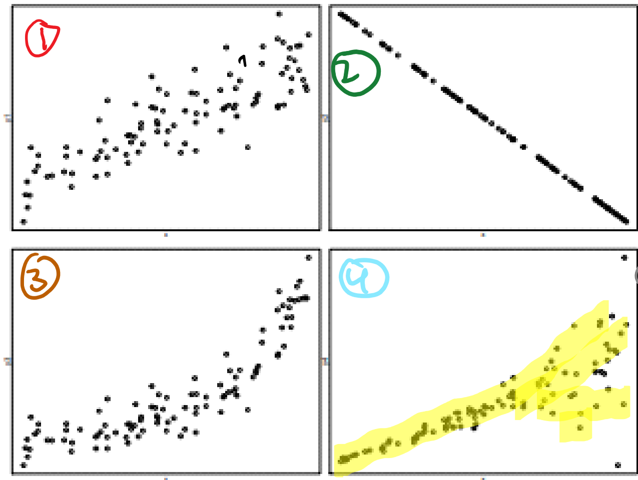
Preliminary activity II



note: trend 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

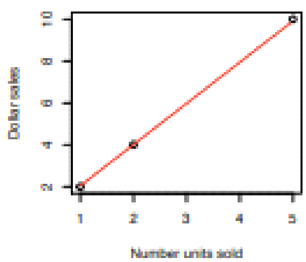
- Statistical methodology that utilizes the relation between variables.
- Predicts a response variable (or outcome) from the relation between the response and other variables.
- Regression analysis is used in many disciplines such as:

  - 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):

$Y = 2x$

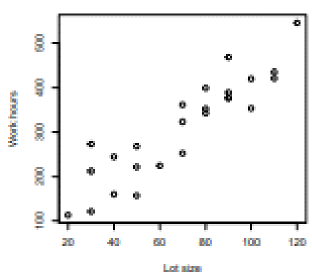


This is an example of a perfect relationship.

All observations fall on the line of functional relationship.

Statistical relation

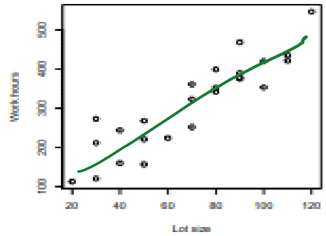
- Not a perfect relation.
- Example: A company produces replacement parts. It produces lots of varying size. The relation between the lot size and work hours is a statistical relation.



Note: There is still a trend but a greater amount of noise

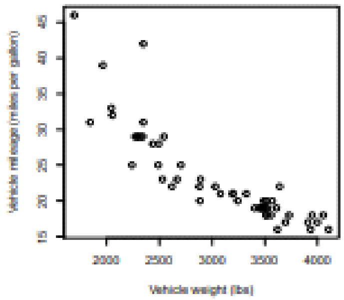
Statistical relation

- 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

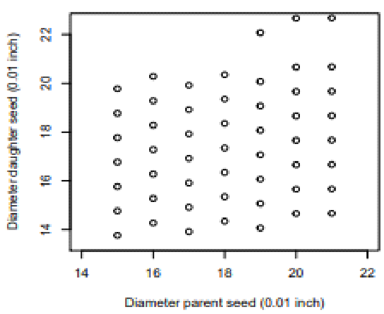
- Example: Weight and mileage for 54 cars.
- Functional or statistical relation?
- Linear statistical relation?



Galton's early considerations of regression

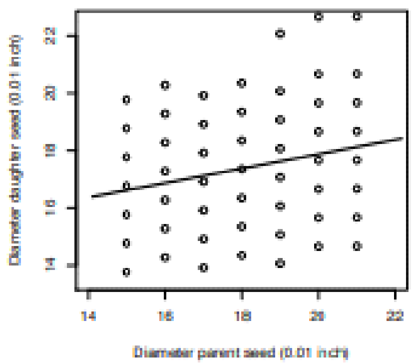
- Sir Francis Galton, English Victorian statistician, sociologist, psychologist, anthropologist, etc.
- Work on inherited characteristics of sweet peas $\Rightarrow$ initial conceptualization of linear regression.
- In 1975, Galton distributed packets of sweet pea seeds to seven friends who harvested 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: know 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 a 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

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

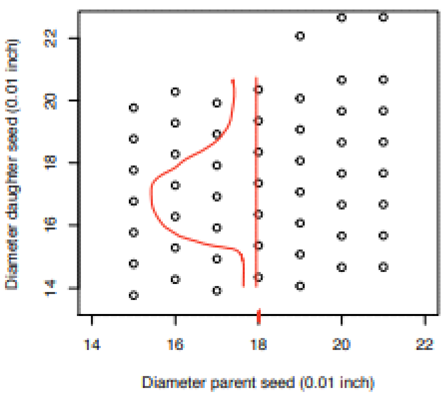
Basic concepts

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 corrolation.

- **Observational study**
  - Investigator has no control over the explanatory variables (X)
  - Limitation: not adequate for cause-and-effect
    A strong association does not necessarily means 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 maybe 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 week; 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:

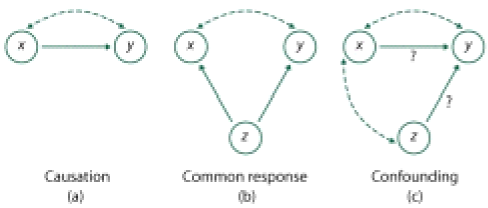


Figure 1: The dashed arrows represent association and the solid ones cause and effect link. The variable x is explanatory, y is response, and z is a lurking variable.

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

Overview of the steps in regression analysis

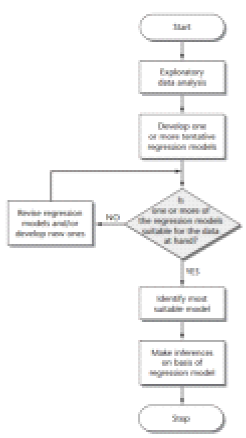


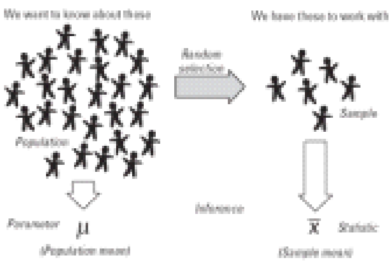
Figure 2: The steps in regression analysis [Kutner et al., 2004, p.14]

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.



Recall

$$\bar{x} = \frac{1}{N} \sum x_i$$

Toluca company example[1]

- 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	Lot size X_i	Work hours Y_i
1	80	399
2	30	121
...	...	...
24	80	342
25	70	323

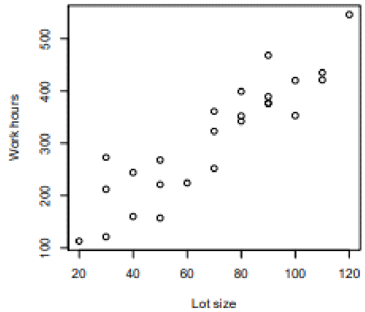
← how much time it takes to produce the items

ex/ 399 hours to create 80 fridges

[1] From [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:

$$\boxed{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,}$$

where

4 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 $\mathrm{E}(\varepsilon_i) = 0$, with common variance $\mathrm{Var}(\varepsilon_i) = \sigma^2$, 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) $Var(\varepsilon_i) = \sigma^2$
3) pairwise independent; thus $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$

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

- $i = j$:

wts $E(\varepsilon_i \cdot \varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$

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

So $Var(\varepsilon_i) = E(\varepsilon_i^2) - (\underbrace{E(\varepsilon_i)}_{0})^2$, first assumption says $E(\varepsilon_i) = 0$, second assumption says $Var(\varepsilon_i) = \sigma^2$

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

$\sigma^2 = \sigma^2$

Recall

$Var(x) = E(x^2) - (E(x))^2$

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

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

$0 = E(\varepsilon_i \varepsilon_j)$

Recall

$Cov(\varepsilon_i, \varepsilon_j) = 0$

Important features

Simple linear model

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

constant ⇒ understand constant as not random.

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

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)

   y_i is constant + random = 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 $X = X_i$.

   $E(y_i) = E(B_0 + B_1 x_i + \varepsilon_i) = E(B_0) + E(B_1 x_i) + \underbrace{E(\varepsilon_i)}$, linearity

   $= B_0 + B_1 X_i$

   Thus, the functional relationship between the true mean of Y_i and X_i is a straight line with intercept B_0 and slope B_1

January 11, 2019 8:45 AM

Important features

Simple linear model

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

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

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

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

i.e $y_i \perp$ 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

Var (constant + r.V) = 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 $\mathrm{E}(X) = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$ 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

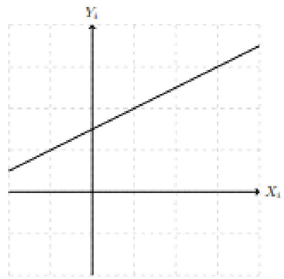
Simple linear model with normal errors

- 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

Least square estimators

Find $\widehat{\beta}_0$ and $\widehat{\beta}_1$ 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).

