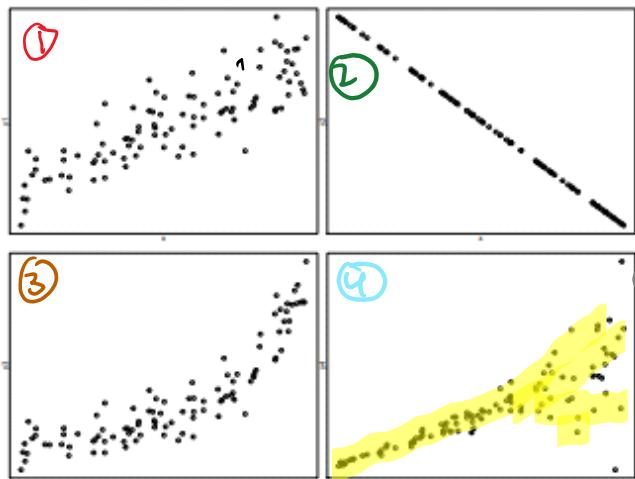


Preliminary activity II



note: friend in difference does not matter which x you pick has different noise, while others have the same noise.

① Description: - increasing/positive relationship
- noise
- linear

③ Description: - increasing/positive
- has noise
- curved

② Description: - decreasing/negative
- no noise

④ Description: - increasing/positive
- has noise

Heteroscedasticity: the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it.

Regression analysis

- statistical methodology that utilizes the relation between variables.
- Predicts a response variable (or outcome) from the relation between the response and other variables.
- Regression analysis is used in many disciplines such as:
 - Business:
 - i) Forecasting: predicting future demand for a product.
 - ii) Optimization: fine tune manufacturing and delivery processes

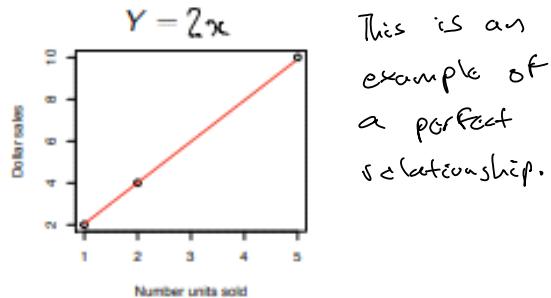
Functional relation

- Relation of the form

$$Y = f(X),$$

where X , Y are variables, and f is a function.

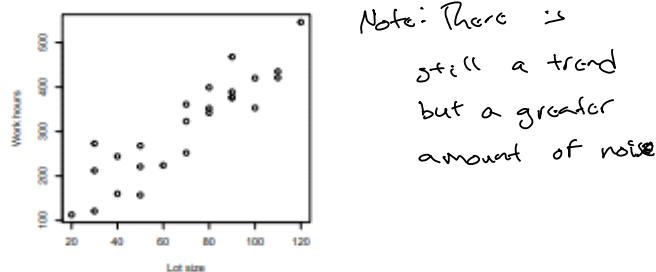
- Example: Relation between dollar sales (Y) of a product sales sold \$2 per unit and number of units sold (X):



All observations fall on the line of functional relationship.

Statistical relation

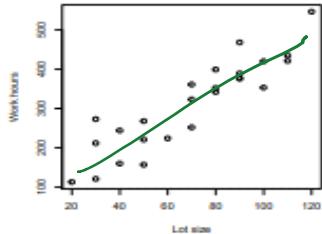
- Not a perfect relation.
- Example: A company produces replacement parts. It produces lots of varying size. The relation between the lot size and work hours is a statistical relation.



Statistical relation

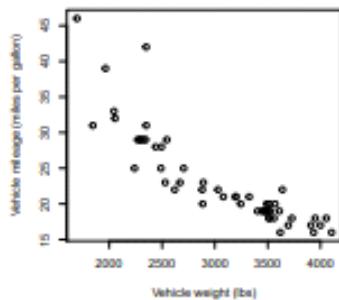
- ▶ Example (contd): There is a relation between X and Y : the higher the lot size, the higher the work hours tend to be.
- ▶ Perfect relation?
No! since there is noise and data points are scattered around the trend.
- ▶ Two lots with $X = 40$ have different Y .
- ▶ Linear or non-linear statistical relation?

Linear statistical relation



Statistical relation

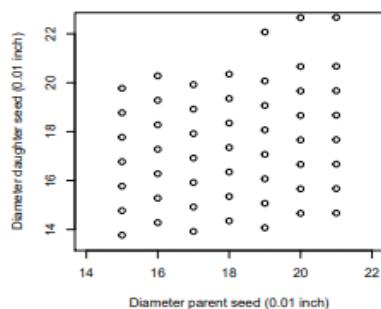
- ▶ Example: Weight and mileage for 54 cars.
- ▶ Functional or statistical relation?
- ▶ Linear statistical relation?



Galton's early considerations of regression

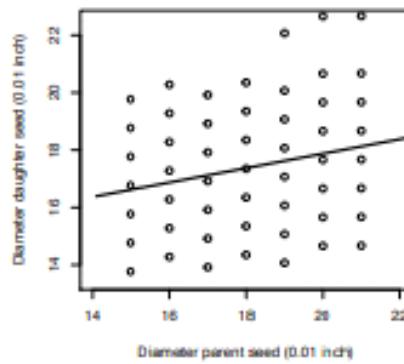
- ▶ Sir Francis Galton, English Victorian statistician, sociologist, psychologist, anthropologist, etc.
- ▶ Work on inherited characteristics of sweet peas ⇒ initial conceptualization of linear regression.
- ▶ In 1875, Galton distributed packets of sweet pea seeds to seven friends who harvested seeds from the new generations of plants and returned them to Galton.
- ▶ Galton plotted the diameter of the daughter seeds against the diameter of the mother seeds [Galton, 1894].

Galton's early considerations of regression



Galton's early considerations of regression

- ▶ Mean diameter of daughter seeds from a particular diameter of mother seed approximately a straight line with positive slope
Tendency of diameter of daughter seeds to vary with diameter of mother seeds
- ▶ Constant variability for diameter of daughter seeds from a particular diameter of mother seed
Random scatter around this tendency



Notation and general concepts

- ▶ **Model:** mathematical expression to describe the behavior of a random variable of interest
- ▶ **Response variable or outcome Y :** variable of interest
- ▶ **Predictor or independent variables X :** known constant variables thought to provide information on the behavior of Y
- ▶ Subscript on Y and X identifies the particular unit from which the observation was taken (X_5 for unit 5)
- ▶ **Parameters:** control behavior of the model; usually represented by Greek letters (β, σ); unknown constants to be estimated from the sample
- ▶ **Linear model:** model linear in the parameters

Note: A model is a representation of reality.
No model is 100% accurate but can be close to matching reality.

Examples

- Dollar sales of a product sales sold \$2 per unit and number of units sold

$$Y = \beta X$$

- Diameter of daughter seeds and diameter of mother seeds

$$Y = \beta X + \varepsilon$$

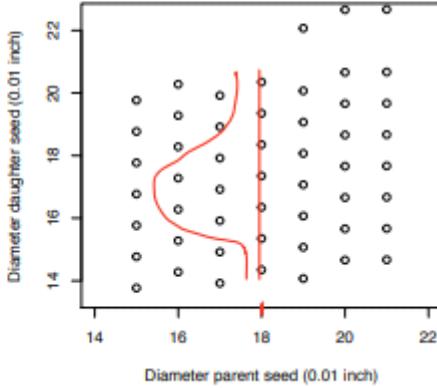
Basic concepts

Two characteristics of a statistical relation:

1. Tendency of Y to vary with X
2. Random scatter around this tendency

In a regression model:

1. The mean of Y vary in a systematic fashion with X
2. Probability distribution of Y for any given value of X



Data collection for regression analysis

Note: observational studies can't conclude cause and effect, only correlation.

