



BIOESTATÍSTICA

SUPOSIÇÕES DOS MODELOS
PARA ANÁLISE DE DADOS.



MODELO MATEMÁTICO

- OS TESTES ESTATÍSTICOS PARA ANÁLISE DE DADOS UTILIZAM UM MODELO MATEMÁTICO
- $y_i = m + e_i$ (modelo mais simples)
- $y_{ij} = m + t_i + e_{ij}$ (inteiramente casualizado)
- $y_{ijk} = m + t_i + b_j + e_{ijk}$ (blocos ao acaso)



O QUE É MODELO?

- REPRESENTAÇÃO DE UM FENÔMENO OU DE UM OBJETO.
- TIPOS DE MODELOS:
 - ICÔNICOS (EX.: MAQUETE, ÍCONE)
 - ANALÓGICOS (EX.: RELÓGIO, TERMÔMETRO)
 - MATEMÁTICOS (EX.: ANÁLISE DE VARIÂNCIA, DE REGRESSÃO)



MODELOS MATEMÁTICOS

- DETERMINISTAS (EX.: ANÁLISE FINANCEIRA)
- ESTOCÁSTICOS (EX.: EXPERIMENTO, AMOSTRAGEM)
- DETERMINISMO: princípio segundo o qual tudo no universo, até mesmo a vontade humana, está submetido a leis necessárias e imutáveis, de tal forma que o comportamento humano está totalmente predeterminado pela natureza, e o sentimento de liberdade não passa de uma ilusão subjetiva (DICIONÁRIO HOUAISS)



MODELOS ESTOCÁSTICOS

- ESTOCASTICIDADE: doutrina que considera o acaso uma ocorrência objetiva, inerente aos processos e eventos da natureza, e não uma mera incapacidade de compreensão científica ou uma expressão da ignorância humana em relação as verdadeiras causas de um fenômeno; diz-se do que depende ou resulta de uma variável aleatória (DICIONÁRIO HOUAISS).



MODELOS ESTOCÁSTICOS

- LINEARES: $y_{ij} = m + t_i + e_{ij}$
- NÃO LINEARES:
 - $y_{ij} = m \cdot t_i + e_{ij}$
 - $y_{ij} = m + t_i^c + e_{ij}$



MODELO MATEMÁTICO

- CONJUNTO DE DADOS: 5, 8, 11, 21, 16, 5
- MÉDIA = $m = 11,0$
- MODELO MATEMÁTICO →
$$y_i = m + e_i$$



OUTRO EXEMPLO:

- TRAT 1: 5, 4, **8**, 2, 11 (MÉDIA=6,0)
- TRAT 2: 3, 4, 11, 5, 21 (MÉDIA = 8,8)
- MÉDIA GERAL = 7,4
- EFEITO TRAT 1 → $6,0 - 7,4 = -1,4$
- EFEITO TRAT 2 → $8,8 - 7,4 = +1,4$
- MODELO MATEMÁTICO:
 $y_{ij} = m + t_i + e_{ij}$ (e_{ij} = resíduo ou erro ou ruído)
- **8** = $7,4 - 1,4 + 2,0$



TESTES ESTATÍSTICOS

- CLASSIFICAÇÃO: PARAMÉTRICOS E NÃO PARAMÉTRICOS
- PARAMÉTRICOS: TESTE t , E F
- TESTE t → INFERÊNCIA SOBRE A MÉDIA DA POPULAÇÃO
- TESTE F → INFERÊNCIA SOBRE A VARIÂNCIA DA POPULAÇÃO
- **TESTE t** : 2 TRATAMENTOS E **TESTE F** PARA MAIS DE 2 TRATAMENTOS



SUPOSIÇÕES

- NORMALIDADE DOS DADOS OU DOS RESÍDUOS
- HOMOGENEIDADE DAS VARIÂNCIAS (HOMOSCEDASTICIDADE)
- ADITIVIDADE DOS EFEITOS (TRATAMENTOS, BLOCOS, ETC.)
- INDEPENDÊNCIA DOS TRATAMENTOS

O QUE OCASIONA ESTES PROBLEMAS?