1) $\frac{\partial Q}{\partial B_0} = 2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i) \quad (-1)$

$0 = -2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)$

$\frac{\partial Q}{\partial B_1} = 2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)\cdot(-x_i)$

$0 = 2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)\cdot(-x_i)$

2) B_0 : Critical points

$0 = -2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)$

$0 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} B_0 - B_1 \sum_{i=1}^{n} x_i$

$\sum_{i=1}^{n} B_0 = \sum_{i=1}^{n} y_i - B_1 \sum_{i=1}^{n} x_i$

$n B_0 = \sum_{i=1}^{n} y_i - B_1 \sum_{i=1}^{n} x_i$

$B_0 = \frac{\sum_{i=1}^{n} y_i}{n} - B_1 \frac{\sum_{i=1}^{n} x_i}{n}$

$\hat{B}_0 = \bar{y} - B_1 \bar{x}$

B_1 : Critical point

$0 = 2\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)\cdot(-x_i)$

$0 = \sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)(x_i)$

$0 = \sum_{i=1}^{n}(y_i x_i - B_0 x_i - B_1 x_i^2)$

$0 = \sum_{i=1}^{n} y_i x_i - B_0 \sum_{i=1}^{n} x_i - B_1 \sum_{i=1}^{n} x_i^2$

$B_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - B_0 \sum_{i=1}^{n} x_i$

$B_1 = \frac{\sum_{i=1}^{n} y_i x_i - B_0 \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}$

$\sum_{i=1}^{n} x_i^2 B_1 = \sum_{i=1}^{n} y_i x_i - \left(\frac{\sum_{i=1}^{n} y_i}{n} - B_1 \frac{\sum_{i=1}^{n} x_i}{n}\right)\left(\sum_{i=1}^{n} x_i\right)$

$\sum_{i=1}^{n} x_i^2 B_1 = \sum_{i=1}^{n} y_i x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + \frac{1}{n} B_1 \sum_{i=1}^{n} x_i$

$\sum_{i=1}^{n} x_i^2 B_1 - \frac{1}{n} B_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$

$B_1\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} y_i x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$

$B_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i}$

Least square estimators

Least square estimators of β_1 and β_0:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2},$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X},$$

where

- $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and
- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

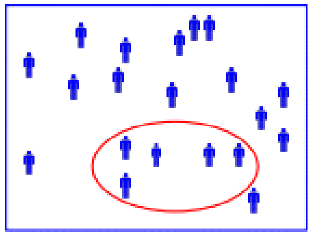
are the sample mean of Y and X, respectively.

Estimation of the parameters

- Postulated model:

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

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



- 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

$\Rightarrow$ Estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$

$\wedge$ ← hat means estimate, not true value.

Method of least squares

Simple linear model:

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

We want to minimize the sum of the errors to get the best one

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

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

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

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

Least square estimators

Exercise 2

Show that

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

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

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

Least square estimators

Least square estimators of $\widehat{\beta}_1$ and $\widehat{\beta}_0$:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2},$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}.$$

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

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

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

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

Regression equation

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $E(y_i) = \beta_0 + \beta_1 x_i$; True mean of y when $X = x$

to understand as $E(y_i \mid x_i)$

Regression equation or **fitted regression line**

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

where $\widehat{Y}$ 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:* $\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$, *and*
- *There are uncorrelated* $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$.

Then the least squares estimators $\widehat{\beta}_0$ *and* $\widehat{\beta}_1$ *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

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

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

CHECK NEXT

pagé! ☺

Parameter θ

Estimator $\hat{\theta}$

$\hat{\theta}$ is an <u>unbiased</u> 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 shit is biased

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

$$E(\hat{\beta}_1) = E\left(\frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}\right)$$, # use linearity dawg.

<u>Note</u>

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

usually!

$$= \frac{1}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}\left\{E\left(\sum x_i y_i\right) - E\left(\frac{1}{n}\sum x_i \sum y_i\right)\right\}$$

$$= \frac{1}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}\left\{\sum x_i E(y_i) - \frac{1}{n}\sum x_i \sum E(y_i)\right\}$$

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

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

$$= \frac{1}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}\left\{\underbrace{\sum x_i(\beta_0 + \beta_1 x_i)}_{\beta_0 \sum x_i + \beta_1 \sum x_i^2} - \frac{1}{n}\sum x_i \underbrace{\sum(\beta_0 + \beta_1 x_i)}_{n\beta_0 + \beta_1 \sum x_i}\right\}$$

$$= \frac{\beta_1 \sum x_i^2 - \frac{1}{n}\beta_1(\sum x_i)^2}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}$$

$$= \beta_1$$

<u>Conclusion</u>

$E(\hat{\beta}_1) = \beta_1$, i.e.

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

January 16, 2019 8:48 AM

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

$E(\hat{\beta}_0) = E(\bar{y}) - E(\hat{\beta}_1)\bar{x}$, linearity of expectation thing

Recall $\hat{\beta}_1$

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

$E(\bar{y}) = E(\frac{1}{n}\sum y_i) = \frac{1}{n}\sum E(y_i) = \frac{1}{n}\sum(\beta_0 + \beta_1 x_i)$

linearity of E

$= \frac{1}{n} n\beta_0 + \frac{1}{n}\beta_1 \sum x_i$

$= \beta_0 + \beta_1 \bar{x}$

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

Conclusion: $E(\hat{\beta}_1) = \beta_0$, i.e. $\hat{\beta}_0$ is an unbiased estimator of β_0

Toluca company example

Using R, we find:

$$\sum_{i=1}^{n} X_i = 1750 \quad \sum_{i=1}^{n} Y_i = 7807 \quad \sum_{i=1}^{n} X_i Y_i = 617180$$

$$\sum_{i=1}^{n} X_i^2 = 142300 \quad n = 25$$

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. $B_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \dfrac{617180 - \frac{1}{25}(1750)(7807)}{142300 - \frac{1}{25}(1750)^2}$

$= 3.5702020202$

$B_0 = \bar{y} - \hat{B}_1 \bar{x} = \frac{1}{25}\left[(7807) - 3.570202(1750)\right]$

$= \frac{1}{n}\sum y_i - 3.5702020202 \frac{1}{n}\sum x_i$

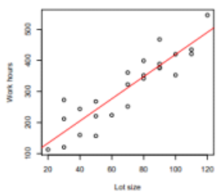
$= 62.36586$

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.

2.



$\hat{B}_1 = 3.5702$

$\hat{B}_0 = 62.37$

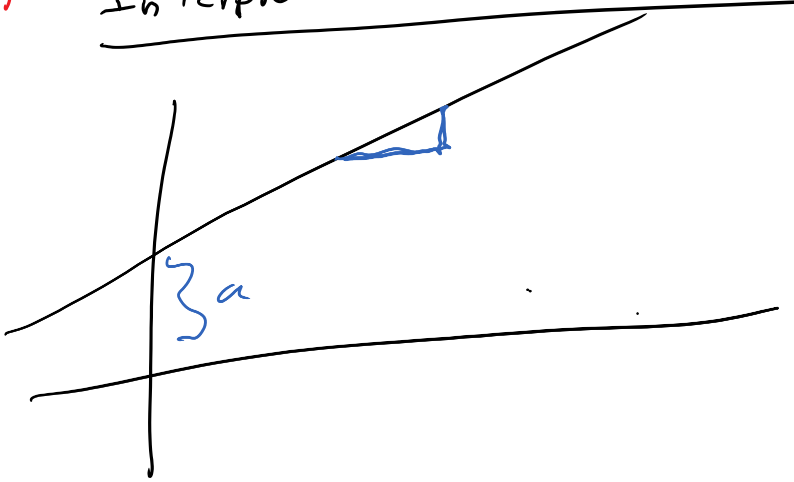
$\hat{y} = 62.37 + 3.5702\,X$

estimation mean of y.

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

January 16, 2019 9:00 AM

3. 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 k! :)

Toluca company example: R output

$Hours = B_0 + B_1 Size + \varepsilon_i$

y X

```
Call:
lm(formula = Hours ~ Size, data = 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

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
```

name of the data set

linear model

or

useless according to prof

* rolls go and p val < 0.05

$\hat{B}_0$ $s(\hat{B}_0)$ p value corresponding

t-test: t^* of $H_0: B_0 = 0$, $H_A: B_0 \neq 0$

always two sided and zero for a sample

B_1 $s(B_1)$ t^* of $H_0: B_1 = 0$, $H_A: B_1 \neq 0$

s n-2

ignore this for now

F^* of $H_0: B_1 = 0$, $H_A: B_1 \neq 0$

DF R^2

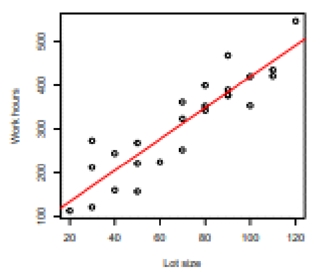
p val says how extreme a value is, you can have very strong evidence against or for the null. can tell small and large pvalues strong or weak your null you know is

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?

