# WORKSHOP ON THE ANALYSIS OF NEURAL DATA 2001

# MARINE BIOLOGICAL LABORATORY

# WOODS HOLE, MASSACHUSETTS

## A REVIEW OF STATISTICS

## PART 3: REGRESSION ANALYSIS AND THE GENERALIZED LINEAR MODEL

EMERY N. BROWN

NEUROSCIENCE STATISTICS RESEARCH LABORATORY
DEPARTMENT OF ANESTHESIA AND CRITICAL CARE
MASSACHUSETTS GENERAL HOSPITAL

DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
HARVARD MEDICAL SCHOOL / MIT

## Regression Analysis

### A. Simple Regression

### B. Model Assumptions

### C. Model Fitting

### D. Properties of Parameter Estimates

### E. Model Goodness-of-Fit

**F-test**
**$R^2$**
**Analysis of Residuals**

### F. The Geometry of Regression (Method of Least-Squares)

## A. Simple Regression

Assume we have a data consisting of pairs of two variables and we denote them as $(x_1, y_1)$, $(x_2, y_2)$, …, $(x_n, y_n)$. For example, x and y might be measurements of height and weight of a set of individuals from a well-defined cohort. Let's assume that there is a linear relation between x and y. We assume that the linear relation may be written as

$$y = \alpha + \beta x$$

**Example 1.**

For example let us consider this example taken from Draper and Smith (1981), *Applied Regression Analysis*

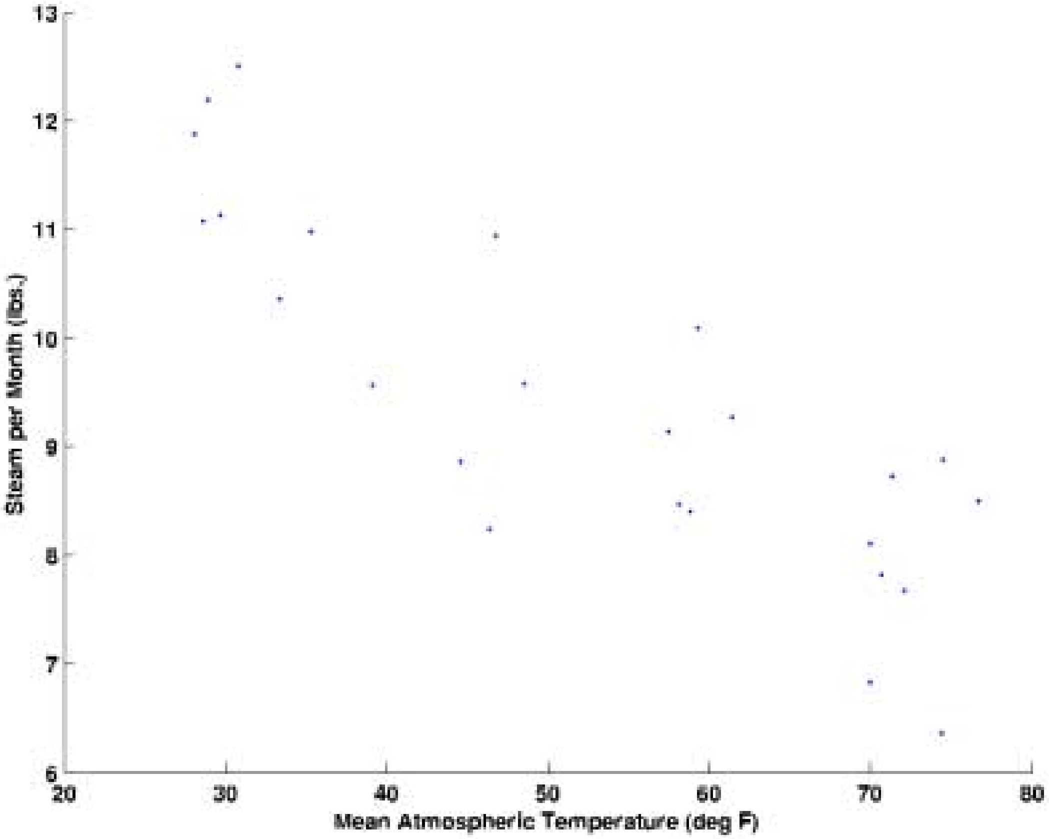


**Figure 1. Relation between Monthly Steam Production and Mean Atmospheric Temperature.**

The variable y is the amount of steam produced per month in a plant and variable x is the mean atmospheric temperature. There is an obvious negative relation.

**B. Model Assumptions**
We assume

i) $E[y \mid x] = \alpha + \beta x$
ii) The x's are fixed non-random covariates
iii) The y's are independent Gaussian random variables with mean $\alpha + \beta x_i$ and variance $\sigma^2$

**C. Model Fitting**
Our objective is to estimate the parameters $\alpha, \beta$ and $\sigma^2$. Because y is assumed to have a Gaussian distribution conditional on x, a logical approach is to use maximum likelihood estimation. For these data the joint probability density (likelihood) is

$$f\left(y \mid \alpha, \beta, \sigma^2, x\right) =$$

$$\prod_{i=1}^{N} f\left(y_i \mid \alpha + \beta x_i, \sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2}\right\}$$

The log likelihood is

$$logf\left(y \mid x, \alpha, \beta, \sigma^2\right) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \alpha - \beta x)^2}{\sigma^2}$$

Differentiating with respect to $\alpha$ and $\beta$ yields

$$\frac{\partial \log f\left(y \mid x, \alpha, \beta, \sigma^2\right)}{\partial \alpha} = -2\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)$$

$$\frac{\partial \log f\left(y \mid x, \alpha, \beta, \sigma^2\right)}{\partial \beta} = -2\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)x_i$$

Setting the derivative equal to zero yields the normal equations

$$\alpha N + \beta\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i$$
$$\alpha\sum_{i=1}^{N} x_i + \beta\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i$$

Or in matrix form

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}\begin{bmatrix}\alpha \\ \beta\end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

$$\begin{bmatrix}\alpha \\ \beta\end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}^{-1}\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

The solution for $\beta$ and $\alpha$ are

$$\hat{\beta} = \frac{\sum_{i=1}^{N} x_1 y_1 - \left( \sum_{i=1}^{N} x_1 \right) / N}{\sum_{i=1}^{N} x_i^2 - \left( \sum_{i=1}^{N} x_i \right)^2 / N} = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$$

$$\hat{\alpha} = \overline{y} - \beta \overline{x}$$

We may write any estimate as

$$\hat{y} = \hat{\alpha} + \hat{\beta} x = \overline{y} - \hat{\beta} \overline{x} + \hat{\beta} x = \overline{y} + \hat{\beta} (x - \overline{x})$$

If we go back and differentiate the log likelihood with respect to $\sigma^2$, we obtain the maximum likelihood estimate

$$\hat{\sigma}^2 = \sum_{i=1}^{N} \left( y_i - \hat{\alpha} - \hat{\beta} x_i \right)^2 / N$$

**Remarks**

1. Choosing $\alpha$ and $\beta$ by the maximum likelihood is equivalent to the method of least square in this case. By the method of least squares, we minimize the sum of the squared initial deviations of the data from the regression line.

