Erasmus School of Economics

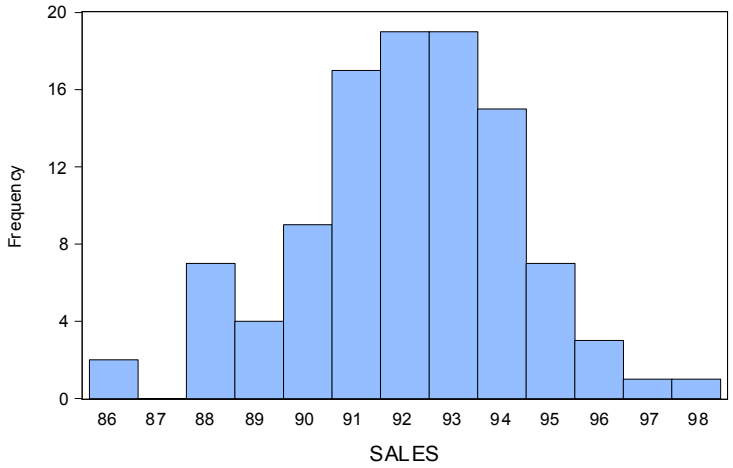
MOOC Econometrics

Lecture 1.1 on Simple Regression: Motivation

Philip Hans Franses

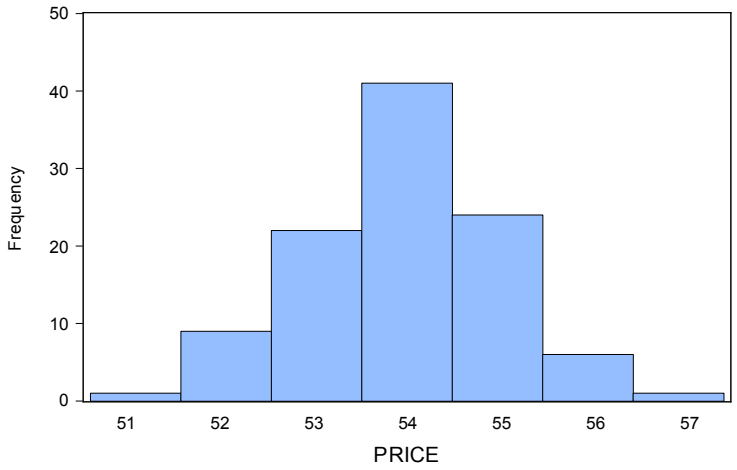
Erasmus University Rotterdam

Histogram of 104 sales data



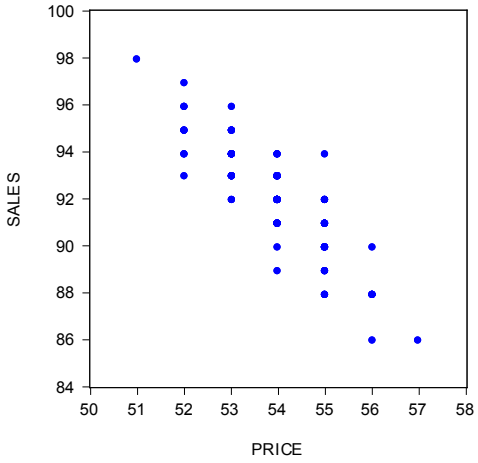
Frequency is number of weeks

Histogram of 104 price data



Frequency is number of weeks

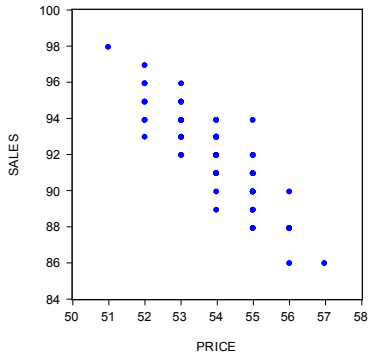
Scatter diagram of sales against price

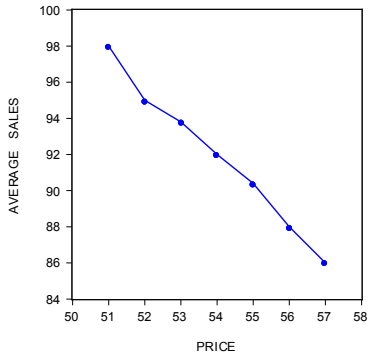


104 pairs of weekly observations on sales and price
(some sales-price combinations occur multiple times)

