CHAPTER 1

Applied Linear Statistical Models 5th ed. Neter, Li, Kutner, Nachtsheim

Functional Relationship between Two Variables

- Expressed by a mathematical formula.
- If X denotes the independent variable and Y the dependent variable, a functional relation is of the form: $Y = f(X)$
- Given a particular value of X, the function f indicates the corresponding value of Y.

Statistical Relationship Between Two Variables

- Not a perfect relationship.
- In general, the observations for a statistical relation do not fall directly on the curve of a relationship

Regression Analysis serves three major purposes: (1) description (2) control (3) Prediction

Model

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Y_i is the value of the response in the i^{th} trial
β_0 and β_1 are parameters
X_i is a known constant, the value of the predictor variable in i^{th} trial
ε_i is a random error term with mean $E[\varepsilon_i] = 0$ and variance $\sigma^2[\varepsilon_i] = \sigma^2$, ε_i and ε_j are uncorrelated so that their covariance is zero for all i, j where $i \neq j$, $i = 1 \ldots n$

Important Features of Model

1 The response Y_i in the i^{th} trial is the sum of two components, (1) the constant term $\beta_0 + \beta_1 X_i$ and (2) the random term ε_i. Hence Y_i is a random variable

2 Since $E[\varepsilon_i] = 0$, it follows:
$E[Y_i] = E[\beta_0 + \beta_1 X_i + \varepsilon_i] = \beta_0 + \beta_1 X_i + E[\varepsilon_i] = \beta_0 + \beta_1 X_i$

2

Thus, the response Y_i, when the level of X in the i^{th} trial is X_i, comes from a probability distribution whose mean is

$$E[Y_i] = \beta_0 + \beta_1 X_i$$

Therefore the regression model is $E[Y] = \beta_0 + \beta_1 X_i$. Since the regression function relates the means of the probability distributions of Y for a given X to the level of X

3. The response Y_i in the i^{th} trial exceeds or falls short of the value of the regression function by the error term ε_i

4. The error terms ε_i are assumed to have constant variance σ^2. It therefore follows that the responses Y_i have the same constant variance:

$$\sigma^2[Y_i] = \sigma^2$$

we have $\sigma^2[\beta_0 + \beta_1 X_i + \varepsilon_i] = \sigma^2[\varepsilon_i] = \sigma^2$

Thus the regression model assumes that the probability distributions of Y have the same variance σ^2, regardless of the level of the predictor variable X.

5. The error terms are assumed to be uncorrelated. Since the error terms ε_i and ε_j are uncorrelated, so are the responses Y_i and Y_j.

6. In summary, model implies that the responses Y_i come from probability distributions whose means are $E[Y_i] = \beta_0 + \beta_1 X_i$ and whose variance are σ^2, the same for all levels of X. Y_i, Y_j - uncorrelated

Meaning of Regression Parameters

- β_0 and β_1 are called regression coefficients
- β_1 is the slope of the regression line.
- It indicates the change in the mean of the probability distribution of Y per unit increase in X.
- β_0 is the Y intercept of regression line
- When $X=0$, β_1 gives the mean probability distribution of Y at $X=0$

- When scope of model does not cover $X=0$, β_0 does not have any particular meaning as a seperate term in the regression model

Method of least squares

The method of least squares considers the deviation of Y_i from its expected value: $Y_i - (\beta_0 + \beta_1 X_i)$

In particular, the method of least squares requires that we consider the sum of n squared deviations. Denoted by Q

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

According to the method of least squares, the estimators of β_0 and β_1 are those values b_0 and b_1 that minimise the criterion Q for the given sample observations

$$b_1 = \frac{\Sigma(x_i - \bar{x})\Sigma(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Obtained by differentiating and equating to zero and manipulating

Properties of Least Squares Estimate

Markov theorem states: Under the conditions of regression model, the least squares estimators b_0 and b_1 are unbiased and have minimum variance among all unbiased linear estimators

Hence. $E[b_0] = \beta_0$ and $E[b_1] = \beta_1$

$$\hat{Y} = b_0 + b_1 X_i$$

We call a value of the response variable a response and $E[Y]$ the mean response

Residuals

- The i^{th} residual is the difference between the observed value Y_i and the corresponding fitted value $\hat{Y}_i$.
- The residual denoted by e_i and is defined as:

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

- From the model → $e_i = Y_i - (b_0 + b_1 X_i)$
$= Y_i - b_0 - b_1 X_i$

- We need to distinguish between the model error term value $\varepsilon_i = Y_i - E[Y_i]$ and the residual $e_i = Y_i - \hat{Y}_i$

