Ensaios de Hipóteses Não Paramétricos

Estatística II - 2024/2025
ISCTE-IUL

Afonso Moniz Moreira[1,2]

[1]ISCTE-IUL, Departamento de Métodos Quantitativos para a Economia e Gestão
[2]CMVM - Comissão do Mercado de Valores Mobiliários, Departamento de Supervisão de Mercados

Aviso/Disclaimer

- Este conjunto de slides não é, nem pretende ser uma substituição à bibliografia principal da cadeira de Estatística II.
- Este conjunto de slides não é, nem pretende ser uma fonte rigorosa de estudo dos tópicos da cadeira.
- O único propósito deste conjunto de slides é ajudar o autor a guiar as aulas da forma mais coloquial possível sem ter de carregar formalismos desnecessários.
- Assim sendo, o formalismo estatístico é eliminado sempre que possível para agilizar uma primeira aprendizagem por parte dos estudantes.

Teoria dos Ensaios de Hipóteses Não Paramétricos I

Definição

- Então qual é exatamente a diferença entre um teste de hipótese paramétrico e um teste não paramétrico ?

Teoria dos Ensaios de Hipóteses Não Paramétricos I

Definição

- Então qual é exatamente a diferença entre um teste de hipótese paramétrico e um teste não paramétrico ?
- Os testes paramétricos permitem fazer inferência sobre um ou mais parâmetros que caracterizam uma ou várias populações.

Teoria dos Ensaios de Hipóteses Não Paramétricos I

Definição

- Então qual é exatamente a diferença entre um teste de hipótese paramétrico e um teste não paramétrico ?
- Os testes paramétricos permitem fazer inferência sobre um ou mais parâmetros que caracterizam uma ou várias populações.
- Contrariamente ao conceito de teste anterior, a definição de teste não paramétrico é um assunto que ainda gera discussão e discordância académica...

Teoria dos Ensaios de Hipóteses Não Paramétricos I

Definição

- Então qual é exatamente a diferença entre um teste de hipótese paramétrico e um teste não paramétrico ?
- Os testes paramétricos permitem fazer inferência sobre um ou mais parâmetros que caracterizam uma ou várias populações.
- Contrariamente ao conceito de teste anterior, a definição de teste não paramétrico é um assunto que ainda gera discussão e discordância académica...
- Tendo em consideração que a definição de teste não paramétrico é um conceito muito lato, seria mais fácil defini-lo como qualquer teste cujos objectos em análise não são parâmetros de uma população.

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:
  1. Pode ser utilizado com dados na escala **nominal**.

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:
  1. Pode ser utilizado com dados na escala **nominal**.
  2. Pode ser utilizado com dados na escala **ordinal**.

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:
  1. Pode ser utilizado com dados na escala **nominal**.
  2. Pode ser utilizado com dados na escala **ordinal**.
  3. Pode ser utilizado com dados na escala de **intervalos** ou **rácios**, desde que:

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:
  1. Pode ser utilizado com dados na escala **nominal**.
  2. Pode ser utilizado com dados na escala **ordinal**.
  3. Pode ser utilizado com dados na escala de **intervalos** ou **rácios**, desde que:
     a) A função de distribuição da variável aleatória que gera os dados **não está especificada**, ou

Teoria dos Ensaios de Hipóteses Não Paramétricos II

Definição

- No entanto, na bibliografia principal da cadeira, é apresentada a definição de Conover (1980):
- Um método estatístico diz-se **não-paramétrico** se satisfaz pelo menos uma das seguintes condições:
  1. Pode ser utilizado com dados na escala **nominal**.
  2. Pode ser utilizado com dados na escala **ordinal**.
  3. Pode ser utilizado com dados na escala de **intervalos** ou **rácios**, desde que:
     a) A função de distribuição da variável aleatória que gera os dados **não está especificada**, ou
     b) Está especificada a menos de um número **infinito de parâmetros desconhecidos**.

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)
  - Assumem que os dados seguem uma **distribuição específica** (por exemplo, normal).

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)
  - Assumem que os dados seguem uma **distribuição específica** (por exemplo, normal).
  - Dependem de um número **finito de parâmetros** (por exemplo, média μ e variância σ^2).

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)
  - Assumem que os dados seguem uma **distribuição específica** (por exemplo, normal).
  - Dependem de um número **finito de parâmetros** (por exemplo, média μ e variância σ^2).
  - Testes paramétricos exigem que os dados sejam numéricos e tenham propriedades que permitam operações aritméticas.

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)
  - Assumem que os dados seguem uma **distribuição específica** (por exemplo, normal).
  - Dependem de um número **finito de parâmetros** (por exemplo, média μ e variância σ^2).
  - Testes paramétricos exigem que os dados sejam numéricos e tenham propriedades que permitam operações aritméticas.
  - Pressupõe uma distribuição conhecida.

Teoria dos Ensaios de Hipóteses Não Paramétricos III

Definição

- Por que razão um teste paramétrico não se enquadra nesta definição? Porque os testes paramétricos...
  - Não são aplicáveis a escalas nominais (Azul, Verde, Vermelho) nem ordinais (Bom, Mau, Excelente)
  - Assumem que os dados seguem uma **distribuição específica** (por exemplo, normal).
  - Dependem de um número **finito de parâmetros** (por exemplo, média μ e variância σ^2).
  - Testes paramétricos exigem que os dados sejam numéricos e tenham propriedades que permitam operações aritméticas.
  - Pressupõe uma distribuição conhecida.
  - Trabalham com poucos parâmetros.

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.
- Ou seja, estão a responder à seguinte pergunta:

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.
- Ou seja, estão a responder à seguinte pergunta:
- Os dados da amostra aleatória podem ser bem ajustados, têm fit, à densidade que eu estou a estudar ?

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.
- Ou seja, estão a responder à seguinte pergunta:
- Os dados da amostra aleatória podem ser bem ajustados, têm fit, à densidade que eu estou a estudar ?
- As hipóteses do teste são formalizadas da seguinte maneira:

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.
- Ou seja, estão a responder à seguinte pergunta:
- Os dados da amostra aleatória podem ser bem ajustados, têm fit, à densidade que eu estou a estudar ?
- As hipóteses do teste são formalizadas da seguinte maneira:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$, ou seja:

Testes de Ajustamento

O que são ?

- Este tipo de testes também são denominados por testes de bondade de ajustamento, porquê ?
- Porque queremos perceber se uma determinada amostra aleatória, $(X_1, ..., X_n)$, pode ser considerada como proveniente de uma população com uma certa função densidade de probabilidade teórica $f_0(x)$.
- Ou seja, estão a responder à seguinte pergunta:
- Os dados da amostra aleatória podem ser bem ajustados, têm fit, à densidade que eu estou a estudar ?
- As hipóteses do teste são formalizadas da seguinte maneira:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$, ou seja:
- A densidade da população X, $f(x)$, é $f_0(x)$ VS não é $f_0(x)$.

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.
- Então o teste de ajustamento, tal como anteriormente mencionado, é dado por:

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.
- Então o teste de ajustamento, tal como anteriormente mencionado, é dado por:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.
- Então o teste de ajustamento, tal como anteriormente mencionado, é dado por:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$
- Como vai funcionar o teste do Qui-Quadrado?

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.
- Então o teste de ajustamento, tal como anteriormente mencionado, é dado por:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$
- Como vai funcionar o teste do Qui-Quadrado?
- Vou construir C classes de valores assumidos pela variável X, $A_1, A_2, ..., A_C$, de forma a que tenhamos uma partição.

