Data Analysis and Machine-Learning

Chapter 11.

ANOVA (One-Way, Two-way) Using JAMOVI



1. Introduction

Data Source for exercise: ‘Telco Customer Churn’ from Kaggle.

In particular, we will use “Total Charges” as the dependent variable, and “Payment Method” as the independent variable with four categories (‘bank transfer (automatic)’, ‘credit card (automatic)’, ‘electronic check’, and ‘mailed check’).

2. Few Additional Concepts regarding ANOVA

Applicable data types for ANOVA:

Independent: Continuous data

Dependent: Discrete data, Categorical data

ANOVA Steps

e.g., One-way ANOVA => Post Hoc Tests

**η²:**

= SS(between variance) / [SS(betweenV)+SS(withinV)] = SS(betweenV) / SS(TotalV)

The ratio of variance (with regards to the total variance of the independent variable) explained by the independent variable (similarly to the concept of R2, i.e., the explanatory power of the model). e.g., accordingly to the following ANOVA test, the ‘payment method’ variable explains about 12.3% of the variance of the dependent variable (‘total charges’).

Bonferroni Correction (P.bonferroni):

e.g., for p(0.05), significance test is adjusted as 0.05/N

3. Descriptives

Jamovi: Exploration => Descriptives

Variables: Total Charges

Split by: Payment Method

**Descriptives**

| Descriptives | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | | **PaymentMethod** | | **TotalCharges** | |
| N |  | Bank transfer (automatic) |  | 1542 |  |
|  |  | Credit card (automatic) |  | 1521 |  |
|  |  | Electronic check |  | 2365 |  |
|  |  | Mailed check |  | 1604 |  |
| Mean |  | Bank transfer (automatic) |  | 3079 |  |
|  |  | Credit card (automatic) |  | 3071 |  |
|  |  | Electronic check |  | 2091 |  |
|  |  | Mailed check |  | 1054 |  |
|  | | | | | |

4. ANOVA (One-Way)

Analyses => ANOVA

Dependent Variable: Total Charges

Fixed Factors: Payment Method

| ANOVA - TotalCharges | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | |
| PaymentMethod |  | 4.43e0+9 |  | 3 |  | 1.48e+9 |  | 328 |  | < .001 |  | 0.123 |  |
| Residuals |  | 3.17e+10 |  | 7028 |  | 4.51e+6 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |

Mean Squared (Payment Method) = Sum of Squares / df

Mean Squared (Payment Method) = Between Variance

Mean Squared (Residuals) = Within Variance

F-value = Mean-Square(Paymentmethod) / Mean-Square(Residuals) = 328

Accordingly to the following ANOVA test, the ‘payment method’ variable explains about 12.3% of the variance of the dependent variable (‘total charges’).

ANOVA result implies statistical significance (p<.001), which then requires post hoc test.

5. Post Hoc Test

In Post Hoc Tests Column, input Paymentmethod

Corrections: Tukey, Bonferroni

## Post Hoc Tests

| Post Hoc Comparisons - PaymentMethod | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Comparison** | | | | | |  | | | | | | | | | | | |
| **PaymentMethod** | |  | | **PaymentMethod** | | **Mean Difference** | | **SE** | | **df** | | **t** | | **ptukey** | | **pbonferroni** | |
| Bank transfer (automatic) |  | - |  | Credit card (automatic) |  | 7.90 |  | 76.7 |  | 7028 |  | 0.103 |  | 1.000 |  | 1.000 |  |
|  |  | - |  | Electronic check |  | 988.43 |  | 69.5 |  | 7028 |  | 14.220 |  | < .001 |  | < .001 |  |
|  |  | - |  | Mailed check |  | 2024.82 |  | 75.7 |  | 7028 |  | 26.734 |  | < .001 |  | < .001 |  |
| Credit card (automatic) |  | - |  | Electronic check |  | 980.53 |  | 69.8 |  | 7028 |  | 14.048 |  | < .001 |  | < .001 |  |
|  |  | - |  | Mailed check |  | 2016.91 |  | 76.0 |  | 7028 |  | 26.537 |  | < .001 |  | < .001 |  |
| Electronic check |  | - |  | Mailed check |  | 1036.38 |  | 68.7 |  | 7028 |  | 15.087 |  | < .001 |  | < .001 |  |
| Note. Comparisons are based on estimated marginal means | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | |

Bank Transfer variable is not statistically different to the Credit Card variable (p=1.00).

Bank Transfer is statistically different to Electronic Check and Mailed Check (<.001).