► Observational study

- Investigator has no control over the explanatory variables (X)
- Limitation: not adequate for cause-and-effect
A strong association does not necessarily mean a cause-and-effect relationship

► Experiment

- Investigator exercises control over the explanatory variables (X) through random assignment
- Random assignment balances out effect of other variables that might affect Y
- Gold standard for cause-and-effect conclusions

Example of observational study

Study the relationship between age of employees (X) and number of days of illness last year (Y)

- Observational data because we can't control age or # of sick days-
- An observed association between X and Y does not necessarily imply that X explains Y

- Note: There may be other factors that we have not looked at.

Example of experiment

Study the relationship between productivity and length of training of analysts working in a bank:

1. 30 analysts considered
2. randomly select 10 analysts that will be trained for 2 weeks; randomly select 10 other analysts that will be trained for 5 weeks; the 10 remaining will be trained for 8 weeks
3. productivity of the 30 analysts observed for a fixed time after the training

- Experiment because investigators can manipulate the value of x
 - e.g. 8 weeks
 - This could have cause and effect.

Cause-and-effect / Causation

- We observe an association between Y and X
- Does changing one of the variables imply the other to change?
- Mechanisms that can result in an observed association between Y and X :

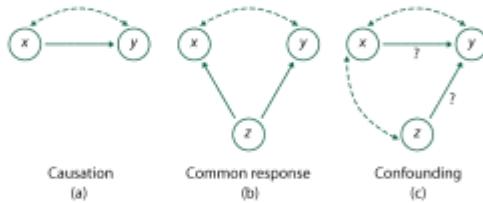


Figure 1: The dashed arrows represent association and the solid ones cause and effect link. The variable x is explanatory, y is response, and z is a lurking variable.

Regression analysis by itself provides no information about causation. Be careful in drawing causal conclusions

Overview of the steps in regression analysis

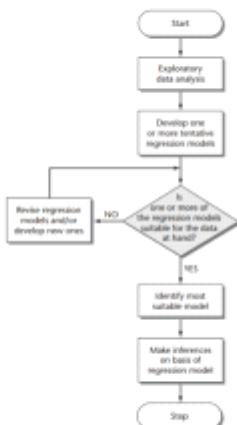


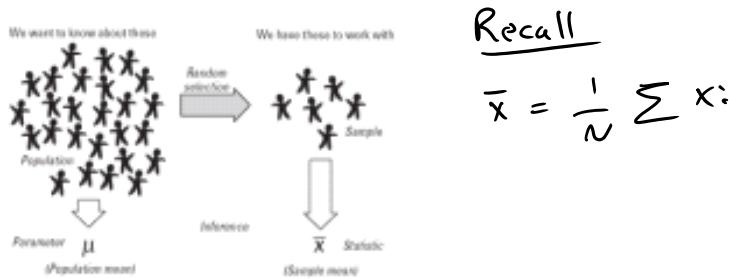
Figure 2: The steps in regression analysis [Kutner et al., 2004, p.14]

Three main purposes of regression analysis

1. **Describe:** describe the relation between diameter of daughter seeds and diameter of mother seeds.
2. **Control:** control the length of training to maximize productivity constrained by costs.
3. **Predict:** predict future demand for a product.

Parameters, estimators, and estimates

- ▶ **Parameter:** quantity of interest, quantity describing a population (or model).
A parameter is a constant (constant/random) quantity.
- ▶ **Estimator:** rule for calculating an estimate of parameter.
An estimator is a random (constant/random) quantity.
- ▶ **Estimate:** result of the estimator (for a given sample).
An estimate is a constant (constant/random) quantity.



Toluca company example¹

- ▶ Toluca Company produces replacement parts for refrigeration equipment
- ▶ Produces lots of varying size
- ▶ Cost improvement: find optimal lot size
- ▶ Key input: relationship between lot size and labor hours
- ▶ Data: lot size X and work hours Y for 25 production runs

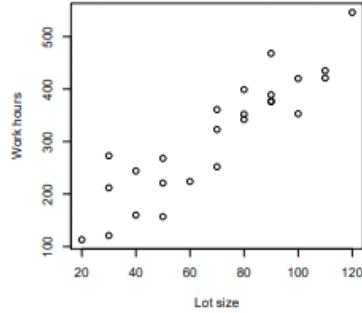
Run <i>i</i>	Lot size X_i	Work hours Y_i
1	80	399
2	30	121
...
24	80	342
25	70	323

← how much time it takes to produce the items
ex / 399 hours to create 80 fridges

¹From [Kutner et al., 2004], page 19

Toluca company example

From the scatter plot:
- looks like a linear model



Simple linear model

Suppose we have n observed pairs (X_i, Y_i) , $i = 1, \dots, n$. The simple linear model is:

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

where

- ↳ assumptions
- ▶ Y_i is the observed value of Y on unit i ,
 - ▶ β_0 and β_1 are parameters,
 - ▶ X_i is the observed value of X on unit i , and
 - ▶ ε_i are random errors that have zero mean $E(\varepsilon_i) = 0$, with common variance $\text{Var}(\varepsilon_i) = \sigma^2$, and pairwise independent.

$$\varepsilon_i \perp \varepsilon_j, i \neq j$$

Simple linear model

Exercise 1

Show that the random errors satisfy

$$E(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{if } i \neq j \\ \sigma^2 & \text{if } i = j \end{cases}$$

Recall Assumptions about random errors

- 1) $E(\varepsilon_i) = 0$
- 2) $\text{Var}(\varepsilon_i) = \sigma^2$
- 3) pairwise independent; thus $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$

For this proof there are 2 cases $i=j$ & $i \neq j$

- $i=j$:

$$\text{wts } E(\varepsilon_i + \varepsilon_i) = E(2\varepsilon_i) = \sigma^2$$

$$\text{we know: } \text{Var}(\varepsilon_i) = \sigma^2$$

$$\text{so } \text{Var}(\varepsilon_i) = E(\varepsilon_i^2) - \underbrace{(E(\varepsilon_i))^2}_0 \quad ; \begin{array}{l} \text{first assumption says } E(\varepsilon_i) = 0 \\ \text{second assumption says } \text{Var}(\varepsilon_i) = \sigma^2 \end{array}$$

$$\sigma^2 = \sigma^2 - 0$$

$$\sigma^2 = \sigma^2$$

- $i \neq j$: we want to show $E(\varepsilon_i \varepsilon_j) = 0$

$$0 = E(\varepsilon_i \varepsilon_j) - \underbrace{E(\varepsilon_i)E(\varepsilon_j)}_0, \text{ first assumption}$$

$$0 = E(\varepsilon_i \varepsilon_j)$$

$$\left\{ \begin{array}{l} \text{Recall} \\ \text{Var}(x) = E(x^2) - (E(x))^2 \\ \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \end{array} \right.$$

Important features

Simple linear model

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

constant \Rightarrow variables and constant are not random.

1. The response Y_i is a sum of two terms:

- A constant term
- A random term

The outcome Y_i is random (constant/random)

2. $E(Y_i) = \beta_0 + \beta_1 X_i$, where $E(Y_i)$ is a shortcut for $E(Y_i | X_i)$
the mean of Y when $X = X_i$.

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_i + \varepsilon_i) = E(\beta_0) + E(\beta_1 X_i) + \underbrace{E(\varepsilon_i)}_{\text{linearity}} \\ &= \beta_0 + \beta_1 X_i \end{aligned}$$

Parameters are always constant. We don't know them but they are constant

Y_i is constant + random = random

Thus, the functional relationship between the true mean of Y_i and X_i is a straight line with intercept β_0 and slope β_1

Important features

Simple linear model

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

3. $\text{Var}(Y_i) = \sigma^2$, where $\text{Var}(Y_i)$ is a shortcut for $\text{Var}(Y_i|X_i)$ the mean of Y when $X = X_i$.

