Module Code	ST3451
Module Name	APPLIED LINEAR STATISTICAL METHODS I
Module Short Title	STATISTICAL METHODS 1
ECTS weighting	5
Semester/term taught	Michaelmas
Contact Hours	3 hours per week, some of which will be tutorials
Module Personnel	Jason Wyse wyseja@tcd.ie
Learning Outcomes	When students have successfully completed this module they should be able to:

1. Derive and apply estimators, tests and confidence intervals for the parameters for a range of linear regression models.
2. Derive and construct ANOVA tables.
3. Examine the fit of a regression model through regression diagnostics and test the assumptions of the model.
4. Build an appropriate linear regression model for a given data set.

Module Learning Aims The student will learn about the simple linear regression (SLR) model in detail. This will include derivation of least squares estimators and their properties, sampling distributions of the estimators in the case of Gaussian errors, and tests of significance. The student will also learn about ANOVA- decomposition of the error sum of squares. The matrix approach to linear regression will follow where multiple regression will be discussed. Various diagnostics of fit will be explored, with illustration of how these can be used in practice. Some modifications of the usual regression model will be discussed as well as model building through

variable selection.

Module Content

1. Simple linear regression
2. Multiple regression
3. Regression diagnostics
4. Variable selection

Recommended Reading List

- A second course in statistics regression analysis- Mendenhall & Sincich
- Classical & modern regression with applications- R. H. Meyers
- Introduction to linear regression analysis- Montgomery, Peck & Vining
- Applied regression analysis- Draper & Smith

Assessment Details Final exam (90%) and assignments (10%)

Academic Year of Data 2015/16

28/04/15

ALSM I

1.0 SIMPLE LINEAR REGRESSION

Basic Model

- Let y be a random variable which follows some mean μ and variance σ^2
- We can write y as $y = \mu + \varepsilon$ where μ is a fixed parameter and ε is a r.v.

$\mathbb{E}[\varepsilon] = 0$ $\quad$ $\mathrm{Var}[\varepsilon] = \sigma^2$

$y = \mu + \varepsilon$ $\quad$ $\mathbb{E}[y] = \mathbb{E}[\mu] + \mathbb{E}[\varepsilon] = \mu$

$\mathrm{Var}[y] = \mathrm{Var}[\varepsilon] = \sigma^2$

- If $y \sim N(\mu, \sigma^2)$ then $\varepsilon \sim N(0, \sigma^2)$, then the term ε is refered to as the error
- We could write $\varepsilon = y - \mu$ so that ε is the random deviation of y about it's mean

- Suppose there is an iid sample $y_1, \ldots, y_n$
- We could use this basic model by writing $y_i = \mu + \varepsilon_i \quad i = 1, \ldots, n$
- When $y_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $\varepsilon_i \sim N(0, \sigma^2)$ (by independence)
- Then using this model is equivalent to situations where a one sample t-test is appropriate for testing the hypothesis: $H_0: \mu = 100$ vs $H_1: \mu \neq 100$
- Gives a $100(1-\alpha)\%$ CI for the mean:

$$\bar{y} \pm \left(t_{n-1, 1-\alpha/2}\right)\left(s/\sqrt{n}\right)$$

Estimators: parameter $\rightarrow$ Estimator

$\mu \rightarrow \bar{y}$

$\sigma^2 \rightarrow s^2$

With $y_1, \ldots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad$ $\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ $\quad$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Simple model $y_i = \mu + \varepsilon_i \quad i = 1, \ldots, n$

$\mathbb{E}[\varepsilon_i] = 0$ $\quad$ $\mathrm{Var}[\varepsilon_i] = \sigma^2$

28/09/15

Estimator for μ: $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Estimator for σ^2: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$\hat{y}_i = \hat{\mu} + \hat{\varepsilon}_i \quad <=> \quad \hat{\varepsilon}_i = y_i - \hat{\mu}$

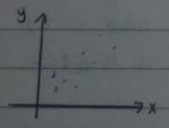
$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\hat{\varepsilon}^2$

Simple linear Regression Model

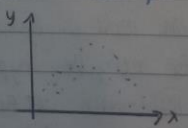
- When we have one independent variable, we call the model Simple linear regression SLR
- Dependent variable y_i (continuous)
- Independent variable x (continuous)
- x is incorporated into the model as a predictor or explanatory variable
  - → It explains "in some way" the value of observed y
  - → The values of the independent variable are assumed known (x is not a r.v.)

- For a given value of x, we consider y to be distributed with mean $E[y|x] = \mu_{y|x}$ i.e. $y = \mu_{y|x} + \varepsilon$

- Now, how does $\mu_{y|x}$ depend on x?
- Collect some data (x_1, y_1) ... (x_n, y_n) and create a scatter plot of x and y



$\mu_{y|x} = \beta_0 + \beta_1 x$



$\mu_{y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$

- Simplest model is SLR model; $\mu_{y|x} = E[y|x] = \beta_0 + \beta_1 x$
- β_0 and β_1 are fixed constant (parameters) → call these regression coefficients
  - → β_0: intercept parameter - mean of y when x is zero
  - → β_1: Slope parameter - change in <u>mean</u> value of y for a one unit increase in x
    $\mu_{y|x+1} - \mu_{y|x} = \beta_0 + \beta_1(x+1) - (\beta_0 + \beta_1 x) = \beta_1$

Note: When we say linear models, we mean linear in the parameters
eg. $y_i = e^x + \varepsilon_i$ is not linear

28/09/15

ALSM 1

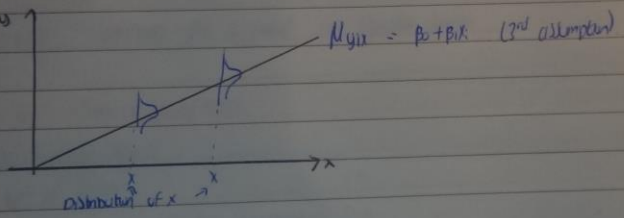
Assumptions:

1. For a fixed value of x, y is a random variable with a finite mean and variance:
   $|E[y|x]| < \infty$ $\quad Var[y|x] < \infty$ (i.e. it exists)

2. The values of y are uncorrelated

3. The model is linear i.e. $E[y|x] = \beta_0 + \beta_1 x_i$

4. Homoscedacity - Variance of y is not dependent on x, i.e. variance is constant.
   $Var[y|x] = Var[y] = \sigma^2$



Statement of the model:

- Data $(x_1, y_1) \dots (x_n, y_n)$
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, \dots, n$
- $E[\varepsilon_i] = 0 \quad \forall i$
- $Var[\varepsilon_i] = \sigma^2 \; \forall i$ (Homoscedastic)
- $Cov[\varepsilon_i, \varepsilon_j] = 0 \quad$ if $i \neq j$ (y_i, y_j uncorrelated)

Models as data descriptors

- In data analysis, we use a model to describe the data generating process
- This will allow us to say something about the patterns in the data generation
- We make inference about the model parameters
- "All models are wrong, but some are useful"

Basic Model: $y = \mu + \varepsilon_i \quad i = 1, \ldots, n$

$E[\varepsilon_i] = 0 \qquad \text{Var}[\varepsilon_i] = \sigma^2$

$\text{Cov}[\varepsilon_i, \varepsilon_j] = 0 \quad i \neq j$

SLR: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (instead of μ in BM)

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

1/15

ALSM 1

1

1.5 PARAMETER ESTIMATION

- We will use the collected data $(x_1, y_1) \ldots (x_n, y_n)$ to estimate β_0, β_1
- The objective is to find the "best" straight line through the data

The Method of Least Squares

- Suppose that $\hat{\beta}_0, \hat{\beta}_1$ are the estimates of β_0, β_1 (In general $\hat{\beta}_0 \neq \beta_0$ $\hat{\beta}_1 \neq \beta_1$)
- When $x = x_i$, the model predicts $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ Prediction error is $y_i - \hat{y}_i$
- Want to minimize this error for all point $i = 1, \ldots, n$

- The Least Squares (LS) criterion is to choose $\hat{\beta}_0, \hat{\beta}_1$ to minimize:

$$SS[\text{Error}] = SSE = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = Q(\hat{\beta}_0, \hat{\beta}_1)$$

