Planejamento de Experimentos

Analise de Variância (ANOVA) com um Fator

Planejamento de Experimentos

- Muitas vezes é necessário obter informações sobre produtos e processos *empiricamente*.
- Trabalho assemelha-se ao de pesquisadores ou cientistas que precisam projetar experimentos, coletar dados e analisá-los.

Planejamento de Experimentos

- Experimentos são empregados para:
  - resolver problemas;
  - decidir entre diferentes processos, conceitos, etc;
  - entender a influência de determinados fatores, etc.

Planejamento de Experimentos

- Importância:
  - Intensa base tecnológica dos produtos,
  - Exigências governamentais e de clientes,
  - Necessidade de emprego de experimentos durante todas as etapas do ciclo de vida de produto.

Planejamento de Experimentos

- Planejar (delinear) experimentos:
  - Definir quais dados, em que quantidade e em que condições devem ser coletados durante um determinado experimento, buscando, satisfazer dois grandes objetivos: *maior precisão* estatística possível na resposta e o *menor custo*.

ANOVA com um Fator

- Análise para experimentos com uma variável resposta e um fator.
- É usada para testar a afirmação de que três ou mais médias populacionais são iguais.
- É uma extensão do teste t para duas populações independentes

ANOVA com um Fator

- A *variável resposta* é aquela que estamos comparando.
- O *fator* é a variável qualitativa usada para definir os grupos.
  - Assumiremos *k* amostras (grupos)
- O termo “com um fator” é devido a de cada valor possui apenas uma classificação.
  - Exemplos: comparações por sexo, raça, etc.

ANOVA com um Fator

- Condições e pressuposições
  - Os dados são provenientes de amostras aleatórias *independentes*.
  - As variâncias de cada amostra é assumidamente igual. Essa propriedade é conhecida como *homoscedasticidade*.
  - Os resíduos (ou erros) são normalmente distribuídos. Ou seja, cada uma das amostras possui distribuição normal.

ANOVA com um Fator

- A hipótese nula é que todas as médias são iguais

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

- A hipótese alternativa é que ao menos uma das médias é diferentes das demais
  - Pense sobre quatro jogos eletrônicos onde três são para PlayStation 2, mas um não é como os outros, é exclusivo para PlayStation 3.
  - Não é necessário que todos sejam diferentes, mas apenas um deles.

ANOVA com um Fator

- Uma sala de aula de estatística foi dividida em três partes: frente (F), meio (M), e fundão(B).
- O professor notou que quanto mais longe os alunos sentavam, mais provável que perdessem aulas ou usassem mensagens instantâneas durante as aulas. Ele deseja saber se os alunos que sentam mais longe foram pior nos exames.

ANOVA com um Fator

A ANOVA não testa que uma média é menor que outra, apenas se elas são todas iguais ou ao menos uma é diferente.

$$H_0: \mu_F = \mu_M = \mu_B$$

$$H_1: \mu_i \neq \mu_j, para\ algum\ i \neq j.$$

ANOVA com um Fator

- Uma amostra aleatória de alunos de cada parte da sala foi selecionada
- As notas para os alunos na segunda prova foram anotadas:
  - Frente (F): 82, 83, 97, 93, 55, 67, 53
  - Meio (M): 83, 78, 68, 61, 77, 54, 69, 51, 63
  - Fundão (B): 38, 59, 55, 66, 45, 52, 52, 61

ANOVA com um Fator

As medidas resumo para as partes da sala são apresentadas na tabela abaixo:

Parte	Frente (F)	Meio (M)	Fundão (B)
n	7	9	8
Média	75,71	67,11	53,50
Desv. Pad.	17,63	10,95	8,96
Variância	310,90	119,86	80,29

One-Way ANOVA

- Variação
  - Variação é a soma de quadrados dos desvios entre um valor e a média dos valores.
  - Soma de quadrados, ou SQ, é seguida pela identificação da variável a qual se refere entre parênteses como SQ(E) ou SQ(D).
  - Soma de Quadrados é um outro nome para variação.

ANOVA com um Fator

- Os valores são iguais?
  - Não, pois há alguma variação nos dados.
  - Conhecida como variação *total* e denotada por SQ(Total) é a Soma de Quadrados (variação) Total.

ANOVA com um Fator

- As médias amostrais são iguais?
  - Não, pois existe alguma variação entre um grupo e outro, conhecida como variação *entre* os grupos.
  - Algumas vezes chamada de variação devida ao *fator* e denotada por SQ(E) para Soma de Quadrados (variação) Entre os grupos.