2. The estimate of $\hat{\alpha}$ shows that every regression line goes through the point $(\overline{x}, \overline{y})$.

3. The residuals are $y_i - \hat{y}_i$ and are the components in the data which the model does not explain. We note that

$$\hat{y}_i = \overline{y} + \beta (x_i - \overline{x})$$

$$\sum_{i=1}^{N} (y_i - \hat{y}_i) = \sum_{i=1}^{N} (y_i - \overline{y}) + \beta \sum_{i=1}^{N} (x_i - \overline{x}) = 0$$

## 4. The Pythagorean Relation

$$y_i - \hat{y}_i = (y_i - \overline{y}) - (\hat{y}_i - \overline{y})$$

$$\sum (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \{(y_i - \overline{y}) - (\hat{y}_i - \overline{y})\}^2$$

$$= \sum_{i=1}^{N} (y_i - \overline{y})^2 + \sum_{i=1}^{N} (y_i - \overline{y})^2 - 2\sum_{i=1}^{N} (y_i - \overline{y})(\hat{y}_i - \overline{y})$$

*N.B.*

$$-2\sum_{i=1}^{N} (y_i - \overline{y})\,\beta\,(x_i - \overline{x}) = -2\beta\sum_{i=1}^{N} (y_i - \overline{y})(x_i - \overline{x})$$

$$= -2\beta^2 \sum_{i=1}^{N} (x_i - \overline{x})^2$$

$$= -2\sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2 \ ,$$

and

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - \overline{y})^2 - \sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2 \ .$$

Or

$$\sum_{i=1}^{N} (y_i - \overline{y})^2 = \sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

| | | | | |
|---|---|---|---|---|
| Total sum of squares<br>(TSS) | = | Explained sum of squares<br>(ESS) | + | Residual sum of squares<br>(RSS) |

## 5. Correlation and Regression

Note that the correlation coefficient for x and y is

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2\right]^{\frac{1}{2}}}$$

and recall that

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

Hence,

$$\beta = \left\{\frac{\sum_{i=1}^{N}(y_i - \overline{y})^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}\right\}^{\frac{1}{2}} r_{xy}$$

The regression coefficient is a scale version of the correlation coefficient.

## D. Properties of the Parameter Estimates

The variances of the parameter estimates

$$\text{Var }\left(\hat{\beta}\right) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$\text{Var }(\hat{\alpha}) = \frac{\sum_{i=1}^{N} x_i^2 \sigma^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$$

If we estimate $\sigma^2$ by $\hat{\sigma}^2$ then $1-\rho$ confidence intervals for the parameters based on the ***t***-distribution are

$$\hat{\beta} \pm \frac{t_{n-2,1-\rho/2}\hat{\sigma}}{\left\{\sum_{i=1}^{N} (x_i - \overline{x})^2\right\}^{\frac{1}{2}}}$$

$$\hat{\alpha} \pm t_{n-2,1-\rho/2} \left\{ \frac{\sum_{i=1}^{N} x_i^2}{N \sum_{i=1}^{N} (x_i - \overline{x})^2} \right\}^{\frac{1}{2}} \hat{\sigma}$$

We also invert the above statistics to test hypotheses about the regression coefficients by constructing a ***t***-test.

To construct a confidence interval for a predicted $y_k$ value at a given x value, $x_k$, we note that the

$$\text{var}(y_k) = \frac{\sigma^2}{N} + \frac{(x_k - \overline{x})^2 \sigma^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$$

and hence an approximate $1-\rho$ confidence interval is

$$\hat{y}_k \pm t_{n-2,1-\rho/2} \left[ \frac{1}{N} + \frac{(x_k - \overline{x})^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \right]^{\frac{1}{2}} \hat{\sigma}$$

**Example 1. (continued)**

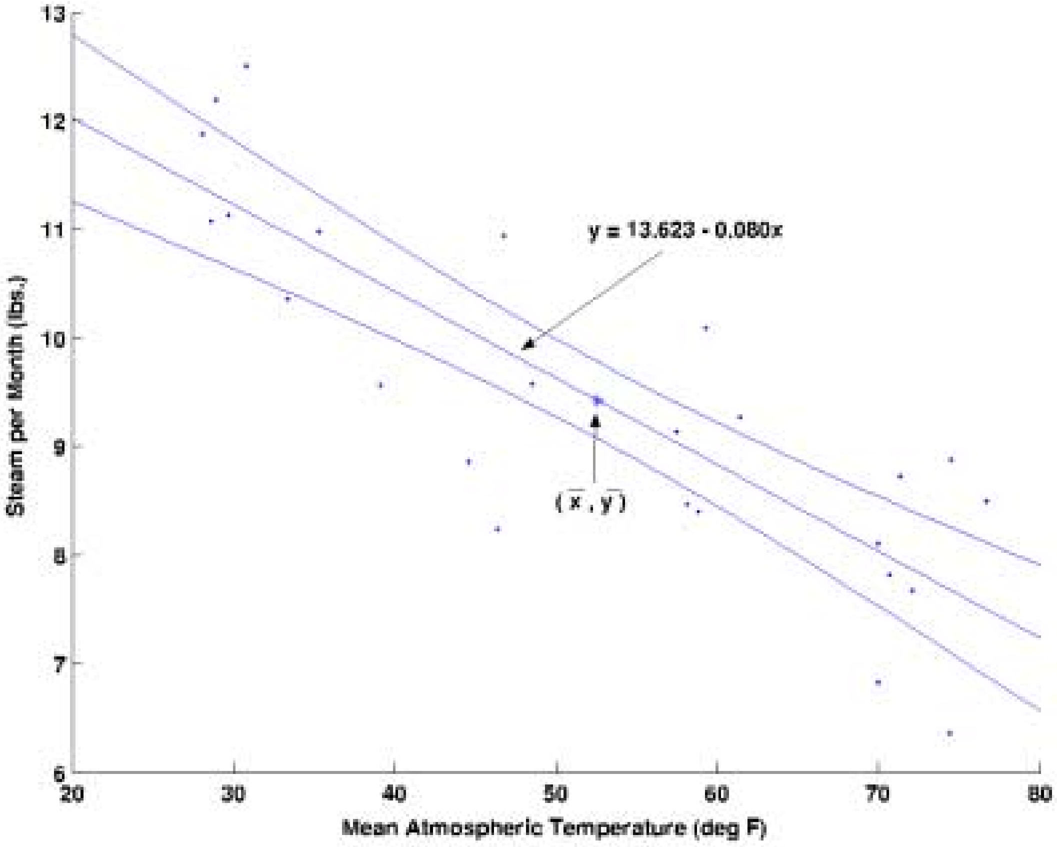


**Figure 2. Fit and Confidence Intervals for a Simple Linear Regression Model**

## E. Model Goodness-of-Fit

A crucial (if not the most crucial) step in statistical analyses of data is measuring goodness of fit. That is, how well does the model agree with the data. We discuss three graphical and two statistical measures of goodness-of-fit.