$\hat{y} = 62.37 + 3.5402 \cdot 60 = 276.582$

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

$\hat{y} = 62.37 + 3.5402 \cdot 60 = 276.582$

Predicted values and residuals

- Fitted regression line:

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

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

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

  Fitted value $\widehat{Y}_i$ used as:
  - **Prediction** of the value of Y for particular value X_i of X. Sometimes written $\widehat{Y}_{pred_i}$.
  - **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 $\widehat{Y}_i$

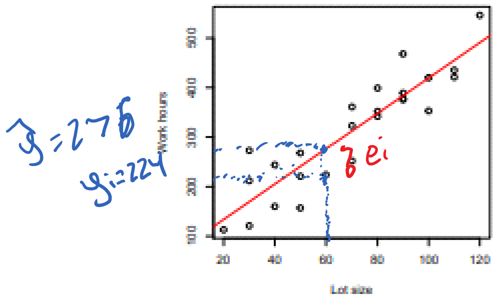
$$e_i = Y_i - \widehat{Y}_i$$

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 textbook[2]
  - 1.20
  - 1.21
- From the slides[3]
  - Slide 26
  - Slide 37
  - Slide 40
  - Slide 41
  - Slide 47

Go do this or slap yourself!
karen style slide questions dawg.

[2]Solutions in the student manual (CD) provided with the textbook
[3]Partial solutions posted on Quercus

Proofs

1) $\sum e_i = \sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1} y_i - \sum_{i=1} \hat{\beta}_0 - \sum_{i=1} \hat{\beta}_1 x_i = \sum_{i=1} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum x_i = \sum y_i - \sum y_i + \hat{\beta}_1 \sum x_i - \hat{\beta}_1 \sum x_i = 0$

$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i)$ sub $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $= \sum y_i - n\bar{y} + n\hat{\beta}_1 \bar{x} - \hat{\beta}_1 \sum x_i$

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

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

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

4.)

$\sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x_i (y_i - \hat{y}_i)$, definition of $e_i = y_i - \hat{y}$

$= \sum_{i=1}^{n} (x_i y_i - x_i \hat{y}_i) = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \hat{y}_i$, summation properties + expand

$= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum x_i^2$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ + summation properties

$= \sum_{i=1}^{n} x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum x_i^2$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \bar{x} \sum x_i - \hat{\beta}_1 \sum x_i^2$, expand

$= \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum x_i \sum y_i + \left(\frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2} \right) \frac{1}{n} (\sum x_i)^2 - \left(\frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2} \right) \sum x_i^2$, $\bar{y} = \frac{\sum y_i}{n}$, $\bar{x} = \frac{\sum x_i}{n}$, $\hat{b}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}$

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

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

$= \left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right) [\underbrace{1 - 1}_{0}]$

$= 0$

January 18, 2019 9:06 AM

Properties of the fitted regression line

Exercise 1

Prove the following properties:

1. $\sum_{i=1}^{n} e_i = 0$
2. $\sum_{i=1}^{n} e_i^2$ *is a minimum of function* $Q(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$
3. $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i$
4. $\sum_{i=1}^{n} X_i e_i = 0$
5. $\sum_{i=1}^{n} \widehat{Y}_i e_i = 0$
6. *The regression line passes through* $(\overline{X}, \overline{Y})$

Term Test I: Monday Jan 28th, 5pm IC 1

Unbiased estimator of the of the population mean of Y

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

} Type of exercise you might find on tests!

Proof of unbiased of something else

we have parameter θ

$\theta = \beta_0 + \beta_1 x_i$

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

we wts:

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

$E(\hat{\theta}) = E(\hat{y}) = E(\underbrace{\hat{\beta}_0}_{\text{random}} + \underbrace{\hat{\beta}_1}_{\text{random}} \underbrace{x_i}_{\text{assumption} \Rightarrow \text{constant}})$

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

$= \theta$

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

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

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

Recall
mean of y given a value of X then
$E(y_i) = \beta_0 + \beta_1 x_i$
where $y_i = \beta_0 + \beta_1 x_i + \varepsilon$
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

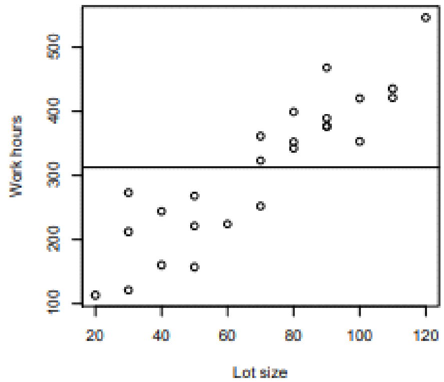
$$SSTot = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$$

3 Types of Sum of Squares

- Total sum of squares: total deviation in the response variable.
- distance between y and observed

January 18, 2019 9:07 AM

Total sum of squares



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

Partitioning

Sum of squares

total

SSReg SSRes

Partitioning of the total sum of squares

$$SSTot = SSReg + SSRes \qquad (1)$$

where

- $SSReg = \sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2$ is the regression sum of squares
- $SSRes = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2$ is the residual sum of squares

Solve this exercise!

January 18, 2019 9:08 AM

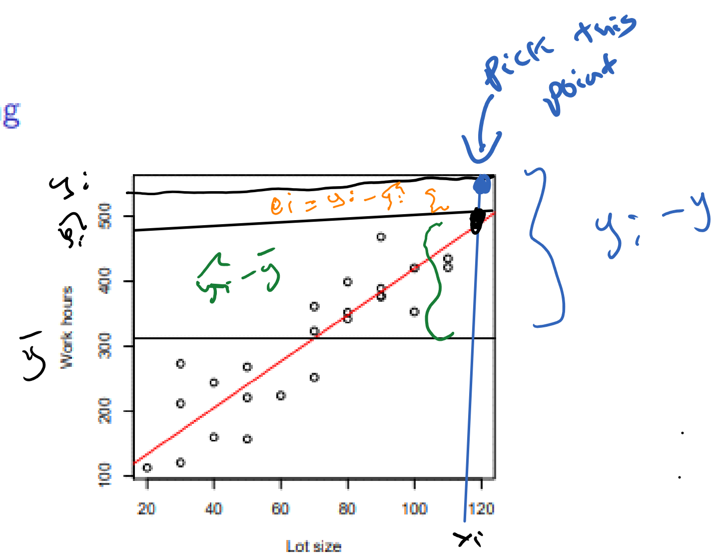
Partitioning

← Do this!

Exercise 3

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

Partitioning



$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$$

Toluca example

We had:

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

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

Exercise 4

Compute SSTot, SSReg, and SSRes.

Hint: we have (to prove)

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

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 .... n times unknown, think we already computed $\bar{y}$. So last one is fixed.

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

$\sum(y_i-\hat{y}_i)^2=\sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)$, n random y_i, although $\hat{\beta}_0$ and $\hat{\beta}_1$ are random its based on y_i

- $SSReg$ has 1 degrees of freedom

$\sum(\hat{y}_i-\bar{y})^2=\sum(\hat{\beta}_0+\hat{\beta}_1x_i-\bar{y})$, $\hat{\beta}_0$, $\hat{\beta}_1$ 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:
> $$R^2 = \frac{SSReg}{SSTot} = 1 - \frac{SSRes}{SSTot}$$

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 TT1

:)

Simple linear model with normal errors

Model supposed in this section:

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

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

sometimes add an additional assumption

Admin: test monday - starting from here not on the test

Exercise 6

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

y_i is normally distributed: ε_i normally distributed and $y_i = cst + \varepsilon_i$

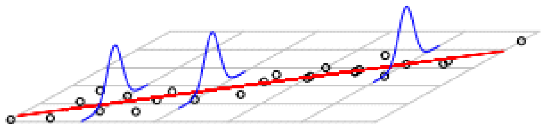
constant + normal distribution = normal distribution that is shifted

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

$Var(y_i) = \sigma^2$

previously shown :)

Simple linear model with normal errors



distribution of random error ε_i for a specific x. Don't worry about this

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 $Y_1, Y_2, \ldots, Y_n$. Show that the likelihood function for the simple linear model with normal errors is

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

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

$$\prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n}$$

Recall

$$\prod_{i=1}^{n} \exp(u_i) = \exp\left\{\sum_{i=1}^{n} u_i\right\}$$

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

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

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

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 $\widehat{\beta}_0$ and $\widehat{\beta}_1$

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.
- $\Rightarrow \overline{Y}, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{Y}_i, e_i$ 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

$$\mathrm{Var}(U) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(Y_i) + \underbrace{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i a_j \mathrm{Cov}(Y_i, Y_j)}_{0}.$$

0 since pairwise independently cause $Cov(y_i, y_j) = 0$ when $i \neq j$)

Pairwise $\perp$

$y_i \perp y_j$ when $i \neq j$

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

$$\boxed{\mathrm{Var}(U) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(y_i)}$$

$$= \sigma^2 \sum_{i=1}^{n} a_i^2$$

shove this up your head.

