MOOC Econometrics

Erasmus
School of
Economics

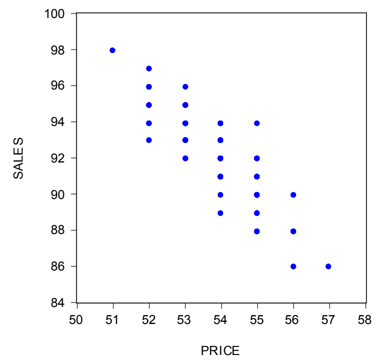
Lecture 1.5 on Simple Regression: Application

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

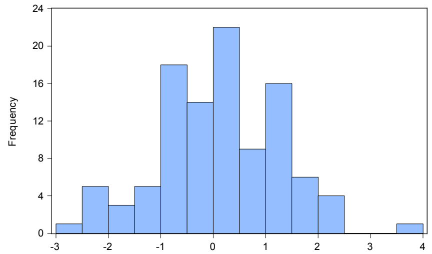
- $R^2 = 0.725$, $s = 1.189$
- 95% confidence interval β:

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

- 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 $a = 186.5$ ($\mathrm{SE}_a = 5.767$), $b = -1.750$ ($\mathrm{SE}_b = 0.107$), $s = 1.189$.
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: $S_0 = \alpha + \beta P_0 + \varepsilon_0 = \alpha + \beta \times (-\frac{\alpha}{2\beta}) + \varepsilon_0 = \frac{\alpha}{2} + \varepsilon_0$

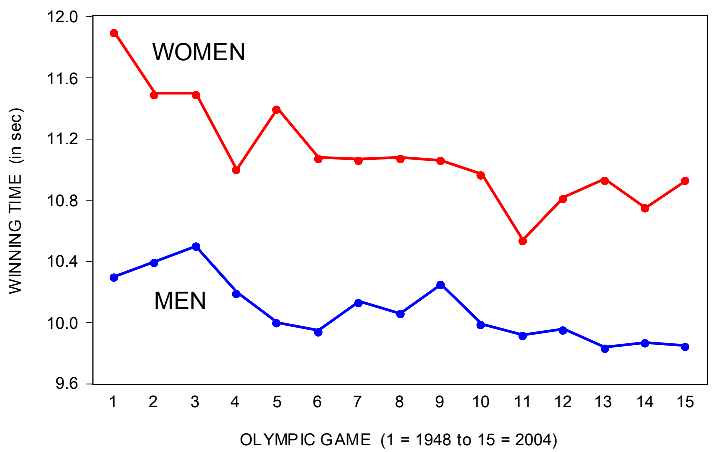
  95% interval for α: $a \pm 2 \times \mathrm{SE}_a = 186.5 \pm 2 \times 5.767 = (175, 198)$

  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).

Correction Lecture 1.5 Slide 6 (Video at 2 minutes and 45 seconds)

The answer of the test is not fully correct, as the standard error of the forecast should not be obtained by adding the two standard errors but by taking the square root of the added variances. The correct answer reads as follows.

$S_0 = \alpha/2 + \varepsilon_0$ and $\widehat{S}_0 = a/2$, so that the forecast error $f = S_0 - \widehat{S}_0 = (\alpha - a)/2 + \varepsilon_0$ is the sum of two components that are uncorrelated because of Assumption A5. Then $\mathrm{var}(f) = (1/4)\mathrm{var}(a) + \mathrm{var}(\varepsilon_0)$, with estimated value $(1/4)\mathrm{SE}_a^2 + s^2 = (1/4)(5.767)^2 + (1.189)^2 = 9.728$, so the standard error of the forecast is $\sqrt{9.728} \approx 3.1$.

The point forecast is $a/2 = 186.5/2 \approx 93.3$, so an approximate 95% interval is $93.3 \pm 2 \times 3.1$. This interval runs from 87.1 to 99.5, with width 12.4. After rounding, we find the interval [87, 100], which is somewhat smaller than the interval [85, 101] presented in the lecture.

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

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).