### 1. F-test

Given the null hypothesis $H_0 : \beta_1 = 0$, i.e. that there is no linear relation in x and y, we can test this explicitly using an F-test defined as

$$F_{p-1,N-p} = \frac{ESS/(p-1)}{RSS/(N-p)} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2 \Big/ (p-1)}{\sum_{i=1}^{N}(y_i - \hat{y})^2 \Big/ (N-p)}$$

where ESS is the explained sum of squares; RSS is the residual sum. We reject this null hypothesis for large values of the F statistic. This suggests that if the amount of the variance in the data that the regression explains is large relative to the amount which is unexplained then we reject the null hypothesis.

**2. R-Squared**

Another measure of goodness-of-fit is the $R^2$. It is defined as

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}$$

The $R^2$ measures the fraction of variance in the data explained by the regression equation. The R-Squared and the F-statistic are related as

$$R^2 = \frac{F\left(p-1\right)/\left(N-p\right)}{1+F\left(p-1\right)/\left(N-p\right)}$$

Clearly, as the F-statistic increases, so does the $R^2$.

**Example 1. Data Analysis Summary (Continued)**

| **Parameter Estimate** | **Standard Error** | **t-Statistic** |
|---|---|---|
| $\hat{\alpha} =$ **13.62** | **0.582** | **23.41** |
| $\hat{\beta} =$ **-0.0798** | **0.010** | **-7.60** |

**TSS = 63.82**
**ESS = 45.69**
**RSS = 17.13**

**F-statistic (1, 23) = 57.52, $R^2$= 0.7144**
Factoid: The square of the t-statistic (n-p degrees of freedom) is the F-statistic (1, n-p).

**3. Graphical Measures of Goodness-of-Fit**

1. Plot the Raw Data Is the relation linear?
2. Plot the Fit vs. the ResidualsIs there a relation between the explained and the unexplained structure in the data?
3. Plot the Residuals Lack of fit of the model

