**Chaper 5:** Matrix Approach to Simple Linear Regression

Matrix:
A *m by n matrix* **B** is a grid of numbers with m rows and n columns.

$$\mathbf{B} = \begin{matrix} b_{11} \ldots & b_{1n} \\ \vdots & \vdots \\ b_{m1} \ldots & b_{mn} \end{matrix}$$

Element $b_{ik}$ is from the ith row and kth column.

A *vector* **b** is a matrix with 1 column:

$$\mathbf{b} = \begin{matrix} b_1 \\ \vdots \\ b_n \end{matrix}$$

A transpose of a *m by n matrix* **B** is a *n by m matrix* **B′**

$$\mathbf{B}' = \begin{matrix} b_{11} \ldots & b_{m1} \\ \vdots & \vdots \\ b_{1n} \ldots & b_{mn} \end{matrix}$$

A product of a *m by n* matrix (**B**) and a *n by p* matrix (**A**) is:

$$\mathbf{BA} = \begin{matrix} (b_{11}a_{11}+\ldots+b_{1n}a_{n1}) \ldots (b_{11}a_{1p}+\ldots+b_{1n}a_{np}) \\ \vdots \\ (b_{m1}a_{11}+\ldots+b_{mn}a_{n1}) \ldots (b_{m1}a_{1p}+\ldots+b_{mn}a_{np}) \end{matrix}$$

which has dimension m by p. (See board for detail.)

An *n by n* (square) matrix **I** is called the *identity* matrix if

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}$$ (We say "1 on the diagonal")

Note that $\mathbf{Ib} = \mathbf{b}$ (assuming the dimensions line up).

Some square matrices have an inverse: $\mathbf{B}^{-1}$ which is matrix such that $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$.

A regression model can be expressed as:

$\mathbf{y} = \mathbf{X}\beta + \varepsilon$ where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

The matrix **X** is called the **design matrix** and $\beta$ are the coefficients.

**e** is a random vector. It has mean $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

and covariance matrix: $\text{cov}(\mathbf{e}) =$

$$\begin{pmatrix} \sigma^2 & 0 & \dots & \dots & 0 \\ 0 & \sigma^2 & \dots & \dots & 0 \\ \vdots & & & & \vdots \\ & & & \sigma^2 & 0 \end{pmatrix}$$

0 … … 0 $\sigma^2$ The ikth entry is $\text{cov}(\varepsilon_i, \varepsilon_k)$.

A nice thing about matrices is that the regression coefficient estimates now have a simple form:

$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ where $\mathbf{b} = (b_0\ b_1)'$.
(Derive on board.)

$\mathbf{Y\text{-}hat} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} = \mathbf{Hy}$ and $\mathbf{H}$ is called the "hat" matrix.

$\text{cov}(\mathbf{b}) = \sigma^2 (\mathbf{X'X})^{-1}$ which is estimated by $\text{MSE}(\mathbf{X'X})^{-1}$.

# Chapter 6: Multiple Regression I

In the simple (i.e. only one covariate) linear regression
y = beta0 + beta1 x + e, the model was motivated by fitting a line through a scatterplot.

A multiple linear regression models the mean of y as a linear function of more than one covariate:

y = beta0 + beta1 x1 + … + betap xp + e

with e independent, normal(0,sigma^2)

Note that the regression model is:
E(y) = beta0 + beta1 x1 + … + betap xp

or (write matrix formulation)

These are *first order models* since they do not include products of xs.

(i.e. E(y) = beta0 + beta1 x1 + beta2 x2 + beta3 (x1*x2) is a second order model.)

Polynomial models are multiple regressions:

E(y) = beta0 + beta1 x + beta2 x^2

When p = 2, the betas are estimated by fitting a plane through a cloud of data (see figure). For p>2, it's a higher dimensional generalization.

