STAT 500

Simple Linear Regression: Model and Estimates

Research Questions

- Study the relationship of two or more quantitative variables.
  - quantitative: number, usually continuous
  - qualitative: classes, identify groups
- Is there a significant linear relationship between the response variable and the explanatory variable?
- What mean value of response would we predict for a given value of the explanatory variable?
- What value of response would we predict for a given value of the explanatory variable?

Simple Linear Regression

$$Y_i = \beta_o + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$$

- $i = 1, \ldots, n$ is the number of observations
- Y_i is the *response* or dependent variable
- X_i is the predictor, *explanatory variable*, or independent variable, treated as known and fixed
- ϵ_i is the *random error* term representing individual variation and measurement error

Linear Model Notation

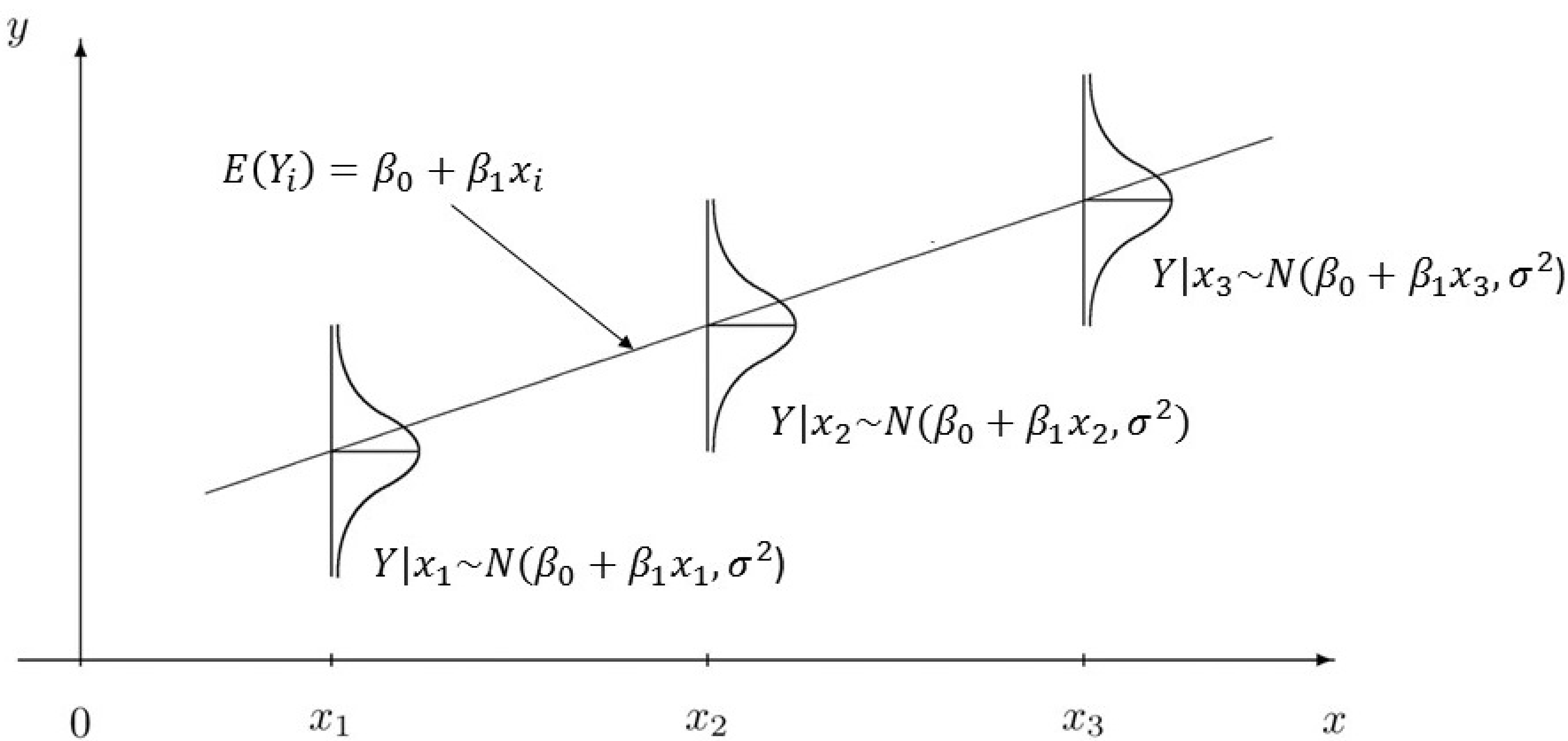
Write SLR model as a linear model $\mathbf{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where