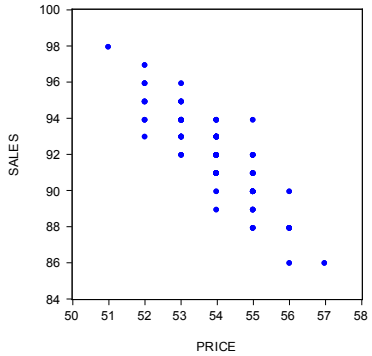
Average sales for given price

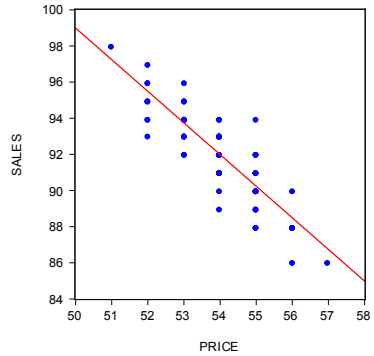
Price	51	52	53	54	55	56	57
Number of weeks	1	9	22	41	24	6	1
Average sales	98.0	95.0	93.8	92.0	90.4	88.0	86.0





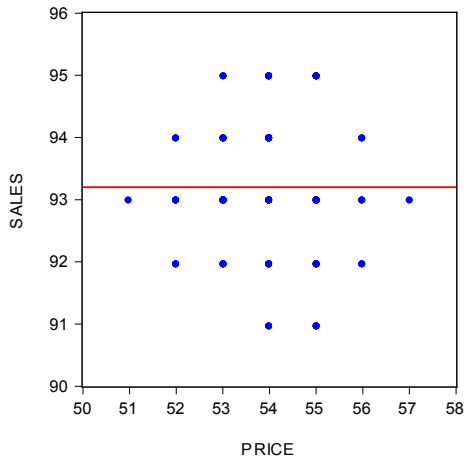
Fitting a straight line in a scatter diagram



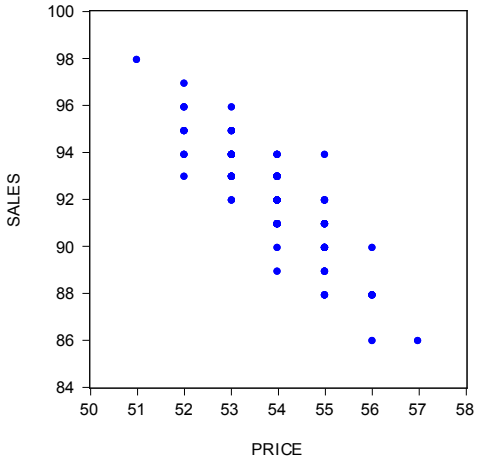


- Predicted Sales $= a + b \times$ Price
- Residual $e =$ Actual Sales $-$ Predicted Sales

Scatter diagram for data without price effect (b = 0)



Scatter diagram of sales against price



How to maximize turnover

Test

Define turnover as product of price and sales, where Sales $= a + b \times$ Price with $a > 0$ and $b < 0$. If a and b are known, the store manager can determine for which price turnover is maximal.
Derive the formula for the optimal price in terms of a and b.

- Answer: Let $P =$ Price and $T =$ Turnover $=$ Price $\times$ Sales, then

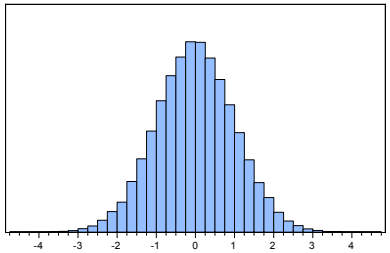
  $T = P(a + bP) = aP + bP^2$

  $\frac{dT}{dP} = a + 2bP = 0$

- Optimal price: $P = -\frac{a}{2b}$

Normal Distribution

- Sales $\sim NID(\mu, \sigma^2)$
- Standard normal distribution: $\mu = 0$ and $\sigma^2 = 1$



Density function (discretized; area is 1)

- Estimator of population mean μ: sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