- CARACTERÍSTICAS DE ALGUMAS VARIÁVEIS EM ESTUDO (EX.: DISTRIBUIÇÃO DE ESPÉCIES RARAS)
- VALORES PERDIDOS (EX.: FALTA DE MATERIAL DE ESTUDO)
- ERROS NA CONDUÇÃO DO EXPERIMENTO (EX.: TEMPERATURA FORA DO PADRÃO)
- ESCOLHA DOS TRATAMENTOS

TR1	TR2	TR3	TR4
0	0	1	100
0	0	1	100
0	0	1	100

O QUE OCASIONA ESTES PROBLEMAS?

- POUCAS RÉPLICAS (EX.: 2 OU 3 QUANDO NECESSÁRIAS 6 A 10 RÉPLICAS)
- AUSÊNCIA DE CASUALIZAÇÃO:

SEM CASUALIZAÇÃO



<i>TR 1</i>	<i>TR 2</i>	<i>TR 3</i>	<i>TR 4</i>
01	02	03	04
01	02	03	04





CASUALIZAÇÃO

COM CASUALIZAÇÃO



<i>TR 1</i>	<i>TR 2</i>	<i>TR 3</i>	<i>TR 4</i>
01	03	04	02
03	04	02	01





RÉPLICA OU REPETIÇÃO?

- REPLICAR: fazer réplica ou imitação de; reproduzir (PARA UM EXPERIMENTO)
- REPETIR: tornar a fazer, a usar, a executar; refazer (PARA UM TRATAMENTO)
- SEGUNDO HOUAISS

AINDA, O QUE OCASIONA ESTES PROBLEMAS?



- TAMANHO E FORMA DAS UNIDADES EXPERIMENTAIS (UNIDADES AMOSTRAIS OU PARCELAS)
- EX.: UM TANQUE PODE SER UMA UNIDADE EXPERIMENTAL COMPOSTA DE ALGUNS INDIVÍDUOS. QUANTOS?
- UNIDADES EXPERIMENTAIS PEQUENAS → MAIS PROBLEMAS (alta variação)
- UNIDADES EXPERIMENTAIS GRANDES → MAIS CUSTO
- DESAFIO: ENCONTRAR O PONTO DE EQUILÍBRIO



VALORES EXTREMOS (OUTLIERS):

- INERENTE AO MATERIAL EM ESTUDO
- ERRO NA COLETA DE DADOS
- PROBLEMAS NA CONDUÇÃO DO EXPERIMENTO (PRAGAS E DOENÇAS, P. EX.)



BIOESTATÍSTICA

NORMALIDADE



TESTE F

- O TESTE F É CONSIDERADO MUITO ROBUSTO AOS DESVIOS DA NORMALIDADE
- ROBUSTEZ: INSENSIBILIDADE AOS DESVIOS DAS SUPOSIÇÕES.
- SEMPRE QUE POSSÍVEL “REFUGIE-SE” NO TEOREMA DO LIMITE CENTRAL. EX.: ESTUDO DE ESSÊNCIAS AROMÁTICAS (5 A 10 PROVADORES OBTENDO-SE A MÉDIA).
- ASSIMETRIA E CURTOSE PRÓXIMAS DE ZERO = NORMAL.

