

SME 0810 - Métodos Não Paramétricos

Trabalho Prático 1/2

Testes para Duas Amostras Relacionadas

Aluno(a) 1: Ada Maris Pereira Mário..... Nº USP: 12725432

Aluno(a) 2: Emanuel Victor da Silva Favorato..... Nº USP: 12558151

RELATÓRIO

1 Introdução e Objetivos

Neste projeto, serão utilizadas duas bases de dados para a realização de testes não-paramétricos para dados pareados.

A primeira [Kaggle, 2022] diz respeito a notificações de diabetes reportadas pelo "National Institute of Diabetes and Digestive and Kidney Diseases" entre mulheres maiores de 21 anos descendentes de indígenas da etnia Pima. Estamos interessados em ajustar um modelo de regressão logística para o diagnóstico da comorbidade e queremos verificar se os erros de predição se devem ao acaso, ou se são provenientes de falhas sistemáticas no ajuste, as variáveis a serem utilizadas neste estudo são:

- **Pregnancies:** variável discreta. Indica quantas vezes uma determinada mulher esteve grávida.
- **Glucose:** variável discreta. Indica o nível de glicose no sangue.
- **BloodPressure:** variável discreta. Corresponde à medição da pressão sanguínea.
- **SkinThickness:** variável discreta. Indica o quão densa é a pele de uma determinada paciente.
- **Insulin:** variável discreta. Expressa os níveis de insulina no sangue.
- **BMI:** variável contínua. Corresponde ao índice de massa corporal.
- **DiabetesPedigreeFunction:** variável contínua. Indica a porcentagem de diabetes.
- **Age:** variável discreta. Corresponde à idade das pacientes.
- **Outcome:** variável dicotômica. É a nossa variável resposta e assume os valores 0 (se a paciente não possui diabetes) ou 1 (se esta está acometida pela enfermidade).

Já o segundo conjunto de dados [Kaggle, 2019] é relativo a um experimento sobre efeitos de medicamentos ansiolíticos benzodiazepínicos na memória. Os medicamentos testados foram são Alprazolam (A), Triazolam (T), e um grupo Placebo (S). Cada medicamento tem dosagens que vão de 1 a 3, com 1 sendo a mais baixa e 3 a mais alta. É constituído por 9 colunas e 199 entradas. Neste estudo, o interesse principal é em relação aos efeitos do Alprazolam (A) em dose baixa (1) na pontuação de memória. Nesse sentido, as variáveis a serem analisadas são:

- **Mem_Score_Before:** variável contínua. Quanto tempo (segundos) foi necessário para o indivíduo finalizar um teste de memória antes de tomar o medicamento.
- **Mem_Score_After:** variável contínua. Quanto tempo (segundos) foi necessário para o indivíduo finalizar um teste de memória após tomar o medicamento.

2 Metodologia

Para a realização de testes, será utilizado $\alpha = 5\%$. Para os testes de McNemar serão utilizados os dados da base [Kaggle, 2022]. Para os demais, será utilizada a base [Kaggle, 2019].

As hipóteses a serem testadas serão:

- **Teste de McNemar:** Os falsos positivos e negativos do modelo de regressão para diagnóstico de diabetes são resultantes de falhas sistemáticas.

$$H_0 : P(FP) = P(FN) \quad X \quad H_1 : P(FP) \neq P(FN)$$

- **Teste dos Sinais:** O Alprazolam em doses baixas causa efeitos colaterais na cognição, deixando a memória mais lenta.

$$H_0 : P(+) = P(-) \quad X \quad H_1 : P(+)<P(-)$$

- **Teste de Wilcoxon:** O Alprazolam em doses baixas causa efeitos colaterais na cognição, deixando a memória mais lenta.

$$H_0 : E(D) = 0 \quad X \quad H_1 : E(D) < 0$$

Para o tratamento, exploração e visualização dos dados, utilizaram-se métodos das bibliotecas Pandas e Matplotlib da linguagem de programação Python. Para os testes, utilizaram-se métodos das bibliotecas statsmodels, SciPy e SKLearn da mesma linguagem.

3 Resultados

Inicialmente, os dados foram pré-processados verificando a existência de valores nulos, inválidos ou duplicados; os quais não estavam presentes. Para os testes do sinal e de Wilcoxon, o conjunto foi filtrado de modo a incluir apenas as entradas referentes ao medicamento A na dosagem 1.

Realizados os testes, encontraram-se os seguintes resultados:

- **Teste de McNemar:** A estatística de teste calculada foi de $t = 0,075$. Observou-se $p\text{-valor} = 0,783$, não se rejeitando H_0 .

Conclusão: Ao nível de significância 5%, os dados fornecem evidências de que os erros de diagnóstico são provenientes do acaso e não devido a falhas sistemáticas no modelo de regressão logística.

- **Teste dos Sinais:** A estatística de teste calculada foi de $t = 9$ dentre um n corrigido de $n = 23$. Observou-se $p\text{-valor} = 0,202$, não se rejeitando H_0 .

Conclusão: Ao nível de 5% de significância, os dados evidenciam que o Alprazolam em doses baixas não provoca efeitos na cognição de modo a deixar mais lenta a memória.

- **Teste de Wilcoxon:** A estatística de teste calculada foi de $t = 132$. Observou-se $p\text{-valor} = 0,435$, não se rejeitando H_0 .

Conclusão: Ao nível de 5% de significância, os dados evidenciam que o Alprazolam em doses baixas não provoca efeitos na cognição de modo a deixar mais lenta a memória.

As informações estão sumarizadas na Tabela 1:

Teste	Estatística de Teste	$p\text{-valor}$	Decisão
Teste de McNemar	$t = 0,075$	0,783	Não Rejeitar H_0
Teste dos Sinais	$t = 9$	0,202	Não Rejeitar H_0
Teste de Wilcoxon	$t = 132$	0,435	Não Rejeitar H_0

Tabela 1: Sumário dos testes não-paramétricos realizados

4 Considerações Finais

Neste trabalho, foram aplicados três testes não-paramétricos para analisar dados pareados em dois cenários distintos: o diagnóstico de diabetes utilizando um modelo de regressão logística e os efeitos do Alprazolam na memória de indivíduos em uma dosagem baixa. Os testes de McNemar, dos Sinais e de Wilcoxon foram realizados com base nos objetivos de verificar se os erros do modelo de diagnóstico de diabetes eram sistemáticos e se o Alprazolam em baixa dosagem afetava negativamente a memória.

Os resultados dos testes não levaram à rejeição das hipóteses nulas em todos os casos, indicando que os erros de predição no diagnóstico de diabetes parecem ser atribuíveis ao acaso, sem falhas sistemáticas no modelo. Além disso, os efeitos colaterais do Alprazolam em doses baixas sobre a memória também não apresentaram evidências de impacto significativo na cognição dos indivíduos.

Esses resultados são consistentes com o esperado para o nível de significância adotado ($\alpha = 5\%$), reforçando que, tanto para a predição do diagnóstico de diabetes quanto para o desempenho da memória após o uso do Alprazolam, não há indícios fortes de problemas que requeiram modificações nos modelos ou preocupações com os efeitos do medicamento na dose investigada.

Referências

- Kaggle. Memory test on drugged islanders data, 2019. Disponível em: <https://www.kaggle.com/datasets/steveahn/memory-test-on-drugged-islanders-data/data>. Acesso em: 08 de outubro de 2024.
- Kaggle. Diabetes dataset, 2022. Disponível em: <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>. Acesso em: 08 de outubro de 2024.