Variance of a linear function

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

Exercise 9

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

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a_i y_i = \sum_{i=1}^{n} \frac{1}{n} y_i$ ← I slapped myself trying to figure this out

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

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

Variance of the sample mean $\overline{Y}$

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

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

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

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

$$\boxed{\mathrm{Var}(\overline{Y}) = \sigma^2 \sum_{i=1}^{n} a_i^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 =} \; \sigma^2 \cdot \frac{n}{n^2} = \frac{\sigma^2}{n}$$

} variance of sample mean.

- Well-known result!

January 23, 2019 7:52 AM

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

prove this

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

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right) Y_i}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$$

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

trick used to change index

see this dude as a constant

We have

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

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

, now we can identify a_i

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

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

my solution :)

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

slide 29

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

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

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

$$\boxed{\mathrm{Var}(\widehat{\beta}_1) = \sigma^2\left(\frac{1}{\sum_{j=1}^{n}(x_j - \bar{x})}\right)^2}$$

Profs solution

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

slide 29

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

could write i too

January 23, 2019 7:52 AM

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

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

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

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

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

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

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

$$\mathrm{Var}(\widehat{\beta}_0) =$$

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

Exercise 10

Show that

$$\mathrm{Var}(\widehat{Y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right\}.$$

not instead of i

so it's not confusing

Ex for you to do! solve this

Hint:

1. Use
$$\widehat{Y}_0 = \overline{Y} + \widehat{\beta}_1(X_0 - \overline{X})$$
to write $\widehat{Y}_0$ 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.

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

January 23, 2019 7:56 AM

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

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

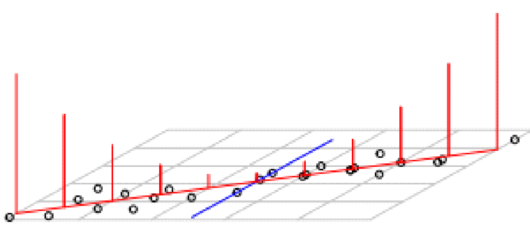
$$\boxed{\mathrm{Var}(\widehat{Y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right\}}$$

Comment 2

- The minimum value of $\mathrm{Var}(\widehat{Y}_0)$ is $\sigma^2 \cdot \frac{1}{n}$ which is attained when $x_0 = \bar{x}$.
- $\mathrm{Var}(\widehat{Y}_0)$ increases as x_0 moves away from $\bar{x}$.
- This formula gives the variance when $\widehat{Y}_0$ is considered as an estimate of the population mean of Y for a value X_0 of X.

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

don't need to know this but try to :)



January 23, 2019 7:57 AM

Variance of a prediction $\widehat{Y}_{pred_0}$

- A prediction of a future observation of Y for a given value X_0 of X is
$$\widehat{Y}_{pred_0} = \widehat{\beta}_0 + \widehat{\beta}_1 X_0.$$
- An estimate of the population mean of Y for a given value X_0 of X is
$$\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 X_0.$$
- However, it is possible to show that
$$\mathrm{Var}(\widehat{Y}_{pred_0}) = \mathrm{Var}(\widehat{Y}_0) + \sigma^2.$$

Variance of a prediction $\widehat{Y}_{pred_0}$

$Var(\hat{y}_0) = \sigma^2 \left\{ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma (x_i - \bar{x})^2} \right\}$

- **Idea**:
$$\begin{array}{lll} \widehat{Y}_0 & \text{is an estimator of} & \beta_0 + \beta_1 X_0 \\ \widehat{Y}_{pred_0} & \text{is an estimator of} & \beta_0 + \beta_1 X_0 + \varepsilon_0, \end{array}$$

take variance for this to make sense

where $\mathrm{Var}(\varepsilon_0) = \sigma^2$.
- Finally,
$$\boxed{\mathrm{Var}(\widehat{Y}_{pred_0}) = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right\}}$$

just add σ^2

Estimates and precision: summary

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

usefull stuff write on cheat sheet dawg! :)

we don't know $\beta_1, \beta_0, \sigma^2$

population parameter that is unknown.

so, we estimate:

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

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

- It is an unbiased estimator, i.e. $\mathrm{E}(s^2) = \sigma^2$.
- 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 $\widehat{\beta}_1$ is

estimated variance of β_1

$$s^2(\widehat{\beta}_1) = s^2 \frac{1}{\sum\left(X_i - \overline{X}\right)^2}.$$

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

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

Toluca example: estimated variances

Ex for you to do!

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.*

Exercises

- From the textbook[1]
  - 2.25, part (a)
  - 2.24, part (a). Only set up the ANOVA table in the format of Table 2.2
- From the slides[2]
  - Slide 1
  - Slide 3
  - Slide 7
  - Slide 9
  - Slide 18
  - Slide 19
  - Slide 24
  - Slide 25
  - Slide 30
  - Slide 37
  - Slide 45

[1] Partial solutions in the student manual (CD) provided with the textbook
[2] Partial solutions posted on Quercus

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 $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.
- Reminder: we showed that under this model $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 X_i, \sigma^2)$.

Most common hypothesis test in simple linear regression

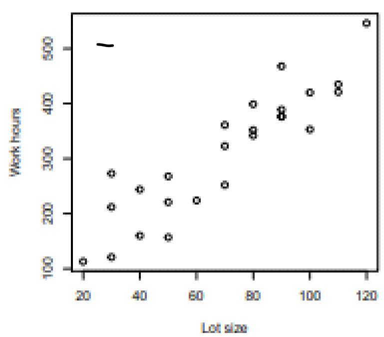
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

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

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

- 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

- In toluca example, we obtained $\widehat{\beta}_1 = 3.57$. $\longleftarrow$ previously $\widehat{\beta}_1 \neq 0$
- Is this departure from zero only due to randomness of $\widehat{\beta}_1$?

but just $\widehat{\beta}_1 \neq 0$ does not mean its linear cause it could be due to randomness.

Chi-squared distribution

$z_1, \ldots, z_k \sim (N(0,1)$

$\chi^2_{(k)} = \sum_{i=1}^{k} z_i^2$

- If $Z_1, \ldots, Z_k$ 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 $Q \sim \chi^2_k$.

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

January 25, 2019 9:33 AM

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

Let us prove that

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

with the following steps:

1. Write
$$\frac{\widehat{\beta}_1 - \beta_1}{s(\widehat{\beta}_1)} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}(\widehat{\beta}_1)}} \Big/ \sqrt{\frac{s^2(\widehat{\beta}_1)}{\mathrm{Var}(\widehat{\beta}_1)}}.$$
2. Show that $\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}(\widehat{\beta}_1)}} \sim \mathcal{N}(0,1)$.
3. Show that $\frac{(n-2)s^2(\widehat{\beta}_1)}{\mathrm{Var}(\widehat{\beta}_1)} \sim \chi^2_{n-2}$.
4. Conclude using that $\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}(\widehat{\beta}_1)}}$ and $\frac{s^2(\widehat{\beta}_1)}{\mathrm{Var}(\widehat{\beta}_1)}$ are independent.

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

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

std deviation

We have:

$$\hat{\beta}_1 \sim N(E(\hat{\beta}_1), \mathrm{var}(\hat{\beta}_1))$$

$$\hat{\beta}_1 \sim N(\beta_1, \mathrm{var}(\hat{\beta}_1))$$

, $\hat{\beta}_1$ normally distributed because $\hat{\beta}_1$ is linear fact of y_i's and the y_i's are normally distributed

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

we are standardizing

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

January 25, 2019 9:49 AM

$$3.\ \frac{(n-2)s^2(\hat{\beta}_1)}{Var(\hat{\beta}_1)} \underset{\text{excercise}}{=} \frac{(n-2)s^2}{\sigma^2} \leftarrow \text{prove this}$$

replace s^2

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

claim $\sum\left(\frac{e_i}{\sigma}\right)^2$ is a standard normal distribution with cov of e_i.
because e_i is N distributed (sec 2) for full definition.

