Advances in Social Science, Education and Humanities Research, volume 176

2nd International Conference on Management, Education and Social Science (ICMESS 2018)

Research on Project Teaching of Statistics Course Based on Data Visualization

Zhang Bin
School of Education, Hubei University for Nationalities

***Abstract*—With the wide applications of big data, in order to improve the systematization and practicability of statistics courses, and to better cultivate applied and comprehensive talents, this paper discusses the harmonious and unified relationship between project teaching and data visualization. Taking mathematical statistics as an example, the project teaching mode of statistics courses is reconstructed based on data visualization approach. Finally, we use a teaching example to illustrate the importance of visualization in project teaching research.**

Keywords—Statistics courses; Project teaching; Data visualization

I. INTRODUCTION

With the arrival of Big Data era, big data processing methods are widely applied in various industries and fields, such as Internet, economy, finance, e-commerce and logistics distribution. The increasing demand for big data processing talents needed by these industries has brought new opportunities and challenges for the cultivation of talents in universities. The reform of all professional personnel training modes is imminent, and in particular the reform of statistical courses has become an urgent task [1]. At present, there are mainly two teaching modes of statistical courses in China. One is teacher-centered, which is based on classroom teaching or case teaching. The advantage of this teaching mode is that the knowledge system is complete, but it is quite different from the actual application. The other is student-centered, which refers to in the form of grouping to analyze the report. One of the advantages of this teaching mode is that it can stimulate students' subjective initiative and fully mobilize their enthusiasm. However, it weakens students’ in-depth understanding of the main theoretical knowledge. Most students understand the how, not the why, so that they usually become puzzled in coping with the complex data. In order to cultivate applied talents, whether a statistics course is in statistical major or not, it is necessary to change the teaching philosophy and training mode.

Based on the above situation, this paper builds a teacher-oriented and student-centered project teaching mode from the perspective of the talent training model. By introducing data visualization methods, teachers can more effectively and efficiently conduct statistical teaching. Meanwhile, students can process data more systematically and accurately by using statistical methods.

II. PROJECT TEACHING AND DATA VISUALIZATION

Project teaching method refers to teaching activities that are conducted through joint research on the complete project work. The earliest project teaching method can be traced back to European work-study education in the 18th century, and later it developed into cooperative education. By the mid to late 20th century, it gradually became a mode of cultivating practical talents. Its core is that teachers reorganize the textbook knowledge according to the project form, and students complete relevant projects or tasks after class according to their interests so as to stimulate students' interest in learning and improve their ability to solve problems systematically. Data visualization is by using graphical means to convey and display data information clearly and effectively. The basic idea is to represent each data item in the database as a single graphical element. A large number of data sets constitute the data image, and at the same time, the various attribute values of the data are expressed in the form of multidimensional data. By observing the data from different dimensions, more in-depth observation and analysis of the data can be obtained [2-3].

The statistical course takes the collection and processing of actual data as the main goal, by using statistical methods to understand the laws behind the data. It can not be separated from the actual application background, but also need the image and popular way to show the results of the study. In statistical course teaching, it is necessary to make project the basic form, supplemented by data visualization tools to present the conclusions [4-5]. Therefore, these two methods are naturally unified in the statistical course and cannot be separated from each other. The lack of project is out of the actual application background. The lack of visualization makes the results abstract and difficult to understand. For example, when studying the factors affecting the income of rural residents, a unitary linear regression is used to conduct statistical modelling of rural household income and its influencing factors (education expenditure or health care expenditure, etc.). Firstly, the linear correlation coefficient between the explanatory variables and the income of rural residents needs to be calculated. The linear correlation coefficient is a figure between -1 and 1, and its value closer to -1 or 1 indicates more linear correlation, while closer to 0 shows less linear correlation. However, the middle number, such as 0.68, it is difficult to explain how the linear correlation is. At that time, the trend of the scatter plot of these two variables can clearly show the linear relationship between them, and it also benefits data preprocessing. For example, if the scatter plot of the two variables shows an exponential growth trend, the logarithm transformation can make the transformed data more linear. Secondly, the residual or goodness of fit of linear regression needs to be calculated. As the specific value is difficult to reflect the effect of linear regression, the residual plot is used to reflect the fitting effect of the data, and the regression equation is needed to visually reflect the relationship between explanatory variables and response variables [6]. The above example shows that data visualization can assist statistical teaching very well. Therefore, it is necessary to build a project teaching mode based on data visualization.

