Princípios de Bioestatística

Análise de Variância

Enrico A. Colosimo

Departamento de Estatística
Universidade Federal de Minas Gerais
http://www.est.ufmg.br/~enricoc

2011

Introdução

- Existem muitas situações nas quais um pesquisador deseja comparar **mais do que dois grupos** experimentais com relação a uma **variável quantitativa**.
- **Exemplos:**
    - Comparar três drogas analgésicas para reduzir a **dor pós-operatória** em pacientes submetidos à mesma intervenção cirúrgica.
    - Comparar a **idade** de pacientes entre três grupos de risco (baixo, médio, alto).

Introdução

- A primeira vista, pode parecer correto realizar vários testes t entre os grupos, comparando-os *dois a dois*.
- No caso da comparação de três grupos (grupo A, grupo B e grupo C), haveria **três testes** t de comparação entre médias:

  μ_A *vs* μ_B, μ_A *vs* μ_C e μ_B *vs* μ_C.
- Na comparação de quatro grupos, haveria **seis testes** t de comparação entre médias.
- Se o número de grupos é igual a 10, precisaríamos de **45 testes** t dois a dois.

- **Observação 1:** O número de testes aumenta conforme o número de grupos aumenta. Para k grupos temos $\binom{k}{2}$ comparações.
- **Observação 2:** Tal procedimento (a realização de todas as comparações dois a dois) é estatisticamente incorreto.
  - O teste t foi delineado para, **em um mesmo experimento**, comparar-se uma média A com apenas outra, B, com probabilidade α de se concluir, incorretamente, por uma diferença que não existe. Se forem feitas mais de uma comparação envolvendo a média A, a probabilidade de um erro deste tipo **passa a ser maior do que** α.

Introdução

- O procedimento correto para se evitar esse aumento no nível global de significância do experimento consiste em utilizar a técnica da **Análise de Variância** (ANOVA).
- Este método compara todas as médias em um único teste e visa a identificar a existência de **ao menos uma diferença** entre grupos, se alguma existir.
- Caso o resultado seja significativo, aplica-se posteriormente uma das várias técnicas existentes de **comparações múltiplas** entre as médias. Estes procedimentos permitem identificar quais as populações são diferentes entre si, mantendo controlado o nível de significância do teste.

Exemplo: Volume expiratório forçado (FEV)

- Desejamos comparar o *volume expiratório forçado* de pacientes com doença coronária oriundos de três centros médicos diferentes (21 pacientes da Johns Hopkins University School of Medicine, 16 pacientes do Rancho Los Amigos Medical Center e 23 pacientes da St. Louis University School of Medicine).
- Estamos interessados em testar $H_0 : \mu_1 = \mu_2 = \mu_3$ contra a alternativa que pelo menos uma das médias populacionais difere das outras. Os dados são apresentados a seguir.

Exemplo: Volume expiratório forçado (FEV)

Tabela: Volume expiratório forçado em 1 segundo para pacientes com doença coronária de três diferentes centros médicos.

Johns Hopkins		Rancho Los Amigos		St. Louis	
3,23	2,57	3,22	2,61	2,79	3,17
3,47	2,08	2,88	3,39	3,22	2,23
1,86	2,47	1,71	3,17	2,25	2,19
2,47	2,47	2,89		2,98	4,06
3,01	2,74	3,77		2,47	1,98
1,69	2,88	3,29		2,77	2,81
2,10	2,63	3,39		2,95	2,85
2,81	2,53	3,86		3,56	2,43
3,28		2,64		2,88	3,20
3,36		2,71		2,63	3,53
2,61		2,71		3,38	
2,91		3,41		3,07	
1,98		2,87		2,81	
$\bar{x}_1 = 2,63$ litros		$\bar{x}_2 = 3,03$ litros		$\bar{x}_3 = 2,88$ litros	
$s_1 = 0,496$ litros		$s_2 = 0,523$ litros		$s_3 = 0,498$ litros	

Fontes de variação

- Como o nome sugere, a análise de variância depende de estimativas da dispersão.
- Quando trabalhamos com diferentes populações, podemos calcular dois tipos de medidas de variância:
  - a variação dos valores dos indivíduos em torno das médias populacionais (variância dentro dos grupos);
  - e a variação das médias populacionais em torno da média global (variância entre os grupos).
- Se a variabilidade dentro das k diferentes populações é pequena em relação a variabilidade entre suas respectivas médias, isto sugere que as médias populacionais são de fato diferentes.

Fontes de variação

- Para testar a hipótese nula

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

para um conjunto de k populações, primeiro precisamos encontrar uma medida de variabilidade das observações individuais em torno de suas médias populacionais.