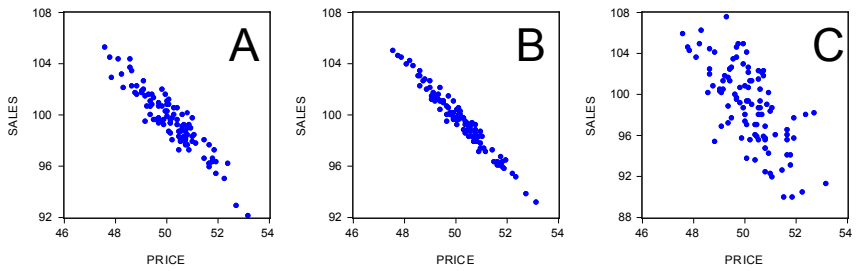
Overview of coming lectures

- Lecture 1.2: Simple regression model
- Lecture 1.3: The technique of regression
- Lecture 1.4: Assumptions and statistical properties
- Lecture 1.5: Two applications
- Modules 2-6: Various extensions
- Simple regression provides fundamental basis

Prediction

Test

Which situation is easiest to predict sales for given price?



- B is easiest: least variation around line.

TRAINING EXERCISE 1.1

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

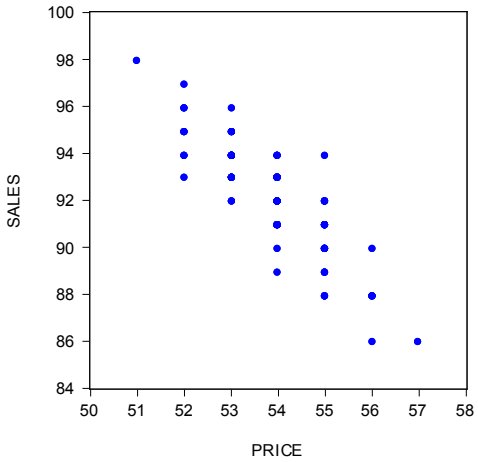
MOOC Econometrics

Lecture 1.2 on Simple Regression: Representation

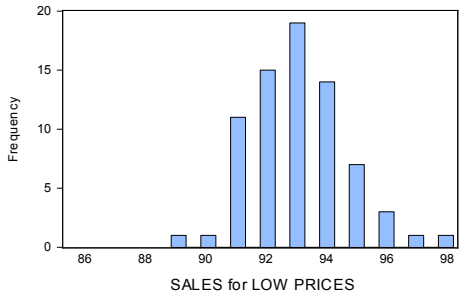
Philip Hans Franses

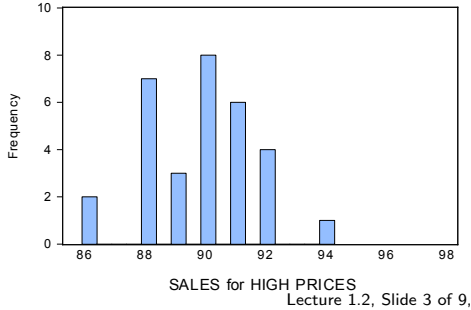
Erasmus University Rotterdam

Scatter diagram of sales against price



Two histograms of sales





Simple regression: conditional mean

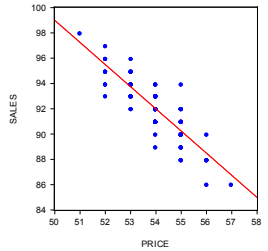
- y_i Normally Identically Distributed with mean μ and variance σ^2:
  $y_i \sim NID(\mu, \sigma^2)$
- Expected value: $E(y_i) = \mu$
  Variance: $E(y_i - \mu)^2 = \sigma^2$
- Sample estimates: $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
  (see Building Blocks) $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
- Simple regression: replace unconditional mean $E(y_i) = \mu$
  by conditional mean $E(y_i) = \alpha + \beta x_i$

Simple regression equation

- Unconditional prediction with $y_i \sim N(\mu, \sigma^2)$: $E(y_i) = \mu$
- Conditional prediction with $y_i \sim N(\alpha + \beta x_i, \sigma^2)$: $E(y_i) = \alpha + \beta x_i$
- An alternative format is: $y_i = \alpha + \beta x_i + \varepsilon_i$
  $\varepsilon_i \sim N(0, \sigma^2)$
- If x_i is fixed (non-random), $\varepsilon_i \sim N(0, \sigma^2)$, and $y_i = \alpha + \beta x_i + \varepsilon_i$, then y_i has mean $\alpha + \beta x_i$ and variance σ^2. (see Building Blocks)

Prediction

- Consider again scatter plot of sales and price:



Test

Predict sales for a price of 50, and also for a price of 58.

- Answer: Fit straight line to the observed data.
- Price $= 50 \rightarrow$ Sales ≈ 99; Price $= 58 \rightarrow$ Sales ≈ 85.