- In the optimization literature, this is called an objective function

- SSE known as:
  - Sum of Squared errors
  - Residual Sum of Squares
  - Error Sum of Squares
  - Sum of Squared errors about the regression line

- $SSE = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ SSE a function of $\hat{\beta}_0, \hat{\beta}_1$
- Minimize SSE WRT: $\hat{\beta}_0, \hat{\beta}_1$

$$\frac{d(SSE)}{d(\hat{\beta}_0)} = \frac{d}{d(\hat{\beta}_0)} \left[(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + \ldots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2 \right]$$

$$= -2(y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1) + \ldots + -2(y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)$$

$$= -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \quad (1)$$

$$\frac{d(SSE)}{d(\hat{\beta}_1)} = \sum \frac{d}{d(\hat{\beta}_1)} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$= \sum (-2x_i)(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \quad (2)$$

- If we set ① and ② $=0$, we get normal equations

① $\Sigma y_i - n\hat{\beta}_0 - \hat{\beta}_1 \Sigma x_i = 0$

② $\Sigma x_i y_i - \hat{\beta}_0 \Sigma x_i - \hat{\beta}_1 \Sigma x_i^2 = 0$

① $\hat{\beta}_0 = \frac{\Sigma y_i}{n} - \hat{\beta}_1 \frac{\Sigma x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$

Using ① in ② $\Sigma x_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = 0$

$\Sigma x_i (y_i - \bar{y}) - \hat{\beta}_1 \Sigma x_i (x_i - \bar{x}) = 0$

$\hat{\beta}_1 = \frac{\Sigma (x_i)(y_i - \bar{y})}{\Sigma (x_i)(x_i - \bar{x})} = \frac{S_{xy}}{S_{xx}}$

$\Sigma x_i (x_i - \bar{x})$

$\Sigma (x_i)(x_i) - \Sigma (x_i)(\bar{x})$

$= \Sigma (x_i)^2 - \bar{x}^2$

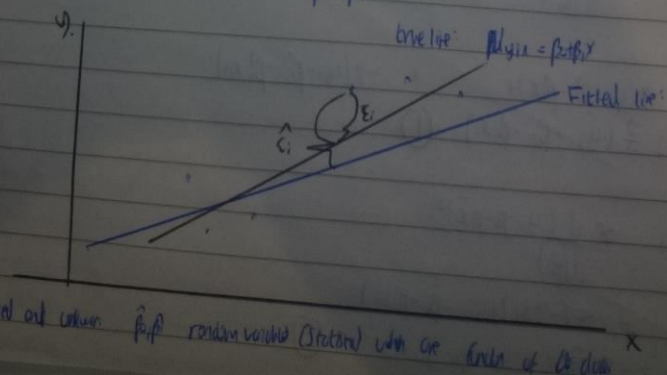
$\Sigma (x_i - \bar{x}) = \Sigma x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$

and similarly: $\Sigma x_i (y_i - \bar{y}) = \Sigma (y_i)(x_i - \bar{x}) = \Sigma (x_i - \bar{x})(y_i - \bar{y})$

The least squares estimators are:

$\hat{\beta}_1 = S_{xy}/S_{xx}$ $\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Fitted regression line is: $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$



β_0, β_1 fixed but unknown; $\hat{\beta}_0, \hat{\beta}_1$ random variable (statistic) whose one value of the data

05/10/15 ALSM1

Last Week:

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, \ldots, N$

(y_i: dependent variable; β_0, β_1: parameters; x_i: independent variable; ε_i: error)

$E[\varepsilon_i] = 0 \qquad Var[\varepsilon_i] = \sigma^2$

$Cov[\varepsilon_i, \varepsilon_j] = 0 \quad \text{for } i \neq j$

Want to predict mean value of y at any given x. $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Example

Basic Model: $y_i = \beta_0 + \varepsilon_i \quad (\beta_1 = 0)$

What is the LS estimator of β_0?

$\Sigma \varepsilon_i^2 = \Sigma (y_i - \hat{y}_i)^2$

$\Sigma (y_i - \beta_0)^2 \qquad dQ/d\beta_0 = -2 \Sigma (y_i - \hat{\beta}_0) = 0$

$\Sigma y_i - n\hat{\beta}_0 = 0$

$\hat{\beta}_0 = \bar{y}$

Example:

The corrected sum of ~~Squares~~ cross product and Squares can be computed in an efficient manner. E.g. $S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y})$ [corrected → mean has been [illegible]]

$S_{xy} = \Sigma (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})$

$= \Sigma x_i y_i - \Sigma x_i \bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$

$= \Sigma x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$

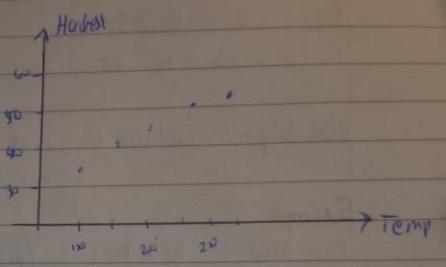
$= \Sigma x_i y_i - n\bar{x}\bar{y}$

$= \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}$

Similarly $S_{xx} = \Sigma (x_i - \bar{x})(x_i - \bar{x}) = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$

Example:
A Smartphone has an outer shield protective case which is made of hybrid materials. The hardness of the material depends on temperature which materials are merged. Durability is indicated by hardness. The materials were merged at 5 different temperatures & measured hardness.

Hardness	Temperature
36	150
41	175
43	200
50	225
51	250



The Scatterplot suggest a SLR model
Fit the model y = hardness x = temperature

$n=5$ $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ $S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

$\sum x_i = 1000$ $= 45775 - \frac{(1000)(221)}{5} = 975$

$\sum y_i = 221$ $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 205000 - \frac{1000^2}{5} = 6250$

$\sum x_i^2 = 205000$

$\sum y_i^2 = 9941$ $\hat{\beta}_1 = \frac{975}{6250} = 0.156$

$\sum x_i y_i = 45775$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $= \frac{\sum y_i}{n} - \hat{\beta}_1 \frac{\sum x_i}{n} = \frac{221}{5} - 0.156\left(\frac{1000}{5}\right) = 13.0$

Fitted line $= \hat{y} = 13 + 0.156x$

- 0.156 is the estimated increase in mean hardness for every extra degree celcius
- 13 is the estimated mean hardness at temperature zero
- In this case, intercept does not have meaningful interpretation
- However, by having the intercept in the model, we are allowing the line more flexibility in it's fit of the data

10/15 ALSM1

Try to find the SSE for $\hat{\beta}_0, \hat{\beta}_1$ $SSE = \sum(y_i - \hat{y}_i)^2 = \sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$= \sum(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i)^2$

$= \sum((y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x}))^2$

$= \sum(y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum(x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum(x_i - \bar{x})^2$

$= S_{yy} - 2\hat{\beta}_1 S_{xy} + \hat{\beta}_1^2 S_{xx}$

$= S_{yy} - \frac{2S_{xy}}{S_{xx}} S_{xy} + \frac{S_{xy}^2}{S_{xx}^2} S_{xx}$

$= S_{yy} - S_{xy}^2/S_{xx}$

For our example: $S_{yy} = 9927 - \frac{221^2}{5} = 153.8$

SSE = 6.7 which is the sum of squared residuals about the fitted line

$SSE = S_{yy} - S_{xy}^2/S_{xx} \Rightarrow S_{yy} = SSE + S_{xy}^2/S_{xx}$

variability in the data — variability accounted for by fitting the line

05/10/15

ALSM 1

$\Sigma(y_i-\bar{y})^2 = \Sigma(y_i-\hat{y}_i)^2 + \frac{S_{xy}^2}{S_{xx}}$

SSE for base model SSE for SLR

1.7 ANALYSIS OF VARIANCE (ANOVA) IN THE DEPENDENT VARIABLE

The variation in y can be quantified by either

$\Sigma(y_i-0)^2 = \Sigma y_i^2$ → Total Uncorrected Sum of Squares SS (uncorrected)

or more usually

$\Sigma(y_i-\bar{y})^2 = S_{yy}^2(n-1)$ → Total corrected Sum of Squares

Now $\Sigma(y_i-\bar{y})^2 = \Sigma(y_i^2-2y_i\bar{y}+\bar{y}^2)$

$= \Sigma y_i^2 - 2n\bar{y}^2 + n\bar{y}^2$

$= \Sigma y_i^2 - n\bar{y}^2$

= SS (uncorrected) − correction

Total corrected SS = Total Uncorrected SS − Correction

Consider writing SS(uncorrected) in terms of fitted value and residual $(y_i = \hat{y}_i + \hat{\varepsilon}_i)$