EFEITO DOS DESVIOS DA NORMALIDADE

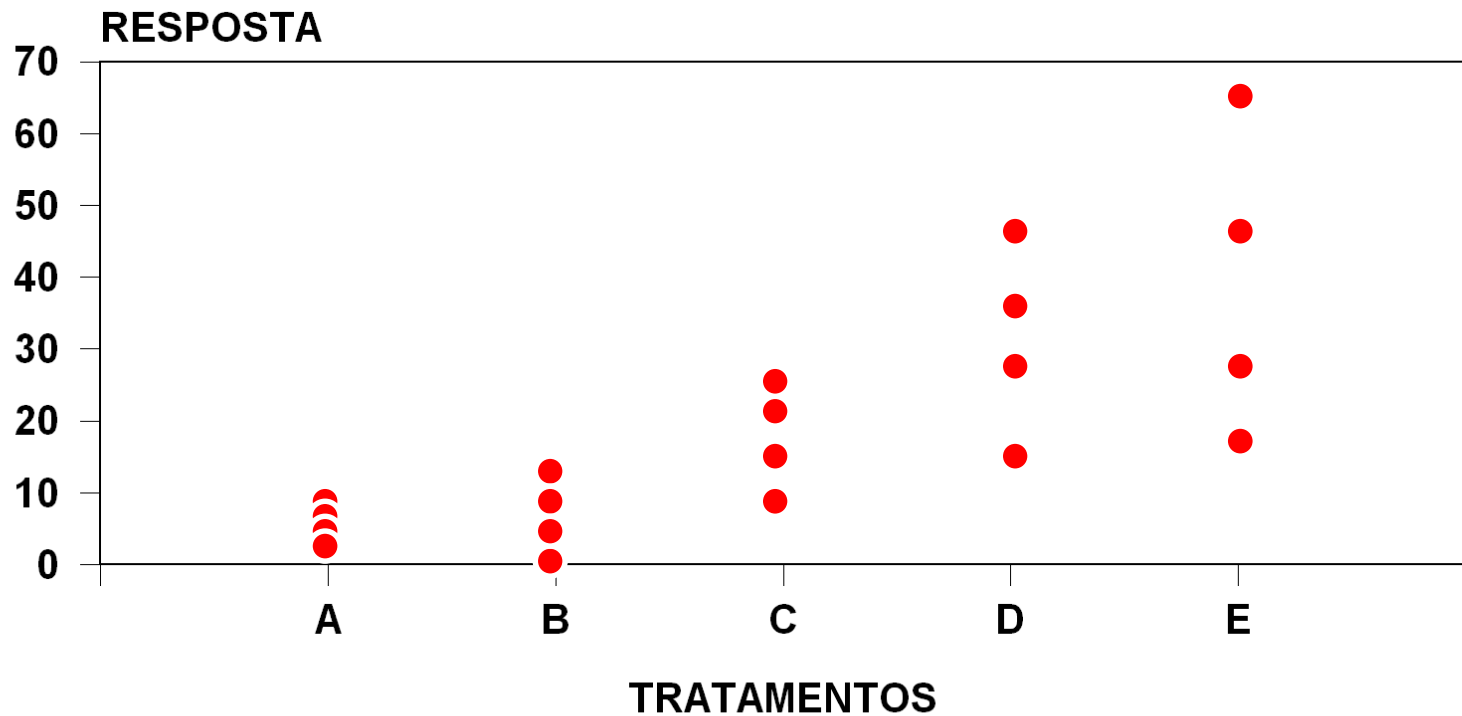
<i>Curtose</i>	<i>Teste F</i>	<i>Consequência</i>
> 0	Muito pequeno	Não rejeita a hipótese nula, embora ela seja incorreta
< 0	Muito grande	Rejeita a hipótese nula embora ela seja correta

A ASSIMETRIA NÃO É PARTICULARMENTE GRAVE PARA O TESTE F, MAS SIM PARA O TESTE t. NO CASO DE ASSIMETRIA DEVE-SE USAR TESTES NÃO PARAMÉTRICOS.

TESTE RECOMENDADO PARA VERIFICAR NORMALIDADE (GERALMENTE AMOSTRAS PEQUENAS) É O TESTE DE SHAPIRO-WILK

HETEROSCEDASTICIDADE

- NOS TESTES t E F, AS VARIÂNCIAS DOS TRATAMENTOS SÃO CONSIDERADAS COMO SENDO IDÊNTICAS (HOMOSCEDASTICIDADE OU HOMOGENEIDADE DAS VARIÂNCIAS).