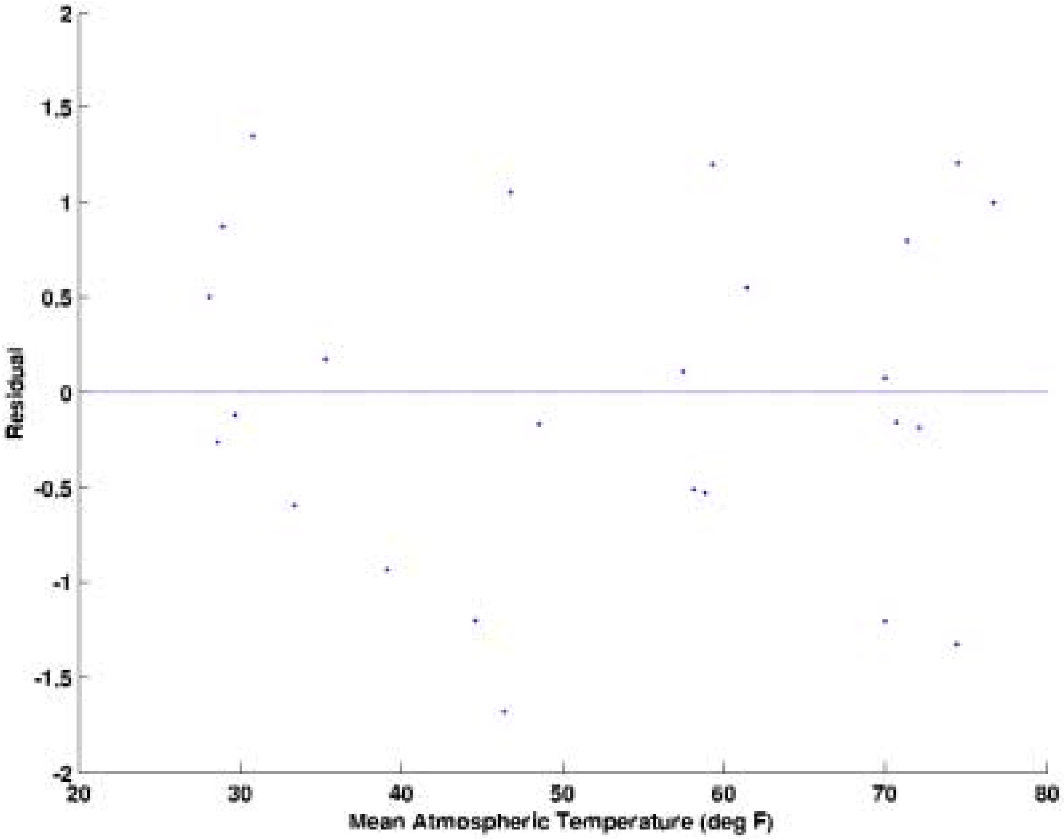


**Figure 3. Plot of Residuals versus x.**

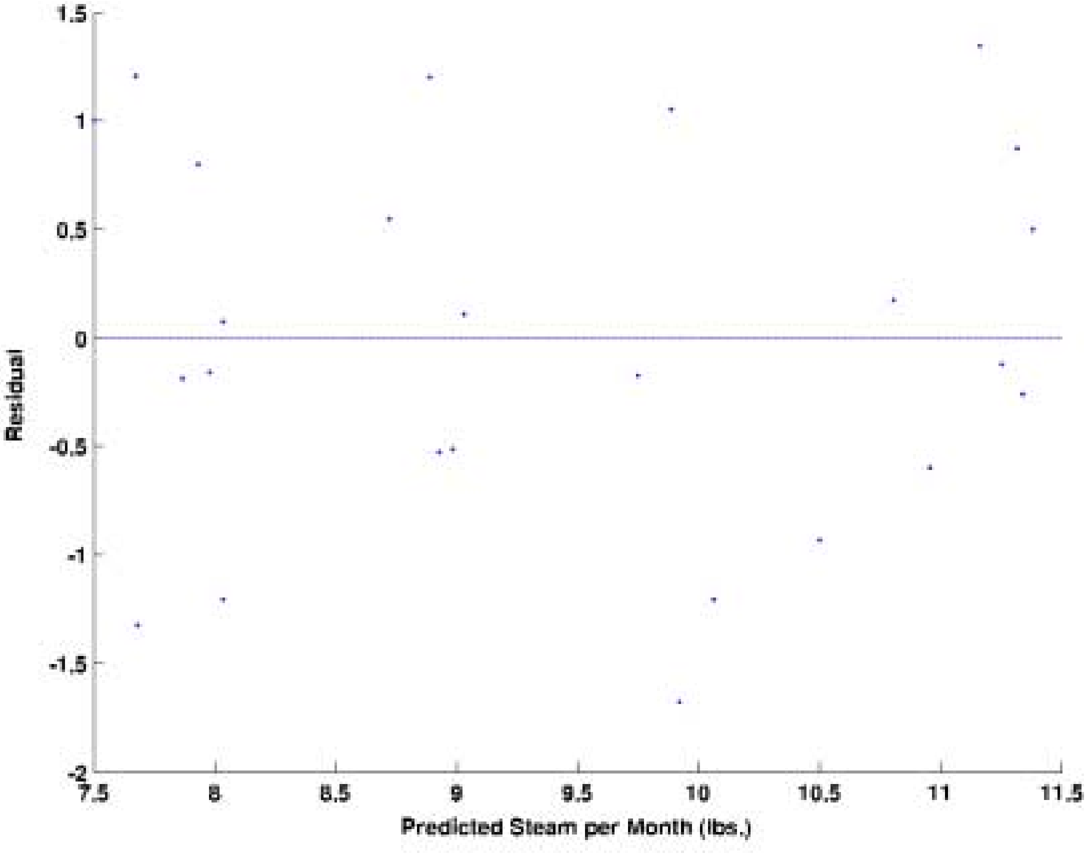
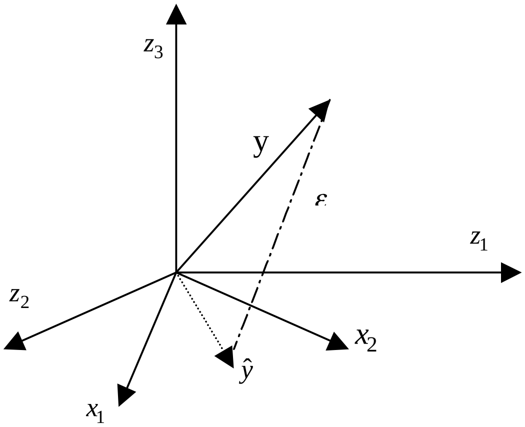


**Figure 4. Plot of the Residuals against the Predicted y values.**

**F. The Geometry of Regression Analysis (Method of Least-Squares)**
Regression analysis has an intuitively appealing geometric interpretation.



$$R^2 = \frac{\|\hat{y}\|^2}{\|y\|^2} \qquad\qquad F - statistic \propto \frac{\|\hat{y}\|^2}{\|\varepsilon\|^2}$$

**I. Derivation and Properties of the Least Squares Estimates**

The procedure used to estimate the parameters in the linear regression models in this study is least squares. In this section we first derive the least squares estimates, and then by making some distributal assumptions about our model, we next derive the statistical properties of the estimates.

We begin by assuming a model of the form

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{1}$$

where $\underset{\sim}{Y}$ is an $n \times l$ vector of dependent observations.

$\underset{\sim}{X}$ is an $n \times p$ matrix of explanatory observations, i.e. the matrix contains $p$ $n \times l$ dimensional vectors.

$\underset{\sim}{\beta}$ is an $n \times l$ vector of parameters.

$\underset{\sim}{\varepsilon}$ is an $n \times l$ vector of errors.

In order to estimate our vector of parameters $\underset{\sim}{\beta}$ using least squares, we assume that $\underset{\sim}{X}$ is of rank $\underset{\sim}{p}$ or, in other words, that the $p$ column vectors in $\underset{\sim}{X}$ are linearly independent. Therefore, we want to find the estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{\hat{\beta}}$, which minimizes

$$\begin{aligned}\underset{\sim}{\hat{\varepsilon}}'\underset{\sim}{\hat{\varepsilon}} &= (\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}}) \\ &= \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{Y} - \underset{\sim}{Y}'\underset{\sim}{X}\underset{\sim}{\hat{\beta}} + \underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\hat{\beta}}\end{aligned} \tag{1.2}$$

Because $\underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{Y}$ is a scalar it equals its transpose $\underset{\sim}{Y}'\underset{\sim}{X}\underset{\sim}{\hat{\beta}}$. Consequently (1.2) simplifies to

$$\underset{\sim}{\hat{\varepsilon}}'\underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{Y}'\underset{\sim}{Y} - 2\underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{Y} + \hat{\beta}'\underset{\sim}{X}\underset{\sim}{\hat{\beta}} \tag{1.3}$$

To find the least squares estimates of $\underset{\sim}{\beta}$ we differentiate (1.3) with respect to $\underset{\sim}{\hat{\beta}}$ and set the derivative equal to zero.

$$\frac{\partial(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}})}{\partial\hat{\underset{\sim}{\beta}}} = -2\underset{\sim}{X}'\underset{\sim}{Y} + (\underset{\sim}{X}'\underset{\sim}{X})\hat{\underset{\sim}{\beta}} \tag{1.4}$$

Solving for $\hat{\underset{\sim}{\beta}}$ gives

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} \tag{1.5}$$

That

$$\frac{\partial\left(\hat{\underset{\sim}{\beta}}'\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}}\right)}{\partial\hat{\underset{\sim}{\beta}}} = (\underset{\sim}{X}'\underset{\sim}{X})\hat{\underset{\sim}{\beta}}$$

follows from the fact that $\hat{\underset{\sim}{\beta}}'(\underset{\sim}{X}'\underset{\sim}{X})\hat{\underset{\sim}{\beta}}$ is a quadratic form. Furthermore, we know that $(\underset{\sim}{X}'\underset{\sim}{X})$ has an inverse since we initially assumed $\underset{\sim}{X}$ had rank $p$. Hence, $(\underset{\sim}{X}'\underset{\sim}{X})$ is positive definite of rank $p$and has inverse $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$. The second order conditions for the minimization of $\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}$ follows directly since $(\underset{\sim}{X}'\underset{\sim}{X})$ is positive definite. In order to test hypotheses about our data we need to make some distributional assumptions about our model. The standard assumptions for the linear regression model are:

**I.** $\underset{\sim}{\varepsilon} \sim N\left(\underset{\sim}{0}, I\underset{\sim}{\sigma}^2\right)$. The $\varepsilon_i$ ‘s are independent, normally distributed random errors with mean 0 and common variance $\sigma^2$.

**II.** The $p$ $n$-dimensional vectors comprising $\underset{\sim}{X}$ are taken as fixed constants.
From these two assumptions we see that $\underset{\sim}{Y}$ being a linear function of $\underset{\sim}{\varepsilon}$, is also normally distributed with mean $\underset{\sim}{X}\underset{\sim}{\beta}$ and variance $\underset{\sim}{I}\sigma^2$. Using these distributional properties we now derive the distributions of the least squares estimates.

**Proposition 1.1** The least squares estimate of $\underset{\sim}{\beta}, \hat{\underset{\sim}{\beta}}$, is normally distributed with means $\underset{\sim}{\beta}$ and variance $\sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$.