Teste de Ajustamento do Qui-Quadrado

- Neste teste vamos considerar uma amostra aleatória $(X_1, ..., X_n)$ que foi retirada de uma população X cuja densidade $f(x)$ é desconhecida.
- Considerem também que temos uma densidade, $f_0(x)$, que é conhecida e que está totalmente especificada.
- Então o teste de ajustamento, tal como anteriormente mencionado, é dado por:
- $H_0 : f(x) = f_0(x)$ VS $H_0 : f(x) \neq f_0(x)$
- Como vai funcionar o teste do Qui-Quadrado?
- Vou construir C classes de valores assumidos pela variável X, $A_1, A_2, ..., A_C$, de forma a que tenhamos uma partição.
- Agora que temos as classes vou poder proceder ao cálculo de dois objectos.

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja
- O_i é o número de elementos da amostra que pertencem a A_i (frequências observadas).

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja
- O_i é o número de elementos da amostra que pertencem a A_i (frequências observadas).
- Mas também temos a densidade $f_0(x)$ de H_0, que é conhecida e que está totalmente especificada, logo posso calcular:

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja
- O_i é o número de elementos da amostra que pertencem a A_i (frequências observadas).
- Mas também temos a densidade $f_0(x)$ de H_0, que é conhecida e que está totalmente especificada, logo posso calcular:
- A probabilidade associada a cada classe se a verdadeira distribuição for $f_0(x)$:

$$P_i^* = \mathbb{P}[A_i | H_0 \text{ é verdadeira }]$$

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja
- O_i é o número de elementos da amostra que pertencem a A_i (frequências observadas).
- Mas também temos a densidade $f_0(x)$ de H_0, que é conhecida e que está totalmente especificada, logo posso calcular:
- A probabilidade associada a cada classe se a verdadeira distribuição for $f_0(x)$:

$$P_i^* = \mathbb{P}[A_i | H_0 \text{ é verdadeira }]$$

- Então a frequência absoluta de individuos, que deveria estar na classe A_i, sobe H_0, é:

Teste de Ajustamento do Qui-Quadrado I

- Usando a amostra aleatória $(X_1, ..., X_n)$ podemos calcular as frequência absolutas observadas, O_i de cada classe A_i, ou seja
- O_i é o número de elementos da amostra que pertencem a A_i (frequências observadas).
- Mas também temos a densidade $f_0(x)$ de H_0, que é conhecida e que está totalmente especificada, logo posso calcular:
- A probabilidade associada a cada classe se a verdadeira distribuição for $f_0(x)$:

$$P_i^* = \mathbb{P}[A_i | H_0 \text{ é verdadeira }]$$

- Então a frequência absoluta de individuos, que deveria estar na classe A_i, sobe H_0, é:

$$E_i = n \times P_i^*$$

Teste de Ajustamento do Qui-Quadrado II

- Se H_0 for verdadeira, então a diferença entre O_i e E_i $\forall i$ não pode ser muito grande... No entanto, tal como outros testes, o que é uma diferença grande ?

Teste de Ajustamento do Qui-Quadrado II

- Se H_0 for verdadeira, então a diferença entre O_i e E_i $\forall i$ não pode ser muito grande... No entanto, tal como outros testes, o que é uma diferença grande ?
- A estatística de teste que vai avaliar esta diferença será:

Teste de Ajustamento do Qui-Quadrado II

- Se H_0 for verdadeira, então a diferença entre O_i e E_i $\forall i$ não pode ser muito grande... No entanto, tal como outros testes, o que é uma diferença grande ?
- A estatística de teste que vai avaliar esta diferença será:

$$T = \sum_{i=1}^{C} \frac{(O_i - E_i)^2}{E_i} \xrightarrow{d} \chi^2_{(C-1)} \tag{1}$$

Teste de Ajustamento do Qui-Quadrado II

- Se H_0 for verdadeira, então a diferença entre O_i e E_i $\forall i$ não pode ser muito grande... No entanto, tal como outros testes, o que é uma diferença grande ?
- A estatística de teste que vai avaliar esta diferença será:

$$T = \sum_{i=1}^{C} \frac{(O_i - E_i)^2}{E_i} \xrightarrow{d} \chi^2_{(C-1)} \tag{1}$$

- Ou seja, para todas C classes, vamos avaliar a diferença entre C_i e o E_i, tomamos o quadrado porque o sinal de diferença não interessa, verificamos o peso desta diferença quadrada em E_i, e no final agrego tudo num único número.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:
- Rejeita-se H_0 a um nível α, caso o valor de T supere $\chi^2_{(C-1);\alpha}$, portanto caso $T \geq \chi^2_{(C-1);\alpha}$.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:
- Rejeita-se H_0 a um nível α, caso o valor de T supere $\chi^2_{(C-1);\alpha}$, portanto caso $T \geq \chi^2_{(C-1);\alpha}$.
- Este teste necessita de alguns pressupostos verificados sob pena da aproximação à chi-quadrado não ser eficaz:

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:
- Rejeita-se H_0 a um nível α, caso o valor de T supere $\chi^2_{(C-1);\alpha}$, portanto caso $T \geq \chi^2_{(C-1);\alpha}$.
- Este teste necessita de alguns pressupostos verificados sob pena da aproximação à chi-quadrado não ser eficaz:
  a) Não mais de 20% das classes com E_i inferior a 5.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:
- Rejeita-se H_0 a um nível α, caso o valor de T supere $\chi^2_{(C-1);\alpha}$, portanto caso $T \geq \chi^2_{(C-1);\alpha}$.
- Este teste necessita de alguns pressupostos verificados sob pena da aproximação à chi-quadrado não ser eficaz:
  a) Não mais de 20% das classes com E_i inferior a 5.
  b) Todas as classes com E_i superior ou igual a 1.

Teste de Ajustamento do Qui-Quadrado III

- Este número é comparado com o valor critico da qui-quadrado determinado pelo nível significância (α) que se pretende.
- Para grandes diferenças $T > 1$ para diferenças menores $T < 1$.
- Trata-se de um teste que é sempre unilateral direito e portanto:
- Rejeita-se H_0 a um nível α, caso o valor de T supere $\chi^2_{(C-1);\alpha}$, portanto caso $T \geq \chi^2_{(C-1);\alpha}$.
- Este teste necessita de alguns pressupostos verificados sob pena da aproximação à chi-quadrado não ser eficaz:
  a) Não mais de 20% das classes com E_i inferior a 5.
  b) Todas as classes com E_i superior ou igual a 1.
- Se estas regras não se verificarem então agregam-se as classes contíguas até se verificar a regra.

Exercício Nº1

O Recenseamento de 320 famílias com 5 filhos conduziu aos seguintes resultados:

Rapazes	5	4	3	2	1	0
Famílias	18	56	110	88	40	8

Verifique se estes resutlados são compatíveis com a hipótese do número de rapazes numa família de 5 filhos ser uma variável aleatória com distribuição binomial, admitindo a equiprobabilidade dos sexos, ao nível de significância de 0,01.

