mean function in Eq.(1 corresponds to a plane in three dimensions. When $p > 2$, the fitted mean function is a hyperplane, the generalization of a $p$-dimensional plane in a $(p + 1)$-dimensional space.

We can write the multiple linear regression model in matrix notation as

$$\mathbf{Y} = E(Y|X) + \mathbf{e} = \mathbf{X}\beta + \mathbf{e} \tag{2}$$

where the errors $\mathbf{e} = \mathbf{Y} - E(Y|X)$ depend on unknown parameters in the mean function and so are not obervable quantities.

**Example 1.1.** (Straight-line regression) For the straight-line regression model, Eq.(1) becomes

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Here $\mathbf{X}$ is a $n \times 2$ matrix and $\beta$ a $2 \times 1$ vector of parameters.

**Example 1.2.** (Polynomial regression) Suppose that the response is a polynomial function of a single covariate

$$Y_j = \beta_0 + \beta_1 X_j + \cdots + \beta_{p-1} X_j^{p-1} + e_j$$

then we have

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^{p-1} \\ 1 & X_2 & X_2^2 & \cdots & X_2^{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \cdots & X_n^{p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

where $\mathbf{X}$ has dimension $n \times p$. Note that the polynomial regression can be written in form Eq.(2) because of its linearity in $\beta$.

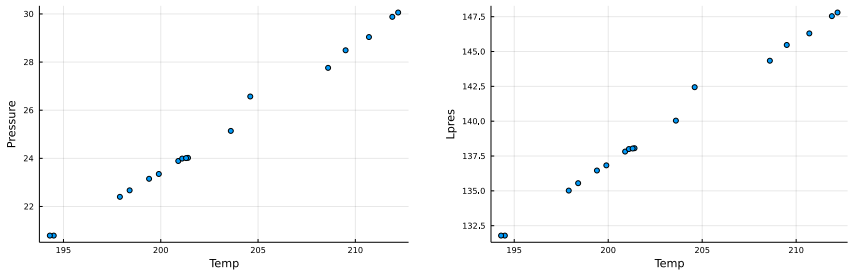
We will start with the straight-line regression problem, and thus the multiple regression model reduces to

$$\begin{aligned} \mathrm{E}(Y|X = x) &= \beta_0 + \beta_1 x \\ \mathrm{Var}(Y|X = x) &= \sigma^2 \end{aligned} \tag{3}$$

We make two important assumptions concerning the errors. First, we assume that $E(e_i|X = x_i) = 0$. The second assumption is that the errors are all independent, meaning that the value of the error for one case gives no information about the value of the error for another case.

## 2 Forbes Data

To illustrate the analysis and computation, we include the real-world Forbes' data in Table 1. It presents the measured response between boiling point and pressure in the Alps and Scotland in 1857. Two sets of pressure data, i.e., the measured pressure in Inch of mercury, and $Lpres = 100 \times \log_{10}(Pressure)$, can be seen in the table.




(a) Pressure-Temp (b) Lpres-Temp

Figure 1: Scatter plots between boiling point and barometric pressure in Forbes' 1857 data.

Table 1: Forbes' 1857 data on boiling point and barometric pressure for 17 locations in the Alps and Scotland.

| Case Number | Temp (°F) | Pressure ( Inches Hg) | Lpres = 100 × log( Pressure ) |
|---|---|---|---|
| 1 | 194.5 | 20.79 | 131.79 |
| 2 | 194.3 | 20.79 | 131.79 |
| 3 | 197.9 | 22.40 | 135.02 |
| 4 | 198.4 | 22.67 | 135.55 |
| 5 | 199.4 | 23.15 | 136.46 |
| 6 | 199.9 | 23.35 | 136.83 |
| 7 | 200.9 | 23.89 | 137.82 |
| 8 | 201.1 | 23.99 | 138.00 |
| 9 | 201.4 | 24.02 | 138.06 |
| 10 | 201.3 | 24.01 | 138.04 |
| 11 | 203.6 | 25.14 | 140.04 |
| 12 | 204.6 | 26.57 | 142.44 |
| 13 | 209.5 | 28.49 | 145.47 |
| 14 | 208.6 | 27.76 | 144.34 |
| 15 | 210.7 | 29.04 | 146.30 |
| 16 | 211.9 | 29.88 | 147.54 |
| 17 | 212.2 | 30.06 | 147.80 |

