Why do you need statistics and regression analysis?

- Ability to read market research papers
- Analyze your own data in a simple way
- Assist you in pricing and marketing decisions
- Read scientific research

- We will cover first intuitive principles only, which are familiar from statistics course.

Go to the Schwerpunktfach „Angewandte Ökonometrie und Statistik“ of our Department to learn in detail.
Alternative methods of estimating demand (getting data)

- Look at the literature
- Consumer interviews
- Experience of the past
- Market experiments in different subsidiaries or across time
- Scanner data make it much more easy nowadays
  - Sales of single products can be easily tracked
  - Prices are easily changed because of electronic systems
  - ==> simple experiments can easily be done
- Use data from “Vorteilscard”, discount cards, city card, etc. consumer information systems
- E-business is collecting automatically information, e.g. site visits, log-in information, etc.

But basically with all these methods, once you gathered data, you need methods to get something meaningful out of them.
Suppose you have TV advertisements in different market segments (countries) and have collected sales in all these markets.

Now you would like to know, how many products you could sell more, if you added ten seconds of TV ad.

Measured sales of your product

Regression line

Duration of TV ads in seconds
How to find this relation?

- Given the points you have collected: how can you best fit a straight line through these points?

- Well, you just try to find a straight line, which **minimizes** the squared sum of the vertical **distances** of your points to this line
  - Squared sum because negative and positive distances should be considered
Regression analysis

- A statistical technique that describes the way in which one variable is related to another
- Used to estimate demand (but also to lots of other problems)

- Simplest version: LHS (Sales, left hand side variable or dependent variable – but I find this easier to remember) is determined linearly by the RHS (Advertisements, right hand side var. or independent variable)

\[ S_i = \alpha + \beta A_i + e_i \]

(we first try it with one RHS variable, but it’s not more difficult if we use several ones)

- \( e_i \) is an error term, capturing other issues (variables) not covered in the analysis or pure randomness.
Errors e should be minimal. β is the slope coefficient we are looking for.

\[ S_2 = \alpha + \beta A \]

\[ e_1 = -1 \]

\[ e_2 = 1.5 \]
Multiple regression

Includes two or more independent variables

\[ S_i = \alpha + \beta_1 A_i + \beta_2 P_i + e_i \]

where:
- \( S_i \) = sales
- \( A_i \) = Advertising expenses
- \( P_i \) = price

- Independent influence of one variable is measured
- i.e. holding the other variable constant:
- partial effect of one variable can be identified.
How good is your estimate?

- Coefficient of determination: “R-squared”
- Measures goodness of fit of the estimated regression line
- Proportion of total variance explained by your variables
- Varies between 0 and 1
FIGURE 5.6 Six Regression Lines: Coefficient of Determination Equals 0, 0.2, 0.4, 0.6, 0.8, and 1.0
Unfortunately parameters $\alpha$ and $\beta$ are not always reliably estimated

- It could be that you don’t have enough data or that the relationship is unstable, so that your estimated parameter $\beta_1$ is not reliable (i.e. the value you found is just random).

- T-statistic (t-value) tells us if we can say with some confidence that the parameter $\beta_1$ is different from zero, i.e. that there is a relation between the two variables in question.

- Rule of Thumb: as long as the t-statistic is greater than 2, you can be sure, that the found coefficient is not just random. Coefficients with t’s below 1.5 or so, should be disregarded
  - (you can say: there is no significant relation between these two variables; technically that means that the chance is high (more than 10% that the found coefficient is just random)
Example

Sales = 20 + 0.5 Income + 1.3 Advert – 0.45 Price

(3.4)   (4.2)   (0.7)   (2.2)

T-values are in parentheses

- If income of the population rises by one unit (the unit you had in your data), then sales increase by 0.5 units – holding advertising expenses and prices constant.
- If price rises by one unit, then sales decrease by 0.45 units – again holding the other two variables constant.
- Advertising does not seem to have a significant effect on sales (t-value is too low). There is a very high chance (around 50%) that the positive effect, you measured has arisen just by chance.
Interpreting regression results

- Most of the time you will not do the statistics yourself, but you should be able to read and interpret the stuff.

- Usual statistics programs can do the job, like Stata, SPSS, even Excel.

- Main advantage of regression over other statistical tools like simple correlation or scatter graphs, etc.:
  - The impact of several variables can be checked simultaneously.
  - Ceteris paribus condition: the coefficient $\beta_1$ reports the influence of this variable, **holding all other variables unchanged**
    - This is exactly what is done in all economic models routinely!!
    - This is also what you would like to know, if you are trying to “manipulate” the market: e.g. you would like to reduce prices by 1€ (holding all other market conditions constant) -> what will be the effect on demand???
Example (1)

- You collected data on sales of Ford cars for each of the last 60 months.
- Now you want to find out how the sales of your (Ford) cars react to price changes. You might have had some special discount periods among these 60 months.
- The price rose more or less over this period and still sales increased. What does this tell you about your demand function?
Example (continued)

- Did you use real price?
- What happened to income of your potential customers?
- What happened to prices of your main competitors (Opel, VW, ...)?
- What happened to taxes, ...?