SS (uncorrected) $= \Sigma y_i^2 = \Sigma(\hat{y}_i+\hat{\varepsilon}_i)^2$

$= \Sigma(\hat{y}_i^2 + 2\hat{y}_i\hat{\varepsilon}_i + \hat{\varepsilon}_i^2)$

$= \Sigma\hat{y}_i^2 + \Sigma\hat{\varepsilon}_i^2 + 2\Sigma\hat{y}_i\hat{\varepsilon}_i$

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

$= \Sigma\hat{y}_i(y_i-\beta_0-\beta_1x_i)$

$= \Sigma\hat{y}_i(y_i-\bar{y}+\beta_1\bar{x}-\beta_1x_i)$

$= \Sigma\hat{y}_i[(y_i-\bar{y})+\beta_1(\bar{x}-x_i)]$

$= [\beta_0+\beta_1x_i][(y_i-\bar{y})+\beta_1(\bar{x}-x_i)]$

$= \Sigma[\bar{y}+\beta_1(x_i-\bar{x})][(y_i-\bar{y})+\beta_1(\bar{x}-x_i)]$

$= \Sigma\bar{y}(y_i-\bar{y}) + \Sigma\bar{y}\beta_1(\bar{x}-x_i) + \beta_1\Sigma(x_i-\bar{x})(y_i-\bar{y}) + \beta_1^2\Sigma(x_i-\bar{x})(\bar{x}-x_i)$

$= \bar{y}\Sigma(y_i-\bar{y}) - \bar{y}\beta_1\Sigma(x_i-\bar{x}) + \beta_1\Sigma(x_i-\bar{x})(y_i-\bar{y}) - \beta_1^2\Sigma(x_i-\bar{x})^2$

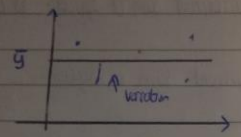
$= \bar{y}(0) - \bar{y}\beta_1(0) + \beta_1S_{xy} - \beta_1^2S_{xx}$

$= \frac{S_{xy}}{S_{xx}} S_{xy} - \frac{S_{xy}^2}{S_{xx}^2} S_{xx} = \frac{S_{xy}^2}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}} = 0$

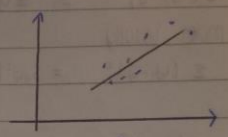
$SS(\text{Uncorrected}) = \Sigma \hat{y}_i^2 + \Sigma \hat{\varepsilon}_i^2 + 2\Sigma \hat{y}_i \hat{\varepsilon}_i$

$= \Sigma \hat{y}_i^2 + \Sigma \hat{\varepsilon}_i^2$

$= SS(\text{Model}) + SSE$



SSE BASIC



SSE SLR

$SS(\text{Model}) = \Sigma \hat{y}_i^2 = \Sigma [\hat{y}_i - \bar{y} + \bar{y}]^2$

$= \Sigma [(\hat{y}_i - \bar{y})^2 + 2\bar{y}(\hat{y}_i - \bar{y}) + \bar{y}^2]$

$= \Sigma (\hat{y}_i - \bar{y})^2 + 2\bar{y}\Sigma(\hat{y}_i - \bar{y}) + n\bar{y}^2$

Aside: $\bar{y}\Sigma(\hat{y}_i - \bar{y})$

$= \bar{y}\Sigma \hat{y}_i - \bar{y}\Sigma \bar{y}$

$= \bar{y}\Sigma \hat{y}_i - n\bar{y}^2$

$= \Sigma y_i^2 / n - \frac{n(\Sigma y_i)^2}{n^2} = 0$

Aside: $\bar{y}\Sigma(\hat{y}_i - \bar{y}) = \bar{y}\Sigma(\beta_0 + \beta_1 x_i - \bar{y})$

$= \bar{y}\Sigma(\bar{y} - \beta_1 \bar{x} + \beta_1 x_i - \bar{y})$

$= \bar{y}\Sigma(x_i - \bar{x})\beta_1$

$= \beta_1 \bar{y}\Sigma(x_i - \bar{x})$

$\downarrow 0 = 0$

$SS(\text{Model}) = \Sigma(\hat{y}_i - \bar{y})^2 + n\bar{y}^2$

$SS(\text{Regression}) + \text{correction}$

The regression sum of squares represents the reduction in variation (around the predicted mean) by adding the $\beta_1 x_i$ term to a model containing β_0

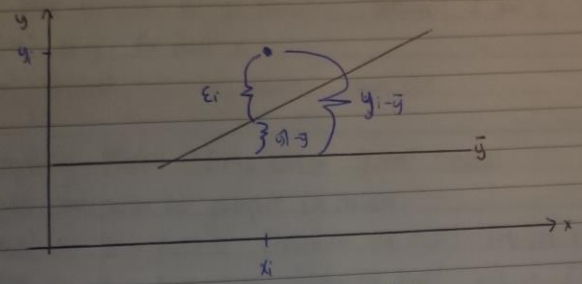
05/10/15 ALSM

$SS(\text{Uncorrected}) = SS(\text{Reg}) + n\bar{y}^2 + SSE$

If we subtract the correction from each side:

$SS(\text{Uncorrected}) - n\bar{y}^2 = SS(\text{Reg}) + SSE$

$SS(\text{Corrected}) = SS(\text{Reg}) + SSE$



$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y})^2 + \sum (\hat{y}_i - \bar{y})^2$

07/05/15

ALM 1

$$\underset{SS(Corr)}{\sum(y_i-\bar{y})^2} = \underset{SS(Reg)}{\sum(\hat{y}_i-\bar{y})^2} + \underset{SSE}{\sum(y_i-\hat{y}_i)^2}$$

(model; error in general)

- If there is little or no relationship between X and y. In the case $\hat{y}_i$ will be "close" to $\bar{y}$ for each x_i
- The SS(Reg) will be smaller than SSE

Comparing Models - Extra Sum of Squares Methods

Suppose we wish to compare the models:

Model 1: $y_i = \beta_0 + \varepsilon_i$ (horizontal line model) $i=1,\dots,n$ (reduced model)

Model 2: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (Sloping line Model) $i=1,\dots,n$ (full model)

Model 1 is a special case of Model 2 ($\beta_1=0$) and can be viewed as having less structure as it only describes deviations about a mean β_0

Model 1: $\hat{\beta}_0 = \bar{y}$ $\quad SSE = \sum(y_i-\bar{y})^2$

Model 2: $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ $\quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ $\quad SSE = \sum(y_i-\hat{y}_i)^2$ where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Note: β_0 is different in Model 1 and Model 2

Model 2 is less restricted than Model 1 $SSE_2 \leq SSE_1$. Always can get closer to the data with a complex model, and $SSE_1 - SSE_2$ is a measure of how much better model 2 is than model 1 → better in the sense of accounting for the variability in the data.

$$SSE_1 - SSE_2 = \sum(y_i-\bar{y})^2 - \sum(y_i-\hat{y}_i)^2$$
$$= SS(\text{Corrected}) - SSE_2$$
$$= SS(\text{Regression}) \Rightarrow \text{Extra Sum of Squares}$$

- Hence SS(Reg) indicates the importance of the $\beta_1 x_1$ term in the model
- we can use the "R" notation, $SS(Reg) = R(\beta_1|\beta_0)$ adds β_1, already have β_0

Now a Second Comparison:

Model 0: $y_i = \varepsilon_i \quad i=1,\dots,N$ (Reduced model)

Model 1: $y_i = \beta_0 + \varepsilon_i \quad i=1,\dots,N$ (Full Model)

Model 0: $SSE_0 = \Sigma (y_i - 0)^2 = \Sigma (y_i)^2$

Model 1: $SSE_1 = \Sigma (y_i - \bar{y})^2$

Thus $SSE_1 \leq SSE_0$ and $SSE_0 - SSE_1$ is a measure of how much better model 1 is