# 3 Ordinary Least Squares Estimation

In our analysis of these data, the response will be taken to be Lpres = 100 × log 10(Pressure), and the predictor is Temp. Neither multiplication by 100 nor the base of the logarithms has important effects on the analysis. Multiplication by 100 avoids using scientific notation for numbers we display in the text, and changing the base of the logarithms merely multiplies the logarithms by a constant.

The criterion function for obtaining estimators is based on the residuals, which geometrically are the vertical distances between the fitted line and the actual y-values. The residuals reflect the inherent asymmetry in the roles of the response and the predictor in regression problems. The ordinary least squares (OLS) estimators are those values $\beta_0$ and $\beta_1$ that minimize the residual sum of squares

$$RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \tag{4}$$

The least squares estimates can be derived in many ways. One method of finding the minimizer is to differentiate with respect to $\beta_0$ and $\beta_1$, set the

derivatives equal to 0, and solve

$$\frac{\partial RSS\left(\beta_0,\beta_1\right)}{\beta_0} = -2\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1 x_i\right)=0$$
$$\frac{\partial RSS\left(\beta_0,\beta_1\right)}{\beta_1} = -2\sum_{i=1}^{n} x_i\left(y_i-\beta_0-\beta_1 x_i\right)=0$$

Upon rearranging terms, we get

$$\begin{aligned}\beta_0 n+\beta_1\sum x_i &= \sum y_i\\ \beta_0\sum x_i+\beta_1\sum x_i^2 &= \sum x_i y_i\end{aligned} \tag{5}$$

Eq.(5) is called the normal equations for the simple linear regression model in Eq.(3). It can be seen that the normal equations depend on the data only through the sufficient statistics $\sum x_i$, $\sum y_i$, $\sum x_i^2$ and $\sum x_i y_i$. Solving Eq.(5), we can get

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x},\quad \hat{\beta}_1=\frac{SXY}{SXX} \tag{6}$$

The definition of $SXX$, $SXY$ and other statistic quantities can be found in Table 2.

Table 2: Definition of symbols.

| Quantity | Definition | Description |
|---|---|---|
| $\bar{x}$ | $\sum x_i/n$ | Sample average of $x$ |
| $\bar{y}$ | $\sum y_i/n$ | Sample average of $y$ |
| $SXX$ | $\sum\left(x_i-\bar{x}\right)^2=\sum\left(x_i-\bar{x}\right)x_i$ | Sum of squares for the $x$'s |
| $\mathrm{SD}_x^2$ | $SXX/(n-1)$ | Sample variance of the $x$'s |
| $\mathrm{SD}_x$ | $\sqrt{SXX/(n-1)}$ | Sample standard deviation of the $x$'s |
| $SYY$ | $\sum\left(y_i-\bar{y}\right)^2=\sum\left(y_i-\bar{y}\right)y_i$ | Sum of squares for the $y$'s |
| $\mathrm{SD}_y^2$ | $SYY/(n-1)$ | Sample variance of the $y$'s |
| $\mathrm{SD}_y$ | $\sqrt{SYY/(n-1)}$ | Sample standard deviation of the $y$'s |
| $SXY$ | $\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=\sum\left(x_i-\bar{x}\right)y_i$ | Sum of cross-products |
| $s_{xy}$ | $SXY/(n-1)$ | Sample covariance |
| $r_{xy}$ | $S_{xy}/\left(\mathrm{SD}_x\mathrm{SD}_y\right)$ | Sample correlation |

Using Forbes' data, we will write $x$ to be the sample mean of $Temp$ and $y$ to be the sample mean of $Lpres$. The quantities needed for computing the least squares estimators are

$$\bar{x}=202.95294,\quad SXX=530.78235,\quad SXY=475.31224$$
$$\bar{y}=139.60529,\quad SYY=427.79402$$

The quantity $SYY$, although not yet needed, is given for completeness. Using Eq.(6), we find

$$\hat{\beta}_1=0.895,\quad \hat{\beta}_0=-42.138$$

and thus the estimated line is given by

$$\hat{E}(Lpres|Temp) = -42.138 + 0.895Temp$$

Figure 2 shows that the fit of this line to the data is excellent.