- ε_i is the vertical deviation of Y_i from the ~~exact~~ unknown true regression line and hence is unknown

- The residual is the vertical deviation of Y_i from the fitted value $\hat{Y}_i$ on the estimated regression line and it is known

Properties of Fitted Regression Line

1. The sum of the residuals is zero: $\sum_{i=1}^{n} e_i = 0$

2. The sum of the squared residuals $\sum e_i^2$ is a minimum. This was the requirement to be satisfied in deriving the least squares estimates of the regression parameters since the criterion Q to be minimized equals $\sum e_i^2$ when the least squares estimates b_0 and b_1 are used for estimating β_0 and β_1

3. The sum of the observed values Y_i equals the sum of the fitted values $\hat{Y}_i$:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

It follows that the mean of the fitted values $\hat{Y}_i$ is the same as the mean of the observed values Y_i namely $\bar{Y}_i$

5.

4. The sum of the weighted residuals is zero when the residual in the i^{th} trial is weighted by the level of the predictor variable in the i^{th} trial:

$$\sum X_i e_i = 0$$

5. As a consequence of previous properties, the sum of the weighted residuals is zero when the residual in the i^{th} trial is weighted by the fitted value of the response variable for the i^{th} trial.

$$\sum \hat{Y}_i e_i = 0$$

6. The regression line always goes through the point $(\bar{x}, \bar{y})$

Point Estimator of σ^2

For single population it is $s^2 = \frac{\sum(y_i - \bar{y})^2}{n-1}$

In regression model:

$$y_i - \hat{y}_i = e_i$$

The appropriate sum of squares $SSE = \sum(y_i - \hat{y}_i)^2 = \sum e_i^2$

SSE = error sum of squares or residual sum of squares

SSE has n-2 degrees of freedom. 2 are lost because β_0 and β_1 are estimated used in obtaining $\hat{y}_i$.

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2} = \frac{\sum e_i^2}{n-2}$$

MSE is error mean square or residual mean square

$$E[MSE] = \sigma^2$$

6.

Estimation of Parameters by Method of Maximum Likelihood

When the functional form of the probability distribution of the error term is specified, estimators of the parameters β_0, β_1 and σ^2 can be obtained by method of maximum likelihood

The method chooses as estimates those values of the parameters that are most consistent with the sample data

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$

Parameter.	Max likelihood Estimator
β_0	Same as LS
β_1	Same as LS
σ^2	$\hat{\sigma}^2 = \frac{\sum(y_i - \hat{y}_i)^2}{2}$

The maximum likelihood estimator $\hat{\sigma}^2$ is biased normally use MSE

CHAPTER 2

Inference Concerning β_1

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

- The reason for testing $\beta_1 = 0$ is that when $\beta_1 = 0$ there is no linear association between Y and X

- Implies Our only is there no linear association between Y and X but also that there is no relation of any type between Y and X since the probability distribution of Y are then identical at all levels of X

Sampling Distribution of b_1

The sampling distribution of b_1 refers to the different values of b_1 that would be obtained with repeated sampling when the level of the predictor variable X are held constant from sample to sample

For normal error regression model, the sampling distribution of b_1 is normal, with mean and variance

$$E[b_1] = \beta_1$$

$$\sigma^2[b_1] = \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2} = \frac{MSE}{\Sigma(X_i - \bar{X})^2}$$

Sampling Distribution of b_0

Sampling distribution of b_0 refers to the different values of b_0 that would be obtained with repeated sampling when the level of the predictor variable X are held constant from sample to sample

$$E[b_0] = \beta_0$$

$$\sigma^2[b_0] = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i - \bar{X})^2}\right]$$

$\frac{b_0 - \beta_0}{s[b_0]}$ is distributed as $t(n-2)$

$$b_0 \pm t(1-\alpha/2; n-2)\, s[b_0]$$

R^2

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

The measure R^2 is called the coefficient of determination. Since $0 \le SSE \le SSTO$ it follows: $0 \le R^2 \le 1$.

We may interpret R^2 as the proportionate reduction of total variation associated with the use of the predictor variable X

Thus, the larger R^2 is, the more the total variation of Y is reduced by introducing the predictor variable X.

The closer it is to 1, the greater is said to be the degree of linear association between X and Y.

Properties of Residuals

Mean: The mean of n residuals e_i for a simple linear regression model $\bar{e} = \frac{\sum e_i}{n} = 0$

Where $\bar{e}$ denotes the mean of the residuals. Thus since $\bar{e}$ is always 0, it provides no information as to whether the true errors ε_i have expected value $E[\varepsilon_i] = 0$

Variance: The variance of the n residuals e_i is defined as:

$$s^2 = \frac{\sum (e_i - \bar{e})^2}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{SSE}{n-2} = MSE$$

If the model is appropriate, MSE is an unbiased estimator of variance of the error terms σ^2.

