Testes Não Paramétricos

Nos testes abordados até agora, ditos testes paramétricos, as hipóteses envolvem apenas parâmetros populacionais, como a média, a variância, uma proporção, etc. Além disso, em geral, estes testes comportam uma diversidade de suposições fortes a que o seu emprego deve subordinar-se de que são exemplo:

- as observações devem ser extraídas de populações com distribuição especificada;
- as variáveis em estudo devem ser medidas em escala intervalar ou de rácios, de modo a que seja possível utilizar operações aritméticas sobre os valores obtidos das amostras (adição, multiplicação, ...),
- etc.

Vamos agora abordar um conjunto de testes que nos permitem testar outro tipo de hipóteses que não apenas sobre parâmetros populacionais (e.g., se a distribuição populacional em estudo pode ser considerada Normal). Estes são chamados testes não paramétricos.

Estes testes são, em geral, fáceis de aplicar, pois podem ser usados quando as hipóteses exigidas por outras técnicas não são satisfeitas. Apesar de haver certas suposições básicas associadas à maioria das provas não paramétricas, essas suposições são em menor número e mais fracas do que as associadas às provas paramétricas. A maior parte das provas não paramétricas servem para pequenas amostras e, além disso, aplicam-se a dados medidos em escala ordinal, e alguns mesmo a dados em escala nominal.

Tabelas de Contingência

Teste do Qui-quadrado de Independência

Suponha que numa amostra aleatória de tamanho ***n*** de uma dada população são observados **dois atributos** ou **características** **A** e **B** (qualitativas ou quantitativas), uma com ***r*** e outra com ***s*** modalidades ou categorias, respectivamente $A_1, A_2, \ldots, A_r$ e $B_1, B_2, \ldots, B_s$.

Cada indivíduo da amostra é classificado numa e numa só categoria (ou classe) de **A** e numa e numa só categoria (ou classe) de **B**.

A classificação dos elementos da amostra dá origem a uma tabela de dupla entrada, designada por **tabela de contingência r×s**, com o seguinte aspecto:

	B_1	B_2	...	B_s
A_1	O_{11}	O_{12}	...	O_{1s}
A_2	O_{21}	O_{22}	...	O_{2s}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_r	O_{r1}	O_{r2}	...	O_{rs}

$\mathbf{O_{ij}}$ (i=1,...,r e j=1,...,s) → número de elementos classificados simultaneamente nas categorias A_i de A e B_j de B, numa amostra de tamanho ***n***.

Sejam:

- $O_{i\bullet} = \sum_{j=1}^{s} O_{ij}$ (i=1,...,r) → nº de elementos na amostra com modalidade A_i;

- $O_{\bullet j} = \sum_{i=1}^{r} O_{ij}$ (j=1,...,s) → nº de elementos na amostra com modalidade B_j.

Tem-se,

$$\boxed{n = \sum_{i=1}^{r}\sum_{j=1}^{s} O_{ij} = \sum_{i=1}^{r} O_{i\bullet} = \sum_{j=1}^{s} O_{\bullet j}}$$

onde ***n*** é a dimensão da amostra.

O objectivo a que nos propomos é o de tentar inferir sobre a existência ou não de qualquer relação ou associação entre os atributos (variáveis) A e B, mais concretamente, inferir se A e B são ou não independentes.

Hipóteses a testar:

H_0: A e B são independentes

H_1: A e B não são independentes

Denote-se por:

- $p_{ij}=P(A_i \cap B_j)$ ($i=1,..,r$ e $j=1,...,s$) a probabilidade (desconhecida) de um indivíduo da população ser classificado simultaneamente nas categorias A_i de A e B_j de B;

- $p_{i\bullet}=P(A_i)$ ($i=1,...,r$) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria A_i de A;

- $p_{\bullet j}=P(B_j)$ ($j=1,...,s$) a probabilidade (desconhecida) de um indivíduo da população ser classificado na categoria B_j de B.

$$\boxed{1=\sum_{i=1}^{r}\sum_{j=1}^{s}p_{ij}=\sum_{i=1}^{r}p_{i\bullet}=\sum_{j=1}^{s}p_{\bullet j}}.$$

Ora, se os atributos são independentes, verifica-se a conhecida relação,

$$P(A_i \cap B_j) = P(A_i)\, P(B_j),$$

isto é,

$$\mathbf{p_{ij} = p_{i\bullet} \times p_{\bullet j}}$$

Assim, as hipóteses anteriores podem ser formuladas do seguinte modo:

H_0: $p_{ij} = p_{i\bullet} \times p_{\bullet j}$ (para todo i e j)

H_1: $p_{ij} \neq p_{i\bullet} \times p_{\bullet j}$ (para algum $i \neq j$).