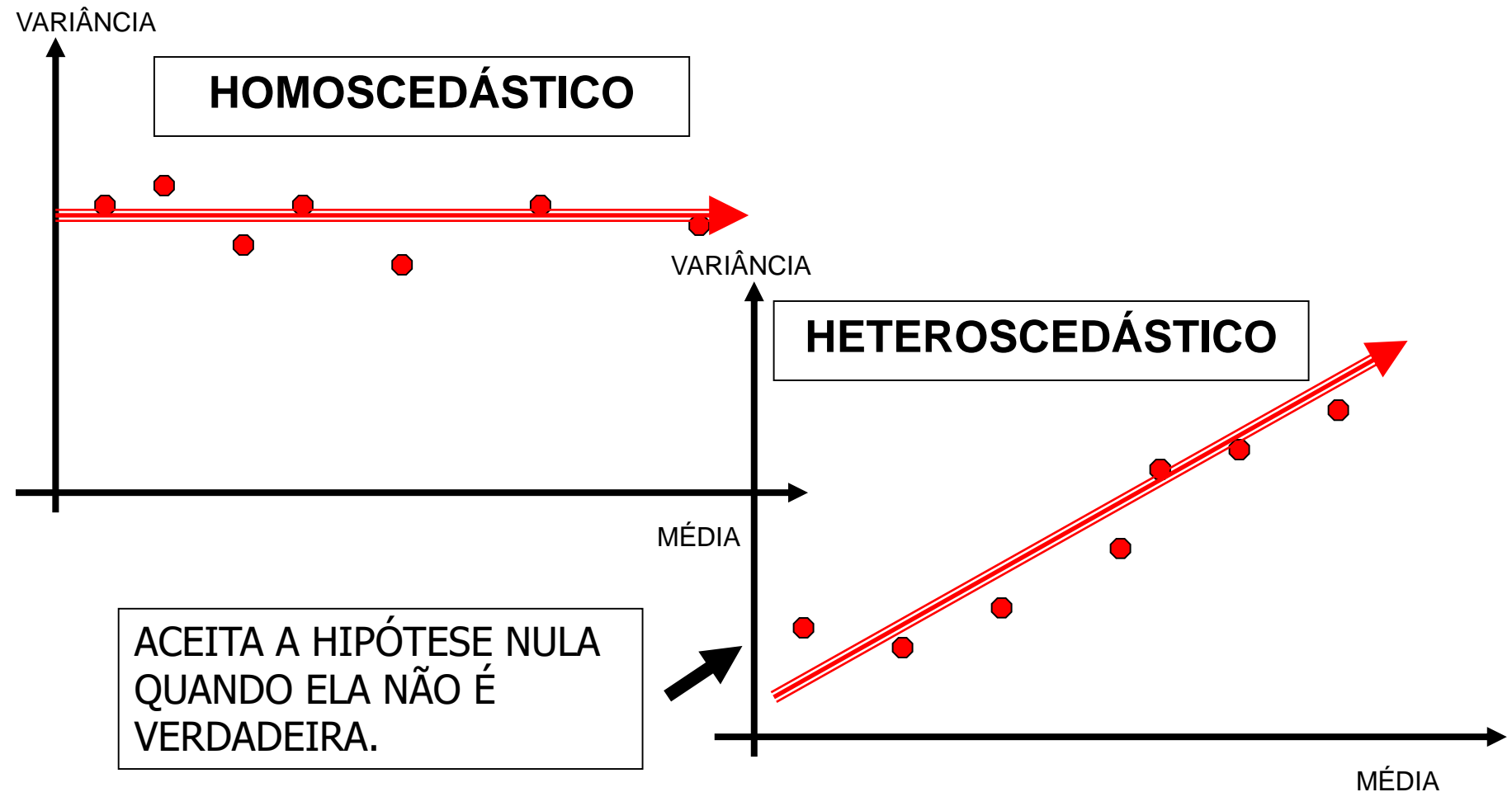




HETEROSCEDASTICIDADE

- OUTRO MEIO DE ESTUDAR A HETEROSCEDASTICIDADE É VERIFICAR A CORRELAÇÃO ENTRE A MÉDIA E A VARIÂNCIA DOS TRATAMENTOS

CORRELAÇÃO MÉDIA-VARIÂNCIA



ADITIVIDADE

- OS EFEITOS DOS TRATAMENTOS DEVEM SER ADITIVOS

Efeito aditivo

	<i>Fator A</i>		
<i>Fator B</i>	<i>Nível 1</i>	<i>Nível 2</i>	<i>Nível 3</i>
<i>Nível 1</i>	10	20	25
<i>Nível 2</i>	20	30	35
<i>s</i>	7,07	7,07	7,07

+10

Assimetria = -0,25

Curtose = -0,01

MAS PODE NÃO SER

Efeito multiplicativo

	<i>Fator A</i>		
<i>Fator B</i>	<i>Nível 1</i>	<i>Nível 2</i>	<i>Nível 3</i>
<i>Nível 1</i>	10	30	60
<i>Nível 2</i>	20	60	120
<i>s</i>	7,07	21,21	42,43

x 2

Assimetria = 1,15

Curtose = 1,29

SOLUÇÃO?

