

Bài tập Hồi quy bội

13/10/2022

1. Oil extraction from peanut data

Pressure (bars)	Temp. (°C)	Moisture (% by weight)	Flow Rate (L/min)	Particle Size (mm)	Yield
415	25	5	40	1.28	63
550	25	5	40	4.05	21
415	95	5	40	4.05	36
550	95	5	40	1.28	99
415	25	15	40	4.05	24
550	25	15	40	1.28	66
415	95	15	40	1.28	71
550	95	15	40	4.05	54
415	25	5	60	4.05	23
550	25	5	60	1.28	74
415	95	5	60	1.28	80
550	95	5	60	4.05	33
415	25	15	60	1.28	63
550	25	15	60	4.05	21
415	95	15	60	4.05	44
550	95	15	60	1.28	96

Source: “An Application of Fractional Experimental Designs,” by M. B. Kilgo, *Quality Engineering*, **1**, pp. 19–23.

An engineer performed an experiment to determine the effect of CO₂ pressure, CO₂ temperature, peanut moisture, CO₂ flow rate, and peanut particle size on the total yield of oil per batch of peanuts. Table above summarizes the experimental results.

- **a.** Fit a multiple linear regression model relating yield to these regressors.
- **b.** Test for significance of regression. What conclusions can you draw?
- **c.** Use t tests to assess the contribution of each regressor to the model. Discuss your findings
- **d.** Calculate R^2 and R_{Adj}^2 for this model. Compare these values to the R^2 and R_{Adj}^2 for the multiple linear regression model relating yield to temperature and particle size. Discuss your results.

2. Life expectancy

- Rossman [1994] presents an interesting study of average life expectancy of 40 countries. Sheet “LifeExp” gives the data. The study has three responses: LifeExp is the overall average life expectancy. LifeExpMale is the average life expectancy for males, and LifeExpFemale is the average life expectancy for females. The regressors are People - per - TV, which is the average number of people per television, and People - per - Dr, which is the average number of people per physician.
 - a. Fit different multiple linear regression models for each response.
 - b. Test each model for significance of regression. What conclusions can you draw?
 - c. Use t tests to assess the contribution of each regressor to each model. Discuss your findings.
 - d. Calculate R^2 and R_{Adj}^2 for each model.
 - e. Find a 95% CI for the regression coefficient for People - per - Dr in each model.
 - f. Create an indicator variable for gender. Perform a thorough analysis of the overall average life expectancy. Discuss the results of this analysis relative to your previous analyses of these data.

3. Patient Satisfaction Data

The sheet “PatientSatisfaction” contains some of the data that have been collected on a random sample of 25 recently discharged patients. The response variable is satisfaction, a subjective response measure on an increasing scale. The potential regressor variables are patient age, severity (an index measuring the severity of the patient’s illness), an indicator of whether the patient is a surgical or medical patient (0 = surgical, 1 = medical), and an index measuring the patient’s anxiety level.

- Fit an appropriate regression model to the satisfaction response using age and severity as the regressors and account for the medical versus surgical classification of each patient with an indicator variable. Has adding the indicator variable improved the model? Is there any evidence to support a claim that medical and surgical patients differ in their satisfaction?

4. Young Wine quality

- Consider the wine quality of young red wines data in the sheet “youngwine”. The winemakers believe that the sulfur content has a negative impact on the taste (thus, the overall quality) of the wine. Perform a thorough analysis of these data. Do the data support the winemakers ’ belief?
- For the purposes of this exercise, ignore regressor x_1 . Perform a thorough analysis of these data. What conclusions do you draw from this analysis?

5. Wine quality

The quality of Pinot Noir wine is thought to be related to the properties of clarity, aroma, body, flavor, and oakiness. Data for 38 wines are given in the sheet “wine quality”.

- **a.** Fit a multiple linear regression model relating wine quality to these regressors.
- **b.** Test for significance of regression. What conclusions can you draw?
- **c.** Use t tests to assess the contribution of each regressor to the model. Discuss your findings.
- **d.** Calculate R^2 and R^2_{Adj} for this model. Compare these values to the R^2 and R^2_{Adj} for the linear regression model relating wine quality to aroma and flavor. Discuss your results.
- **e.** Find a 95 % CI for the regression coefficient for flavor for both models in part d. Discuss any differences.
- **f.** Fit a regression model using clarity, aroma, body, flavor, and oakiness as the regressors. Investigate this model for influential observations and comment on your findings.