$$\begin{bmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \\ \boldsymbol{Y_3} \\ \vdots \\ \vdots \\ \boldsymbol{Y_n} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \boldsymbol{x_1} \\ \mathbf{1} & \boldsymbol{x_2} \\ \mathbf{1} & \boldsymbol{x_3} \\ \vdots & \vdots \\ \vdots & \vdots \\ \mathbf{1} & \boldsymbol{x_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_0} \\ \boldsymbol{\beta_1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \\ \boldsymbol{\epsilon_3} \\ \vdots \\ \vdots \\ \boldsymbol{\epsilon_n} \end{bmatrix}$$

Model Assumptions

- x's are fixed (or conditioned upon)
- The expected response is a linear function of the explanatory variable : $E(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i$
- additive random errors $Y_i = E(Y_i|X_i = x_i) + \epsilon_i$
- independent (uncorrelated) random errors
- homogeneous error variance: $Var(\epsilon_i) = \sigma^2$
- normally distributed random errors: $\epsilon_i \sim N(0, \sigma^2)$

Model and Assumptions



Model and Assumptions

The conditional distribution of Y given that $X = x$ is

$$N(\beta_o + \beta_1 x, \sigma^2)$$

- $\beta_1 =$ slope, is the change in the conditional mean of Y for a one unit increase in x
- β_o is the conditional mean of Y when $X = 0$
- If we replace x by $x - x_o$ to obtain $Y = \beta_o + \beta_1(x - x_0) + \epsilon$, then β_o is the conditional mean of Y when $X = x_o$
- σ^2 is the variation of responses about the conditional mean for any specific value of the explanatory variable

Relationship to ANOVA

- ANOVA: each group (each level of explanatory variable) has its own mean
- Each x_i in regression defines its own group, but...
  - too many groups with too few observations per group
  - Linear regression analysis makes stronger assumption about the means (linear structure)

A bit of history

Sir Francis Galton coined the term “regression”.

- biometrician, geneticist, 1870-1920s
- compared the heights of children to their parents
- parents and children had similar means
- short parents had short children, tall parents had tall children
- children were closer to average than their parents
- “regression” to the mean

Least Squares Estimation

Use data $(Y_i, x_i), i = 1, 2, \cdots, n$ to estimate the regression coefficients in the model

$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

- Choose estimates b_o and b_1 to minimize
$$g(b_o, b_1) = \sum_{i=1}^{n} [Y_i - (b_o + b_1 x_i)]^2$$
- Why squared errors?
  - Tradition (Gauss invented least squares estimation)
  - Equivalent to maximum likelihood estimation when errors are independent and normally distributed with constant variance

Least Squares Estimation

Results:

$$
\begin{aligned}
b_o &= \bar{Y} - b_1\bar{x} \\
b_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}
$$

- These are best linear unbiased estimators (blue)
- Predicted (fitted) values: $\hat{Y}_i = b_o + b_1 x_i$
- Residuals: $e_i = Y_i - \hat{Y}_i$

Least Squares Estimation

- Choose b_o, b_1 to minimize

$$g(b_o, b_1) = \sum_{i=1}^{n} (Y_i - (b_o + b_1 x_i))^2$$

- Taking derivatives and setting them equal to zero yields the normal equations

$$\begin{aligned} b_o n + b_1 \sum x_i &= \sum Y_i \\ b_o \sum x_i + b_1 \sum x_i^2 &= \sum x_i Y_i \end{aligned}$$

- The normal equations can also be written as

$$\begin{aligned} \sum e_i &= \sum (Y_i - (b_o + b_1 x_i)) = 0 \\ \sum x_i e_i &= \sum x_i (Y_i - (b_o + b_1 x_i)) = 0 \end{aligned}$$

