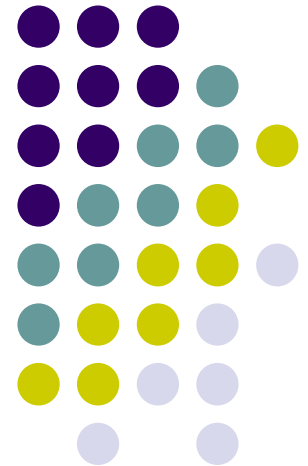
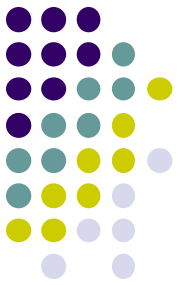


BIOSTATÍSTICA

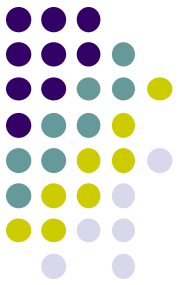
DELINEAMENTO EM BLOCOS AO ACASO



PREMISSAS PARA O USO



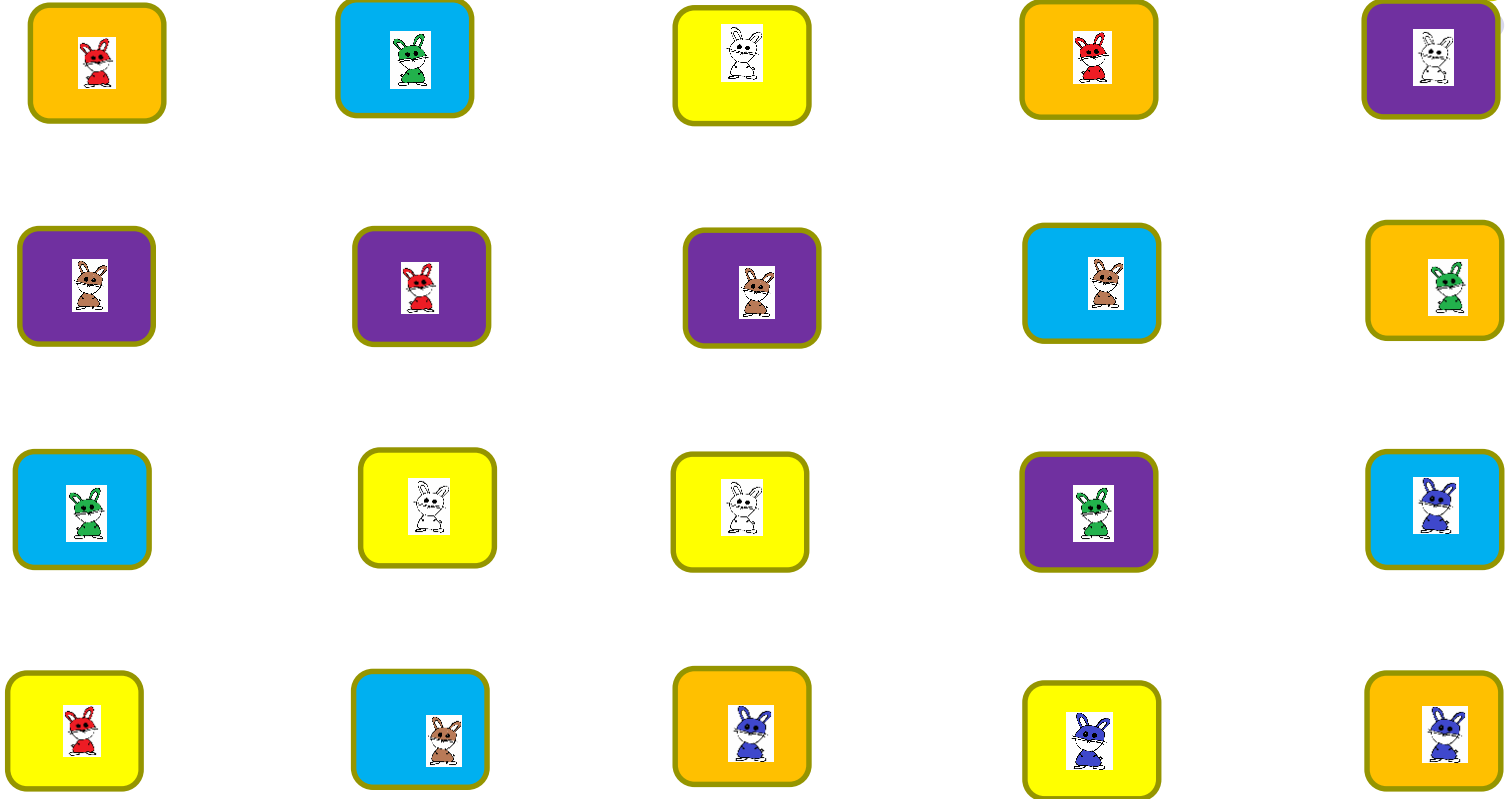
- PRATICAMENTE IMPOSSÍVEL CONTROLAR TODOS OS FATORES QUE INFLUENCIAM UM EXPERIMENTO.
- EXPERIMENTOS DE CAMPO SÃO AINDA MAIS DIFÍCEIS DE CONTROLAR (EFEITO DO LOCAL, CLIMA, E OUTROS FATORES EXTERNOS).
- PLANEJAMENTO ENTÃO DEVE SER CUIDADOSO E DEVEMOS USAR TODAS AS FERRAMENTAS QUE OS ESTATÍSTICOS NOS COLOCAM A DISPOSIÇÃO.
- O IMPORTANTE É CONSEGUIRMOS DETECTAR OS EFEITOS DESEJADOS (FAZER UM TESTE DE HIPÓTESE COM BAIXO ERRO EXPERIMENTAL).



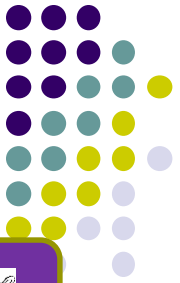
EXEMPLO:

- VAMOS ESTUDAR O EFEITO DE 4 TIPOS DE RAÇÃO PARA ALIMENTAÇÃO DE ANIMAIS SILVESTRES E TEMOS 20 ANIMAIS DE 5 PROLES DIFERENTES (4 ANIMAIS / PROLE).
- COMO DELINEAR ESTE ENSAIO?
- APRESENTAMOS 4 POSSIBILIDADES (A, B, C e D): Qual seria o melhor delineamento?