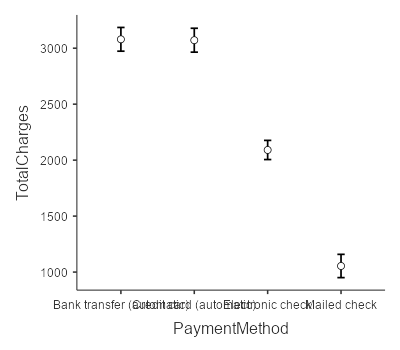
Credit Card is statistically different to Electronic Check and Mailed Check (<.001).

Electronic Check is statistically different to Mailed Check (<.001).

The result is more comprehensible with visualization, as follows.

## Estimated Marginal Means

### PaymentMethod



At first sight, it can be implied that bank transfer has virtually no difference with regards to the credit card.

6. Two-Way ANOVA

F-values for two main effects (variables: ‘Payment Method’, ‘Contract’)

F-value for interaction effect

Thus, Three F-values are to be considered.

| ANOVA - TotalCharges | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | |
| PaymentMethod |  | 3.74e0+9 |  | 3 |  | 1.25e+9 |  | 354.5 |  | < .001 |  | 0.106 |  |
| Contract |  | 6.17e0+9 |  | 2 |  | 3.09e+9 |  | 877.7 |  | < .001 |  | 0.175 |  |
| PaymentMethod ✻ Contract |  | 7.74e0+8 |  | 6 |  | 1.29e+8 |  | 36.7 |  | < .001 |  | 0.022 |  |
| Residuals |  | 2.47e+10 |  | 7020 |  | 3.52e+6 |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | |

All F-values (main effects, interaction effect) are statistically significant (<.001), which calls for post-hoc test, as follows.

7. Post Hoc Tests (for two main effects)

| Post Hoc Comparisons - PaymentMethod | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Comparison** | | | | | |  | | | | | | | | | |
| **PaymentMethod** | |  | | **PaymentMethod** | | **Mean Difference** | | **SE** | | **df** | | **t** | | **ptukey** | |
| Bank transfer (automatic) |  | - |  | Credit card (automatic) |  | 47.7 |  | 68.8 |  | 7020 |  | 0.694 |  | 0.900 |  |
|  |  | - |  | Electronic check |  | -410.4 |  | 77.6 |  | 7020 |  | -5.290 |  | < .001 |  |
|  |  | - |  | Mailed check |  | 1820.7 |  | 70.7 |  | 7020 |  | 25.752 |  | < .001 |  |
| Credit card (automatic) |  | - |  | Electronic check |  | -458.1 |  | 77.7 |  | 7020 |  | -5.897 |  | < .001 |  |
|  |  | - |  | Mailed check |  | 1773.0 |  | 70.8 |  | 7020 |  | 25.035 |  | < .001 |  |
| Electronic check |  | - |  | Mailed check |  | 2231.1 |  | 79.4 |  | 7020 |  | 28.102 |  | < .001 |  |
| Note. Comparisons are based on estimated marginal means | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | |

| Post Hoc Comparisons - Contract | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Comparison** | | | | | |  | | | | | | | | | |
| **Contract** | |  | | **Contract** | | **Mean Difference** | | **SE** | | **df** | | **t** | | **ptukey** | |
| Month-to-month |  | - |  | One year |  | -1573 |  | 59.5 |  | 7020 |  | -26.4 |  | < .001 |  |
|  |  | - |  | Two year |  | -2414 |  | 61.7 |  | 7020 |  | -39.1 |  | < .001 |  |
| One year |  | - |  | Two year |  | -841 |  | 71.2 |  | 7020 |  | -11.8 |  | < .001 |  |
| Note. Comparisons are based on estimated marginal means | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | |

Bank Transfer variable is not statistically different to the Credit Card variable (p=1.00).

Bank Transfer is statistically different to Electronic Check and Mailed Check (<.001).

