

Bus 274: Further Statistics for Business

Spring 2010

PROBLEM SHEET # 10

Problem 1: File `IBM_internet_study_data.xls` contains data recorded in an international study¹ about trust and the Internet. The meaning of the variables is:

country:	country
IP1:	percent of households with Internet access
IP2:	Internet subscribers per 100
trust:	percent trust
income:	per capita income in dollars (1,000s)
price:	average Internet access price in dollars
age:	percent population 60 and older
edu:	average years of education
urban:	percent urban
lines:	phone lines per 1000
PCs:	PCs per 1000 population

Here, **trust** means the percentage of respondents who answered that people could be trusted to the question

“Generally speaking, would you say that most people can be trusted, or that you can’t be too careful in dealing with people?”

The following *preliminary* analysis uses only two variables: **trust** and **IP1**.

- Draw a scatterplot with **trust** on the abscissa axis and **IP1** on the ordinate axis.
- Compute the correlation between the two variables.
- Compute the regression line of **IP1** w.r.t. **trust**.
- Compute and interpret the coefficient of determination.
- Suppose a country was found to have a **trust** value of 51. How would you predict this country’s **IP1** value? (Use a point prediction.)

Problem 2: File `speed_dist.xls` contains a famous dataset from the 1920s with 50 observations on two variables:

- **speed**, the top speed (in mph) a car could travel,
 - **dist**, the stopping distance (in ft) of the car (from its top speed).
- Draw a scatterplot with **speed** on the abscissa axis and **dist** on the ordinate axis.
 - Compute the regression line of **dist** w.r.t. **speed** and interpret your result. (What is its coefficient of determination etc.)
 - Compute the regression line of **dist** w.r.t. *squared speed* (this is still a simple linear regression, because it is linear in the regression coefficient) and interpret your result. Which regression you computed is better?

¹Huang H, Keser C, Leland J, Shachat J (2003): Trust, the Internet, and the digital divide. IBM Systems Journal 42, 507–518.