Absolute and relative changes

Test

Retail store A has a sales level of 150 units and store B of 250 units. The two store managers start an advertising campaign, after which store A sells 225 units and store B sells 400.
Which store has the largest relative increase in sales?

- Answer: Relative change for A: $\frac{225-150}{150} = 0.5$
  Relative change for B: $\frac{400-250}{250} = 0.6 \quad \leftarrow$ largest
- If current level x increases by dx, then relative increase is $\frac{dx}{x}$

Elasticity

- For $y = \alpha + \beta x$, the slope (marginal effect) is $\beta = \frac{dy}{dx}$
- Definition of elasticity: $\frac{dy/y}{dx/x}$
  So: relative change in y divided by relative change in x
- Elasticity in $y = \alpha + \beta x$:
  $$\text{elasticity} = \frac{dy/y}{dx/x} = \frac{dy}{dx} \times \frac{x}{y} = \beta \times \frac{x}{y}$$
- Fact: elasticity in $\log(y) = \alpha + \beta \log(x)$ is equal to β.

TRAINING EXERCISE 1.2

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

Lecture 1.3 on Simple Regression: Estimation

Philip Hans Franses

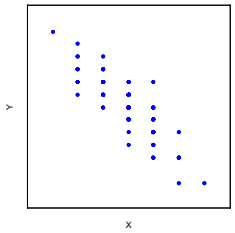
Erasmus University Rotterdam

Model and data

- Simple regression model: $y_i = \alpha + \beta x_i + \varepsilon_i$
- In econometrics, we do not know α and β (and ε_i)
  we do have observations on x_i and y_i
- Use observed data on x_i and y_i to find "optimal" values of a and b so that $y_i \approx a + bx_i$.
- The line $y = a + bx$ is called the regression line.

Data and regression line

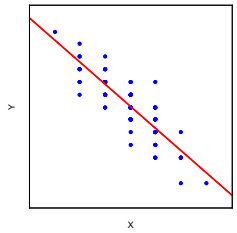
- Data: n pairs of observations (x_i, y_i) for $i = 1, 2, \ldots, n$



- Fitted line: $y_i = a + bx_i$
  - a: intercept
  - b: slope
- Residuals: $e_i = y_i - a - bx_i$
- Choose fitted line such that e_i are small.

Data and regression line

- Data: n pairs of observations (x_i, y_i) for $i = 1, 2, \ldots, n$



- Fitted line: $y_i = a + bx_i$
  - a: intercept
  - b: slope
- Residuals: $e_i = y_i - a - bx_i$
- Choose fitted line such that e_i are small.

Least squares

- Least squares criterion: find a and b by minimizing

  $S(a,b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$

- Get a and b by solving $\frac{\partial S}{\partial a} = 0$ and $\frac{\partial S}{\partial b} = 0$.

- We first analyze $\frac{\partial S}{\partial a} = 0$ (and later we consider $\frac{\partial S}{\partial b} = 0$):

  $0 = \frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = -2 \sum_{i=1}^{n} e_i$

- Note: One residual follows from the other n-1 residuals:

  $e_n = -(e_1 + e_2 + ... + e_{n-1})$

Solving $\frac{\partial S}{\partial a} = 0$

- $0 = \frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = -2 \sum_{i=1}^{n} y_i + 2na + 2b \sum_{i=1}^{n} x_i$

- Denote sample means by $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$,

  then above equation gives: $-2\bar{y} + 2a + 2b\bar{x} = 0$

- So: $a = \bar{y} - b\bar{x}$

Test question

Test

Suppose we apply least squares on de-meaned data, with dependent variable $y_i^* = y_i - \bar{y}$ and explanatory factor $x_i^* = x_i - \bar{x}$.
Which values do a and/or b take in this special case?

- Answer:

  Check that $\bar{y}^* = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{1}{n} \sum_{i=1}^{n} \bar{y} = \bar{y} - \bar{y} = 0$,

  and likewise $\bar{x}^* = 0$.

- So: $a = \bar{y}^* - b\bar{x}^* = 0 - b \times 0 = 0$.

- Later we will see that b is not affected by de-meaning.

