LISTA DE EXERCICIOS PARA ENTREGAR INDIVIDUALMENTE

RECOMENDÁVEL FAZER COM SUA DUPLA

LEMBREM-SE:

* IMPORTANTE EXPLANAR E EXPLICAR O ENTENDIMENTO NA SUA ENTREGA.
* O NOME DO ARQUIVO DEVE SER SEU **PRIMEIRONOME\_SOBRENOME** (Exemplo: Adriana\_Silva)
* VOCÊ FARÁ A ENTREGA DO TRABALHO EM WORD COM TODAS AS **EXPLICAÇÕES, RACIONAIS, CÓDIGOS E RESULTADOS DO PYTHON (SIM, A LISTA É PARA SER FEITA NO PYTHON, pode dar print e colar aqui junto com sua resposta)**.
* NÃO USAR COR VERMELHA, POIS ELA FICA DESTINADA ÀS CORREÇÕES!!
* ESSA ENTREGA SERÁ VIA MOODLE.

1. (1,0) Fale com suas palavras. Como fazer a escolha entre testes paramétricos e testes não paramétricos?

Escolher entre testes paramétricos e não paramétricos depende principalmente de dois fatores: as características dos seus dados e os objetivos da sua análise.

**Testes Paramétricos:**

São usados quando seus dados atendem a certas condições, são elas:

1. Normalidade: os dados seguem uma distribuição normal. Exemplo de uma normal  
   A graph with a blue line

   Description automatically generated  
   A screen shot of a computer program

   Description automatically generated
2. Homogeneidade de variâncias: os grupos que são analisados devem ter variâncias semelhantes.
3. Independência: as observações são independentes umas das outras.
4. Escala de medida: a os dados devem ser variáveis quantitativas. Se seus dados atendem a essas suposições, os testes paramétricos são geralmente mais poderosos e precisos, significando que eles têm uma melhor chance de detectar uma diferença verdadeira se ela realmente existir.

**Testes não Paramétricos:**

São mais flexíveis, e são usados para testes com várias qualitativas que não atendem aos requisitos do teste paramétrico ou para variáveis qualitativas.

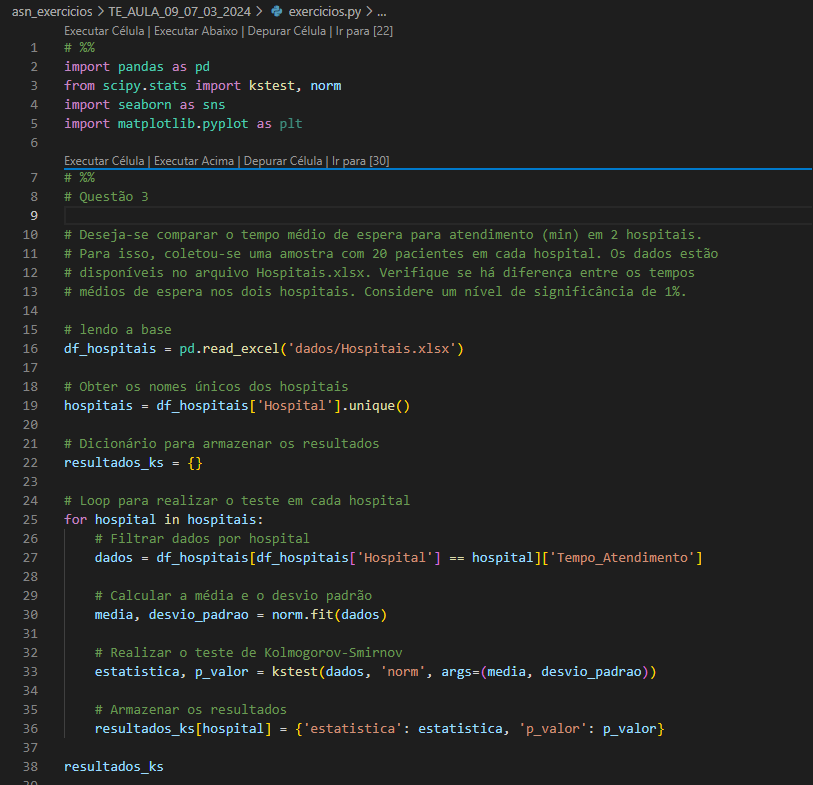
Os dados são ordinais, nominais ou não são medidos em uma escala intervalar ou de razão. Isso significa que, se estamos lidando com categorias em vez de números contínuos, ou se seus dados são ranqueados, os testes não paramétricos podem ser a melhor escolha. Importante, para pequenas amostras, dependendo do seu apetite ao risco, por mais que atendam aos requisitos do paramétrico, pode ser mais interessante que os testes sejam feitos no não paramétricos.

