Formulário de Estatística Não Paramétrica

Luís Israel

3 de julho de 2018

1. Binomial

Teste exato para amostras com variáveis binárias.
$H0 : p = p_0 ...$
$H1 : p \neq p_0 ...$

$$P(X = i) = \binom{n}{i} p_0^i (1 - p_0)^{n-i}$$

2. Runs

Teste para amostras com variáveis binárias.
$H0$: a sequência foi produzida aleatoriamente
$H1$: a sequência não foi produzida aleatoriamente

1. Se necessário, classificar o vetor como valores positivos e negativos, normalmente em relação à mediana
2. Contar o o número de sequências r, o número de valores positivos $n1$ e o número de valores negativos $n2$
3. Comparar r com seus respectivos valores críticos na tabela binomial

3. Independência qui-quadrado

Teste de independência entre amostras não pareadas.
$H0$: as variáveis são independentes
$H1$: as variáveis são dependentes
$\phi = \sqrt{\frac{\chi^2}{n}}$
$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$
$T = \sqrt{\frac{\phi^2}{\sqrt{(r-1)(c-1)}}}$
$V = \sqrt{\frac{\chi^2}{nt}},\ t = \min(r-1, c-1)$

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

$$E_{i,j} = \frac{n_i n_j}{n}$$

$$\chi_0^2 > \chi^2_{\alpha,(r-1)(c-1)}$$

4. McNemar

Teste de homogeneidade de frequências marginais para amostras pareadas.
$H0 : p_b = p_c$
$H1 : p_b \neq p_c$

T1	T2	Total	
B1	a	b	a + b
B2	c	d	c + d
Total	a + c	b + d	n

$$\chi_0^2 = \frac{(b - c)^2}{b + c}$$

$$\chi_0^2 > \chi^2_{\alpha,1}$$

5. Sinal

Teste de comparação entre amostras pareadas.
$H0 : p = p_0 ...$
$H1 : p \neq p_0 ...$

1. Calcular a diferença entre os vetores
2. Classificar o novo vetor como valores positivos e negativos
3. Fazer o teste binomial com o novo vetor

Wilcoxon

Teste de comparação entre amostras pareadas.
$H0$: as amostras são de populações iguais
$H1$: as amostras são de populações diferentes

1. Para $i = 1, ..., N$ calcular $|x_{2,i} - x_{1,i}|$ e $sgn(x_{2,i} - x_{1,i})$
2. Excluir pares onde $|x_{2,i} - x_{1,i}| = 0$. N_r é o tamanho da amostra reduzida
3. Ordenar os pares e atribuir um rank R_i a eles

$$W = \sum_{i=1}^{N_r} sgn(x_{2,i} - x_{1,i}) R_i$$

$$|W| > W_{\alpha, N_r}$$

6. Q de Cochran

Teste de comparação entre amostras pareadas com variáveis binárias.
$H0$: os tratamentos tem efeitos iguais
$H1$: os tratamentos tem efeitos diferentes

Tratamento	1 Tratamento	2 Tratamento	k
Bloco 1	x_11	x_12	x_1k
Bloco 2	x_21	x_22	x_2k
Bloco b	x_b1	x_b2	x_bk

$$T = k(k-1) \frac{\sum_{j=1}^{k} (X_{oj} - \frac{N}{k})^2}{\sum_{i=1}^{b} X_{io}(k - X_{io})}$$

$$T > \chi^2_{\alpha, k-1}$$

onde
k é o número de tratamentos
X_{oj} é o número total de colunas para o j^o tratamento
b é o número de blocos
X_{io} é o número total de linhas para o i^o bloco
N é o total absoluto

7. Mann-Whitney

Teste de comparação entre dois grupos de amostras não pareadas.
$H0$: as amostras são de populações iguais
$H1$: as amostras são de populações diferentes

Ordenar os dados de todos os grupos de 1 a N. Em caso de empate, o rank dos itens é a média dos ranks que receberiam caso não houvesse empate. O menor U é a estatística usada no teste.

$$U_i = R_i - \frac{n_i(n_i + 1)}{2}$$

onde
R_i é a soma dos ranks no grupo i
n_i é o número de observações no grupo i

8. H de Kruskall-Wallis

Teste de comparação entre i grupos de amostras não pareadas.
$H0$: as amostras são de populações iguais
$H1$: as amostras são de populações diferentes

Ordenar os dados de todos os grupos de 1 a N. Em caso de empate, o rank dos itens é a média dos ranks que receberiam caso não houvesse empate.

$$H = (N-1) \frac{\sum_{i=1}^{g} n_i (\bar{r}_{io} - \bar{r})^2}{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

$$H > \chi^2_{\alpha, g-1}$$

onde
n_i é o número de observações no grupo i
r_{ij} é o rank da observação j do grupo i
N é o número total de observações entre os grupos
$\bar{r}_{io} = \frac{\sum_{j=i}^{n_i} r_{ij}}{n_i}$ é o rank médio de todas as observações no grupo i
$\bar{r} = \frac{1}{2}(N+1)$ é a média de todos os r_{ij}

Mediana de Mood

Teste de comparação entre amostras não pareadas.
$H0$: as amostras são de populações com medianas iguais
$H1$: as amostras são de populações com medianas diferentes

1. Calcular a mediana de todos os dados.
2. Dividir cada grupo em duas categorias: uma com a contagem de itens menores ou iguais à mediana; outra com a contagem de itens maiores que a mediana.
3. Formatar numa tabela de contingência e realizar o teste qui quadrado.