**Proof:** To show that $\underset{\sim}{\beta}$ is normally distributed we have only to notice that $\underset{\sim}{\beta}$ is a linear function of $\underset{\sim}{Y}$, since linear functions of independent normally distributed random variables are also normally distributed.

To find the mean of the distribution of $\underset{\sim}{\beta}$ consider

$$E\left(\underset{\sim}{\hat{\beta}}\right) = E\left\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}\right\} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{E}(\underset{\sim}{\varepsilon}) \tag{1.6}$$

But from Assumption I, we know that $E(\underset{\sim}{\varepsilon}) = 0$, and since $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{I}$, it follows that $E\left(\underset{\sim}{\hat{\beta}}\right) = \underset{\sim}{\beta}$. To determine the variance of $\underset{\sim}{\hat{\beta}}$, we consider

$$\begin{aligned} Var(\underset{\sim}{\hat{\beta}}) &= Var\left\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}\right\} \\ &= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}''Var(\underset{\sim}{Y}\underset{\sim}{Y}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \end{aligned} \tag{1.7}$$

But $Var(\underset{\sim}{Y}\underset{\sim}{Y}') = \sigma^2\underset{\sim}{I}$, so we can rewrite (7) as

$$\begin{aligned} Var(\underset{\sim}{\hat{\beta}}) &= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{X})(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2\underset{\sim}{I} \\ &= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2\underset{\sim}{I} \end{aligned} \tag{1.8}$$

**Proposition 1.2** The least squares estimate of $\underset{\sim}{Y}, \underset{\sim}{\hat{Y}}$, where $\underset{\sim}{\hat{Y}} = \underset{\sim}{X}\underset{\sim}{\hat{\beta}}$ is $N\left(\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\right)$.

**Proof:** the fact that $\underset{\sim}{\hat{Y}}$ has a normal distribution follows directly from the fact that it is a linear combination of $\underset{\sim}{Y}$. We find the mean of $\underset{\sim}{\hat{Y}}$ by considering

$$E(\underset{\sim}{\hat{Y}}) = E\left\{\underset{\sim}{X}\underset{\sim}{\hat{\beta}}\right\} = E\left\{\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'E(\underset{\sim}{\varepsilon})\right\}. \tag{1.9}$$

Because $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{X}) = \underset{\sim}{I}$ and $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, we get $E(\underset{\sim}{\hat{Y}}) = \underset{\sim}{X}\underset{\sim}{\hat{\beta}}$.
To find the variance of $\underset{\sim}{\hat{Y}}$ we write

$$\begin{aligned} Var(\underset{\sim}{\hat{Y}}) &= Var\left\{\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}\right\} \\ &= \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'Var(\underset{\sim}{Y}\underset{\sim}{Y}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' \end{aligned} \tag{1.10}$$

Since we know that $Var(\underset{\sim}{Y}\underset{\sim}{Y}') = \sigma^2\underset{\sim}{I}$, substituting this into (1.10) and rearranging terms gives

$$\begin{aligned} Var(\underset{\sim}{\hat{Y}}) &= \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{X})(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\sigma^2\underset{\sim}{I} \\ &= \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\sigma^2\underset{\sim}{I} \end{aligned} \tag{1.11}$$

**Proposition 1.3** The least squares estimate of $\underset{\sim}{\varepsilon}, \underset{\sim}{\hat{\varepsilon}}$, where $\underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{Y} - \underset{\sim}{\hat{Y}}$ is $N(\underset{\sim}{0}, (\underset{\sim}{I} - \underset{\sim}{H})\sigma^2)$, where $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$. To complete the proof of this Proposition, we need the fact that the matrix $\underset{\sim}{I} - \underset{\sim}{H}$ is symmetric and idempotent. A matrix $\underset{\sim}{A}$ is said to be idempotent is $\underset{\sim}{A}^2 = \underset{\sim}{A}$ and $\underset{\sim}{A}$ is symmetric if $A' = A$.

**Proof:** The fact that $\underset{\sim}{\hat{\varepsilon}}$ has a normal distribution follows directly since it is a linear function of $\underset{\sim}{Y}$. In order to find the mean of $\underset{\sim}{\hat{\varepsilon}}$ to consider

$$E(\underset{\sim}{\hat{\varepsilon}}) = E\left\{\underset{\sim}{Y} - \underset{\sim}{\hat{Y}}\right\} = E(\left\{\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\right\}\underset{\sim}{Y}) \tag{1.12}$$

We recall that $E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta}$. Therefore, (1.12) simplifies to

$$E(\underset{\sim}{\hat{\varepsilon}}) = \tilde{X}\underset{\sim}{\beta} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{X}\underset{\sim}{\beta} = 0 \tag{1.13}$$

We find the variance of $\underset{\sim}{\hat{\varepsilon}}$ by first rewriting $\underset{\sim}{\hat{\varepsilon}}$ as

$$\underset{\sim}{\hat{\varepsilon}} = \underset{\sim}{Y} - \underset{\sim}{\hat{Y}} = (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{Y} \tag{1.14}$$

Hence,

$$Var(\underset{\sim}{\hat{\varepsilon}}) = (\underset{\sim}{I} - \underset{\sim}{H})Var(\underset{\sim}{Y}\underset{\sim}{Y}')(\underset{\sim}{I} - \underset{\sim}{H})' \tag{1.15}$$

But $\underset{\sim}{I} - \underset{\sim}{H}$ is symmetric, and $Var(\underset{\sim}{Y}\underset{\sim}{Y}') = \sigma^2\underset{\sim}{I}$, so from (1.15) we get

$$Var(\underset{\sim}{\varepsilon}) = (\underset{\sim}{I} - \underset{\sim}{H})(\underset{\sim}{I} - \underset{\sim}{H})\sigma^2\underset{\sim}{I}\,. \tag{1.16}$$

We recall that $\underset{\sim}{I} - \underset{\sim}{H}$ is idempotent and finally we have

$$Var(\underset{\sim}{\varepsilon}) = (\underset{\sim}{I} - \underset{\sim}{H})\sigma^2$$

We notice that all the estimates just derived are unbiased in that the expected values of their distributions are the parameters they estimate. Now we can find an unbiased estimate of $\sigma^2$. In the next Proposition we demonstrate that such an estimate is $\sigma^2$, where

**Proposition 1.4**

$$\hat{\sigma}^2 = \frac{\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\hat{\beta}}\underset{\sim}{X}'\underset{\sim}{Y}}{(n-p)}$$

is an unbiased estimate of $\sigma^2$. In order to complete the proof of this Proposition we need the following two results from the theory of matrix algebra, and the distribution theory of quadratic forms: For two conforming matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$ trace $(\underset{\sim}{A}\underset{\sim}{B}) =$ trace $(\underset{\sim}{B}\underset{\sim}{A})$. If $\underset{\sim}{A}$ is a symmetric matrix and $\underset{\sim}{e}$ is a vector of uncorrelated random variables with common variance $\sigma^2$ and mean $0$,

$$E(\underset{\sim}{e}'\underset{\sim}{A}\underset{\sim}{e}) = \sigma^2 \text{ trace } (\underset{\sim}{A}).$$