3

Non independence The residuals e_i are not independent random variable because they involve the fitted values $\hat{Y}_i$ which are based on the same fitted regression function.

As a result, subject to two constraints, the sum of $e_i = 0$ and products $X_i e_i$ must sum to 0.

1 Solve 1a for b_0

First we get rid of (-2) by multiplying both sides by $(-\frac{1}{2})$

$$0 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)$$

Next distribute the summation operator through all of the terms in the expression in parentheses

$$0 = \sum_{i=1}^{n}(y_i) - \sum_{i=1}^{n}(b_0) - \sum_{i=1}^{n}(b_1 x_i)$$

Bring middle summation to LHS

$$\sum_{i=1}^{n}(b_0) = \sum_{i=1}^{n}(y_i) - \sum_{i=1}^{n} b_1 x_i$$

Because b_0 and b_1 are constant ⇒

$$N b_0 = \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i$$

To isolate b_0 on the LHS, divide by N

$$b_0 = \left(\frac{\sum_{i=1}^{n} y_i}{N}\right) - b_1\left(\frac{\sum_{i=1}^{n} x_i}{N}\right) \quad (1c)$$

This equation 1c will be handy later.
Note that first term $\left(\frac{\sum y_i}{N}\right)$ is the mean of y_i or $\bar{y}$
likewise $\frac{\sum x_i}{n}$ is mean of x_i or $\bar{x}$
We get:

$$b_0 = \bar{y} - b_1 \bar{x}$$

2 Solve 1b for b_1 $\quad 0 = -2\sum_{i=1}^{N} x_i (y_i - b_0 - b_1 x_i)$

Get rid of (-2) by multiplying across by $(-\frac{1}{2})$

$$0 = \sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i)$$

Distribute x_i through parenthesis

$$0 = \sum_{i=1}^{N} (y_i x_i - b_0 x_i - b_1 x_i^2)$$

Regression

3\. One way of obtaining estimates b_0, b_1 of β_0, β_1 is by minimising:

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

in terms of β_0, β_1 where (b_0, b_1) is the value of (β_0, β_1) corresponding to minimal value of S.

Here S is called the sum of squares of errors and this method is called the least square method.

Under certain general conditions, the estimators (b_0, b_1) obtained this way also turn out to be the minimum unbiased estimators as well as the maximum likelihood estimators

We can determine (b_0, b_1) by differentiating S with respect to (β_0, β_1) setting them to 0 and solving for (β_0, β_1) giving

$$dS/d\beta_0 = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$dS/d\beta_1 = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1) x_i = 0$$

(5)

Resulting in

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}.$$

See next part for derivation of two formulas:

We have: $0 = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$ (1a)

and $0 = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1) x_i = 0$ (1b)

5.

Distribute the summation operator through all of the terms in the eqn

$$0 = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} b_0 x_i - \sum_{i=1}^{n} b_1 x_i^2$$

Bring constants in front of sums:

$$0 = \sum_{i=1}^{N} y_i x_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2$$

Take third term to LHS

$$b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - b_0 \sum_{i=1}^{n} x_i$$

Retrieve value for b_0 equation 1C

$$b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - \left[\left(\frac{\sum y_i}{N}\right) - b_1\left(\frac{\sum x_i}{N}\right)\right] \sum x_i$$

Multiplying out the last term on the right we get:

$$b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{N} - b_1 \left(\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{N}\right)$$

Take last term to LHS

$$b_1 \sum x_i^2 + b_1 \frac{(\sum x_i)^2}{N} = \sum y_i x_i - \frac{\sum y_i \sum x_i}{N}$$

Factor out b_1

$$b_1 \left[\sum x_i^2 + \frac{(\sum x_i)^2}{N}\right] = \sum y_i x_i - \frac{\sum y_i \sum x_i}{N}$$

Divide both sides by eqn in large bracket and isolate b_1

$$b_1 = \frac{\sum y_i x_i - \frac{\sum y_i \sum x_i}{N}}{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}$$

1. Regression Extra Notes

How some sets of variables effect others

Assume a functional, parametric relationship between the variables, typically linear in unknown parameters which are to be estimated from the available data

Two sets of ~~pattern~~ variable can be distinguished: Predictor variables and response

Predictor variables are those that can be either set to a desired value (controlled) or else take values that can be observed without any error.