Variance

$$\text{Var}(y_i) = \text{Var}(\underbrace{\beta_0 + \beta_1 x_i}_{\text{constant}} + \varepsilon_i) = \text{Var}(\varepsilon_i) = \sigma^2$$

4. The outcomes Y_i are pairwise independent because the errors ε_i are pairwise independent.

i.e. $y_i \perp \text{from } y_j$ when $i \neq j$

Recall

- Variance is not linear.
- you could use expectation but too hard.
- $\beta_0 + \beta_1 x_i$ is a constant

$$\text{Var}(\text{constant} + r.v) = \text{Var}(r.v)$$

Reminder: normal distribution

A random variable X is normal if its probability density function is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\},$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ are the parameters of the distribution. We say that X is normally distributed with mean $E(X) = \mu$ and variance $\text{Var}(X) = \sigma^2$ and we write

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Additional Assumption

Sometimes we make an additional assumption that random errors are normally distributed.

- This assumption is only used when explicitly stated

Simple linear model with normal errors

- ▶ The random errors are sometimes assumed to be normally distributed.
- ▶ Simple linear model with normal errors:

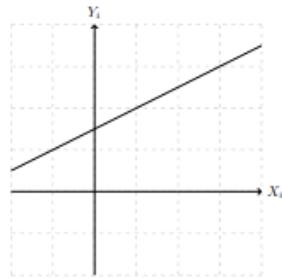
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

where

- ▶ β_0 and β_1 are parameters,
- ▶ ε_i are independently and identically distributed (i.i.d.) with normal distribution with mean 0 and variance σ^2 .

- ▶ In what follows, we suppose a simple linear model (errors not necessarily normal) unless otherwise specified.

Interpretation of the regression parameters



- ▶ If the scope of the model includes $X = 0$, the **intercept β_0** is the mean of Y when $X = 0$ (no meaning otherwise)
- ▶ The **slope β_1** is the change in the mean of Y per unit increase of X

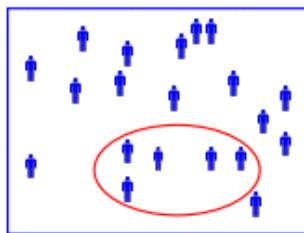
?

Estimation of the parameters

- Postulated model:

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

- Observed values (X_i, Y_i)
- Parameters β_0 and β_1 unknown and to be estimated from the sample.
- Two estimation methods:
 1. Least squares
 2. Maximum likelihood
- ⇒ Estimates $\hat{\beta}_0$ and $\hat{\beta}_1$



- See the model as describing the population
- Select sample at random
- observe x & y values
- estimate β_0 and β_1 from the sample

$\hat{\cdot}$ means estimate, not true value.

Method of least squares

Simple linear model:

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

- Parameters β_0 and β_1 to be estimated from the data.
- Goal: find the best estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ given the data.
- What does best mean?
- Least square: best by criterion

We want to minimize the sum of the errors to get the best fit.

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \underbrace{\beta_0 - \beta_1 X_i}_{y_i - \bar{E}(y_i)})^2$$

$y_i - \beta_0 - \beta_1 x_i$ is the deviation of Y_i from its expected value.

- Least square estimators of β_0 and β_1 : $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize criterion Q .

Least square estimators

Find $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize criterion

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

given the data.

1. Write the normal equations (derivatives of Q set to 0).
2. Find the critical points (solution of the normal equations).
3. Determine whether the critical point is a maximum or a minimum (we will skip this step).

$$\begin{aligned} 1) \quad & \frac{\partial Q}{\partial \beta_0} = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \cdot (-1) \\ & 0 = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \\ \frac{\partial Q}{\partial \beta_1} & = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) \\ 0 & = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) \end{aligned}$$

2) β_0 : Critical points

$$\begin{aligned} 0 & = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \\ 0 & = \sum_{i=1}^n y_i - \beta_0 n - \beta_1 \sum_{i=1}^n x_i \\ \sum_{i=1}^n \beta_0 & = \sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i \\ n \beta_0 & = \sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i \\ \beta_0 & = \frac{\sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i}{n} \\ \beta_0 & = \bar{Y} - \beta_1 \bar{X} \end{aligned}$$

$$\begin{aligned} \text{3) } \beta_1 & : \text{critical point} \\ 0 & = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) \\ 0 & = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \cdot x_i \\ 0 & = \sum_{i=1}^n (y_i x_i - \beta_0 x_i - \beta_1 x_i^2) \\ 0 & = \sum_{i=1}^n y_i x_i - \beta_0 \sum_{i=1}^n x_i - \beta_1 \sum_{i=1}^n x_i^2 \\ \beta_1 \sum_{i=1}^n x_i^2 & = \sum_{i=1}^n y_i x_i - \beta_0 \sum_{i=1}^n x_i \\ \beta_1 & = \frac{\sum_{i=1}^n y_i x_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2} \\ \hat{x} \hat{x} \beta_1 & = \sum_{i=1}^n y_i x_i - (\sum_{i=1}^n \frac{y_i}{n} - \beta_0 \sum_{i=1}^n \frac{x_i}{n})(\sum_{i=1}^n x_i) \\ \sum_{i=1}^n x_i^2 \beta_1 & = \sum_{i=1}^n y_i x_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n x_i + \beta_0 \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i^2 \beta_1 - \frac{1}{n} \beta_0 \sum_{i=1}^n x_i & = \sum_{i=1}^n y_i x_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \\ \beta_1 \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n x_i \right) & = \sum_{i=1}^n y_i x_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \\ \beta_1 & = \frac{\sum_{i=1}^n y_i x_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n x_i} \end{aligned}$$

Least square estimators

Least square estimators of β_1 and β_0 :

$$\begin{aligned} \hat{\beta}_1 & = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2}, \\ \hat{\beta}_0 & = \bar{Y} - \hat{\beta}_1 \bar{X}, \end{aligned}$$

where

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

are the sample mean of Y and X , respectively.

Least square estimators

Exercise 2

Show that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

Least square estimators

Least square estimators of β_1 and β_0 :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2}, \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X},\end{aligned}$$

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

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

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

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

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

$$\begin{aligned}&= \frac{\sum_{i=1}^n x_i y_i - \frac{2}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i + \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{2}{n} (\sum_{i=1}^n x_i)^2 + \frac{1}{n} (\sum_{i=1}^n x_i)^2} \\ &= \frac{\sum_{i=1}^n x_i y_i - \frac{2}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i + \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2} \\ &= \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i)^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}\end{aligned}$$



Regression equation

$$\text{Model: } y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad E(y_i) = \beta_0 + \beta_1 x_i; \text{ True mean of } y \text{- when } x = x$$

$\underbrace{+ \varepsilon_i}_{\text{to understand as } E(y_i | x_i)}$

Regression equation or fitted regression line

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

where \hat{Y} is the estimated mean of the response variable at level X of the explanatory.

Gauss-Markov theorem

Theorem 1

Consider the simple linear model

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

Suppose that the following assumptions concerning the random errors (called Gauss-Markov assumptions) are satisfied:

- ▶ They have mean zero: $E(\varepsilon_i) = 0$,
- ▶ They are homoscedastic: $\text{Var}(\varepsilon_i) = \sigma^2 < \infty$, and
- ▶ There are uncorrelated $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$.

Then the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased and have minimum variance among all unbiased linear estimators.

Proof of the Gauss-Markov theorem

Step 1 Exercise: Prove that the least squares estimators are unbiased, i.e. prove that

$$E(\hat{\beta}_1) = \beta_1 \quad \text{and} \quad E(\hat{\beta}_0) = \beta_0$$

Step 2 To be proven later: The least squares estimators have minimum variance among all unbiased linear estimators.

(Heck NEX)

page 11

Parameter θ
 Estimator $\hat{\theta}$
 $\hat{\theta}$ is an unbiased estimator of θ
 if $E(\hat{\theta}) = \theta$
 otherwise, $\hat{\theta}$ is biased and its bias is $E(\hat{\theta}) - \theta$

Prof asks to
 prove stuff
 if shift is
 biased

$$\begin{aligned}
 \hat{\beta}_1 &= \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \\
 E(\hat{\beta}_1) &= E\left(\frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right), \text{ use linearity} \\
 &= \frac{1}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \left\{ E(\sum x_i y_i) - E\left(\frac{1}{n} \sum x_i \sum y_i\right) \right\} \\
 &= \frac{1}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \left\{ \sum x_i E(y_i) - \frac{1}{n} \sum x_i \sum E(y_i) \right\} \\
 &= \frac{1}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \left\{ \sum x_i (B_0 + \beta_1 x_i) - \frac{1}{n} \sum x_i \sum (B_0 + \beta_1 x_i) \right\} \\
 &= \frac{\beta_1 \sum x_i^2 - \frac{1}{n} \beta_1 (\sum x_i)^2}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \\
 &= \beta_1
 \end{aligned}$$

Note
 $E\left(\frac{x}{y}\right) \neq \frac{E(x)}{E(y)}$
 usually!