Solving $\frac{\partial S}{\partial b} = 0$

- $S(a,b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$

- $0 = \frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = -2 \sum_{i=1}^{n} x_i e_i$

- Note: if $x_1 \neq 0$, then $e_1 = -(x_2 e_2 + x_3 e_3 + ... + x_n e_n)/x_1$

- $\begin{aligned} 0 &= \sum_{i=1}^{n} x_i (y_i - a - bx_i) \\ &= \sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} x_i y_i - (\bar{y} - b\bar{x}) \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y} + b \sum_{i=1}^{n} x_i \bar{x} - b \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} x_i (y_i - \bar{y}) - b \sum_{i=1}^{n} x_i (x_i - \bar{x}) \end{aligned}$

- So: $b = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$

Solving $\frac{\partial S}{\partial b} = 0$

- $b = \frac{\sum_{i=1}^n x_i(y_i-\bar{y})}{\sum_{i=1}^n x_i(x_i-\bar{x})}$
- Use that $\sum_{i=1}^n (y_i - \bar{y}) = \sum_{i=1}^n y_i - n\bar{y} = 0$, and similarly
  $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = 0$, hence
  $\bar{x}\sum_{i=1}^n (y_i - \bar{y}) = 0$ and $\bar{x}\sum_{i=1}^n (x_i - \bar{x}) = 0$
- We get:
  $b = \frac{\sum_{i=1}^n x_i(y_i-\bar{y})}{\sum_{i=1}^n x_i(x_i-\bar{x})} = \boxed{\frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2}}$

Test

What value does b take if all observations of y_i are equal to 93?

- Answer: $b = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2}$ with $y_i - \bar{y} = 93 - 93 = 0$.
  So: $b = 0$.

R-squared

- Data (x_i, y_i) give numerical values of a and b, with $a = \bar{y} - b\bar{x}$.
- Then $y_i = a + bx_i + e_i = \bar{y} - b\bar{x} + bx_i + e_i$, so
  $y_i - \bar{y} = b(x_i - \bar{x}) + e_i$ $\quad (*)$
- Deviation $y_i - \bar{y}$ partly explained by $x_i - \bar{x}$ $\quad$ (e_i is unexplained).
- Seen before: $\sum_{i=1}^n e_i = 0$ and $\sum_{i=1}^n x_i e_i = 0$, hence
  $\sum_{i=1}^n (x_i - \bar{x})e_i = \sum_{i=1}^n x_i e_i - \bar{x}\sum_{i=1}^n e_i = 0$.
- Squaring and Summing (SS) both sides of $(*)$ therefore gives:
  $\sum_{i=1}^n (y_i - \bar{y})^2 = b^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n e_i^2$
  SSTotal $=$ SSExplained $+$ SSResidual
- $R^2 = \frac{\text{SSExplained}}{\text{SSTotal}} = \boxed{1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i-\bar{y})^2}}$

Estimate of error variance

- $y_i = \alpha + \beta x_i + \varepsilon_i$ with $\varepsilon_i \sim NID(0, \sigma^2)$
- Unknown σ^2 is estimated from residuals $e_i = y_i - a - bx_i$.
- Residuals $e_i, i = 1, 2, ..., n$, have $n - 2$ free values (seen before).
- $s^2 = \frac{1}{n-2}\sum_{i=1}^n (e_i - \bar{e})^2$.
- Seen before: $\sum_{i=1}^n e_i = 0$, so $\bar{e} = 0$.
- Therefore: $\boxed{s^2 = \frac{1}{n-2}\sum_{i=1}^n e_i^2}$
- (see Building Blocks for case $n - 1$)

TRAINING EXERCISE 1.3

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

Lecture 1.4 on Simple Regression: Evaluation

Philip Hans Franses

Erasmus University Rotterdam

Prediction interval

- Least squares: data $(x_i, y_i), i = 1, 2, \ldots, n \quad \rightarrow \quad a$ and b
- Regression line: $y = a + bx$
  Residuals: $e_i = y_i - a - bx_i$
  Residual standard deviation: $s = \sqrt{s^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2}$
- Question: Predict outcome of y_0 for new value of x_0.
- Actual value: $y_0 = \alpha + \beta x_0 + \varepsilon_0$
  Point prediction: $\hat{y}_0 = a + bx_0$
  Interval for ε_0: $(-ks, ks)$.
- Prediction interval for y_0 : $(\hat{y}_0 - ks, \hat{y}_0 + ks)$

Test question

- Prediction interval for y_0 : $(\hat{y}_0 - ks, \hat{y}_0 + ks)$.

Test

Which prediction interval has the highest confidence to contain y_0: for $k = 1$ or $k = 2$?