(write matrix formulation for estimators)

**Example: Cheddar Cheese**
**Reference:** Moore, David S., and George P. McCabe (1989). *Introduction to the Practice of Statistics*.
**Authorization:**
**Description:** As cheese ages, various chemical processes take place that determine the taste of the final product. This dataset contains concentrations of various chemicals in 30 samples of mature cheddar cheese, and a subjective measure of taste for each sample. The variables "Acetic" and "H2S" are the natural logarithm of the concentration of acetic acid and hydrogen sulfide respectively. The variable "Lactic" has not been transformed.
**Variable Names:**

1. Case: Sample number
2. Taste: Subjective taste test score, obtained by combining the scores of several tasters
3. Acetic: Natural log of concentration of acetic acid
4. H2S: Natural log of concentration of hydrogen sulfide
5. Lactic: Concentration of lactic acid

Our goal is to predict Taste.

First model:

$Taste_i = beta0 + beta1\ Acetic_i + beta2\ H2S_i + beta3\ Lactic_i + e_i$

```
Call:
lm(formula = taste ~ Acetic + H2S + Lactic, data = cheese)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
Acetic        0.3277     4.4598   0.073  0.94198
H2S           3.9118     1.2484   3.133  0.00425 **
Lactic       19.6705     8.6291   2.280  0.03108 *

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-Squared: 0.6518,   Adjusted R-squared: 0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

Go through what the various things in the table above mean.

Do parameter estimates change depending on what else is in the model?

Yes! (when the covariates are correlated themselves)

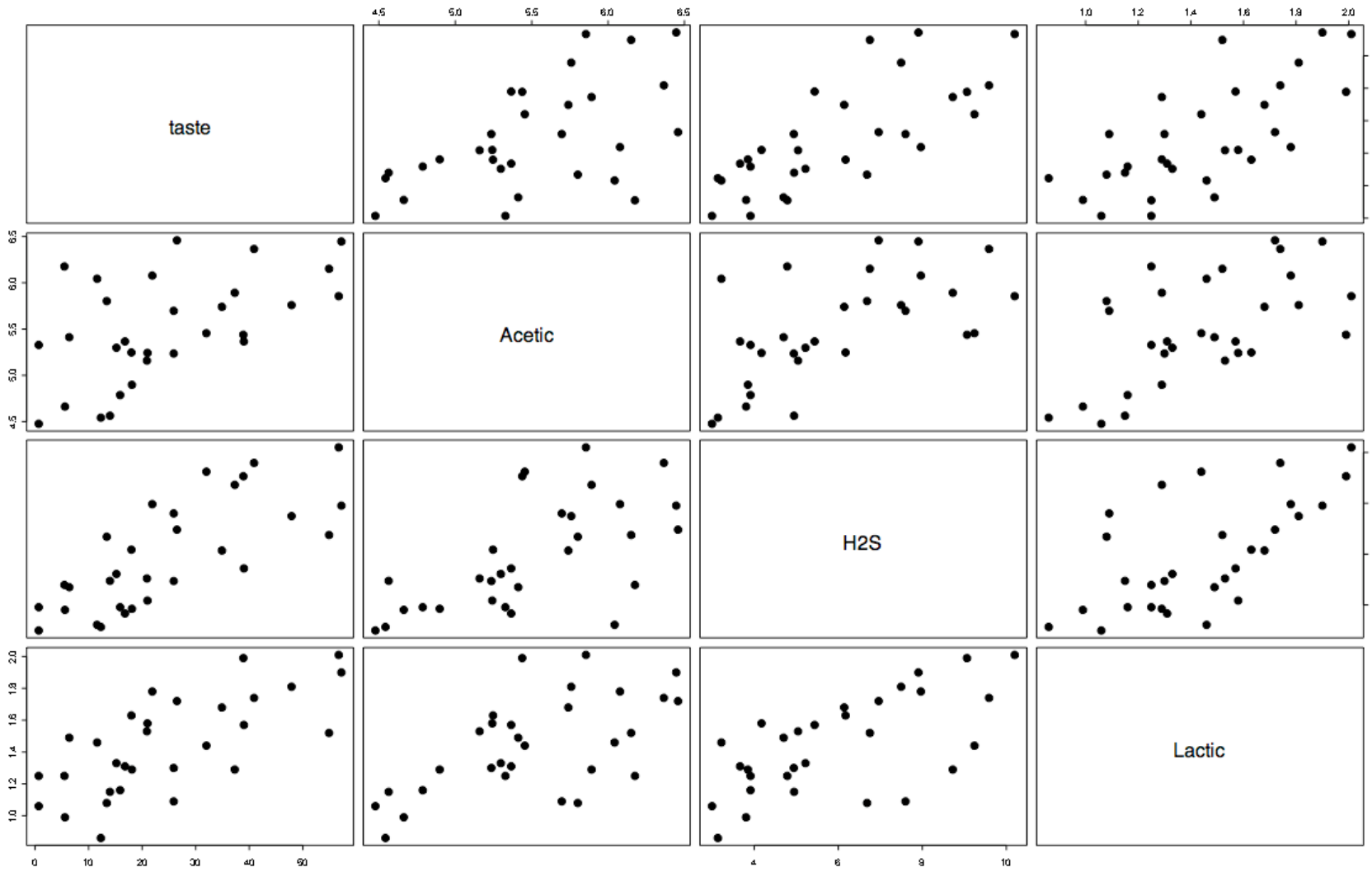
```
Call:
lm(formula = taste ~ Acetic, data = cheese)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -61.499     24.846  -2.475  0.01964 *
Acetic        15.648      4.496   3.481  0.00166 **
Residual standard error: 13.82 on 28 degrees of freedom
Multiple R-Squared: 0.302,    Adjusted R-squared: 0.2771
F-statistic: 12.11 on 1 and 28 DF,  p-value: 0.001658
```

$Taste_i = beta0 + beta1\ Acetic_i + beta2\ H2S_i + beta3\ Lactic_i + e_i$