$y_i = B_0 + \beta_1 x_i + \varepsilon_i$
 $E(y_i) = B_0 + \beta_1 x_i$



Conclusion
 $E(\hat{\beta}_1) = \beta_1$, i.e.
 $\hat{\beta}_1$ is an unbiased estimator of β_1 .

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

$$E(\hat{\beta}_0) = E(\beta_0) - E(\hat{\beta}_1)\bar{x}$$

, linearity of expectation

Recall $\hat{\beta}_1$

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

$$\begin{aligned} E(y) &= E\left(\frac{1}{n} \sum y_i\right) = \frac{1}{n} \sum E(y_i) = \frac{1}{n} \sum (\beta_0 + \beta_1 x_i) \\ &= \frac{c}{n} n\beta_0 + \frac{1}{n} \beta_1 \sum x_i \\ &= \beta_0 + \beta_1 \bar{x} \end{aligned}$$

$$E(\hat{\beta}_1) = \beta_1 \rightarrow \text{so } \hat{\beta}_1 \text{ is an unbiased estimator.}$$

Conclusion: $E(\hat{\beta}_0) = \beta_0$, i.e.

$\hat{\beta}_0$ is an unbiased estimator of β_0

Toluca company example

Using R, we find:

$$\begin{aligned}\sum_{i=1}^n X_i &= 1750 & \sum_{i=1}^n Y_i &= 7807 & \sum_{i=1}^n X_i Y_i &= 617180 \\ \sum_{i=1}^n X_i^2 &= 142300 & n &= 25\end{aligned}$$

Exercise 3

1. Compute the least squares estimates of β_1 and β_0 .
2. What is the regression equation?
3. Interpret the parameters.

Take formula in plug in

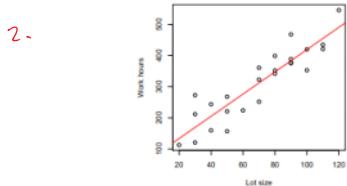
$$1. \hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n} (\sum_{i=1}^n X_i)^2} = \frac{617180 - \frac{1}{25}(1750)(7807)}{142300 - \frac{1}{25}(1750)^2} = 3.570202$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{25} \sum_{i=1}^n Y_i - 3.570202 \frac{1}{25} \sum_{i=1}^n X_i = \frac{1}{25}(7807) - 3.570202(175) = 62.36586$$

Test is more proof based

Toluca company example

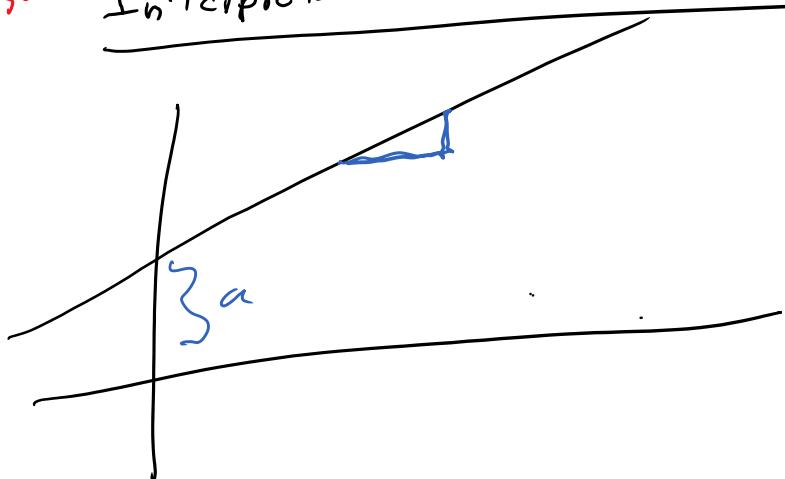
Check if 0 is in the range otherwise don't interpret the intercept because $x=0$ is not in the range of the observed x values.



$$\begin{aligned}\hat{\beta}_1 &= 3.5702 \\ \hat{\beta}_0 &= 62.36 \\ \hat{y} &= 62.36 + 3.5702 \times \text{Estimation mean of } y.\end{aligned}$$

The estimated work time increases by 3.57 hours when the lot size increases by 1 unit (one additional unit produced).

3c Interpretation of the regression parameters



$$y = a + bx$$

a : obtained y when
 $x = 0$

$$b = \frac{\Delta y}{\Delta x}$$

if $\Delta x = 1$

$$b = \Delta y$$

b is the difference in Y when x increases by 1 unit.

Copy down the slides, later! ↗

Toluca company example: R output

$\text{Hours} = \beta_0 + \beta_1 \text{Size} + \epsilon$

Y *X*

Call:
 $\text{lm}(\text{formula} = \text{Hours} \sim \text{Size}, \text{data} = \text{toluca})$

Residuals:

Min	1Q	Median	3Q	Max
-83.876	-34.088	-5.982	38.826	103.528

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	62.366	26.177	2.382	0.0259 *
Size	3.570	0.347	10.290	4.45e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

β_0

β_1

Residual standard error: 48.82 on 23 degrees of freedom
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8138
F-statistic: 105.9 on 1 and 23 DF, p-value: 4.449e-10

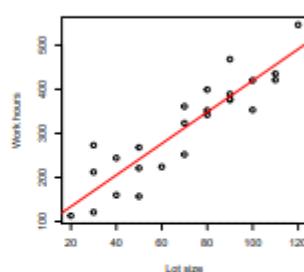
Preliminary exercise: Toluca company example

- ▶ Regression equation (or fitted regression line)

$$\hat{Y} = 62.37 + 3.5702X,$$

where

- ▶ X is the lot size, and
- ▶ Y is the work hours.



What is:

- ▶ The predicted work hours for a new production run for a lot size of 60?
 $\hat{Y} = 62.37 + 3.5702 \cdot 60 = 276.582$

- ▶ The estimated population mean work hours for a lot size of 60?

$$\hat{Y} = 62.37 + 3.5702 \cdot 60 = 276.582$$

Predicted values and residuals

- ▶ Fitted regression line:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

- ▶ **Fitted value:** value of Y computed from the regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

Fitted value \hat{Y}_i used as:

- ▶ **Prediction** of the value of Y for particular value X_i of X .
Sometimes written $\hat{Y}_{\text{pred},i}$.
- ▶ **Estimate** of the population mean of Y for particular value X_i of X .
- ▶ A **residual** is the deviation of the observed value Y_i from the fitted value \hat{Y}_i

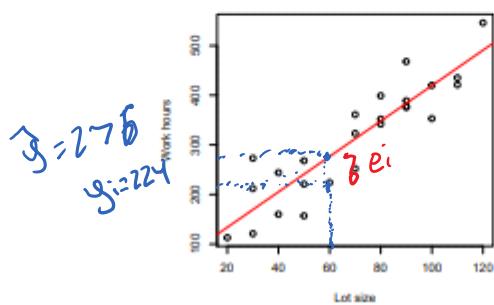
$$e_i = Y_i - \hat{Y}_i$$

Toluca company example

For production run 6, the lot size was 60 and 224 work hours were required.

Exercise 4

1. What is the fitted value for this observation?
2. What is the residual?
3. Where can we read the observed work hours (Y), the fitted work hours, and the residual in the scatterplot?



Exercises

- ▶ From the textbook²
 - ▶ 1.20
 - ▶ 1.21
- ▶ From the slides³
 - ▶ Slide 26
 - ▶ Slide 37
 - ▶ Slide 40
 - ▶ Slide 41
 - ▶ Slide 47

go do this or slap
your self!
karen style slide
questions dawg.