- Answer: $k = 2$, as the interval is wider.

Assumptions

- A1: Data Generating Process is $y_i = \alpha + \beta x_i + \varepsilon_i$.
- A2: The n observations of x_i are fixed numbers.
- A3: The n error terms ε_i are random, with $E(\varepsilon_i) = 0$.
- A4: The variance of the n errors is fixed, $E(\varepsilon_i^2) = \sigma^2$.
- A5: The errors are uncorrelated, $E(\varepsilon_i \varepsilon_j) = 0$ for all $i \neq j$.
- A6: α and β are unknown, but fixed for all n observations.
- A7: $\varepsilon_1, \ldots, \varepsilon_n$ are jointly normally distributed;
  with A3, A4, A5: $\varepsilon_i \sim NID(0, \sigma^2)$.

Statistical properties of b: preliminaries

- Least squares slope estimator: $b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
- Derive properties of b from those of error terms ε_i (see A1-A7).
- We will show that $b = \beta + \sum_{i=1}^{n} c_i\varepsilon_i$, where $c_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ are fixed numbers (see A2).
- Next slide shows steps needed for this result.

Derivation of the constants c_i

- $b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ (A)
- A1: $y_i - \bar{y} = (\alpha + \beta x_i + \varepsilon_i) - (\alpha + \beta\bar{x} + \bar{\varepsilon})$
  $= \beta(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})$ (B)
- $b \overset{(A,B)}{=} \beta + \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ (C)
- $\sum_{i=1}^{n}(x_i - \bar{x})\bar{\varepsilon} = \bar{\varepsilon}\sum_{i=1}^{n}(x_i - \bar{x}) = \bar{\varepsilon}(\sum_{i=1}^{n} x_i - n\bar{x})$
  $= \bar{\varepsilon}(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i) = 0$ (D)
- $b \overset{(C,D)}{=} \beta + \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta + \sum_{i=1}^{n} c_i\varepsilon_i$ with $c_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

The mean of b: unbiased

- $b = \beta + \sum_{i=1}^{n} c_i\varepsilon_i$ with c_i fixed (due to A2).
- $E(b) = E(\beta) + \sum_{i=1}^{n} E(c_i\varepsilon_i)$.
- A6: β fixed, hence $E(\beta) = \beta$.
- c_i fixed, so $E(c_i\varepsilon_i) = c_iE(\varepsilon_i) = 0$ (due to A3).
- Hence: $E(b) = \beta + \sum_{i=1}^{n} 0 = \beta$.
- So b is unbiased estimator of slope parameter β.

Formula for $\sigma_b^2 = \text{var}(b)$

- $b = \beta + \sum_{i=1}^{n} c_i\varepsilon_i$ with $c_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
- $\sigma_b^2 = E((b - E(b))^2) = E((b-\beta)^2)$
  $= E((\sum_{i=1}^{n} c_i\varepsilon_i)^2) = E(\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\varepsilon_i\varepsilon_j)$
  $\overset{(A2)}{=} \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jE(\varepsilon_i\varepsilon_j)$
  $\overset{(A5)}{=} \sum_{i=1}^{n} c_i^2E(\varepsilon_i^2)$
  $\overset{(A4)}{=} \sigma^2\sum_{i=1}^{n} c_i^2$
  $\overset{(*)}{=} \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
- Notes: Step A5: $E(\varepsilon_i\varepsilon_j) = 0$ for all $i \neq j$
  Step (*): $\sum_{i=1}^{n} c_i^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(\sum_{i=1}^{n}(x_i-\bar{x})^2)^2} = \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

Test on slope parameter β

- Seen before: $b = \beta + \sum_{i=1}^{n} c_i \varepsilon_i$, $E(b) = \beta$, and $\sigma_b^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- A7: ε_i normal, then b normal (see Building Blocks)
- So: $b \sim N(\beta, \sigma_b^2)$ and $Z = \frac{b-\beta}{\sigma_b} \sim N(0,1)$
- Replace unknown σ^2 by $s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$, then $s_b^2 = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- $t_b = \frac{b-\beta}{s_b} \sim t(n-2)$ (compare Building Blocks)
- t-test on H_0: $\beta = 0$ based on $t_b = \frac{b}{s_b}$
  Rule-of-thumb for large n: reject H_0 if $t_b < -2$ or $t_b > 2$.