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:
- $X-$ Número de rapazes, numa família de 5 filhos. Onde $X \in \{0, 1, 2, 3, 4, 5\}$

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:
- $X-$ Número de rapazes, numa família de 5 filhos. Onde $X \in \{0, 1, 2, 3, 4, 5\}$
- Portanto as hipóteses em confronto são:

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:
- $X-$ Número de rapazes, numa família de 5 filhos. Onde $X \in \{0, 1, 2, 3, 4, 5\}$
- Portanto as hipóteses em confronto são:
- $H_0 : X \sim \mathcal{B}(n = 5, p = \frac{1}{2})$ VS $H_0 : X \nsim \mathcal{B}(n = 5, p = \frac{1}{2})$

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:
- $X-$ Número de rapazes, numa família de 5 filhos. Onde $X \in \{0, 1, 2, 3, 4, 5\}$
- Portanto as hipóteses em confronto são:
- $H_0 : X \sim \mathcal{B}(n = 5, p = \frac{1}{2})$ VS $H_0 : X \nsim \mathcal{B}(n = 5, p = \frac{1}{2})$
- Admite-se a equiprobabilidade dos sexos, logo $p = \frac{1}{2}$.

Exercício Nº1 - Resolução

- Portanto temos um teste de ajustamento do qui-quadrado, onde verificamos se os dados da amostra aleatória concretizada têm um bom ajustamento (i.e., têm fit) com uma distribuição Binomial com 5 tentativas e uma probabilidade de 50%, portanto $\mathcal{B}(5, \frac{1}{2})$.
- A variável para a qual estamos a realizar o ensaio é dada por:
- $X-$ Número de rapazes, numa família de 5 filhos. Onde $X \in \{0, 1, 2, 3, 4, 5\}$
- Portanto as hipóteses em confronto são:
- $H_0 : X \sim \mathcal{B}(n = 5, p = \frac{1}{2})$ VS $H_0 : X \nsim \mathcal{B}(n = 5, p = \frac{1}{2})$
- Admite-se a equiprobabilidade dos sexos, logo $p = \frac{1}{2}$.
- Vamos usar a amostra para obter uma estimativa da estatística de teste.

Exercício Nº1 - Resolução

- A estatística de teste é dada por:

$$T = \sum_{i=1}^{C} \frac{(O_i - E_i)^2}{E_i} \xrightarrow{d} \chi^2_{(C-1)}$$

- Calculam-se os valores da massa de probabilidade da binomial para valor de X:
- Sabemos que $P_i = \binom{5}{i} \cdot (0,5)^i \cdot (0,5)^{5-i} = \binom{5}{i} \cdot (0,5)^5 = \frac{\binom{5}{i}}{32}$
- $P_0 = \binom{5}{0} \cdot \frac{1}{32} = 1 \cdot \frac{1}{32} = \frac{1}{32} = 0,03125$
- $P_1 = \binom{5}{1} \cdot \frac{1}{32} = 5 \cdot \frac{1}{32} = \frac{5}{32} = 0,15625$
- $P_2 = \binom{5}{2} \cdot \frac{1}{32} = 10 \cdot \frac{1}{32} = \frac{10}{32} = 0,3125$
- $P_3 = \binom{5}{3} \cdot \frac{1}{32} = 10 \cdot \frac{1}{32} = \frac{10}{32} = 0,3125$
- $P_4 = \binom{5}{4} \cdot \frac{1}{32} = 5 \cdot \frac{1}{32} = \frac{5}{32} = 0,15625$
- $P_5 = \binom{5}{5} \cdot \frac{1}{32} = 1 \cdot \frac{1}{32} = \frac{1}{32} = 0,03125$

Exercício Nº1 - Resolução

- A tabela que resume todos os cálculos encontra-se de seguida:

X_i	O_i	P_i	$E_i = nP_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	8	0,0312	9,984	0,39426
1	40	0,1563	50,016	2,00576
2	88	0,3125	100,000	1,44
3	110	0,3125	100,000	1,44
4	56	0,1563	50,016	0,71594
5	18	0,0312	9,984	6,43592
Total	320	1,0000	320	**11,99188**

Exercício Nº1 - Resolução

- A tabela que resume todos os cálculos encontra-se de seguida:

X_i	O_i	P_i	$E_i = nP_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	8	0,0312	9,984	0,39426
1	40	0,1563	50,016	2,00576
2	88	0,3125	100,000	1,44
3	110	0,3125	100,000	1,44
4	56	0,1563	50,016	0,71594
5	18	0,0312	9,984	6,43592
Total	320	1,0000	320	**11,99188**

- A estatística de teste/obsevada/realizada é dada por: $T^* = 11,99$.

Exercício Nº1 - Resolução

- A tabela que resume todos os cálculos encontra-se de seguida:

X_i	O_i	P_i	$E_i = nP_i$	$\frac{(O_i-E_i)^2}{E_i}$
0	8	0,0312	9,984	0,39426
1	40	0,1563	50,016	2,00576
2	88	0,3125	100,000	1,44
3	110	0,3125	100,000	1,44
4	56	0,1563	50,016	0,71594
5	18	0,0312	9,984	6,43592
Total	320	1,0000	320	**11,99188**

- A estatística de teste/obsevada/realizada é dada por: $T^* = 11,99$.
- R.C. $= [15{,}1, +\infty[$ e R.A. $= [0, 15{,}1[$ para $\alpha = 0.01$.

Exercício Nº1 - Resolução

- A tabela que resume todos os cálculos encontra-se de seguida:

X_i	O_i	P_i	$E_i = nP_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	8	0,0312	9,984	0,39426
1	40	0,1563	50,016	2,00576
2	88	0,3125	100,000	1,44
3	110	0,3125	100,000	1,44
4	56	0,1563	50,016	0,71594
5	18	0,0312	9,984	6,43592
Total	320	1,0000	320	**11,99188**

- A estatística de teste/obsevada/realizada é dada por: $T^* = 11,99$.
- R.C. $= [15{,}1, +\infty[$ e R.A. $= [0, 15{,}1[$ para $\alpha = 0.01$.
- Como $T^* = 11,99 < 15,1$, não rejeitamos H_0. Os dados da amostra são compatíveis, ou têm um bom ajustamento, com uma binomial com $n = 5$ e $p = 0{,}5$.

Exercícios de Output

- Exercício Nº13 - Exercícios de Ouputs - Parte B

Teste de Kolmogorov-Smirnov I

- Neste teste pretende-se verificar se a distribuição empírica dos dados $F(x)$ pode ser ajustada a uma determinada distribuição $F_0(x)$.

Teste de Kolmogorov-Smirnov I

- Neste teste pretende-se verificar se a distribuição empírica dos dados $F(x)$ pode ser ajustada a uma determinada distribuição $F_0(x)$.
- Está especialmente arquitectado para dados contínuos não agrupados.

Teste de Kolmogorov-Smirnov I

- Neste teste pretende-se verificar se a distribuição empírica dos dados $F(x)$ pode ser ajustada a uma determinada distribuição $F_0(x)$.
- Está especialmente arquitectado para dados contínuos não agrupados.
- Por esta razão usa-se na verificação do ajustamento a uma distribuição normal.

Teste de Kolmogorov-Smirnov I

- Neste teste pretende-se verificar se a distribuição empírica dos dados $F(x)$ pode ser ajustada a uma determinada distribuição $F_0(x)$.
- Está especialmente arquitectado para dados contínuos não agrupados.
- Por esta razão usa-se na verificação do ajustamento a uma distribuição normal.
- O teste de Kolmogorov-Smirnov, tal como o teste da Qui-Quadrado, implica que a distribuição sob teste, $F_0(x)$, esteja totalmente especificada...