²Solutions in the student manual (CD) provided with the textbook

³Partial solutions posted on Quercus

Proofs

$$\text{1) } \sum e_i = \sum (y_i - \hat{B}_0 - \hat{B}_1 x_i) = \sum y_i - \sum \hat{B}_0 - \sum \hat{B}_1 x_i = \sum y_i - n \hat{B}_0 - \hat{B}_1 \sum x_i = \sum y_i - \sum y_i + \hat{B}_1 \sum x_i - \hat{B}_1 \sum x_i = 0$$

$$e_i = y_i - \hat{y}_i = y_i - (\hat{B}_0 - \hat{B}_1 x_i)$$

$$\text{sub } \hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x} \Rightarrow \sum y_i - n \bar{y} + n \hat{B}_1 \bar{x} - \hat{B}_1 \sum x_i$$

3) To prove: $\sum y_i = \sum \hat{y}_i$

$$\text{From ex1/ we know } \sum e_i = \sum (y_i - \hat{y}_i) = 0 \Leftrightarrow \sum y_i - \sum \hat{y}_i = 0$$

$$\Rightarrow \sum y_i = \sum \hat{y}_i$$

4.)

$$\begin{aligned} \sum x_i e_i &= \sum x_i (y_i - \hat{y}_i) \quad , \text{ definition of } e_i = y_i - \hat{y}_i \\ &= \sum (x_i y_i - x_i \hat{y}_i) = \sum x_i y_i - \sum x_i \hat{y}_i \quad , \text{ summation properties + expand} \\ &= \sum x_i y_i - \sum x_i (\hat{B}_0 + \hat{B}_1 x_i) = \sum x_i y_i - \hat{B}_0 \sum x_i - \hat{B}_1 \sum x_i^2 \quad , \quad \hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i \quad , \quad \text{summation properties} \\ &= \sum x_i y_i - (\hat{B}_0 \bar{x}) \sum x_i - \hat{B}_1 \sum x_i^2 \quad , \quad \hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x} \\ &= \sum x_i y_i - \bar{y} \sum x_i - \hat{B}_1 \sum x_i^2 \quad , \quad \text{expand} \\ &= \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i + \left(\frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right) \cdot \frac{1}{n} (\sum x_i)^2 - \left(\frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right) \sum x_i^2 \quad , \quad \bar{y} = \frac{\sum y_i}{n} \quad , \quad \hat{B}_1 = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \\ &= \left(\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right) \left[1 + \frac{\frac{1}{n} (\sum x_i)^2 - \sum x_i^2}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right] \quad , \quad \text{factor } \left(\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right) \\ &= \left(\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right) \left[1 - \frac{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right] \quad , \quad \text{factor } (-1) \text{ from} \\ &= \left(\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right) [1 - 1] \\ &= 0 \end{aligned}$$

Properties of the fitted regression line

Exercise 1

Prove the following properties:

$$1. \sum_{i=1}^n e_i = 0$$

$$2. \sum_{i=1}^n e_i^2 \text{ is a minimum of function}$$

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

$$3. \sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$$

$$4. \sum_{i=1}^n X_i e_i = 0$$

$$5. \sum_{i=1}^n \hat{Y}_i e_i = 0$$

6. The regression line passes through (\bar{X}, \bar{Y})

Term Test I: Monday
Jan 28th, 2019
IC1

Unbiased estimator of the population mean of Y Exercise 2 $\hat{\theta}$ Show that \hat{Y}_i is an unbiased estimator of the population mean of Y given a value X_i of X .

Proof of unbiased of something else

we have parameter θ

$$\theta \leftarrow \text{estimator} = \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

we want:

$$E(\hat{\theta}) = \theta \leftarrow \text{To prove}$$

$$E(\hat{\theta}) = E(\hat{y}_i) = E(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= E(\hat{\beta}_0) + x_i E(\hat{\beta}_1)$$

$$\equiv \theta$$

, linearity

, $\hat{\beta}_0$ unbiased estimator of β_0 , $\hat{\beta}_1$ unbiased estimator of β_1

Conclusion: \hat{y}_i is an unbiased estimator of $\beta_0 + \beta_1 x_i$
 pop/true func of y when
 $x = x_i$)

} Type of exercise you might find on tests!

Recall

mean of y given a value of x then

$$E(y_i) = \beta_0 + \beta_1 x_i$$

$$\text{where } y_i = \beta_0 + \beta_1 x_i + \varepsilon$$

$$\text{so, } \theta = \beta_0 + \beta_1 x_i$$

Motivation

- ▶ How do we usually measure the variation?

- ▶ Deviation of an observation Y_i from the mean of Y :
- ▶ Measure of total variation: the **total sum of squares**

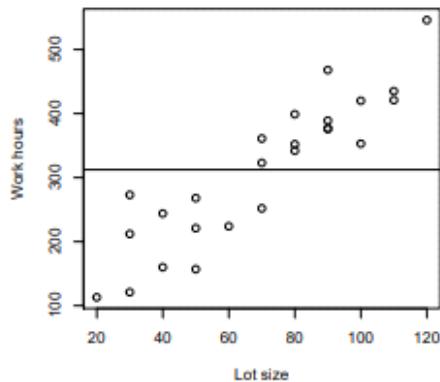
$$SSTot = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

3 Types of Sum of Squares

Total sum of squares: total deviation in the response variable.

Distance between y and observed

Total sum of squares



$$SSTot = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Partitioning

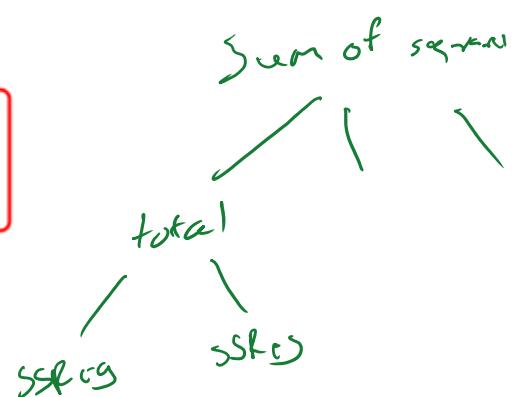
Partitioning of the total sum of squares

$$SSTot = SSReg + SSRRes \quad (1)$$

where

- $SSReg = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ is the **regression sum of squares**
- $SSRRes = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ is the **residual sum of squares**

Goal is to exercise



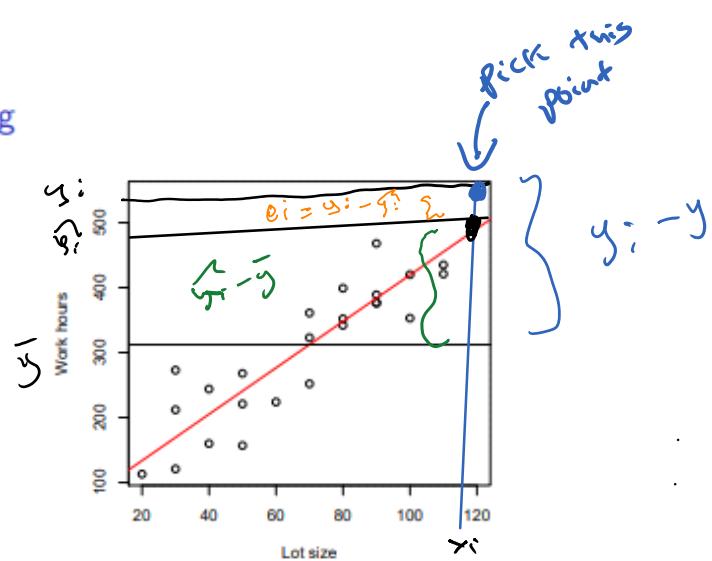
Partitioning

Do this!