```
Call:
lm(formula = taste ~ Acetic + H2S + Lactic, data = cheese)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
Acetic        0.3277     4.4598   0.073  0.94198
H2S           3.9118     1.2484   3.133  0.00425 **
Lactic       19.6705     8.6291   2.280  0.03108 *
Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-Squared: 0.6518,   Adjusted R-squared: 0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```


taste
Acetic
H2S
Lactic

Analysis of Variance Table

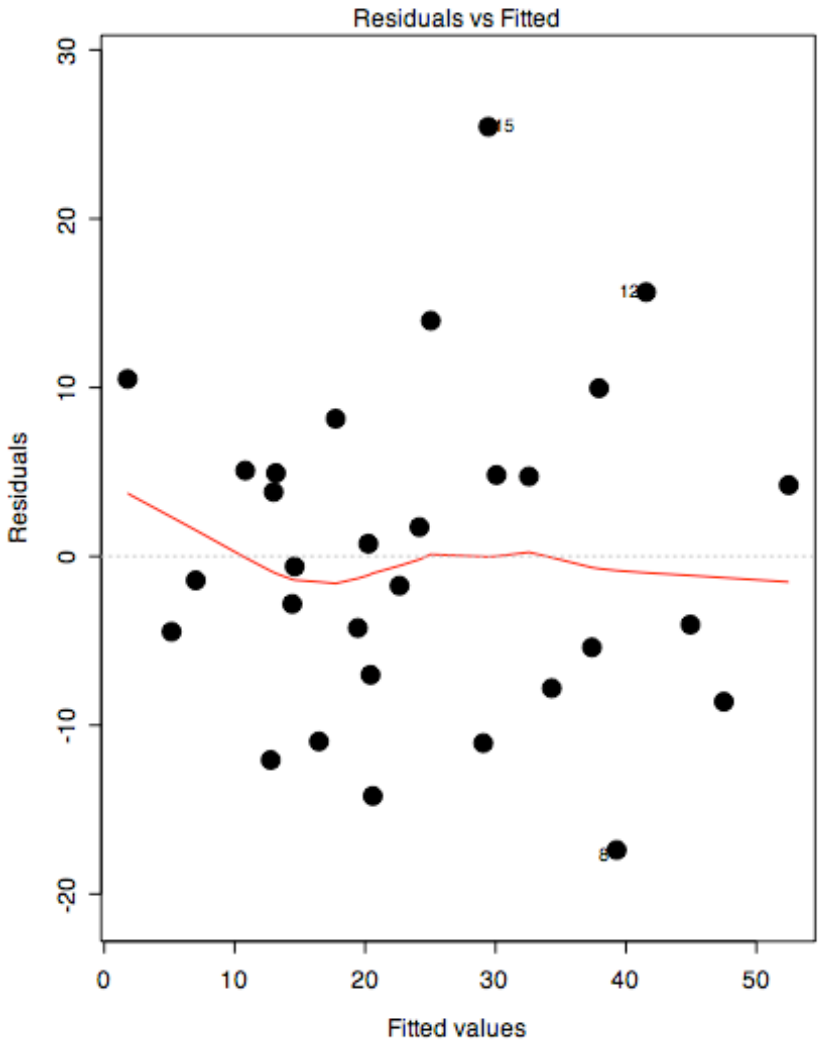
Response: taste

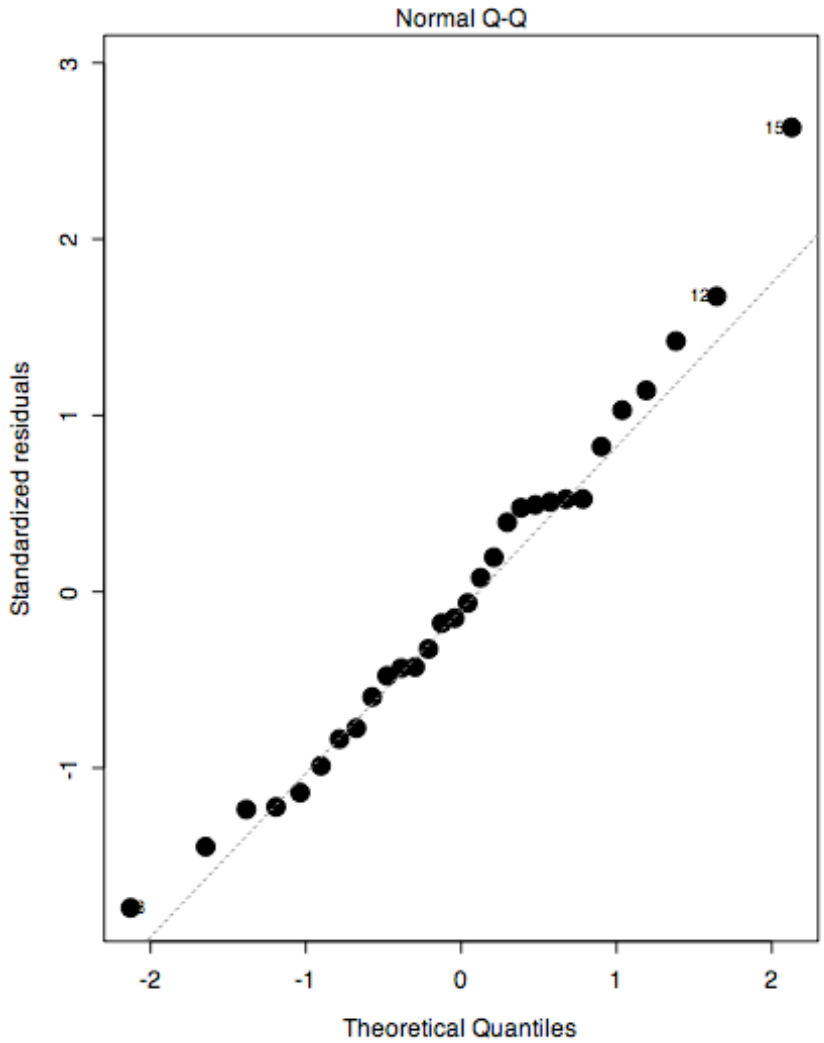
| | Df | Sum Sq | Mean Sq |
|---|---|---|---|
| Acetic | 1 | 2314.14 | 2314.14 |
| H2S | 1 | 2147.02 | 2147.02 |
| Lactic | 1 | 533.32 | 533.32 |
| Residuals | 26 | 2668.41 | 102.63 |

---

Response: taste

| | Df | Sum Sq | Mean Sq |
|---|---|---|---|
| Acetic | 1 | 2314.14 | 2314.14 |
| Lactic | 1 | 1672.68 | 1672.68 |
| H2S | 1 | 1007.66 | 1007.66 |
| Residuals | 26 | 2668.41 | 102.63 |


Residuals vs Fitted
Residuals
Fitted values
30
20
10
0
-10
-20
0
10
20
30
40
50
15
12
6


Normal Q-Q
Standardized residuals
Theoretical Quantiles
3
2
1
0
-1
-2
-1
0
1
2
15
12
6