Teste de Kolmogorov-Smirnov I

- Neste teste pretende-se verificar se a distribuição empírica dos dados $F(x)$ pode ser ajustada a uma determinada distribuição $F_0(x)$.
- Está especialmente arquitectado para dados contínuos não agrupados.
- Por esta razão usa-se na verificação do ajustamento a uma distribuição normal.
- O teste de Kolmogorov-Smirnov, tal como o teste da Qui-Quadrado, implica que a distribuição sob teste, $F_0(x)$, esteja totalmente especificada...
- Nos outputs apresentados este teste cálcula os parâmetros à custa da amostra. É por isto que é aplicada a correção de Lilliefors ao nível de significância. Aparece nas notas de rodapé dos outputs do SPSS.

Teste de Kolmogorov-Smirnov II

- Portanto quando a distribuição não está bem parameterizada e é necessário usar a amostra para estimar os parâmetros, caso mais comum, o teste em causa é dado por:

Teste de Kolmogorov-Smirnov II

- Portanto quando a distribuição não está bem parameterizada e é necessário usar a amostra para estimar os parâmetros, caso mais comum, o teste em causa é dado por:
- $H_0 : F(x) = F_0(x)$ VS $H_1 : F(x) \neq F_0(x)$, em que $F_0(x) = \mathcal{N}(\bar{X}, S^2)$, $\bar{X}$ e S^2 correspondem à distribuição normal em teste, à média e variância amostrais, respectivamente.

Teste de Kolmogorov-Smirnov II

- Portanto quando a distribuição não está bem parameterizada e é necessário usar a amostra para estimar os parâmetros, caso mais comum, o teste em causa é dado por:
- $H_0 : F(x) = F_0(x)$ VS $H_1 : F(x) \neq F_0(x)$, em que $F_0(x) = \mathcal{N}(\bar{X}, S^2)$, $\bar{X}$ e S^2 correspondem à distribuição normal em teste, à média e variância amostrais, respectivamente.
- Decisão do teste vai ser feita com base nos outputs.

Teste de Kolmogorov-Smirnov II

- Portanto quando a distribuição não está bem parameterizada e é necessário usar a amostra para estimar os parâmetros, caso mais comum, o teste em causa é dado por:
- $H_0 : F(x) = F_0(x)$ VS $H_1 : F(x) \neq F_0(x)$, em que $F_0(x) = \mathcal{N}(\bar{X}, S^2)$, $\bar{X}$ e S^2 correspondem à distribuição normal em teste, à média e variância amostrais, respectivamente.
- Decisão do teste vai ser feita com base nos outputs.
- Deve ser usado para amostras grandes, no nosso caso, a regra prática é $N \geq 50$.

Teste de Kolmogorov-Smirnov II

- Portanto quando a distribuição não está bem parameterizada e é necessário usar a amostra para estimar os parâmetros, caso mais comum, o teste em causa é dado por:
- $H_0 : F(x) = F_0(x)$ VS $H_1 : F(x) \neq F_0(x)$, em que $F_0(x) = \mathcal{N}(\bar{X}, S^2)$, $\bar{X}$ e S^2 correspondem à distribuição normal em teste, à média e variância amostrais, respectivamente.
- Decisão do teste vai ser feita com base nos outputs.
- Deve ser usado para amostras grandes, no nosso caso, a regra prática é $N \geq 50$.
- Também é comum aparecerem os quantile-quantile (Q-Q) plots no teste de Kolmogorov-Smirnov. Este gráficos permitem avaliar os quantis observados contra os quantis esperados dados pela distribuição que estamos a ajustar.

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.
- Este teste faz uso dos quantis ao invés de usar a distribuição ou respectiva densidade como os anteriores.

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.
- Este teste faz uso dos quantis ao invés de usar a distribuição ou respectiva densidade como os anteriores.
- Portanto este teste é formalizado como o teste de ajustamento do qui-quadrado mas só permite avaliar para uma distribuição normal ao invés de qualquer distribução:

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.
- Este teste faz uso dos quantis ao invés de usar a distribuição ou respectiva densidade como os anteriores.
- Portanto este teste é formalizado como o teste de ajustamento do qui-quadrado mas só permite avaliar para uma distribuição normal ao invés de qualquer distribução:
- $H_0 : X \sim \mathcal{N}(\bar{X}, S^2)$ VS $H_1 : X \nsim \mathcal{N}(\bar{X}, S^2)$, em que $\bar{X}$ e S^2 correspondem à média e variância amostrais, respectivamente.

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.
- Este teste faz uso dos quantis ao invés de usar a distribuição ou respectiva densidade como os anteriores.
- Portanto este teste é formalizado como o teste de ajustamento do qui-quadrado mas só permite avaliar para uma distribuição normal ao invés de qualquer distribução:
- $H_0 : X \sim \mathcal{N}(\bar{X}, S^2)$ VS $H_1 : X \not\sim \mathcal{N}(\bar{X}, S^2)$, em que $\bar{X}$ e S^2 correspondem à média e variância amostrais, respectivamente.
- Decisão do teste vai ser feita com base nos outputs.

Teste de Shapiro-Wilk

- Ao contrário do teste Kolmogorov-Smirnov, o teste de Shapiro-Wilk só permite verificar a bondade de ajustamento à distribuição normal e não a qualquer distribuição $F_0(x)$.
- Este teste faz uso dos quantis ao invés de usar a distribuição ou respectiva densidade como os anteriores.
- Portanto este teste é formalizado como o teste de ajustamento do qui-quadrado mas só permite avaliar para uma distribuição normal ao invés de qualquer distribução:
- $H_0 : X \sim \mathcal{N}(\bar{X}, S^2)$ VS $H_1 : X \nsim \mathcal{N}(\bar{X}, S^2)$, em que $\bar{X}$ e S^2 correspondem à média e variância amostrais, respectivamente.
- Decisão do teste vai ser feita com base nos outputs.
- Deve ser usado para amostras pequenas, no nosso caso, a regra prática é $N < 50$.

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:
- H_0 : As diferentes amostras são provenientes de populações com a mesma distribuição. VS H_1 : Pelo menos uma das amostras é proveniente de uma população com distribuição diferente das restantes.

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:
- H_0 : As diferentes amostras são provenientes de populações com a mesma distribuição. VS H_1 : Pelo menos uma das amostras é proveniente de uma população com distribuição diferente das restantes.
- Atenção! As amostras são **provenientes de** uma população com uma determinada distribuição, não a distribuição da própria amostra!!

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:
- H_0 : As diferentes amostras são provenientes de populações com a mesma distribuição. VS H_1 : Pelo menos uma das amostras é proveniente de uma população com distribuição diferente das restantes.
- Atenção! As amostras são **provenientes de** uma população com uma determinada distribuição, não a distribuição da própria amostra!!
- Em geral, não interessa saber qual é a distribuição subjacente a cada população, apenas testar se é mesma para todas as amostras.

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:
- H_0 : As diferentes amostras são provenientes de populações com a mesma distribuição. VS H_1 : Pelo menos uma das amostras é proveniente de uma população com distribuição diferente das restantes.
- Atenção! As amostras são **provenientes de** uma população com uma determinada distribuição, não a distribuição da própria amostra!!
- Em geral, não interessa saber qual é a distribuição subjacente a cada população, apenas testar se é mesma para todas as amostras.
- Estes testes permitem relaxar as hipóteses subjacentes aos respectivos congéneres paramétricos.