TRANSFORMAÇÃO DOS DADOS

Efeito aditivo (transformação logarítmica)

	<i>Fator A</i>		
<i>Fator B</i>	<i>Nível 1</i>	<i>Nível 2</i>	<i>Nível 3</i>
<i>Nível 1</i>	1,00	1,48	1,78
<i>Nível 2</i>	1,30	1,78	2,08
<i>s</i>	0,21	0,21	0,21

+ 0,30

Assimetria = -0,27

Curtose = -0,63

INDEPENDÊNCIA DOS TRATAMENTOS

- ESTÁ RELACIONADA A DISPOSIÇÃO DOS TRATAMENTOS E COLETA DOS DADOS EXPERIMENTAIS

A 4x4 grid representing an experimental design. The grid contains the following treatments in each cell:

T1	T3	T2	T4
T3	T2	T1	T4
T2	T4	T3	T1
T3	T4	T1	T2

Four blue arrows point downwards to the top of each column. Four green arrows point horizontally to the left of each row.

COMO COLETAR OS DADOS, SE TEMOS A
POSSIBILIDADE DE COLETAR APENAS 4
UNIDADES EXPERIMENTAIS (PARCELAS) POR
DIA?



OUTRO EXEMPLO:

- COMPARANDO 3 LOCAIS PARA AVALIAR A DENSIDADE POPULACIONAL DE CAPIVARAS. CADA LOCAL COM 4 REPETIÇÕES.

LOCAL	LOCAL	LOCAL
1	2	3
A	D	C
B	A	D
C	B	A
D	C	B

A, B, C e D = REPETIÇÕES

SÓ PODEMOS OBSERVAR 4 UNIDADES AMOSTRAIS / DIA

- OPÇÃO 1: 1A, 1B, 1C, 1D
- OPÇÃO 2: 1A, 2A, 3A, 1B
- OPÇÃO 3: 1A, 1B, 2A, 2B
- OPÇÃO 4: NENHUMA DELAS

LOCAL	LOCAL	LOCAL
1	2	3
A	D	C
B	A	D
C	B	A
D	C	B

A, B, C e D = REPETIÇÕES



TRANSFORMAÇÃO DE DADOS

- DOIS MODOS EM QUE AS SUPOSIÇÕES DA ANAVA **NÃO** SÃO VÁLIDAS:
- A) OS DADOS CONSISTEM EM MEDIÇÕES FEITAS NA ESCALA ORDINAL OU NOMINAL (DISCRETAS);
- B) MESMO QUANDO A MEDIÇÃO É FEITA NA ESCALA INTERVALAR ESTAS NÃO OBTÊM AS SUPOSIÇÕES ENUNCIADAS (NORMALIDADE, HOMOGENEIDADE DA VARIÂNCIA, ADITIVIDADE DO MODELO, INDEPENDÊNCIA DOS TRATAMENTOS).



SOLUÇÃO?

- USAR TESTES NÃO PARAMÉTRICOS
- TRANSFORMAR OS DADOS ANTES DA ANÁLISE.
- SE OPTAR PELA TRANSFORMAÇÃO DOS DADOS, MESMO APÓS TRANSFORMADOS, OS DADOS DEVEM OBDECER AS SUPOSIÇÕES DO MODELO.



TIPOS DE TRANSFORMAÇÃO

- MAIS USADAS:
- RAIZ QUADRADA
- ARCO SENO
- LOGARÍTIMO



RAIZ QUADRADA

- USADA PARA DADOS DE CONTAGEM QUE POSSUEM DISTRIBUIÇÃO PRÓXIMA DA POISSON: $\sqrt{y_i}$
- SE O CONJUNTO DE DADOS POSSUI MUITOS VALORES ZERO OU PRÓXIMOS DE ZERO RECOMENDA-SE A TRANSFORMAÇÃO: $\sqrt{y_i + 0,375}$



EXEMPLO:

- NÚMERO DE COLEÓPTEROS EM AMOSTRAS TOMADAS EM DIFERENTES TIPOS DE SOLO E LOCAIS.