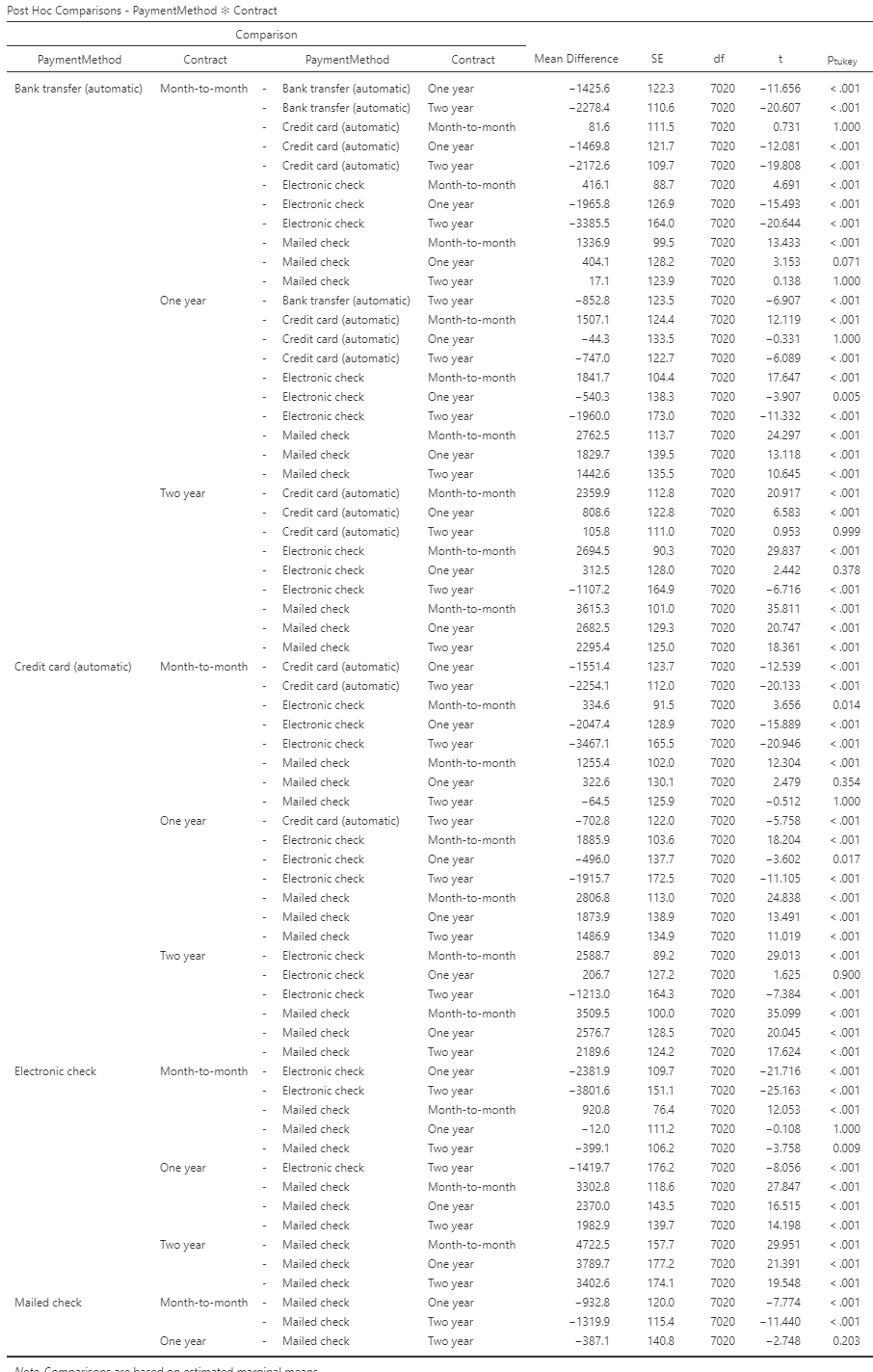
Credit Card is statistically different to Electronic Check and Mailed Check (<.001).

Electronic Check is statistically different to Mailed Check (<.001).

Month-to-month contract is statistically different to one-year contract and two-year contract.

One-year contract is statistically different to two-year contract.

8. Post Hoc Tests (for Interaction Effect)



9. Visualization

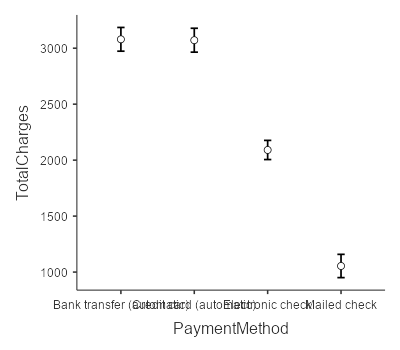
### PaymentMethod ✻ Contract



In contrast to the graph that we have seen for one-way ANOVA (below), total charges for electronic checks with one-year and two-year contracts are high. When we only considered payment method as the single variable for ANOVA, total charges for electronic check seemed rather low compared to other categories. But when considered together with another variable, contract, total charges were high when payment method was electronic check, which means that the interaction effect was significant.

## Estimated Marginal Means

### PaymentMethod



To sum up, interestingly for One-year contract and Two-year contract, total charges under electronic check is high (i.e., when considering together with the ‘contract’, people with one-year & two-year contract have high total charges with electronic checks). In other words, it can be implied that the interaction effect was significant.