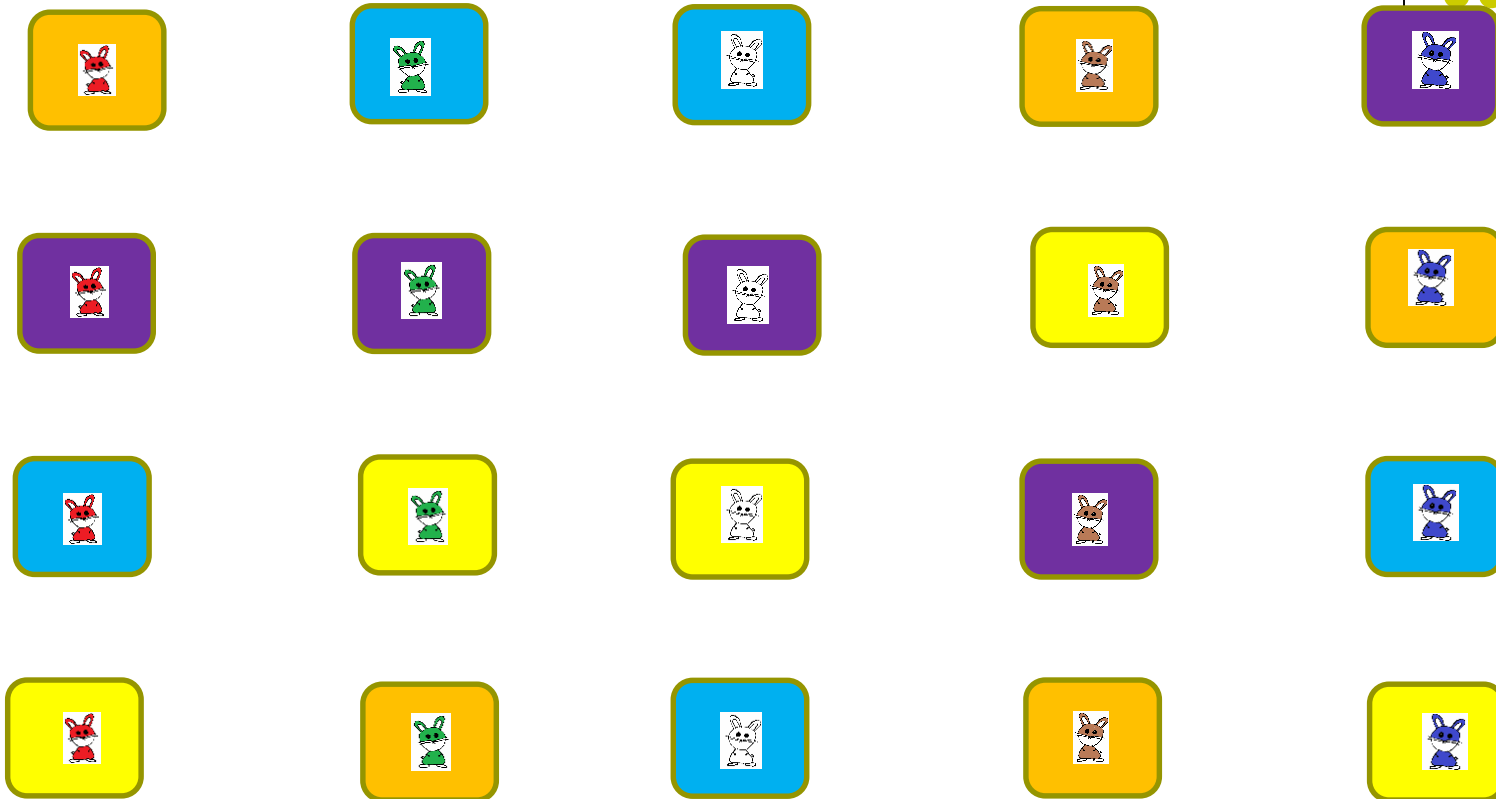
A



 Ração 1  Ração 2  Ração 3  Ração 4



B



Ração 1



Ração 2



Ração 3

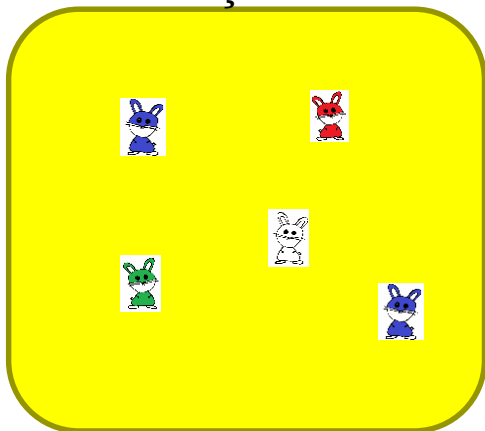


Ração 4

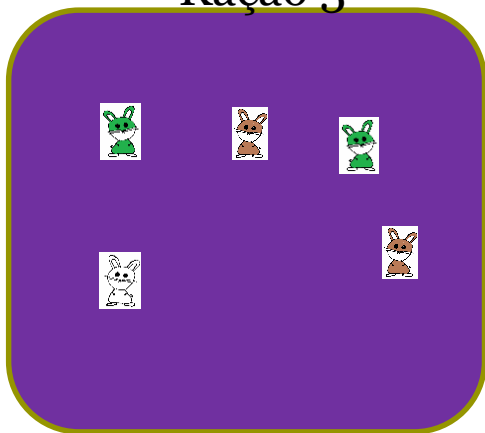


C

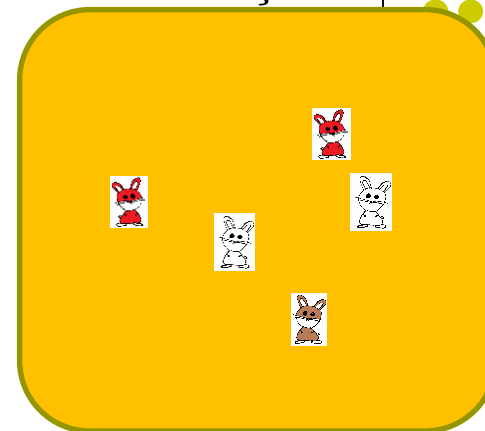
Ração 1



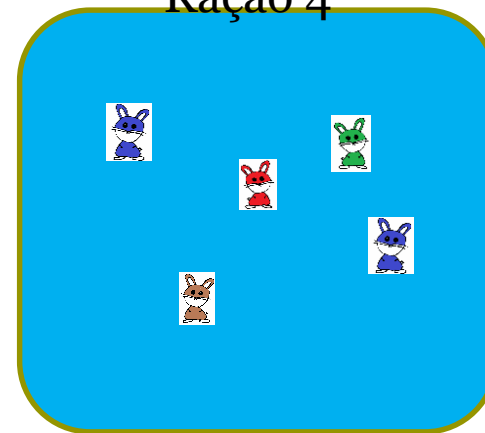
Ração 3



Ração 2

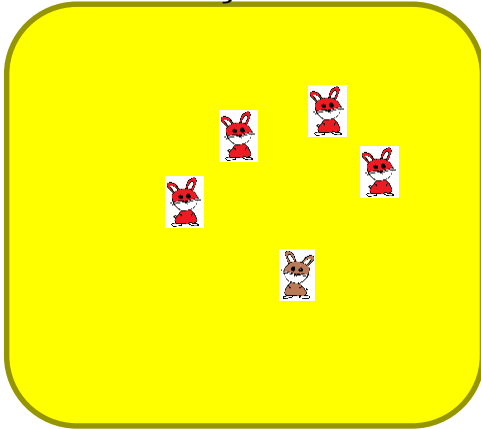


Ração 4

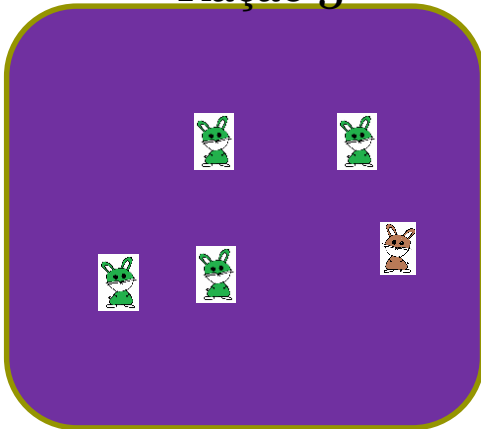


D

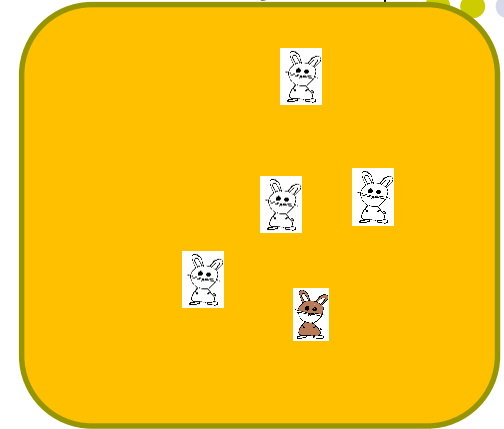
Ração 1



Ração 3



Ração 2



Ração 4



QUADRO DA ANÁLISE DE VARIÂNCIA PARA INTEIRAMENTE AO ACASO



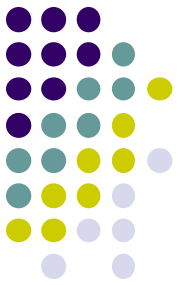
C.V.	g.l.	S.Q.	Q.M.	F
RAÇÃO	3			
RESÍDUO	36			
TOTAL	39			

QUADRO DE ANÁLISE DE VARIÂNCIA PARA BLOCOS AO ACASO:

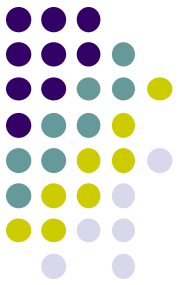


C.V.	g.l.	S.Q.	Q.M.	F
PROLE	9			
RAÇÃO	3			
RESÍDUO	27			
TOTAL	39			

DELINEAMENTO EM BLOCOS AO ACASO:



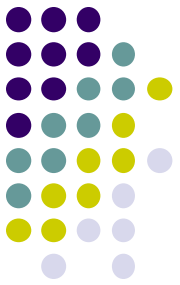
- RANDOMIZED COMPLETE BLOCK DESIGN.
- PRINCIPAL RESTRIÇÃO: DENTRO DE CADA BLOCO DEVE ESTAR UMA RÉPLICA DE CADA TRATAMENTO.
- AS CONDIÇÕES INTERNAS DE CADA BLOCO DEVEM SER HOMOGÊNEAS.
- QUANDO FOR POSSÍVEL UTILIZAR ESTE DELINEAMENTO DÊ PREFERÊNCIA EM RELAÇÃO AO INTEIRAMENTE AO ACASO.



EXEMPLO:

- SE ESTAMOS COLETANDO DADOS EM UM EXPERIMENTO INSTALADO EM UM LABORATÓRIO SURGEM DUAS POSSIBILIDADES PARA A COLETA DESSES DADOS:
- A) COLETAR UMA RÉPLICA OU BLOCO POR DIA;
- B) POR FACILIDADE, COLETAR TODAS AS UNIDADES EXPERIMENTAIS DE CADA TRATAMENTO POR DIA.
- QUAL A ALTERNATIVA MAIS INDICADA?

ANÁLISE ESTATÍSTICA:



MODELO

$$y_{ij} = m + t_i + b_j + e_{ij}$$

$i = 1, 2, \dots, a$ tratamento s

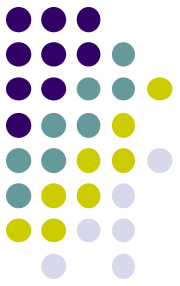
$j = 1, 2, \dots, b$ blocos

$$\sum_{i=1}^a t_i = 0;$$

$$\sum_{j=1}^b b_j = 0;$$

$$\sum_{i=1}^a \sum_{j=1}^b e_{ij} = 0$$

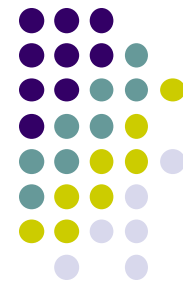
TESTE DE HIPÓTESE:



$$H_0 : t_1 = t_2 = \dots = t_a$$

H_A : pelo menos dois tratamentos diferem entre si.

EFICIÊNCIA RELATIVA (ER) (FISHER – 1947)



$$ER = \frac{(f_1 + 1).(f_2 + 3).QMRi}{(f_2 + 1).(f_1 + 3).QMRb} . 100$$

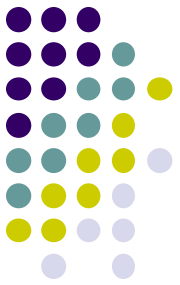
ER = eficiência relativa do delineamento em blocos em relação ao inteiramente ao acaso (%)

f_1 = g.l. resíduo para o delineamento em blocos ao acaso

f_2 = g.l. resíduo para o delineamento inteiramente ao acaso

QMRi = quadrado médio do resíduo do delineamento inteiramente ao acaso

QMRb = quadrado médio do resíduo do delineamento em blocos ao acaso.



Exercício: Com os dados abaixo fazer um programa SAS para analisar o experimento em Blocos Casualizados

DATA RAC1;

INPUT PROLE RACAO1 RACAO2 RACAO3 RACAO4;

DATALINES;

1 15.77 22.98 36.35 52.62

2 43.51 12.40 31.67 44.27

3 17.74 27.95 49.64 35.27

4 29.89 25.55 39.85 46.60

5 24.17 17.46 40.38 54.17

;;;

Teste normalidade e homogeneidade das variâncias.

PROC GLM DATA=.....;

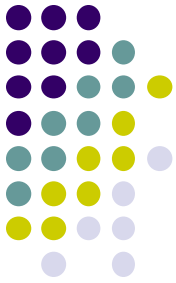
CLASS PROLE RACAO;

MODEL GANHO = PROLE RACAO;

MEANS RACAO / TUKEY LINES;

RUN;

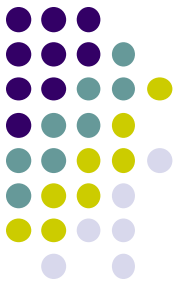
RESULTADO



Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	2084.666390	297.809484	3.47	0.0284
Error	12	1028.544330	85.712027		
Corrected Total	19	3113.210720			

Means with the same letter are not significantly different.				
Tukey Grouping		Mean	N	RACAO
	A	46.586	5	4
	A			
B	A	39.578	5	3
B				
B	C	26.216	5	1
	C			
	C	21.268	5	2

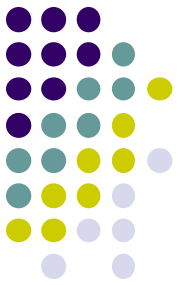
EFICIÊNCIA RELATIVA (exemplo)



$$ER = \frac{(12 + 1).(16 + 3).200}{(16 + 1).(12 + 3).100} . 100 = 194$$

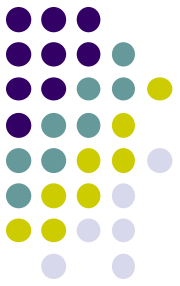
ISTO SIGNIFICA QUE O DELINEAMENTO EM BLOCOS AO ACASO É CERCA DE 2 VEZES MAIS EFICIENTE QUE O INTEIRAMENTE AO ACASO, OU SEJA, SE ESTE EXPERIMENTO FOSSE DELINEADO COMO INTEIRAMENTE AO ACASO, NECESSITARÍAMOS DE NO MÍNIMO 10 PROLES, EM VEZ DE 5.