Testes à igualdade de duas ou mais distribuições

- Os testes que vamos falar de seguida, de um modo geral, têm o objectivo de testar:
- H_0 : As diferentes amostras são provenientes de populações com a mesma distribuição. VS H_1 : Pelo menos uma das amostras é proveniente de uma população com distribuição diferente das restantes.
- Atenção! As amostras são **provenientes de** uma população com uma determinada distribuição, não a distribuição da própria amostra!!
- Em geral, não interessa saber qual é a distribuição subjacente a cada população, apenas testar se é mesma para todas as amostras.
- Estes testes permitem relaxar as hipóteses subjacentes aos respectivos congéneres paramétricos.
- Continuidade das variáveis, ainda que este pressuposto seja ignorado, e o conhecimento da distribuição da população.

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.
- Tal como mencionado anteriormente as hipóteses são dadas por:

Teste de Mann-Whitney I

- O teste de Mann-Whitey é uma das alternativas não paramétricas ao teste da diferença de duas médias:
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.
- Tal como mencionado anteriormente as hipóteses são dadas por:
- H_0: As duas amostras são provenientes de populações com a mesma distribuição: $F(x) = G(x), \forall x$
- H_1: As duas amostras são provenientes de populações com distribuições distintas: $\exists x : F(x) \neq G(x)$

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.
- Tal como mencionado anteriormente as hipóteses são dadas por:

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.
- Tal como mencionado anteriormente as hipóteses são dadas por:
- H_0: As duas amostras são provenientes de populações com a mesma distribuição: $F(x) = G(x), \forall x$
- H_1: As duas amostras são provenientes de populações com distribuições distintas: $\exists x : F(x) \neq G(x)$

Teste de Mann-Whitney II

- O modo como o teste está arquitectado, torna-o especialmente sensível às diferenças nas medianas das distribuições.
- Temos assim duas amostras aleatórias provenientes de duas populações diferentes:
- $X_1, X_2, \ldots, X_{n_1}$ da população X e com distribuição $F(x)$
- $Y_1, Y_2, \ldots, Y_{n_2}$ da população Y e com distribuição $G(x)$
- Sem qualquer perda de generalidade, assuma-se também que $n_1 < n_2$.
- Tal como mencionado anteriormente as hipóteses são dadas por:
- H_0: As duas amostras são provenientes de populações com a mesma distribuição: $F(x) = G(x), \forall x$
- H_1: As duas amostras são provenientes de populações com distribuições distintas: $\exists x : F(x) \neq G(x)$
- O teste de Mann-Whitney é especialmente sensível às diferenças de medianas entre as duas distribuições.

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.
- $H_0 : \theta_1 \geq \theta_2$ VS $H_1 : \theta_1 < \theta_2$ - Unilateral Esquerdo

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.
- $H_0 : \theta_1 \geq \theta_2$ VS $H_1 : \theta_1 < \theta_2$ - Unilateral Esquerdo
- Neste caso, estamos a avaliar, sobe a H_1, se a mediana da distribuição 1 é menor que a mediana da distribuição 2.

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.
- $H_0 : \theta_1 \geq \theta_2$ VS $H_1 : \theta_1 < \theta_2$ - Unilateral Esquerdo
- Neste caso, estamos a avaliar, sobe a H_1, se a mediana da distribuição 1 é menor que a mediana da distribuição 2.
- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente inferiores aos da população 2.

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.
- $H_0 : \theta_1 \geq \theta_2$ VS $H_1 : \theta_1 < \theta_2$ - Unilateral Esquerdo
- Neste caso, estamos a avaliar, sobe a H_1, se a mediana da distribuição 1 é menor que a mediana da distribuição 2.
- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente inferiores aos da população 2.
- $H_0 : \theta_1 \leq \theta_2$ VS $H_1 : \theta_1 > \theta_2$ - Unilateral Direito

Teste de Mann-Whitney III

- Assim sendo, designando θ_1 e θ_2 por a mediana da população 1 e da população 2, respectivamente, o teste pode ser formalizado por:
- $H_0 : \theta_1 = \theta_2$ VS $H_1 : \theta_1 \neq \theta_2$ - Bilateral
- Neste caso, estamos a avaliar se a mediana da distribuição 1 é igual à mediana da distribuição 2.
- $H_0 : \theta_1 \geq \theta_2$ VS $H_1 : \theta_1 < \theta_2$ - Unilateral Esquerdo
- Neste caso, estamos a avaliar, sobe a H_1, se a mediana da distribuição 1 é menor que a mediana da distribuição 2.
- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente inferiores aos da população 2.
- $H_0 : \theta_1 \leq \theta_2$ VS $H_1 : \theta_1 > \theta_2$ - Unilateral Direito
- Neste caso, estamos a avaliar, sobe a H_1, se a mediana da distribuição 1 é maior que a mediana da distribuição 2.

Teste de Mann-Whitney IV

- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente maiores aos da população 2.

Teste de Mann-Whitney IV

- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente maiores aos da população 2.
- Portanto a especificação correta para este teste pode ser qualquer uma das anteriores e vai depender do problema concreto que estamos a enfrentar.

Teste de Mann-Whitney IV

- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente maiores aos da população 2.
- Portanto a especificação correta para este teste pode ser qualquer uma das anteriores e vai depender do problema concreto que estamos a enfrentar.
- Decisão do teste vai ser feita com base em outputs.

Teste de Mann-Whitney IV

- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente maiores aos da população 2.
- Portanto a especificação correta para este teste pode ser qualquer uma das anteriores e vai depender do problema concreto que estamos a enfrentar.
- Decisão do teste vai ser feita com base em outputs.
- O teste tem uma estatística para amostras finitas, a estatística de Mann-Whitney U, e uma estatística assimptótica, a distribuição Normal. Cada uma com o respectivo p-value.

Teste de Mann-Whitney IV

- Ou seja, sobe a H_1, os valores da população 1 são tendencialmente maiores aos da população 2.
- Portanto a especificação correta para este teste pode ser qualquer uma das anteriores e vai depender do problema concreto que estamos a enfrentar.
- Decisão do teste vai ser feita com base em outputs.
- O teste tem uma estatística para amostras finitas, a estatística de Mann-Whitney U, e uma estatística assimptótica, a distribuição Normal. Cada uma com o respectivo p-value.
- Neste teste uma amostra pequena é quando os dois grupos não têm mais de 20 observações ou quando um dos grupos tem menos de 5 observações.

Teste de Mann-Whitney V

- Exercício Nº9 - Exercícios de Outputs SPSS - Parte B

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:
- Temos assim k amostras aleatórias provenientes de k populações diferentes:

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:
- Temos assim k amostras aleatórias provenientes de k populações diferentes:
- $X_{11}, X_{12}, \ldots, X_{1n_1}$ da população X e com distribuição $F_1(x)$

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:
- Temos assim k amostras aleatórias provenientes de k populações diferentes:
- $X_{11}, X_{12}, \ldots, X_{1n_1}$ da população X e com distribuição $F_1(x)$
- $X_{21}, X_{22}, \ldots, X_{2n_2}$ da população Y e com distribuição $F_2(x)$

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:
- Temos assim k amostras aleatórias provenientes de k populações diferentes:
- $X_{11}, X_{12}, \ldots, X_{1n_1}$ da população X e com distribuição $F_1(x)$
- $X_{21}, X_{22}, \ldots, X_{2n_2}$ da população Y e com distribuição $F_2(x)$
- ...

