

Métodos Computacionais para Estatística II

Prof: Jony Arrais Pinto Junior

Lista 05

1. Leia o banco de dados *Banco Escalas Psicologia* e faça as análises necessárias para responder os seguintes questionamentos. Faça todos os testes no R e no SPSS (quando possível). No banco de dados todos os valores 999 significam que não se conhece as informações. O banco é composto de variáveis sócio demográficas idade, gênero e escolaridade, o teste do QI e testes que foram aplicados para verificar a cognição dos pacientes: RAVLT, WCST, Stroop, Digitos e Reprodução Visual. Foram examinados 58 pacientes controle (sem transtorno), 73 pacientes com trauma e 83 pacientes com TEPT (transtorno de stress pós-traumático). Quanto maior a pontuação no QI, no WCST e no teste de Digitos melhor está o paciente, nos demais testes é o contrário.
 - (a) Existe diferenças entre as idades dos 3 grupos?
 - (b) As distribuições de sexo nos três grupos são iguais?
 - (c) Os grupos possuem escolaridade semelhante?
 - (d) Pode-se dizer que existe diferença entre os QI dos grupos?
 - (e) Existe diferença entre as pontuações dos três grupos nos testes RAVLT, WCST, Stroop, Digitos e Reprodução visual?

Legenda:

Grupo: 0 - Controle, 1 - Trauma e 2 - TEPT

Sexo: 1 - Masculino e 2 - Feminino

Apresente as hipóteses de cada teste aplicado e apresente por escrito sua conclusão.

2. O interesse agora é comparar o grupo *controle* com o grupo *algum trauma* (Trauma ou TEPT), ou seja, crie uma variável chamada *novo grupo* no banco de dados e refaça as comparações da questão anterior com relação ao *novo grupo*.
3. Dois medicamentos, *A* e *B*, foram aplicados em grupos distintos de pessoas com o objetivo de verificar se eram igualmente eficientes em cicatrização de queimaduras de 3º grau. Uma área de $2\text{cm} \times 2\text{cm}$ destes pacientes foi tratada durante 1 mês com esses medicamentos e foi coletada a porcentagem da área tratada que ficou cicatrizada após o tratamento. Abaixo seguem os valores obtidos para os dois medicamentos.

A	31	21	24	25	23	26	25	23	26	30
B	25	31	31	32	22	26	15	21	26	29

- (a) Utilizando o SPSS apresente toda a análise necessária para responder a pergunta se os dois medicamento são igualmente eficazes. Adote um nível de significância de 5%. Apresente por escrito a sua conclusão. Apresente também um gráfico apropriado para acompanhar a análise.
- (b) Utilizando o R responda. Qual seria o poder do teste aplicado em (a) se a verdadeira diferença entre os dois medicamentos fosse de 3%?

4. Os dados da base *EMBRAPA* são referentes a um estudo para avaliar o efeito de quatro tratamentos no crescimento (peso em kg) de bezerros durante as primeiras 26 semanas. Os quatro tratamentos considerados foram: 1 - Pasto aos 8 dias de idade sem suplementação volumosa; 2 - Pasto aos 8 dias de idade mais capim elefante; 3 - Pasto aos 56 dias de idade sem suplementação volumosa e 4 - Pasto aos 56 dias de idade mais capim elefante.
- (a) Você diria que os pesos dos bezerros passadas as 26 semanas dos 4 tratamentos diferem a um nível de significancia de 5%?
 - (b) Desconsiderando os tratamentos, voce diria que o aumento de peso entre a semana 0 e a semana 26 foi superior a 88 kg? Adote um nível de significância de 4%.
5. Um dos conceitos de vulnerabilidade social envolve a renda do indivíduo. Como esta é uma variável de difícil acesso, pois nem sempre os entrevistados gostam de fornecê-la, um sociólogo estava interessado em saber se a situação do imóvel de moradia (1 - alugado, 2 - próprio) e a escolaridade da mãe (1 - baixa, 2 - média e 3 - alta) estão associadas à vulnerabilidade social para que num futuro próximo estas medidas pudessem formar um indicador de vulnerabilidade social, sem que seja necessário perguntar a renda dos indivíduos. Deste modo, foram coletadas informações referentes a 40 indivíduos das seguintes variáveis: vulnerabilidade social, situação do imóvel, escolaridade da mãe e anos de escolaridade do indivíduo. Faça as análises dos itens (a) e (b) no SPSS e (c) no R e lembre-se de fornecer a melhor análise possível. Forneça um gráfico para cada análise e apresente as hipóteses do teste utilizado, bem como uma conclusão por escrito. Use um nível de significância de 5%
- (a) Pode-se dizer que as variáveis situação do imóvel e escolaridade da mãe estão associadas com vulnerabilidade social? Comente os resultados.
 - (b) Pode-se concluir que existe uma diferença entre as idades de quem possui e não possui vulnerabilidade social?
6. Foram coletadas 15 amostras do modelo de ar condicionado A e 25 amostras do modelo B e foi medida a pressão sonora dos mesmos em decibéis. Foi fornecida a seguinte análise pelo estatístico da empresa. Use um nível de significância de 5% para responder os itens abaixo.