ANOVA com um Fator

- Os valores dentro de cada grupo são iguais?
  - Não, pois existe alguma variação dentro de cada um dos grupos, conhecida como variação *dentro* dos grupos.
  - Algumas vezes é chamada de variação resíduo (erro) e é denotada por SQ(D) para Soma de Quadrados (variação) Dentro dos grupos.

ANOVA com um Fator

- Existem duas fontes de variação:
  - A variação *entre* os grupos, SQ(E), ou a variação devido ao *fator*;
  - A variação *dentro* dos grupos, SQ(D), ou a variação que não pode ser explicada pelo fator então é chamada de *resíduo*, ou erro.

ANOVA com um Fator

- Tabela base de uma ANOVA com um fator

Fonte de variação	SQ	gl	QM	F	p
Entre					
Dentro				--	--
Total			--	--	--

ANOVA com um Fator

- Média geral
  - A média aritmética de todos os valores quando o fator é ignorado.
  - Ou ainda, a média ponderada das médias amostrais de cada grupo pelo tamanho de cada amostra.

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i} \qquad \bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

ANOVA com um Fator

- Média geral para o nosso exemplo é 65.08

$$\bar{\bar{x}} = \frac{7(75.71)+9(67.11)+8(53.50)}{7+9+8}$$

$$\bar{\bar{x}} = \frac{1562}{24}$$

$$\bar{\bar{x}} = 65.08$$

ANOVA com um Fator

- Variação Entre Grupos, SQ(E)
  - A SQ(E) é a variação entre cada *média amostral* , de cada grupo, e a *média geral*.
  - A variação de cada grupo é ponderada pelo tamanho amostral.

$$SS(B) = \sum_{i=1}^{k} n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2$$

$$SS(B) = n_1\left(\bar{x}_1 - \bar{\bar{x}}\right)^2 + n_2\left(\bar{x}_2 - \bar{\bar{x}}\right)^2 + \cdots + n_k\left(\bar{x}_k - \bar{\bar{x}}\right)^2$$

ANOVA com um Fator

A Variação Entre Grupos para o nosso exemplo é SQ(E)=1902

$$SS(B) = 7(75.71 - 65.08)^2 + 9(67.11 - 65.08)^2 + 8(53.50 - 65.08)^2$$

$$SS(B) = 1900.8376 \approx 1902$$

ANOVA com um Fator

- Variação Dentro dos Grupos, SQ(D)
  - A SQ(D) é o total das variações individuais ponderadas.
  - A ponderação é realizada com os graus de liberdade
  - Os gl da amostra de cada grupo é o tamanho amostral menos uma unidade.

ANOVA com um Fator

Variação Dentro dos Grupos, SQ(D)

$$SS(W) = \sum_{i=1}^{k} df_i s_i^2$$

$$SS(W) = df_1 s_1^2 + df_2 s_2^2 + \cdots + df_k s_k^2$$

ANOVA com um Fator

- A variação dentro dos grupos para o nosso exemplo é 3386

$$SS(W) = 6(310.90) + 8(119.86) + 7(80.29)$$

$$SS(W) = 3386.31 \approx 3386$$

ANOVA com um Fator

- Após preencher as SQ temos:

Fonte de variação	SQ	gl	QM	F	p
Entre	1902				
Dentro	3386			--	--
Total	5288		--	--	--

ANOVA com um Fator

- Graus de liberdade, gl
  - Os gl são a quantidade de valores que podem variar antes que o restante dos valores sejam predeterminados.
  - Por exemplo, se você tem 6 números que possuem uma média 40, você saberia que o total deves ser 240. Cinco dos seis números poderiam ser quaisquer, mas uma vez que os 5 primeiros são conhecidos, o último é fixado para que a soma seja 240. Os gl seriam 6-1=5.

ANOVA com um Fator

- Os gl *entre* os grupos é o número de grupos menos 1
  - Temos três grupos, logo gl(E) = 2
- Os gl *dentro* dos grupos é a soma dos gl individuais de cada grupo
  - Tamanho das amostras são 7, 9 e 8
  - gl(D) = 6 + 8 + 7 = 21
- Os gl *total* é o tamanho da amostra geral menos 1
  - gl(Total) = 24 – 1 = 23

ANOVA com um Fator

- Preenchendo os gl temos ...