$SSE_0 - SSE_1$ = SS(Uncorrected) − SS(corrected) $= n\bar{y}^2$ (the correction)

Hence, $n\bar{y}^2$ represents the importance of β_0 in the model

$R(\beta_0) = n\bar{y}^2$ (the reduction)

Finally, let's compare:

Model 0: $y_i = \varepsilon_i \quad i=1,\dots,n$

Model 2: $y_i = \beta_0 + \beta_1 x_i \quad i=1,\dots,n$

Where Model 0 is a special case of Model 2 ($\beta_0 = 0, \beta_1 = 0$)

Model 0: $SSE_0 = \Sigma y_i^2$

Model 2: $SSE_2 = \Sigma (y_i - \hat{y}_i)^2$

We have $SSE_2 \leq SSE_0$ and $SSE_0 - SSE_2$ = SS(uncorrected) − SSE_2

= SS(Reg) + $n\bar{y}^2$

= SS(Model) − difference from adding β_0, β_1 to model with no structure

Here SS(Model) represents the importance of β_0 and $\beta_1 x_i$ in the model and

$R(\beta_1, \beta_1) = SS(Model) = R(\beta_0) + R(\beta_1 | \beta_0)$

SS(uncorrected)	$= \Sigma y_i^2$	n
SS(corrected)	$= \Sigma (y_i - \bar{y})^2$	$n-1$
Correction	$= n\bar{y}^2$	1
SS(Model)	$= R(\beta_0, \beta_1)$	2
SS(Regression)		1
SSE	$= \Sigma (y_i - \hat{y})^2$	$n-2$

12/10/15

ALSM I

Deviation between predicted mean from regression line and sample mean

$SS(Reg) = \Sigma(\hat{y}_i - \bar{y})^2$ (1)

$SSE \rightarrow$ residuals (not explained by model) $= \Sigma(y_i - \hat{y}_i)^2$ (n-2)

$SS(Corrected) = \Sigma(y_i - \bar{y})^2$ (n-1)

$SS(Reg) = \Sigma(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1 S_{xy}$

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

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

$\hat{\beta}_1 = S_{xy}/S_{xx}$

$= \Sigma(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2$

$= \Sigma(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})^2$

$= \Sigma(\hat{\beta}_1 (x_i - \bar{x}))^2$

$= \hat{\beta}_1^2 \Sigma(x_i - \bar{x})^2$

$= \hat{\beta}_1^2 S_{xx}$

$= \frac{S_{xy}^2}{S_{xx}^2} S_{xx} = \hat{\beta}_1 S_{xy}$

Anova Table

An analysis of variance (ANOVA) table gives a summary of the sources and contributions to SS(Corrected) or SS(Uncorrected) by various parts of the model

Source	DF	SS	Mean Square = SS/DF
Regression on X	1	$\Sigma(\hat{y}_i - \bar{y})^2$	MS(Reg)
Residual Error	n-2	$\Sigma(y_i - \hat{y}_i)^2$	MSE
Total Corrected	n-1	$\Sigma(y_i - \bar{y})^2$	

Source	DF	SS	MS = SS/DF
Correction R(β_0)	1	$n\bar{y}^2$	
Regression on X R($\beta_1|\beta_0$)	1	$\Sigma(\hat{y}_i - \bar{y})^2$	MS(Reg)
Residual Error	n-2	$\Sigma(y_i - \hat{y}_i)^2$	MSE
Total Uncorrected	n	Σy_i^2	

- The ANOVA Table gives $R(\beta_1|\beta_0)$ and $R(\beta_0)$ without having to fit extra models

Coefficient of Determination

- The contribution of the $\beta_1 x_i$ term in the model, measured by SS(Reg) and this has an upper limit of SS(Corrected)
- Hence we can quantify the contribution of x by the means of the coefficient of determination R^2

Where $R^2 = \frac{SS(Reg)}{SS(Corr)}$ $\quad SS(Corr) = SS(Reg) + SSE$

$\rightarrow SS(Reg) = SS(Corrected) - SSE$

$R^2 = \frac{SS(Corr) - SSE}{SS(Corr)} = 1 - \frac{SSE}{SS(Corr)}$ $\quad 0 \leq R^2 \leq 1$

- If $R^2 = 0.76$, then 76% of the variation in y is "explained" by its linear relationship with x.

NOTE: We can get large R^2 for poor model

Example

$\Sigma x_i = 1000$ $\quad \Sigma y_i = 221$ $\quad \Sigma x_i^2 = 206250$ $\quad \Sigma y_i^2 = 9927$ $\quad \Sigma x_i y_i = 45775$

$SS(Corr) = SS(Corrected) - n\bar{y}^2$

$= 9972 - 221^2/5 = 158.8$

$\beta_1 = .156$ $\quad S_{xy} = 975$

$SS(Reg) = \beta_1 S_{xy} = 0.156(975) = 152.1$

$SSE = SS(Corr) - SS(Reg) = 158.8 - 152.1 = 6.7$

ANOVA Table

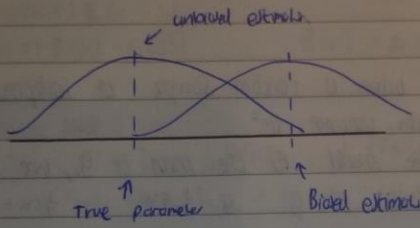
Source	DF	SS	MS
Regression	1	152.1	152.1
Residual	3	6.7	2.23
Corrected	4	158.8	

$R^2 = \frac{SS(Reg)}{SS(Corr)} = \frac{152.1}{158.8} = 0.957$

115 12/10/15 ALSM 1

1.8 PROPERTIES OF ESTIMATORS

- An estimator is a sample statistic $[\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \hat{\beta}_1 = S_{xy}/S_{xx}]$ and as a result has a sampling distribution
- If the mean of the sampling distribution is NOT equal to the true parameter value, being estimated, then the estimator is biased.
- Otherwise unbiased



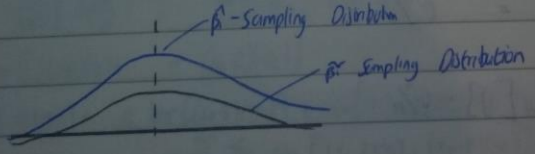
- Apart from being unbiased, it is also desirable for an estimator to have the smallest possible variance

Could do: $\hat{\beta}_1 = \frac{y_n - y_1}{x_n - x_1}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ (suggested estimators)

$E[\hat{\beta}_1] = E\left[\frac{y_n - y_1}{x_n - x_1}\right] = \frac{1}{x_n - x_1} E[y_n - y_1]$

$= \frac{1}{x_n - x_1} [\beta_0 + \beta_1 x_n - \beta_0 - \beta_1 x_1]$

$= \beta_1 \Rightarrow$ unbiased estimator



Revision of Random Variable Properties

- Let μ and v be 2 r.v.

$E[\mu]$: mean of μ

$\text{Var}[\mu] = E[\mu^2] - [E[\mu]]^2$

$\text{Corr}[\mu, v] = \frac{\text{Cov}[\mu, v]}{\sqrt{\text{Var}[\mu]\,\text{Var}[v]}}$

- If μ and v are uncorrelated or independent, the $Cov[\mu, v] = 0$
- $Cov[\mu, \mu] = Var[\mu]$
- $E[a\mu + bv] = a\,E[\mu] + b\,E[v]$
  $Var[a\mu + bv] = a^2 Var[\mu] + b^2 Var[v] + 2ab\,Cov[\mu, v]$

- Consider another random variable z and constant c
- $Cov[a\mu + bv, cz] = ac\,Cov[\mu, z] + bc\,Cov[v, z]$

For example:
- Consider a situation where a random sample of independent observations $y_1, \dots, y_n$ each with variance σ^2
- Then, irrespective of the model for the mean of y_i, we can consider the variance of the sampling distribution of $\bar{y}$ $\quad \bar{y} = \frac{1}{n}\sum y_i = \frac{1}{n}(y_1 + \dots + y_n)$
  $= \frac{1}{n}y_1 + \dots + \frac{1}{n}y_n$

