

ME111 - Atividade 07

Profa. Tatiana Benaglia

03/06/2020 - 1S2020

Testes de Hipótese para Dados Categóricos

Nessa atividade vocês irão trabalhar os conceitos de testes de hipóteses para variáveis categóricas, nos casos de uma variável ou duas variáveis. Ao final dessa atividade, o aluno deverá:

- entender o que é um teste de hipótese;
- saber identificar as hipóteses a serem testadas e o teste apropriado;
- aplicar os diversos testes em situações reais e tirar conclusões;
- realizar esses testes em algum software como R ou Excel.

Problema 1 - Teste Exato de Fisher

Nós vimos em aula o experimento “Uma senhora que toma chá” realizado por Fisher. Nesse experimento, 8 xícaras de chá foram utilizadas para testar se a senhora sabia diferenciar o que havia sido colocado primeiro na xícara: chá ou leite.

Você, como Estatístico(a), questiona o número de xícaras utilizado e acredita que 8 xícaras é muito pouco para tirar conclusões. Sendo assim, você decide refazer o experimento utilizando 20 xícaras (10 com chá sendo colocado primeiro e 10 com leite sendo colocado primeiro). Nesse caso, a senhora escolheu corretamente 8 das 10 xícaras feitas chá primeiro.

- a) Resuma o resultado desse experimento em uma tabela de contingência 2x2.
- b) Apresente as hipóteses sendo testadas.
- c) A estatística do teste utilizada é a quantidade de acertos (X) dentre as 10 xícaras feitas com chá primeiro. Sob a hipótese nula, X segue uma distribuição hipergeométrica com que parâmetros? Veja que cada valor possível de X resulta em uma tabela 2x2 diferente e temos associado a esse valor uma probabilidade. Apresente um gráfico com a distribuição de X . Dica: a função `dhyper()` calcula as probabilidades de uma distribuição hipergeométrica no R.
- d) Calcule o p-valor para esse teste. Dica: função `fisher.test()` no R.
- e) Assumindo um nível de significância $\alpha = 0.05$, qual a sua conclusão?
- f) Simule a distribuição de X , ou seja, gere números de acertos segundo a configuração do experimento. Repita essa simulação 1000 vezes e apresente um gráfico com a distribuição de X .
- g) Calcule o p-valor obtido através da simulação e compare-o com o p-valor exato.

Problema 2 - Lançamento de um Dado

- a) Para esse problema, vocês devem jogar um dado 60 vezes (manualmente mesmo!) e anotar quantas vezes que cada face ocorreu. Apresente isso em uma tabela.
- b) Você está desconfiado que o seu dado apresenta uma tendência de sair mais alguma(s) face(s) que outras. Sendo assim, aplique um teste de aderência nos dados observados para verificar se eles trazem evidência contra a hipótese de que o dado é não viesado, ou seja, todas as faces ocorrem com a mesma frequência. Apresente as hipóteses sendo testadas, a estatística do teste e o p-valor.

- c) Qual a sua conclusão a um nível de significância $\alpha = 0.05$? O que você observou traz evidência de que o seu dado é não viesado?

Problema 3 - Sabor de Skittles

Uma amostra de 70 pessoas foi selecionada e essas pessoas responderam qual era o seu Skittle favorito por cor (verde, laranja, roxo, vermelho ou amarelo). Uma outra pesquisa independente amostrou 90 pessoas, perguntando-lhes o seu Skittle favorito, mas ao invés de cor, eles pediram que a escolha foi feita pelo sabor (limão, laranja, uva, morango e limão siciliano, respectivamente). A tabela abaixo mostra os resultados das duas pesquisas. A maneira como as pessoas escolhem seus tipo de Skittle favorito, por cor ou sabor, parece estar relacionado a qual tipo é escolhido?

	Verde (Limão)	Laranja	Roxo (Uva)	Vermelho (Morango)	Amarelo (Siciliano)	Total
Cor	18	9	17	15	11	70
Sabor	13	16	19	34	8	90
Total	31	25	36	49	19	160

- Pelo tipo de amostragem realizado, aqui deve ser aplicado um teste de independência ou homogeneidade? Justifique.
- Quais são as hipóteses sendo testadas?
- Faça um gráfico que mostre as proporções das preferências de acordo com o critério (cor ou sabor).
- Apresente uma tabela com os frequências esperadas para cada uma das 10 células da tabela.
- As suposições para o teste qui-quadrado são satisfeitas, ou seja, as frequências esperadas são grandes o suficiente?
- Calcule a estatística do teste e o p-valor. Dica: use a função `chisq.test()` no R.
- O que você conclui desse teste? Existe evidência nos dados que o método de escolha (cor ou sabor) afeta nas respostas?

Problema 4 - Discriminação de Gênero

Você acredita que existe discriminação de gênero e que em um banco, por exemplo, homens são promovidos com mais frequência que mulheres, mesmo tendo as mesmas qualificações?

Um experimento foi realizado para verificar esse fato. Um total de 48 gerentes de banco participaram do estudo. Os gerentes foram questionados se eles recomendariam um candidato em particular para promoção ou não. Cada gerente recebeu um currículo e determinou se o candidato deveria ser promovido. Os currículos que eles receberam eram idênticos, exceto pelo fato que 24 deles foram identificados como sendo de homens e 24 deles sendo de mulheres. Desses 48 candidatos, 34 foram recomendados para promoção e 14 não foram. Dentre os 24 homens, 21 foram recomendados para promoção. Enquanto que dentre as 24 mulheres, 13 foram recomendadas para promoção. Os resultados estão resumidos na tabela abaixo.

	Promoção	Não Promoção	Total
Homem	21	3	24
Mulher	13	11	24
Total	34	14	48

- Apresente as hipóteses a serem testadas nesse caso para verificar se existe discriminação de gênero.
- Calcule as proporções de promoção dentro de cada gênero e verifique a diferença entre essas duas proporções. Parece haver indício de que homens são promovidos com mais frequência que mulheres?

- c) Faça um gráfico de barras que mostre a proporção de promoção para cada gênero. Visualmente, existe uma diferença entre gêneros?
- d) Faça um teste de homogeneidade. Apresente a estatística do teste qui-quadrado e o p-valor. Dica: use a função `chisq.test()` no R.
- e) Assumindo um nível de significância de $\alpha = 0.05$, existe evidência nos dados de que existe discriminação de gênero?

Problema 5 - Medidas dos Pés e Mãos

Para esse problema, iremos trabalhar com os comprimentos dos pés e das mãos de cada um de vocês. Você deve utilizar uma fita métrica ou qualquer outra ferramenta para medir o comprimento do seu pé e sua mão e o mesmo critério deve ser aplicado a ambos os lados (direito e esquerdo). A precisão não é importante, pois o interesse é apenas na medida relativa, ou seja, dizer que mão ou pé é maior.

Por favor, enviem seus dados preenchendo o formulário disponível no link:

<https://forms.gle/Tmoa5AkoA2DjZFVN9>

Nesse formulário temos informações sobre gênero, mão dominante (destro/canhoto), e qual pé e mão tem maior comprimento. A planilha com os dados será disponibilizada pelo Moodle na Atividade 07.

Com base nesses dados, façam o teste de associação apropriado entre os pares de variáveis em cada um dos itens abaixo, explicitando as hipóteses sendo testadas, estatística do teste, p-valor e conclusão.

- a) Tamanhos do pé e mão.
- b) Destro/canhoto e tamanho do pé.
- c) Gênero e tamanho do pé.
- d) Gênero e ser destro/canhoto.