Teste de Kruskall-Wallis I

- Este teste pode ser visto como uma alternativa não-paramétrica à ANOVA.
- Corresponde a uma generalização do teste de Mann-Whitney para $k > 2$ populações, portanto:
- Temos assim k amostras aleatórias provenientes de k populações diferentes:
- $X_{11}, X_{12}, \ldots, X_{1n_1}$ da população X e com distribuição $F_1(x)$
- $X_{21}, X_{22}, \ldots, X_{2n_2}$ da população Y e com distribuição $F_2(x)$
- ...
- $X_{k1}, X_{k2}, \ldots, X_{kn_k}$ da população Y e com distribuição $F_k(x)$

Teste de Kruskall-Wallis II

- Este teste tem a mesma sensibilidade face à posição dos dados que o teste de Mann-Whitney, pelo que têm a mesma sensibilidade à mediana. Logo a formalização do teste é:

Teste de Kruskall-Wallis II

- Este teste tem a mesma sensibilidade face à posição dos dados que o teste de Mann-Whitney, pelo que têm a mesma sensibilidade à mediana. Logo a formalização do teste é:
- $H_0 : \theta_1 = \theta_2 = ... = \theta_k$ VS $H_1 : \exists_{i,j} : \theta_i \neq \theta_j$, em que θ_j é a mediana da distribuição j-ésima população.

Teste de Kruskall-Wallis II

- Este teste tem a mesma sensibilidade face à posição dos dados que o teste de Mann-Whitney, pelo que têm a mesma sensibilidade à mediana. Logo a formalização do teste é:
- $H_0 : \theta_1 = \theta_2 = ... = \theta_k$ VS $H_1 : \exists_{i,j} : \theta_i \neq \theta_j$, em que θ_j é a mediana da distribuição j-ésima população.
- Como se pode verificar pela especificação, este teste tem semelhanças com o teste ANOVA.

Teste de Kruskall-Wallis II

- Este teste tem a mesma sensibilidade face à posição dos dados que o teste de Mann-Whitney, pelo que têm a mesma sensibilidade à mediana. Logo a formalização do teste é:
- $H_0 : \theta_1 = \theta_2 = ... = \theta_k$ VS $H_1 : \exists_{i,j} : \theta_i \neq \theta_j$, em que θ_j é a mediana da distribuição j-ésima população.
- Como se pode verificar pela especificação, este teste tem semelhanças com o teste ANOVA.
- Decisão do teste vai ser feita com base em outputs.

Teste de Kruskall-Wallis III

- Exercício Nº10 - Exercícios de Outputs SPSS - Parte B

Teste de Independência do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.

Teste de Independência do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.
- Podem ser usados com quaisquer tipos de dados, desde que possam ser transformados em categorias, como é o caso dos dados quantitativos discretos e contínuos.

Teste de Independência do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.
- Podem ser usados com quaisquer tipos de dados, desde que possam ser transformados em categorias, como é o caso dos dados quantitativos discretos e contínuos.
- Vamos então considerar uma amostra aleatória de dimensão n classificada segundo duas variáveis qualitativas/categóricas, a primeira com r categorias e segundo com c categorias:

Teste de Indepêndencia do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.
- Podem ser usados com quaisquer tipos de dados, desde que possam ser transformados em categorias, como é o caso dos dados quantitativos discretos e contínuos.
- Vamos então considerar uma amostra aleatória de dimensão n classificada segundo duas variáveis qualitativas/categóricas, a primeira com r categorias e segundo com c categorias:
- $A_i, i = 1, ..., r$ é uma categoria da primeira variável.

Teste de Independência do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.
- Podem ser usados com quaisquer tipos de dados, desde que possam ser transformados em categorias, como é o caso dos dados quantitativos discretos e contínuos.
- Vamos então considerar uma amostra aleatória de dimensão n classificada segundo duas variáveis qualitativas/categóricas, a primeira com r categorias e segundo com c categorias:
- $A_i, i = 1, ..., r$ é uma categoria da primeira variável.
- $B_j, j = 1, ..., c$ é uma categoria da segunda variável.

Teste de Indepêndencia do Qui-Quadrado I

- Serve para avaliar a dependência entre duas variáveis categóricas/qualitativas... Ou seja duas variáveis que podem ter várias categorias.
- Podem ser usados com quaisquer tipos de dados, desde que possam ser transformados em categorias, como é o caso dos dados quantitativos discretos e contínuos.
- Vamos então considerar uma amostra aleatória de dimensão n classificada segundo duas variáveis qualitativas/categóricas, a primeira com r categorias e segundo com c categorias:
- $A_i, i = 1, ..., r$ é uma categoria da primeira variável.
- $B_j, j = 1, ..., c$ é uma categoria da segunda variável.
- $O_{i,j}$ é a frequência absoluta de casos, na amostra, para os quais se observou A_i na 1a variável e B_j na 2a variável.

Teste de Independência do Qui-Quadrado II

Variável 1 $\downarrow \backslash$ **Variável 2** $\rightarrow$	B_1	B_2	$\cdots$	B_c	**Total (linha)**
A_1	$O_{1,1}$	$O_{1,2}$	$\cdots$	$O_{1,c}$	$\sum_{j=1}^{c} O_{1,j}$
A_2	$O_{2,1}$	$O_{2,2}$	$\cdots$	$O_{2,c}$	$\sum_{j=1}^{c} O_{2,j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	$O_{r,1}$	$O_{r,2}$	$\cdots$	$O_{r,c}$	$\sum_{j=1}^{c} O_{r,j}$
Total (coluna)	$\sum_{i=1}^{r} O_{i,1}$	$\sum_{i=1}^{r} O_{i,2}$	$\cdots$	$\sum_{i=1}^{r} O_{i,c}$	$\sum_{i=1}^{r}\sum_{j=1}^{c} O_{i,j} = n$

Tabela: Tabela de Contigência para a Variável 1 e 2

- Vamos supor que a probabilidade de um elemento da amostra pertencer a A_i e B_j simultaneamente é p_{ij}.

Teste de Independência do Qui-Quadrado II

Variável 1 $\downarrow \backslash$ **Variável 2** $\rightarrow$	B_1	B_2	$\cdots$	B_c	**Total (linha)**
A_1	$O_{1,1}$	$O_{1,2}$	$\cdots$	$O_{1,c}$	$\sum_{j=1}^{c} O_{1,j}$
A_2	$O_{2,1}$	$O_{2,2}$	$\cdots$	$O_{2,c}$	$\sum_{j=1}^{c} O_{2,j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	$O_{r,1}$	$O_{r,2}$	$\cdots$	$O_{r,c}$	$\sum_{j=1}^{c} O_{r,j}$
Total (coluna)	$\sum_{i=1}^{r} O_{i,1}$	$\sum_{i=1}^{r} O_{i,2}$	$\cdots$	$\sum_{i=1}^{r} O_{i,c}$	$\sum_{i=1}^{r}\sum_{j=1}^{c} O_{i,j} = n$

Tabela: Tabela de Contigência para a Variável 1 e 2

- Vamos supor que a probabilidade de um elemento da amostra pertencer a A_i e B_j simultaneamente é p_{ij}.
- Então, o número de individuos da amostra que se espera observar no par (A_i, B_j) é: $E_{ij} = n \times p_{ij}$.