Os verdadeiros valores das probabilidades $p_{i\bullet}$ e $p_{\bullet j}$ são estimadas, a partir dos dados amostrais, por

$$\boxed{\hat{p}_{i\bullet} = \frac{O_{i\bullet}}{n}} \quad \text{e} \quad \boxed{\hat{p}_{\bullet j} = \frac{O_{\bullet j}}{n}},$$

Notação: $e_{ij}=n\, p_{ij}$ → número esperado de indivíduos na classe A_i de A e B_j de B.

Quando H_0 é verdadeira, i.e, $p_{ij}= p_{i\bullet}\times p_{\bullet j}$, temos

$$e_{ij}=n\, p_{ij} = n\, p_{i\bullet}\times p_{\bullet j} \xrightarrow{\textit{estimado por}} \boxed{\hat{e}_{ij} = n\ \hat{p}_{i\bullet}\, \hat{p}_{\bullet j} = \frac{O_{i\bullet}\times O_{\bullet j}}{n}}$$

A estatística do teste de independência é então:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(O_{ij}-\hat{e}_{ij})^2}{\hat{e}_{ij}},$$

que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assimptótica do Qui-quadrado com (r-1)(s-1) graus de liberdade.

Vimos que quando H_0 é verdadeira e_{ij} pode ser estimado por $\boxed{\hat{e}_{ij} = n\,\hat{p}_{i\bullet}\,\hat{p}_{\bullet j}}$. Logo, a diferença entre O_{ij} (frequência observada) e $\hat{e}_{ij}$ (estimativa da frequência esperada supondo a independência) não deve ser grande.

Assim, a estatística teste, mede o afastamento dos dados em relação à hipótese de independência. Trata-se então de um **teste unilateral à direita**.

Exemplo: Um supermercado quer testar ao nível de significância de 5% a hipótese de que o modo de pagamento dos clientes nesse estabelecimento é independente do período do dia em que fazem as compras. Existem três modos de efectuar os pagamentos: por cheque, dinheiro e cartão de débito/crédito.

A seguinte tabela de contingência 3×3 apresenta os resultados obtidos numa amostra de 4000 clientes:

	PERÍODO DO DIA		
MODO DE PAGAMENTO	**Manhã**	**Tarde**	**Noite**
Cheque	750	1500	750
Dinheiro	125	300	75
Cartão de débito/Crédito	125	200	175

Denotando por A o atributo *Modo de pagamento* e por B o atributo *Período do dia em que faz as compras*, as hipóteses as testar são

H_0: A e B são independentes

H_1: A e B não são independentes

Uma vez que A e B assumem cada uma 3 modalidades, sob H_0, a estatística teste tem distribuição assimptótica do Qui-quadrado com (r-1)(s-1)=(3-1)(3-1)= 4 graus de liberdade.

Ao nível de significância de 0.05, a região crítica é então [9.49, +∞[(consulte tabela ou faça no EXCEL: INV.CHI(0,05;4)).

	PERÍODO DO DIA			
MODO DE PAGAMENTO	Manhã	Tarde	Noite	**Totais**
Cheque	750	1500	750	3000
Dinheiro	125	300	75	500
Cartão de Crédito	125	200	175	500
Totais	1000	2000	1000	4000

Cálculo das frequências esperadas: $\hat{e}_{ij} = n\ \hat{p}_{i\bullet}\ \hat{p}_{\bullet j} = n \dfrac{O_{i\bullet}}{n} \dfrac{O_{\bullet j}}{n} = \dfrac{O_{i\bullet} O_{\bullet j}}{n}$

$\hat{e}_{11} = (3000 \times 1000)/4000 = 750$

$\hat{e}_{12} = (3000 \times 2000)/4000 = 1500$

$\hat{e}_{13} = (3000 \times 1000)/4000 = 750.$

$\vdots$

Frequências esperadas

	PERÍODO DO DIA			
MODO DE PAGAMENTO	Manhã	Tarde	Noite	**Totais**
Cheque	750	1500	750	3000
Dinheiro	125	250	125	500
Cartão de Crédito	125	250	125	500
Totais	1000	2000	1000	4000

Valor observado da estatística teste:

$$\chi^2{}_{obs} = \frac{(750-750)^2}{750} + \frac{(1500-1500)^2}{1500} + \ldots + \frac{(200-250)^2}{250} + \frac{(175-125)^2}{125} = 60.$$

Uma vez que 60 excede o valor crítico 9.49, ao nível de significância de 0.05, rejeitamos a hipótese de que o modo de pagamento é independente do período do dia em que as compras são feitas.

Medidas de Associação

No teste do Qui-Quadrado apresentado, se for rejeitada a hipótese de independência entre os atributos, pode interessar medir a intensidade da associação entre os mesmos, através de uma medida adequada.

Uma vez que a estatística do teste mede o afastamento em relação à hipótese de independência, o seu valor observado poderá ser usado para avaliar o grau de associação entre os atributos.

