Planejamento e Análise Estatística de Experimentos com um único fator: análise de dados de experimentos completamente aleatorizados

Prof. Caio Azevedo

- PCA: cada unidade experimental é alocada aleatoriamente à um determinado tratamento (fator).
- Exemplo 1 (das árvores): é selecionado aleatoriamente 30 árvores de cada tipo.
- Exemplo 2: Uma bioquímica (Tecnologia de Alimentos) está interessada em estudar a extração de pigmentos naturais, com aplicação como corante em alimentos. Numa primeira etapa tem-se a necessidade de escolher o melhor solvente extrator. A escolha do(s) melhor(es) solventes foi realizada através da medida da absorbância de um pigmento natural do fruto de baguaçú.
  Fator = tipos de solvente; k=5 níveis; n_k=5 repetições.

Descrição do Exemplo 2

- Quanto maior a absorbância, melhor o solvente.
- Unidade experimental: 10 gramas de polpa do fruto de baguaçú.
- Casualização: a partir de 1 kg de polpa, foram sendo retiradas amostras de 10 gramas, onde foram aplicados os tratamentos, numa ordem aleatória.
- Assim como no exemplo das árvores, em princípio, o fator é qualitativo.
- Experimento balanceado : mesmo número de observações (unidades experimentais) por nível do fator.
- Possível dependência entre as undidades experimentais?

Conjunto de dados

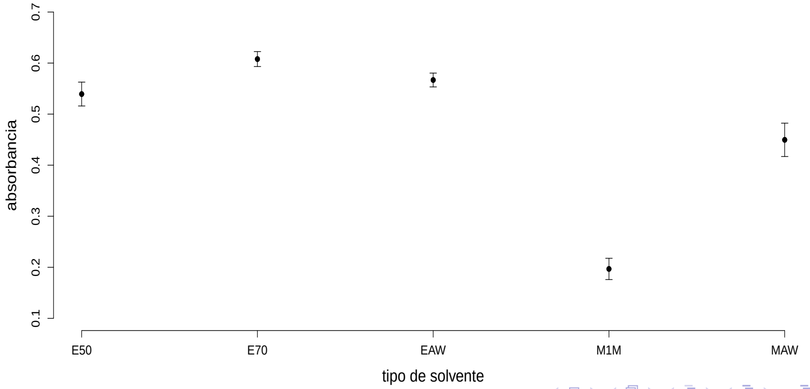
Solvente	Absorbância (Observação)				
	1	2	3	4	5
E50	0,5553	0,5623	0,5585	0,5096	0,5110
EAW	0,5436	0,5660	0,5860	0,5731	0,5656
MAW	0,4748	0,4321	0,4309	0,5010	0,4094
E70	0,6286	0,6143	0,5826	0,6079	0,6060
M1M	0,1651	0,1840	0,2144	0,2249	0,1954

Análise descritiva

Não há sentido em construir box-plots ou histogramas.

Solvente	Medida descritiva					
	Média	**DP**	**Var.**	**CV%**	**Mínimo**	**Máximo**
E50	0,539	0,026	0,0007	4,937	0,510	0,562
E70	0,608	0,017	0,0003	2,744	0,583	0,629
EAW	0,567	0,015	0,0002	2,717	0,544	0,586
M1M	0,197	0,024	0,0006	12,107	0,165	0,225
MAW	0,450	0,037	0,0014	8,283	0,409	0,501

Gráfico de perfis médios



Modelo (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij}, i = 1, 2, .., 5$$

(grupos); $j = 1, ..., 5$(unidades experimentais)

- Erros $\xi_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$, μ, α_i não aleatório.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- $\mu + \alpha_i$: média populacional relacionada ao i-ésimo fator, $\alpha_1 = 0$.
- $Y_{ij} \overset{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.

Suposições do modelo

- Homocedasticidade.
- Normalidade dos erros.
- Independência entre as observações.

