Data Science

Home Work 2.

1. Practice the data set DS\_HW1\_R\_and\_R\_data.xlsx.

Use ANOVA to analyze the factor effects of Inspector and Part. Each treatment has 3 replicates (tests). Explain the analysis results for all the following questions, including the ANOVA table and the test results. Now the parts and inspectors were randomly sampled from two corresponding populations. Now, the 10 parts and 3 inspectors were randomly sampled from two corresponding populations. For example, there are 100 parts in the part population and 10 parts are randomly sampled. Also, there are 30 inspectors in the inspector population and 3 inspectors are randomly sampled. Use the whole data set of DS\_HW1\_R\_and\_R\_data.xlsx for the analysis.

1. Give the ANOVA table with the interaction effect.

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

1. Are Inspector and Part the significant factors for impedance? Why?

Yes, both Inspector and Part are significant factors for impedance.

For Inspector:

* F-statistic = 7.284929
* PR (p-value) = 0.00481

For Part:

* F-statistic = 162.27027
* PR (p-value) = 0.0 (approximately)

In both cases, the p-values are significantly smaller than 0.05, indicating that both Inspector and Part are highly significant factors affecting impedance. Therefore, they have a significant impact on impedance.

1. Is there an interaction effect between the Inspector and Part of the impedance? Why?

Yes, there is a significant interaction effect between Inspector and Part on impedance.

For Inspector:Part interaction effect:

* F-value = 5.272947
* p-value (PR) ≈ 0.000001 (much less than 0.05)

The extremely small p-value (<0.05) for the interaction effect indicates that the interaction between Inspector and Part has a statistically significant impact on impedance.

1. Exam the data for model assumptions.
   * Normality of Residuals use Shapiro-Wilk test

* Shapiro-Wilk Test Statistic: 0.8783404231071472
* Shapiro-Wilk p-value: 0.0000004913499651593

p-value<0.05 we reject the null hypothesis. In the case of the Shapiro-Wilk test, the null hypothesis is that the data follows a normal distribution.

一張含有 文字, 行, 螢幕擷取畫面, 繪圖 的圖片

自動產生的描述

* + Homogeneity of Variance
* Levene's Test Statistic: 0.5184678522571821
* Levene's p-value: 0.5972601919997482

p-value >0.05 we fail to reject the null hypothesis. In Levene's Test, the null hypothesis is that the variances across different groups or categories are equal (homogeneity of variances).

一張含有 文字, 螢幕擷取畫面, 行, 圖表 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 文字, 圖表, 行 的圖片

自動產生的描述

* + Independence

Durbin-Watson Test Statistic（DW-value）：0.485797608043757

This value is not significantly close to 2. In this case, we can consider that there is some degree of autocorrelation, but it is not very strong.

The answer is no Independence.

1. Estimate the variance of repeatability, reproducibility, and gauge.
   * Repeatability
   * Reproducibility
   * Gauge

*+*

1. Compare the random factor analysis to the fixed factor analysis. What are the difference?
   * **Random Factor Analysis (Mixed-Effects Model):**
     + Allows for the inclusion of random effects, which account for variability associated with the levels of the factors that are not of direct interest (e.g., specific inspectors).
     + Provides estimates of both fixed and random effects, allowing for a more comprehensive understanding of variability.
   * **Fixed Factor Analysis:**
     + Assumes that all variability in the data is due to the fixed factors being studied.
     + Does not account for potential variability associated with specific levels of the factors.

In summary, the primary distinction lies in their treatment of variability: Random Factor Analysis (RFA) explicitly models and estimates random effects, whereas Fixed Factor Analysis (FFA) assumes that all variability is caused by fixed factors. Although it may be challenging to clearly discern the differences between these two analytical approaches in this case, a thorough understanding of domain knowledge allows for the consideration of random effects, potentially avoiding lurking variables and facilitating a more comprehensive analysis.

Code：<https://colab.research.google.com/drive/1BjLPo93FPUQDUv8yYoGVqqjgQrWorkiy?usp=sharing>