Exercise 3

- Prove Equation (1)
- Show that $SSReg = \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2$

Partitioning



$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Toluca example

We had:

$$\sum_{i=1}^n X_i = 1750 \quad \sum_{i=1}^n Y_i = 7807 \quad \sum_{i=1}^n X_i^2 = 142300$$

$$\sum_{i=1}^n Y_i^2 = 2745173 \quad \hat{\beta}_1 = 3.57 \quad \hat{\beta}_0 = 62.37$$

Exercise 4

Compute $SSTot$, $SSReg$, and $SSRes$.

Hint: we have (to prove)

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

Toluca example

Degrees of freedom

- ▶ Proper definition requires to understand the underlying geometry of the problem
- ▶ For now: the number of values in the calculation of a statistic that can freely vary
- ▶ $SSTot$ has $n-1$ degrees of freedom

$$\sum (y_i - \bar{y})^2, y_i \dots n \text{ times unknown, there are } n-1 \text{ degrees of freedom}$$

- ▶ $SSRes$ has $n-2$ degrees of freedom

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{B}_0 - \hat{B}_1 x_i)^2, \text{ one is fixed.}$$

- ▶ $SSReg$ has 1 degrees of freedom

$$\sum (\hat{y}_i - \bar{y})^2 = \sum (\hat{B}_0 + \hat{B}_1 x_i - \bar{y})^2$$

, \hat{B}_0 and \hat{B}_1 are random i.e. based on y_i

, \hat{B}_0, \hat{B}_1 are random but subtract \bar{y} , so $2-1=1$

Mean squares

- ▶ **Mean squares:** sum of squares divided by its associated degrees of freedom
- ▶ **Regression mean squares:**

$$MSReg = \frac{SSReg}{1} = SSReg$$

- ▶ **Residual mean squares:**

$$MSRes = \frac{SSRes}{n-2}$$

Comment 1

- ▶ The total mean squares $\frac{SSTot}{n-1}$ is _____
- ▶ $MSReg$ and $MSRes$ do not add to $\frac{SSTot}{n-1}$

Analysis of variance table

Analysis of variance (ANOVA) table to display the sum of squares and degrees of freedom

Source of variation	Sum of squares	df	Mean squares
Regression	$SSReg$	1	$MSReg = \frac{SSReg}{1}$
Residual	$SSRes$	$n - 2$	$MSRes = \frac{SSRes}{n-2}$
Total	$SSTot$	$n - 1$	

Toluca example

Analysis of variance (ANOVA) table for Toluca example

Source of variation	Sum of squares	df	Mean squares
Regression	$SSReg$	1	$MSReg = \frac{SSReg}{1}$
Residual	$SSRes$	$n - 2$	$MSRes = \frac{SSRes}{n-2}$
Total	$SSTot$	$n - 1$	

Coefficient of determination

fraction of total variation
of y explained by the model.

Coefficient of determination:

$$R^2 = \frac{SS_{Reg}}{SSTot} = 1 - \frac{SS_{Res}}{SSTot}$$

Fraction of the variation in Y explained by the model (i.e. its linear relationship with X).

Exercise 5

Compute and interpret the coefficient of determination for the toluca example.

study
up till here
for TFI
!!

Simple linear model with normal errors

Model supposed in this section:

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{constant}} + \varepsilon_i, \quad \varepsilon_i \sim \text{Normal distribution}$$

where

- β_0 and β_1 are parameters
- ε_i are i.i.d. with normal distribution with mean 0 and variance σ^2 , i.e. $\varepsilon_i \sim N(0, \sigma^2)$

Exercise 6

Show that Y_i follows a $N(\beta_0 + \beta_1 X_i, \sigma^2)$.

sometimes

Admin: test
monthly r starting
from here not on
the test

add an additional assumption

y_i is normally distributed; ε_i normally distributed and
 $y_i = \text{est} + \varepsilon_i$

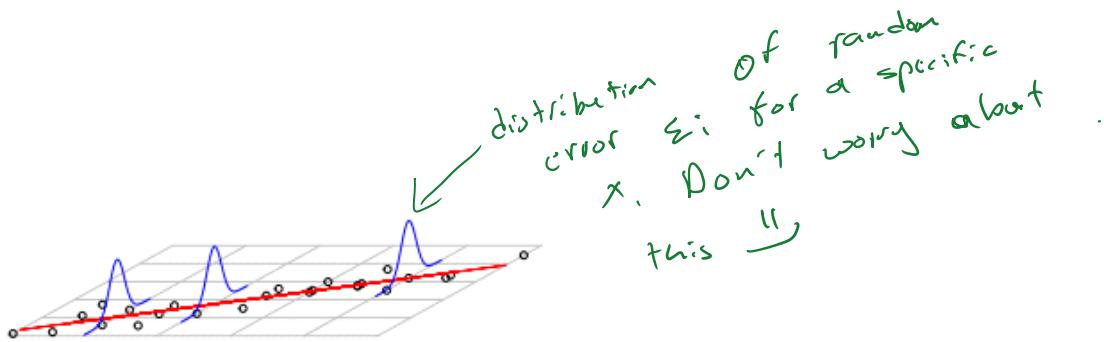
constant + normal distribution = normal distribution that
is shifted

$$E(y_i) = \beta_0 + \beta_1 x_i$$

$$\text{Var}(y_i) = \sigma^2$$

↑ previously shown !!

Simple linear model with normal errors



Maximum likelihood estimation of the parameters

- ▶ When the probability distribution of the errors is specified, estimators of the parameters can be obtained by **maximum likelihood estimation (mle)**.
- ▶ **General idea:** choose the values of the parameters that are the most likely given the sample data.
- ▶ With mle the sample data are viewed as fixed and the parameters are allowed to vary freely.
- ▶ For a sample of n iid observations z_1, z_2, \dots, z_n coming from a distribution with density function f depending on a parameter (or parameter vector) θ , the **likelihood function** is

$$\mathcal{L}(\theta | z_1, z_2, \dots, z_n) = \prod_{i=1}^n f(z_i)$$

Maximum likelihood estimation of the parameters

- ▶ The value of θ that maximizes \mathcal{L} is the **maximum likelihood estimator**.
- ▶ It is more convenient to work with the natural logarithm of the likelihood function, called the log-likelihood function $\log \mathcal{L}$.
- ▶ \mathcal{L} and $\log \mathcal{L}$ are maximized for the same values of the parameters.

Likelihood function

- ▶ Probability distribution function (pdf) of a $\mathcal{N}(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- ▶ Probability distribution function (pdf) of Y_i :

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

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right\}$$

Likelihood function

Exercise 7

Suppose that we have n observations Y_1, Y_2, \dots, Y_n . Show that the likelihood function for the simple linear model with normal errors is

$$\mathcal{L}(\beta_0, \beta_1, \sigma^2 | Y_1, \dots, Y_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right\}$$

Hint: the Y_i 's are iid $\mathcal{N}(\beta_0 + \beta_1 X_i, \sigma^2)$

$$\prod_{i=1}^n f(y_i) = \prod_{i=1}^n \underbrace{\exp(u_i)}_{\text{Recall}} = \exp \left\{ \sum_{i=1}^n u_i \right\}$$

Maximum likelihood estimators

Exercise 8

Find the maximum likelihood estimators of β_0 , β_1 , and σ^2 .

Hint:

1. Write down the log-likelihood function.
2. Write down the normal equations, i.e. the partial derivatives of the log-likelihood function with respect to the three parameters.
3. Solve the normal equations.

3. The MLE are

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{x})(x_i - \bar{y})}{\sum (y_i - \bar{x})^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n} = \frac{\sum (y_i - \bar{y})^2}{n}$$



This is a biased estimator of σ^2

Hint:

$$\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}, \frac{\partial L}{\partial \sigma^2}$$

Properties of the maximum likelihood estimators

The maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

1. are **unbiased**
2. have **minimum variance among all unbiased linear estimators**
3. are **consistent** (as the number of data observations increases indefinitely, the resulting sequence of estimates converges in probability to the true value of the parameter)
4. are **sufficient** (no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter)

Random variables

- Model supposed in this section:

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

where

- β_0 and β_1 are parameters,
- ε_i are iid with mean 0 and variance σ^2 .
- Quantities computed from random variables are random variables.
- $\Rightarrow \bar{Y}, \hat{\beta}_0, \hat{\beta}_1, \hat{Y}_i, e_i$ are random variables.
- **They have variability!**
- This section gives measures of precision of these random variables.