Teste de Bartlett para igualdade de variâncias

- $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ vs $H_1 : \sigma_i^2 \neq \sigma_j^2$ para pelo menos um $i \neq j$
- Estatística do teste:

$$Q_B = \frac{q}{c},$$

em que

$$q = (n-k)\ln S_p^2 - \sum_{i=1}^{k}(n_i - 1)\ln S_i^2, S_p^2 = QMR = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{n-k}$$

$$c = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}(n_i - 1)^{-1} - (n-k)^{-1}\right]$$

- Sob H_0, $Q_B \approx \chi^2_{(k-1)}$. Rejeita-se H_0 quando $P(Q_B > q_B | H_0) < \alpha$, q_B valor calculado e α e o nível de significância.

Teste de Levene para igualdade de variâncias

- $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ vs $H_1 : \sigma_i^2 \neq \sigma_j^2$ para pelo menos um $i \neq j$
- Estatística do teste:

$$Q_L = \frac{(n-k)\sum_{i=1}^{k} n_i \left(Z_{i.} - Z_{..}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{i=1}^{n_i}\left(Z_{ij} - Z_{i.}\right)^2},$$

em que

$$Z_{ij} = |Y_{ij} - \overline{Y}_{i.}|; Z_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}; Z_{..} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$$

- Sob H_0, $Q_L \approx F_{(k-1,n-k)}$. Rejeita-se H_0 quando

  $P(Q_L > q_L | H_0) < \alpha$, q_L valor calculado e α e o nível de significância.

Testes para homocedasticidade

- Teste de Bartlett : 3,772 (0,4378).
- Teste de Levene : 0,696 (0,6033).
- Hipótese de homocedasticidade parace razoável.

Suposição de normalidade

- De acordo com o modelo, $\xi_{ij}^{*} = \frac{Y_{ij} - X_{ij}'\boldsymbol{\beta}}{\sigma} \sim N(0,1)$, em que X_{ij}' é a i,j-ésima linha da matriz de planejamento e $\sigma = \sqrt{\sigma^2}$.
- Problema $\boldsymbol{\beta}$ e σ^2 são desconhecidos, tem-se que usar estimadores apropriados.
- Neste caso $R_{ij} = \frac{Y_{ij} - \widehat{Y}_{ij}}{\widehat{\sigma}} \approx N(0,1)$, $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$, em que $\widehat{Y}_{ij} = X_{ij}'\widehat{\boldsymbol{\beta}}$.
- Contudos, os $R_i's$ não são independentes.
- Solução : utilizar o resíduo "studentizado", $T_{ij} = \frac{Y_{ij} - \widehat{Y}_{ij}}{(1 - h_{ij})\widehat{\sigma}} \approx N(0,1)$, em que h_{ij} é o ij-ésimo elemento da diagonal principal.
- Os $T_{ij}'s$ apresentam uma "dependência desprezível", sob as suposições do modelo.

Análise de resíduos

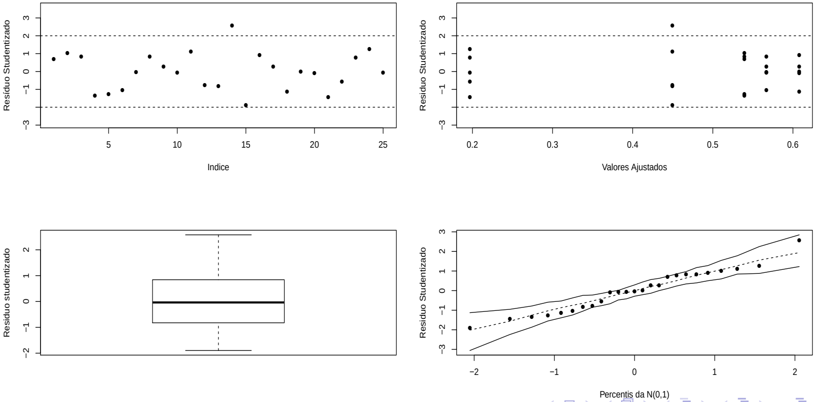


Tabela ANOVA

FV	SQ	GL	QM	Estatística F	pvalor
Solvente	0,541	4	0,135	212,81	$<$ 0,0001
Resíduo	0,012	20	$<$ 0,001		
Total	0.553	24			

Rejeita-se H_0.

Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
μ (E50)	0,539	0,011	[0,517; 0,561]	47,826	$<$ 0,0001
α_2 (E70)	0,069	0,0160	[0,037 ; 0,010]	4,298	0,0003
α_3 (EAW)	0,028	0,0160	[-0,004 ; 0,059]	1,726	0,0998
α_4 (M1M)	-0,343	0,0160	[-0,374; -0,311]	-21,481	$<$ 0,0001
α_5 (MAW)	-0,090	0,0160	[-0,121 ;-0,058]	-5,624	$<$ 0,0001

Parâmetro α_3 não significativo. Isto sugere uma possível equivalência entre os solventes E50 e EAW.

Modelo reduzido (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij}, i = 1, 2, .., 5$$

(grupos); $j = 1, ..., 5$(unidades experimentais)

- Erros $\xi_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$, μ, α_i não aleatório.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- $\mu + \alpha_i$: média populacional relacionada ao i-ésimo fator,

  $\alpha_1 = \alpha_3 = 0$.
- $Y_{ij} \overset{ind.}{\sim} N(0, \sigma^2)$.

Análise de resíduos

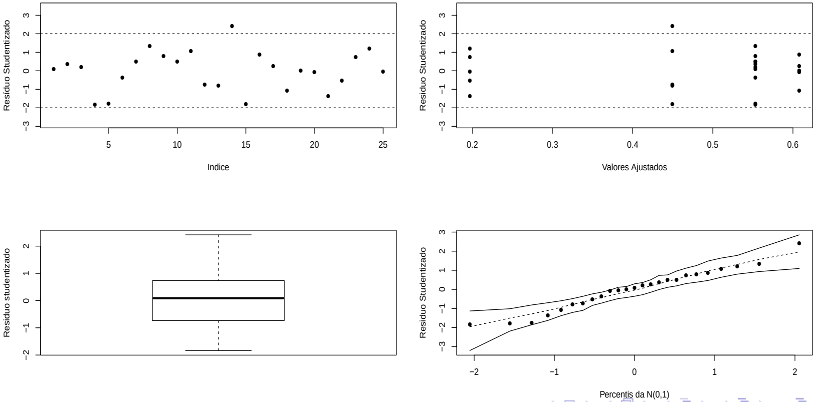


Tabela ANOVA do modelo reduzido

FV	SQ	GL	QM	Estatística F	pvalor
Solvente	0,539	3	0,180	258,41	$<$ 0,0001
Resíduo	0,014	21	$<$ 0,001		
Total	0,553	24			

Rejeita-se H_0.

Estimativas dos parâmetros do modelo

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
μ (E50/EAW)	0,553	0,008	[0,537;0,569]	66,310	$< 0{,}0001$
α_2 (E70)	0,055	0,0114	[0,026;0,083]	3,792	0,0011
α_4 (M1M)	-0,356	0,0114	[-0,385;-0,328]	-24,665	$< 0{,}0001$
α_5 (MAW)	-0,103	0,0114	[-0,132;-0,075]	-7,161	$< 0{,}0001$

Todos os incrementos α são significativos e todos parecem distintos entre si.

Estimativas finais das médias

Solvente	Estimativa	EP	IC(95%)
E50/EAW	0,553	0,008	[0,537;0,569]
E70	0,607	0,012	[0,584;0,631]
M1M	0,197	0,012	[0,173;0,220]
MAW	0,450	0,012	[0,426;0,472]

- Melhor solvente: E70.
- Pior solvente: M1M.
- Os soventes E50 e EAW são equivalentes.