- A estimativa *combinada* da variância comum σ^2 é tal medida. Seja $n = n_1 + n_2 + \ldots + n_k$ (tamanho total da amostra), então

$$s_D^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k},$$

em que s_j^2 é a variância amostral do grupo j.

- Esta quantidade é simplesmente a média ponderada das k variâncias amostrais. O subscrito D se refere a variabilidade "dentro de grupos".

Fontes de variação

- Precisamos de uma expressão que estime a variação das médias em torno da média global, ou seja, a variância "entre grupos".
- Se a hipótese nula é verdadeira, esta quantidade também estima a variância comum σ^2

$$s_E^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2}{k-1},$$

em que $\bar{x}_j$ é a média amostral do grupo j e $\bar{x}$ é a média global das n observações

$$\begin{aligned}
\bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k} \\
&= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n}.
\end{aligned}$$

Fontes de variação

- Agora que temos as estimativas das variâncias, gostarímos de responder a seguinte questão: as médias amostrais variam em torno da média global mais do que as observações individuais variam em torno das médias amostrais?
- Se sim, isto implica que as correspondentes médias populacionais são diferentes.
- Para testar a hipótese nula que as médias populacionais são idênticas, usamos a seguinte estatística de teste

$$F = \frac{s_E^2}{s_D^2}.$$

Fontes de variação

- Sob a hipótese nula, tanto s_E^2 quanto s_D^2 estimam a variância comum σ^2, e F é próximo de 1.
- Se existe uma diferença entre as populações, então a variância entre os grupos é maior que a variância dentro dos grupos, e F é maior que 1.
- Sob H_0, a razão F tem uma distribuição F com $k-1$ e $n-k$ graus de liberdade.

Fontes de variação

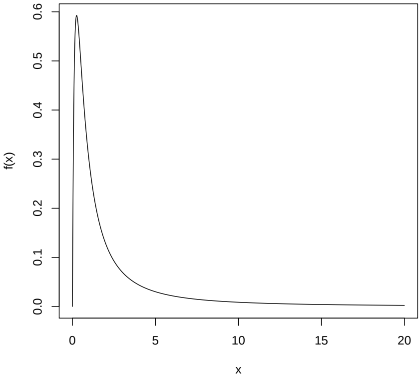

Figura: Distribuição F com 4 e 2 graus de liberdade.

- **Observação:** Podemos organizar o procedimento na seguinte tabela

Tabela: Tabela da Análise de Variância.

Fontes de variação	SQ	GL	QM	F
Entre os grupos	SQ entre	$k-1$	QM entre=SQ entre$/k-1$	QM entre/QM dentro
Dentro dos grupos	SQ dentro	$n-k$	QM dentro=SQ dentro$/n-k$	-
Total	SQ Total	$n-1$	-	-

em que:

- SQ entre é a "soma de quadrados" entre os grupos e é numerador de s_E^2, ou seja, $SQ \quad \text{entre} = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_k(\bar{x}_k - \bar{x})^2$.
- SQ dentro é a "soma de quadrados" dentro dos grupos e é numerador de s_D^2, ou seja, $SQ \quad \text{dentro} = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2$.
- Note que QM entre= s_E^2 e QM dentro= s_D^2.

Fontes de variação

- Retomando o exemplo, estamos interessados em testar se a média da variável FEV é igual para os pacientes dos três diferentes centros médicos

$$H_0 : \mu_1 = \mu_2 = \mu_3.$$

- Para começar, calculamos a estimativa da variância dentro dos grupos

$$\begin{aligned} s_D^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_k^2}{n_1+n_2+n_3-3} \\ &= \frac{(21-1)(0,496)^2 + (16-1)(0,523)^2 + (23-1)(0,498)^2}{21+16+23-3} \\ &= 0,254 \quad \text{litros}^2. \end{aligned}$$

Fontes de variação

- Temos que a média global é

$$\begin{aligned}\bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} \\ &= \frac{21(2,63)16(3,03)23(2,88)}{21 + 16 + 23} \\ &= 2,83 \quad \text{litros},\end{aligned}$$

e assim a estimativa da variância entre os grupos é

$$\begin{aligned}s_E^2 &= \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2}{3 - 1} \\ &= \frac{21(2,63 - 2,83)^2 + 16(3,03 - 2,83)^2 + 23(2,88 - 2,83)^2}{3 - 1} \\ &= 0,769 \quad \text{litros}^2.\end{aligned}$$

Fontes de variação

- Desta forma, a estatística de teste é

$$
\begin{aligned}
F &= \frac{s_E^2}{s_D^2} \\
&= \frac{0,769}{0,254} \\
&= 3,028.
\end{aligned}
$$

- Para a distribuição F com $k-1=3-1=2$ e $n-k=60-3=57$ graus de liberdade, $p-valor=0,0563$.
- Se rejeita a hipótese nula ao nível de 10% de significância, mas não se rejeita a hipótese nula ao nível de 5% de significância.
- Possivelmente haja alguma diferença entre as médias dos valores do FEV entre estas três populações.