ARCO SENO

- ARCO SENO RAIZ QUADRADA DA PROPORÇÃO
- USADA PARA PROPORÇÕES DA DISTRIBUIÇÃO BINOMIAL

$$y_i = \text{arco seno} \sqrt{p}$$

$$p = \text{proporção}$$

SAS → $y_i = \text{arsin} (\text{sqrt} (p));$



ARCO SENO

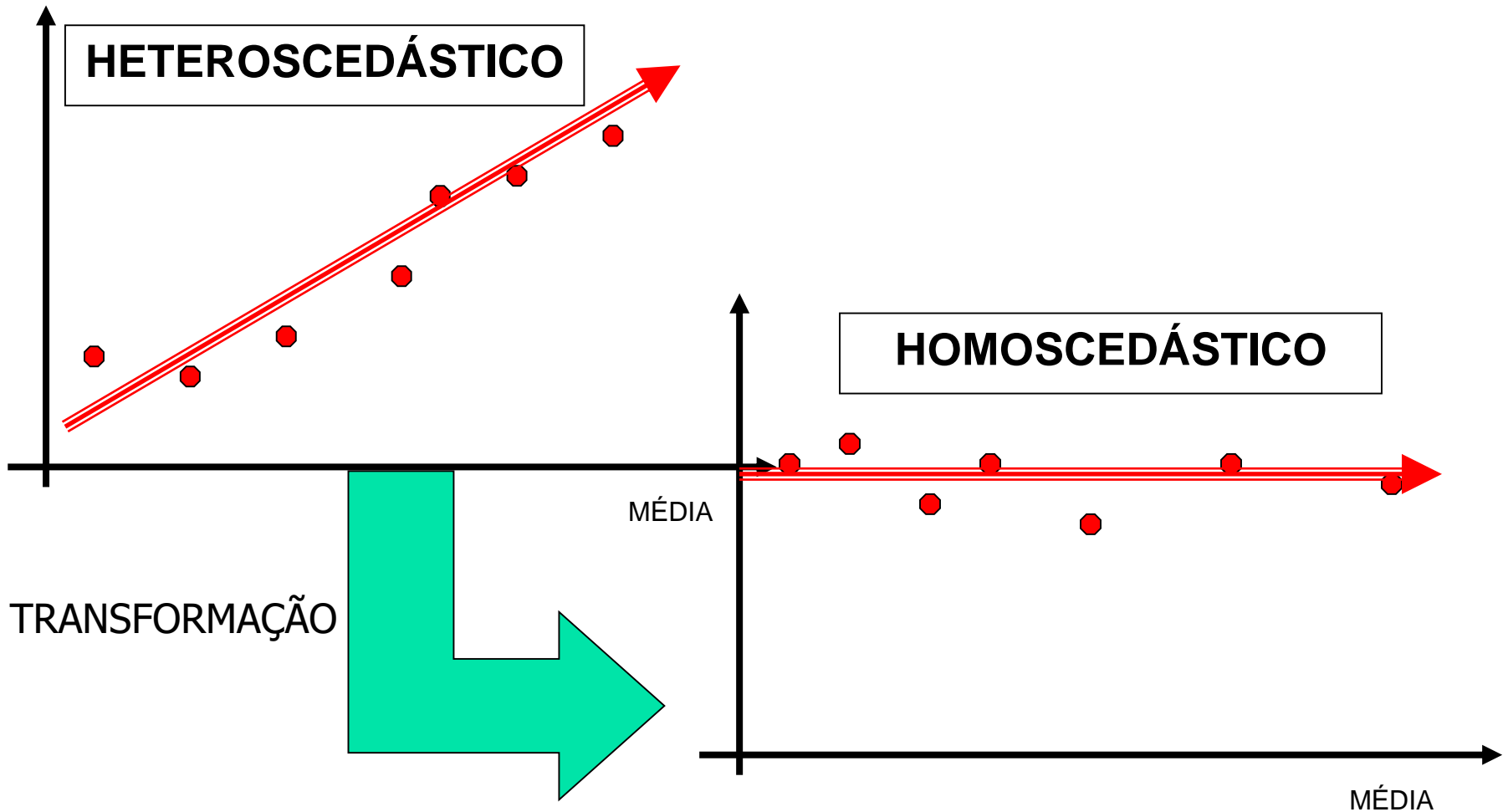
- PODE TAMBÉM SER USADA PARA PORCENTAGENS E PROPORÇÕES CUJOS VALORES VARIAM DE 0 A 30 OU DE 70 A 100%.
- ENTRE 30 E 70 % A TRANSFORMAÇÃO RARAMENTE É NECESSÁRIA.
- The Arcsine Transformation:
Has the time come for retirement?
Emily Wilson, Melanie Underwood,
Olivia Puckrin, Karla Letto,
Rebecca Doyle, Holly Caravan,
Stacey Camus, Kate Bassett (2010, U OF CALIFORNIA)