Confidence and prediction intervals

- Approximate 95% confidence interval for b:
  $\frac{b-\beta}{s_b} \approx N(0,1)$ with interval $(-2, 2)$
- $-2 \leq \frac{b-\beta}{s_b} \leq 2 \quad \rightarrow \quad b - 2s_b \leq \beta \leq b + 2s_b$
- Approximate 95% prediction interval for y (see before):
  $a + bx - 2s \leq y \leq a + bx + 2s$
- Note: $-2s \leq \varepsilon \leq 2s$ is approximate 95% confidence interval for ε, uncertainty in a and b is neglected here.

Test question

Test

Let measurement scale of the dependent variable y be fixed, and compare two scales for the explanatory factor x: first x is measured in 10 units (recorded value of 5 corresponds to 50 units), and later in 100 units (recorded value of 5 corresponds to 500 units).
Which case gives the widest confidence interval for b?

- Answer: For U units, the value of x changes from $\frac{U}{10}$ to $\frac{U}{100}$, so x becomes 10 times as small.
- As $b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, b is multiplied by $\frac{\frac{1}{10}}{\frac{1}{100}} = 10$.
- So b becomes 10 times as large, and same for confidence interval.

Seven assumptions

Assumption	Violation	Lecture
A1: $y_i = \alpha + \beta x_i + \varepsilon_i$	More than one x var.	2
	Choose among x-var.	3
	Binary y_i (0 or 1)	5
A2: x_i fixed	Random x_i	4
A3/A6: $E(\varepsilon_i) = 0$, α, β fixed	Parameter breaks	3
A4: Homoskedastic	Heteroskedastic errors	6
A5: Uncorrelated	Correlated errors	6
A7: ε normal	Often not needed	2-6

TRAINING EXERCISE 1.4

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

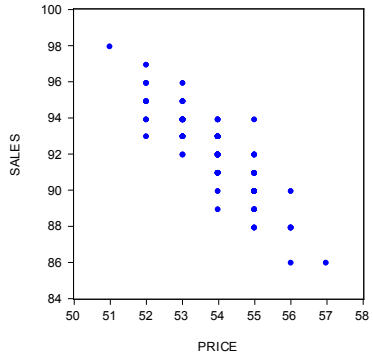
Lecture 1.5 on Simple Regression: Application

Philip Hans Franses

Erasmus University Rotterdam

Effect of price on sales

- 104 weekly data



- Model: Sales $= \alpha + \beta$Price $+ \varepsilon$

Estimation results

- Regression equation: Sales $= a + b$Price $+ e$

Variable	Coefficient	Standard error	t-Statistic	p-value
Intercept	$a = 186.507$	5.767	32.339	0.000
Price	$b = -1.750$	0.107	-16.380	0.000

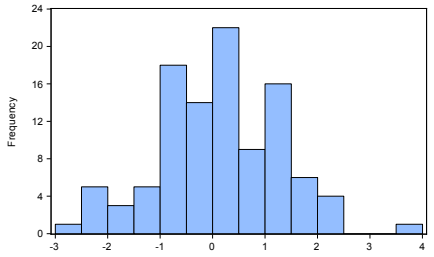
- $R^2 = 0.725$, $s = 1.189$
- 95% confidence interval β:

$$-1.750 - 2 \times 0.107 \leq \beta \leq -1.750 + 2 \times 0.107$$
$$-1.964 \leq \beta \leq -1.536$$

- On average: price 1 unit $\downarrow$ $\rightarrow$ sales 1.5 - 2.0 units $\uparrow$
- Price effect on sales is highly significant.

Histogram of residuals

- $e =$ Sales $- a - b$Price



Mean	$= 0.000$	(normal: 0, see Building Blocks)
Standard dev.	$= 1.183$	
Skewness	$= 0.029$	(normal: 0, see Building Blocks)
Kurtosis	$= 3.225$	(normal: 3, see Building Blocks)

