THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Analysis of Engineering and Scientific Data

Semester 1 – 2019

Sabrina Streipert s.streipert@uq.edu.au

Chapter 11: Simple Linear Regression

- Aim: Study or analysis of the relationship between two or more variables
  e.g. Pressure of a gas in a container versus its temperature

> We examine a dependent variable and one or more independent variables (= predictors).
> $\Longrightarrow$ Regression Analysis

- Key importance (conditional expectation):

$$\mathbb{E}[Y \mid x] = \mu_{Y|x}$$

Suppose for now, the variable Y depends linearly on only one predictor, i.e.:

$$\mathbb{E}[Y \mid x] = \mu_{Y|x} = \beta_0 + \beta_1 x$$

$\Longrightarrow$

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

where:

- x is a (non-random) predictor, and
- ϵ is a R.V.(=noise) with $\mathbb{E}[\epsilon] = 0$, $\mathrm{Var}(\epsilon) = \sigma^2$.

Assumptions:

- Normality of residuals,
- Constant variance, and,
- Independence of observations

Method:

- Collect data:

$$(x_1, y_1), \ldots, (x_n, y_n).$$

- Assume linear relation:

$$y \approx \beta_0 + \beta_1 x \qquad \leftrightarrow \qquad y = \beta_0 + \beta_1 x + \epsilon$$

- Since we do not have all possible tuples, we can only estimate β_0 and β_1 by $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively, i.e.,

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, \ldots, n.$$

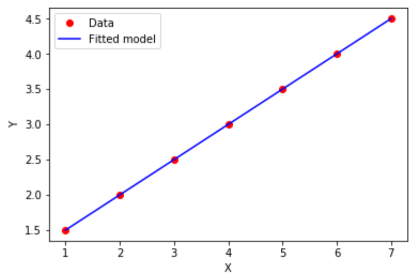
- e_i = **residual**.

Use $\hat{\beta}_0, \hat{\beta}_1$ for predictions.

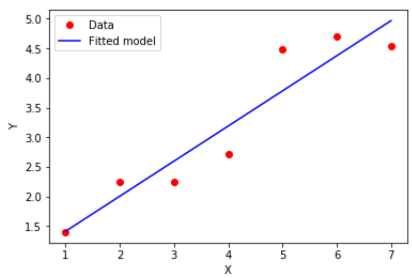
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

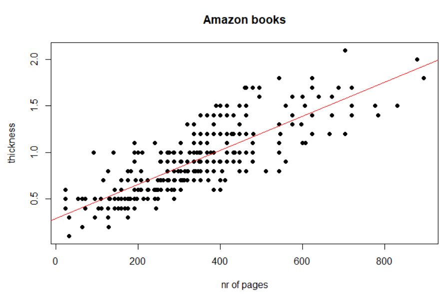
Note that we can also compute predicted observations for our data $(x_i, y_i)_{\{1 \leq i \leq n\}}$.

Ideally, we would like to find $\hat{\beta}_0$ and $\hat{\beta}_1$, such that $y_i = \hat{y}_i$, that is, $e_i = 0$.



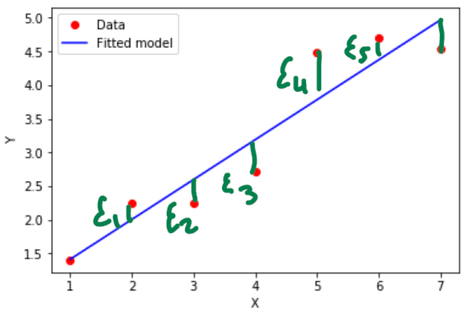
Much more likely:






```
D <- read.delim("amazon-books.txt")
plot(D$NumPages, D$Thick)
abline(lm(D$Thick~D$NumPages), col='red')
```

Total mean squared error



Total Mean Squared Error:

$$L = SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \longrightarrow \min$$

The least squares estimators

- To find the best estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, we would like to minimize

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

- Specifically, solve $\hat{\beta}_0, \hat{\beta}_1 = \text{argmin}_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$.

- The solution, called the *least squares estimators* must satisfy:

  Since we want to minimize L, we take the (partial) derivative and set them equal to zero.

