Resumo 12 – Testes de hipóteses não paramétricos

Os métodos não-paramétricos fazem poucas suposições sobre a natureza das distribuições dos dados. Não exige que as distribuições nas populações sejam normais, nem são baseados em parâmetros (estimadores) da população. Os testes não-paramétricos são denominados métodos de distribuição livre (*free distribution*).

Vantagens: As técnicas não-paramétricas têm diversas vantagens sobre os métodos paramétricos de inferência estatística. Uma delas é que não incorporam todas as suposições restritivas características dos testes paramétricos, nem exigem que as populações sejam normalmente distribuídas. Os testes não paramétricos, em geral, exigem apenas que a variável em estudo seja ordenável, podendo ser aplicado também para variáveis semiquantitativas e qualitativas. Além disso, por tratarem de ordenações em vez de valores reais das observações, os testes paramétricos podem ser realizados de modo relativamente rápido para pequenas amostras. O uso de postos torna as técnicas não-paramétricas menos sensíveis aos erros de medidas do que os testes paramétricos, permitindo assim o uso de ordinais em vez de dados contínuos.

Desvantagens: Se as suposições básicas de um teste paramétrico são satisfeitas, os testes não-paramétricos são menos poderosos do que a técnica paramétrica correspondente, o que significa que, se a hipótese nula for falsa, o teste não-paramétrico exigirá uma amostra maior para fornecer evidências suficientes para rejeitá-la. Além disso, hipóteses testadas por testes não-paramétricos tendem a ser menos específicas do que as testadas por testes tradicionais. Como contam com ordenações em vez de valores das observações, os testes não paramétricos não usam toda informação disponível sobre a distribuição. Finalmente, se uma grande proporção de observações são empatadas os desvios-padrão, superestimam os valores da estatística de teste, e correções precisam ser adicionadas aos cálculos.

12.1. Teste de aderência (Qui-quadrado)

1) útil para verificar se a distribuição das freqüências observadas dos dados se ajusta a um modelo teórico pré-determinado;
2) recomendado para amostras grandes ($n > 50$) e tem por finalidade comparar se as freqüências observadas na amostra estão próximas das freqüências esperadas para um modelo de probabilidade proposto (normal, binomial, poisson, etc);
3) se as freqüências esperadas não diferirem estatisticamente das freqüências observadas, pode-se inferir que a característica em estudo da população seguem a distribuição proposta; de outra maneira, possui distribuição diversa.

O teste que mede a eficiência do ajuste da distribuição, ou seja, o quanto a freqüência observada está próxima da freqüência esperada, daí o nome de aderência, é o teste de Qui-quadrado (χ^2) .

As hipóteses a serem testadas são:

H_0 : As freqüências observadas não diferem das freqüências esperadas em relação à distribuição proposta, ou seja, a característica em estudo da população segue a distribuição proposta.

vs

H_A : As freqüências observadas diferem das freqüências esperadas em relação à distribuição proposta,ou seja, a característica em estudo da população não segue a distribuição proposta.

A estatística de teste, que tem distribuição de Qui-quadrado com v = k – p – 1 graus de liberdade (sendo k = ao número de classes da distribuição de freqüência e p = ao número de parâmetros estimados), supondo H_0 verdadeira, é dada por:

$$\chi^2 = \sum\left[\frac{(O_i - E_i)^2}{E_i}\right],$$

onde

O_i : freqüências observadas na i-ésima classe;

$E_i = n \cdot p_i$: freqüências esperadas na i-ésima classe, supondo H_0 verdadeira, sendo pi é a probabilidade de ocorrência na classe considerada.

A hipótese nula (H_0), será rejeitada, ao nível de significância estipulado, quando $\chi^2_{obs} > \chi^2_{(v;\alpha)}$, caso contrário, não rejeita-se H_0.

Exemplo 12.1: Em uma certa população, 100 descendentes estudados forneceram os seguintes dados: 26 indivíduos com genótipo AA, 45 com o genótipo Aa e 29 com o genótipo aa. Verificar se o modelo genético proposto é adequado para essa população. ($\chi^2_{obs} = 1{,}18$).