Exercício

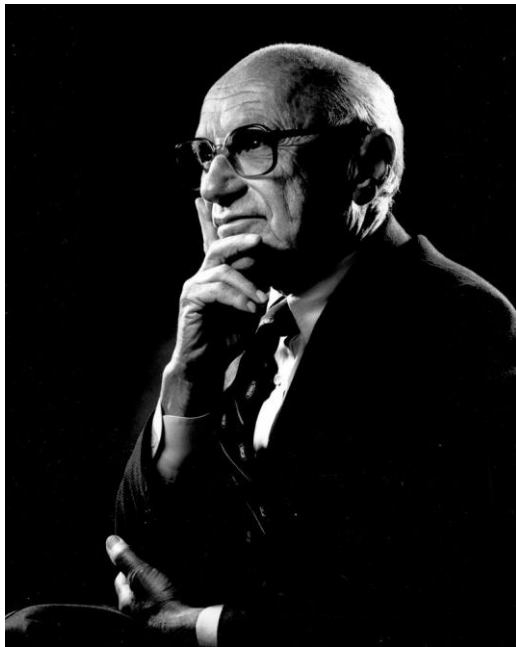


- Um pesquisador deseja estudar camuflagem de sementes de Trigo através do tratamento com corantes inertes. Para tanto, em um viveiro de pássaros selvagens com 85 aves, colocou 10 pratos com 6 compartimentos cada. Em cada compartimento colocou 100 gramas de sementes de trigo tratadas com os seguintes corantes: 1- Marron; 2- Azul; 3- Verde; 4- Vermelho; 5- Amarelo e 6 – Sem tratamento. Após 5 dias coletou os pratos e pesou a quantidade de sementes remanescente em cada compartimento. Os dados estão no arquivo CAMUFLAGEM_SEMENTES.xls, na página da disciplina. Utilize o Proc Format assim como o comando Format nos procedimentos utilizados.

Teste não paramétrico de Friedman

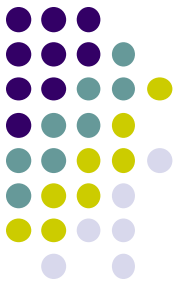


- O correspondente não paramétrico a análise de variância de delineamentos em blocos casualizados é o Teste de Friedman.



Milton Friedman

Nasceu em Nova York em 1912 e faleceu em São Francisco em 2006. Foi um economista, estatístico e escritor norte-americano que lecionou na Universidade de Chicago por mais de três décadas. Recebeu o Premio Nobel de Economia em 1976.

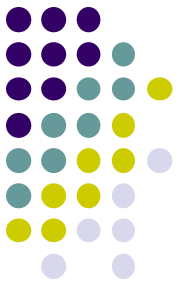


Teste de Friedman.

- Ordenar as observações da variável em estudo dentro de cada bloco de 1 a k. As observações empatadas são atribuídos valores médios das ordens que foram atribuídas se não estivessem empatados.
- Calcular em seguida o valor de
 - T_i = soma das ordens para o tratamento i, $i=1,2, \dots, k$.
- Calcular o valor do teste e comparar com a tabela

$$F_r = \frac{12}{bk(k+1)} \sum T_i^2 - 3b(k+1)$$

Cálculo do valor do teste de Friedman



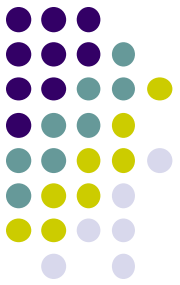
<u>B</u>	<u>Test A</u>	<u>Test B</u>	<u>Test C</u>
1	102	101	90
2	106	114	105
3	98	120	96
4	92	99	97

<u>B</u>	<u>Test A</u>	<u>Test B</u>	<u>Test C</u>
1	3	2	1
2	2	3	1
3	2	3	1
4	1	3	2
<i>T_i</i>	8	11	5

= 24

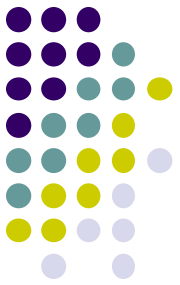
$$F_r = \frac{12}{bk(k+1)} \sum T_i^2 - 3b(k+1)$$

F_r tem distribuição de χ^2 com $(k-1)$ gl



Teste de Friedman no SAS

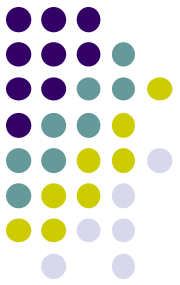
- O teste de Friedman é idêntico a análise de variância do teste estatístico CMH (Cochran-Mantel-Haenszel), quando a análise usa ordens (ordena os dados). (SCORES=RANK).
- Tiramos um exemplo do livro de Lehmann, 1975, p. 264.
- Lehmann, Erich L. (1975) Nonparametrics: Statistical Methods Based on Ranks. São Francisco, Holden-Day, 5634pp



Hipnose (Lehmann, 1975)

- Estudo feito para investigar se a hipnose tem o mesmo efeito sobre o potencial elétrico da pele (medido em milivolts) para quatro tipos de emoções: MEDO, ALEGRIA, TRISTEZA e CALMA.
- Oito indivíduos foram hipnotizados e solicitados a demonstrar essas emoções, quando foram medidos o potencial elétrico da pele.

Programa SAS

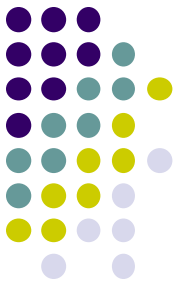


```
DATA HIPNOSE;
INPUT INDIV EMOCAO $ PELE_MV @@;
    DATALINES;
1 MEDO 23.1 1 ALEGRIA 22.7 1 TRISTEZA 22.5 1 CALMA 22.6
2 MEDO 57.6 2 ALEGRIA 53.2 2 TRISTEZA 53.7 2 CALMA 53.1
3 MEDO 10.5 3 ALEGRIA 9.7 3 TRISTEZA 10.8 3 CALMA 8.3
4 MEDO 23.6 4 ALEGRIA 19.6 4 TRISTEZA 21.1 4 CALMA 21.6
5 MEDO 11.9 5 ALEGRIA 13.8 5 TRISTEZA 13.7 5 CALMA 13.3
6 MEDO 54.6 6 ALEGRIA 47.1 6 TRISTEZA 39.2 6 CALMA 37.0
7 MEDO 21.0 7 ALEGRIA 13.6 7 TRISTEZA 13.7 7 CALMA 14.8
8 MEDO 20.3 8 ALEGRIA 23.6 8 TRISTEZA 16.3 8 CALMA 14.8
;
ODS GRAPHICS ON;
TITLE1 C=RED H=17PT 'Teste não paramétrico de Friedman -
Delineamento em blocos casualizados';
PROC FREQ DATA=HIPNOSE ;
    TABLES INDIV*EMOCAO*PELE_MV/
            CMH2 SCORES=RANK NOPRINT;
RUN;
TITLE1 C=RED H=17PT 'Teste não paramétrico de Friedman -
Delineamento inteiramente casualizado';
PROC FREQ DATA=HIPNOSE ;
    TABLES EMOCAO*PELE_MV/
            CMH2 SCORES=RANK NOPRINT;
RUN;
ODS GRAPHICS OFF;
```

INDIV = bloco

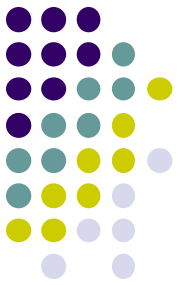
Não tem bloco
(INDIV)

Teste de hipóteses



- H_0 : Não há diferença entre as emoções.
- H_A : Pelo menos uma das emoções difere de outra.

Indivíduos usados como bloco



Teste não paramétrico de Friedman - Delineamento em blocos casualizados

The FREQ Procedure

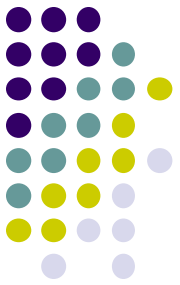
Summary Statistics for EMOCAO by PELE_MV
Controlling for INDIV

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.3750	0.5403
2	Row Mean Scores Differ	3	6.4500	0.0917

Significativo a 10%

Total Sample Size = 32

Delineamento inteiramente casualizado



Teste não paramétrico de Friedman - Delineamento inteiramente casualizado

The FREQ Procedure

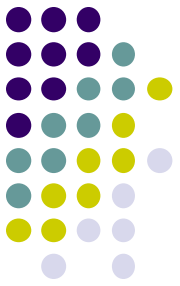
Summary Statistics for EMOCAO by PELE_MV

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0011	0.9731
2	Row Mean Scores Differ	3	0.5678	0.9038

Não significativo

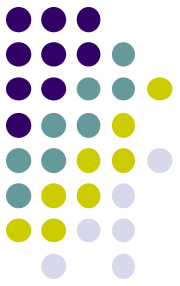
Total Sample Size = 32

Comparação entre os tratamentos



- Se rejeitamos a hipótese nula, o próximo passo é verificar quais tratamentos diferem entre si. Neste caso teríamos que usar o teste das Ordens Assinaladas de Wilcoxon (PROC UNIVARIATE), para todas as combinações (2 a 2) dos tratamentos. Para poucos tratamentos (3 por exemplo) isto não é um problema pois teremos 3 combinações. Mas quando temos 10 ou mais tratamentos o número de combinações aumenta exponencialmente (45 combinações para 10 tratamentos). Isto é tedioso e aumenta muito a programação e o tempo de trabalho.

Qual a solução?



Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics

W. J. Conover; Ronald L. Iman

The American Statistician, Volume 35, Issue 3 (Aug., 1981), 124-129.

TEXTBOOK

Practical Nonparametric Statistics, 3rd Edition

W. J. Conover

ISBN: 978-0-471-16068-7

592 pages

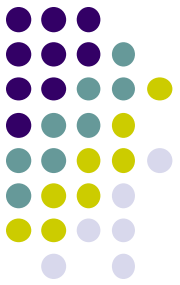
December 1998, ©1999



Ronald L Iman



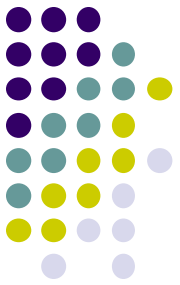
William J Conover



Aproximação pelo teste F

- Análise da variância dos dados ordenados.
- Uso do PROC RANK.
- Uso dos testes de comparação de médias, para os dados ordenados.