Fonte de variação	SQ	gl	QM	F	p
Entre	1902	2			
Dentro	3386	21		--	--
Total	5288	23	--	--	--

ANOVA com um Fator

- Variância
  - As variâncias são chamadas Quadrados Médios abreviadamente QM, acompanhadas da identificação da variável a que se refere QM(E) ou QM(D)
  - São os quadrados médios dos desvios das médias, encontrados dividindo a variação (SQ) pelos seus respectivos gl.
  - QM = SQ / gl

$$Variance = \frac{Variation}{df}$$

ANOVA com um Fator

- QM(E) = 1902 / 2 = 951,0
- QM(D) = 3386 / 21 = 161,2
- QM(T) = 5288 / 23 = 229,9
  - Note que the QM(Total) não é a soma dos QM(E) e QM(D).
  - Isso vale para as Somas de Quadrados SS(Total) = SQ(E) + SQ(D), mas não para os quadrados médios.
  - O QM(Total) não costuma ser usado.

ANOVA com um Fator

- Completando the QM temos...

Fonte de variação	SQ	gl	QM	F	p
Entre	1902	2	951,0		
Dentro	3386	21	161,2	--	--
Total	5288	23	229,9	--	--

ANOVA com um Fator

- Variâncias especiais
  - O QM(E) é conhecido como o *estimador da variâcia* uma vez que é a média ponderada das variâncias das amostras. Denotada por S^2_e.
  - O QM(Total) é o estimador da *variância da variável resposta*.
    - Não é utilizado na técnica da ANOVA, mas é útil para avaliações posteriores.
  - O *Coeficiente de Explicação* R^2*=SQ(E)/SQ(Total)*, representa a proporção de variação explicada pela ANOVA considerando o fator e quanto maior melhor.

ANOVA com um Fator

- Estatística F
  - Uma estatística teste F é a razão de duas variâncias amostrais
  - O QM(E) e QM(D) são as variâncias amostrais que divididas uma pela outra fornecem o valor de F.
  - F = QM(E) / QM(D)
- Para nossos dados
  - F = 951,0 / 161,2 = 5,9

ANOVA com um Fator

- Adicionando F a tabela...

Fonte de variação	SQ	gl	QM	F	p
Entre	1902	2	951,0	5,9	
Dentro	3386	21	161,2	--	--
Total	5288	23	229,9	--	--

ANOVA com um Fator

- A estatística F é unicaudal a direita
- A estatística F vem da distribuição F com gl(E) no numerador e gl(D) no denominador
- O valor-p é a área a direita do valor da estatística F
- $P(F_{2,21} > 5,9) = 0,009$

ANOVA com um Fator

- Completando a tabela com o valor-p

Fonte de variação	SQ	gl	QM	F	p
Entre	1902	2	951,0	5,9	0,009
Dentro	3386	21	161,2	--	--
Total	5288	23	229,9	--	--

ANOVA com um Fator

- O valor-p = 0,009 é menor que o nível de significância de 0,05, então rejeitamos a hipótese nula.
- Isto é, rejeitamos que as médias das três partes da sala foram as mesmas.
- Então ao menos uma das partes da sala possui média diferente.

ANOVA com um Fator

- A ANOVA não diz qual média é diferente.
- Você precisa de verificações post hoc para isso.
- Essas verificações são conhecidos como Testes de Comparações Múltiplas

Comparações Múltiplas

- Via Intervalo de Confiança para diferença:

$$IC(\mu_i - \mu_j) = \left[(\bar{y}_i - \bar{y}_j) \pm t_{\alpha/2(n_i+n_j-2)} \frac{S_e}{\sqrt{\left({}^{1}/_{n_i} + {}^{1}/_{n_j} \right)}} \right]$$

- Via testes de hipótese:
  - Estatística do teste:

$$\frac{(\bar{y}_i - \bar{y}_j)\sqrt{\left({}^{1}/_{n_i} + {}^{1}/_{n_j} \right)}}{S} \sim t_{(n_i+n_j-2)}$$

Comparações Múltiplas

- IC's para nossos dados:
  - $\bar{y}_F = 75{,}7,\ \bar{y}_M = 67{,}1,\ \bar{y}_B = 53{,}5\ , n_F = 7, n_M = 9, n_B = 8$ e $S^2{}_e = 951$
  - $E = t_{\alpha/2(n_i+n_j-2)} \dfrac{S_e}{\sqrt{\left({}^{1}/_{n_i} + {}^{1}/_{n_j} \right)}}$

Comparações		Diferenças	GL	t	E	LI	LS
F	M	8,6	14	1,895	115,93	-107,33	124,53
F	B	22,21	13	1,943	357,42	-335,21	379,63
M	B	13,61	15	1,895	151,85	-138,24	165,46