IT2120 - Probability and Statistics

Department of Information Technology, Faculty of Computing

Year 2 semester 1 (2025)

Tutorial 11

1. A university wants to predict student GPA (0-4 scale) based on various study habits and background factors. Factors they considered were weekly study hours, class attendance percentage, SAT_Score and read hours (weekly hours spent reading non-course materials). They have randomly selected 200 students for the study. Following is the R output for the multiple linear regression model.

```
Call:
lm(formula = GPA ~ Study_Hours + Attendance + SAT_Score + Read_Hours,
    data = academic_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.88234 -0.19834 -0.00983  0.20147  0.91873

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.998e+00  3.456e-01   5.781 3.08e-08 ***
Study_Hours  7.954e-02  8.912e-03   8.925  < 2e-16 ***
Attendance   2.012e-02  3.456e-03   5.821 2.49e-08 ***
SAT_Score    9.876e-04  1.234e-04   8.003 1.67e-13 ***
Read_Hours   4.892e-03  1.234e-02   0.396    0.692
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2987 on 195 degrees of freedom
Multiple R-squared:  0.6248,    Adjusted R-squared:  0.6171
F-statistic: 81.08 on 4 and 195 DF,  p-value: < 2.2e-16
```

   i. Write the equation for the fitted regression model.
   ii. Are all the regression coefficients significant in the fitted model? Justify your answer.
   iii. What is the coefficient of determination? Interpret the value.
   iv. Is the fitted regression model significant? Justify your answer.

2. A financial analyst wants to understand the drivers of company profitability (ROA: Return on Assets(%) - measure of profitability). The model uses following factors:

   - **R_D_Spend:** Research & Development spending (in millions)
   - **Employee_Productivity:** Revenue per employee (in thousands)
   - **Debt_Ratio:** Total debt to total assets ratio
   - **Asset_Turnover:** Sales revenue divided by total assets

   Data was collected from 150 companies. Following are the R outputs for the regression model:

```
Call:
lm(formula = ROA ~ R_D_Spend + Employee_Productivity + Debt_Ratio +
    Asset_Turnover, data = company_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1234 -0.9567 -0.0789  0.9234  4.0567

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           7.89234    1.23456   6.394 1.23e-09 ***
R_D_Spend             0.24891    0.02345  10.614  < 2e-16 ***
Employee_Productivity 0.07923    0.00789  10.042  < 2e-16 ***
Debt_Ratio          -12.45678    1.56789  -7.945 9.87e-13 ***
Asset_Turnover        4.18901    0.45678   9.171  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.467 on 145 degrees of freedom
Multiple R-squared:  0.8123,    Adjusted R-squared:  0.8071
F-statistic: 156.7 on 4 and 145 DF,  p-value: < 2.2e-16
```

   i. Write the equation for the fitted regression model.
   ii. Are all the regression coefficients significant in the fitted model? Justify your answer.
   iii. What is the coefficient of determination? Interpret the value.
   iv. Is the fitted regression model significant? Justify your answer.