**Applied Statistic – Turma 9DTSR**

Trabalho 4

**Habilidades desenvolvidas: Análise de correlação de Pearson. Teste Qui-quadrado. Regressão Linear Simples.**

**Parte 1:**

Os sistemas de compartilhamento de bicicletas são uma nova geração de aluguel de bicicletas tradicional, onde todo o processo de associação, locação e devolução tornou-se automático. Através destes sistemas, o usuário pode facilmente alugar uma bicicleta a partir de uma determinada posição e retornar em outra posição. Atualmente, existem cerca de 500 programas de compartilhamento de bicicletas em todo o mundo, compostos por mais de 500 mil bicicletas. Hoje, existe um grande interesse nesses sistemas devido ao seu importante papel no trânsito, questões ambientais e de saúde.

Identificar quais são os principais fatores que fazem as pessoas alugarem mais bicicletas, para que como organização possamos traçar o plano de logística de disponibilização, plano de manutenção e aquisição de novos ativos.

Com a base de dados disponibilizada, utilizamos as variáveis meses, dias da semana e temperatura para traçarmos o modelo estatístico para nos explicar a variável resposta cnt (quantidade total de aluguéis de bicicletas).

* Fonte de dados: <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>

1) Classifique o tipo de variável

|  |  |  |  |
| --- | --- | --- | --- |
| **Variável** | **Descrição** | **Tipo de Variável** | **Regra de data mining (Preditora ou Target)** |
| **instant** | record index | ID | - |
| **dteday** | date | ID | - |
| **season** | season (1:winter, 2:spring, 3:summer, 4:fall) | Qualitativa Ordinal | Preditora |
| **yr** | year (0: 2011, 1:2012) | Qualitativa Ordinal | Preditora |
| **mnth** | month ( 1 to 12) | Qualitativa Ordinal | Preditora |
| **holiday** | weather day is holiday or not | Qualitativa Nominal | Preditora |
| **weekday** | day of the week | Qualitativa Ordinal | Preditora |
| **workingday** | if day is neither weekend nor holiday is 1, otherwise is 0. | Qualitativa Nominal | Preditora |
| **weathersit** | 1: Clear, Few clouds, Partly cloudy, Partly cloudy; 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist;3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds;4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog | Qualitativa Nominal | Preditora |
| **temp** | Normalized temperature in Celsius. The values are derived via (t-t\_min)/(t\_max-t\_min), t\_min=-8, t\_max=+39 (only in hourly scale) | Quantitativa Continua | Preditora |
| **atemp** | Normalized feeling temperature in Celsius. The values are derived via (t-t\_min)/(t\_max-t\_min), t\_min=-16, t\_max=+50 (only in hourly scale) | Quantitativa Continua | Preditora |
| **hum** | Normalized humidity. The values are divided to 100 (max) | Quantitativa Continua | Preditora |
| **windspeed** | Normalized wind speed. The values are divided to 67 (max) | Quantitativa Continua | Preditora |
| **casual** | count of casual users | Quantitativa Discreta | Target |
| **registered** | count of registered users | Quantitativa Discreta | Target |
| **cnt** | count of total rental bikes including both casual and registered | Quantitativa Discreta | Target |

1. Existe correlação entre CNT e temperatura?
   1. Apresente o gráfico de dispersão entre as variáveis cnt (total) versus temperatura. Interprete os resultados.

Gráfico, Gráfico de dispersão

O conteúdo gerado por IA pode estar incorreto.

* 1. Apresente a H0 e H1 do teste de hipóteses.

**H0: r = 0, a correlação é igual a zero, ausência de correlação**

**H1: r <> 0, dependência**

**Erro de decisão: alpha = 0,05 (nível significância)**

* 1. Apresente os resultados do Teste de Correlação de Pearson. Interprete os resultados do teste de hipóteses da saída do Python.

Texto

O conteúdo gerado por IA pode estar incorreto.

**Valor p < 0,05, dentro do nível de significância. Logo as variáveis são dependentes devido correlação de 0.64 (moderada)**

1. Existe correlação entre CNT e umidade?
   1. Apresente o gráfico de dispersão entre as variáveis cnt (total) versus temperatura. Interprete os resultados.

Gráfico, Gráfico de dispersão

O conteúdo gerado por IA pode estar incorreto.

* 1. Apresente a H0 e H1 do teste de hipóteses.

**H0: r = 0, a correlação é igual a zero, ausência de correlação**

**H1: r <> 0, dependência**

**Erro de decisão: alpha = 0,05 (nível significância)**

* 1. Apresente os resultados do Teste de Correlação de Pearson. Interprete os resultados do teste de hipóteses da saída do Python.

Texto

O conteúdo gerado por IA pode estar incorreto.

**Valor p < 0,05, dentro do nível de significância. Logo as variáveis são dependentes devido correlação fraca de -0.10.**

1. Existe correlação entre CNT e Estação do ano?
   1. Apresente o gráfico Box Plot das variáveis cnt (total) versus estação do ano. Interprete os resultados.

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

**O Boxplot das variáveis CNT versus Estação do Ano mostra que a mediana do número total de ocorrências (CNT) varia entre as diferentes estações. Além disso, os quartis, bem como os valores máximos e mínimos, apresentam variações significativas entre as estações.**

**Isso sugere que pode haver uma associação entre a estação do ano e o total de ocorrências (CNT).**

* 1. Apresente a tabela e o gráfico 100% empilhado da variável cnt (total) versus a variável estação do ano. Interprete os resultados.

**Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.Pelo gráfico empilhado reforça a existência de variação de CNT entre as Estações do Ano.**

* 1. Apresente a H0 e H1 do teste de hipóteses.

**H0: r = 0, a correlação é igual a zero, ausência de correlação**

**H1: r <> 0, dependência**

**Erro de decisão: alpha = 0,05 (nível significância)**

* 1. Apresente os resultados do Teste Qui-quadrado calculado no Python. Interprete os resultados do teste de hipóteses da saída do Python.

Texto

O conteúdo gerado por IA pode estar incorreto.

**Rejeitamos H0, logo as variáveis são dependentes.**

1. Existe correlação entre CNT e dia da semana?
   1. Apresente o gráfico Box Plot das variáveis cnt(total) versus dia da semana do ano. Interprete os resultados.

**Boxplot das variáveis CNT versus Dia da Semana do Ano mostra que a mediana do número total de ocorrências (CNT) é semelhante entre os dias de semana. Além disso, os quartis, bem como os valores máximos e mínimos, não apresentam variações significativas.**

**Isso sugere que não há uma associação clara entre o dia da semana e o total de ocorrências (CNT), indicando que a variável dia da semana pode não influenciar diretamente a variação de CNT.**

Gráfico, Gráfico de caixa estreita

O conteúdo gerado por IA pode estar incorreto.

* 1. Apresente a tabela e o gráfico 100% empilhado da variável cnt (total) versus a variável estação do ano. Interprete os resultados.

**Pelo gráfico empilhado reforça a não existência de variação de CNT entre** Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.**e Dia da Semana**

* 1. Apresente a H0 e H1 do teste de hipóteses.

**H0: r = 0, a correlação é igual a zero, ausência de correlação**

**H1: r <> 0, dependência**

**Erro de decisão: alpha = 0,05 (nível significância)**

* 1. Apresente os resultados do Teste Qui-quadrado calculado no Python Interprete os resultados do teste de hipóteses da saída do Python

Texto

O conteúdo gerado por IA pode estar incorreto.

**Falhamos em rejeitar H0, logo variáveis não são dependentes.**

Data de entrega: 13/02/2025

Arquivo formato Word.

Regina Bernal

29/01/2025