9. Friedman

Teste de comparação entre amostras pareadas.
$H0$: os tratamentos tem efeitos iguais
$H1$: os tratamentos tem efeitos diferentes

Dada uma matriz $(x_{ij})_{n\times k}$ com n blocos e k tratamentos, calcular os ranks *dentro* de cada bloco. Em caso de empate, o rank dos itens é a média dos ranks que receberiam caso não houvesse empate. Criar uma nova matriz $(r_{ij})_{n\times k}$ onde cada r_{ij} é um rank.

$$Q = \frac{SS_t}{SS_e}$$

$$Q > \chi^2_{\alpha,k-1}$$

onde
$SS_t = n\sum_{j=1}^{k}(\bar{r}_{\cdot j} - \bar{r})^2$
$SS_e = \frac{1}{n(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}(r_{ij} - \bar{r})^2$
$\bar{r}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$
$\bar{r} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} r_{ij}$

10. Coeficiente de Spearman

Coeficiente de correlação entre rankings de variáveis.
$H0 : \rho = 0$
$H1 : \rho \neq 0$

Ordenar as variáveis em rankings de acordo com o grupo. A segunda fórmula só funciona quando não há empates.

$$\rho = r_s = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

11. Siegel-Tukey

Teste de comparação para amostras não pareadas.
$H0 : \sigma_A^2 = \sigma_B^2$
$H1 : \sigma_A^2 > \sigma_B^2$

1. Ordenar os dados de todos os grupos de 1 a N de forma alternada (rank 1 é o menor, 2 e 3 são os dois maiores, 4 e 5 são os próximos menores, etc)
2. Aplicar o teste de Mann-Whitney

12. Coeficiente de Kendall

Coeficiente de correlação entre rankings de variáveis.
$H0 : \tau = 0$
$H1 : \tau \neq 0$

$$\tau = \frac{1}{n(n-1)}\sum_{i\neq j} sgn(x_i - x_j)(y_i - y_j)$$

Coeficiente Kappa

Coeficiente de concordância entre classificadores.
$H0 : \tau = 0$
$H1 : \tau \neq 0$
p_o : concordância relativa observada
p_e : concordância relativa esperada

$$\kappa = 1 - \frac{1 - p_o}{1 - p_e}$$

13. Kolmogorov-Smirnor

Ordenar os dados no formato $\{Y_1 < \cdots < Y_n\}$.

$$D = max_{1 \leq i \leq N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right)$$

Teste para distribuições contínuas de uma dimensão. A estatística deve ser comparada aos valores críticos de uma tabela própria.

$H0$: os dados seguem a distribuição especificada

$H1$: os dados não seguem a distribuição especificada

F é a função da distribuição cumulativa teórica

Anderson-Darling

Ordenar os dados no formato $\{Y_1 < \cdots < Y_n\}$.

$$A^2 = -n - S$$

$$S = \sum_{i=1}^{n} \frac{2i-1}{n}[\ln(\Phi(Y_i)) + \ln(1 - \Phi(Y_{n+1-i}))]$$

Teste para distribuições contínuas de uma dimensão. A estatística deve ser comparada aos valores críticos da distribuição teórica.

$H0$: os dados seguem a distribuição especificada

$H1$: os dados não seguem a distribuição especificada

Referências

Material do prof Quinino

https://de.wikipedia.org/wiki/Binomialtest

https://itl.nist.gov/div898/handbook/eda/section3/eda35d.htm

https://en.wikipedia.org/wiki/Chi-squared_test

https://en.wikipedia.org/wiki/McNemar%27s_test

https://en.wikipedia.org/wiki/Sign_test

https://en.wikipedia.org/wiki/Cochran%27s_Q_test

https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance

https://en.wikipedia.org/wiki/Median_test

https://www.isixsigma.com/tools-templates/hypothesis-testing/understanding-uses-moods-median-test/

https://en.wikipedia.org/wiki/Friedman_test

https://en.wikipedia.org/wiki/Siegel%E2%80%93Tukey_test

https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

https://en.wikipedia.org/wiki/Siegel%E2%80%93Tukey_test

https://en.wikipedia.org/wiki/Cohen%27s_kappa

http://uregina.ca/~gingrich/ch11a.pdf

https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

https://en.wikipedia.org/wiki/Anderson%E2%80%93Darling_test

https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test

https://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test