III. CONSTRUCTING STATISTICS COURSE PROJECT TEACHING MODE BASED ON DATA VISUALIZATION

Statistics courses are a general term for the application of statistical methods to various professional issues, including statistics, economic statistics, education statistics, etc. Due to the wide application of statistical methods and the different backgrounds of these courses, the statistical methods used are very different. As long as it involves the use of statistical methods, it is indispensable to deal with all kinds of data, and it will inevitably inseparable from the intuitive understanding of data visualization [7-8]. Constructing statistical course project teaching mode based on data visualization is divided into three aspects. The first aspect is teaching research. According to the specific requirements of talents training, the knowledge points are re-sorted, classified and summarized, and the series and organic combination of knowledge points are realized by introducing specific background and data [9]. The second aspect is the implementation of teaching. It refers to explain the main theories and knowledge points based on the project, including the introduction of project background and data, explaining statistical theory and methods, data visualization and conclusion analysis. The third aspect is student learning. First of all, students need to learn the statistical theory and methods involved in the project and its scope of application, and then complete project task in the form of grouping. Figure 1 shows the structure of the project.

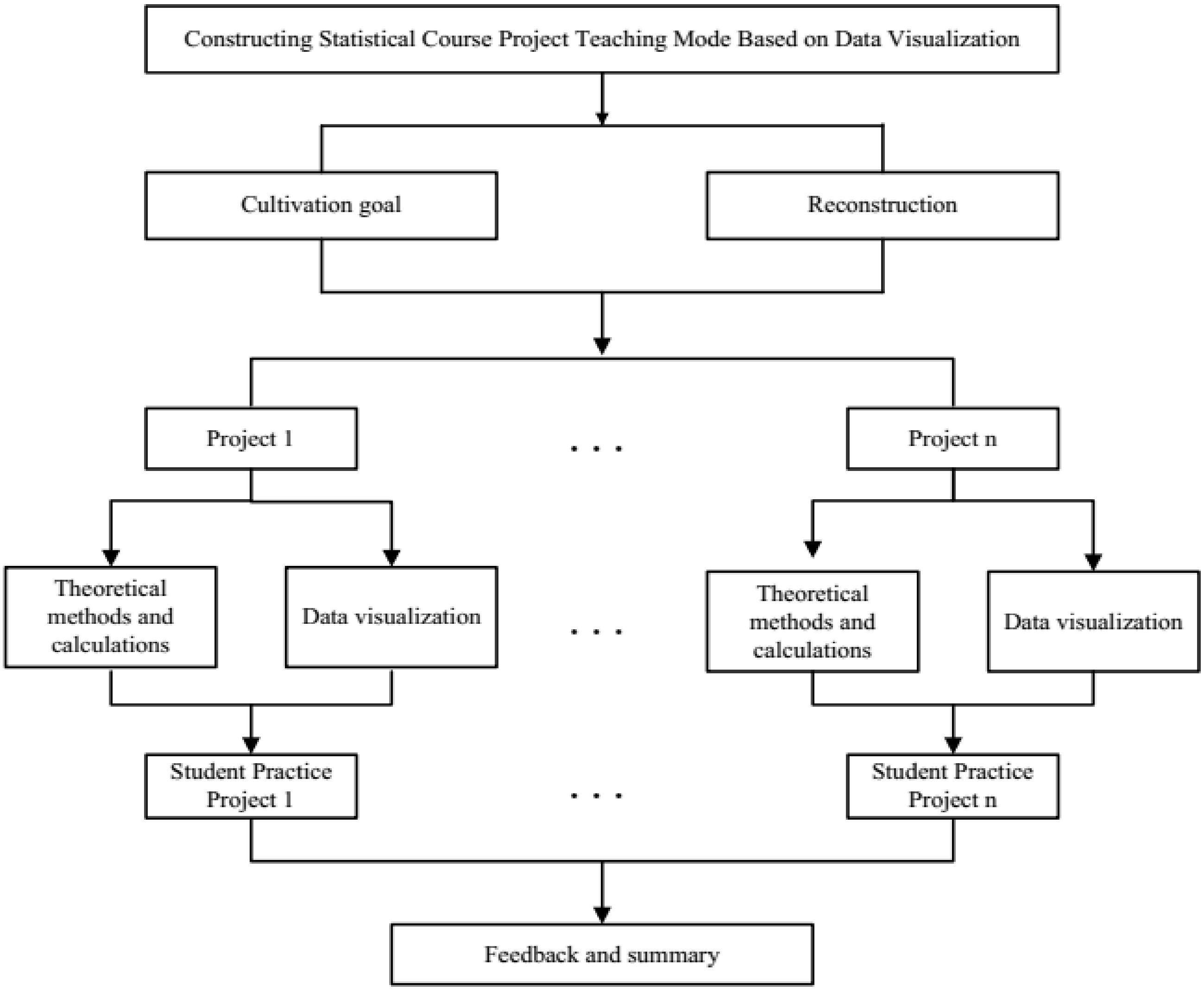
Constructing Statistical Course Project Teaching Mode Based on Data Visualization
Cultivation goal
Reconstruction
Project 1
. . .
Project n
Theoretical methods and calculations
Data visualization
. . .
Theoretical methods and calculations
Data visualization
Student Practice Project 1
. . .
Student Practice Project n
Feedback and summary