- Only multiple regression can tell you, if your discount actions have been successful
Identification problem

- How can you identify a demand curve, if you have only data points on price and quantity?
  - These data points arise both from shifts in demand and supply.
- Very important issue for econometricians, but difficult to handle in practice
- You want to estimate price elasticity of demand, but may not be able to do so, because demand was not stable over time (or the regions you looked at)
- To identify demand properly, you need to assume, that all the variation in your data come from changes in supply only
- Possible problem: you fail to distinguish between movements along the demand curve (say downwards) and shifts thereof!!!
Identification problem

FIGURE 5.1  Price Plotted against Quantity, 1997–1999
Identification problem
Identification problem

**Figure 5.3** Fixed Demand Curve and Shifting Supply Curve

Copyright © 2002 by W.W. Norton & Company
Chapter 5/Question 1

The Klein Corporation’s marketing department, using regression analysis, estimates the firm’s demand function, the result being

\[ Q = -104 - 2.1P + 3.2I + 1.5A + 1.6Z \]

\[ R^2 = 0.89 \]

Sample = 200 observations

where Q is the quantity demanded, P is the price, I is disposable income, A is advertising expenditure and Z is the price of a competing product.

a) If I = 5,000, A = 20 and Z = 1,000, what is the Klein Corporation's demand curve?

b) If P = 500 (and the conditions in part a hold), estimate the quantity demanded.

c) How well does this regression equation fit the data?
Chapter 5/Question 2

The Strunk Company hires a consultant to estimate the demand function for its product. The result of a regression analysis is:

\[ \log Q = 1.48 - 0.16 \log P + 0.43 \log Z \]

\[ (2.5) \quad (3.4) \]

\[ R^2 = 0.95 \]
\[ \text{Sample = 89 observations} \]

where \( Q \) is the quantity demanded, \( P \) is the price and \( Z \) is the price of a competing product.

a) Strunk’s president argues that a 2 percent price reduction will result in a 3.2 percent increase in the number of units sold by the firm. Do you agree?

b) Calculate the cross price elasticity of demand between Strunk’s product and the rival product.

c) Comment on the goodness of fit of the regression.
Chapter 5/Solutions

**Question 1**

a) \[ Q = -104 - 2.1P + 3.2 (5000) + 1.5 (20) + 1.6 (1000) \]
   \[ Q = 17,526 - 2.1P \] demand function

b) \[ Q = 17,526 - 2.1 (500) \]
   \[ Q = 16,476 \]

c) The model explains 89 percent of the variation in Q!

**Question 2**

a) The price elasticity of demand equals -0.16, a 2 percent price reduction will result in
   0.32 percent increase in quantity demanded.

b) The cross price elasticity of demand between Strunk’s product and the rival product
   equals 0.43. A 1 percent increase in the price of the rival’s product will result in a 0.43
   percent increase in quantity demanded.

c) The model explains 95 percent of the variation in Q!
Since the Hawkins Company’s costs (other than advertising) are essentially all fixed costs, it wants to maximize its total revenue (net of advertising expenses). According to a regression analysis (based on 124 observations) carried out by a consultant hired by the Hawkins Company,

\[ Q = -23 - 4.1P + 4.2I + 3.1A, \]

where \( Q \) is the quantity demanded (in dozens), \( P \) is the price of the firm’s product (in dollar per dozen), \( I \) is per capita income (in dollars), and \( A \) is advertising expenditure (in dollars).

a) If the price of the product is $10 per dozen, should the firm increase its advertising?

b) If the advertising budget is fixed at $10,000, and per capita income equals $8,000, what is the firm’s marginal revenue curve?

c) If the advertising budget is fixed at $10,000, and per capita income equals $8,000, what price should the Hawkins Company charge?
2a) Total revenue less advertising expenditures is
\[
TR - A = -23P - 4.1P^2 + 4.2IP + 3.1AP - A \\
= -23P - 4.1P^2 + 4.2IP + (3.1P - 1)A \\
= -640 + 42I + 30A \quad \text{if} \quad P = 10 \\
\frac{d(TR - A)}{dA} = 30 > 0 \quad \text{if} \quad P = 10
\]

Advertising expenditures increase net revenues and should be increased.

2b) \[
P = \frac{1}{4.1} \times (-23 - Q + 33,600 + 31,000) \\
= 15,750 - 0.244Q \quad \text{MR} = 15,750 - 0.488Q.
\]

2c) Total revenue is maximized where marginal revenue equals zero. Given the levels of advertising and per capita income, marginal revenue is zero at an output of \( Q = 15,750/0.488 = 32,275; \)
this implies \( P = 15,750 - 0.244 \times 3,2275 = 7,875. \)