Variance of a linear function

- Consider an arbitrary linear function of the Y_i 's

$$U = \sum_{i=1}^n a_i Y_i,$$

where a_i are constants.

- General formula for the variance of U

$$\text{Var}(U) = \sum_{i=1}^n a_i^2 \text{Var}(Y_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{Cov}(Y_i, Y_j).$$

- The Y_i 's are pairwise independent and $\text{Var}(Y_i) = \sigma^2$ for all i .
The variance of U reduces to

$$\text{Var}(U) = \sum_{i=1}^n a_i^2 \text{Var}(Y_i)$$

$$= \sigma^2 \sum_{i=1}^n a_i^2 \quad \leftarrow \text{show this up your hand.}$$

Variance of a linear function

$$\text{Linear fct of the } Y_i \text{'s} = \sum_{i=1}^n a_i Y_i$$

Exercise 9

Show that $\bar{Y}, \hat{\beta}_0, \hat{\beta}_1, \hat{Y}_i$ are all linear functions of the Y_i 's. Give the corresponding constants a_i 's for every estimator.

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \sum_{i=1}^n a_i Y_i = \sum_{i=1}^n \frac{1}{n} Y_i \quad \leftarrow \begin{matrix} \text{I slapped myself trying to} \\ \text{figure this out} \end{matrix}$$

where $a_i = \frac{1}{n}$

Important note: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{j=1}^n (X_j - \bar{X})^2$. \leftarrow Same shift, same quantities but different index. So they are equal.

Variance of the sample mean \bar{Y}

- Sample mean as a linear function of the Y_i 's:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \sum_{i=1}^n a_i Y_i$$

with $a_i = \frac{1}{n}$.

- Applying the formula for the variance of a linear function of the Y_i 's:

$$\text{Var}(\bar{Y}) = \sigma^2 \sum_{i=1}^n a_i^2 = \sigma^2 \sum_{i=1}^n \left(\frac{1}{n}\right)^2 =$$

$$\sigma^2 \cdot \frac{n}{n^2} = \frac{\sigma^2}{n} \quad \left\{ \begin{matrix} \text{variance} \\ \text{of} \\ \text{sample} \\ \text{mean.} \end{matrix} \right.$$

- Well-known result!

Variance of the slope $\hat{\beta}_1$

- Slope $\hat{\beta}_1$ as a linear function of the Y_i 's:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sum_{j=1}^n (X_j - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{j=1}^n (X_j - \bar{X})^2}$$

formula for $\hat{\beta}_1$
proved earlier

We have

$$\hat{\beta}_1 = \sum_{i=1}^n a_i Y_i \quad \text{with} \quad a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^n (X_j - \bar{X})^2}.$$

*trick used
to change index
see this
dude as a constant*

$$\hat{\beta}_1 = \sum_{i=1}^n \frac{(x_i - \bar{y})}{\sum_{j=1}^n (x_j - \bar{y})^2} a_j; \quad / \quad \text{now we can identify } a_i$$

Variance of the slope $\hat{\beta}_1$

- Applying the formula for the variance of a linear function of the Y_i 's:

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \sigma^2 \sum_{i=1}^n a_i^2 = \sigma^2 \sum_{i=1}^n \left\{ \frac{(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \right\}^2 \\ &= \sigma^2 \sum_{i=1}^n \left(\frac{(x_i - \bar{x})^2}{\left(\sum_{j=1}^n (x_j - \bar{x})^2 \right)^2} \right) = \frac{\sigma^2}{\left(\sum_{j=1}^n (x_j - \bar{x})^2 \right)^2} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{\sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sigma^2 \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{aligned}$$

$$\begin{aligned} \text{Proof solution} \\ \text{slide 2a} \\ \text{Var}(\hat{\beta}_1) &= \sigma^2 \sum_{i=1}^n a_i^2 = \sigma^2 \sum_{i=1}^n \left\{ \frac{(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \right\}^2 \\ &= \sigma^2 \left\{ \sum_{i=1}^n \left(\frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \right) \right\} = \sigma^2 \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ &\quad \text{could write} \end{aligned}$$

Variance of the intercept $\hat{\beta}_0$

- The intercept $\hat{\beta}_0$ is a linear function of the Y_i 's. Indeed, we have (see exercise on slide 30)

$$\hat{\beta}_0 = \sum_{i=1}^n a_i Y_i \quad \text{with} \quad a_i = \frac{1}{n} - \frac{X_i - \bar{X}}{\sum_{j=1}^n (X_j - \bar{X})^2} \bar{X}.$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad ; \text{ known: } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} y_i$$

Variance of the intercept $\hat{\beta}_0$

- Applying the formula for the variance of a linear function of the Y_i 's:

Variance of the intercept $\hat{\beta}_0$

$$\text{Var}(\hat{\beta}_0) =$$

Variance of \hat{Y}_0 as estimate of the population mean

Exercise 10

Show that

$$\text{Var}(\hat{Y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\}$$

*not independent
of i*

Hint:

*so it's not
confusing*

1. Use

$$\hat{Y}_0 = \bar{Y} + \hat{\beta}_1(X_0 - \bar{X})$$

to write \hat{Y}_0 as a linear function of the Y_i 's (or use exercise on slide 30).

2. Apply the formula for the variance of a linear function of the Y_i 's.
3. Simplify.

*Ex for you
to do! solve
this*

Variance of \hat{Y}_0 as an estimate of the population mean

Variance of \hat{Y}_0 as estimate of the population mean

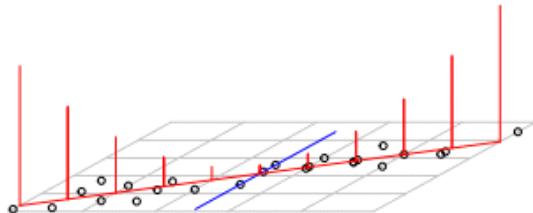
$$\text{Var}(\hat{Y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\}$$

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

Comment 2

- ▶ The minimum value of $\text{Var}(\hat{Y}_0)$ is $\sigma^2 \cdot \frac{1}{n}$ which is attained when $X_0 = \bar{X}$.
- ▶ $\text{Var}(\hat{Y}_0)$ increases as X_0 moves away from \bar{X} .
- ▶ This formula gives the variance when \hat{Y}_0 is considered as an estimate of the population mean of Y for a value X_0 of X .

Variance of \hat{Y}_0 as estimate of the population mean



don't need
to know this
but try to ☺

Variance of a prediction \hat{Y}_{pred_0}

- A prediction of a future observation of Y for a given value X_0 of X is

$$\hat{Y}_{pred_0} = \hat{\beta}_0 + \hat{\beta}_1 X_0.$$

- An estimate of the population mean of Y for a given value X_0 of X is

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0.$$

- However, it is possible to show that

$$\text{Var}(\hat{Y}_{pred_0}) = \text{Var}(\hat{Y}_0) + \sigma^2.$$

Variance of a prediction \hat{Y}_{pred_0}

$$\text{Var}(\hat{Y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{\sum (x_i - \bar{x})^2} \right\}$$

Idea: 

\hat{Y}_0 is an estimator of $\beta_0 + \beta_1 X_0$

\hat{Y}_{pred_0} is an estimator of $\beta_0 + \beta_1 X_0 + \varepsilon_0$,

*3 take variance for
to make sense*

where $\text{Var}(\varepsilon_0) = \sigma^2$. 

- Finally,

$$\text{Var}(\hat{Y}_{pred_0}) = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right\}$$

just add σ^2

Estimates and precision: summary

Quantity	Estimator	Variance of the estimator
β_1	$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$	$\sigma^2 \frac{1}{\sum(X_i - \bar{X})^2}$
β_0	$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$	$\sigma^2 \left\{ \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2} \right\}$
$E(Y_0)$	$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$	$\sigma^2 \left\{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right\}$
Y_0	$\hat{Y}_{pred_0} = \hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$	$\sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right\}$

we don't know $\beta_1, \beta_0, \sigma^2$
 population parameter
 that is unknown.

so, we estimate:

usefull stuff
 write on
 check sheet
 drawing!

