Questions (2078)

* Parametric test and non-parametric tests

↓	↓
Tests used when data follows particular distribution. for eg: normal distribution. Common parametric tests include t-tests, ANOVA, linear regression, Pearson correlation.	Non-parametric tests are used when a particular distribution can't be assumed. for eg: Mann-Whitney U test, Kruskal-Wallis test, Spearman Rank Correlation

* Correlation analysis

Correlation is a statistical technique used to measure the strength and direction of relationship between two variables.

for eg: correlation ← cor(x, y, method = "pearson")

* Measures of relative position are statistical measures that describe the position of particular value within a dataset relative to the other values.
  eg: percentiles, deciles, median
  eg: quantile(x, 0.9) ⇒ 90th percentile

* Loops are utilized to execute a code block a number of times until the specified condition is met.

```
eg: for (i in 1:5) {
        print(i)
    }
```

* Functions are blocks of code that can be reused to perform specific task. Makes code more organized, clean and less redundant.

```
eg: square ← function (x) {
        return (x*2)
    }

result ← square(5)
```

* Pipes in R are used to pass the result of an (%>%) operation to a function in the right hand side as its first argument. They are used to chain together multiple operations.

```
eg: x ← c (1, 2, 3, 4)
    mean_value ← x %>% mean()
    print (mean_value)
```

(2080)

* Variance Inflation Factor (VIF) and Lasso Regression both technique are ~~both techniques~~ used in regression analysis.

VIF is a measure used to detect a multicollinearity in regression analysis.

Lasso Regression is a technique used for variable selection and regularization in regression analysis. It is equivalent to controlling VIF.

* ANOVA post-hoc tests are required when there is a significant difference in means between three or more groups.

Post-hoc tests: Fisher LSD (Every group t-test)

Tukey HSD ✓

anova_result ← aov(a~b, data=...)
tukey_test ← TukeyHSD(anova_result)

Bonferroni test: more correction

* Outliers: Outliers are data points that deviate significantly from the overall pattern of the data.

Leverage points: Leveraged points are observations that have a large influence in the estimation of regression co-efficients.

Influential observations: Influential observations are those data points that have a substantial impact on the regression analysis, often due to the combination leverage and outlying behavior.

We can identify them using diagnostic tools and techniques.