Fig. 1. Constructing Statistical Course Project Teaching Mode Based on Data Visualization

According to the project teaching requirements, each course needs to be reconstructed based on its knowledge system. According to the requirements of talents training specifications, various knowledge points need to be scientifically reclassified and synthesized, and the concatenation and mastery of various knowledge points need to be achieved through the introduction of appropriate projects. (See Figure 2).

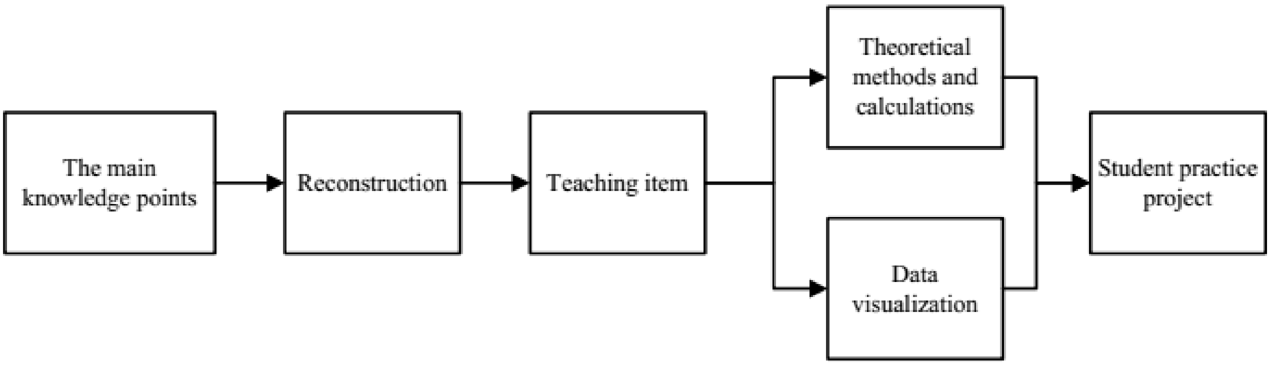

Fig. 2. Project teaching process based on data visualization

The following uses statistics as an example to build a project teaching mode based on data visualization.

A. *The main knowledge points (in order)*

Ensemble, sample, empirical distribution function, statistics, sampling distribution, mean value, variance, three major sampling distributions, point estimation, moment estimator, maximum likelihood estimation, consistency, unbiasedness, validity, CR inequality, CR lower bound, interval estimation, hypothesis test basic idea, significance test, linear regression, least squares estimation, one-way analysis of variance, parameter estimation

B. *Reconstruction of Knowledge Points Based on Project*

Unit 1: population, population moment, sample, sampling distribution; Unit 2: empirical distribution function, statistics, sample moments

Unit 3: moment estimation, maximum likelihood estimation, consistency, unbiasedness, validity, CR inequality; Unit 4: significance test, rejection domain, p-value, interval estimation; Unit 5: linear regression, least squares estimation; Unit 6: variance analysis, parameter estimation

C. *Composition of teaching items*

Item 1: distribution function and its feature (background and data: generated by R simulation)

Theoretical methods and calculations: give the ensemble and calculate ensemble average, variance, kurtosis and skewness;

Data visualization: Use R to simulate the density function map of the distribution; Simulate the density function of the three major sampling distributions and mark the commonly used quantiles;

Students practice project: Generate a random number for the specified distribution by simulation, thereby realizing visualization.

Item 2: Important statistics and its distribution functions (background and data: A student’s grade in a course)

Theoretical methods and calculations: Calculate sample mean, variance, and other statistics; find the distribution of common statistics based on the ensemble distribution;

Data visualization: Draw a histogram of the sample, mark the location of the sample mean, plot the empirical distribution function curve, and simulate the distribution function and density function curve of the important statistics such as the mean value;

Students practice project: Respectively calculate sample mean, variance, and other statistics using the heights of the class boys and girls as samples, thereby realizing visualization.