Least Squares Estimation

Normal equations can be written in matrix form

$$\begin{bmatrix} b_o n + b_1 \Sigma x_i \\ b_o \Sigma x_i + b_1 \Sigma x_i^2 \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma x_i Y_i \end{bmatrix}$$

that is equivalent to

$$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma x_i Y_i \end{bmatrix}$$

and can be written as $X^T X \, \mathrm{b} = X^T \mathrm{Y}$

where $\mathrm{b} = \begin{bmatrix} b_o \\ b_1 \end{bmatrix} \quad \mathrm{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathrm{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$

Least Squares Estimation

Solution to the normal equations

$$\begin{bmatrix} \boldsymbol{b_o} \\ \boldsymbol{b_1} \end{bmatrix} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} \bar{\boldsymbol{Y}} - \boldsymbol{b_1}\bar{\boldsymbol{x}} \\ \dfrac{\sum\limits_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{Y_i} - \bar{\boldsymbol{Y}})}{\sum\limits_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2} \end{bmatrix}$$

Least Squares Estimation

Variance-covariance matrix of the least squares estimator:

$$
\begin{aligned}
Var\begin{bmatrix} b_o \\ b_1 \end{bmatrix} &= \begin{bmatrix} Var(b_o) & Cov(b_o, b_1) \\ Cov(b_o, b_1) & Var(b_1) \end{bmatrix} \\
&= Var\left(\left(X^T X\right)^{-1} X^T Y\right) \\
&= \left(X^T X\right)^{-1} X^T Var(Y) \left[\left(X^T X\right)^{-1} X^T\right]^T \\
&= \left(X^T X\right)^{-1} X^T \left[\sigma^2 I\right] X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1}
\end{aligned}
$$

Least Squares Estimation

The the variance-covariance matrix of the least squares estimator for the regression coefficients is

$$
\begin{aligned}
Var\begin{bmatrix} b_o \\ b_1 \end{bmatrix} &= \begin{bmatrix} Var(b_o) & Cov(b_o, b_1) \\ Cov(b_o, b_1) & Var(b_1) \end{bmatrix} \\
&= \sigma^2 \left(X^T X\right)^{-1} \\
&= \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\Sigma_i (x_i - \bar{x})^2} & \frac{-\bar{x}}{\Sigma_{i=1}^n (x_i - \bar{x})^2} \\ \frac{-\bar{x}}{\Sigma_{i=1}^n (x_i - \bar{x})^2} & \frac{1}{\Sigma_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix}
\end{aligned}
$$

Least Squares Estimation

- Matrix of second partial derivatives of $g(b_o, b_1)$

$$\begin{bmatrix} \frac{\partial^2 g(b_o,b_1)}{\partial b_o^2} & \frac{\partial^2 g(b_o,b_1)}{\partial b_o \partial b_1} \\ \frac{\partial^2 g(b_o,b_1)}{\partial b_o \partial b_1} & \frac{\partial^2 g(b_o,b_1)}{\partial b_1^2} \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = X^T X$$

  Since this matrix is positive definite (if we have at least two different x_i values), it guarantees we have a minimum.

- Note: least squares estimate of the slope is different if there is no intercept in the model

Multivariate Normal Distribution

<u>Definition:</u> Suppose $\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_m \end{bmatrix}$ is a random vector whose elements are independently distributed standard normal random variables.
For any $\boldsymbol{n} \times \boldsymbol{m}$ matrix $\boldsymbol{A}$, We say that

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{A}\mathbf{Z}$$

has a *multivariate normal distribution* with mean vector

$$\boldsymbol{E}(\boldsymbol{Y}) = \boldsymbol{E}(\boldsymbol{\mu} + \boldsymbol{A}\mathbf{Z}) = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{E}(\mathbf{Z}) = \boldsymbol{\mu} + \boldsymbol{A}\mathbf{0} = \boldsymbol{\mu}$$

and variance-covariance matrix

$$\boldsymbol{Var}(\mathbf{Y}) = \boldsymbol{A}[\boldsymbol{Var}(\mathbf{Z})]\boldsymbol{A}^{\boldsymbol{T}} = \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} \equiv \boldsymbol{\Sigma}$$

Multivariate Normal Distribution

We will use the notation

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

When $\boldsymbol{\Sigma}$ is positive definite, the joint density function is

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\ e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & Cov(Y_1, Y_3) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & Cov(Y_2, Y_3) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & Cov(Y_n, Y_3) & \cdots & Var(Y_n) \end{bmatrix}$$

Multivariate Normal Distribution

The multivariate normal distribution has some useful properties. One is that normality is preserved under linear transformations:

If $\mathbf{Y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\mathbf{W} = \boldsymbol{c} + \boldsymbol{B}\mathbf{Y} \sim \boldsymbol{N}(\mathbf{c} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^{\boldsymbol{T}})$$

for any non-random $\mathbf{c}$ and $\boldsymbol{B}$.

