**ISOM 2600 In-class Exercise (Topic 5)**

Employee retention is becoming a big issue for organizations. Recruiting a new employee can cost more than keeping an existing worker, including the money and time cost of interview process, training, job errors, etc.

If a company can identify whether an employee is going to resign, they can implement retention program to prevent it from happening.

In this exercise, you are given a dataset for human resource analysis, with the following variables:

Predictors:

* Gender: Dummy, 1 for male, 0 for female
* DaysLateLast30: Times the employee was late to work in last 30 days.
* Salary: Yearly salary (US$)
* Absences: Counts of absences
* WorkingPeriod: Days since was hired
* PerfScore: Performance score by manager.

Response variable: *EngagementSurvey*

For the dataset, the company provides a list of existing employees with basic information. They also hired an external company to conduct an engagement survey. As part of the retention process, the company would like to know

1. what factor might affect the employees’ satisfaction?

2. a regression model to predict staff’s engagement.

Question 1

They run a simple regression model with the predictor “Salary”. Here is the regression output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 4.0038 | 0.149 | 26.852 | 0.000 | 3.710 | 4.298 |
| **Salary** | 1.642e-06 | 1.96e-06 | 0.836 | 0.404 | -2.23e-06 | 5.51e-06 |

Table

Description automatically generated

Which of the following is the fitted model?

a. Y = 4.0038 + 26.852\*Salary

b. Y = 4.0038 + 1.642\*10-6\*Salary

c. Y = 1.642\*10-6 + 4.0038\*Salary

d. Y = 1.642\*10-6 + 0.836\*Salary

Answer: B

Question 2

In the regression output, it is also given that R2 = 0.003. Which of the following description about R2 is correct?

a. R2 shows that the model is insignificant (useless) since it is lower than 0.05.

b. Adding extra predictor to the model, for example, *WorkingPeriod*, might result in a lower value of R2.

c. Only 0.3% of the variation in the response variable explained by the model.

d. Using R2 can help us to compare the performance of different models. If a model with two predictors generates a better R2, then the model is better performed than the simple regression model.

Answer: C

Question 3

Given the residual plot of the model below:

Calendar

Description automatically generated

Which assumption of residuals in linear regression is violated?

A. Normality B. Linearity

C. Constant variance D. Independence

Answer: A

Question 4

Now we consider a multiple regression model with all predictors. The regression result is given below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 3.2644 | 0.369 | 8.853 | 0.000 | 2.537 | 3.992 |
| **Gender** | -0.1269 | 0.087 | -1.452 | 0.148 | -0.299 | 0.045 |
| **DaysLateLast30** | -0.2779 | 0.059 | -4.718 | 0.000 | -0.394 | -0.162 |
| **Salary** | 2.998e-07 | 1.59e-06 | 0.189 | 0.851 | -2.84e-06 | 3.44e-06 |
| **Absences** | -0.0088 | 0.007 | -1.178 | 0.240 | -0.023 | 0.006 |
| **WorkingPeriod** | -2.11e-05 | 6.14e-05 | -0.344 | 0.731 | -0.000 | 9.99e-05 |
| **PerfScore** | 0.3722 | 0.106 | 3.507 | 0.001 | 0.163 | 0.581 |

Table

Description automatically generated

Which predictor is significant at alpha = 0.05?

A. Gender B. DaysLateLast30

C. Salary D. WorkingPeriod

Answer: B