Estimates and precision: true and estimated variances

- Derived variances are true variances; they depend on the population (or true) variance of the random errors σ^2 .
- In the simple linear regression model, an estimator of σ^2 is

$$s^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = MSRes.$$

- It is an unbiased estimator, i.e. $E(s^2) = \sigma^2$.
- Estimated variances are obtained by replacing σ^2 with s^2 in the formulae of the true variances.
- We write s^2 instead of Var to indicate an estimated variance instead of a true variance, e.g. the estimated variance of $\hat{\beta}_1$ is

estimated variance $\hat{\beta}_1$

$$s^2(\hat{\beta}_1) = s^2 \frac{1}{\sum(X_i - \bar{X})^2}.$$

- The square root of the estimated variance is sometimes called standard error.

$\sqrt{s^2(\hat{\beta}_1)}$
 = standard error
 of $\hat{\beta}_1$

Toluca example: estimated variances

Exercise 11

- ▶ Compute the standard error of the slope and intercept of the fitted regression line.
- ▶ Compute the point estimate of the mean work hours for a lot of size 85. Compute its standard error.
- ▶ Compute the point estimate of a prediction of the work hours for a lot of size 85. Compute its standard error.

Ex for
you to do!

Exercises

- ▶ From the textbook¹
 - ▶ 2.25, part (a)
 - ▶ 2.24, part (a). Only set up the ANOVA table in the format of Table 2.2
- ▶ From the slides²
 - ▶ Slide 1
 - ▶ Slide 3
 - ▶ Slide 7
 - ▶ Slide 9
 - ▶ Slide 18
 - ▶ Slide 19
 - ▶ Slide 24
 - ▶ Slide 25
 - ▶ Slide 30
 - ▶ Slide 37
 - ▶ Slide 45

¹Partial solutions in the student manual (CD) provided with the textbook

²Partial solutions posted on Quercus

49

Simple linear model with normal errors

- Model supposed in this section:

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

where

- β_0 and β_1 are parameters,
- ε_i are i.i.d. with **normal distribution** with mean 0 and variance σ^2 , i.e. $\varepsilon_i \sim N(0, \sigma^2)$.

- Reminder: we showed that under this model
 $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$.

Most common hypothesis test in simple linear regression

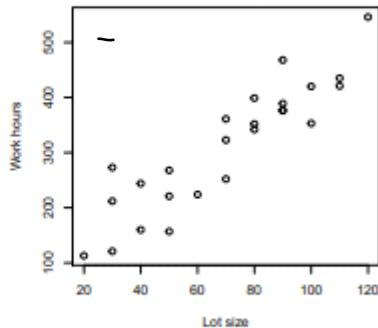
$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\ E(y_i) &= \beta_0 + \beta_1 x_i \end{aligned}$$

- Two sided test:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases}$$

- What does H_0 state?

H_0 states that there is no linear relationship between $E(y_i)$ and x_i :



if $\beta_1 = 0$
 there is no
 linear relationship
 between ...

- In toluca example, we obtained $\hat{\beta}_1 = 3.57$. ← previously $\hat{\beta}_1 \neq 0$
- Is this departure from zero only due to randomness of $\hat{\beta}_1$?
 but just $\hat{\beta}_1 \neq 0$
 does not mean it
 linear cause it
 could be due to
 randomness.

Chi-squared distribution

$$Z_1, \dots, Z_k \sim (\mathcal{N}(0, 1)) \quad \chi^2_{(k)} = \sum_{i=1}^k Z_i^2$$

- If Z_1, \dots, Z_k are independent standard normal ($\mathcal{N}(0, 1)$) random variables, then

$$Q = \sum_{i=1}^k Z_i^2$$

is distributed according to the **chi-squared distribution with k degrees of freedom**.

- This is denoted as $Q \sim \chi_k^2$.

Student's distributions

- The **Student's t-distribution with k degrees of freedom** can be defined as the distribution of the random variable T

$$T = \frac{Z}{\sqrt{V/k}} = Z \sqrt{\frac{k}{V}},$$

where

- $Z \sim \mathcal{N}(0, 1)$,
- $V \sim \chi_k^2$,
- Z and V are independent.
- This is denoted as $T \sim t_k$.

Sampling distribution of $(\hat{\beta}_1 - \beta_1)/s(\hat{\beta}_1)$

Let us prove that

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} \sim t_{n-2}$$

with the following steps:

1. Write

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} / \sqrt{\frac{s^2(\hat{\beta}_1)}{\text{Var}(\hat{\beta}_1)}}.$$

2. Show that $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \sim N(0, 1)$.

3. Show that $\frac{(n-2)s^2(\hat{\beta}_1)}{\text{Var}(\hat{\beta}_1)} \sim \chi^2_{n-2}$.

4. Conclude using that $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}}$ and $\frac{s^2(\hat{\beta}_1)}{\text{Var}(\hat{\beta}_1)}$ are independent.

Sampling distribution of $(\hat{\beta}_1 - \beta_1)/s(\hat{\beta}_1)$

$$2. \quad \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \sim N(0, 1) \quad , \quad \hat{\beta}_1 \text{ normally distributed because } \hat{\beta}_1 \text{ is linear function of } y_i \text{'s and the } y_i \text{'s are normally distributed}$$

std deviation

We have:

$$\hat{\beta}_1 \sim N(E(\hat{\beta}_1), \text{Var}(\hat{\beta}_1)) \quad , \quad E(\hat{\beta}_1) \text{ is an unbiased estimator}$$

$$\hat{\beta}_1 \sim N(\beta_1, \text{Var}(\hat{\beta}_1)) \quad , \quad \text{so } E(\hat{\beta}_1) = \beta_1$$

we are standardizing

$$\therefore \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \sim N(0, 1)$$

$$\text{Ex. } \frac{(n-2)s^2(\hat{\beta}_1)}{\text{var}(\hat{\beta}_1)} = \frac{(n-2)\sigma^2}{\sigma^2} \quad \leftarrow \text{prove this } \Downarrow$$

exercisit

replace s^2

$$\frac{(n-2) \sum_{i=1}^n \frac{e_i^2}{\sigma^2}}{\sigma^2} = \frac{\sum e_i^2}{\sigma^2} = \sum \left(\frac{e_i}{\sigma} \right)^2$$

If C claims $\sum \left(\frac{e_i}{\sigma} \right)^2$ is a standard normal distribution
because e_i is N distributed with cov of e_i .

$\frac{e_i}{\sigma} \sim N(0, 1)$ (see 2) for full definition.

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \sim N(0, 1)$$

$$\frac{(n-2) \sum \left(\frac{e_i}{\sigma} \right)^2}{\sqrt{\text{var}(\hat{\beta}_1)}} \sim \chi_{n-2}^2$$

$n-2$

Hypothesis test concerning β_1

1. Two sided test:

$$\begin{cases} H_0 : \beta_1 = \beta_1^0 \\ H_a : \beta_1 \neq \beta_1^0 \end{cases}$$

2. Test statistic:

$$t^* = \frac{\hat{\beta}_1 - \beta_1^0}{s(\hat{\beta}_1)}$$

3. Decision rule:

- ▶ Reject H_0 if $|t^*| > t_{1-\alpha/2;n-2}$
- ▶ Do not reject H_0 if $|t^*| \leq t_{1-\alpha/2;n-2}$

where

- ▶ $\alpha = \Pr(\text{Reject } H_0 \text{ when it is true})$ is the probability of a Type I error,
- ▶ $t_{1-\alpha/2;n-2}$ is the $1 - \alpha/2$ -percentile of a t_{n-2} distribution, i.e. the value such that $\Pr(t_{n-2} \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$.

Toluca example

Exercise 1

Is the slope significantly different from zero in the Toluca example?

Interpret the result.