# ASSIGNMENT 2

Instruction:

Following are the questions from the Chapter 2 Exercise.

This time the plots required for any linear regression question should be graphed either using R or MS

Dataset required for the problems are attached in the folder.

## MEIS dataset

**2.152 Dwelling permits and sales for 21 European countries.**

The Organisation for Economic Co-operation and Development collects data on Main Economic

Indicators (MEIs) for many countries. Each variable is recorded as an index with the year 2000

serving as a base year. This means that the variable for each year is reported as a ratio of the value for the year divided by the value for 2000. Use of indices in this way makes it easier to compare values for different countries.

**MEIS**

(a) Make a scatterplot with sales as the response variable and permits issued for new dwellings as the explanatory variable. Describe the relationship. Are there any outliers or influential observations?

(b) Find the least-squares regression line and add it to your plot.

(c) What is the predicted value of sales for a country that has an index of 160 for dwelling permits?

(d) The Netherlands has an index of 160 for dwelling permits. Find the residual for this country.

(e) What percent of the variation in sales is explained by dwelling permits?

**2.153 Dwelling permits and production.**

Refer to the previous exercise. **MEIS**

(a) Make a scatterplot with production as the response variable and permits issued for new dwellings as the explanatory variable. Describe the relationship. Are there any outliers or influential observations?

(b) Find the least-squares regression line and add it to your plot.

(c) What is the predicted value of production for a country that has an index of 160 for dwelling

permits?

(d) The Netherlands has an index of 160 for dwelling permits. Find the residual for this country.

(e) What percent of the variation in production is explained by dwelling permits? How does this

value compare with the value that you found in the previous exercise for the percent of variation in sales that is explained by building permits.

**2.154 Sales and production.**

Refer to the previous two exercises. **MEIS**

(a) Make a scatterplot with sales as the response variable and production as the explanatory variable. Describe the relationship. Are there any outliers or influential observations?

(b) Find the least-squares regression line and add it to your plot.

(c) What is the predicted value of sales for a country that has an index of 125 for production?

(d) Finland has an index of 125 for production. Find the residual for this country.

(e) What percent of the variation in sales is explained by production? How does this value compare with the percents of variation that you calculated in the two previous exercises?

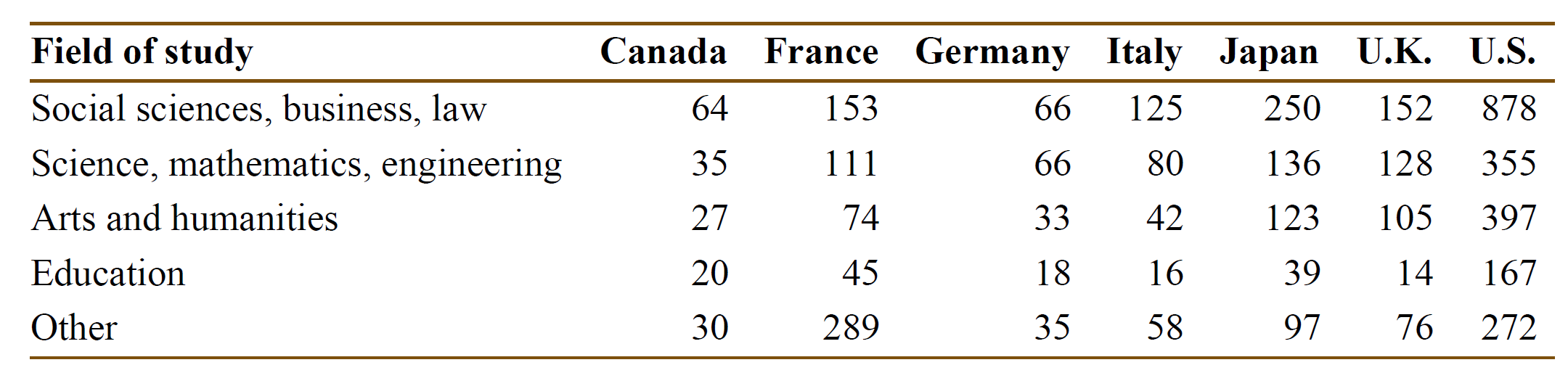
## FOS Dataset

**2.157 Fields of study for college students.**

The following table gives the number of students (in thousands) graduating from college with

degrees in several fields of study for seven countries:39

**Field of study Canada France Germany Italy Japan U.K. U.S.**



(a) Calculate the marginal totals and add them to the table.