- Reasonably normal

Optimal price for maximal turnover

- Store manager can use regression outcomes to set price.
- Objective: maximize Turnover = Price $\times$ Sales.
- Optimal price $P_0 = \frac{-a}{2b}$ (see Lecture 1.1).
- $a = 186.5$ and $b = -1.75$, so $P_0 = \frac{186.5}{3.5} = 53.3$.
- Associated predicted sales S_0:

$$S_0 = a + bP_0 = 186.5 - 1.75 \times 53.3 \approx 93.$$

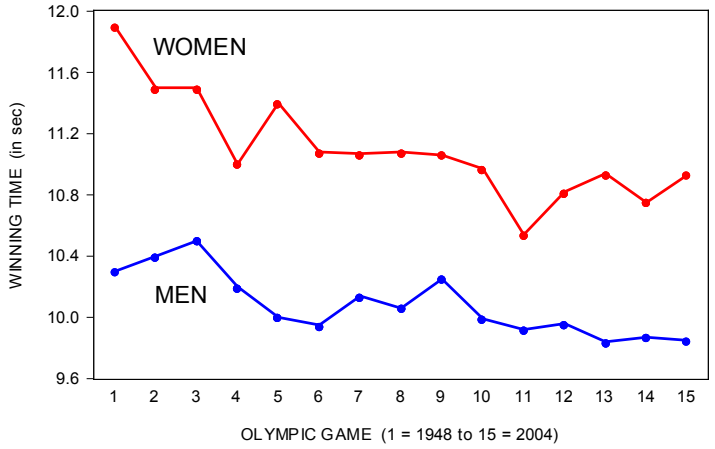
Confidence interval for optimal sales level

Test

Let S = Sales and P = Price, with model $S = \alpha + \beta P + \varepsilon$. Optimal price is $P_0 = -\frac{\alpha}{2\beta}$, with associated sales S_0. Regression gives $a = 186.5$ ($\text{SE}_a = 5.767$), $b = -1.750$ ($\text{SE}_b = 0.107$), $s = 1.189$.
Find the (approximate) 95% confidence interval for sales if the store manager sets the price at the optimal level P_0.
Hint: First show that $S_0 = \frac{\alpha}{2} + \varepsilon_0$.

- Answer: $S_0 = \alpha + \beta P_0 + \varepsilon_0 = \alpha + \beta \times (-\frac{\alpha}{2\beta}) + \varepsilon_0 = \frac{\alpha}{2} + \varepsilon_0$
  95% interval for α: $a \pm 2 \times \text{SE}_a = 186.5 \pm 2 \times 5.767 = (175, 198)$
  95% interval for ε_0: $\pm 2 \times 1.189 = (-2.4, 2.4)$
- Optimal sales: lower bound: $(175/2) - 2.4 \approx 85$
  upper bound: $(198/2) + 2.4 \approx 101$

Olympic winning times 100 meter (athletics)



Olympic winning times 100 meter (athletics)

- W = winning time (seconds), G = game (from 1=1948 to 15=2004)
- Simple regression: $W_i = \alpha + \beta G_i + \varepsilon_i$ (with $G_i = i$ for $i = 1, \ldots, 15$)
- Estimation results:

	a	SE_a	b	SE_b	R^2
Men	10.386	0.067	-0.038	0.007	0.673
Women	11.606	0.111	-0.063	0.012	0.672

- 95% confidence intervals for b: men: -0.038 ± 0.014
  women: -0.063 ± 0.024
- Women seem to have made most progress.
  Model assumes fixed gain β (in seconds per game).

Model with fixed relative gains

- Maybe nonlinear trend is better?
- If $W_i = \gamma e^{\beta G_i}$, then $\frac{W_{i+1}}{W_i} = e^{\beta(G_{i+1}-G_i)} = e^{\beta}$ is fixed.
- Then $\log(W_i) = \alpha + \beta G_i + \varepsilon_i$ (with $G_i = i$ and $\alpha = \log(\gamma)$)
- Outcomes:

	a	SE_a	b	SE_b	R^2
Men	2.341	0.0065	-0.0038	0.0007	0.677
Women	2.452	0.0099	-0.0056	0.0011	0.673

- Again, women made most progress.

Forecast of winning times for 2008 and 2012

Test

Use the four models shown below to forecast winning times (in seconds) of men and women in the Olympic games of 2008 (with $G_i = i = 16$) and 2012 (with $G_i = i = 17$).

Men: $W_i = 10.386 - 0.038 G_i + e_i$ $\log(W_i) = 2.341 - 0.0038 G_i + e_i$
Women: $W_i = 11.606 - 0.063 G_i + e_i$ $\log(W_i) = 2.452 - 0.0056 G_i + e_i$

Note: 'log' denotes the natural logarithm.

Answer:

	Men		Women	
	2008	2012	2008	2012
Actual time	9.69	9.63	10.78	10.75
Linear trend	9.78	9.74	10.60	10.54
Nonlinear trend	9.78	9.74	10.62	10.56

TRAINING EXERCISE 1.5

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).