$E[\bar{y}] = \frac{1}{n}E[y_1] + \dots + \frac{1}{n}E[y_n]$
$= \frac{1}{n}\mu + \dots + \frac{1}{n}\mu$
$= \mu$

$Var[\bar{y}] = Var\left[\frac{1}{n}y_1 + \dots + \frac{1}{n}y_n\right]$
$= Var\left[\frac{1}{n}y_1\right] + \dots + Var\left[\frac{1}{n}y_n\right] + Cov\left[\frac{1}{n}y_1, \frac{1}{n}y_2\right] + \dots + Cov\left[\frac{1}{n}y_{n-1}, \frac{1}{n}y_n\right]$
$= \frac{1}{n^2}Var[y_1] + \dots + \frac{1}{n^2}Var[y_n] + 2\sum_{i=1}^{n-1}\sum_{j>i}\frac{1}{n^2}Cov[y_i, y_j]$
$= \frac{1}{n^2}\sigma^2 + \dots + \frac{1}{n^2}\sigma^2 + 0 \quad$ (Cov=0 by way of independance)
$= n\sigma^2/n^2 = \sigma^2/n$

$E[\bar{y}] = \mu,\ Var[\bar{y}] = \sigma^2/n$

$y_i \sim N(\mu, \sigma^2) \Rightarrow \bar{y} \sim N(\mu, \sigma^2/n)$

σ^2 unknown: $\frac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

115 12/10/15 ALSM1

1.9 PROPERTIES OF THE LEAST SQUARES ESTIMATOR

Consider the sampling distribution of $\hat{\beta}_1$.

$$E[\hat{\beta}_1] = E\left[\frac{S_{xy}}{S_{xx}}\right] = E\left[\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{S_{xx}}\right]$$

$$= \frac{1}{S_{xx}} E\left[\Sigma(x_i-\bar{x})(y_i-\bar{y})\right] \quad x_i\text{'s known} \Rightarrow \text{assumed to be fixed}$$

$$= \frac{1}{S_{xx}} E\left[(x_1-\bar{x})(y_1-\bar{y}) + \dots + (x_n-\bar{x})(y_n-\bar{y})\right]$$

$$= \frac{1}{S_{xx}} \left[(x_1-\bar{x})E(y_1-\bar{y}) + \dots + (x_n-\bar{x})E[y_n-\bar{y}]\right] \quad X\text{'s assumed non-random}$$

$$= \frac{1}{S_{xx}} \left[\Sigma(x_i-\bar{x})\, E[y_i-\bar{y}]\right]$$

$$= \frac{1}{S_{xx}} \left[\Sigma(x_i-\bar{x})\, E[y_i] - E[\bar{y}]\right] \quad E[y_i] = \beta_0+\beta_1 x_i \text{ (From Model)}$$

Assuming model is correct:

$$E[\bar{y}] = E\left[\frac{\Sigma y_i}{n}\right] = \frac{1}{n} E[y_i] = \frac{1}{n}\Sigma[\beta_0+\beta_1 x_i]$$

$$= \frac{1}{n}\left[n\beta_0 + \beta_1 \Sigma x_i\right] = \beta_0 + \beta_1 \bar{x}$$

$$= \frac{1}{S_{xx}} \left[\Sigma(x_i-\bar{x}) \left[\underbrace{\beta_0+\beta_1 x_i}_{E[y_i]} - \underbrace{\beta_0-\beta_1\bar{x}}_{E[\bar{y}]}\right]\right]$$

$$= \frac{1}{S_{xx}} \Sigma(x_i-\bar{x})\beta_1(x_i-\bar{x})$$

$$= \beta_1 \frac{1}{S_{xx}} \Sigma(x_i-\bar{x})(x_i-\bar{x})$$

$$= \beta_1 \frac{S_{xx}}{S_{xx}}$$

$$= \beta_1$$

$\hat{\beta}_1$ is an unbiased estimator of β_1.

Variance

$$Var[\hat{\beta}_1] = Var\left[\frac{S_{xy}}{S_{xx}}\right] = \frac{1}{S_{xx}^2} Var\left[\Sigma(x_i-\bar{x})y_i\right]$$

$$= \frac{1}{S_{xx}^2} Var\left[(x_1-\bar{x})(y_1) + \dots + (x_n-\bar{x})(y_n)\right]$$

$$= \frac{1}{S_{xx}^2}\left[Var[(x_1-\bar{x})(y_1)] + \dots + Var[(x_n-\bar{x})(y_n)] + Cov\left[(x_1-\bar{x})(y_1), (x_2-\bar{x})(x_2)\right] + \dots + Cov\left[(x_{n-1}-\bar{x})(y_{n-1}), (x_n-\bar{x})(y_n)\right]\right]$$

$$= \frac{1}{S_{xx}^2}\left[\Sigma Var[(x_i-\bar{x})(y_i)] + 2\sum_{i=1}^{n-1}\sum_{j>i} Cov\left[(x_i-\bar{x})(y_i), (x_j-\bar{x})(y_j)\right]\right]$$

$$= \frac{1}{S_{xx}^2}\left[\Sigma(x_i-\bar{x})^2 Var[y_i] + 2\Sigma\Sigma(x_i-\bar{x})(x_j-\bar{x}) Cov[y_i, y_j]\right]$$

$$= \frac{1}{S_{xx}^2}\left[\Sigma(x_i-\bar{x})\sigma^2 + 0\right] \quad \text{(by independence)}$$

$$= \sigma^2 S_{xx}/S_{xx}^2 = \sigma^2/S_{xx}$$

14/10/15 ALSM 1

$E[\hat{\beta}_0] = \beta_0$ $\quad E[\beta_1] = \beta_1$

$Var[\hat{\beta}_0] = ?$ $\quad Var[\beta_1] = \frac{\sigma^2}{S_{xx}}$

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

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

$= Var[\bar{y}] + Var[\hat{\beta}_1 \bar{x}] - 2Cov[\bar{y}, \hat{\beta}_1 \bar{x}]$

$= \underline{Var[\bar{y}]} + \underline{Var[\hat{\beta}_1 \bar{x}]} - 2\bar{x}^2 Cov[\bar{y}, \hat{\beta}_1]$ Assume $\bar{x}$ known $\Rightarrow$ not random

$Cov[\bar{y}, \hat{\beta}_1] = Cov\left[\frac{1}{n}\Sigma y_i, \frac{S_{xy}}{S_{xx}}\right]$ (x's assumed not random)

$= \frac{1}{n S_{xx}} Cov[\Sigma y_i, S_{xy}]$

$= \frac{1}{n S_{xx}} Cov[y_1 + \dots + y_n, S_{xy}]$

$= \frac{1}{n S_{xx}} [Cov[y_1, S_{xy}] + \dots + Cov[y_n, S_{xy}]]$

$= \frac{1}{n S_{xx}} [\Sigma Cov[y_i, S_{xy}]]$

$= \frac{1}{n S_{xx}} \Sigma Cov\left[y_i, \Sigma_j (x_j - \bar{x}) y_j\right]$

$= \frac{1}{n S_{xx}} \Sigma Cov[y_i, (x_1 - \bar{x}) y_1] + \dots + Cov[y_i, (x_i - \bar{x}) y_i] + \dots + Cov[y_i, (x_n - \bar{x}) y_n]$

$= \frac{1}{n S_{xx}} \Sigma Cov[y_i, (x_i - \bar{x}) y_i]$

$= \frac{1}{n S_{xx}} \Sigma (x_i - \bar{x}) Cov[y_i, y_i]$

$= \frac{\sigma^2}{n} \frac{\Sigma (x_i - \bar{x})}{S_{xx}} = 0 \quad \Rightarrow (\Sigma (x_i - \bar{x}) = 0)$

$Var[\hat{\beta}_0] = Var[\bar{y}] + Var[\hat{\beta}_1 \bar{x}]$

$= \sigma^2/n + \bar{x}^2 \sigma^2 / S_{xx} + 0$

$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$

Note: $Cov[\hat{\beta}_0, \hat{\beta}_1] = Cov[\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1]$

$= Cov[\bar{y}, \hat{\beta}_1] - \bar{x} Cov[\hat{\beta}_1, \hat{\beta}_1]$

$= 0 - \bar{x} Var[\hat{\beta}_1]$

$= \frac{-\bar{x}\sigma^2}{S_{xx}}$

1.10 GAUSS-MARKOV THEORM

Assuming $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, \dots, n$

$E[\varepsilon_i] = 0 \quad Var[\varepsilon_i] = \sigma^2 \quad Cov[\varepsilon_i, \varepsilon_j] = 0 \quad i \neq j$

19/10/

- Then the LS estimators have minimum variance amongst all unbiased linear estimates i.e. LS estimates
- BLUE : Best Linear Unbiased Estimator

III ESTIMATION OF σ^2

When the model is $y_i = \mu + \varepsilon_i$ we use $S^2 = \frac{1}{n-1}\Sigma(y_i - \bar{y})^2$ as an estimator of σ^2

When the model is $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ we use $\frac{1}{n-2}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

MSE = SSE/n-2

Is this estimator unbiased?