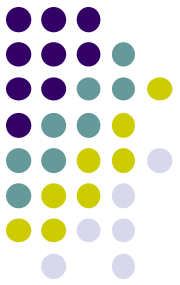
Programa SAS



```
DATA HIPNOSE;
  LENGTH EMOCACAO $10.;
  INPUT INDIV EMOCACAO $ PELE_MV @@;
  DATALINES;
1 MEDO 23.1 1 ALEGRIA 22.7 1 TRISTEZA 22.5 1 CALMA 22.6
2 MEDO 57.6 2 ALEGRIA 53.2 2 TRISTEZA 53.7 2 CALMA 53.1
3 MEDO 10.5 3 ALEGRIA 9.7 3 TRISTEZA 10.8 3 CALMA 8.3
4 MEDO 23.6 4 ALEGRIA 19.6 4 TRISTEZA 21.1 4 CALMA 21.6
5 MEDO 11.9 5 ALEGRIA 13.8 5 TRISTEZA 13.7 5 CALMA 13.3
6 MEDO 54.6 6 ALEGRIA 47.1 6 TRISTEZA 39.2 6 CALMA 37.0
7 MEDO 21.0 7 ALEGRIA 13.6 7 TRISTEZA 13.7 7 CALMA 14.8
8 MEDO 20.3 8 ALEGRIA 23.6 8 TRISTEZA 16.3 8 CALMA 14.8
;
ODS GRAPHICS ON;
PROC SORT DATA=HIPNOSE;
BY INDIV;
TITLE1 C=RED H=17PT 'Análise da variância dos dados transformados em ordens -
Aproximação do teste não paramétrico de Friedman';
PROC RANK DATA=HIPNOSE OUT=HIPNOSE1;
BY INDIV;
VAR PELE_MV;
RANKS RPELE_MV;
PROC GLM DATA=HIPNOSE1;
CLASS INDIV EMOCACAO;
MODEL RPELE_MV = INDIV EMOCACAO;
MEANS EMOCACAO / TUKEY LINES ALPHA=0.1;
RUN;
ODS GRAPHICS OFF;
```

O PROC RANK
atribui ordem aos
dados por bloco
(INDIV) e grava-
os no arquivo
HIPNOSE1

Resultado I



The GLM Procedure

Dependent Variable: RPELE_MV Rank for Variable PELE_MV

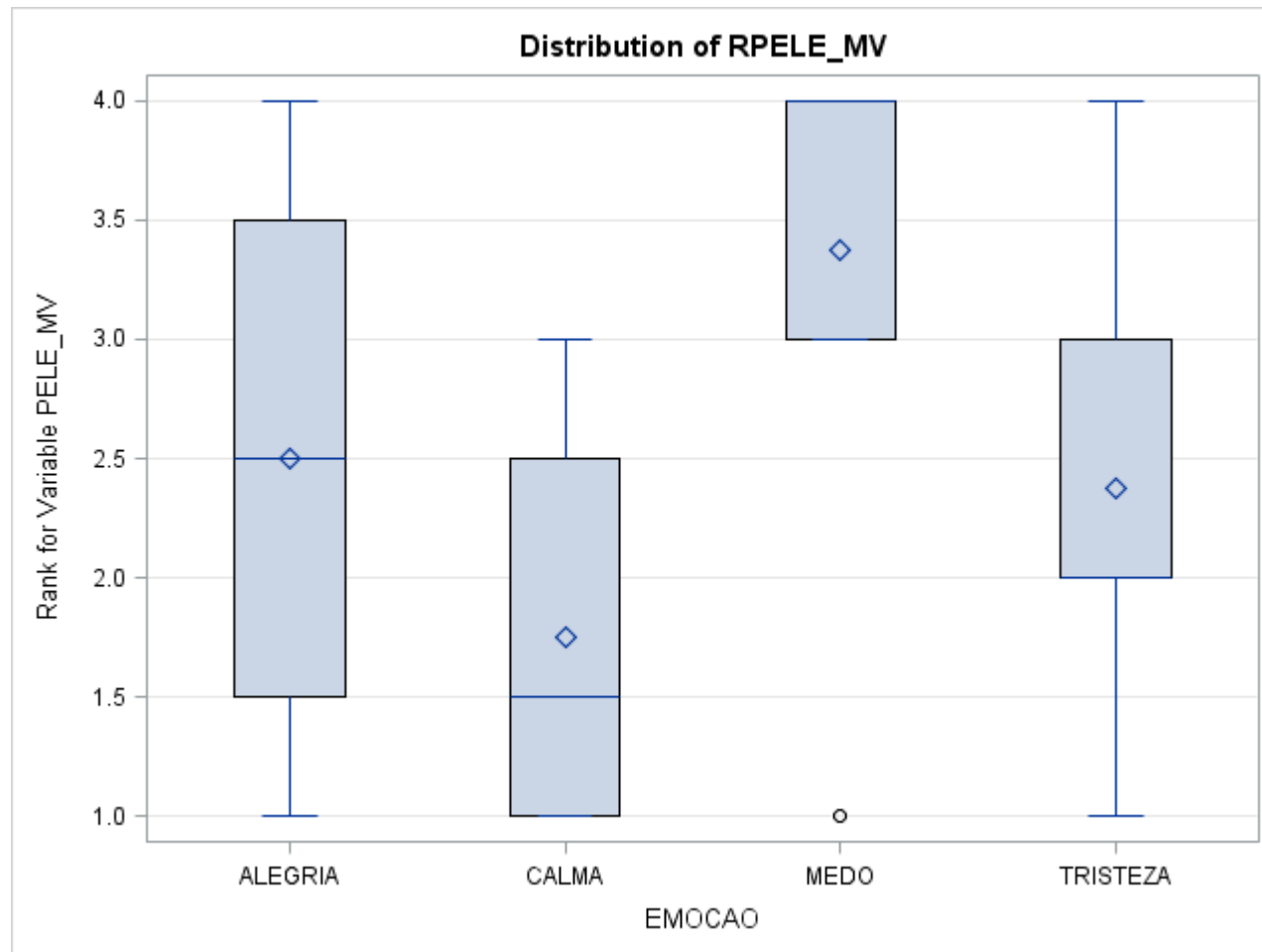
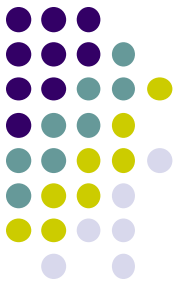
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	10.75000000	1.07500000	0.77	0.6540
Error	21	29.25000000	1.39285714		
Corrected Total	31	40.00000000			

R-Square	Coeff Var	Root MSE	RPELE_MV Mean
0.268750	47.20775	1.180194	2.500000

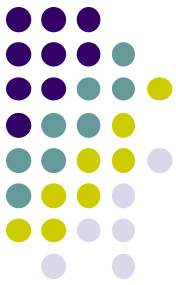
Source	DF	Type I SS	Mean Square	F Value	Pr > F
INDIV	7	0.00000000	0.00000000	0.00	1.0000
EMOCAO	3	10.75000000	3.58333333	2.57	0.0813

Source	DF	Type III SS	Mean Square	F Value	Pr > F
INDIV	7	0.00000000	0.00000000	0.00	1.0000
EMOCAO	3	10.75000000	3.58333333	2.57	0.0813

Resultado II



Resultado III



Análise da variância dos dados transformados em ordens - Aproximação do teste não paramétrico de Friedman

The GLM Procedure

Tukey's Studentized Range (HSD) Test for RPELE_MV

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.1
Error Degrees of Freedom	21
Error Mean Square	1.392857
Critical Value of Studentized Range	3.45055
Minimum Significant Difference	1.4398

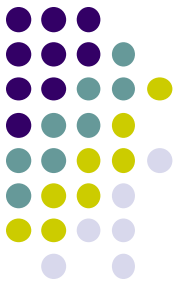
Means with the same letter are not significantly different.				
Tukey Grouping		Mean	N	EMOCAO
	A	3.3750	8	MEDO
	A			
B	A	2.5000	8	ALEGRIA
B	A			
B	A	2.3750	8	TRISTEZA
B				
B		1.7500	8	CALMA

Exercício



Uma rede de restaurantes faz periodicamente uma avaliação dos seus estabelecimentos. Para este trabalho possui 3 avaliadores que de forma independente fazem as avaliações. As notas variam de 0 a 20 pontos. A Diretoria deseja saber se existe diferença entre os restaurantes para investir na melhoria dos que apresentaram pior desempenho. **Os dados devem ser colocados em Excel da mesma forma que estão apresentados na tabela abaixo (pode mudar o nome das colunas).** Com estas análises seria possível recomendar o(s) restaurante(s) que deveria(m) entrar num programa de treinamento para melhoria do desempenho?

AVALIADOR	RESTAURANTE	NOTA	AVALIADOR	RESTAURANTE	NOTA	AVALIADOR	RESTAURANTE	NOTA
ANT	R1	17	NOR	R1	18	TAR	R1	16
ANT	R2	18	NOR	R2	18	TAR	R2	18
ANT	R3	18	NOR	R3	17	TAR	R3	17
ANT	R4	16	NOR	R4	16	TAR	R4	15
ANT	R5	17	NOR	R5	12	TAR	R5	17
ANT	R6	18	NOR	R6	17	TAR	R6	18
ANT	R7	16	NOR	R7	17	TAR	R7	15
ANT	R8	19	NOR	R8	18	TAR	R8	16
ANT	R9	17	NOR	R9	18	TAR	R9	16
ANT	R10	17	NOR	R10	17	TAR	R10	18
ANT	R11	16	NOR	R11	17	TAR	R11	16
ANT	R12	18	NOR	R12	15	TAR	R12	13



O PROC BOXPLOT

- Cria lado a lado gráficos de caixa e “bigode” (whiskers) de medidas organizadas em grupo. Um gráfico de caixa mostra a média, quartis, e os valores mínimos e máximos (BOXSTYLE=SKELETAL).
- PROC BOXPLOT DATA=A0;
- PLOT Y*X / BOXSTYLE=SKELETAL
HORIZONTAL;
- RUN;

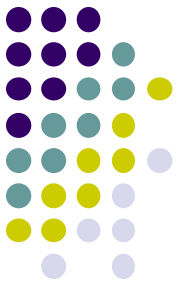
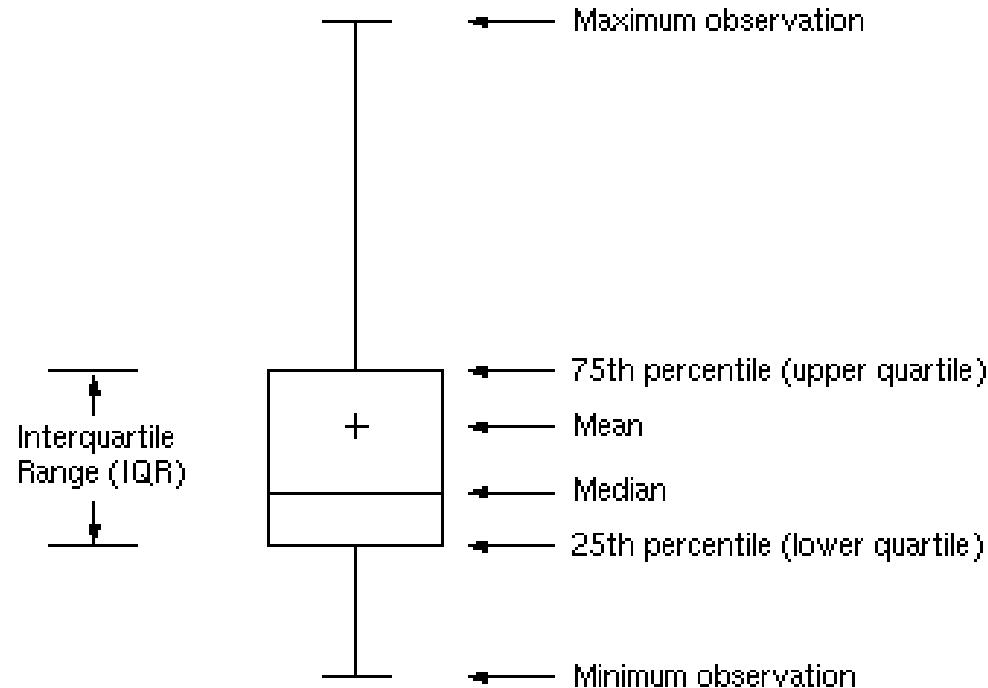
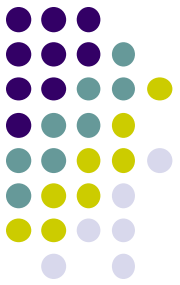


Gráfico de Caixa



1) McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots," *The American Statistician*, 32, 12–16.

2) Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

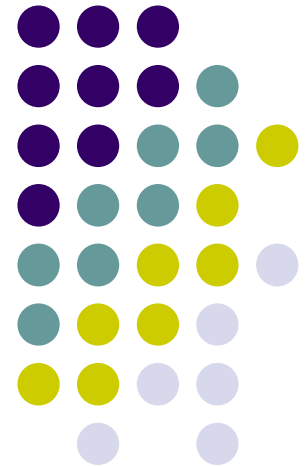


Exercício

- Com os dados de casos de dengue notificados em algumas regiões do estado de São Paulo, no arquivo do Exercício 4 (página da disciplina – cuidado com o nome das variáveis!!!), fazer um gráfico de caixa por região (todas as regiões no mesmo gráfico), em dois modelos (HORIZONTAL e VERTICAL).

BIOESTATÍSTICA

O TESTE QUI-QUADRADO

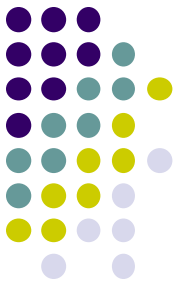


Usado para análise de dados categóricos

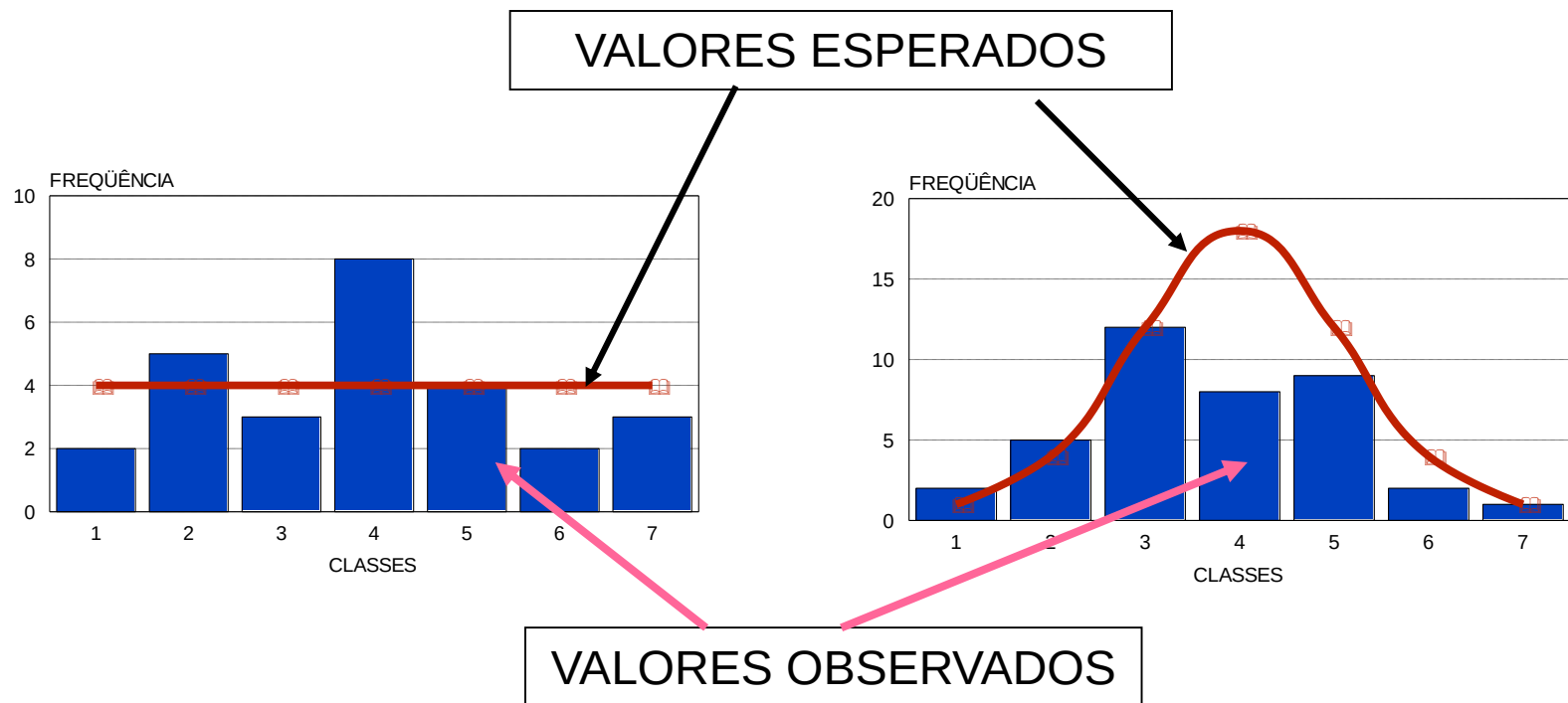


- CATEGÓRICOS: dividido em categorias; ordenado, classificado (HOUAISS).
- EX.: ALUNOS: graduação e pós-graduação.
- CLASSE DE SALÁRIO: A, B, C, e D
- DANO: baixo, médio e severo

O TESTE χ^2

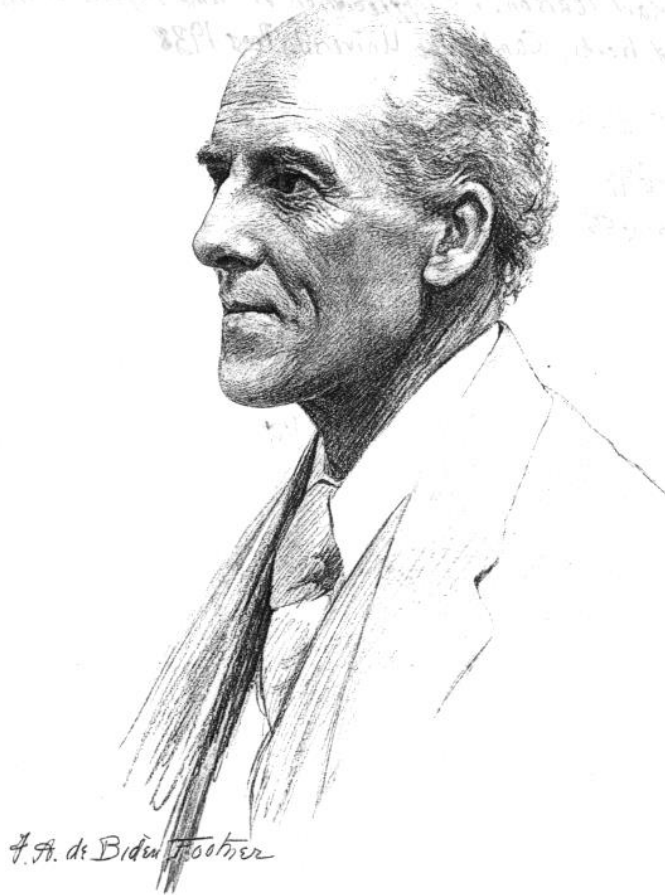


- O TESTE χ^2 É USADO PARA VERIFICAR SE UM CONJUNTO DE DADOS POSSUI UMA DISTRIBUIÇÃO ESPECÍFICA. PORTANTO É CHAMADO TAMBÉM DE TESTE DE AJUSTAMENTO.



Karl Pearson

Nasceu em 27 de março de 1857 em
Londres, Inglaterra
Falecido em 27 de abril de 1936 em
Surrey, Inglaterra

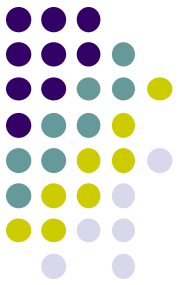


Pearson aplicou a estatística a problemas biológicos da hereditariedade e evolução.

Estudou em Cambridge University e trabalhou a maior parte de sua vida na University College, Londres.

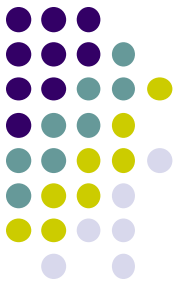
De 1893 a 1912 ele escreveu 18 trabalhos que contribuíram para a análise de regressão, o coeficiente de correlação e o teste qui-quadrado (1900). Foi Pearson que apresentou o termo 'desvio padrão'.

Pearson teve grande disputa científica com Fisher, enquanto ele defendia grandes amostras Fisher defendia pequenas amostras para os experimentos.



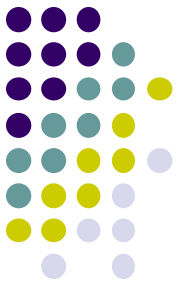
Tipos de teste χ^2

- **Teste de ajustamento** (goodness-of-fit): usado para testar se uma distribuição observada se ajusta a uma outra distribuição teórica. Também chamado teste de uma amostra.
- **Teste de independência (associação)**: usado para testar se existe ou não uma relação entre duas variáveis nominais ou categorizadas. Também chamado de Análise de Tabela de Contingência.



Suposições do Teste χ^2

- **Dados provêm de uma amostra aleatória.**
- **Independência:** observações devem ser independentes e mutuamente exclusivas.
- **Amostra de tamanho suficientemente grande.** Pequenas amostras expõe o pesquisador ao erro tipo II.
- **Tamanho de células adequado.** Recomenda-se ter células de no mínimo 5, mas nunca zero. Quando isto não é conseguido aplica-se o Teste Exato de Fischer.
- **Dados sempre na escala nominal ou ordinal** (categorizados).



TESTES DE AJUSTAMENTO

- OUTROS TESTES APARECERAM:
KOLMOGOROV-SMIRNOV, TESTE G^2 (RAZÃO DA MÁXIMA VEROSSIMILHANÇA) E SHAPIRO-WILK.
- FÓRMULA DO TESTE χ^2 :

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i},$$

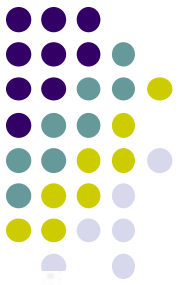
f_i = frequência observada

$\hat{f}_i$ = frequência esperada



EXEMPLO:

- TESTE DAS SUPOSIÇÕES DA HERANÇA MENDELIANA (GENÉTICA).
- VAMOS SUPOR QUE SEGUINDO AS LEIS DE MENDEL, UMA POPULAÇÃO DEVE TER A FREQUÊNCIA 9:3:3:1.
- NUM EXPERIMENTO COM UMA DETERMINADA POPULAÇÃO OS VALORES OBTIDOS FORAM 72:31:43:4.
- PERGUNTA: ESSA POPULAÇÃO ESTÁ DE ACORDO COM A TEORIA DA HERANÇA MENDELIANA?

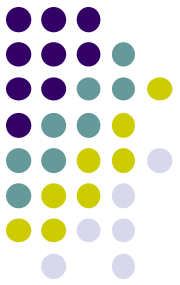


Gregório Mendel

- Monge Agostiniano austríaco que publicou em 1865 o livro Tratado sobre Plantas Híbridas.
- Marca não somente o início da genética como uma disciplina científica, mas também do uso sistemático da matemática, medidas quantitativas e estatística aplicadas na BIOLOGIA.
- É considerado o pai da Genética.