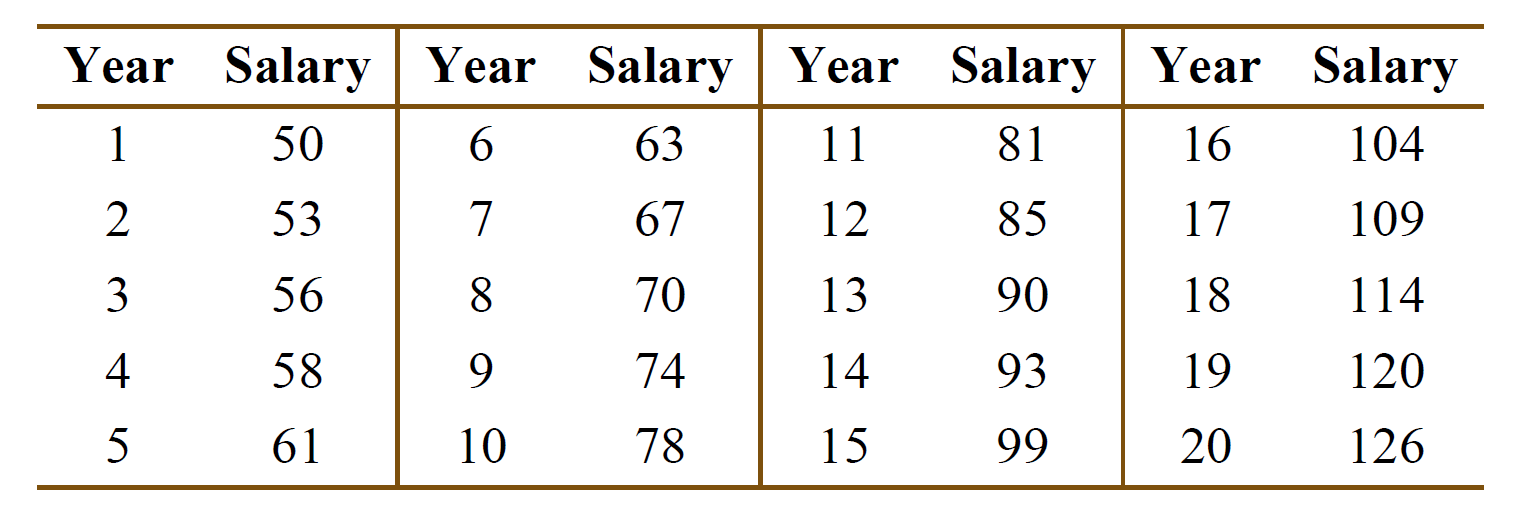
(b) Find the marginal distribution of country and give a graphical display of the distribution.