Forbes Data

Weisberg, Sanford, *Applied Linear Regression*, Wiley, 1980.

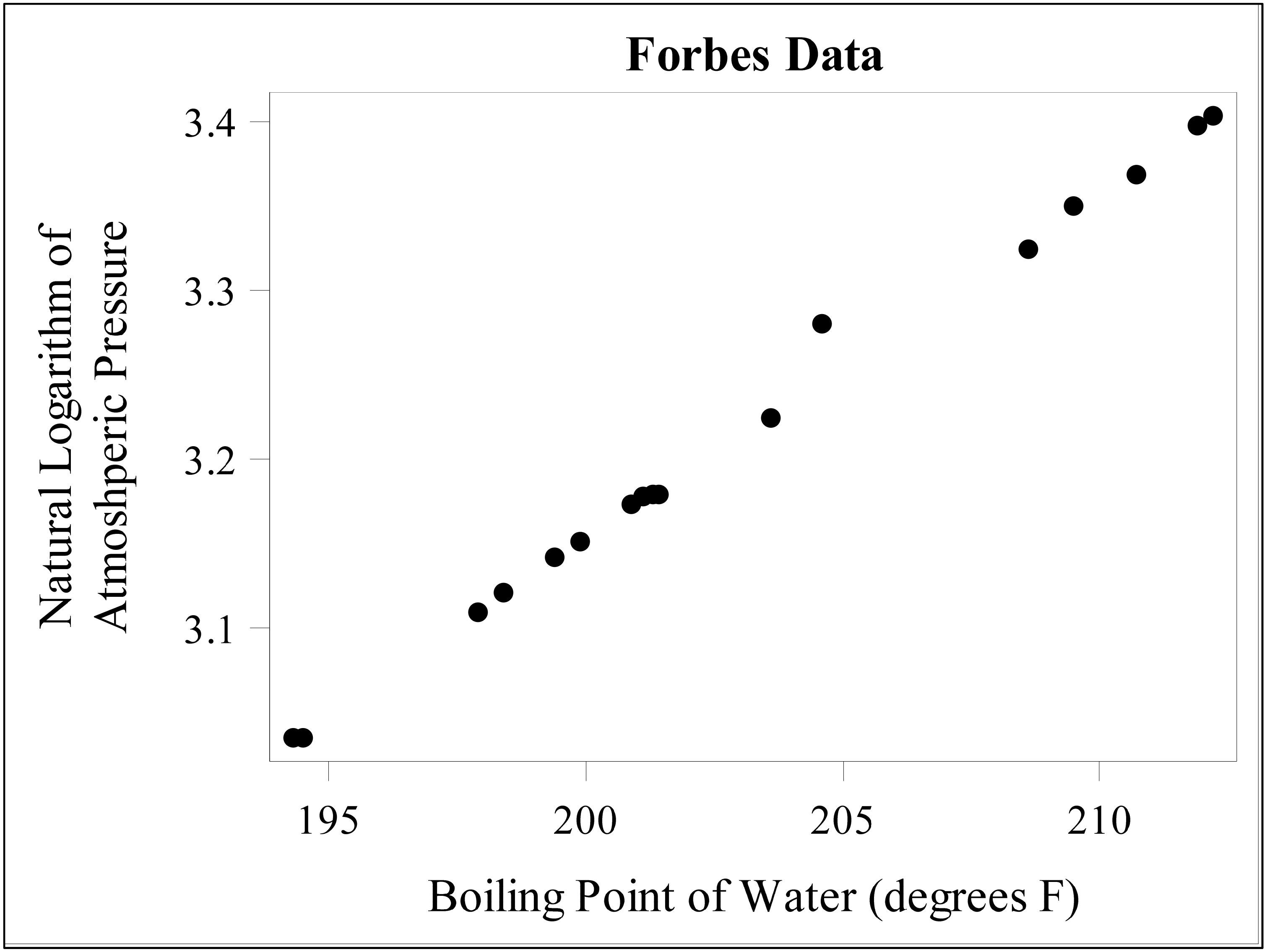
- James D. Forbes collected data in the mountains of Scotland
- n=17 locations (at different altitudes)
- Objective: Predict barometric pressure (in inches of mercury) from boiling point of water (X) in ^{o}F.
- Use Y=log(barometric pressure)
- Motivation: Fragile barometers were difficult to transport

