Introdução à Estatística Não Paramétrica

Anderson Rodrigo da Silva

Instituto Federal Goiano

Conteúdo

Generalidades e uso

- Em geral os métodos são simples e versáteis
- Há pouca ou nenhuma especificação ou exigência acerca dos parâmetros da população da qual a amostra foi obtida
- Dispensam, por exemplo: normalidade, homocedasticidade
- São especialmente úteis em situações em que é difícil estabelecer uma escala quantitativa (variáveis ordinais)
- São em geral mais poderosos que os paramétricos quando as exigências do modelo não são atendidas

Restrições

- Em geral não levam em conta a magnitude dos dados
- Em geral não nos permite testar interações
- Requerem, frequentemente, o uso de tabelas um pouco mais complexas

Intervalo de confiança para uma proporção

- O **Teste Binomial** é aplicado a amostras de variáveis binárias, por exemplo: presença/ausência, férteis/inférteis, macho/fêmea, etc.
- Os dados consistem do resultado de n tentativas repetidas e independentes de Bernoulli (sucesso/fracasso).
- O objetivo é avaliar a probabilidade π de sucesso.

Método

- Considere X a variável aletória que representa o número de sucessos em n tentativas independentes
- Dizemos que X tem distribuição binomial com parâmetros n e π, $X \sim Bin(n, \pi)$, então:

$$IC_{1-\alpha}(\pi) : P_i(\alpha) \leq \pi \leq P_s(\alpha)$$

em que $P_i(\alpha)$ e $P_s(\alpha)$ são os limites inferior e superior do intervalo

Método

$$P_i(\alpha) = \frac{X}{X + (n - X + 1)F_{1-\alpha/2,\ 2(n-X+1),\ 2X}}$$

$$P_s(\alpha) = 1 - \frac{n - X}{n - X + (X + 1)F_{1-\alpha/2,\ 2(X+1),\ 2(n-X)}}$$

Os valores de F são obtidos da tabela da distribuição F

Exemplo

Avaliando-se o nível de infestação de certa doença de raiz num cultivo de algodão, 25 plantas foram tomadas ao acaso, constatando-se que 2 delas apresentavam sintomas. Admitindo que uma infestação abaixo de 30% é controlável, verifique a possibilidade de se efetivar o controle. Use $\alpha = 5\%$
As hipóteses de interesse são

$$H_0 : \pi = 0.3 \quad vs \quad H_1 : \pi < 0.3$$

ANOVA não paramétrica

Quando as pressuposições do modelo de ANOVA (normalidade, homocedasticidade) não são verificadas, ou quando há *outliers*, a hipótese de igualdade de tratamentos pode ainda ser avaliada por meio dos seguintes testes não paramétricos:

- Teste de Kruskal-Wallis (DIC)
- Teste de Friedman (DBC)

Teste de Kruskal-Wallis

Os centros (medianas, $\tilde{\mu}$) de de k amostras independentes (tratamentos) podem ser comparados por meio do teste de Kruskal-Wallis.
As seguintes hipóteses são avaliadas:

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \ldots = \tilde{\mu}_k$$

contra a hipótese H_1 de que ao menos duas amostras diferem entre si

Método

- Considere k amostras, cada uma com n_i observações. Temos que $N = \sum_{i=1}^{k} n_i$
- Atribua ranks de 1 a N, conjuntamente, às observações. Havendo impates, calcule um rank médio.
- Calcule a estatística:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

em que R_i é a soma dos ranks da i-ésima amostra.
Para $k \geq 3$ e $n_i \geq 6$, a seguinte aproximação é feita:

$$H \sim \chi^2_{k-1}$$

Obs.: para $k = 2$ o teste de Wilcoxon deve ser aplicado; para $k = 3$, tabelas específicas devem ser utilizadas (ver Campos, 1983)

Exemplo

Num experimento sobre eficiência de inseticidas no feijoeiro, foram obtidos os seguintes resultados de rendimento (kg/ha)

Testemunha	Ekatin	Diazinon	EPN
1451	1451	1368	1534
1534	1575	1534	1658
1493	1534	1534	1700
1285	1493	1534	1700
5763	6053	5970	6592
1441	1513	1492	1648

Teste de Friedman

- O teste de Friedman é análogo ao teste F da ANOVA para dados provenientes de um DBC.
- Além dos tratamentos, há mais um fator de classificação (blocos)
- As mesmas hipóteses apresentadas no teste de Kruskal-Wallis são avaliadas
- O procedimento consiste em atribuir ranks às observações de cada bloco, separadamente.

Teste de Friedman

Estatística de teste:

$$X^2 = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1)$$

em que b é o número de blocos, k é o número de tratamentos, R_i é a soma dos ranks do i-ésimo tratamento

H_0 é rejeitada quando: $X^2 \geq \chi^2_{k-1,\ 1-\alpha}$

Exemplo

Num ensaio de adubação nitrogenada (salitre) em alface, 4 tratamentos foram estudados (Trat.1 = testemunha). Os resultados de produção (peso de 12 pés, em gramas) são dados na tabela:

Bloco	Trat.1	Trat.2	Trat.3	Trat.4
I	3640	4200	4700	5300
II	4890	4550	6020	5900
III	4800	5320	5250	5150
IV	4460	5500	5580	5560

Comparações múltiplas de tratamentos - teste de Kruskal-Wallis

A hipótese $H_0 : \tilde{\mu}_i = \tilde{\mu}_j$ de que dois tratamentos são iguais, será rejeitada quando

$$|\bar{R}_i - \bar{R}_j| \geq t_{N-k}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

sendo $\bar{R}_i$ o rank médio do i-ésimo tratamento, t_{N-k} é o valor da tabela de t-Student ao nível $1 - \alpha/2$

Comparações múltiplas de tratamentos - teste de Friedman

A hipótese $H_0 : \tilde{\mu}_i = \tilde{\mu}_j$ de que dois tratamentos são iguais, será rejeitada quando

$$|R_i - R_j| \geq t_{(k-1)(b-1)} \sqrt{\frac{bk(k+1)}{12}}$$

sendo R_i a soma dos ranks do i-ésimo tratamento, $t_{(k-1)(b-1)}$ é o valor da tabela de t-Student ao nível $1 - \alpha/2$

Transformação Box-Cox

Quando as pressuposições do modelo de ANOVA não são atendidas, outra alternativa é a transformação de dados, como a transformação *potência ótima de Box-Cox*.

Considerando $Y_1, Y_2, ..., Y_n$ os dados originais, a transformação de Box-Cox consiste em encontrar um λ tal que os dados transformados $Y_1^*, Y_2^*, ..., Y_n^*$ se aproximem de uma distribuição normal com variâncias constante. Esta transformação é dada por:

$$Y_i^* = \begin{cases} \log(Y_i), & \text{se } \lambda = 0 \\ \dfrac{Y_i^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \end{cases}$$

Exercícios

1. Numa área cultivada com algodão foi observado que dentre 20 maçãs examinadas, 8 estavam atacadas pela lagarta rosada. Verifique a necessidade de controle químico, admitindo que o nível crítico seja 25%.
2. Num experimento de competição de cultivares de cana-de-açúcar, foram registrados os seguintes rendimentos (kg/ha) de açúcar por tonelada de cana:

Co413	Co419	Co421	Co678
116.1	105.3	121.6	123.8
114.4	99.9	122.0	122.9
113.6	104.8	124.9	118.5
101.6	105.1	123.3	
112.7	100.2		

Estabeleça comparações entre as variedades.