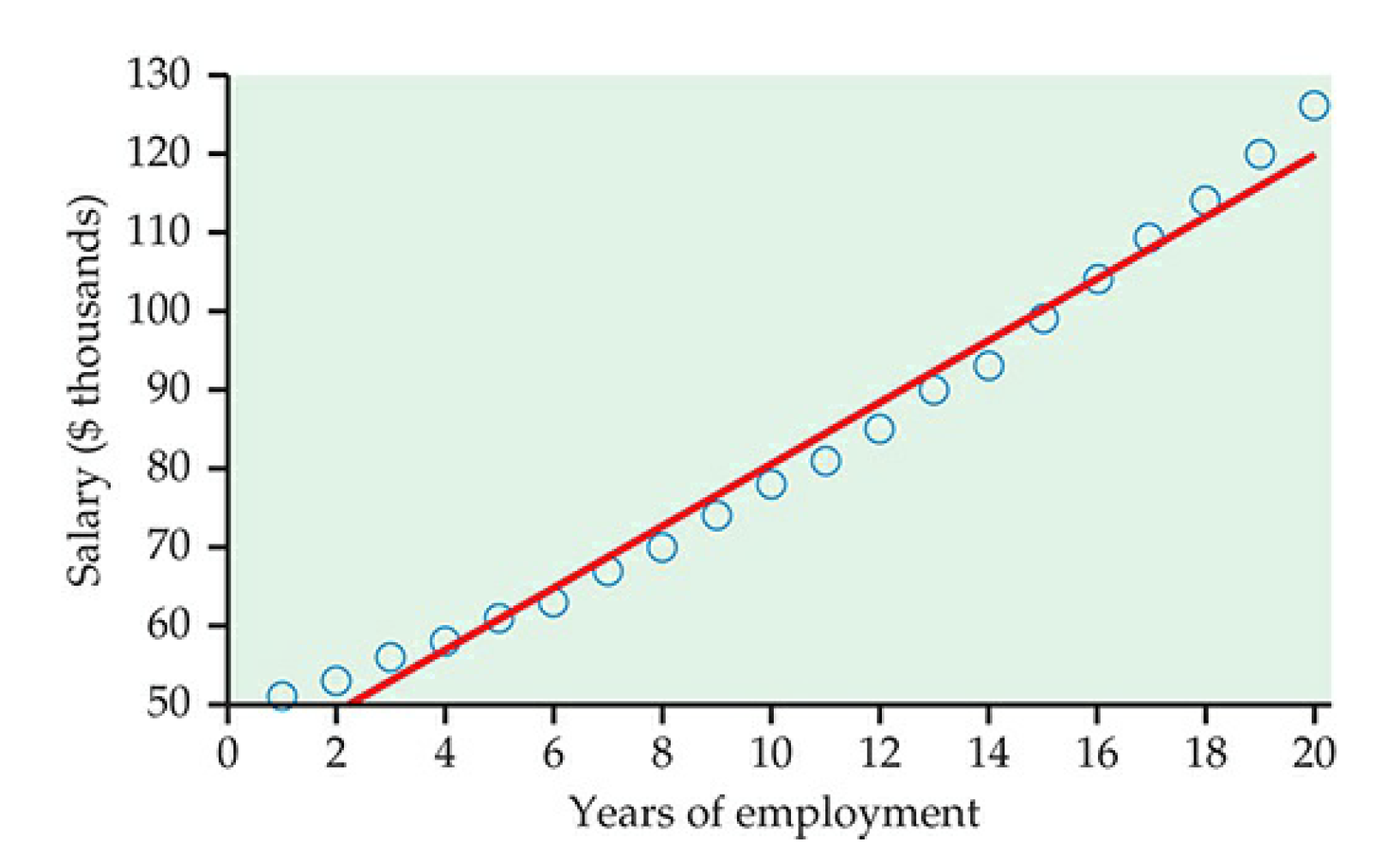
(c) Do the same for the marginal distribution of field of study.

**2.162 Salaries and raises.**

For this exercise we consider a hypothetical employee who starts working in Year 1 with a salary of $50,000. Each year her salary increases by approximately 5%. By Year 20, she is earning $126,000.

The following table gives her salary for each year (in thousands of dollars): **RAISES**





**FIGURE 2.34**

Plot of salary versus year for an individual who receives approximately a 5% raise each

year for 20 years, with the least-squares regression line, for Exercise 2.162.

(a) Figure 2.34 is a scatterplot of salary versus year, with the least-squares regression line. Describe the relationship between salary and year for this person.

(b) The value of *r*2 for these data is 0.9832. What percent of the variation in salary is explained by year? Would you say that this is an indication of a strong linear relationship? Explain your answer.

**2.163 Look at the residuals.**

Refer to the previous exercise.