Our objective is to find out how changes in the predictor variables effect the values of the response variables

Predictor variables: - Input Variables
- x-variables
- Regressors
- Independent variables

Response Variable: - Output Variable
- y Variable
- Dependent Variable

We shall be concerned with relationships of the form:

Response variable = Linear model function in terms of input variables + random error

In the simplest case we have data $(y_1, x_1)(y_2, x_2)\ldots(y_n, x_n)$ the linear function of the form: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \ldots N$ (1)

7. Regr?

If further (X,Y) have joint normal distribution, then $E[Y|X=x]$ is a linear function of the form

$$E[Y|X=x] = \beta_0 + \beta_1 x$$

Hence $\hat{y}_i = b_0 + b_1 x_i$ can and should be seen as $\hat{E}[Y|X=x_i]$, the estimator of this conditional mean

2 ε_i and ε_j for all $i \neq j$ are uncorrelated, hence y_i, y_j are also uncorrelated

3 $\varepsilon_i \sim N(0, \sigma^2)$ thus they are independent. Hence:

$$Y | X = x_i = x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

and y_i are independent but not identically distributed random variables (with some abuse in notation, let

$$y_i = (Y|X=x_i) \sim N(\beta_0 + \beta_1 x_i, \sigma^2))$$

Note that in regression model, we specify the conditional distribution of Y given $X = x_i$; full inference on (Y,X) would require the specification of the joint distribution for (Y,X)

2.2. Some Distributional Theory

When examining the regression equation, we will need to test various hypotheses which in general will depend on the distributional properties of sums of squares of independent normal variables and their ratios

We give a brief summary of distributional results of the quadratic form

1. Normal density: Random variable X has a normal distribution with mean μ and variance σ^2 if it has the density

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

8.
For $-\infty \leq x \leq \infty$. $Z = \left(\frac{X-\mu}{\sigma}\right)$ transforms X to a standard normal variable with mean 0 and variance 1

2 (central) t-distribution: X has a t-distribution with v degrees of freedom (denote by $t(v)$) if it has density:

$$f_v(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$$

for $-\infty \leq t \leq \infty$. Here, $\Gamma(q) = \int_0^\infty e^{-x}x^{q-1}\,dx$ is the gamma function.

In general t-distribution looks like a normal distribution with heavy tails. As $v \to \infty$, t-distribution tends to the normal distribution and in fact $t(\infty) = N(0,1)$. For practical purposes, when $v \geq 30$ they are equal

3 X is said to have a χ^2 distribution with n degrees of freedom $(\chi^2(n))$ if it has the density function:

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{n/2}}\, e^{-\frac{x}{2}} e^{\frac{n}{2}-1} \qquad \text{for } 0 < x \leq \infty$$

How do these distributions appear in regression analysis?
Most of the tests of hypothesis as well as estimators of model parameters will depend on sums of squares of independent, normally distributed random variables and their ratios, and these sums usually have χ^2 distribution, whereas the ratio of independent random variables with χ^2 distribution have F distribution

2

is used to relate y to x. We will also write down this model in general terms as:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Here ϵ is a random quantity measuring the error of any included y may fall off the regression line, or is a random quantity measuring the variation in y not explained by x.

We also assume that the input variable x is either controlled or measured without error, thus is not a random variable (As long as error in measurement that may exist in x is smaller than the measurement error in y, then this assumption is fairly robust.)

If the relation between y and x is more complex than the linear relationship given in (1) then models of the form:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p + \epsilon \qquad (2)$$

can be used

We say the model is linear in the sense that the model is linear in parameters.

For example: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots \beta_p x^p + \epsilon$ is linear

Whereas: $y = \sigma^2 \theta + a_1 x_1^{\beta} x_2^{\gamma} + \epsilon$ is not.

2. Straight Line

Suppose we observe the data $(y_1, x_1), (y_2, x_2), \dots (y_n, x_n)$ and we think the model (1)

$$y_i = \beta_0 + \beta_1 x_i + \epsilon \qquad i = 1, 2, \dots n$$

is the right model.

Here β_0, β_1 are fixed but unknown model parameters to be estimated from our data

9 Regression

1 If $X_1 \dots X_n$ are independent normally distributed random variables with means $(\mu_1, \mu_2, \dots \mu_n)$ and common variance σ^2 then $\sigma^{-2} Z^t A Z$ where $Z^t = (x_1 - \mu_1, \dots x_n - \mu_n)$ and A is any symmetric matrix with $r = tr(A)$, has a (central) χ^2 distribution with r degrees of freedom.
Sum of diagonal elements of any symmetric matrix A is called the trace of the matrix and denoted by $tr(A)$