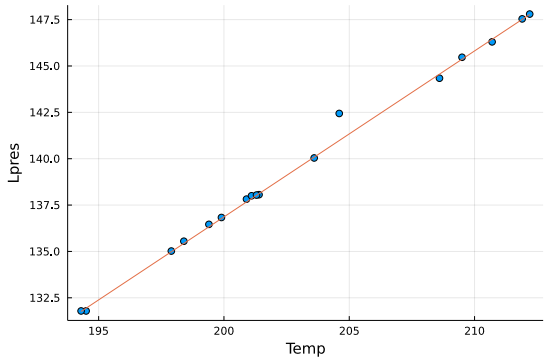



Figure 2: OLS estimation of Forbes' 1857 data.

The above OLS estimation can be performed with the help of statistical softwares. Example codes can be found in the GitHub repository of this lecture [1]

**Example 3.1.** By executing the following command in bash

```
julia forbes.jl
```

we get

```
lpres ~ 1 + bp

Coefficients:
------------------------------------------------------------------------------
             Coef.        Std. Error  t        Pr(>|t|) Lower 95%  Upper 95%
------------------------------------------------------------------------------
(Intercept)  -42.1378     3.3402     -12.62    <1e-08  -49.2572    -35.0183
bp             0.895494   0.0164518   54.43    <1e-17    0.860428    0.93056
------------------------------------------------------------------------------
```

Besides the coefficients $\beta$s, the program also provides some statistics which help evaluate the performance of the linear regression model. We will discuss some of these quantities in the following sections.

[1] `https://github.com/vavrines/ALR`

# 4 Estimating $\sigma^2$

Since the variance $\sigma^2$ is essentially the average squared size of the $e_i^2$, we should expect that its estimator $\hat{\sigma}^2$ is obtained by averaging the squared residuals. Under the assumption that the errors are uncorrelated random variables with zero means and common variance $\sigma^2$, an unbiased estimate of $\sigma^2$ is obtained by dividing residual sum of squares (RSS) by its degrees of freedom (df), where df equals the number of cases minus the number of parameters in the mean function. For simple regression, df $= n - 2$, so the estimate of $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{RSS}{n-2}$$

This quantity is called the residual mean square. The RSS can be computed by Eq.(4). With the help of Eq.(6), it can also be computed via

$$RSS = SYY - \frac{SXY^2}{SXX} = SYY - \hat{\beta}_1^2 SXX \tag{7}$$

which results

$$RSS = 427.79402 - \frac{475.31224^2}{530.78235} = 2.15493$$
$$\sigma^2 = \frac{2.15493}{17-2} = 0.14366$$

The square root of $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{0.14366} = 0.37903$ is often called the standard error of regression. It is in the same units as is the response variable.

# 5 Properties of Least Squares Estimates

The OLS estimates depend on data only through the statistics given in Table 2. This is both an advantage, making computing easy, and a disadvantage, since any two data sets for which these are identical give the same fitted regression, even if a straight-line model is appropriate for one but not the other. The estimates $\beta_0$ and $\beta_1$ can both be written as linear combinations of $y_1, \ldots, y_n$. For example, employing Eq.(6) and writing $c_i = (x_i - x)/SXX$, we have

$$\hat{\beta}_1 = \sum \left( \frac{x_i - \bar{x}}{SXX} \right) y_i = \sum c_i y_i$$