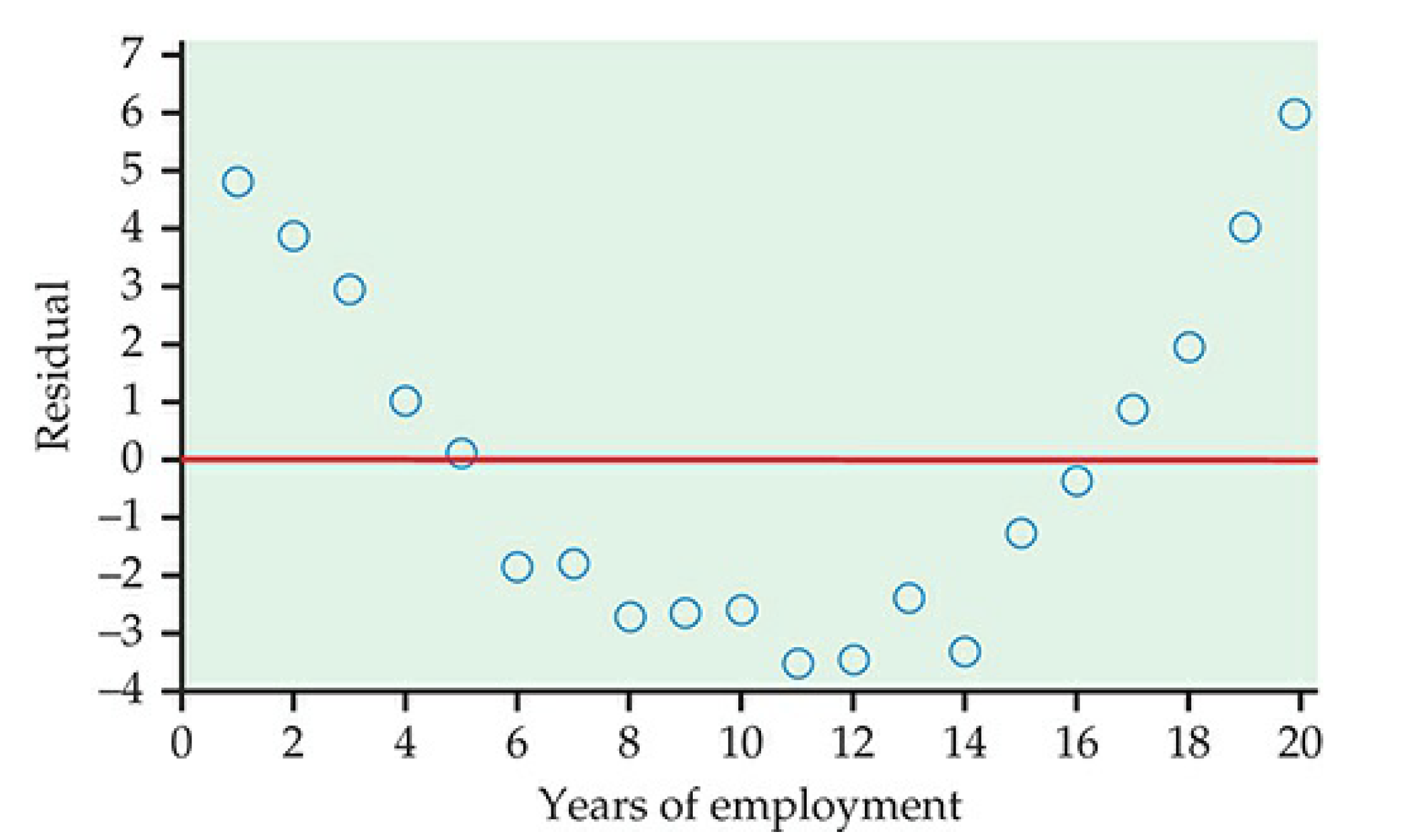
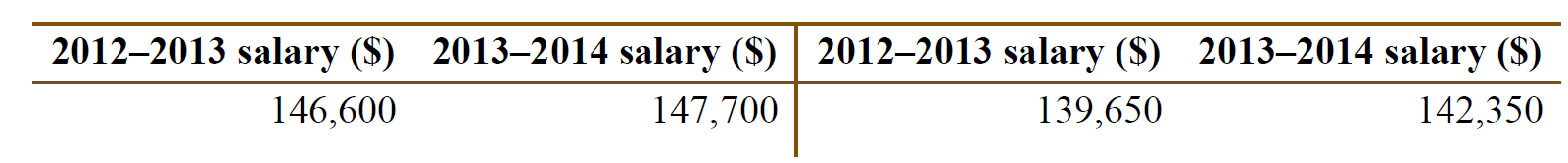