**Proof:** We consider first

$$E\left(\hat{\sigma}^2\right) = \frac{E\left\{\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{Y}'\right\}}{(n-p)} \tag{1.17}$$

Substituting for $\underset{\sim}{\hat{\beta}}'$ gives

$$E\left(\hat{\sigma}^2\right) = \frac{E\left\{\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}\underset{\sim}{Y}\right\}}{(n-p)} \tag{1.18}$$

If we now substitute for $\underset{\sim}{Y}'$ and $\underset{\sim}{Y}$, (1.18) becomes

$$E(\hat{\sigma}^2$$

$$0 = \frac{E\left\{(\underset{\sim}{\beta}'\underset{\sim}{X}' + \underline{\varepsilon}')(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) - (\underset{\sim}{\beta}'\underset{\sim}{X}' + \underset{\sim}{\varepsilon}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}\right\}}{n-p} \tag{1.19}$$

Expanding (1.19), taking expectations and simplifying yields

$$E(\sigma^2) = \frac{E\left\{\underset{\sim}{\varepsilon}'\underset{\sim}{I}\underset{\sim}{\varepsilon} - \underset{\sim}{\varepsilon}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}\right\}}{(n-p)} \tag{1.20}$$

In equation (1.20) we have the expectation of two quadratic forms. Applying the first of two results stated initially, we get

$$E(\hat{\sigma})^2 = \frac{\left\{\sigma^2 \text{ trace}(\underset{\sim}{I}) - \sigma^2 \text{trace } (\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}')\right\}}{(n-p)} \tag{1.21}$$

Since $\underset{\sim}{I}$ is an $nxn$ identity matrix its trace is $n$. Applying the second of the two results, we stated, we find that

$$\text{trace } (\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}') = \text{ trace } (\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}) . \tag{1.22}$$

Because $(\underset{\sim}{X}'\underset{\sim}{X})$ and $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ are $pxp$, trace $(\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}) = p$ .
Substituting back into (1.21) gives

$$E(\hat{\sigma}^2) = \frac{\sigma^2 n - \sigma^2 p}{n-p} = \sigma^2 \tag{1.23}$$

Having found the least squares estimates for our linear regression model and derived their statistical properties, we now would like to demonstrate an important optimality property associated with our estimate $\underset{\sim}{\hat{\beta}}$. Specifically, we want to show that $\underset{\sim}{\hat{\beta}}$ is the best linear unbiased estimate of $\underset{\sim}{\beta}$. That is, in the class of linear unbiased estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{\hat{\beta}}$ has the smallest variance-covariance matrix. This well-known property of the least squares estimates is known as the Gauss-Markov Proposition. It is stated and proved below as Proposition 1.5.

**Proposition 1.5** In the class of linear unbiased estimates of $\underset{\sim}{\beta}$, the least squares estimate $\underset{\sim}{\hat{\beta}}$ has the smallest variance.

**Proof:** Consider an alternative linear estimate of $\underset{\sim}{\beta}$, and call it $\underset{\sim}{\beta}*$, where

$$\underset{\sim}{\beta}* = \underset{\sim}{A}\underset{\sim}{Y}$$

Let $\underset{\sim}{A} = (\underset{\sim}{X}\underset{\sim}{X})^{-1}\underset{\sim}{X} + \underset{\sim}{C}$. We now want to determine $\underset{\sim}{C}$ such that $\underset{\sim}{\beta}*$is unbiased with the smallest variances of all linear estimates. Therefore, we consider

$$\begin{aligned} E(\underset{\sim}{\beta}*) &= \underset{\sim}{A}E(\underset{\sim}{Y}) = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{\beta} + \underset{\sim}{C}\underset{\sim}{X}\underset{\sim}{C}' \\ &= \underset{\sim}{\beta} - \underset{\sim}{C}\underset{\sim}{X}\underset{\sim}{C}' \end{aligned} \tag{1.24}$$

In order for $\underset{\sim}{\beta}*$ to be unbiased we require that $\underset{\sim}{C}\underset{\sim}{X}=0$. Now we want to find the variance of $\underset{\sim}{\beta}*$

$$Var(\underset{\sim}{\beta}*)=E\left\{(\underset{\sim}{\beta}*-\underset{\sim}{\beta})(\underset{\sim}{\beta}*-\underset{\sim}{\beta}')\right\} \tag{1.25}$$

Substituting for $\underset{\sim}{\beta}*$ in (1.25) gives

$$Var(\underset{\sim}{\beta}*)=E\left\{\left[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}+\underset{\sim}{C}\right](\underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{\varepsilon})-\underset{\sim}{\beta}\right\}\left\{\left[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}+\underset{\sim}{C}\right]\left[\underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}\right]-\underset{\sim}{\beta}\right\}' \tag{1.26}$$

Expanding terms, taking expectations and simplifying yields

$$Var(\underset{\sim}{\beta}*)=(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}+\underset{\sim}{C}\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'\underset{\sim}{C}'=\sigma^2\left[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}+\underset{\sim}{C}\underset{\sim}{C}'\right] \tag{1.27}$$

We now notice that

$$Var(\underset{\sim}{\beta}*)-Var(\underset{\sim}{\beta})\geq 0 \tag{1.28}$$

The equality in (1.28) holds if and only if we choose $\underset{\sim}{C}\underset{\sim}{C}'=0$. This implies that the elements of $\underset{\sim}{C}$ must all be zero. We have demonstrated that the least squares estimate of $\underset{\sim}{\beta}$, $\hat{\underset{\sim}{\beta}}$ has the smallest variance-covariance matrix among the class of all linear unbiased estimates. We note however that $\hat{\underset{\sim}{\beta}}$ is not the only estimate in this class with the smallest variance. Moreover, it is possible to construct biased linear estimates or unbiased nonlinear estimates whose variance is smaller than $\hat{\underset{\sim}{\beta}}$. By making similar arguments for $\hat{\underset{\sim}{Y}}$ and $\hat{\underset{\sim}{\varepsilon}}$, we can also prove that they being linear functions of $\hat{\underset{\sim}{\beta}}$, they are the b.l.u.e. of $\underset{\sim}{Y}$ and $\underset{\sim}{\varepsilon}$, respectively.

We now illustrate one final mathematical property of our least squares estimates, which is helpful for understanding the theory behind this method of estimation.

**Proposition 1.6** $\hat{\underset{\sim}{\varepsilon}}'(\hat{\underset{\sim}{Y}})=0$

To prove this Proposition, we need the same result for the matrix $\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}$ that we stated in the proof of Proposition 1.3 for the matrix $\underset{\sim}{I}-\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\tilde{X}'$. That is $\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$ is also symmetric and idempotent.