Since we are conditioning on the values of $X$, the $c_i$ are fixed numbers, and

$$\begin{aligned} \mathrm{E}\left(\hat{\beta}_1 | X\right) &= \mathrm{E}\left(\sum c_i y_i | X = x_i\right) = \sum c_i \mathrm{E}\left(y_i | X = x_i\right) \\ &= \sum c_i \left(\beta_0 + \beta_1 x_i\right) \\ &= \beta_0 \sum c_i + \beta_1 \sum c_i x_i \end{aligned}$$

Direct summation $\sum c_i = 0$ and $\sum c_i x_i = 1$ gives

$$\mathrm{E}(\hat{\beta}_1|X) = \beta_1$$

which shows that $\hat{\beta}_1$ is unbiased. A similar computation will show that $\mathrm{E}(\hat{\beta}_0|X = \beta_0)$.

The variance of $\hat{\beta}_1$ can be obtained via

$$\begin{aligned} \mathrm{Var}\left(\hat{\beta}_1|X\right) &= \mathrm{Var}\left(\sum c_i y_i|X = x_i\right) \\ &= \sum c_i^2 \,\mathrm{Var}\left(Y|X = x_i\right) \\ &= \sigma^2 \sum c_i^2 \\ &= \sigma^2/SXX \end{aligned} \tag{8}$$

Computing the variance of $\hat{\beta}_0$ again requires an application of Eq.(6). We write

$$\begin{aligned} \mathrm{Var}\left(\hat{\beta}_0|X\right) &= \mathrm{Var}\left(\bar{y} - \hat{\beta}_1\bar{x}|X\right) \\ &= \mathrm{Var}(\bar{y}|X) + \bar{x}^2\,\mathrm{Var}\left(\hat{\beta}_1|X\right) - 2\bar{x}\,\mathrm{Cov}\left(\bar{y}, \hat{\beta}_1|X\right) \end{aligned}$$

To complete this computation, we need to compute the covariance

$$\begin{aligned} \mathrm{Cov}\left(\bar{y}, \hat{\beta}_1|X\right) &= \mathrm{Cov}\left(\frac{1}{n}\sum y_i, \sum c_i y_i\right) \\ &= \frac{1}{n}\sum c_i \,\mathrm{Cov}\left(y_i, y_i\right) \\ &= \frac{\sigma^2}{n}\sum c_i \\ &= 0 \end{aligned}$$

because $y_i$s are independent and $\sum c_i = 0$. Therefore we have

$$\mathrm{Var}\left(\hat{\beta}_0|X\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right) \tag{9}$$

Following Eq.(8) and (9), the Forbes' data gives

$$\begin{aligned} &\mathrm{Var}(\hat{\beta}_1) = 0.14366/530.78235 = 0.00027 \\ &\mathrm{Var}(\hat{\beta}_0) = 0.14366 * (1/17 + 202.95294^2/530.78235) = 11.15679 \end{aligned}$$

The square root of an estimated variance is called a standard error (Se)

$$\mathrm{Se}(\hat{\beta}_1) = 0.01645, \quad \mathrm{Se}(\hat{\beta}_0) = 3.34018$$

which is consistent with the results in Example 3.1.

# 6 Analysis of Variance

The analysis of variance provides a convenient method of comparing the fit of two or more mean functions for the same set of data. The methodology developed here is very useful in multiple regression and, with minor modification, in most regression problems.

An elementary alternative to the simple regression model suggests fitting the mean function

$$\mathrm{E}(Y|X = x) = \beta_0 \tag{10}$$

The mean function Eq.(10) is the same for all values of X. Fitting with this mean function is equivalent to finding the best line parallel to the horizontal or $x$-axis. The OLS estimate of the mean function is $\mathrm{E}(Y|X) = \hat{\beta}_0$, where $\hat{\beta}_0$ is the value of $\beta_0$ that minimizes $\sum(y_i - \beta_0)^2$. The minimizer is given by

$$\hat{\beta}_0 = \bar{y}$$

The residual sum of squares is

$$\sum \left(y_i - \hat{\beta}_0\right)^2 = \sum (y_i - \bar{y})^2 = SYY \tag{11}$$

This residual sum of squares has $n - 1$ df, $n$ cases minus one parameter in the mean function.