$\frac{e_i}{\sigma} \sim N(0,1)$ (sec 2)

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} = \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}}}_{\sim N(0,1)\ Z} \Bigg/ \sqrt{\frac{\overbrace{\frac{(n-2)s^2(\hat{\beta}_1)}{\sqrt{Var(\hat{\beta}_1)}}}^{\chi^2_{n-2}}}{\underbrace{n-2}_{K}}} \sim t_{n-2}$$

January 25, 2019 9:34 AM

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**:
$$t^* = \frac{\widehat{\beta}_1 - \beta_1^0}{s(\widehat{\beta}_1)}$$
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
   - $\alpha = \underbrace{\mathrm{Pr}}_{\text{probability}}(\text{Reject } H_0 \text{ when it is true})$ 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 $\mathrm{Pr}(t_{n-2} \leq t_{1-\alpha/2;n-2}) = 1 - \alpha/2$.

Toluca example

Exercise 1

Is the slope significantly different from zero in the Toluca example? Interpret the result.

2019-01-30 8:05 AM

we calculated earlier: $\hat{\beta}_1 = 3.5$, $\underbrace{s(\hat{\beta}_1)}_{\text{standard error}} = 0.347$, $n = 25$, $\alpha = 5\%$

$\hat{\beta}_1^0 = 0$

1. $\begin{cases} H_0: \beta_1 = 0 \\ H_A: \beta_1 \neq 0 \end{cases}$

2. $t^* = \dfrac{\hat{\beta}_1 - \beta_1^0}{s(\hat{\beta}_1)} = \dfrac{3.5}{0.347} = 10.08453$

3. $t_{1-\frac{\alpha}{2}, \underbrace{n-2}_{DF}} = t_{0.975, 23} = \underbrace{2.069}_{\text{value of quantile}}$

$\alpha = 0.05$ (usually)

$n = 25$

4. So since test statistic $(10.08453 > 2.069)$ we reject the null hypothesis

$|t^*| = 10.086 > 2.069$ (means value from sample is extreme)

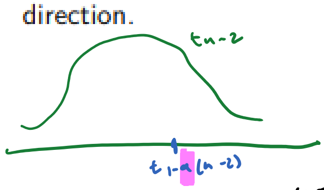
5. Conclusion: slope is significantly different from zero, so we reject null.

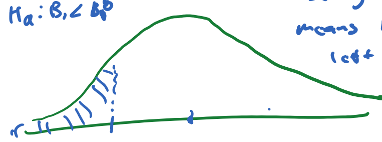
One-sided tests versus two-sided tests

- Choice not always clear.
- One-sided test: only concerned with deviations from H_0 in one direction.

$H_a: \beta_1 > \beta_1^0$ $H_a: \beta_1 < \beta_1^0$

being extreme means being left of the value



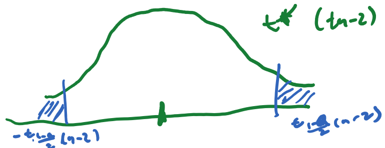


↖ just $1-\alpha$ density cause only $1-\alpha$.

One sided is more prone to reject the null hypothesis. We don't usually use one sided. If in doubt use two sided

- Two-sided test: concerned with deviations from H_0 in either direction.

$H_A: \beta_1 \neq \beta_1^0$



- If you are ever in doubt use a two sided-test.

Sampling distribution of $(\widehat{\beta}_0 - \beta_0)/s(\widehat{\beta}_0)$ ← practice exercise

Exercise 2
Show that

$$\boxed{\frac{\widehat{\beta}_0 - \beta_0}{s(\widehat{\beta}_0)} \sim t_{n-2}.}$$

Hypothesis test concerning β_0

1. **Two sided test**:
$$\begin{cases} H_0 : \beta_0 = \beta_0^0 \\ H_a : \beta_0 \neq \beta_0^0 \end{cases}$$
2. **Test statistic**:
$$t^* = \frac{\widehat{\beta}_0 - \beta_0^0}{s(\widehat{\beta}_0)}$$
3. **Decision rule**:
   - Reject H_0 if $|t^*| > t_{1-\alpha/2;n-2}$
   - Do not reject H_0 if $|t^*| \leq t_{1-\alpha/2;n-2}$

Toluca example

Exercise 3

Is the intercept significantly different from 50 in the Toluca example?

Hypothesis test concerning $\widehat{Y}_i$

Exercise 4

Using the same procedure as for β_0 and β_1, construct a hypothesis test for the population mean of Y for a value X_i of X.

Follow these steps:

1. Sampling distribution: show that

$$\boxed{\frac{\widehat{Y}_i - \mathrm{E}(Y_i)}{s(\widehat{Y}_i)} \sim t_{n-2}.}$$

2. Set up the hypothesis, test statistic, and decision rule.

Comment 1

We could do the same for a hypothesis test concerning a prediction $\widehat{Y}_{pred_i}$.

Hypothesis test concerning $\widehat{Y}_i$

Hypothesis tests when σ^2 is known

Would we still use a t-test if σ^2 were known? If no, which test would we use and why? Determine what test statistic you would use, show what is the sampling distribution of the test statistic, and set up the decision rule.

what do we use as a distribution if σ^2 were known? Normal

Use normal distribution

Build test using for β_1: $\frac{\hat{\beta}_1 - \beta_1}{Var(\hat{\beta}_1)} \sim N(0,1)$

known if σ^2 known

F test for β_1

- Construct a hypothesis test for β_1 with ANOVA called F test.
- For single linear regression, the F test is equivalent to the t test for $H_0 : \beta_1 = 0$ considered earlier.
- But ANOVA provides a battery of tests for multiple linear regression.

← only for simple linear regression not multiple!

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	$SSTot$	$n-1$	

January 30, 2019 8:09 AM

F distribution and sampling distribution of $MSReg/MSRes$

$\Rightarrow$ fisher distribution

- The F-distribution with k_1 and k_2 degrees of freedom can be defined as the distribution of the random variable F

$$F = \frac{V_1/k_1}{V_2/k_2},$$

ration of $\perp$ chi-square distribution

where
  - $V_1 \sim \chi^2_{k_1}$,
  - $V_2 \sim \chi^2_{k_2}$,
  - V_1 and V_2 are independent.
- This is denoted as $F \sim F_{k_1,k_2}$.
- We can show that under $H_0 : \beta_1 = 0$,

$$\boxed{\frac{MSReg}{MSRes} \sim F(1, n-2).}$$

$\Leftarrow$ we can prove this but later than

F test for β_1

- **Two sided-test**:

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

- **Test statistic**:

$$F^* = \frac{MSReg}{MSRes}$$

- **Decision rule**:
  - Reject H_0 if $F^* > F_{1-\alpha;1,n-2}$
  - Do not reject H_0 if $F^* \leq F_{1-\alpha;1,n-2}$

  where $F_{1-\alpha;1,n-2}$ is the $1-\alpha$-percentile of a $F(1, n-2)$ distribution
- F test only for two-sided test

same null as ex 1, but different test thats

Equivalent

$F^* = \frac{MSReg}{MSRes}$

$F^* = 105.9$ (use r code output)

$F_{1-a;1,n-2} = F_{0.95;1,23} = 4.26$ / use fisher distribution

$a = 0.05$
$n = 25$

(a stands for 1-a)

$q = 0.95$

$\frac{v_2}{v_1}$ } degrees of freedom

$v_1 = 1$ ⇐ column

$v_2 = 23$ ⇐ row

$F^* > 4.26$, we reject null. Same conclusion as earlier.

Confidence intervals for β_0, β_1, and $\mathrm{E}(Y_i)$

- We know: $(\widehat{\beta}_0 - \beta_0)/s(\widehat{\beta}_0) \sim t_{n-2}$. Using the general formula, we obtain for a confidence interval (CI)

$$\boxed{(1-\alpha)\% \text{ CI for } \beta_0: \qquad \widehat{\beta}_0 \pm t_{1-\alpha/2,n-2}\, s(\widehat{\beta}_0)}$$

- We proceed similarly for β_1 and $\mathrm{E}(Y_i)$ and obtain:

$$\boxed{\begin{array}{ll} (1-\alpha)\% \text{ CI for } \beta_1: & \widehat{\beta}_1 \pm t_{1-\alpha/2,n-2}\, s(\widehat{\beta}_1) \\ (1-\alpha)\% \text{ CI for } \mathrm{E}(Y_i): & \widehat{Y}_i \pm t_{1-\alpha/2,n-2}\, s(\widehat{Y}_i) \end{array}}$$

- A CI for a prediction is called a prediction interval and is obtained using a similar construction:

$$\boxed{\begin{array}{l} (1-\alpha)\% \text{ prediction interval for a prediction of } Y \text{ when} \\ X = X_i: \qquad \widehat{Y}_{pred_i} \pm t_{1-\alpha/2,n-2}\, s(\widehat{Y}_{pred_i}) \end{array}}$$

Equivalence of F and t test

Exercise 5

Show that, for a given α, the F test of $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ is equivalent algebraically to the two-tailed t test.