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

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

to (1): $$0 = \cancel{2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = \sum y_i - \sum \beta_0 - \sum \beta_1 x_i$$

$$0 = n \cdot \bar{y} - \beta_0 \cdot n - \beta_1 n \cdot \bar{x}$$

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

to (2):
$$0 = \sum y_i x_i - \sum \beta_0 x_i - \sum \beta_1 x_i^2$$
$$0 = \sum y_i x_i - \beta_0 n \cdot \bar{x} - \beta_1 \sum x_i^2$$
$$\beta_1 = \frac{\sum y_i x_i - \beta_0 n \bar{x}}{\sum x_i^2} = \frac{\sum y_i x_i - \bar{y} n \bar{x} + \beta_1 \bar{x}^2 n}{\sum x_i^2}$$

$$\beta_1 - \beta_1 \frac{\bar{x}^2 n}{\sum x_i^2} = \frac{\sum y_i x_i - \bar{y} n \bar{x}}{\sum x_i^2}$$
$$\beta_1 \left[\frac{\sum x_i^2 - \bar{x}^2 n}{\sum x_i^2} \right] = \frac{\sum y_i x_i - \bar{y} n \bar{x}}{\sum x_i^2}$$
$$\hat{\beta}_1 = \frac{\sum y_i x_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

The least squares solution

Using the sample means, x and y

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

the estimators are:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

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

Additional quantities of interest

$\sum x_i^2 - \bar{x}^2 \cdot n =$

$$S_{XX} = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) = \sum x_i^2 - \sum 2\bar{x}x_i + \sum \bar{x}^2$$

$= \sum x_i^2 - 2\bar{x}(n\bar{x}) + \bar{x}^2 \cdot n =$

$= \sum x_i^2 - \bar{x}^2 \cdot n$

$$S_{XY} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}$$

That is,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} = \frac{S_{XY}}{S_{XX}}.$$

In addition, we have:

total sum of squares

$$SS_T = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum\left(y_i^2 - 2y_i\bar{y} + \bar{y}^2\right) = \sum y_i^2 - n \cdot \bar{y}^2$$

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

regression sum of squares

$$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 = \ldots = \hat{\beta}_1^2 S_{xx} = \hat{\beta}_1 \cdot \hat{\beta}_1 S_{xx} = \hat{\beta}_1 \cdot \frac{S_{xy}}{\cancel{S_{xx}}} \cdot \cancel{S_{xx}} = \hat{\beta}_1 S_{xy}$$

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

error sum of squares

$$L = SS_E = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \ldots = SS_T - \hat{\beta}_1 S_{xy} = SS_T - SS_R$$

It holds that

$$SS_T = SS_R + SS_E,$$

The Analysis of Variance

- We did not consider the final unknown parameter in our regression model:

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

  namely, the $\mathrm{Var}(\epsilon) = \sigma^2$.

- We use the residuals $e_i = \hat{y}_i - y_i$, to obtain an estimate of σ^2.

- Specifically,

$$SS_E = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2,$$

  and it can be shown that

$$\boxed{\mathbb{E}[SS_E] = (n-2)\sigma^2,}$$

  so:

$$\boxed{\hat{\sigma}^2 = \frac{SS_E}{n-2}.}$$

Literature: S.R. Searle "Linear Models"

How good is my regression model?

A widely used measure for a regression model is the following ratio of sum of squares, which is often used to judge the adequacy of a regression model:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T},$$

$R^2 \in [0,1]$

fe $R^2 = 0.877$

" model accounts for 87.7% of the variability in data".

Properties of least square estimator

$$\mathbb{E}[\hat{\beta}_0] = \beta_0, \quad \mathrm{Var}\left(\hat{\beta}_0\right) = \sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{S_{XX}}\right],$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1, \quad \mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{XX}},$$

Therefore, the estimated standard error of the slope and the estimated standard error of the intercept are:

$$se\left(\hat{\beta}_0\right) = \sqrt{\sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{S_{XX}}\right]},$$

$$se\left(\hat{\beta}_1\right) = \sqrt{\frac{\sigma^2}{S_{XX}}}.$$

EXAMPLE

A study considers the microstructure of the ultrafine powder of partially stabilized zirconia as a function of temperature. The data is as follows:

x (Temperature)	1100	1200	1300	1100	1500	1200	1300
y (Porosity)	30.8	19.2	6.0	13.5	11.4	7.7	3.6

$n = 7$

$\bar{x} = 1242.9$ $\quad\bar{y} = 13.1714$

$\sum_{i=1}^{7} x_i^2 = 1093 \cdot 10^4$ $\quad\sum y_i^2 = 1737.7$

$\sum x_i y_i = 110590$

Given $\bar{x}$ find $\sum x_i = \bar{x} \cdot n$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

Find the least square estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{110590 - \frac{n \cdot \bar{x} \cdot \cancel{n} \cdot \bar{y}}{\cancel{n}}}{10930000 - \frac{\cancel{n} \cdot \bar{x}^2}{\cancel{n}}} =$$

$$= \frac{110590 - 7 \cdot 1242.9 \cdot 13.1714}{10930000 - 7 \cdot (1242.9)^2} = -0.0344$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 13.1714 - (-0.0344) \cdot 1242.9 =$$

$$= 55.9272$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Estimate the porosity for a temperature of 1400 degrees Celcius.