**Proof:** Again let $\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'=\underset{\sim}{H}$ and recall from (1.14) that $\hat{\underset{\sim}{\varepsilon}}=(\underset{\sim}{I}-\underset{\sim}{H})\underset{\sim}{Y}$. Hence,

$$\hat{\underset{\sim}{\varepsilon}}'\left(\hat{\underset{\sim}{Y}}\right)=\left[(\underset{\sim}{I}-\underset{\sim}{H})\underset{\sim}{Y}\right]'\underset{\sim}{H}\underset{\sim}{Y}=\underset{\sim}{Y}'(\underset{\sim}{I}-\underset{\sim}{H})'\underset{\sim}{H}\underset{\sim}{Y} \tag{1.29}$$

since $\underset{\sim}{\hat{Y}} = \underset{\sim}{H}\underset{\sim}{Y}$ . But both $\underset{\sim}{I} - \underset{\sim}{H}$ and $H$ are symmetric and idempotent, so that

$$\underset{\sim}{\hat{\varepsilon}}'(\underset{\sim}{Y}) = \underset{\sim}{Y}'(\underset{\sim}{H} - \underset{\sim}{H})\underset{\sim}{Y} = 0 \tag{1.30}$$

What this result says is that the vector of estimated errors is orthogonal to the vector of predicted values of the dependent variable. In other words, regression decomposes $\underset{\sim}{Y}$ into two orthogonal components $\underset{\sim}{\hat{Y}}$ and $\underset{\sim}{\hat{\varepsilon}}$. The predicted values of $\underset{\sim}{Y}$ , $\underset{\sim}{\hat{Y}}$, is the orthogonal projection of $\underset{\sim}{Y}$ on to the subspace determined by $\underset{\sim}{X}$ . We illustrate this geometrically in Figure A.1 for the case when n=3 and p=2. We return to this geometric interpretation of regression in this next section, where we discuss the F-test.

## I. The F-Test

Our primary method for testing hypotheses about our regression models is the F-test. In this section, we outline its derivation and give its geometric interpretation.

Let's write $\underset{\sim}{\beta}$ as

$$\underset{\sim}{\beta} = \begin{bmatrix} \underset{\sim}{\beta_1} \\ \underset{\sim}{\beta_2} \end{bmatrix} \quad \text{where } \underset{\sim}{\beta_1} \text{ is a } k \times l \text{ column vector} \tag{2.1}$$

$$\underset{\sim}{\beta_2} \text{ is a } (p-k)\times l \text{ column vector}$$

If we wish to test the hypothesis that $\underset{\sim}{\beta_1} = \underset{\sim}{0}$ we can do so by using the following steps to construct the F-test.

1) Compute the Regression Sum of Squares (SSR$_0$) under $H_0$ that $\underset{\sim}{\beta_1} = \underset{\sim}{0}$ .
2) Subtract (SSR$_0$) from the Regression Sum of Squares (SSR$_A$) obtained under the alternative hypothesis $H_A$ $\underset{\sim}{\beta_1} \neq 0$ .
3) Divide this difference by $k$ , the number of parameters in $\underset{\sim}{\beta_1}$ .
4) Divide this new statistic by the Mean-Squared Error under H$_A$(MSE$_A$).

If the normality assumptions regarding $\varepsilon$ stated in Section I hold, under H$_0$ the ratio

$$\frac{(SSR_A - SSR_0)/k}{MSE_A} \tag{2.2}$$

has an F-distribution with *k* and *n-p* degrees of freedom. We reject H$_0$ at the $\alpha$ level of significance if our observed F-statistic exceeds the $(1-\alpha)$ quantile of the F-distribution with *k, n-*

*p* degrees of freedom. Although we do not prove here that this statistic has an F-distribution, we show geometrically why this ratio is reasonable for the test of this hypothesis.

For the geometric argument, we again consider the case where $n = 3$ $p = 2$, shown in Figure A.2. Here the null hypothesis is that the coefficient of $X_1$ is zero, and therefore $K = 1$.

From Figure A.2 we see that

$$SSR_A = \overline{\overrightarrow{OA}}^2 - n\overline{Y}^2$$

$$SSR_0 = \overline{\overrightarrow{OB}}^2 - n\overline{Y}^2 \qquad (2.3)$$

$$MSE_A = \overline{\overrightarrow{AC}}^2$$

To test the hypothesis that $\beta_1 = 0$, we find that the appropriate F statistic is

$$F_{1,1} = \overline{\overrightarrow{AB}}^2 \Big/ \overline{\overrightarrow{AC}}^2 \qquad (2.4)$$

We notice that $\overrightarrow{AB}$ is the difference between the projections of $Y$ under $H_A$ and $Y$ under $H_0$. This geometric description of the F-test shows that $H_0$ is rejected if $Y$ is significantly closer to the subspace spanned by $X_1$ and $X_2$ than the one spanned by $X_2$ alone.

**G. Generalized Linear Model (GLM)**
**Linear Model**

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

$y = \mu + \varepsilon$ where $\mu$ is the mean of

$y$ $n \times 1$ $X$ $n \times p$ $\beta$ $p \times 1$

**Components of the Generalized Linear Model**

1. Random Component for $y$ is chosen from the exponential family
2. Systematic Component are covariates $x_1, x_2, \ldots, x_p$ produce a linear predictor $\eta$ given by

$$\eta = \sum_{j=1}^{p} x_j \beta_j \ .$$

3. The link between the random and systematic components $\eta = g(\mu)$ where $g$ is monotonic differentiable function.

**Revisit the Exponential Family of Distribution**

Given $y$ a data vector from a probability distribution belonging to the exponential family. We may express the probability density or mass function of $y$ as

$$f_y(y,\theta,\phi) = \exp\left\{y\theta - b(\theta))/a(\phi) + c(y,\phi)\right\}.$$

$a(\,), b(\,)$ and $c(\,)$ functions to be specified if $\phi$ is known then we have an exponential family model with a canonical parameter.

**Example 2: Gaussian Model**

$$f_y(y,\theta,\phi) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-(y-\mu)^2/2\sigma^2\right\}$$

$$= \exp\left\{(y\mu - \mu^2/2)/2\sigma^2 - \frac{1}{2}(y^2/\sigma^2 + \log(2\pi\sigma^2))\right\},$$

where $\theta = \mu, \quad \phi = \sigma^2, \quad b(\theta) = \theta^2/2, \quad c(y,\phi) = -\frac{1}{2}\left\{y^2/\sigma^2 + \log(2\pi\sigma^2\right\}$.

