

Questão 1

Assuma que a distribuição da diferença (D) entre os resultados do método novo e padrão seja Normal. Se os métodos fossem equivalentes, suas medidas seriam muito parecidas, e a diferença, D , teria média 0. Suponha, para simplificar, que o desvio-padrão amostral igual a 3,5 possa ser considerado como o desvio-padrão populacional, σ , e que possamos desprezar a imprecisão na obtenção de cada medida. Nessas condições quantos cães esperaríamos encontrar nessa amostra com diferenças (em módulo, ou seja, tanto para mais quanto para menos) maiores do que $6 U/L$?

Alternativas:

- (a) de 3 a 4
- (b) de 1 a 2
- (c) de 8 a 9
- (d) nenhum
- (e) menos de 1

Solução: **Alternativa a.**

Assumindo que a distribuição da diferença D entre os resultados do método novo e padrão seja Normal com média 0 e com desvio-padrão amostral aproximadamente 3,5 — isto é, $D \sim N(0; 3,5^2)$ —, tem-se:

$$P(|D| > 6) = P(D < -6) + P(D > 6) = P(D < -6) + (1 - P(D < 6)),$$

ou ainda, dada a simetria da distribuição Normal,

$$P(|D| > 6) = 2 * P(D < -6) \approx 2 * 0,043.$$

Assim, o número esperado de cães para essa amostra, com diferenças maiores do que $6 U/L$ é

$$\mathbb{E}(P(|D| > 6)) = n * P(|D| > 6) \approx 40 * 0,086 \approx 3,44.$$

Questão 2

Um estudo para relacionar tamanho do cérebro, gênero e inteligência, coletou em uma amostra de 40 indivíduos, variáveis referentes à inteligência (com resultados resumidos em 3 variáveis que refletem o QI), ao tamanho do cérebro (via ressonância magnética) e pessoais (peso, altura e gênero). Os dados encontram-se no arquivo QI, na página da disciplina. Para responder às perguntas abaixo, leve em consideração as suposições dos métodos e se os dados as satisfazem.

Alternativas:

1. Que método é adequado para comparar as variâncias dos pesos de homens e mulheres?
2. Que método é adequado para verificar se existe relação entre os valores de QIs verbal e de os dados de RM?

3. Que método é adequado para comparar o QI médio verbal de mulheres e homens?
4. Que método é adequado para prever o QI_Total do indivíduo a partir de variáveis pessoais, de inteligência e dimensões do cérebro?
5. Que método é adequado para comparar os QIs médios verbal e de performance?
 - (a) correlação de Pearson
 - (b) teste de Mann-Whitney
 - (c) teste de Levene
 - (d) regressão linear
 - (e) teste t de Student pareado

Solução: 1.(c); 2.(a); 3.(b); 4.(d); 5.(e).

Note que:

- O coeficiente de correlação de Pearson mede o grau da correlação linear entre duas variáveis quantitativas; assume valores entre -1 e 1 , sendo que:
 - 1 significa uma correlação perfeita positiva entre as duas variáveis;
 - -1 significa uma correlação negativa perfeita entre as duas variáveis (se uma aumenta, a outra sempre diminui);
 - 0 significa que as duas variáveis não dependem linearmente uma da outra; no entanto, pode existir uma outra dependência que seja “não linear”, cujo resultado deve ser investigado por outros meios.
- O teste de Mann-Whitney é usado quando se dispõe de uma amostra pequena de duas populações independentes das quais não há informações a respeito de suas distribuições (ou estas são sabidamente não normais), e quer-se testar se as distribuições são iguais em localização, isto é, há interesse em saber se uma população tende a ter valores maiores do que a outra, ou se elas têm aproximadamente a mesma mediana.
- O teste de Levene é um teste para a igualdade de variâncias de dois ou mais grupos (com qualquer distribuição contínua).
- Regressão linear múltipla é uma coleção de técnicas estatísticas para construir modelos que descrevem de maneira razoável relações entre várias variáveis explicativas de um determinado processo. A diferença entre a regressão linear simples e a múltipla é que na múltipla são tratadas duas ou mais variáveis explicativas.
- O teste t de Student pareado é utilizado para realizar testes de igualdade de variâncias ou de médias, em que trabalha-se com amostras dependentes (pareadas) — para cada unidade amostral realiza-se duas medições.

Questão 3

Em Estatística, tenta-se, sempre que possível, testar conjuntamente hipóteses sobre parâmetros de um modelo. Para testar a hipótese de que os métodos padrão e novo são equivalentes, um das análises desenvolvidas em aula foi verificar se a reta de regressão que os relaciona é a identidade ($x = y$). Para tanto, basta verificar se o intercepto do modelo é igual a 0 e a inclinação é igual a 1 . No entanto, em aula, esses testes de hipóteses sobre os dois parâmetros foram feitos separadamente.

Juntamente com o código disponibilizado em aula, utilize o comando:

```
hypotheses = '(Padrão = 1), (Intercept = 0) '  
print(res.f_test(hypotheses))
```

para o teste conjunto e escolha qual das alternativas abaixo está INCORRETA:

Alternativas:

- (a) Se fixarmos o mesmo nível de significância para todos os testes de hipóteses envolvidos em nossa conclusão, a probabilidade de cometer o erro tipo I interpretando o resultado do teste de hipóteses conjunto será menor do que probabilidade de cometer o erro tipo I na conclusão global baseada nos dois testes separados.
- (b) Para o teste de hipóteses conjunto desses dois parâmetros, o nível de significância do teste deverá ser duplicado.
- (c) A um nível de significância de 1%, rejeita-se a hipótese de que a reta de regressão que relaciona os dois métodos é a identidade ($x = y$).
- (d) com o comando `f_test`, podemos testar várias hipóteses simultaneamente, mais de 2, inclusive.
- (e) outra opção para testar hipóteses desse tipo é o teste de Wald.

Solução: **Alternativa b.**

Note que:

- Um teste conjunto de uma hipótese deve ser usado se uma inferência ou conclusão exigir que múltiplas hipóteses sejam simultaneamente verdadeiras. No contexto de regressão, uma hipótese conjunta impõe restrições a múltiplos coeficientes de regressão, diferentemente da realização de testes individuais, nos quais uma restrição é imposta a um único coeficiente. O nível de significância usado para um teste individual é a probabilidade marginal de rejeitar falsamente a hipótese, independentemente de as hipóteses restantes serem rejeitadas. Assim, a taxa de erro do tipo 1 ao usar os testes individuais é muito grande, a menos que os níveis de significância correspondentes sejam divididos pelo número de hipóteses, pois nesse caso, esse tipo de ajuste é suficiente para testes independentes ¹.
- O teste F compara os ajustes de diferentes modelos lineares forma de verificar a contribuição de uma ou mais variáveis explicativas no modelo; é um caso especial do teste de Wald. Seu uso possibilita controlar os erros tipo I aplicando apenas um teste para verificar uma hipótese nula linear conjunta, mantendo a probabilidade de rejeitar H_0 incorretamente igual a α .
- O teste de Wald baseia-se na distribuição assintótica normal dos estimadores de máxima verossimilhança dos parâmetros do modelo e é usado para avaliar a significância de cada coeficiente (β) no modelo.

¹Mais em <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1713204/>.