Item 3: Parameter estimation method and its statistical properties (background and data: generated by R simulation)

Theoretical methods and calculations: Moment estimates and maximum likelihood estimates of unknown parameters based on samples, and describes its statistical properties;

Data visualization: Simulate point estimate density function curve to verify statistical properties such as unbiasedness and consistency;

Student practice project: Calculate the moment and maximum likelihood estimates of unknown parameters in a given distribution, thereby realizing visualization.

Item 4: Hypothesis Testing (Background and Data: Testing Parameters in a Product Design)

Theoretical methods and calculations: According to the actual situation, test statistics are selected and the distribution and quantile points are calculated. Based on this, the rejection domain and p-value of the hypothesis test are calculated and then the interval estimate of unknown parameters is calculated.

Data visualization: Simulate the generation of rejection fields of unilateral and bilateral inspection, mark the location of p-values, and label the interval estimates of the parameters;

Student Practice Project: Verify that the mean and variance of a product are consistent with the design parameters, thereby realizing visualization.

Item 5: Linear Regression Model (Background and Data: Car Data and Resident Income Data)

Theoretical methods and calculations: Establish a linear regression model, use least squares method to estimate the parameters and calculate the sum of residuals and residual sum;

Data visualization: Draw scatter plots and regression lines between variables, and plot residuals and goodness of fit for each sample;

Student Practice Project: Multiple linear regression modeling of air quality and its influencing factors, thereby realizing visualization.

Item 6: Analysis of variance (background and data: China's GDP data)

Theoretical methods and calculations: Calculate the sum of squared deviations, the sum of squares between groups, the sum of squares within a group, and its degrees of freedom. Calculate the value of the F-test statistic and compare the quantile points of the F distribution.

Data visualization: Draw the density function graph of the F-test statistic, and mark the location of the statistic value and rejection domain;

Student practice project: Analyze the variance of the province's GDP data, thereby realizing visualization.

IV. Teaching examples

The following is an example of the item 5 in the statistics course, specifically expounding the importance of data visualization in project teaching.

Example: Linear Regression Model (real data)

Car data is taken from R's basic package database and the data is called cars. It records the speed and braking distance (dist) of 50 cars in the 1920s [10-11]. Here we try to establish a statistical model of this data.

Solution: First, use the scatter plot to analyze the correlation between vehicle speed and braking distance, as shown in the following figure:

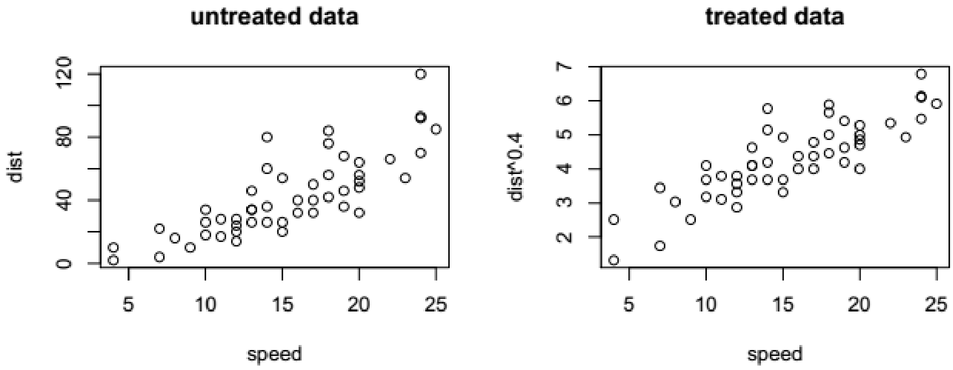

Fig. 3. Scatter plot of raw data and transformed data

The left figure of FIG. 3 shows the scatter plot of braking distance (dist) and speed. It can be found that the change trend of the two variables is in a convex model. For this purpose, the one-to-one power transformation of the braking distance of the response variable is as follows: $y = dist^{0.4}, x = speed$. The right figure shows the scatter plot of y and x, and it looks more linear. Therefore, the following one-dimensional linear regression model is established, where y is response variable and x is explanatory variable: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

The least squared estimate of the coefficient β_0, β_1 is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n, \bar{y} = \sum_{i=1}^{n} y_i / n,$

Calculated that: $\hat{\beta}_0 = 1.48, \hat{\beta}_1 = 0.18,$ thus the linear model is $y = 1.48 + 0.18x$.

The goodness of fit is R^2 = 0.7132.