$E[MSE] = \frac{1}{n-2}\Sigma E[(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2]$

Remember $E[y^2] = Var[y] + [E[y]]^2$

$E[MSE] = \frac{1}{n-2}\Sigma\left[Var[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i] + (E[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i])^2\right]$

$= \frac{1}{n-2}\Sigma\left[Var[y_i] + Var[\hat{\beta}_0] + x_i^2 Var[\hat{\beta}_1] - 2\,Cov[y_i, \hat{\beta}_0] - 2x_i\,Cov[y_i, \hat{\beta}_1] + 2x_i\,Cov[\hat{\beta}_0, \hat{\beta}_1] + [[\hat{\beta}_0 + \hat{\beta}_1 x_i - \beta_0 - \beta_1 x_i]^2\right]$

$= \frac{1}{n-2}\Sigma\left[\sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) + \frac{x_i^2\sigma^2}{S_{xx}} - 2\,Cov[y_i, \hat{\beta}_0] - 2x_i\,Cov[y_i, \hat{\beta}_1] + 2x_i\,Cov[\hat{\beta}_0, \hat{\beta}_1]\right]$

① ② ③

① : $Cov[y_i, \hat{\beta}_1] = Cov[y_i, \frac{S_{xy}}{S_{xx}}]$

$= \frac{1}{S_{xx}} Cov[y_i, \Sigma(x_i - \bar{x})y_i]$ Will only pick up i^{th} term (independence)

$= \frac{1}{S_{xx}} Cov[y_i, (x_i - \bar{x})y_i]$

$= \frac{(x_i - \bar{x})}{S_{xx}} Cov[y_i, y_i] = \frac{\sigma^2(x_i - \bar{x})}{S_{xx}}$

② : $Cov[y_i, \hat{\beta}_0] = Cov[y_i, \bar{y} - \hat{\beta}_1\bar{x}]$

$= Cov[y_i, \frac{\Sigma y_i}{n} - \hat{\beta}_1\frac{\Sigma y_i}{n}]$

$= Cov[y_i, \bar{y}] - Cov[y_i, \hat{\beta}_1\bar{x}]$

$= \frac{1}{n} Cov[y_i, \Sigma y_i]$

$= \frac{1}{n} Cov[y_i, y_i] = \sigma^2/n$

$\frac{\sigma^2}{n} - \frac{\sigma^2(x_i - \bar{x})}{S_{xx}}$

1/10/15 ALSM 1

$$E[MSE] = \sum\left[\sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{\bar{X}}{S_{XX}}\right) + X_i^2 + \frac{X_i^2\sigma^2}{S_{XX}} - 2\left(\frac{\sigma^2}{n} - \frac{\bar{X}(X_i-\bar{X})\sigma^2}{S_{XX}}\right) - \frac{2X_i(X_i-\bar{X})\sigma^2}{S_{XX}} + 2X_i\left(\frac{-\bar{X}\sigma^2}{S_{XX}}\right)\right]$$

$$= \frac{1}{n-2}\sum\left[\sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} - \frac{X_i^2\sigma^2}{S_{XX}} - \frac{\bar{X}^2\sigma^2}{S_{XX}} - \frac{2X_i\bar{X}\sigma^2}{S_{XX}}\right]$$

$$= \sum\left[\sigma^2 + \sigma^2/n \; 2\sigma^2/n - \sigma^2/S_{XX}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right]$$

$$= \sum\left[\sigma^2 - \sigma^2/n - \sigma^2/S_{XX}(X_i-\bar{X})^2\right]$$

$$= \frac{1}{n-2}\left[n\sigma^2 - \sigma^2 - \sigma^2/S_{XX}\left(\sum(X_i-\bar{X})^2\right)\right]$$

$$= \frac{1}{n-2}\left[n\sigma^2 - \sigma^2 - \frac{\sigma^2}{S_{XX}}S_{XX}\right]$$

$$= \frac{1}{n-2}\left[n\sigma^2 - \sigma^2 - \sigma^2\right]$$

$$= \frac{n-2}{n-2}\sigma^2 = \sigma^2 \Rightarrow$$ unbiased estimator of σ^2 is σ^2 w/ MSE

Aside:

$Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad Y_1, \ldots, Y_n$

$Z_i = \frac{Y_i - \mu}{\sigma} \sim N(0,1)$ $\quad Z_i^2 \sim \chi_1^2$ chi-square, 1 d.f.

$Z_1^2 + \ldots + Z_n^2 \sim \chi_n^2$ $\quad n$ d.f.

$\bar{Y} \sim N(\mu, \sigma^2/n)$

$\mu = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}(\bar{Y} - \mu)/\sigma \sim N(0,1)$

$\mu^2 \sim \chi_1^2$

$R = \frac{1}{\sigma^2}\sum(Y_i - \mu)^2 = \sum\left(\frac{Y_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$ $\quad n$ d.f.

$$R = \frac{1}{\sigma^2}\sum(Y_i - \mu)^2 = \frac{1}{\sigma^2}\sum\left[(Y_i - \bar{Y}) + (\bar{Y} - \mu)\right]^2$$

$$= \frac{1}{\sigma^2}\left[\sum(Y_i - \bar{Y})^2 + 2(0) + n(\bar{Y} - \mu)\right]$$

$$= \frac{1}{\sigma^2}\sum(Y_i - \bar{Y})^2 + \frac{n}{\sigma^2}(\bar{Y} - \mu)$$

$$= \frac{1}{\sigma^2}\left[(n-1)S^2\right] + \frac{\sqrt{n}(\bar{Y} - \mu)^2}{\sigma^2}$$

$\uparrow N(0,1)$

$R = \frac{(n-1)S^2}{\sigma^2} + \mu^2$

$R \sim \chi^2_n \quad \mu^2 \sim \chi^2_1$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Show $\chi^2_n \leftrightarrow$ Sum of n Squared $N(0,1)$

$F_{\mu_1,\mu_2} = \frac{\chi^2_{\mu_1}/\mu_1}{\chi^2_{\mu_2}/\mu_2}$ F-test for comparison of variance

$t_d = \frac{N(0,1)}{\sqrt{\chi^2_d/d}}$ t distribution with d degrees of freedom

1/10/15 ALSM

From last week:

$y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$z_i = \frac{y_i - \mu}{\sigma} \sim N(0,1)$

$z_i^2 \sim \chi_1^2$

$\sum z_i^2 \sim \chi_n^2$

$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}$

1R STATISTICAL INFERENCE FROM THE LINEAR MODEL

- In order to make inferences we must make an assumption about the distribution of y_i or ε_i
- We will assume the distribution is normal

$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

$\varepsilon_i \sim N(0, \sigma^2)$

1 $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x}) y_i}{S_{xx}} = \sum c_i y_i$

$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / S_{xx})$

2 $\hat{\beta}_0 = \frac{1}{n}(\sum y_i - \hat{\beta}_1 \sum x_i) = \sum d_i y_i$

$\hat{\beta}_0 \sim N(\beta_0, \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}))$

3 It can be shown that $\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$ (two parameters $\hat{\beta}_0, \hat{\beta}_1$)

$\frac{(n-2)(\frac{SSE}{n-2})}{\sigma^2} = \frac{(n-2)MSE}{\sigma^2} \sim \chi^2_{n-2}$

