

TESTE DE KRUSKAL-WALLIS E TESTE DE FRIEDMAN

i) KRUSKAL-WALLIS

- Generalização do teste de Mann-Whitney para mais de 2 variáveis.
- H_0 : todas as populações são iguais.
- H_1 : Existe diferenças entre populações.

procedimento para o teste:

- classificação conjunta das observações (postos)
- Fazer a soma dos postos em cada amostra
- Calcular a estatística de teste (H) e verificar p-valor
- Se p-valor significativo aplicar teste de comparação dos postos médios

Exemplo: Para o arquivo "dados-aula10" planilha KW,

- verificar se há diferença significativa na % de impureza da água submetidos a diferentes tratamentos.
- Verificar se há diferença significativa no grau de emissão de ruído.

FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

ii) FRIEDMAN

Aplicado para k amostras dependentes → é uma generalização do teste de Wilcoxon.

Procedimento: classificação dos dados dentro de um bloco (da mesma situação) e determinação da estatística de teste (χ^2). Se p-valor significativo aplicar comparação de postos médios.

Exemplo: Para os dados da planilha "Friedman" verificar se a % de decomposição é estatisticamente igual para os diferentes tratamentos.

FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

MEDIDAS DE ASSOCIAÇÃO E COEFICIENTES DE CORRELAÇÃO

Interesse em mensurar a "força" da associação entre as variáveis por meio de cálculo de algum coeficiente.

Duas variáveis qualitativas nominais

coeficiente de contingência (C) ou o coeficiente PHI (r_ϕ)

Montar uma tabela de contingência

Quanto maior o coeficiente maior será o grau de associação entre as variáveis.

O valor máximo do Coeficiente de Contingência depende da dimensão da tabela de contingência. Ex.: Tabela 2x 2 = 0,71.

Coef. Phi tem como valor máximo de 1, entretanto, só é aplicavel em tabelas 2 x 2 (variáveis dicotômicas)

Exemplo de aplicação: Para os dados de Teste1 e Teste2 (dados-aula10.xls "contingencia"). Determinar o coeficiente de correlação entre Teste1 e Teste2, usando Coef de Contingência e Phi.

FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

Duas variáveis qualitativas ordinais

O coeficiente mais utilizado é o coeficiente de correlação de Spearman (r_s)

A significância pode ser obtida por meio da distribuição t com $gl = v = n-2$

Exemplo: Verificar se existe relação entre opinião de pessoas do sexo masculino e do sexo feminino com relação a qualidade ambiental (plan "Spearman")

Duas variáveis quantitativas

O coeficiente utilizado é o coeficiente de correlação de Pearson (r)

Diagrama de dispersão indica a possibilidade de correlação linear.

Avalia a força e a direção do relacionamento LINEAR entre as duas variáveis.

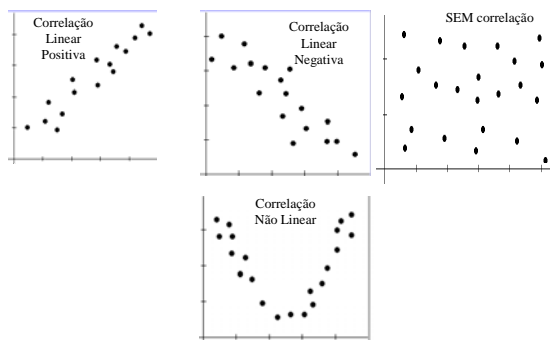
A significância é verificada pelo teste t com $gl = n-2$

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}}$$

FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

Diagrama de Dispersão



FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

Exemplo: Verificar se existe correlação linear entre os atributos físicos do solo apresentados na plan "Pearson"

REGRESSÃO LINEAR SIMPLES

Objetivo: obter uma equação estatística que expresse o relacionamento entre a variável dependente (resposta) e as independentes (explicativas, preditoras).

Regressão simples: apenas duas variáveis (uma dependente e outra independente).

regressão múltipla: uma dependente e mais de uma independente

modelo de regressão linear simples:

$$Y_i = \alpha + \beta X_i + e_i = \beta_0 + \beta_1 X_i + e_i$$

Os erros do modelo devem ser independentes e ter distribuição normal com média zero e variância constante

$$e \sim N(0, \sigma^2)$$

FAMAT/UFPA

Prof. Dr. Ednaêdo Garvão Guimarães

Considerando a equação estimada por meio da amostra como:

$$\hat{Y}_i = b_0 + b_1 x_i$$

Temos, usando o método dos mínimos quadrados:

$$b_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad b_0 = \bar{y} - b_1 \bar{x}$$

As inferências no modelo de regressão linear estão baseadas em:

- i) Anova do modelo \rightarrow se p-valor $< 0,05$ o modelo é considerado adequado para explicar a variação de Y em função de X, ou seja, o coeficiente de regressão (coeficiente angular é significativo).
- ii) Coeficiente de determinação (R^2) \rightarrow indica a % de explicação da variável Y em função de X por meio do modelo obtido.
- iii) Intervalos de confiança dos parâmetros \rightarrow estima os limites de confiança dos parâmetros do modelo (útil na comparação de modelos)

FAMAT/UFPA

Prof. Dr. Edinaldo Carneiro Guimarães

iv) Testes de hipóteses dos parâmetros \rightarrow verifica se os parâmetros do modelo são estatisticamente iguais a zero.

v) Verificar se os resíduos ($Y_{obs} - Y_{est}$) seguem os pressupostos do modelo, ou seja, independentes com média zero, distribuição normal e variância constante.

Exemplo: Determinar o modelo de regressão linear para a Resistência em função da Densidade e também para Resistência em função da umidade (plan: "Regressão").

FAMAT/UFPA

Prof. Dr. Edinaldo Carneiro Guimarães