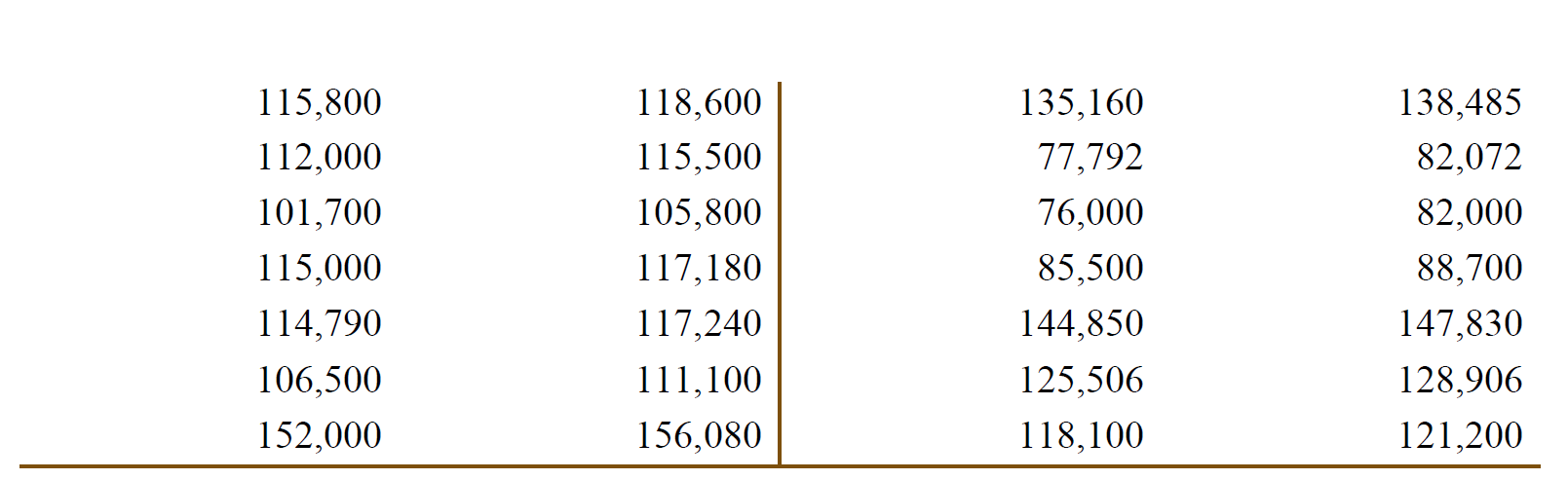
Figure 2.35 is a plot of the residuals versus year. **RAISES**

(a) Interpret the residual plot.

(b) Explain how this plot highlights the deviations from the least-squares regression line that you can see in Figure 2.34.

**2.166 Faculty salaries.**

Here are the salaries for a sample of professors in a mathematics department at a large midwestern university for the academic years 2012–2013 and 2013–2014. **FACULTY**



(a) Construct a scatterplot with the 2013–2014 salaries on the vertical axis and the 2012–2013

salaries on the horizontal axis.

(b) Comment on the form, direction, and strength of the relationship in your scatterplot.

(c) What proportion of the variation in 2013–2014 salaries is explained by 2012–2013 salaries?

**2.177 Class size and class level.**

A university classifies its classes as either “small” (fewer than 40 students) or “large.” A dean sees that 62% of Department A’s classes are small, while Department B has only 40% small classes. She wonders if she should cut Department A’s budget and insist on larger classes. Department A responds to the dean by pointing out that classes for third- and fourth-year students tend to be smaller than classes for first- and second-year students. The following three-way table gives the counts of classes by department, size, and student audience. Write a short report for the dean that summarizes these data. Start by computing the percents of small classes in the two departments and include other numerical and graphical comparisons as needed. Here are the numbers of classes to be analyzed: **CSIZE**