4 It can be shown that $\hat{\beta}_1$ and SSE are independent and $\hat{\beta}_0$ and SSE are independent.

1.13 HYPOTHESIS TESTING

H_0: $\beta_1 =$ some value H_1: $\beta_1 \neq m$ or $\beta_1 > m$ $\beta_1 < m$

Usually we would like to test $m = 0$

$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{S_{xx}})$

Assume that H_0 is true i.e. $\beta_1 = m$

$\frac{\hat{\beta}_1 - m}{\sigma/\sqrt{S_{xx}}} \sim N(0,1)$ distribution (Standardising $\hat{\beta}_1$)

But we don't know what σ is

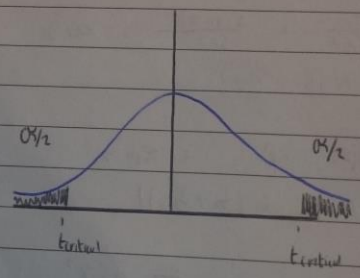
Student's t distribution is defined as follows:

$t_\nu = \frac{Z}{\sqrt{\frac{\chi^2_\nu}{\nu}}}$ $Z \sim N(0,1)$ χ^2_ν > independent

So $t_{n-2} = \frac{\left(\frac{\hat{\beta}_1 - m}{\sigma/\sqrt{S_{xx}}}\right)}{\sqrt{\left(\frac{(n-2)MSE}{\sigma^2}\Big/(n-2)\right)}}$

$t_{n-2} = \frac{\hat{\beta}_1 - m}{\sqrt{MSE/S_{xx}}}$

Look at distribution:



- If we said: "I'll reject H_0 if I was less than 5% likely to have seen this t-value if H_0 was true"

- If $|t_{n-2}| > t_{critical}$ then the probability of observing this data when the null hypothesis is true is less than (or =) α
- We call the region where we would reject, the rejection region
- We can find critical values using tables, statistical packages, compute p-value

Using a similar argument:

$H_0: \beta_1 = C$ $H_1: \beta_1 \neq C$ or $\begin{pmatrix} > \\ < \end{pmatrix} C$

Using the test statistic:

$$t = \frac{\hat{\beta}_1 - C}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}$$

NOTE: The estimated standard deviation of an estimator is called its standard error.

$$SE(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{xx}}}$$

$$SE(\hat{\beta}_0) = \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)}$$

28/10/15 ALSM 1

$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ $\quad \sigma^2 = Var[\varepsilon_i]$

$\frac{(n-2)MSE}{\sigma^2} \sim \chi^2_{n-2}$

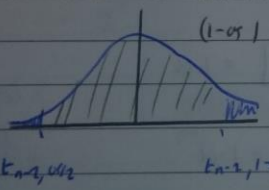
Construct a t-distribution from basis of normal distribution

$t = \frac{z}{\sqrt{\chi^2/d}}$ $\quad z \sim N(0,1)$

$= \frac{(\hat{\beta}_1 - \beta_1)\Big/\sqrt{\frac{\sigma^2}{S_{xx}}}}{\sqrt{\frac{(n-2)MSE}{(n-2)\sigma^2}}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}} \sim t_{n-2}$

If we know that $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE/S_{xx}}} \sim t_{n-2}$ can we it to construct a CI

1.14 CONFIDENCE INTERVALS



$P\left[-t_{n-2,\alpha/2} \le t_{n-2} \le t_{n-2,\alpha/2}\right] = 1-\alpha$

Replace $t \rightarrow P\left[t_{n-2,\alpha/2} \le \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}} \le t_{n-2,\alpha/2}\right]$

$-t_{n-2,\alpha/2}\sqrt{\frac{MSE}{S_{xx}}} \le \hat{\beta}_1 - \beta_1 \le t_{n-2,\alpha/2}\sqrt{\frac{MSE}{S_{xx}}}$

$\hat{\beta}_1 - t_{n-2,\alpha/2}\sqrt{\frac{MSE}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{n-2,\alpha/2}\sqrt{\frac{MSE}{S_{xx}}}$

$\rightarrow \hat{\beta}_1 \pm t_{n-2,\alpha/2}\sqrt{MSE/S_{xx}}$

$\hat{\beta}_1 \pm t_{n-2,\alpha/2}\sqrt{MSE/S_{xx}}$

Confidence Interval for β_0?

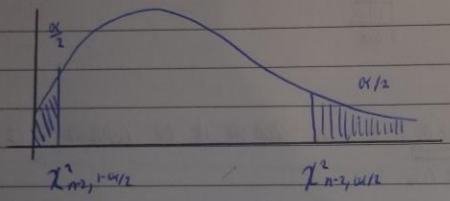
$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2}$

$P\left[-t_{n-2,\alpha/2} \leq \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \leq t_{n-2,\alpha/2}\right] = 1-\alpha$ rearr

$\hat{\beta}_0 \pm \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}\, t_{n-2,\alpha/2}$

Confidence Interval for σ^2?

$\frac{(n-2)MSE}{\sigma^2} \sim \chi^2_{n-2}$ use this fact



Not dealing with a symmetric distribution

$P\left[\chi^2_{n-2,1-\alpha/2} \leq \frac{(n-2)MSE}{\sigma^2} \leq \chi^2_{n-2,\alpha/2}\right] = 1-\alpha$

$P\left[\frac{1}{\chi^2_{n-2,1-\alpha/2}} \leq \frac{\sigma^2}{(n-2)MSE} \leq \frac{1}{\chi^2_{n-2,\alpha/2}}\right] = 1-\alpha$

$P\left[\frac{(n-2)MSE}{\chi^2_{n-2,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-2)MSE}{\chi^2_{n-2,\alpha/2}}\right] = 1-\alpha$

Which is a $100(1-\alpha)\%$ confidence interval for σ^2

1 IS AN ALTERNATIVE METHOD FOR TESTING THE HYPOTHESIS $H_0: \beta_1 = 0$ v $H_1: \beta_1 \neq 0$

- Have possibility of directional alternatives in t-test like $\beta_1 > 0$ or < 0 etc.
- Cannot do this in the F-test

- Recall the model comparison of Section 1.7 using the "extra sum of squares" method

28/10/15 ALSM 1

Model 1: $y_i = \beta_0 + \varepsilon_i$ Intercept only model

Model 2: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$R(\beta_1|\beta_0)$ = Reduction in error sum of squared by adding $\beta_1 x_i$ term to the model

= SS(Reg) "Extra Sum of Squares"

SS(Reg) and SSE are independent

When H_0 is true ($\beta_1 = 0$) then $SS(Reg)/\sigma^2 \sim \chi^2_1$

-But by definition the F distribution is the ratio of two independent χ^2 distributions divided by their d.f.

-An F-dist has 2 d.f.

$$F_{v_1, v_2} = \frac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2} \qquad \chi^2_{v_1}, \chi^2_{v_2} \text{ are independent}$$

So when H_0 is true $\frac{SS(Reg)}{\sigma^2} \sim \chi^2_1$

$MSE(n-2)/\sigma^2 \sim \chi^2_{n-2}$

$$F = \frac{SS(Reg)/\sigma^2 / 1}{MSE(n-2)/\sigma^2 / n-2} = \frac{SS(Reg)}{MSE} \sim F_{1, n-2}$$

-Using this F statistic, we can compare our observation with what would be true if H_0 was true

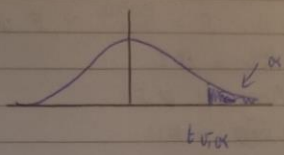
-look up critical F value

-IF $F > F_{1, n-2, \alpha}$ then the probability of this happening under H_0 would be α or less, so in that case we'd reject $H_0 = a$

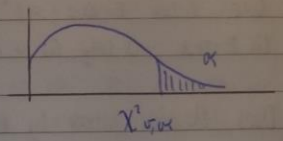
-Evidence for the alternative H_A.