$$\hat{y} = 55.9272 - 0.0344 \cdot 1400 =$$

$$= 7.7672$$

Find SS_E (= error sum of squares).

$$SS_E = SS_T - \hat{\beta}_1 S_{xy} = 385.5230$$

where $SS_T = \sum y_i^2 - n\cdot\bar{y}^2 = 1737.7 - 7\cdot(13.1714)^2$

$S_{xy} = \sum x_i y_i - n\cdot\bar{x}\cdot\bar{y} = 110590 - 7\cdot(1242.9)\cdot(13.1714)$

$\hat{\beta}_1 = -0.0344$

Find the least square estimates for y with respect to the predictor $x_i^* = x_i + \bar{x}$. $x_i^* = x_i + \bar{x}$

Q: Find $\hat{\hat{\beta}}_0, \hat{\hat{\beta}}_1$ s.t. $y \approx \hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 x_i^*$

$$y \approx \hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 (x_i + \bar{x}) = \hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 x_i + \hat{\hat{\beta}}_1 \bar{x}$$

$$y \approx \underbrace{\left(\hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 \bar{x}\right)}_{\hat{\beta}_0} + \hat{\hat{\beta}}_1 x_i \qquad \hat{\beta}_1$$

$\hat{\hat{\beta}}_1 = -0.0344$

$$\hat{\hat{\beta}}_0 + \hat{\hat{\beta}}_1 \bar{x} = \hat{\beta}_0 \Rightarrow \hat{\hat{\beta}}_0 = \hat{\beta}_0 - \hat{\hat{\beta}}_1 \bar{x} = 55.9272 + 0.0344 \cdot 1242.9$$

Question: Penalties on using x instead of X (or the other way around)

① $P(\cancel{X}{=}x)$, $P(\cancel{X}{=}i)$ ⟶ Penalty

② $f_{\cancel{x}X}(x)$, $f_{\cancel{xy}XY}(x,y)$

③ $\cancel{y}Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

} No Penalty

Hypothesis tests in linear regression

- Suppose we would like to test:

$$H_0 : \beta_1 = \beta_{1,0}, \quad H_1 : \beta_1 \neq \beta_{1,0}.$$

2-sided test

- The Test Statistic for the Slope is

recall $T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{S_{XX}}}}. = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$$

$\alpha = 0.01$

reject H_0 if $T < -t_{1-\frac{\alpha}{2}, n-2}$, $T > t_{1-\alpha/2, n-2}$

- Under H_0, the test statistic T follows a t - distribution with $n-2$ degree of freedom.

- Reject H_0 if $|t| > t_{\frac{\alpha}{2}, n-2}$

$t_{1-\frac{\alpha}{2}, n-2} = -t_{\frac{\alpha}{2}, n-2}$

- Suppose we would like to test:

$$H_0 : \beta_0 = \beta_{0,0}, \quad H_1 : \beta_0 \neq \beta_{1,0}.$$

- The Test Statistic for the intercept is

$$T = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right]}}. = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}}$$

- Under H_0, the test statistic T follows a t - distribution with $n-2$ degree of freedom.

- Reject H_0 if $|t| > t_{\frac{\alpha}{2}, n-2}$

An important special case of the hypotheses is:

$$H_0 : \beta_1 = 0, \quad H_1 : \beta_1 \neq 0.$$

If we fail to reject $H_0 : \beta_1 = 0$, this indicates that there is no linear relationship between x and y.

Example:

Suppose we have 20 samples regarding oxygen purity (y) with respect to hydrocarbon levels (x) such that

$$\sum_{i=1}^{20} x_i = 23.92, \quad \sum_{i=1}^{20} y_i = 1,843.21, \quad \bar{x} = 1.1960, \quad \bar{y} = 92.1605$$

$$\sum_{i=1}^{20} y_i^2 = 170,044.5321, \quad \sum_{i=1}^{20} x_i^2 = 29.2892, \quad \sum_{i=1}^{20} x_i y_i = 2,214.6566$$

Test: $H_0 : \quad \beta_1 = 0 \qquad H_1 : \quad \beta_1 \neq 0$

for $\alpha = 0.01$.