1. (1,0) Um fabricante de biscoito afirma que a quantidade de calorias em cada pacotinho é de 60 cal. Para verificar se essa informação é verdadeira, uma amostra aleatória de 36 pacotinhos é coletada, observando-se que a quantidade média de calorias é de 65 cal com desvio-padrão de 3,5. Aplique o teste adequado e verifique se a informação do fabricante é verdadeira, considerando um nível de significância de 5%.

Não consegui finalizar.

1. (1,0) Deseja-se comparar o tempo médio de espera para atendimento (min) em 2 hospitais. Para isso, coletou-se uma amostra com 20 pacientes em cada hospital. Os dados estão disponíveis no arquivo Hospitais.xlsx. Verifique se há diferença entre os tempos médios de espera nos dois hospitais. Considere um nível de significância de 1%.

Primeiramente faremos os testes para saber se a base atende aos pressupostos da normalidade



O primeiro teste é se os grupos atendem a distribuição da normal:

A computer screen with white text

Description automatically generated

De acordo com o p-valor em ambos dos casos, não rejeitamos a hipótese nula, dito isso, ambos as amostras atendem a normalidade.

Vamos aproveitar e fazer uma análise visual dos dados:

HOSPITAL 1

A computer screen shot of text

Description automatically generated

A graph of a normal distribution

Description automatically generated

HOSPITAL 2

A computer screen shot of text

Description automatically generated

A graph of a normal distribution

Description automatically generated

Agora ao fazermos o teste F de Levene para comparar a variabilidade (ou variância) dos tempos de atendimento entre o Hospital 1 e o Hospital 2, tivemos o p-valor de 0.405.

A computer screen with text on it

Description automatically generated

A screenshot of a computer

Description automatically generated

Interpretando o resultado, podemos dizer, com um nível de significância comum de 0.05, que não temos evidências suficientes para rejeitar a hipótese nula do teste de Levene. Dito isso, podemos afirmar que as variâncias dos dois grupos são iguais. Em outras palavras, com base no teste feito, parece que a variabilidade dos tempos de atendimento entre o Hospital 1 e o Hospital 2 é estatisticamente similar.

Analisando visualmente:

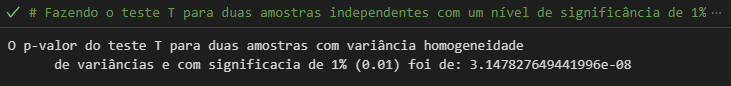
A chart with blue and orange squares

Description automatically generated

Quanto aos outros dois pressupostos, eles são independentes e são contínuos. Sendo assim, faremos o teste estatístico T para duas amostras independentes e com variâncias homogêneas.

A screen shot of a computer program

Description automatically generated

****

Com este resultado, temos evidências suficientes para rejeitar a hipótese nula (H0), que afirma que não há diferença significativa entre as médias dos tempos de atendimento dos dois hospitais. Concluímos que existe uma diferença estatisticamente significativa nos tempos de atendimento entre o Hospital 1 e o Hospital 2, com um nível de confiança de 99%.

1. (1,0) Trinta adolescentes com nível de colesterol total acima do permitido foram submetidos a um tratamento que consistia em dieta e atividade física. A planilha Colesterol.xlsx apresenta os índices de colesterol LDL (mg/dL) antes e depois do tratamento. Verifique se o tratamento é eficaz (com nível de significância de 5%).

Se paramétrico

Teste T para duas amostras Emparelhada

Se não paramétricos = teste Wilcoxon

Primeiro carregamos a base:

A screen shot of a computer

Description automatically generated

Segundo passo é fazer o teste se as bases seguem uma normal, para tal usamos o teste de Shapiro-Wilk

A screen shot of a computer code

Description automatically generated

Com o pvalue de 0.1073339581489563 para o Antes do Tratamento não rejeitamos a hipótese nula e assumimos que a tabela segue uma distribuição normal. O mesmo racional para a base do Depois do tratamento, tendo como resultado: pvalue=0.28345248103141785

Vamos aproveitar e fazer uma validação visual:

A graph with a line

Description automatically generated

A screen shot of a computer

Description automatically generated

A blue line graph with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

Quarto passo é fazer o teste F de Levene para saber se temos igualdade das variâncias:

A screen shot of a computer

Description automatically generated

Teste de Levene:

**Valor p: 0.33235607530146893**

Com esse resultado não rejeitamos a hipótese nula.