02/11/15 ALSM 1

- t distribution with υ d.f.
- Work to get point such that area in the tail to the right is α
- Or as a percentage 100α % of the area
- This percentage point is denoted $t_{\upsilon;\alpha}$



- χ^2 with υ d.f.
- Area to right of $\chi^2_{\upsilon;\alpha}$ is α



Left point: $t_{n-2,1-\alpha/2} = -t_{n-2,\alpha/2}$ (symmetrical)

Right point: $t_{n-2,\alpha/2}$

Example - Material Durability

$n = 5$ $\hat{\beta}_1 = 0.156$ $\hat{\beta}_0 = 13$ $S_{xx} = 6200$

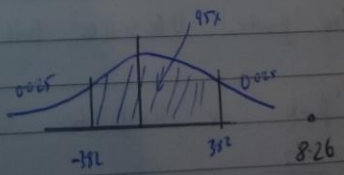
$\bar{x} = 200$ $SS(Reg) = 152.1$ $SSE = 6.7$ $MS(Reg) = 152.1$ $MSE = 2.23$

Non-Directional Hypothesis

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ Testing for relationship between 2 variables

$$t = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{MSE}{S_{xx}}}} = \frac{0.156}{\sqrt{\frac{2.23}{6200}}} = 8.26 \quad t_{calc}$$

- Go to t-distribution with 3 d.f.
- Significance level of 5%
- Critical value $\pm t_{3,0.025} = \pm 3.182$

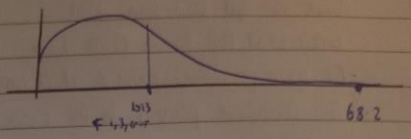


- Test Statistic (8.26) lands in the rejection region
- Evidence to support the alternative $H_1: \beta_1 \neq 0$
- Conclude: Durability of material depends on temperature

02/10/19

F-Statistic

- Under H_0: $F = \frac{MS(Reg)}{MSE} \sim F$ distribution with 1, 3 d.f
- $F = \frac{15.21}{2.27} = 68.2$
- Significance level of 5%
- $F_{1,3,0.05} = 10.13$



→ Reject H_0, evidence against $\beta_1 = 0$

NOTE: $t = 8.26$ $F = 68.2 \approx t^2$

The F-Test is only for a non directional hypothesis

Does Durability increase by 0.15 units for extra degree in temperature or is it increase less?

$H_0: \beta_1 = 0.15$ $\quad H_1: \beta_1 < 0.15$

$t_{calc} = \frac{\hat{\beta}_1 - 0.15}{\sqrt{MSE/S_{xx}}} = \frac{0.156 - 0.15}{\sqrt{2.27/6250}} = 0.32$

- T-table $t_{3,0.05}$ is -2.35
- Rejection region is any $t_{calc} < -2.35$
- Don't reject H_0: no evidence $\beta_1 < 0.15$

P-Value

- P-value: For a test Pr[observing test statistic as extreme as that observed if H_0 is true]
- Here p-value $= P_r[t_3 < 0.32] = 0.61$ (t-table)
- Using p-value, $0.61 > 0.05$; Fail to reject H_0

$H_0: \beta_1 = b$ v $H_1: \beta_1 \neq b$

$t = \frac{\hat{\beta}_1 - b}{\sqrt{MSE/S_{xx}}}$

Assume t is positive $\quad P[|t_{n-2}| > t] = P[t_{n-2} < t] + P[t_{n-2} > t]$

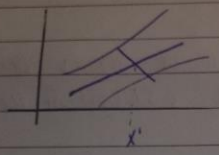
p-value $= 2P[t_{n-2} < -t]$

p-value $= 2P[t_{n-2} > t]$

100(1-α)% Confidence interval for mean of y at x'

$$P\left[-t_{n-2,\alpha/2} \le \frac{\hat{y}' - E[y|x']}{\sqrt{MSE\left(\frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}\right)}} \le t_{n-2,\alpha/2}\right] = 1-\alpha$$

$$\hat{y}' \pm t_{n-2,\alpha/2}\sqrt{MSE\left(\frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}\right)}$$



narrowest at $x' = \bar{x}$

17 PREDICTION

Suppose we wish to predict the value of a new observation when $X = x'$

Predicted value → not mean value $\quad y' = \beta_0 + \beta_1 x' + \varepsilon'$ more uncertainty

- Point prediction of y' would be $\hat{y}' = \hat{\beta}_0 + \hat{\beta}_1 x'$
- Consider the prediction error:
  $y' - \hat{y}'$ (Deviation between actual and predicted)
- $E[y' - \hat{y}'] = \beta_0 + \beta_1 x' - (\beta_0 + \beta_1 x')$
  $= 0$ expected value of 0

- $Var[y' - \hat{y}'] = Var[y'] + Var[\hat{y}'] + 0$ uncorrelated → independence
  $= Var[\varepsilon'] + Var[\hat{y}']$
  $= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}\right)$
  $= \sigma^2\left(1 + \frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}\right)$

Since y' and $\hat{y}'$ are independent normal, then $y' - \hat{y}'$ is normal

$$\Rightarrow \frac{(y'-\hat{y}') - 0}{\sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}\right)}}$$ follow a t-distribution $n-2$ d.f.

02/10/15 ALSM1

How would you do a test of $H_0: \beta_1 = b$ v H_A $\beta_1 < b$ w.r.t. $(\beta_1 > b)$ if you had the p-value above?

Example Continued

95% CI for β_1 = $\hat{\beta}_1 \pm t_{3,0.025}\sqrt{\frac{MSE}{S_{xx}}}$

$= 0.156 \pm 3.182\sqrt{233/6250}$

$0.96 \leq \beta_1 \leq 0.216$

Correspondence between CI and test: 0 isn't in the 95% CI then we would reject $H_0: \beta_1 = 0$, in favour of $H_A: \beta_1 \neq 0$ at 5% significance level

95% CI for β_0 $\hat{\beta}_0 \pm t_{3,0.025}\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$

$0.0793 \leq \beta_0 \leq 3.836$

95% CI for σ^2: $\frac{(n-2)MSE}{\chi^2_{3,0.025}} \leq \sigma^2 \leq \frac{(n-2)MSE}{\chi^2_{3,0.975}}$

$\chi^2_{3,0.025} = 9.348$ $\chi^2_{3,0.975} = 2.158$

$0.716 \leq \sigma^2 \leq 3.1$ wide interval but small n

1.6 INFERENCE CONCERNING $E[y|x']$

From SLR $E[y|x'] = \mu_{y|x'} = \beta_0 + \beta_1 x'$

Point estimate $\hat{y}' = \hat{\beta}_0 + \hat{\beta}_1 x'$

$E[\hat{y}'] = E[\hat{\beta}_0] + x' E[\hat{\beta}_1] = \beta_0 + \beta_1 x'$

$Var[\hat{y}'] = Var[\hat{\beta}_0] + x'^2 Var[\hat{\beta}_1] + 2x' Cov[\hat{\beta}_0, \hat{\beta}_1]$

$= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) + x'^2\left(\frac{\sigma^2}{S_{xx}}\right) - 2x'\frac{\bar{x}\sigma^2}{S_{xx}}$

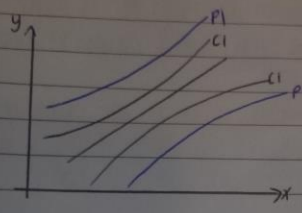
$= \frac{\sigma^2}{n} + \frac{\sigma^2(x'-\bar{x})^2}{S_{xx}}$ SE: $\sigma^2\left(\frac{1}{n}, \frac{(x-\bar{x})^2}{S_{xx}}\right)$

$\Rightarrow \frac{\hat{y}' - E[y|x']}{\sqrt{MSE\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)}}$ follows a t-distribution with n-2 df

02/10/15 ALSM 1

$P[\hat{y}' - t_{n-2,1-\alpha/2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{S_{xx}}\right)} \le y' \le \hat{y}' + t_{n-2,1-\alpha/2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{S_{xx}}\right)}] = 1-\alpha$

and this give a $100(1-\alpha)\%$ PREDICTION interval for y at x'



Since we are estimating the population mean of x' rather than a single value of y, the CI is always narrower than the prediction interval

Both are narrowest at $x' = \bar{x}$

Example

$x = x'$

95% CI for mean durability when temperature is 200°C

$\bar{x} = 200$ $\hat{y} \pm t_{3,0.025}\sqrt{MSE/n}$

42.07 to 46.32

95% PI: $\hat{y} \pm t_{3,0.025}\sqrt{MSE(1+\frac{1}{n})}$

38.995 to 49.405