One-sample Kolmogorov-Smirnov test

```
data: A
D = 0.3977, p-value = 0.05741
alternative hypothesis: two-sided
```

One-sample Kolmogorov-Smirnov test

```
data: B
D = 0.5275, p-value = 0.9876
alternative hypothesis: two-sided
```

Wilcoxon rank sum test

data: ar and grupo

W = 400, p-value = 1.095e-06

alternative hypothesis: true location shift is not equal to 0

- (a) Você conduziria a análise do mesmo modo ou proporia algo diferente? Justifique.
- (b) Apresente os nomes, as hipóteses e a conclusão para os testes realizados pelo estatístico.
7. Um pesquisador da área de educação está interessado em checar se o seu número de publicações esta relacionado com o número de docentes doutores dos quadros das universidades. Para isso o mesmo coletou dados referentes a 15 universidades e observou os seguintes dados. Você diria que existe relação entre estas variáveis?

P	287	306	298	311	298	298	292	301	305	312	302	308	296	307	316
D	49	63	49	55	41	49	35	44	55	58	52	52	48	66	60

Estudo I: Leia o *Banco Cancer Mama*. Foram avaliadas 30 pacientes com câncer de mama e 27 sem câncer de mama. Somente as pacientes com câncer de mama foram submetidas a um tratamento durante os últimos 6 meses com o objetivo de diminuir o número de células cancerígenas. No banco de dados encontram-se as seguintes variáveis:

- Câncer (0 - Não e 1 - Sim)
- Peso (Kg)
- Primeira menstruação antes dos 12 anos (0 - Não e 1 - Sim)
- Grau de parentesco com alguém que já teve câncer de mama na família (0 - 3° Grau, 1 - 2° Grau e 2 - 1° Grau)
- Número de células cancerígenas (antes do tratamento)
- Número de células cancerígenas (depois do tratamento)

Todas as variáveis para os pacientes sem câncer foram medidas em uma única avaliação. Para as pacientes com câncer foram medidas todas as variáveis em uma primeira avaliação e na segunda (depois dos 6 meses de tratamento) foi medida somente o número de células cancerígenas.

Com base em um nível de significância de 5% responda as seguintes perguntas e forneça para cada uma delas um gráfico que ajude a visualizar os resultados inferenciais obtidos (sempre apresente a análise mais completa possível). Deverão ser apresentadas as hipóteses dos testes utilizados e uma conclusão escrita para todas as respostas.

8. Você diria que ter a primeira menstruação antes dos 12 anos é um fator de risco para o desenvolvimento de câncer de mama?
9. Você diria que o tratamento para as pacientes com câncer de mama foi eficiente?
10. A literatura da área médica aponta a obesidade como um fator de risco. Será que as pacientes sem câncer pesam menos do que as que tem?

11. Pode-se dizer que o grau de parentesco com pessoas portadoras de câncer é um fator de risco para o desenvolvimento de câncer de mama?
12. Pode-se dizer que existe relação entre o peso e o número de células cancerígenas antes do tratamento para as pacientes com câncer de mama?

Estudo II: Leia o *Banco Teste e Reteste*. Ele é composto por 46 pacientes depressivos. No banco de dados encontram-se as seguintes variáveis:

- Aparência (teste e re-teste)
- Higiene (teste e re-teste)
- Cuidado (teste e re-teste)
- ILSS
- Saúde (1,2 e 3)
- bom (ANTES e DEPOIS)

As variáveis aparência, higiene e cuidado foram medidas em duas avaliações (teste e re-teste), a primeira avaliação aconteceu na entrevista inicial e a segunda ocorreu depois de 6 meses de terapia. A variável ILSS que mede a saúde do paciente só foi colhida na avaliação inicial. As variáveis aparência, higiene, cuidado e ILSS são obtidas por meio de um questionário de perguntas. Todas essas escalas possuem pontuação variando entre 0 e 1, em que pontuação alta significa que o paciente está menos deprimido.

Uma nova escala chamada saúde está sendo testada, mas ninguém do consultório possui treinamento para poder aplicar a escala. Então três avaliadores distintos aplicaram a escala nos pacientes na primeira avaliação dos mesmos. A pontuação varia de 1 a 10.

Foi perguntado aos pacientes também se os mesmos acreditavam que iriam ficar com o tratamento antes deles iniciarem o tratamento e a mesma pergunta foi feita após a conclusão do tratamento. (0 - Não e 1 - Sim).

Com base em um nível de significância de 5% responda as seguintes perguntas e forneça para cada uma delas um gráfico que ajude a visualizar os resultados inferenciais obtidos (sempre apresente a análise mais completa possível). Deverão ser apresentadas as hipóteses dos testes utilizados e uma conclusão escrita para todas as respostas.

13. Você diria que houve uma melhora com a terapia com relação a aparência dos pacientes?
14. Você diria que houve uma melhora com a terapia com higiene a aparência dos pacientes?
15. Você diria que houve uma melhora com a terapia com relação aos cuidados dos pacientes?
16. Você diria que a nota da aparência na primeira avaliação está relacionada com a pontuação do ILSS?
17. Pode-se dizer que existe relação entre a nota da higiene na primeira avaliação com a pontuação do ILSS?

18. Houve diferença na crença dos pacientes com relação a melhora após a conclusão do tratamento?
19. Houve diferença entre os três avaliadores na escala saúde?