Quinto passo é fazer o T de Student emparelhado, uma vez que estamos olhando para a mesma pessoa ao longo do tempo:

A computer screen shot of a code

Description automatically generated

Teste t de Student emparelhado:

**Valor p: 9.960437580616326e-13**

Dado o resultado, quando fazemos o comparativo do antes e do depois, podemos dizer que o tratamento teve algum tipo de resultado.

Pensei em fazer o teste T de Student olhando para o unicaudal para saber se o Valor depois do tratamento é maior que a média do Colesterol antes do tratamento. Podemos dizer que

A screen shot of a computer

Description automatically generated

Podemos dizer com esse teste que:

***O colesterol após o tratamento é significativamente menor do que o valor de referência (220).***

1. (1,0) Um grupo de 20 consumidores fez um teste de degustação com dois tipos de cerveja (Marca A e Marca B). Ao final, escolheram uma das marcas, como vemos a seguir. Teste a hipótese de não há diferença na preferência dos consumidores, ao nível de significância de 5%.

|  |  |  |  |
| --- | --- | --- | --- |
| Eventos | Marca A | Marca B | Total |
| Frequência | 8 | 12 | 20 |

A screen shot of a computer program

Description automatically generated

De acordo com os testes acima (Valor p: 0.34278171114790873), podemos afirmar que não há preferência entre as marcas.

Também fiz o teste Binomial e o resultado foi o mesmo:

A screenshot of a computer program

Description automatically generated

A black background with white text

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Eventos | Livro A | Livro B | Livro C | Total |
| Frequência | 29 | 15 | 16 | 60 |

1. (1,0) Um grupo de 60 leitores fez uma avaliação de três livros de romance e, ao final, escolheram uma das três opções. Teste a hipótese nula de que não há preferência dos leitores, com nível de significância de 5%.

Não consegui fazer no python.

1. (1,0) Um grupo de 20 adolescentes fez a dieta dos pontos por um período de 1 mês. Verifique se houve redução de peso depois da dieta. Arquivo Dieta.xlsx. Considere nível de significância de 5%.

A computer screen with text on it

Description automatically generated

Fazendo os testes de normalidade:

**Base Antes da Dieta**

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A blue line graph with numbers

Description automatically generated

**Base Depois da Dieta**

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A blue line graph with white text

Description automatically generated

Apesar de parecer uma normal, segundo o teste as bases não seguem uma normal. Dito isso, seguiremos para o teste de McNemar:

A screen shot of a computer program

Description automatically generated

Resultado:

A screenshot of a computer

Description automatically generated

Segundo o teste, podemos dizer que estatisticamente houve mudança no peso com a Dieta.

1. (1,0) Um grupo de 15 consumidores avaliou o nível de satisfação (1=baixo, 2=médio, 3=alto) de três serviços bancários diferentes. Os resultados estão na tabela Banco.xlsx. Verifique se há diferença entre os três serviços. Considere nível de significância de 5%.

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

De acordo com o teste ANOVA, podemos dizer que há diferença entre os serviços dos bancos.

1. (1,0) Conte um problema que você consegue imaginar, explique o contexto e qual teste desenvolveria.

Não consegui fazer.

1. (1,0) Responda o que fazer para:
   1. Verificar associação entre duas variáveis nominais  
      Por exemplo. queremos saber se existe uma associação entre gênero (masculino ou feminino) e preferência por um determinado tipo de música (rock, pop, hip-hop). Usamos o teste do qui-quadrado para determinar se há uma relação significativa entre gênero e preferência musical
   2. Verificar associação entre duas variáveis numéricas  
      Queremos determinar se há uma associação entre a quantidade de horas de estudo por semana e a nota final de um aluno. Utilizamos o coeficiente de correlação de Pearson para medir a força e a direção da relação linear entre essas duas variáveis. Importante também olhar o **GRÁFICO**, somente o número pode te enganar!
   3. Verificar associação entre duas variáveis ordinais  
      Por exemplo, existe uma associação entre a classificação de satisfação do cliente (insatisfeito, neutro, satisfeito) e a frequência de compra (baixa, média, alta)? Para responder podemos utilizar o coeficiente de correlação de Spearman para avaliar a relação entre essas variáveis ordinais.
   4. Verificar associação entre uma variável nominal e outra numérica  
      Queremos saber se há uma diferença significativa na pressão arterial entre pacientes com diabetes e pacientes sem diabetes. Utilizamos o teste t de Student para comparar as médias da pressão arterial entre os dois grupos de pacientes (variável nominal: presença ou ausência de diabetes).