Leis de Mendel

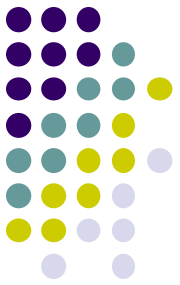


- Trabalhou com linhagens de ervilha (alta e baixa, lisa e rugosa, etc.)
- Fez diversos cruzamentos entre lisas e rugosas obtendo ervilhas lisas ou rugosas apenas.
- Quando cruzou lisa com rugosa obteve 3 ervilhas lisas para cada rugosa (daí a rugosa é considerada recessiva em relação a lisa).

SOLUÇÃO:

H_0 = OS DADOS EXPERIMENTAIS SEGUEM A DISTRIBUIÇÃO DE MENDEL.

H_A = OS DADOS NÃO SEGUEM A DISTRIBUIÇÃO DE MENDEL.



<i>FENÓTIPO</i>	<i>FREQÜÊNCIA OBSERVADA</i>	<i>FREQÜÊNCIA ESPERADA</i>	χ^2
A	72	$9/16 \cdot 150 = 84,37$	1,81
B	31	$3/16 \cdot 150 = 28,12$	0,29
C	43	$3/16 \cdot 150 = 28,12$	7,87
D	4	$1/16 \cdot 150 = 9,37$	3,08
<i>TOTAL</i>	<i>150</i>	<i>150</i>	<i>13,06</i>

TABELA $\chi^2_{(0,05, 3)} = 7,82$

$k = 4, \quad \text{g.l.} = k-1$

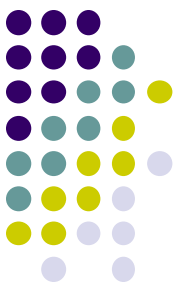
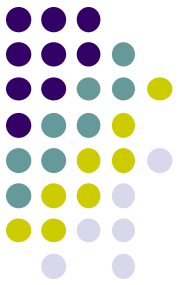


TABELA DE χ^2

Graus de Liberdade	PROBABILIDADE		
	0,10	0,05	0,01
1	2,71	3,84	6,64
2	4,60	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,72
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,80
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	35,17	41,64
24	33,20	36,42	42,98
25	34,38	37,65	44,31
26	35,56	38,88	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89

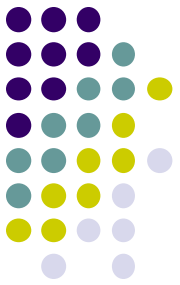


Teste de ajustamento

- Dado jogado 120 vezes.
- Qual deveria ser as frequências dos números 1, 2, 3, 4, 5 e 6?
- Será que o dado está “correto”?



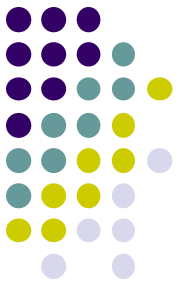
Exemplo de teste de ajustamento:



	“1”	“2”	“3”	“4”	“5”	“6”	Total
Obs	15	23	19	26	21	16	120
Exp							120

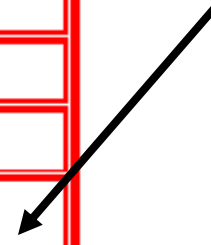


Exemplo de teste de ajustamento (continua)



	O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
“1”	15	20	-5	25	1,25
“2”	23	20	3	9	0,45
“3”	19	20	-1	1	0,05
“4”	26	20	6	36	1,80
“5”	21	20	1	1	0,05
“6”	16	20	-4	16	0,80
Total	120	120			4,4

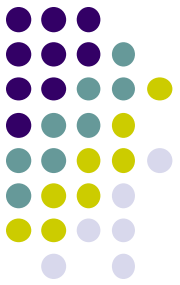
Valor do teste



$$G.L. = k-1 = 6-1 = 5$$

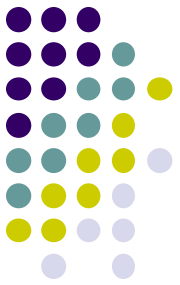
Conclusão ????

$$\text{Tabela de } \chi^2 = 11,070$$



Programa SAS

- ODS PDF FILE='C:\Arquivos2015\Bioestatistica2015\HONEST_DADOS.PDF';
- TITLE1'*** Análise para verificar se os dados são Honestos ***';
- TITLE3'*** Análise de Qui-Quadrado ***';
- **DATA DADO;**
- INPUT FACE VALOR;
- DATALINES;
- 1 15
- 2 23
- 3 19
- 4 26
- 5 21
- 6 16
- ;;;
- **PROC FREQ DATA=DADO;**
- WEIGHT VALOR;
- TABLES FACE/ TESTP=(0.1666 0.1666 0.1666 0.1666 0.1666 0.1667) CHISQ;
- **RUN;**
- ODS PDF CLOSE;



Resultado do Programa

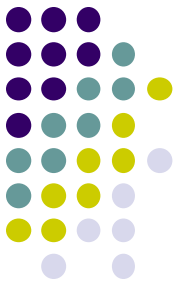
*** *Análise para verificar se os dados são Honestos* ***
The FREQ Procedure

*** *Análise de Qui-Quadrado* ***

FACE	Frequency	Percent	Test Percent
1	15	12.50	16.66
2	23	19.17	16.66
3	19	15.83	16.66
4	26	21.67	16.66
5	21	17.50	16.66
6	16	13.33	16.67

Chi-Square Test for Specified Proportions	
Chi-Square	4.4061
DF	5
Pr > ChiSq	0.4925

Sample Size = 120



Exercício

- Um pesquisador deseja saber se uma determinada vacina apresenta o mesmo tipo de reação adversa para indivíduos da raça humana em diferentes faixas etárias. Para tanto tomou 100 indivíduos em cada faixa etária e verificou se apresentavam febre acima de 39°C após 24 horas da aplicação da vacina. Os dados estão na próxima transparência. Defina e teste as hipóteses.

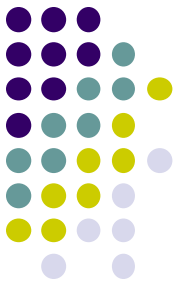
Dados



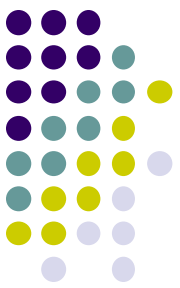
Faixa Etária	Febre	Faixa Etária	Febre
0-5 anos	10	35-45 anos	11
5-15 anos	12	45-55 anos	14
15-25 anos	9	55-65 anos	9
25-35 anos	8	Acima de 65 anos	7

Total de indivíduos com febre= 80

TABELAS DE CONTINGÊNCIA



- DADOS DE CONTAGEM E CLASSIFICADOS EM DUAS OU MAIS CATEGORIAS.
- DESEJA-SE TESTAR A HIPÓTESE DE QUE AS FREQUÊNCIAS DE OCORRÊNCIA NAS VÁRIAS CATEGORIAS SÃO INDEPENDENTES DESSAS CATEGORIAS (Teste de Associação).
- H_0 = AS CATEGORIAS SÃO INDEPENDENTES
- H_A = AS CATEGORIAS NÃO SÃO INDEPENDENTES.



Muito usado na pesquisa médica e social

	Câncer de Pulmão	
	Sim	Não
Fumantes	566	496
Não Fumantes	153	294

$$\chi^2 = 45,8513, \quad p < 0,0001$$

Programa SAS



```
ODS PDF
FILE='C:\Arquivos2015\Bioestatistica2015\FUMANTES.PDF';
TITLE1'*** Análise para verificar a influência do fumo no
cancer de pulmão ***';
TITLE3'*** Análise de Qui-Quadrado ***';
DATA FUMO;
INPUT FUMA $ CANCER $ VALOR;
DATALINES;
SIM SIM 566
SIM NÃO 496
NÃO SIM 153
NÃO NÃO 294
;;;
PROC FREQ DATA=FUMO ;
WEIGHT VALOR;
TABLES FUMA*CANCER/CHISQ;
RUN;
ODS PDF CLOSE;
```

Resultado do Programa

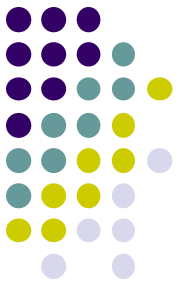
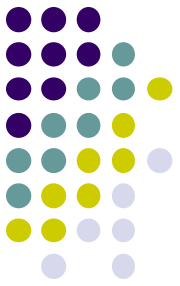


Table of FUMA by CANCER			
FUMA	CANCER		
Frequency			
Percent			
Row Pct			
Col Pct	NÃO	SIM	Total
NÃO	294	153	447
	19.48	10.14	29.62
	65.77	34.23	
	37.22	21.28	
SIM	496	566	1062
	32.87	37.51	70.38
	46.70	53.30	
	62.78	78.72	
Total	790	719	1509
	52.35	47.65	100.00

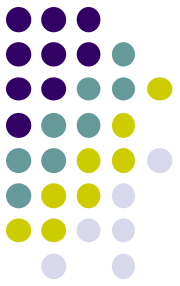
Statistic	DF	Value	Prob
Chi-Square	1	45.8513	<.0001
Likelihood Ratio Chi-Square	1	46.5205	<.0001
Continuity Adj. Chi-Square	1	45.0901	<.0001
Mantel-Haenszel Chi-Square	1	45.8209	<.0001
Phi Coefficient		0.1743	
Contingency Coefficient		0.1717	
Cramer's V		0.1743	

Teste de independência



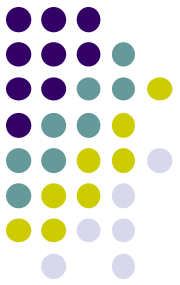
- Um pesquisador está interessado em determinar se existe relação entre a idade das pessoas que freqüentam um parque e a origem das pessoas (Brasil ou estrangeiro). Seleciona 180 pessoas sistematicamente e aplica um questionário, obtendo os seguintes dados:

Teste de independência (continua)



<i>Faixa Etária</i>	<i>Brasil</i>	<i>Exterior</i>	<i>Total</i>
Até 25 anos	22	38	60
25 a 40 anos	27	33	60
Maior que 40 anos	33	27	60
Total	82	98	180

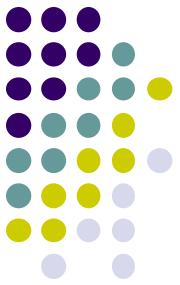
Teste de independência (continua)



$$E = (L_{\text{tot}} \times C_{\text{tot}})/n$$

Células	O	E	O-E	(O-E) ²	(O-E) ² /E
A	22	27,33	-5,33	28,41	1,04
B	27	27,33	-0,33	0,11	0,004
C	33	27,33	5,67	32,15	1,18
D	38	32,67	5,33	28,41	0,87
E	33	32,67	0,33	0,11	0,003
F	27	32,67	-5,67	32,15	0,98
					4,077

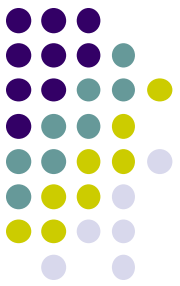
$\chi^2_{\text{CALC}} = 4,077$
 $\chi^2_{\text{CRIT}} = 5,991$
 $\chi^2_{\text{CALC}} > \chi^2_{\text{CRIT}}$ Aceita H_0
 $G.L. = (l-1)(c-1) = 2 * 1 = 2$



EXEMPLO:

- DESEJA-SE SABER SE HÁ INDEPENDÊNCIA NA OCORRÊNCIA DE DUAS ESPÉCIES EM DOIS LOCAIS DIFERENTES. PARA ESTE TESTE DE HIPÓTESES ORGANIZA-SE UMA TABELA DE FREQUÊNCIAS CHAMADA DE TABELA DE CONTINGÊNCIA.
- **HOUAISS: CONTINGÊNCIA:**
 - relação entre dois atributos, quando sua medida é baseada no afastamento de suas frequências em relação aos valores que teriam, na hipótese de independência estatística

EXEMPLOS



	<i>LOCAIS</i>	
<i>ESPÉCIES</i>	<i>1</i>	<i>2</i>
<i>A</i>	<i>10</i>	<i>10</i>
<i>B</i>	<i>10</i>	<i>10</i>

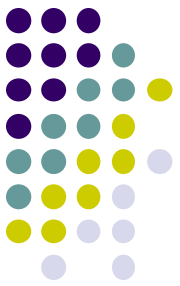
N.S.