Exemplo 12.2: Levantou-se uma amostra de 100 indivíduos de uma população em que se observou a altura das pessoas. A amostra forneceu os seguintes dados (em cm):

Altura	Freqüência
150 |- 155	1
155 |- 160	2
160 |- 165	5
165 |- 170	13
170 |- 175	20
175 |- 180	23
180 |- 185	19
185 |- 190	11
190 |- 195	4
195 |- 200	2

Verificar se os dados se ajustam (aderem) à uma distribuição normal com um nível de 2,5% de significância, considerando as estimativas da média e do desvio padrão populacional iguais a 176,85 cm e 8,75 cm, respectivamente. ($\chi^2_{obs} = 13{,}4125$)

12.2. Teste de independência (qui-quadrado)

Em pesquisas ou levantamentos feitos por meio de entrevistas, questionários ou fichas, quando o pesquisador classifica os indivíduos amostrados segundo duas ou mais variáveis qualitativas categóricas ou ordinais, a apresentação tabular das freqüências observadas pode ser feita através de uma *tabela de contingência*, isto é, uma tabela com duas ou mais entradas, cada entrada relativa a uma das variáveis.

Com a tabela de contingência, conseguimos uma maneira conveniente de fazer descrição dos dados da amostra quando temos duas ou mais variáveis qualitativas a considerar. Passamos agora à análise dos dados fornecidos pela tabela.

Uma indagação que pode ser objeto de um teste bastante simples é se as variáveis qualitativas envolvidas são ou não independentes. Ou seja, podemos desejar testar as hipóteses:

H_0: as variáveis são independentes

vs

H_A: as variáveis não são independentes, ou seja, elas apresentam algum grau de associação entre si.

A estatística de teste, que tem distribuição de Qui-quadrado com v graus de liberdade, supondo H_0 verdadeira, é dada por:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left[\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right],$$

onde

r : o número de linhas do corpo da tabela;

c : o número de colunas do corpo da tabela;

O_{ij} : a freqüência observada na intersecção da linha i com a coluna j;

E_{ij} : a freqüência esperada na intersecção da linha i com a coluna j;

As freqüências esperadas de cada cela da tabela, supondo H_0 verdadeira, são dadas por:

$$E_{ij} = \frac{x_{i+} \cdot x_{+j}}{n},$$

onde

x_{i+} é a freqüência total observada na linha i;

x_{+j} é a freqüência total observada na coluna j;

$n = \sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}$.

A hipótese nula (H_0), será rejeitada, ao nível de significância estipulado, quando $\chi^2_{obs} > \chi^2_{(v;\alpha)}$, caso contrário, não rejeita-se H_0.

A determinação do número de graus de liberdade com que a estatística de teste deverá ser testada, pode ser feita verificando-se quantas das freqüências observadas Oij permanecem “livres” após a determinação das freqüências esperadas. Ora, estas foram determinadas com base na fixação dos totais marginais. Então, respeitados esses totais, o número de valores O_{ij} com grau de liberdade será

$$v = (r - 1)\ (c - 1),$$

pois fatalmente a última freqüência observada a ser considerada em cada linha ou coluna estará determinada pelo total fixado da linha ou coluna, o que equivale a ter uma linha e uma coluna sem graus de liberdade.

Exemplo 12.3: Uma amostra de pacientes com câncer de mama de aparência maligna, mas com um pequeno grau de reação inflamatória, foi classificada segundo uma variável qualitativa ordinal: a faixa etária por ocasião do diagnóstico (menos de 50 anos, entre 50 e 70 anos e mais de 70 anos) e uma variável qualitativa categórica: sobrevivência após três anos (sim e não). Os dados estão apresentados

Faixa Etária	Sobrevivência		Total
	Sim	Não	
Menos de 50 anos	11	6	17
Entre 50 e 70 anos	18	8	26
Mais de 70 anos	15	9	24
Total	44	23	67

Exemplo 12.4: Muitas vezes, uma das variáveis praticamente representa uma classificação dos elementos em populações distintas. Por exemplo, considere uma amostra de 100 pessoas, que foram entrevistadas quanto a suas opiniões sobre um determinado projeto de lei, tendo sido obtidos os resultados:

Sexo	Opinião			Totais
	Favorável	Desfavorável	Indiferente	
Homens	33	12	15	60
Mulheres	7	20	13	40
Totais	40	32	28	100

12.3. Teste de Wilcoxon para duas amostras dependentes

Alternativa para o teste t. Embora não exista nenhuma suposição sobre a forma da distribuição das observações, o teste é mais poderoso para distribuições simétricas. Hipótese formuladas em função das medianas.

H_0 : não existe efeito do tratamento (a diferença mediana é igual a zero ou se a distribuição for simétrica a diferença entre antes e depois é iguas a zero)

vs

H_A : existe efeito do tratamento (a diferença mediana é diferente de zero ou se a distribuição for simétrica a diferença entre antes e depois é diferente de zero)

Para pequenas amostras (até 25 pares), o p-valor deve ser obtido através de uma tabela especial não incluída nesse texto.