From Eq.(7) and (11), we learn that the reduction in residual sum of squares due to enlarging the mean function from Eq.(10) to the simple regression mean function in Eq.(3). This so-called sum of squares due to regression, SSreg, is defined by

$$\begin{aligned} SSreg &= SYY - RSS \\ &= SYY - \left(SYY - \frac{(SXY)^2}{SXX}\right) \\ &= \frac{(SXY)^2}{SXX} \end{aligned} \tag{12}$$

The df associated with $SSreg$ is the difference in df for mean function Eq.(10), $n - 1$, and for Eq.(3), $n - 2$, so the df for $SSreg$ is $(n - 1) - (n - 2) = 1$ for simple regression. These results are often summarized in an analysis of variance table, abbreviated as ANOVA, given in Table 6.3. The column marked "Source" refers to descriptive labels given to the sums of squares. The df column gives the number of degrees of freedom associated with each named source. The next column gives the associated sum of squares. The mean square column is computed from the sum of squares column by dividing sums of squares by the corresponding df. The mean square on the residual line is just $\hat{\sigma}^2$, as already discussed.

The ANOVA is always computed relative to a specific larger mean function, here given by Eq.(3), and a smaller mean function obtained from the larger by setting some parameters to zero, or occasionally setting them to some other known value, here given by Eq.(10).

Table 3: The analysis of variance table for simple regression.

| Source | df | SS | MS | F | $p$-value |
|---|---|---|---|---|---|
| Regression | 1 | $SSreg = 425.6$ | $SSreg/\text{df} = 425.6$ | $MSreg/\hat{\sigma}^2 = 2962.8$ | $\simeq 0$ |
| Residual | $n-2=15$ | $RSS = 2.155$ | $RSS/\text{df} = 0.144$ | | |
| Total | $n-1=16$ | $SYY = 427.8$ | | | |

## 6.1 $F$-test for regression

If the sum of squares for regression $SSreg$ is large, then the simple regression mean function $\mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$ should be a significant improvement over the mean function given by Eq.(10). This is equivalent to saying that the additional parameter in the simple regression mean function $\beta_1$ is different from zero or that $\mathrm{E}(Y|X=x)$ is not constant as $X$ varies. To formalize this notion, we need to be able to judge how large is "large." This is done by comparing the regression mean square, $SSreg$ divided by its df (here equals 1), to the residual mean square $\hat{\sigma}^2$. We call this ratio $F$:

$$F = \frac{(SYY - RSS)/1}{\hat{\sigma}^2} = \frac{SSreg/1}{\hat{\sigma}^2} \tag{13}$$

It is clear that $F$ is just a rescaled version of $SSreg$, with larger values of SSreg resulting in larger values of F. Formally, we can consider testing the null hypothesis (NH) against the alternative hypothesis (AH)

$$\begin{array}{ll} \text{NH}: & \mathrm{E}(Y|X=x) = \beta_0 \\ \text{AH}: & \mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x \end{array} \tag{14}$$

If the errors are $\mathrm{NID}(0, \sigma^2)$ or the sample size is large enough, then under NH Eq.(13) will follow an $F$-distribution with df associated with the numerator and denominator, 1 and $n-2$, for simple regression [2]. This is written as $F \sim F(1, n-2)$. For Forbes' data, we then have

$$F = \frac{425.639}{0.144} = 2963$$

We obtain a significance level or $p$-value for this test by comparing $F$ to the percentage points of the $F(1, n-2)$-distribution. The p-value is shown as "approximately zero," meaning that, if the NH were true, the change of F exceeding its observed value is essentially zero. This is very strong evidence against NH and in favor of AH.