Forbes Data

Obs	BOILING POINT OF WATER (degrees F)	BARAMETRIC PRESSURE (inches Hg)	NATURAL LOG OF BARAMETRIC PRESSURE
1	194.3	20.79	3.03447
2	194.5	20.79	3.03447
3	197.9	22.40	3.10906
4	198.4	22.67	3.12104
5	199.4	23.15	3.14199
6	199.9	23.35	3.15060
7	200.9	23.89	3.17346
8	201.1	23.99	3.17764

Forbes Data

Obs	BOILING POINT OF WATER (degrees F)	BARAMETRIC PRESSURE (inches Hg)	NATURAL LOG OF BARAMETRIC PRESSURE
9	201.3	24.01	3.17847
10	201.4	24.02	3.17889
11	203.6	25.14	3.22446
12	204.6	26.57	3.27978
13	208.6	27.76	3.32360
14	209.5	28.49	3.34955
15	210.7	29.04	3.36867
16	211.9	29.88	3.39719
17	212.2	30.06	3.40320


Forbes Data
Natural Logarithm of
Atmoshperic Pressure
3.4
3.3
3.2
3.1
195
200
205
210
Boiling Point of Water (degrees F)

Analysis of the Forbes Data

- Proposed regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  where $\epsilon_i \sim NID(0,\ \sigma^2)$, $i = 1, 2, ..., 17$

- Y_i=log(pressure)

- X_i=boiling point (^{o}F)

- β_1 is the increase in mean log(pressure) when boiling point of water increases by 1 ^{o}F

- β_0 is the mean log(pressure) when boiling point of water is 0 ^{o}F (Is this extrapolation realistic?)

Predicted Values and Residuals

- Predicted (fitted) values

$$\begin{aligned} \hat{\boldsymbol{Y}}_{\boldsymbol{i}} &= \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{x}_{\boldsymbol{i}} \\ \hat{\mathbf{Y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} \end{aligned}$$

- Residuals

$$\boldsymbol{e}_{\boldsymbol{i}} = \boldsymbol{Y}_{\boldsymbol{i}} - \hat{\boldsymbol{Y}}_{\boldsymbol{i}}$$

Analysis of the Forbes Data

- Estimated regression model

$$\hat{Y} = b_0 + b_1 x = -0.97097 + 0.020623x$$

- Could have subtracted 212 ^{o}F from each boiling point. Then the estimated model is

$$\begin{aligned} \hat{Y} &= b_0 + 212b_1 + b_1(x - 212) \\ &= 3.401106 + 0.020623(x - 212) \end{aligned}$$

- Then 3.401106 is an estimate of the mean log(pressure) at 212 ^{o}F.

Predicted Values: Example

$$\hat{Y}_i = -0.97097 + 0.020623x$$

- Values on the estimated regression line.

- Predict values of Y_i for a given value of x_i

  - $x_i = 201.1$ ^{o}F:
    $-0.97097 + 0.020623(201.1) = 3.176315$

  - $x_i = 210.7$ ^{o}F:
    $-0.97097 + 0.020623(210.7) = 3.374296$

Residuals: Example

$$e_i = Y_i - \hat{Y}_i$$

- Vertical distance between observed value of Y and predicted value of Y.

- Residuals:

  – $x_i = 201.1$ ^{o}F and $Y_i = 3.17764$: $3.17764 - 3.176315 = 0.001325$

  – $x_i = 210.7$ ^{o}F and $Y_i = 3.36867$: $3.36867 - 3.374296 = -0.005626$