Fontes de variação

- De forma, análoga temos a tabela de análise de variância.

Tabela: Tabela da Análise de Variância.

Fontes de variação	SQ	GL	QM	F	*p-valor*
Entre os grupos	14,48	57	$0,769$	3,028	0,0563
Dentro dos grupos	1,54	2	$0,254$	-	-
Total	16,02	59	-	-	-

Procedimentos de comparações múltiplas

- Um valor de F significativo na ANOVA não indica quais são os grupos significativamente diferentes entre si quando comparados dois a dois. Ele apenas mostra que existe ao menos uma diferença entre os grupos estudados.
- A identificação de diferenças particulares entre médias, tomando-as duas a duas, deve ser feita usando um dos vários testes de *Comparações Múltiplas entre Médias* existentes na literatura.
- Estes testes são semelhantes ao t, com a diferença de que controlam o nível de significância ao levar em conta o número de comparações feitas no experimento.

- O procedimento de Bonferroni consiste em corrigir o valor de α, calculando-se
$$\alpha^* = \frac{\alpha}{m}$$
em que α é o nível de significância global e m é o número de comparações a serem realizadas ($m = \binom{k}{2}$, para k grupos).
- Para o caso de $k = 3$ populações, um total de $m = \binom{3}{2} = 3$ testes são necessários. Se definimos o nível de significância global em 10%, devemos utilizar
$$\alpha^* = \frac{0,10}{3} = 0,033$$
para cada teste individual.

- Para realizar um teste da hipótese nula

$$H_0 : \mu_i = \mu_j$$

devemos calcular

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{s_D^2[(1/n_i) + (1/n_j)]}}.$$

- Note que este é um teste t para duas amostras, porém ao invés de estimarmos a variância somente com as duas amostras que estão sendo comparandas, estimamos utilizando a informação das k amostras. Sob a hipótese nula, t_{ij} tem uma distribuição t com $n - k$ graus de liberdade.

- Considerando o exemplo do FEV, começamos comparando as populações 1 e 2, os pacientes da Johns Hopkins e pacientes do Rancho Los Amigos. Neste caso,

$$
\begin{aligned}
t_{12} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_D^2[(1/n_1) + (1/n_2)]}} \\
&= \frac{2,63 - 3,03}{\sqrt{0,254[(1/21) + (1/16)]}} \\
&= -2,39.
\end{aligned}
$$

- Para um teste t com $n - k = 60 - 3$ graus de liberdade, o $p - valor = 0,02$.
- Assim, rejeitamos a hipótese nula ao nível 3,3% e concluímos que a média do FEV dos pacientes da Johns Hopkins é diferente da média do FEV dos pacientes do Rancho Los Amigos.

- Comparando as populações 1 e 3 (os pacientes da Johns Hopkins e os pacientes da St. Louis), temos

$$\begin{aligned} t_{13} &= \frac{\bar{x}_1 - \bar{x}_3}{\sqrt{s_D^2[(1/n_1) + (1/n_3)]}} \\ &= \frac{2,63 - 2,88}{\sqrt{0,254[(1/21) + (1/23)]}} \\ &= -1,64. \end{aligned}$$

- Como o $p - valor > 0,10$, não temos evidências suficientes para concluir que μ_1 difere de μ_3.

- Comparando os pacientes do Rancho Los Amigos e os pacientes da St. Louis (populações 2 e 3), temos

$$
\begin{aligned}
t_{23} &= \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{s_D^2[(1/n_2) + (1/n_3)]}} \\
&= \frac{3,03 - 2,88}{\sqrt{0,254[(1/16) + (1/23)]}} \\
&= 0,91.
\end{aligned}
$$

- Novamente o $p - valor > 0,10$, e não temos evidências suficientes para concluir que μ_2 difere de μ_3.
- Resumindo, encontramos que a média do FEV dos pacientes da Johns Hopikins é um pouco menor que a média do FEV dos pacientes do Rancho Los Amigos. Nenhuma outra distinção entre as populações pode ser feita.

Condições para o uso da ANOVA

- Para que os resultados da Análise de Variância sejam válidos, é necessário que as variâncias amostrais sejam semelhantes nas diferentes amostras. Além disso, a variável a ser comparada deve ter distribuição normal.
- A ANOVA é um procedimento estatístico robusto e fornece resultados confiáveis mesmo com considerável heterocedasticidade (variâncias desiguais), *desde que* os tamanhos amostrais sejam aproximadamente iguais.
- Também é razoavelmente robusto ainda que a variável em estudo tenha uma distribuição bastante desviada da normal, especialmente se os tamanhos amostrais são grandes.