Coeficiente de Contingência de Pearson: $C = \sqrt{\dfrac{\chi^2}{\chi^2 + n}}$

Este coeficiente varia entre 0 e $\sqrt{(q-1)/q}$ onde q=min{r,s} e portanto nunca assume o valor 1.

Valores pequenos de C indicam fraca associação entre os atributos, enquanto que valores grandes de C indicam forte associação.

O facto deste coeficiente não assumir o valor 1 no caso de associação completa é uma sua limitação. Para obviar este problema, Tshuprow propôs o seguinte coeficiente.

Coeficiente de Tshuprow: $T = \sqrt{\dfrac{\chi^2}{n\sqrt{(r-1)\times(s-1)}}}$

Este coeficiente varia entre 0 e 1, tomando o valor 0 no caso de existir independência e o valor 1 quando r=s e houver associação completa.

Por último, referimos o coeficiente proposto por Cramer que atinge o valor 1 quando há associação completa.

Coeficiente V de Cramer: $V = \sqrt{\dfrac{\chi^2}{n(q-1)}}$, com $q=\min\{r,s\}$ $\quad 0 \leq V \leq 1$.

Para o exemplo anterior, rejeitámos a hipótese de independência entre o modo de pagamento e o período do dia em que as compras eram efectuadas.

Para ter uma ideia da intensidade de associação entre estes dois atributos, calculam-se os coeficientes que acabámos de descrever.

Coeficiente de Contingência de Pearson: $C = \sqrt{\dfrac{\chi^2}{\chi^2 + n}} = \sqrt{\dfrac{60}{60+4000}} = 0.122$

$0 \leq C \leq \sqrt{(q-1)/q}$, onde q=min{r,s}=3, i.e, $0 \leq C \leq 0.816$.

Coeficiente de Tshuprow: $T=\sqrt{\dfrac{\chi^2}{n\sqrt{(r-1)\times(s-1)}}}=\sqrt{\dfrac{60}{4000\sqrt{2\times 2}}}$=0.087

Coeficiente V de Cramer: $V=\sqrt{\dfrac{\chi^2}{n(q-1)}}=\sqrt{\dfrac{60}{4000\times 2}}$=0.087

Verificamos, então, que apesar de haver associação entre os atributos, esta pode considerar-se fraca.

Teste de Homogeneidade

Suponha que são recolhidas amostras aleatórias de **s** **populações** (sub-populações ou estratos) $B_1, B_2, \ldots, B_s$, nas quais se observa **um atributo** A com **r** categorias $A_1, A_2, \ldots, A_r$.

Neste contexto, surge também uma tabela de contingência r×s:

	B_1	B_2	$\ldots$	B_s
A_1	O_{11}	O_{12}	$\ldots$	O_{1s}
A_2	O_{21}	O_{22}	$\ldots$	O_{2s}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_r	O_{r1}	O_{r2}	$\ldots$	O_{rs}

$\mathbf{O_{ij}}$ (i=1,...,r e j=1,...,s) → número de elementos da amostra da **população** B_j classificados na categoria A_i de A.

Sejam:

- $\mathbf{O_{i\bullet}} = \sum_{j=1}^{s} O_{ij}$ (i=1,...,r) → nº de elementos na categoria A_i de A em todas as amostras;

- $\mathbf{O_{\bullet j}} = \sum_{i=1}^{r} O_{ij}$ (j=1,...,s) → tamanho da amostra recolhida na população B_j.

Neste caso, cada B_j rotula uma sub-população cujos elementos se distribuem pelas **r** modalidades do atributo A, e o que se pretende saber é **se existe homogeneidade**, isto é, **se não há diferença entre as populações no modo como os seus elementos se distribuem pelas modalidades do atributo A**.

À semelhança do teste de independência, a estatística do teste é

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(O_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}},$$

que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assimptótica do Qui-Quadrado com (r-1)(s-1) graus de liberdade.

Valores muito grandes da estatística de teste traduzem um grande afastamento dos dados em relação à hipótese nula, conduzindo à rejeição desta. Assim, **a estatística de teste mede o afastamento dos dados em relação à hipótese de homogeneidade**.

Para aplicar os testes de independência e de homogeneidade devem ser seguidas as seguintes regras:

- **mais de 20% das frequências esperadas, e_i's, inferiores a 5**

ou,

- **alguma frequência esperada inferior a 1**

devemos proceder à agregação de algumas classes contíguas.

Teste exacto de Fisher

O teste do Qui-quadrado é, como já se disse, baseado numa distribuição assimptótica, o que portanto limita a sua aplicação ao caso de grandes amostras (recorde as limitações sobre as frequências esperadas).

Em tabelas de contingência 2x2, existe uma alternativa ao teste do Qui-quadrado, o *teste de Fisher*, que é um teste exacto, i.e., a distribuição da estatística é exacta (os pontos críticos e *valores-p* são calculados de forma exacta).