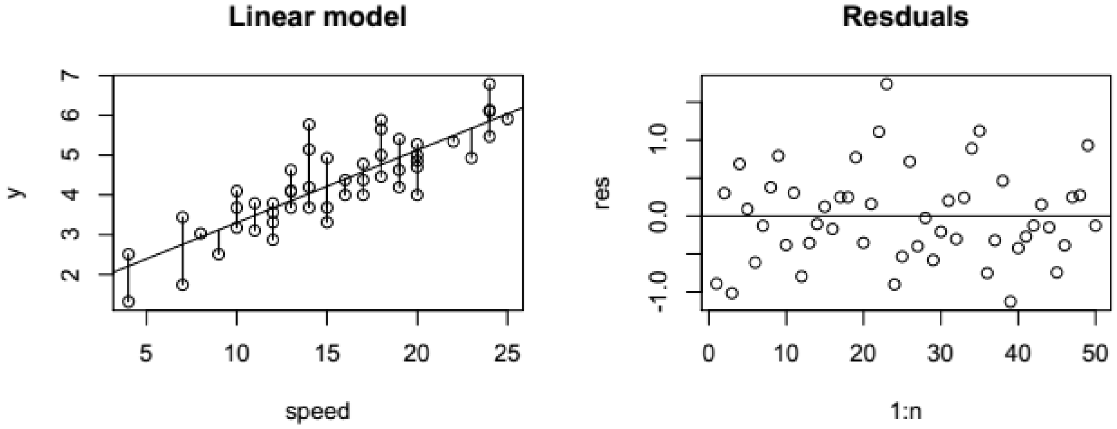

Fig. 4. Linear Model and Fitted Residuals

This example is a good blend of the features of traditional teaching and visual teaching, if not using visualization, it is difficult to find the correlation between speed and braking distance in the original data, so as to make a reasonable transformation. Visualization makes obscure formulas more intuitive and vivid. It can inspire students' interest and imagination and improve classroom teaching quality.

V. Conclusion

The paper explores the relationship between project teaching and data visualization in the teaching of statistical courses. Starting from the talent training model, the project teaching mode is reconstructed based on data visualization. The two projects in the mathematical statistics course are used as examples to demonstrate the basic process of the teaching mode. The actual teaching effect shows that through the reconstruction of various knowledge points, the teaching model proposed in this paper can better balance the theoretical knowledge in the textbook and the data processing in practical problems. It can more vividly show the statistical theories and main methods and can better stimulate students' interest in learning and using statistical methods, so that it will better cultivate applied and comprehensive talents.

Acknowledgment

This work was supported in part by Young fund of Hubei University for Nationalities MY2017Q021.

References

[1] Jiang Shiquan, Liu Zhongxia. Investigation and Experimental Research on Teaching Statistics of Management Statistics in the Perspective of Big Data. Journal of Tongling College, 2017(2), pp. 114-117.

[2] Duan Lian. Research and Practice of Visualization Teaching in Advanced Mathematics. Journal of Jiaxing University, 2012, 24(3), pp. 134-138.

[3] Li Liangshu. A Brief Discussion on the Project Teaching of Higher Vocational "Statistics" Course [J]. Mathematics Learning and Research, 2013(13), pp. 3-4.

[4] Wang Fei, Liang Jiwen. Analysis of project achievements based on statistics from the National Social Science Foundation [J]. Journal of Southwest University for Nationalities (Humanities and Social Sciences), 2017, 38(9), pp. 230-234.

[5] Wei Lili. Discussion on the Teaching Reform of Public Security Statistics Based on Project Teaching Method [J]. Journal of Shanxi Police Academy, 2016, 24(1), pp. 90-93.

[6] Huang Jicong, Lin Shuisheng, Zhang Meihua. Research on the Project Teaching of Higher Vocational Statistics Based on Data Analysis Ability Cultivation [J]. Journal of Beijing Institute of Finance and Trade, 2017, 33(3), pp. 35-37.

[7] LIU Yanqin. Application of Visualization Teaching in the Research Teaching of University Mathematics [J]. Journal of Yulin University, 2012, 22(2), pp. 55-57.

[8] Guo Guoan. Innovative application of visual teaching in advanced mathematics education [J]. Teaching and Education: Higher Education Forum, 2015(30), pp. 58-59.

[9] Fu Zhe. Project-driven teaching research oriented to the whole process: Taking "Statistics" as an example [J]. Education and Teaching Forum, 2015(11), pp.179-180.

[10] Tang Ling, Yang Shuyuan. Application of project-based teaching in teaching of applied colleges and universities——Taking "Basic Statistics" as an example [J]. Journal of Value Engineering, 2016, 35(13), pp. 231-233.

[11] SUN Peng. Research on Teaching Reform of Statistics Course [J]. Quality Education in the West, 2017, 3(10), pp.126-127.