Hint:

- Show that $F^* = (t^*)^2$.
- Use the fact that if $T \sim t_k$ then $T^2 \sim F(1, k)$ to conclude.

Toluca example: R output

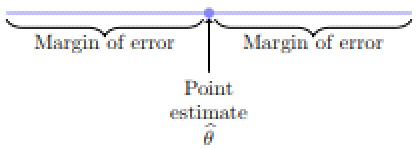
```
Call:
lm(formula = Hours ~ Size, data = toluca)

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

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
```

Confidence intervals

- Confidence interval more informative than point estimates because they reflect the precision of estimates.
- General formula for a $(1-\alpha)\%$ confidence interval for a parameter θ

$$\boxed{\left[\widehat{\theta} - q_{1-\alpha/2}s(\widehat{\theta}); \widehat{\theta} + q_{1-\alpha/2}s(\widehat{\theta})\right] = \widehat{\theta} \pm q_{1-\alpha/2}s(\widehat{\theta}),}$$

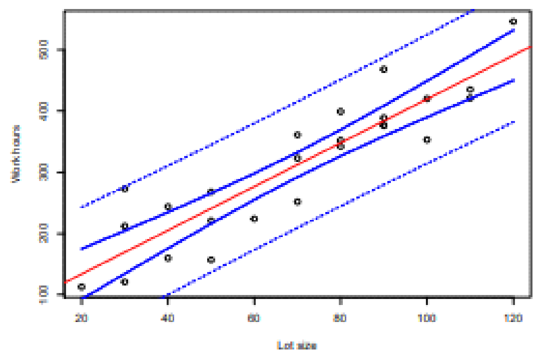
where $\widehat{\theta}$ is an estimator for θ with standard error $s(\widehat{\theta})$ and $(\widehat{\theta} - \theta)/s(\widehat{\theta})$ follows a centered and symmetric distribution with quantile of order α denoted q_α.



Confidence intervals

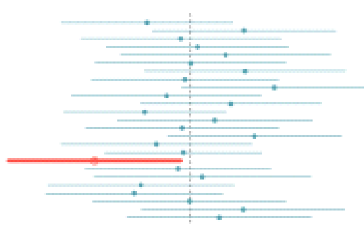
Exercise 6
Where does the formula for a $(1-\alpha)\%$ confidence interval come from?

Confidence bands



Theoretical interpretation of a confidence interval

- 25 samples selected from a population with parameter θ (dashed)
- 95% CI for θ was created for each sample
- Only 1 of these 25 intervals did not include the true value θ



General interpretation of a confidence interval: If we collect a random sample from a population a large number of times, and each time we compute a $(1-\alpha)\%$ confidence interval for a parameter, then $(1-\alpha)\%$ of the confidence intervals will include the true value of the parameter.

January 30, 2019 9:33 AM

Toluca example: confidence intervals

Exercise 7

Give a 95% confidence interval for

- *The slope* β_1,
- *The intercept* β_0,
- *The true mean work hours for a lot of size 85.*

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:* $\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$, *and*
- *There are uncorrelated* $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$.

Then the least squares estimators $\widehat{\beta}_0$ *and* $\widehat{\beta}_1$ *are unbiased and have* ← we already proved first part
minimum variance among all unbiased linear estimators.

Proof of the Gauss-Markov theorem

Step 1 We have already proven (Week 01) that the least squares estimators are unbiased, i.e.

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

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

① Proof of the Gauss-Markov theorem

Step 2 for $\hat{B}_1$

Known: $\hat{B}_1 = \sum_{i=1}^{n} a_i y_i$, $a_i = \frac{x_i - \bar{x}}{\sum_i (x_i - \bar{x})^2}$

Consider another linear unbiased estimator

tilde → $\tilde{B}_1$ of B_1.

if its a linear estimator we can

$\tilde{B}_1 = \sum_i b_i y_i$

we want to show $var(\tilde{B}_1) > var(\hat{B}_1)$

We have $E(\tilde{B}_1) = \sum_i b_i \underbrace{E(y_i)}_{b_0 + b_1 x_i}$

$= \sum_i b_i (b_0 + b_1 x_i)$

$= b_0 \sum_i b_i + b_1 \sum_i b_i x_i$

since we know $\tilde{B}_1$ is unbiased, we know $E(\tilde{b}_1) = b_1$

$= B_1$

So, $\sum_i b_i = 0$, $\sum b_i x_i = 1$

Consider: $d_i = b_i - a_i$, difference between coefficient, we can show $\sum_i a_i d_i = 0$, Excercise do it!

② $Var(\tilde{B}_1) = \sigma^2 \sum b_i^2$

$= \sigma^2 \sum (a_i + d_i)^2$

$= \sigma^2 \sum (a_i^2 + 2 a_i d_i + d_i^2)$

$= \sigma^2 \left[\sum a_i^2 + 2 \underbrace{\sum a_i d_i}_{0} + \sum d_i^2 \right]$

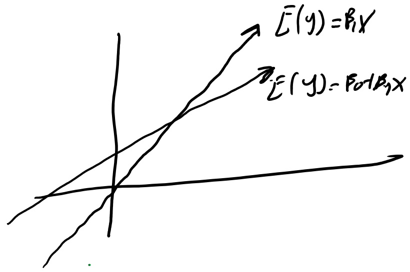
Proof left as an exercise

$= \underbrace{\sigma^2 \sum a_i^2}_{Var(\hat{B}_1)} + \sigma^2 \sum d_i^2$

$= var(\hat{B}_1) + \underbrace{\sigma^2 \sum d_i^2}_{\text{its always positive}} \geq Var(\hat{B}_1)$

Simple linear regression model through the origin

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$



- In some situations, we expect the regression line to pass through the origin. $E(y)$
- That is, the true mean of Y is expected to be __0__ when $X = 0$.
- Linear model with null intercept

$$\boxed{Y_i = \beta_1 X_i + \varepsilon_i,}$$

where

- β_1 is a parameter,
- X_i are known constants,
- ε_i are pairwise independent with mean 0 and common variance σ^2.

- Example: Number of work units performed (X) and Total labor cost (Y).

$E(y)$ should be seen as conditional expectation

Least squares estimator

- For the linear model with null intercept

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

we have

$$\mathrm{E}(Y_i) = \beta_1 X_i.$$

Thus, the functional relationship between the true mean of Y_i and X_i a straight line with intercept __0__ and slope __β_1__

January 30, 2019 9:35 AM

Least square estimators

Exercise 8

Find the least squares estimator $\widehat{\beta}_1$ of β_1.
That is, find $\widehat{\beta}_1$ that minimizes criterion

$$Q(\beta_1) = \sum_{i=1}^{n} (\qquad)^2$$

given the data.

1. *Write the normal equations (derivative of Q set to 0).*
2. *Find the critical point (solution of the normal equations).*
3. *Determine whether the critical point is a maximum or a minimum.*

Least square estimators

$$y_i = \beta_1 x_i + \varepsilon_i$$

1. $$Q(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$$

2. $$\frac{\partial Q}{\partial \beta_1} = 2 \sum_{i=1}^{n} (y_i - \beta_1 x_i) \cdot (-x_i)$$

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

$$0 = \beta_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i$$

$$\beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

name of the critical point

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

January 30, 2019 9:35 AM

Least squares estimators and regression equation

Least squares estimator of β_1:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Regression equation or fitted regression line

$$\widehat{Y} = \widehat{\beta}_1 X$$

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

Important features of the linear model with null intercept

- Unlike the linear model with an intercept, in the linear model with null intercept, the sum of residuals

$$\sum_{i=1}^{n} e_i, \qquad e_i = Y_i - \widehat{Y}_i$$

  is not necessarily zero
- An unbiased estimator of σ^2 is

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

  Reason for denominator $n-1$: only one degree of freedom is lost in estimating the single parameter in the model.