Teste de Independência do Qui-Quadrado II

Variável 1 $\downarrow \backslash$ **Variável 2** $\rightarrow$	B_1	B_2	$\cdots$	B_c	**Total (linha)**
A_1	$O_{1,1}$	$O_{1,2}$	$\cdots$	$O_{1,c}$	$\sum_{j=1}^{c} O_{1,j}$
A_2	$O_{2,1}$	$O_{2,2}$	$\cdots$	$O_{2,c}$	$\sum_{j=1}^{c} O_{2,j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	$O_{r,1}$	$O_{r,2}$	$\cdots$	$O_{r,c}$	$\sum_{j=1}^{c} O_{r,j}$
Total (coluna)	$\sum_{i=1}^{r} O_{i,1}$	$\sum_{i=1}^{r} O_{i,2}$	$\cdots$	$\sum_{i=1}^{r} O_{i,c}$	$\sum_{i=1}^{r}\sum_{j=1}^{c} O_{i,j} = n$

Tabela: Tabela de Contigência para a Variável 1 e 2

- Vamos supor que a probabilidade de um elemento da amostra pertencer a A_i e B_j simultaneamente é p_{ij}.
- Então, o número de individuos da amostra que se espera observar no par (A_i, B_j) é: $E_{ij} = n \times p_{ij}$.
- Usando as distribuições marginais também posso obter as frequências absolutas esperadas em cada classe: $E_i = n \times p_i$ e $E_j = n \times p_j$.

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$

Teste de Indepêndencia do Qui-Quadrado III

- Se existir indepêndencia entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Indepêndencia:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$
- VS
- $H_1 : \exists_{i,j} : P_{ij} \neq p_i \times p_j, i \neq j, i = 1, ..., r; \quad j = 1, ..., c$

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$
- VS
- $H_1 : \exists_{i,j} : P_{ij} \neq p_i \times p_j, i \neq j, i = 1, ..., r; \quad j = 1, ..., c$
- Ou por outro lado, o que estamos a avaliar é:

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$
- VS
- $H_1 : \exists_{i,j} : P_{ij} \neq p_i \times p_j, i \neq j, i = 1, ..., r; \quad j = 1, ..., c$
- Ou por outro lado, o que estamos a avaliar é:
- H_0 : A variável 1 e a variável 2, **são** independentes.

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$
- VS
- $H_1 : \exists_{i,j} : P_{ij} \neq p_i \times p_j, i \neq j, i = 1, ..., r; \quad j = 1, ..., c$
- Ou por outro lado, o que estamos a avaliar é:
- H_0 : A variável 1 e a variável 2, **são** independentes.
- VS

Teste de Independência do Qui-Quadrado III

- Se existir independência entre as duas variáveis então:
- $P_{ij} = p_i \times p_j$
- Então podemos usar o O_{ij} e o E_{ij} para avaliar a Independência:
- $H_0 : P_{ij} = p_i \times p_j, \forall i = 1, ..., r; \quad j = 1, ..., c$
- VS
- $H_1 : \exists_{i,j} : P_{ij} \neq p_i \times p_j, i \neq j, i = 1, ..., r; \quad j = 1, ..., c$
- Ou por outro lado, o que estamos a avaliar é:
- H_0 : A variável 1 e a variável 2, **são** independentes.
- VS
- H_1 : A variável 1 e a variável 2, **não são** independentes.

Teste de Independência do Qui-Quadrado IV

- A estatística de teste é dada por:

Teste de Independência do Qui-Quadrado IV

- A estatística de teste é dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

Teste de Indepêndencia do Qui-Quadrado IV

- A estatística de teste é dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

- Na verdade, não se conhecem as verdadeiras probabilidades marginais envolvidas, então temos de as estimar:

Teste de Indepêndencia do Qui-Quadrado IV

- A estatística de teste é dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

- Na verdade, não se conhecem as verdadeiras probabilidades marginais envolvidas, então temos de as estimar:

$$\hat{p}_i = \frac{o_i}{n} \;\text{ e }\; \hat{p}_j = \frac{o_j}{n} \tag{3}$$

Teste de Independência do Qui-Quadrado IV

- A estatística de teste é dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

- Na verdade, não se conhecem as verdadeiras probabilidades marginais envolvidas, então temos de as estimar:

$$\hat{p}_i = \frac{o_i}{n} \ \text{ e } \ \hat{p}_j = \frac{o_j}{n} \tag{3}$$

- Sob H_0, $E_{ij} = n \times p_i \times p_j$, é estimado por:

Teste de Independência do Qui-Quadrado IV

- A estatística de teste é dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

- Na verdade, não se conhecem as verdadeiras probabilidades marginais envolvidas, então temos de as estimar:

$$\hat{p}_i = \frac{o_i}{n} \;\text{ e }\; \hat{p}_j = \frac{o_j}{n} \tag{3}$$

- Sob H_0, $E_{ij} = n \times p_i \times p_j$, é estimado por:

$$e_{ij} = \hat{E}_{ij} = n \times \hat{p}_i \times \hat{p}_j = n \times \frac{o_i}{n} \times \frac{o_j}{n} \Longleftrightarrow e_{ij} = \frac{o_i \times o_j}{n} \tag{4}$$

Teste de Indepêndencia do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

Teste de Independência do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \overset{approx}{\sim} \chi^2_{(r-1)\times(c-1)} \tag{5}$$

Teste de Indepêndencia do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \overset{approx}{\sim} \chi^2_{(r-1)\times(c-1)} \tag{5}$$

- Este teste é sempre unilateral direito.

Teste de Independência do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \overset{approx}{\sim} \chi^2_{(r-1)\times(c-1)} \tag{5}$$

- Este teste é sempre unilateral direito.
- Devo rejeitar H_0 quando as diferenças quadradas entre o observado e o esperado atingirem um peso demasiado elevado.

Teste de Independência do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \overset{approx}{\sim} \chi^2_{(r-1)\times(c-1)} \tag{5}$$

- Este teste é sempre unilateral direito.
- Devo rejeitar H_0 quando as diferenças quadradas entre o observado e o esperado atingirem um peso demasiado elevado.
- Tal como noutros testes, o que é muito ou pouco é balizado pelo nível de significância que por sua vez estabelece o nível crítico.

Teste de Independência do Qui-Quadrado V

- Então a estatística de teste vai ser dada por:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \overset{approx}{\sim} \chi^2_{(r-1)\times(c-1)} \qquad (5)$$

- Este teste é sempre unilateral direito.
- Devo rejeitar H_0 quando as diferenças quadradas entre o observado e o esperado atingirem um peso demasiado elevado.
- Tal como noutros testes, o que é muito ou pouco é balizado pelo nível de significância que por sua vez estabelece o nível crítico.
- Portanto, rejeito H_0 quando $X^2 > \chi_{(r-1)(c-1);\alpha}$, onde α é o nível de significância.

Teste de Indepêndencia do Qui-Quadrado VI

- Tal como no teste de aderência do Qui-Quadrado, é necessário que se verifiquem certas condições, de forma a que os resultados obtidos sejam fidedignos...

Teste de Independência do Qui-Quadrado VI

- Tal como no teste de aderência do Qui-Quadrado, é necessário que se verifiquem certas condições, de forma a que os resultados obtidos sejam fidedignos...
- Para que o teste do Qui-Quadrado de Independência seja aplicável é necessário:

Teste de Independência do Qui-Quadrado VI

- Tal como no teste de aderência do Qui-Quadrado, é necessário que se verifiquem certas condições, de forma a que os resultados obtidos sejam fidedignos...
- Para que o teste do Qui-Quadrado de Independência seja aplicável é necessário:
    - Não mais de 20% das células tenham frequência absoluta esperada estimada inferior a 5.