	<i>LOCAIS</i>	
<i>ESPÉCIES</i>	<i>1</i>	<i>2</i>
<i>A</i>	<i>5</i>	<i>15</i>
<i>B</i>	<i>5</i>	<i>15</i>

N.S.

	<i>LOCAIS</i>	
<i>ESPÉCIES</i>	<i>1</i>	<i>2</i>
<i>A</i>	<i>15</i>	<i>5</i>
<i>B</i>	<i>5</i>	<i>15</i>

**



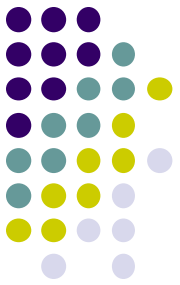
TABELAS DE CONTINGÊNCIA COM MAIOR NÚMERO DE CÉLULAS

- VAMOS SUPOR QUE ESTAMOS ESTUDANDO A OCORRÊNCIA DE CEDRO (*CEDRELA sp*) EM DIVERSAS FORMAÇÕES FLORESTAIS NA REGIÃO AMAZÔNICA E FAZEMOS UMA AMOSTRAGEM NAS DIFERENTES FORMAÇÕES. NESTAS PARCELAS AVALIAMOS A PRESENÇA OU AUSÊNCIA DA *CEDRELA sp*. OBTENDO OS SEGUINTE RESULTADOS:

<i>CEDRELA</i>	<i>ALTA FLORESTA (MT)</i>	<i>ACARÁ (PA)</i>	<i>AÇAILÂNDIA (MA)</i>	<i>BALBINA (AM)</i>
PRESENTE	32	43	16	9
AUSENTE	55	65	64	16

TABELA DE CONTINGÊNCIA: 2 X 4

TESTE DE HIPÓTESE:

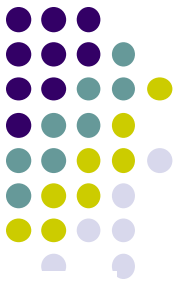


- H_0 = A OCORRÊNCIA DA CEDRELA sp É INDEPENDENTE DO LOCAL.
- H_A = A OCORRÊNCIA DA CEDRELA DEPENDE DO LOCAL.

CEDRELA	AF (MT)	AC(PA)	AÇ(MA)	BA(AM)	TOTAL
PRE SENÇA	32 (29,00)	43 (36,00)	16 (26,67)	9 (8,33)	100
AUSÊNCIA	55 (58,00)	65 (72,00)	64 (53,33)	16 (16,67)	200
TOTAL	87	108	80	25	300

GRAUS DE LIBERDADE: g.l. = (C-1) (L-1) = (4-1) (2-1) = 3

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} = \frac{(32 - 29,00)^2}{29,00} + \frac{(43 - 36,00)^2}{36,00} + \dots + \frac{(16 - 16,67)^2}{16,67} = 8,987$$

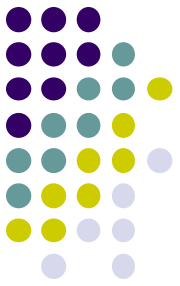


A TABELA 2 X 2

CATEGORIA	A	B	TOTAL
A	f_{11}	f_{12}	R_1
B	f_{21}	f_{22}	R_2
TOTAL	C_1	C_2	n

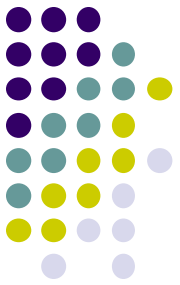
AS TABELAS DE CONTINGÊNCIA COM MAIS DE 2 LINHAS E 2 COLUNAS PODEM SER SUBDIVIDIDAS EM TABELAS 2 X 2 PARA ESTUDAR CASO A CASO. POR EXEMPLO, A TABELA DA CEDRELA PODE SER SUBDIVIDIDA EM DIVERSAS TABELAS 2 X 2, PARA ESTUDAR PARES DE LOCAIS.

Teste exato de Fisher (Fisher's exact test)



- Usado em tabelas de contingência 2×2 (testes de associação) e/ou quando o tamanho da amostra tomada e os valores esperados são pequenos. Útil quando o valor esperado de pelo menos uma célula é menor que 5.
- Amostras independentes e aleatórias.
- Calcula a probabilidade que cada tabela de contingência foi obtida por acaso.

Escolha entre o Qui-quadrado de Pearson e o teste exato de Fisher.



■ Amostra de tamanho pequeno

- ❑ Se menor que 20-40
- ❑ Se o número esperado de observações na célula for menor que 5.
- ❑ Use o Teste Exato de Fisher

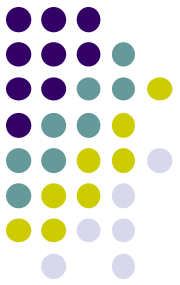
■ Amostra de tamanho pequeno

- ❑ Algumas centenas de dados, use qualquer um

■ Amostra de tamanho grande

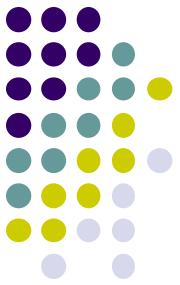
- ❑ Milhares de dados
- ❑ Use qui-quadrado

Exemplo do teste exato de Fisher.



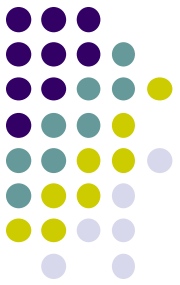
ODS PDF

```
FILE='C:\Arquivos2015\Bioestatistica2015\HIV_AZT.PDF';  
TITLE1'*** Análise para verificar a influência do do AZT na  
progressão do vírus HIV ***';  
TITLE3'*** Análise de Qui-Quadrado ***';  
DATA HIV;  
INPUT TRATAMENTO $ PROGRESSO $ INDIVID;  
DATALINES;  
AZT  SIM    76  
AZT  NÃO   399  
PLACEBO  SIM   129  
PLACEBO  NÃO   332  
;;;  
PROC FREQ DATA=HIV ;  
WEIGHT INDIVID;  
TABLES TRATAMENTO*PROGRESSO/FISHER;  
RUN;  
ODS PDF CLOSE;
```

Resultado do SAS

Table of TRATAMENTO by PROGRESSO			
TRATAMENTO	PROGRESSO		
Frequency Percent Row Pct Col Pct			
	NÃO	SIM	Total
AZT	399	76	475
	42.63	8.12	50.75
	84.00	16.00	
	54.58	37.07	
PLACEBO	332	129	461
	35.47	13.78	49.25
	72.02	27.98	
	45.42	62.93	
Total	731	205	936
	78.10	21.90	100.00



Resultado do teste exato de Fisher.

Statistic	DF	Value	Prob
Chi-Square	1	19.6383	<.0001
Likelihood Ratio Chi-Square	1	19.7994	<.0001
Continuity Adj. Chi-Square	1	18.9440	<.0001
Mantel-Haenszel Chi-Square	1	19.6174	<.0001
Phi Coefficient		0.1448	
Contingency Coefficient		0.1434	
Cramer's V		0.1448	

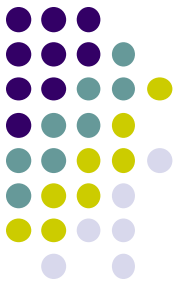
Fisher's Exact Test	
Cell (1,1) Frequency (F)	399
Left-sided Pr $\leq$ F	1.0000
Right-sided Pr $\geq$ F	6.251E-06
Table Probability (P)	3.285E-06
Two-sided Pr $\leq$ P	9.240E-06

O TESTE G^2

- O TESTE χ^2 FOI ENTÃO PROPOSTO POR PEARSON EM 1900 E A SUA APLICAÇÃO FOI PROPOSTA POR FISHER EM 1922. O TESTE G^2 FOI PROPOSTO POR WILKS EM 1935.

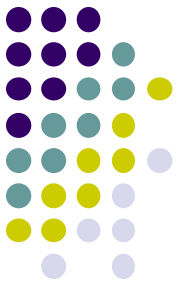


S.S. WILKS



Samuel Stanley Wilks

- NESCEU EM 17 DE JUNHO DE 1906 EM Little Elm, Texas, USA
FALECEU EM 7 DE MARÇO DE 1964 EM Princeton, Nova Jersey, USA
- ESTUDOU NO TEXAS E FEZ SEU DOUTORADO EM IOWA STATE UNIVERSITY.
- PASSOU 2 ANOS NA INGLATERRA (1931 E 1932) QUANDO TRABALHOU COM PEARSON E COM FISHER
- EM 1933 FOI PARA A UNIVERSIDADE DE PRINCETON, NOVA JERSEY, COMO PROFESSOR DE ESTATÍSTICA MATEMÁTICA, ONDE PERMANECEU ATÉ O FIM DA SUA CARREIRA PROFISSIONAL.
- TRABALHOU PARA O GOVERNO E EM 1947 RECEBEU DO PRESIDENTE DOS E.U. O CERTIFICADO DE MÉRITO PELOS TRABALHOS DESENVOLVIDOS NA GUERRA ANTI-SUBMARINA E NA SOLUÇÃO DE PROBLEMAS RELACIONADOS A DISTRIBUIÇÃO DE TROPAS MILITARES.



VOLTAMOS AO TESTE G^2

- UTILIZA A RAZÃO DA VEROSSIMILHANÇA E APRESENTA A SEGUINTE FÓRMULA:

$$G^2 = 2 \cdot \sum f_i \cdot \ln \frac{f_i}{\hat{f}_i} = 4,60517 \cdot \sum f_i \log_{10} \frac{f_i}{\hat{f}_i}$$



TESTE G^2

- RECOMENDA-SE USAR O TESTE G^2 NO LUGAR DO TESTE χ^2 QUANDO:

$$|f_i - \hat{f}_i| < \hat{f}_i \text{ PARA QUALQUER CÉLULA}$$

PARA UMA TABELA DE CONTINGÊNCIA A
FÓRMULA DE G^2 É:

$$G^2 = 2 \left[\sum_i \sum_j f_{ij} \ln f_{ij} - \sum_i R_i \ln R_i - \sum_j C_j \ln C_j + n \ln n \right]$$

TANTO O TESTE G^2 COMO O χ^2 APRESENTAM RESULTADOS SEMELHANTES.