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

Exercise 9

Show that

$$\boxed{\mathrm{Var}(\widehat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}}$$

Recall

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

- $\mathrm{var}(\sum a_i y_i) = \sigma^2 \sum a_i^2$

$$\hat{\beta}_1 = \sum \underbrace{\frac{x_i}{\sum_j x_j^2}}_{a_i} y_i$$

$$\mathrm{Var}(\hat{\beta}_1) = \sigma^2 \sum \frac{x_i^2}{(\sum_j x_j)^2} = \sigma^2 \frac{\sum x_i^2}{(\sum x_j^2)^2}$$

$$= \sigma^2 \left(\frac{1}{\sum x_j^2} \right) = \sigma^2 \left(\frac{1}{\sum x_j^2} \right)$$

second way:

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\frac{\sum x_i y_i}{\underbrace{\sum x_i^2}_{\text{constant}}} \right) = \frac{1}{(\sum x_i^2)^2} \mathrm{Var}(\sum x_i y_i)$$

$$= \frac{1}{(\sum x_i^2)^2} \sigma^2 \sum (x_i)^2$$

$$= \sigma^2 \left(\frac{1}{\sum x_i^2} \right)$$

Variance of $\widehat{Y}_0$ as estimate of the population mean of Y when $X = X_0$

Exercise 10

Show that

$$\boxed{\mathrm{Var}(\widehat{Y}_0) = \sigma^2 \frac{X_0^2}{\sum_{i=1}^{n} X_i^2}}$$

$\hat{y}_0 = \hat{\beta}_1 x_0$

$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

$\mathrm{Var}(\hat{\beta}_1 x_0) = x_0^2 \, \mathrm{Var}(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{\sum x_i^2} \right) x_0^2$

$= \sigma^2 \frac{x_0^2}{\sum x_i^2}$

Variance of a prediction of a new observation of Y when $X = X_0$

Exercise 11

Show that

$$\boxed{\mathrm{Var}(\widehat{Y}_{pred_0}) = \sigma^2 \left[1 + \frac{X_0^2}{\sum_{i=1}^{n} X_i^2}\right]}$$

Estimates and precision: summary

Quantity	Estimator	Variance of the estimator	Estimated variance
β_1	$\widehat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$	$\frac{\sigma^2}{\sum_{i=1}^n X_i^2}$	$\frac{s^2}{\sum_{i=1}^n X_i^2}$
$\mathrm{E}(Y_0)$	$\widehat{Y}_0 = \widehat{\beta}_1 X_0$	$\sigma^2 \frac{X_0^2}{\sum_{i=1}^n X_i^2}$	$s^2 \frac{X_0^2}{\sum_{i=1}^n X_i^2}$
Y_0	$\widehat{Y}_{pred_0} = \widehat{Y}_0 = \widehat{\beta}_1 X_0$	$\sigma^2 \left[1 + \frac{X_0^2}{\sum_{i=1}^n X_i^2}\right]$	$s^2 \left[1 + \frac{X_0^2}{\sum_{i=1}^n X_i^2}\right]$
σ^2	$s^2 = \frac{\sum_{i=1}^n e_i^2}{n-1}$		

Hypothesis tests and confidence intervals

- Similar construction as in the linear regression model with intercept.
- Hypothesis tests based on the sampling distributions:

$$\boxed{\begin{gathered} \frac{\widehat{\beta}_1 - \beta_1}{s(\widehat{\beta}_1)} \sim t_{n-1} \\ \frac{\widehat{Y}_0 - \mathrm{E}(Y_0)}{s(\widehat{Y}_0)} \sim t_{n-1} \end{gathered}}$$

- Confidence intervals:

$$\boxed{\begin{array}{lr} (1-\alpha)\% \text{ CI for } \beta_1: & \widehat{\beta}_1 \pm t_{1-\alpha/2,n-1} s(\widehat{\beta}_1) \\ (1-\alpha)\% \text{ CI for } \mathrm{E}(Y_0): & \widehat{Y}_0 \pm t_{1-\alpha/2,n-1} s(\widehat{Y}_0) \\ (1-\alpha)\% \text{ CI for } Y_0: & \widehat{Y}_{pred_0} \pm t_{1-\alpha/2,n-1} s(\widehat{Y}_{pred_0}) \end{array}}$$

Warehouse example[1]

A plumbing supplies company operates 12 warehouses. A consultant is asked to study the relation between number of work units performed and total labor cost in the warehouses. The data is partially shown below.

Warehouse i	Work units X_i	Labor cost Y_i	$X_i Y_i$	X_i^2
1	20	114	2280	400
2	196	921	180516	38416
3	115	560	64400	13225
...	...	...	...	...
11	182	828	150696	33124
12	160	762	121920	25600
Total	1359	6390	894714	190963

Warehouse example

Exercise 12

1. *Why would we use a linear regression model through the origin for this data set?*
2. *What is the estimate of the slope?*
3. *What is the fitted regression line?*

1. Because it is appropriate to suppose the mean labor cost is 0 when work units is 0

February 1, 2019 9:53 AM

Warehouse example: R output

y ~ x ← with intercept

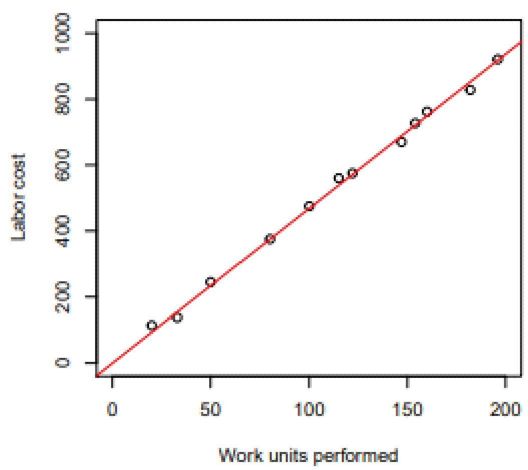
y ~ 0 + x ← without intercept

```
Call:          y          x
lm(formula = cost ~ 0 + units, data = plumbingdata)

Residuals:
    Min      1Q  Median      3Q     Max
-24.720  -4.020   4.432  11.141  21.194

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
units  4.68527    0.03421     137   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.95 on 11 degrees of freedom
Multiple R-squared:  0.9994,Adjusted R-squared:  0.9994
F-statistic: 1.876e+04 on 1 and 11 DF,  p-value: < 2.2e-16
```

note no intercept →

we can't trust this number when we do regression through the original.

Warehouse example



Random vector

- Random vector: vector of random variables

$$\mathbf{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

where z_i's are random variables.

- Example: ε and $\mathbf{Y}$ are random vectors.

Mean of a random vector

- Mean or expectation of $\mathbf{Z}$:

$$\mathrm{E}(\mathbf{Z}) = \begin{pmatrix} \mathrm{E}(z_1) \\ \mathrm{E}(z_2) \\ \vdots \\ \mathrm{E}(z_n) \end{pmatrix}$$

- Example: The mean of the random vector of errors of the regression model is

$$\mathrm{E}(\varepsilon) = \begin{pmatrix} \mathrm{E}(\varepsilon_1) \\ \mathrm{E}(\varepsilon_2) \\ \vdots \\ \mathrm{E}(\varepsilon_n) \end{pmatrix} = \mathbf{0}$$

Variance-covariance matrix of a random vector

- Variance-covariance matrix of **Z**: matrix whose coefficient in the i-th line and j-th column is $\text{Cov}(z_i, z_j) = \text{E}\{[z_i - \text{E}(z_i)][z_j - \text{E}(z_j)]\}$, i.e.

$$\text{Var}(\mathbf{Z}) = \begin{pmatrix} \text{Var}(z_1) & \text{Cov}(z_1, z_2) & \dots & \text{Cov}(z_1, z_n) \\ \text{Cov}(z_1, z_2) & \text{Var}(z_2) & \dots & \text{Cov}(z_2, z_n) \\ \vdots & & \ddots & \vdots \\ \text{Cov}(z_1, z_n) & \dots & \dots & \text{Var}(z_n) \end{pmatrix}$$

Comment 1

We used $\text{Cov}(z_i, z_i) =$

Exercise 1

Show that $\text{Var}(\mathbf{Z}) = \text{E}\{[\mathbf{Z} - \text{E}(\mathbf{Z})][\mathbf{Z} - \text{E}(\mathbf{Z})]'\}$.

Variance-covariance matrix of a random vector

- Variance-covariance matrix of **Z**:

$$\text{Var}(\mathbf{Z}) = \begin{pmatrix} \text{Var}(z_1) & \text{Cov}(z_1, z_2) & \dots & \text{Cov}(z_1, z_n) \\ \text{Cov}(z_1, z_2) & \text{Var}(z_2) & \dots & \text{Cov}(z_2, z_n) \\ \vdots & & \ddots & \vdots \\ \text{Cov}(z_1, z_n) & \dots & \dots & \text{Var}(z_n) \end{pmatrix}$$

Exercise 2

What is the variance-covariance matrix of the random vector of errors of the regression model ε*? And that of* $\mathbf{Y}$*?*

Linear function of $\mathbf{Z}$

- A vector $\mathbf{U}$ of size $k \times 1$ is a linear function of $\mathbf{Z}$ if it can be written
$$\mathbf{U} = \mathbf{AZ}$$
for a matrix of constants $\mathbf{A}$ (i.e. whose elements are not random) of order ____

Exercise 3
Show that $\widehat{\beta}$, $\widehat{\mathbf{Y}}$, and $\mathbf{e}$ are linear functions of $\mathbf{Y}$.