**Remark.** The $F$-distribution with $d_1$ and $d_2$ degrees of freedom is the distribution of

$$X = \frac{S_1/d_1}{S_2/d_2}$$

where $S_1$ and $S_2$ are independent random variables with chi-square distributions with respective degrees of freedom $d_1$ and $d_2$. It can be shown to follow that

the probability density function (pdf) for $X$ is given by

$$f\left(x; d_1, d_2\right) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x\ \mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} = \frac{1}{\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{d_1/2} x^{d_1/2-1} \left(1 + \frac{d_1}{d_2} x\right)^{-(d_1+d_2)/2}$$

for real $x > 0$. Here B is the beta function.

## 6.2 Interpreting $p$-values

Under the appropriate assumptions, the $p$-value is the conditional probability of observing a value of the computed statistic, here the value of $F$, as extreme or more extreme, here as large or larger, than the observed value, given that the NH is true. A small $p$-value provides evidence against the NH.

In some research areas, it has become traditional to adopt a fixed significance level when examining $p$-values. For example, if a fixed significance level of $\alpha$ is adopted, then we would say that an NH is rejected at level $\alpha$ if the p-value is less than $\alpha$. The most common choice for $\alpha$ is 0.05, which would mean that, were the NH to be true, we would incorrectly find evidence against it about 5% of the time, or about 1 test in 20. Accept–reject rules like this are generally unnecessary for reasonable scientific inquiry. Simply reporting $p$-values and allowing readers to decide on significance seems a better approach.

There is an important distinction between statistical significance, the observation of a sufficiently small $p$-value, and scientific significance, observing an effect of sufficient magnitude to be meaningful. Judgment of the latter usually will require examination of more than just the $p$-value.

## 6.3 Power of tests

When the NH is true, and all assumptions are met, the chance of incorrectly declaring an NH to be false at level $\alpha$ is just $\alpha$. If $\alpha = 0.05$, then in 5% of tests where the NH is true we will get a $p$-value smaller than or equal to 0.05.

When the NH is false, we expect to see small $p$-values more often. The power of a test is defined to be the probability of detecting a false NH. For the hypothesis test Eq.(14), when the NH is false, it can be shown that the statistic $F$ given by (2.19) has a noncentral $F$ distribution, with 1 and $n-2$ df, and with noncentrality parameter given by $SXX\beta_1^2/\sigma^2$ [3]. The larger the value of the non centrality parameter, the greater the power. The noncentrality is increased if $\beta_1^2$ is large, if $SXX$ is large, either by spreading out the predictors or by increasing the sample size, or by decreasing $\sigma^2$.

# Problems

1. Show that we can get Eq.(6) by solving Eq.(5).

2. **Height and weight data** The table below gives $Ht$ = height in centimeters and $Wt$ = weight in kilograms for a sample of $n = 10$ 18-year-old girls. Interest is in predicting weight from height.

| $Ht$ | $Wt$ |
|---|---|
| 169.6 | 71.2 |
| 166.8 | 58.2 |
| 157.1 | 56.0 |
| 181.1 | 64.5 |
| 158.4 | 53.0 |
| 165.6 | 52.4 |
| 166.7 | 56.8 |
| 156.5 | 49.2 |
| 168.1 | 55.6 |
| 165.3 | 77.8 |

(1) Draw a scatterplot of $Wt$ on the vertical axis versus Ht on the horizontal axis. On the basis of this plot, does a simple linear regression model make sense for these data? Why or why not?

(2). Compute $\bar{x}$, $\bar{y}$, $SXX$, $SYY$, and $SXY$. Compute estimates of the slope and the intercept for the regression of $Y$ on $X$. Draw the fitted line on your scatterplot.

(3). Obtain the estimate of $\sigma^2$ and find the estimated standard errors of $\beta_0$ and $\beta_1$. Also find the estimated covariance between $\beta_0$ and $\beta_1$. Compute the F-test for the hypotheses that $\beta_1 = 0$ and find the appropriate p-values using two-sided tests.

# References

[1] Sanford Weisberg. *Applied linear regression*, volume 528. John Wiley & Sons, 2005.

[2] Anthony Christopher Davison. *Statistical models*, volume 11. Cambridge university press, 2003.

[3] George AF Seber and Alan J Lee. *Linear regression analysis*. John Wiley & Sons, 2012.