Teste de Indepêndencia do Qui-Quadrado VI

- Tal como no teste de aderência do Qui-Quadrado, é necessário que se verifiquem certas condições, de forma a que os resultados obtidos sejam fidedignos...
- Para que o teste do Qui-Quadrado de Indepêndencia seja aplicável é necessário:
  - Não mais de 20% das células tenham frequência absoluta esperada estimada inferior a 5.
  - Não exista qualquer célula com valor esperado inferior a 1, ou seja, $e_{ij} \geq 1, \forall i, j$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

(a)

- Temos duas variáveis categóricas que são os tipos de dados que o teste de independência do qui-quadrado utiliza.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

(a)

- Temos duas variáveis categóricas que são os tipos de dados que o teste de independência do qui-quadrado utiliza.
- Todas as células têm uma frequência absoluta esperada estimada superior a 5 e logo também não existe nenhuma célula com uma frequência absoluta estimada esperada inferior a 1.

(b)

- H_0 : As variáveis em análise são independentes, isto é, não há uma relação entre sexo e preferência de semanário.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

(a)

- Temos duas variáveis categóricas que são os tipos de dados que o teste de independência do qui-quadrado utiliza.
- Todas as células têm uma frequência absoluta esperada estimada superior a 5 e logo também não existe nenhuma célula com uma frequência absoluta estimada esperada inferior a 1.

(b)

- H_0 : As variáveis em análise são independentes, isto é, não há uma relação entre sexo e preferência de semanário.
- H_1 : As variáveis em análise são dependentes, isto é, há uma relação entre sexo e preferência de semanário.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

(a)

- Temos duas variáveis categóricas que são os tipos de dados que o teste de independência do qui-quadrado utiliza.
- Todas as células têm uma frequência absoluta esperada estimada superior a 5 e logo também não existe nenhuma célula com uma frequência absoluta estimada esperada inferior a 1.

(b)

- H_0 : As variáveis em análise são independentes, isto é, não há uma relação entre sexo e preferência de semanário.
- H_1 : As variáveis em análise são dependentes, isto é, há uma relação entre sexo e preferência de semanário.
- O P-value tem de ser observado para a linha **Pearson Chi-Square**.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

(a)

- Temos duas variáveis categóricas que são os tipos de dados que o teste de independência do qui-quadrado utiliza.
- Todas as células têm uma frequência absoluta esperada estimada superior a 5 e logo também não existe nenhuma célula com uma frequência absoluta estimada esperada inferior a 1.

(b)

- H_0 : As variáveis em análise são independentes, isto é, não há uma relação entre sexo e preferência de semanário.
- H_1 : As variáveis em análise são dependentes, isto é, há uma relação entre sexo e preferência de semanário.
- O P-value tem de ser observado para a linha **Pearson Chi-Square**.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- O valor crítico para o teste em causa seria de $\chi^2_{(3-1)\times(2-1);0.05} = 5.99$.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- O valor crítico para o teste em causa seria de $\chi^2_{(3-1)\times(2-1);0.05} = 5.99$.
- A estatística de teste observada/realizada passa largamente o valor crítico originando um p-value (0.004) consideravelmente baixo.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- O valor crítico para o teste em causa seria de $\chi^2_{(3-1)\times(2-1);0.05} = 5.99$.
- A estatística de teste observada/realizada passa largamente o valor crítico originando um p-value (0.004) consideravelmente baixo.

(c)

- Este teste tem sempre uma região crítica unilateral direita. Pretende-se verificar quando existem discrepâncias entre valores observados e o esperados. O sinal dessa discrepância não interessa.

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- O valor crítico para o teste em causa seria de $\chi^2_{(3-1)\times(2-1);0.05} = 5.99$.
- A estatística de teste observada/realizada passa largamente o valor crítico originando um p-value (0.004) consideravelmente baixo.

(c)

- Este teste tem sempre uma região crítica unilateral direita. Pretende-se verificar quando existem discrepâncias entre valores observados e o esperados. O sinal dessa discrepância não interessa.
- Vamos proceder ao cálculo da principal componente: As estatísticas esperadas:

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- O valor crítico para o teste em causa seria de $\chi^2_{(3-1)\times(2-1);0.05} = 5.99$.
- A estatística de teste observada/realizada passa largamente o valor crítico originando um p-value (0.004) consideravelmente baixo.

(c)

- Este teste tem sempre uma região crítica unilateral direita. Pretende-se verificar quando existem discrepâncias entre valores observados e o esperados. O sinal dessa discrepância não interessa.
- Vamos proceder ao cálculo da principal componente: As estatísticas esperadas:

$$e_{ij} = \frac{(\text{Total da linha}_i) \times (\text{Total da coluna}_j)}{\text{Total geral}}$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

$$\text{(Expresso, Feminino)} : e_{11} = \frac{49 \times 33}{99} = \frac{1617}{99} \approx 16{,}33$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

$$\text{(Expresso, Feminino)} : e_{11} = \frac{49 \times 33}{99} = \frac{1617}{99} \approx 16{,}33$$

$$\text{(Expresso, Masculino)} : e_{12} = \frac{49 \times 66}{99} = \frac{3234}{99} \approx 32{,}67$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

$$\text{(Expresso, Feminino)} : e_{11} = \frac{49 \times 33}{99} = \frac{1617}{99} \approx 16{,}33$$

$$\text{(Expresso, Masculino)} : e_{12} = \frac{49 \times 66}{99} = \frac{3234}{99} \approx 32{,}67$$

$$\text{(Semanário, Feminino)} : e_{21} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

$$\text{(Expresso, Feminino)} : e_{11} = \frac{49 \times 33}{99} = \frac{1617}{99} \approx 16{,}33$$

$$\text{(Expresso, Masculino)} : e_{12} = \frac{49 \times 66}{99} = \frac{3234}{99} \approx 32{,}67$$

$$\text{(Semanário, Feminino)} : e_{21} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

$$\text{(Semanário, Masculino)} : e_{22} = \frac{25 \times 66}{99} = \frac{1650}{99} \approx 16{,}67$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

- i = linha (semanário)
- j = coluna (sexo)
- Total geral = 99

$$\text{(Expresso, Feminino)} : e_{11} = \frac{49 \times 33}{99} = \frac{1617}{99} \approx 16{,}33$$

$$\text{(Expresso, Masculino)} : e_{12} = \frac{49 \times 66}{99} = \frac{3234}{99} \approx 32{,}67$$

$$\text{(Semanário, Feminino)} : e_{21} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

$$\text{(Semanário, Masculino)} : e_{22} = \frac{25 \times 66}{99} = \frac{1650}{99} \approx 16{,}67$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

$$\text{(Sol, Feminino)} : e_{31} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

$$\text{(Sol, Feminino)} : e_{31} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

$$\text{(Sol, Masculino)} : e_{32} = \frac{25 \times 66}{99} = \frac{1650}{99} \approx 16{,}67$$

Exercício Nº12 - Exercícios de Outputs SPSS - Parte B

$$(\text{Sol, Feminino}) : e_{31} = \frac{25 \times 33}{99} = \frac{825}{99} \approx 8{,}33$$

$$(\text{Sol, Masculino}) : e_{32} = \frac{25 \times 66}{99} = \frac{1650}{99} \approx 16{,}67$$

- Então a estatística de teste observada/realizada é dada por:

$$\chi^2 = \frac{(13-16{,}33)^2}{16{,}33} + \frac{(36-32{,}67)^2}{32{,}67} + \frac{(5-8{,}33)^2}{8{,}33}$$

$$+\frac{(20-16{,}67)^2}{16{,}67} + \frac{(15-8{,}33)^2}{8{,}33} + \frac{(10-16{,}67)^2}{16{,}67} = \boxed{\approx 11{,}02}$$