**Example 3: Binomial Model**

$$\Pr(Y = y) = \frac{n!}{(n-y)!\,y!} p^y (1-p)^{n-y}$$
$$= \exp\left(y \log\left(\frac{p}{1-p}\right) + n\log(1-p) + \log\left(\frac{n!}{(n-y)!\,y!}\right)\right)$$

where $\theta = \log(p/(1-p)), \quad \phi = 1, \quad b(\theta) = n\log(1-p), \quad c(y,\phi) = \log\left(\frac{n!}{(n-y)!\,y!}\right)$

**Example 4: Poisson Model**

$$\Pr(Y = y) = \lambda^y \exp(-\lambda)/y!$$
$$= \exp\left(y\log\lambda - \lambda - \log y!\right)$$

where $\theta = \log\lambda, \quad \phi = 1, \quad b(\theta) = -\lambda, \quad c(y,\phi) = \log y!$

**Mean and Variance of $y$ (We will skip this stuff in the lecture; we can revisit it in a tutorial).**

Let $\ell(\theta,\phi,y) = \log f_y(y,\theta,\phi)$. Two standard properties of likelihood functions

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = 0 . \qquad (1)$$

The expected value of the score function is zero

$$E\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) + E\left(\frac{\partial \ell}{\partial \theta}\right)^2 = 0 . \qquad (2)$$

The expected Fisher Information is the negative of the variance of the score function

$$\ell(\theta, y) = \{y(\theta) - b(\theta)\} / a(\phi) + c(y, \phi) , \qquad (3)$$

whence

$$\frac{\partial \ell}{2\theta} = \{y - b'(\theta)\} / a(\phi) \qquad (4)$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -b''(\theta) / a(\phi) . \qquad (5)$$

Eqs. (1) and (4) imply

$$0 = E\left(\frac{\partial \ell}{\partial \theta}\right) = \{\mu - b'(\theta)\} / a(\phi)$$

or

$$E[y] = \mu = b'(\theta) .$$

Using Eqs. 4 and 5 in Eq. 2, we obtain

$$0 = -\frac{b''(\theta)}{a(\phi)} + \frac{Var(y)}{a^2(\phi)} ,$$

Hence

$$Var(y) = b''(\theta) a(\phi) .$$

The variance depends on the mean and a function, which depends on the dispersion parameter; $b''(\theta)$ is the variance function.

**Model Summary Table**

| | **Gaussian** | **Poisson** | **Binomial** | **Gamma** | **Inverse Gaussian** |
|---|---|---|---|---|---|
| **Canonical link:** $g(\mu)$ | $\mu$ | $\log\lambda$ | $\log\left(\frac{p}{1-p}\right)$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |

**MAIN POINT.** **WE MAY WRITE EXPLICITLY THE LINK OF THE SYSTEMATIC AND RANDOM STRUCTURE AS $g(\mu) = X\beta$. IN THIS WAY, WE EXTEND THE SIMPLE REGRESSION MODEL WITH GAUSSIAN ERRORS TO REGRESSION MODELS FOR THE ENTIRE EXPONENTIAL FAMILY OF PROBABILITY DISTRIBUTIONS.**

### Deviance (Goodness of Fit Criterion)

Let $\ell(\mu;y) = \sum_{i=1}^{N} \log f(y_i;\theta_i)$ and define the deviance as

$$D^*(y;\mu) = 2\ell(y;y) - 2\ell(\mu;y)$$

Note that the second expression is the –2log likelihood of the data. Hence minimizing $D^*(y;\mu)$ is equivalent to maximizing the log likelihood. Notice that in the Gaussian case, the deviance is

$$D^*(y;\hat{\mu}) = \sum_{i=1}^{N} (y_i - \hat{\mu}_i)^2 / \sigma^2$$

which is just the residual sum of squares (RSS).

**Analysis of Deviance.** Instead of analysis of variance we use analysis of deviance, which is additive for nested sets of models fit by maximum likelihood. $D^*(y;\mu)$ is analyzed as an approximately chi-squared quantity.

**Deviance Residuals.** The deviance residuals are defined as $D^*(y;\hat{\mu}) = \sum_{i=1}^{N} d_i$ where we define $r_{d,i} = sign(d_i)d_i^{\frac{1}{2}}$. Plots of these quantities can be analyzed in the same way the regression residuals are analyzed.

For generalized linear models in which there is a natural time dependence in the observations we will show that a global measure of model goodness of fit may be constructed using the time-rescaling Proposition (Berman, 1983; Ogata, 1988; Brown et al., 2001).

**Remarks**

1. GLM allows generalization (extension) of linear models to non-Gaussian setting.
2. A unified computational approach.
3. Can use non-canonical link and variance functions.
4. Widely distribution Splus and non Matlab.
5. Now there's Bayesian GLM.
6. Assess goodness-of-fit with an analysis of deviance.

**Neuroscience Data Analysis Examples Using GLM**

1. **Random Threshold Model (Brillinger (1988, 1992))**

Simplified version

$$\log\left[\frac{p_x t}{1-p_x t}\right] = \mu + \sum_{j=1}^{J} \alpha_j x_{t-j} + \sum_{k=1}^{k} \beta_k y_{t-k}$$

$$x_t = \begin{cases} 1 & \text{spike from neuron } x \, at \ t \\ 0 & \text{otherwise} \end{cases}$$

$$y_t = \begin{cases} 1 & \text{spike from neuron } y \text{ at } t \\ 0 & \text{otherwise} \end{cases}.$$

$p_x t$ probability of a spike at $t$ from neuron $x$.

2. **Prospective and Retrospective Encoding by Hippocampal Place Cells (Frank, Brown and Wilson, 2000)**

A rat running on a W-maze, multiple single unit activity is recorded from CA1 region of the hippocampus and the entorhinal cortex.

$t$ indexes trial

$P_t$ probability of going right on an outbound journey from the center arm of the W-maze.

$$\log\left[\frac{p_t}{1-p_t}\right] = \mu + \sum_{j=1}^{J} \alpha_j \lambda_j(t) ,$$

where $\lambda_j(t)$ is the firing rate of the neuron at location $j$ prior to the choice point on pass $t$.

3. **Cockroach Cercal System Response to Wind Stimuli (Davidowitz and Rimberg, 2000)**.

$n_t =$ number of spikes at time $t$ (discrete time bins of 10msec)

| | |
|---|---|
| $v_x(t)$ | $x$ – direction wind velocity |
| $v_y(t)$ | $y$ – direction wind velocity |
| $a_x(t)$ | $x$ – direction wind acceleration |
| $a_y(t)$ | $y$ – direction wind acceleration |

Poisson link function

$$\log \lambda(t) = \mu + \alpha_1 v_x(t) + \alpha_2 v_y(t) + \beta_1 a_x(t) + \beta_2 a_y(t) \,,$$

$\lambda(t)$ spike rate at time $t$ .

**References**

Draper NR, Smith H. (1981). *Applied Regression Analysis*, 2nd ed., Wiley: New York.

McCullagh, P. Nelder JA (1989). *Generalized Linear Models*, 2nd ed, Chapman and Hall: London.