2 In particular, $\sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma^2}$

has $\chi^2(n)$ distribution, whereas, if $\mu_i = \mu$ constant and is estimated by $\bar{x}$ then $\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2}$

has $\chi^2(n-1)$ distribution (this is due to the loss of one degree of freedom in estimating the mean μ by $\bar{X}$)

3 Ratio of two independent χ^2 random variables each divided by their respective degrees of freedom has an F dist. That is if $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ and (X, Y) are independent then $F_{m,n} = \frac{(X/m)}{(Y/m)}$ has a F distribution with m, n degrees of freedom

4 If $X \sim N(\mu, \sigma^2)$ and $Y \sim \chi^2(n)$ and if further X and Y are independent, then $t = \frac{(x-\mu)/\sigma}{\sqrt{Y/n}}$ has a t distribution with n degrees of freedom. We see that $t^2 = \frac{(X-\mu)^2/\sigma^2}{Y/n}$ has F dist with $(1, n)$ degrees of freedom. This dist appears when we want to look at dist of the form $(X-\mu)/\sigma$ when X has a normal dist, but σ is not known and is substituted by the emperical standard deviation

11

The test of hypothesis $H_0: \beta_1 = \beta_1^*$ vs $H_1: \beta_1 \neq \beta_1^*$
can be performed by calculating the test statistic

$$t = \left(\frac{b_1 - \beta_1^*}{s}\right)\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

s = standard deviation
α = significance level

and comparing $|t|$ with table value $(t_{n-2}, 1-\frac{1}{2\alpha})$

Standard error of b_0 can be similarly calculated:

$$s.e(b_0) = \left(\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}\right)\sigma$$

hence $(1-\alpha)100\%$ confidence interval for β_0 is given by

$$b_0 \pm t(n-2, 1-\tfrac{1}{2\alpha})\left(\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}\right)s$$

and the test $H_0: \beta_0 = \beta_0^*$ $H_1: \beta_0 \neq \beta_0^*$
can be performed by comparing the absolute value of

$$t = \left(\frac{b_0 - \beta_0^*}{s}\right)\Big/\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}$$

with $t(n-2, 1-\frac{1}{2\alpha})$

$$\cancel{\sum (X_i - \bar{x})^2}$$
$$\cancel{\sum x_i^2 - 2x_i\bar{x} + \bar{x}^2}$$
$$\cancel{\sum x_i^2 - 2x_i \frac{\sum x_i}{N} + \frac{\sum x_i}{N}}$$

Fact that $\bar{x} = \frac{\sum x_i}{n}$

$n\bar{x}_n = \sum x_i$

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

2.3 Confidence intervals and tests of hypotheses regarding (β_0, β_1)

A simple calculation shows that $b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

hence $V(b_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

In general σ^2 is not known, then a suitable estimator for σ^2 can replace σ^2. If the assumed model (1) is correct, then it is known that under the normality assumption on residuals

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

is the minimum variance unbiased estimator of σ^2

Hence under the assumption that ε_i are normal, we can construct the usual $100(1-\alpha)\%$ interval for β_1:

$$b_1 \pm \frac{t(n-2, 1-\frac{1}{2}\alpha)S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Here $t(n-2, 1-\frac{1}{2}\alpha)$ is the $1-\frac{1}{2}\alpha$ percentage point of a t-dist with $n-2$ degrees of freedom

left to us to clarify that $\frac{b_1-\beta_1}{s.e(b_1)}$ has a t-distribution with $n-2$ df.

s.e. stands for standard error.

Here $\bar{x} = \frac{1}{n}\sum_{i=1} x_i$ is the sample average and (5) are called the normal equation.

We will call $\hat{y}_i = b_0 + b_1 x_i$ the fitted value of y at $x = x_i$ and $\hat{\varepsilon}_i = y_i - \hat{y}_i$ the residual for the i^{th} observation.

Note that $\hat{y}_i = \bar{y} + b_1(x_i - \bar{x})$ Sub in b_0 value

So that $\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

2.1 Examining The Regression Equation

So far we made no assumption involving the probability of ε and hence of y

Basic assumption of model (1):

1. ε_i, $i = 1..n$ are identically distributed uncorrelated random variables with mean $E(\varepsilon_i) = 0$ and variance $V(\varepsilon_i) = \sigma^2$, so that (assuming x variables are measured without error or controlled) y_i are random variables with $E(y_i) = \beta_0 + \beta_1 x_i$ and $V(y_i) = \sigma^2$

In fact correct notation for $E(y_i)$ is $E(Y|X = x_i)$

If both variable X as well as the response variable Y are random variables, then the random variable $f(X)$ that minimises

$~~E[(Y-f(X))^2]~~$

$E[(Y-f(X))^2 | X]$

is given by $E[Y|X]$