Para grandes amostras procedimento é:

1) para cada par, determinar a diferença (di) entre os escores, considerando o sinal da diferença;
2) considerar as diferenças em valor absoluto, ordená-las e atribuir postos a cada uma delas. No caso de empates, calcular a média dos postos e atribuir, a cada diferença (di) empatada, o posto médio;
3) atribuir a cada posto o sinal positivo (+) ou negativo (-) da respectiva diferença (di);
4) calcular a soma dos postos com sinal negativo e a soma dos postos com sinal positivo. A menor dessas somas, em valor absoluto, atribuir o valor da estatística T;
5) contar o número de pares onde as diferenças (di) são diferente de zero e considerar este número como n (pares onde ocorrem empates não contribuem para a diferença entre amostras);
6) se n > 25 (grandes amostras), a estatística do teste (T) tem distribuição aproximadamente normal com média $\mu_T = \frac{(n(n+1))}{4}$ e desvio padrão $\sigma_T = \sqrt{\left(\frac{(n(n+1)(2n+1))}{24}\right)}$, de modo que
$$Z = \frac{(T - \mu_T)}{\sigma_T} \sim N(0;\ 1);$$
7) calcular o valor de Z e verificar sua significância.

Exemplo 12.5: Ratos submetidos situações estressantes desenvolvem úlceras estomacais. O número de pontos de úlcera e o tamanho desses pontos são indicadores da maior ou menor gravidade do problema. Suponha uma situação em que os tamanhos das úlceras sejam mais ou menos uniformes de tal sorte que o número de pontos seja um indicador razoável da gravidade. Para verificar o efeito de dois tipos de estresse térmico, um prolongado a uma temperatura de 5ºC (G1) e outro agudo a uma temperatura de 0ºC (G2), foram utilizados 30 pares de ratos gêmeos, sendo cada elemento do par de gêmeos sorteado para um dos dois grupos. Foram obtidos os resultados apresentado na tabela a seguir. Verifique se existe diferença entre os grupos.

G1	G2
8	10
3	3
0	0
21	20
4	4
0	0
10	6
30	26
8	7
1	0
40	35
17	14
6	1
7	4
18	19
1	0
4	5
95	90
8	0
7	5
2	0
38	36
24	27
5	7
57	56
15	11
60	52
10	8
7	4
15	16

12.4. Teste U de Mann Whitney para duas amostras independentes

Alternativa para o teste t.

Para grandes amostras procedimento é:

1) considerar n_1 como o tamanho amostral do menor grupo e n_2 o tamanho amostral do maior grupo (se $n_1 \, diferente \, de \, n_2$) e n = $n_1 + n_2$;
2) atribuir postos à distribuição conjunta dos valores dos dois grupos. Estes postos irão de 1 até n se não ocorrerem empates. No caso de ocorrência de empates, utilizar a média dos postos;
3) calcular a soma dos postos nos dois grupos (R_1 e R_2);
4) calcular as estatísticas U_1 e U_2, dadas por:
$U_1 = n_1 n_2 + \frac{(n_1(n_1+1))}{2} - R_1$ e $U_2 = n_1 n_2 + \frac{(n_1(n_1+1))}{2} - R_2$;
5) denomina-se U como o menor valor entre U_1 e U_2;
6) para grande amostras (pelo menos 10 em cada grupo), a estatística do teste U tem distribuição aproximadamente normal com média $\mu_U = \frac{(n_1 n_2)}{2}$ e desvio padrão $\sigma_U = \sqrt{\left(n_1 n_2 \frac{(n_1 n_2 + 1)}{12}\right)}$, de modo que $Z = \frac{(U - \mu_U)}{\sigma_U}$ ~N(0; 1);
7) calcular o valor de Z e verificar sua significância.

Exemplo 12.6: A avaliação clínica de pacientes submetidos a cirurgia foi feita considerando a variável estado geral do paciente, que pode assumir os valores: 1(péssimo), 2(ruim), 3(regular), 4(bom) e 5(ótimo). Para comparar dois tipos de cirurgia, 40 pacientes homogêneos quanto à gravidade da doença e outros possíveis fatores interferentes (idade, sexo) foram utilizados, sendo 20 aleatoriamente designados a cada grupo. Um dos pacientes de G1 não foi avaliado. Foram obtidos os dados apresentados na tabela a seguir.

G1	G2
1	1
1	1
2	1
2	2
2	3
2	3
2	3
3	3
3	4
3	4
3	4
3	4
3	4
3	4
3	4
3	5
3	5
5	5
5	5
	5