$n=20$

$n-2=18$

Step1: $t_{1-\frac{\alpha}{2},n-2}$: $\hat{\Phi}\left(t_{1-\frac{\alpha}{2},n-2}\right) = 1-\frac{\alpha}{2} = 0.995$

$t_{1-\frac{\alpha}{2},n-2} \simeq 2.878$

Step2: $T = \dfrac{\hat{\beta}_1 - 0}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}} = \dfrac{14.9475}{\sqrt{\dfrac{(1.18)}{0.6809}}} = 11.35 > 2.878$

$\Rightarrow$ reject

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{20}}{\sum x_i^2 - \frac{(\sum x_i)^2}{20}} = \frac{2214.6566 - \frac{23.92 \cdot 1843.21}{20}}{29.2892 - \frac{(23.92)^2}{20}} = 14.9475$$

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{18} = 1.18$$

$SS_T = \sum y_i^2 - n\cdot\bar{y}^2 = 170044.5321 - 20\cdot(92.1605)^2$

$S_{xy} = \sum x_i y_i - n\cdot\bar{x}\cdot\bar{y} = 10.1774$

$S_{xx} = \sum x_i^2 - n \cdot \bar{x}^2 = 29.2892 - 20 \cdot (1.1960)^2$

Not covered

The F distribution

- An alternative is to use the F statistic as is common in ANOVA (Analysis of Variance) (not covered fully in the course).
- Under H_0, the test statistic

$$F = \frac{SS_R/1}{SS_E/(n-2)} = \frac{MS_R}{MS_E},$$

follows an F - distribution with 1 degree of freedom in the numerator and $n-2$ degrees of freedom in the denominator.

- Here,

$$MS_R = SS_R/1, \quad MS_E = SS_E/(n-2).$$

not covered

Analysis of Variance Table for Testing Significance of Regression

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_E
Error	$SS_E = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_E	
Total	SS_T	$n-1$		

Additional remarks

- There are also confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$ as well as prediction intervals for observations. We do not cover these formulas.

- To check the regression model assumptions, we plot the residuals e_i and check for:
  - Normality,
  - Constant variance, and,
  - Independence

see additional word document

Transformations

Non-linear models can sometimes be "intrinsically" linear.

can apply "nice" transformation to data and get linear relation

Examples:

- $Y = \beta_0 x^{\beta_1} \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.

Why is this "intrinsically" linear?

$$\log(Y) = \log(\beta_0 X^{\beta_1} \varepsilon) = \log(\beta_0) + \log(X^{\beta_1}) + \log(\varepsilon)$$

$$e^{\varepsilon} \sim N(0,\sigma^2)$$

$$\underbrace{\log(Y)}_{Z_i} = \log(\beta_0) + \beta_1 \cdot \log(X) + \log(\varepsilon)$$

$$= \alpha_0 + \alpha_1 W_i + \hat{\varepsilon}$$

$Y_i \neq 0 \neq X_i$

- $Y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.

Why is this "intrinsically" linear?

$$Y = \frac{x}{x\beta_0 + \beta_1 + x\varepsilon}$$

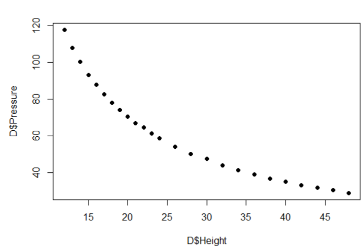
$X_i, Y_i \neq 0$

$$\frac{1}{Y_i} = \frac{x\beta_0 + \beta_1 + x\varepsilon}{x} = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$$

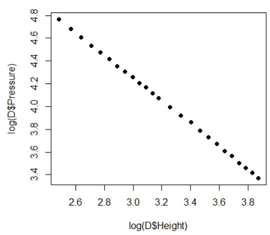
$$z_i = \beta_0 + \beta_1 w_i + \varepsilon_i$$

Boyle's Law

```
D <- read.delim("boyle.txt")
plot(D$Height, D$Pressure, pch=16)
```

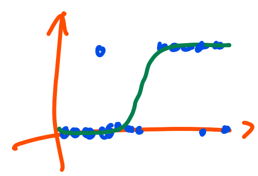


```r
plot(log(D$Height), log(D$Pressure), pch=16)
```



Logistic Regression

- Take the response variable, Y_i as a Bernoulli random variable.
- In this case notice that $\mathbb{E}[Y] = P(Y = 1)$.
- The logit response function has the form



$$\mathbb{E}[Y] = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}.$$

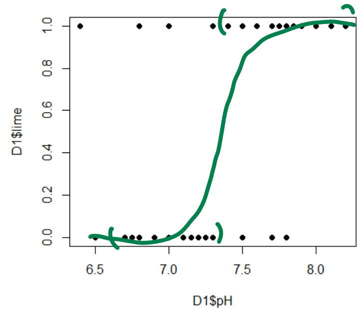
- Fitting a logistic regression model to data yields estimates of β_0 and β_1.
- The following formula is called the odds:

$$\frac{\mathbb{E}[Y]}{1 - \mathbb{E}[Y]} = e^{\beta_0+\beta_1 x}.$$

Example:

Source: *https://dasl.datadescription.com/datafiles/?_sf_s=stream&_sfm_cases=4+59943*

```r
D <- read.delim("streams.txt")
D1 <- D %>% mutate(lime = ifelse(Substrate=='Limestone',1,0)
plot(D1$pH, D1$lime, pch=16)
```



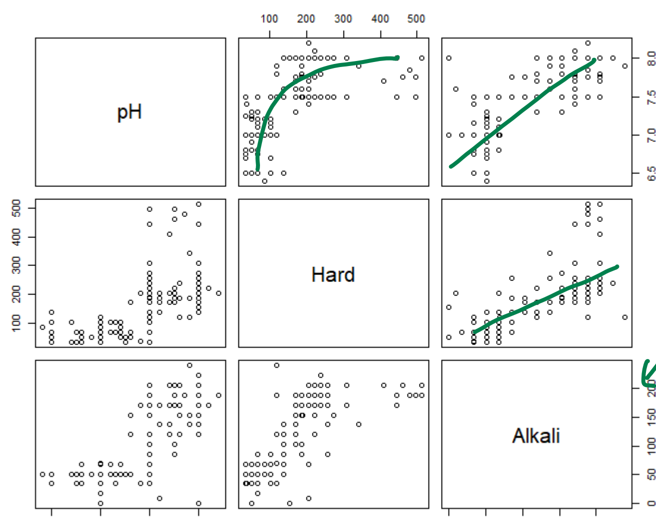
Multiple Regression

What if there are more variables that explain the value of the output?

$\longrightarrow$ Use multiple regression models

Example - revisited:

```
pairs(~pH+Hard+Alkali, data=D)
```



$Alkali = \beta_0 + \beta_1 Hard + \beta_2 pH + \varepsilon$

$Alkali \sim Hard + pH$

End of Stat 2201 material

Following material not covered in class.

Try to fit the following linear model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

with $\epsilon \sim N(0, \sigma^2)$ and $y =$ Hard, $x_1 =$ pH, $x_2 =$ Alkali.

As before, we set up

$$L = \sum_i (y_i - \hat{y}_i)^2 = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \hat{\beta}_2 x_{2,i})^2$$

and try to minimize it.

As before, we find the critical value by setting the partial derivatives equal to zero, that is

$$\begin{aligned}
0 &= \frac{\partial}{\partial \beta_0} L = \sum_i 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \hat{\beta}_2 x_{2,i})(-1) \\
0 &= \frac{\partial}{\partial \beta_1} L = \sum_i 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \hat{\beta}_2 x_{2,i})(-x_{1,i}) \\
0 &= \frac{\partial}{\partial \beta_2} L = \sum_i 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \hat{\beta}_2 x_{2,i})(-x_{2,i})
\end{aligned}$$

which simplifies to

$$\begin{aligned}
0 &= n\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x_1} - \hat{\beta}_2 \bar{x_2} \\
0 &= \sum_i y_i x_{1,i} - \hat{\beta}_0 n \bar{x_1} - \hat{\beta}_1 \sum_i x_{1,i}^2 - \hat{\beta}_2 \sum_i x_{2,i} x_{1,i} \\
0 &= \sum_i y_i x_{2,i} - \hat{\beta}_0 n \bar{x}_2 - \hat{\beta}_1 \sum_i x_{1,i} x_{2,i} - \hat{\beta}_2 \sum_i x_{2,i}^2
\end{aligned}$$

Solving these three equations for the three unknowns $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ yields the estimators.

Back to example:

$n = 172, \sum_i y_i = 24441, \sum_i x_{1,i} = 1259, \sum_i x_{2,i} = 17022, \sum_i x_{1,i}^2 = 9251,$

$\sum_i x_{2,i}^2 = 2338721 \sum_i x_{1,i}x_{2,i} = 128275, \sum_i y_i x_{1,i} = 184664, \sum_i y_i x_{2,i} = 3316930.$

In general, one could express the problem in Matrix notation:

So we have $\vec{y} = X\vec{\beta} + \vec{\epsilon}$

Aim: Find $\vec{\beta}$ so that $L = \sum_i \epsilon_i^2 = \vec{\epsilon}^T \vec{\epsilon}$ is minimal

We have $X^T X \vec{\beta} = X^T \vec{y}$