QUANDO OS RESULTADOS SÃO DIFERENTES (UM SIGNIFICATIVO E OUTRO NÃO, POR EXEMPLO) RECOMENDAM USAR O TESTE G^2 .



```
title1'+-----+';  
title2'|          Estudo da influência da Aspirina no          |';  
title3'|          controle do infarto do miocárdio          |';  
title4'|          New England Journal of Medicine,          |';  
title5'|          318:262-264 (1988)          |';  
title6'+-----+';
```

```
data a;  
input grupo $ infarto $ resposta;  
datalines;  
placebo sim 189  
placebo nao 10845  
aspirina sim 104  
aspirina nao 10933  
;  
proc freq data=a;  
    table grupo*infarto/chisq;  
weight resposta;  
run;
```

PROGRAMA SAS PARA
ANÁLISE ESTATÍSTICA DE
DADOS DE CONTAGEM

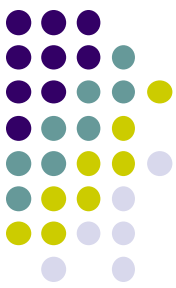
+-----+
 |
 | **Estudo da influência da Aspirina no**
 | **controle do infarto do miocárdio**
 | **New England Journal of Medicine,**
 | **318:262-264 (1988)**
 +-----+

1



TABLE OF GRUPO BY INFARTO

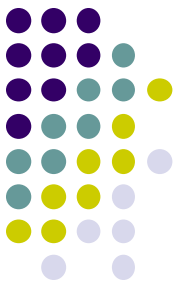
GRUPO	INFARTO		
Frequency			
Percent			
Row Pct			
Col Pct	nao	sim	Total
aspirina	10933	104	11037
	49.54	0.47	50.01
	99.06	0.94	
	50.20	35.49	
placebo	10845	189	11034
	49.14	0.86	49.99
	98.29	1.71	
	49.80	64.51	
Total	21778	293	22071
	98.67	1.33	100.00



STATISTICS FOR TABLE OF GRUPO BY INFARTO

Statistic	DF	Value	Prob
Chi-Square	1	25.014	0.001
Likelihood Ratio Chi-Square	1	25.372	0.001
Continuity Adj. Chi-Square	1	24.429	0.001
Mantel-Haenszel Chi-Square	1	25.013	0.001
Fisher's Exact Test (Left)			1.000
(Right)			3.25E-07
(2-Tail)			5.03E-07
Phi Coefficient		0.034	
Contingency Coefficient		0.034	
Cramer's V		0.034	

Sample Size = 22071



```
proc format;  
value $doe 'p'='PRESENTE'  
           'a'='AUSENTE';  
value esp 1='Acácia'  
         2='Paricá'  
         3='Teca'  
         4='Tatajuba';
```

```
run;
```

```
data a;  
input doenca $ especie arvores;  
datalines;
```

```
p 1 32
```

```
p 2 43
```

```
p 3 16
```

```
p 4 9
```

```
a 1 55
```

```
a 2 65
```

```
a 3 64
```

```
a 4 16
```

```
;
```

```
proc freq data=a;
```

```
weight arvores;
```

```
format doenca $doe. especie esp.;
```

```
label doenca='Doença'
```

```
especie='Espécie';
```

```
table doenca*especie/chisq exact;
```

```
run;
```

PROGRAMA SAS PARA ANÁLISE DE DADOS DE CONTAGENS



The SAS System

1

TABLE OF DOENCA BY ESPECIE

DOENCA(Doença)	ESPECIE(Espécie)				
Frequency Percent Row Pct Col Pct	Acácia	Paricá	Teca	Tatajuba	Total
AUSENTE	55 18.33 27.50 63.22	65 21.67 32.50 60.19	64 21.33 32.00 80.00	16 5.33 8.00 64.00	200 66.67
PRESENTE	32 10.67 32.00 36.78	43 14.33 43.00 39.81	16 5.33 16.00 20.00	9 3.00 9.00 36.00	100 33.33
Total	87 29.00	108 36.00	80 26.67	25 8.33	300 100.00

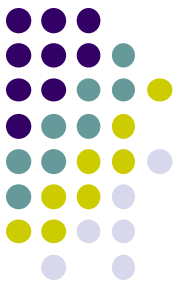


STATISTICS FOR TABLE OF DOENCA BY ESPECIE

Statistic	DF	Value	Prob
Chi-Square	3	8.987	0.029
Likelihood Ratio Chi-Square	3	9.512	0.023
Mantel-Haenszel Chi-Square	1	2.616	0.106
Fisher's Exact Test (2-Tail)			0.024
Phi Coefficient		0.173	
Contingency Coefficient		0.171	
Cramer's V		0.173	

Sample Size = 300

TESTE
G



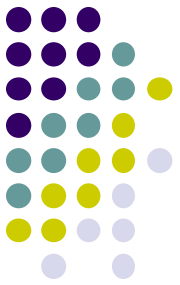
Exemplo

		Resposta		
Trilha	Sexo	Ruim	Média	Boa
A	M	6	5	16
	F	7	2	5
B	M	19	7	6
	F	10	0	1
Total		42	14	28

Teste G^2



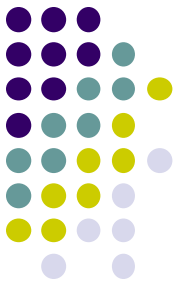
```
TITLE1'+-----+';
TITLE2'|   Estudo de Trilhas de Interpretação   |';
TITLE3'+----- 2015-----+';
DATA A;
INPUT TRILHA $ SEXO $ RESPOSTA $ INDIV;
DATALINES;
A F  RUIM    6
A M  RUIM    7
A F  MÉDIA   5
A M  MÉDIA   2
A F  BOA    16
A M  BOA     5
B F  RUIM   19
B M  RUIM   10
B F  MÉDIA   7
B M  MÉDIA   0
B F  BOA     5
B M  BOA     1
;;;
ODS RTF FILE='C:\Arquivos2014\Bioestatística2014\QUIQUAD.RTF';
ODS GRAPHICS ON;
PROC FREQ DATA=A;
WEIGHT INDIV;
TABLE TRILHA * RESPOSTA/CHISQ PLOTS=FREQPLOT(TYPE=DOT);
RUN;
PROC FREQ DATA=A;
WEIGHT INDIV;
TABLE TRILHA*SEXO/CHISQ PLOTS=FREQPLOT(TYPE=DOT);
RUN;
PROC FREQ DATA=A;
WEIGHT INDIV;
TABLE SEXO*RESPOSTA/CHISQ PLOTS=FREQPLOT(TYPE=DOT);
RUN;
ODS GRAPHICS OFF;
ODS RTF CLOSE;
```



Resultado Teste G^2

Table of TRILHA by RESPOSTA				
TRILHA	RESPOSTA			
Frequency				
Percent				
Row Pct				
Col Pct	BOA	MÉDIA	RUIM	Total
A	21	7	13	41
	25.30	8.43	15.66	49.40
	51.22	17.07	31.71	
	77.78	50.00	30.95	
B	6	7	29	42
	7.23	8.43	34.94	50.60
	14.29	16.67	69.05	
	22.22	50.00	69.05	
Total	27	14	42	83
	32.53	16.87	50.60	100.00

Statistic	DF	Value	Prob
Chi-Square	2	14.4186	0.0007
Likelihood Ratio Chi-Square	2	15.0657	0.0005
Mantel-Haenszel Chi-Square	1	14.1580	0.0002
Phi Coefficient		0.4168	
Contingency Coefficient		0.3847	
Cramer's V		0.4168	



Mais resultados (Teste G^2)

Table of TRILHA by SEXO			
TRILHA	SEXO		
Frequency			
Percent			
Row Pct			
Col Pct	F	M	Total
A	27	14	41
	32.53	16.87	49.40
	65.85	34.15	
	46.55	56.00	
B	31	11	42
	37.35	13.25	50.60
	73.81	26.19	
	53.45	44.00	
Total	58	25	83
	69.88	30.12	100.00

Statistic	DF	Value	Prob
Chi-Square	1	0.6239	0.4296
Likelihood Ratio Chi-Square	1	0.6249	0.4292
Continuity Adj. Chi-Square	1	0.3032	0.5819
Mantel-Haenszel Chi-Square	1	0.6164	0.4324
Phi Coefficient		-0.0867	
Contingency Coefficient		0.0864	
Cramer's V		-0.0867	

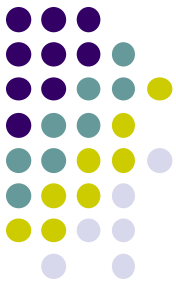
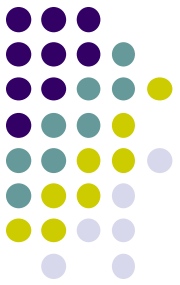


Table of SEXO by RESPOSTA				
SEXO	RESPOSTA			
Frequency Percent Row Pct Col Pct				
	BOA	MÉDIA	RUIM	Total
F	21	12	25	58
	25.30	14.46	30.12	69.88
	36.21	20.69	43.10	
	77.78	85.71	59.52	
M	6	2	17	25
	7.23	2.41	20.48	30.12
	24.00	8.00	68.00	
	22.22	14.29	40.48	
Total	27	14	42	83
	32.53	16.87	50.60	100.00

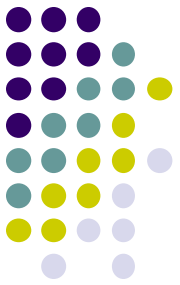
Statistic	DF	Value	Prob
Chi-Square	2	4.6079	0.0999
Likelihood Ratio Chi-Square	2	4.7938	0.0910
Mantel-Haenszel Chi-Square	1	2.9750	0.0846
Phi Coefficient		0.2356	
Contingency Coefficient		0.2293	
Cramer's V		0.2356	



Exercício

- Um analista de mercado fez um estudo com 286 consumidores perguntando a eles se compravam Coca-Cola Zero ou Pepsi Cola Zero ou ambas. Ele(a) deseja saber se existe relação entre o consumo dos dois refrigerantes. Se essa relação é positiva ou negativa. As duas marcas são competidoras? O que dizer para cada classe de renda?

Geral

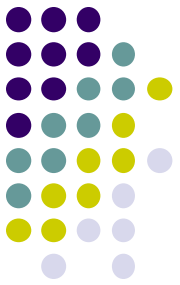


	Pepsi Zero	
Coca Zero	Não	Sim
Não	84	32
Sim	48	122

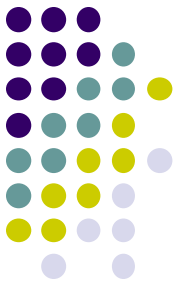
Classe de Renda Alta

	Pepsi Zero	
Coca Zero	Não	Sim
Não	4	30
Sim	40	2

Classe de Renda Baixa



	Pepsi Zero	
Coca Zero	Não	Sim
Não	80	2
Sim	8	120



*ATÉ A PRÓXIMA
SEMANA !!!*