Properties of a linear function of a random vector

Consider a linear function $\mathbf{U} = \mathbf{AZ}$ of a random vector $\mathbf{Z}$. We have

- $\mathrm{E}(\mathbf{U}) = \mathbf{A}\mathrm{E}(\mathbf{Z})$,
- $\mathrm{Var}(\mathbf{U}) = \mathbf{A}\mathrm{Var}(\mathbf{Z})\mathbf{A}'$.

Proof in [Rawlings et al., 1998], pages 83-84

Comment 2
If $\mathrm{Var}(\mathbf{Z}) = \sigma^2\mathbf{I}$, *then*

$$\mathrm{Var}(\mathbf{U}) = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}' = \mathbf{AA}'\sigma^2.$$

Properties of a linear function of a random vector

Exercise 4

Consider $Z_1, Z_2, \ldots, Z_n$ iid with mean μ and variance σ^2.

1. *Show that the mean $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ is a linear function of $\mathbf{Z}$ and give $\mathbf{A}$ such that $\overline{Z} = \mathbf{AZ}$.*
2. *Apply the properties shown on the previous slide to derive the mean and variance of $\overline{Z}$.*

Normal random vector

- The vector of random variables $\mathbf{Z} = (Z_1, \ldots, Z_p)'$ has a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and variance-covariance matrix $\boldsymbol{\Sigma}$ (assume for simplicity that $\boldsymbol{\Sigma}$ has full rank), if its probability distribution function is

$$f_Z(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right]$$

  where $|\cdot|$ is the determinant. We write $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- If $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the marginal probability distribution of each variable Z_i is $\mathcal{N}(\mu_i, \boldsymbol{\Sigma}_{ii})$.

Linear function of normal random vectors

- Consider $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\mathbf{U} = \mathbf{AZ} + \mathbf{b}$, where $\mathbf{A}$ is a $k \times n$ matrix of constants and $\mathbf{b}$ is a $k \times 1$ vector of constants, then $\mathbf{U}$ is normally distributed with
  - mean $\mathrm{E}(\mathbf{U}) =$ $E(AZ+b) = E(AZ) + E(b) = A\underbrace{E(Z)}_{} + b = A\mu + b$
  - variance $\mathrm{Var}(\mathbf{U}) =$ $Var(AZ+b) = var(AZ) = A\underbrace{Var(Z)}_{\Sigma}A'$

  i.e. $\mathbf{U} \sim \mathcal{N}($ $A\mu + b$, $A\Sigma A'$ $)$

Exercise 5

What is the distribution of $\mathbf{Y}$ *in the multiple linear regression with normal errors?*

$Z \rightarrow \varepsilon \sim N(\underbrace{0}_{\text{vector of zero}}, \sigma^2 I)$

$\underset{u}{y} = \underbrace{xB}_{b} + \varepsilon$, $A = I$

xB is constant

$\Rightarrow y \sim N(x\beta, \sigma^2 I)$

Regression estimates are linear functions of $\mathbf{Y}$

- We know (previous section) that $\widehat{\beta}$, $\widehat{\mathbf{Y}}$, and $\mathbf{e}$ are linear functions of $\mathbf{Y}$:
  - $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
  - $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} = \mathbf{PY}$ where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
  - $\mathbf{e} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$
- In this section, we will apply the properties of linear functions of random vectors to study the properties of $\widehat{\beta}$, $\widehat{\mathbf{Y}}$, and $\mathbf{e}$,
- Recall that
$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon.$$
- We already know (previous section) that if $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ then
$$\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \mathbf{I}\sigma^2).$$

February 15, 2019 9:17 AM

Properties of $\widehat{\beta}$

Exercise 6

Show the following propertied of $\widehat{\beta}$

> *The mean and variance-covariance matrix of $\widehat{\beta}$ are*
>
> $$\mathrm{E}(\widehat{\beta}) = \beta \qquad \mathrm{Var}(\widehat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2.$$
>
> *Moreover, when $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ we have*
>
> $$\widehat{\beta} \sim \mathcal{N}(\beta, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2).$$

Comment 3

- *$\widehat{\beta}$ is an unbiased estimator of β* *(if the model is correct!)*
- *The variances and covariances of the estimated regression coefficients are given by the elements of $(\mathbf{X}'\mathbf{X})^{-1}$ multiplied by σ^2.*

Ex. 6

Known $\hat{\beta} = (x'x)^{-1}x'y$

$E(\hat{\beta}) = A\,E(y) = (x'x)^{-1}x' = I\beta = \beta$

$y = x\beta + \varepsilon$

$E(\varepsilon) = 0$

$Var(\hat{\beta}) = A\,var(y)\,A'$

$= (x'x)^{-1}x'\,\sigma^2 I\,x(x'x)^{-1}$

bc $(x'x)^{-1}$ is symmetric

$= \sigma^2 (x'x)^{-1}x'x(x'x)^{-1}$

$= \sigma^2 (x'x)$

if $\varepsilon \sim N$, then $y \sim N$, and if $\hat{\beta} \sim N$ with paramaters $E(\hat{\beta})$ and $Var(\hat{\beta})$

Proof of the properties of $\widehat{\beta}$

Dwaine example

somewhat in the slide

Exercise 7

What is the variance of $\widehat{\beta}_2$ in Dwaine example? What is the covariance between $\widehat{\beta}_0$ and $\widehat{\beta}_2$? Give your answers in function of σ^2 (unknown).

$$
(X'X)^{-1} = \begin{array}{ccc|c} \hat\beta_0 & \hat\beta_1 & \hat\beta_2 & \\ \hline
29.73 & 0.07 & -1.99 & \hat\beta_0 \\
 & 0.0003 & -0.0055 & \hat\beta_1 \\
 & & 0.136 & \hat\beta_2
\end{array}
$$

$$
\hat\beta = \begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} \qquad Var(\hat\beta) = \sigma^2 (X'X)^{-1}
$$

$$Var(\hat\beta_i) = 0.136\sigma^2$$

$$Cov(\hat\beta_0, \hat\beta_2) = -1.99\sigma^2$$

negatively correlated

Properties of $\widehat{\mathbf{Y}}$

Exercise 8

Show the following properties of $\hat{\mathbf{Y}}$.

The mean and variance-covariance matrix of $\hat{\mathbf{Y}}$ *are*

$$\mathrm{E}(\hat{\mathbf{Y}}) = \mathbf{X}\beta,$$
$$\mathrm{Var}(\hat{\mathbf{Y}}) = \mathbf{P}\sigma^2,$$

where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Moreover, when $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ *we have*

$$\hat{\mathbf{Y}} \sim \mathcal{N}(\mathbf{X}\beta, \mathbf{P}\sigma^2).$$

Ex 6

Known $\hat{y} = X\hat{\beta} = Py$

$(AB)' = B'A'$

$P = X(X'X)^{-1}X'$

$E(\hat{y}) =$

$P' = P, \quad P^2 = P$

$Var(\hat{y}) =$

$E(\hat{y}) = E(X\hat{\beta}) = X E(\hat{\beta}) = X\beta$

$Var(\hat{y}) = Var(\underset{\text{constant}}{P}\underset{\text{random}}{y}) = P\underbrace{Var(y)}_{\sigma^2 I}P' = \sigma^2 PP' = \cancel{\sigma}^2 PP = \sigma^2 P^2$

Properties of **e**

Exercise 10

Show the following properties of **e**.

> *The mean and variance-covariance matrix of* **e** *are*
>
> $$\mathrm{E}(\mathbf{e}) = \mathbf{0},$$
> $$\mathrm{Var}(\mathbf{e}) = (\mathbf{I} - \mathbf{P})\sigma^2,$$
>
> *where* $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. *Moreover, when* $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ *we have*
>
> $$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, (\mathbf{I} - \mathbf{P})\sigma^2).$$

Comment 5

The diagonal elements P_{ii} *of* **P** *satisfy* $0 \leq P_{ii} \leq 1$ *because*

$\mathrm{Var}(\hat{y}_i) = \sigma^2 P_{ii} \geq 0$ $\quad \mathrm{Var}(e) = \sigma^2 (I - P)_{ii} = \sigma^2 (1 - P_{ii}) \geq 0$

Comments on the variance

We have
$$\mathrm{Var}(\mathbf{Y}) = \mathrm{Var}(\widehat{\mathbf{Y}}) + \mathrm{Var}(\mathbf{e})$$

Proof.

□

Comment 6

- $\mathrm{Var}(Y_i) = \sigma^2 = \mathrm{Var}(\widehat{Y}_i) + \mathrm{Var}(e_i)$, *i.e. data points having low variance of $\widehat{Y}_i$ have __________ variance on e_i, and vice-versa.*