TRANSFORMAÇÃO LOGARÍTIMICA

- USADA PARA ESTABILIZAR A VARIÂNCIA (HETEROSCEDASTICIDADE).

BOX-COX





Exemplo

- Foi realizado um ensaio para testar a influência da aplicação de 5 tipos fungicidas em morango, com o objetivo de avaliar a quantidade de resíduo na fruta.
- Os valores obtidos são fornecidos em mg/kg.
- Fazer um programa SAS para transformar os dados em logaritmo neperiano:LOG

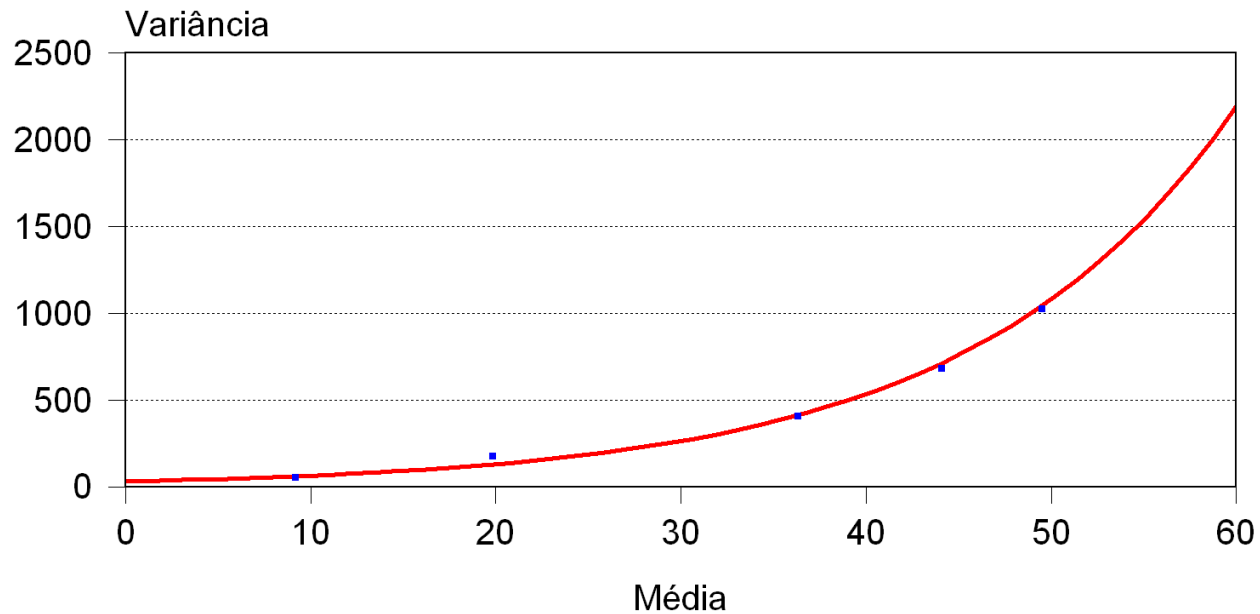


Dados

(Resíduo dos produtos mg/kg)

	Produto A	Produto B	Produto C	Produto D	Produto E
	49,7	7,6	17,7	82,4	41,9
	7,5	8,1	24,8	28,3	27,9
	35,9	11,1	84,7	11,2	8,6
	45,5	20,4	14,8	66,3	16,7
	60,9	3,8	55,5	33,8	17,2
	18,6	4,2	99,5	42,9	6,8
Média	36,34	9,21	49,51	44,12	19,82
Variância	402,68	49,42	1021,87	680,52	172,91

Relação entre média e variância: heteroscedasticidade

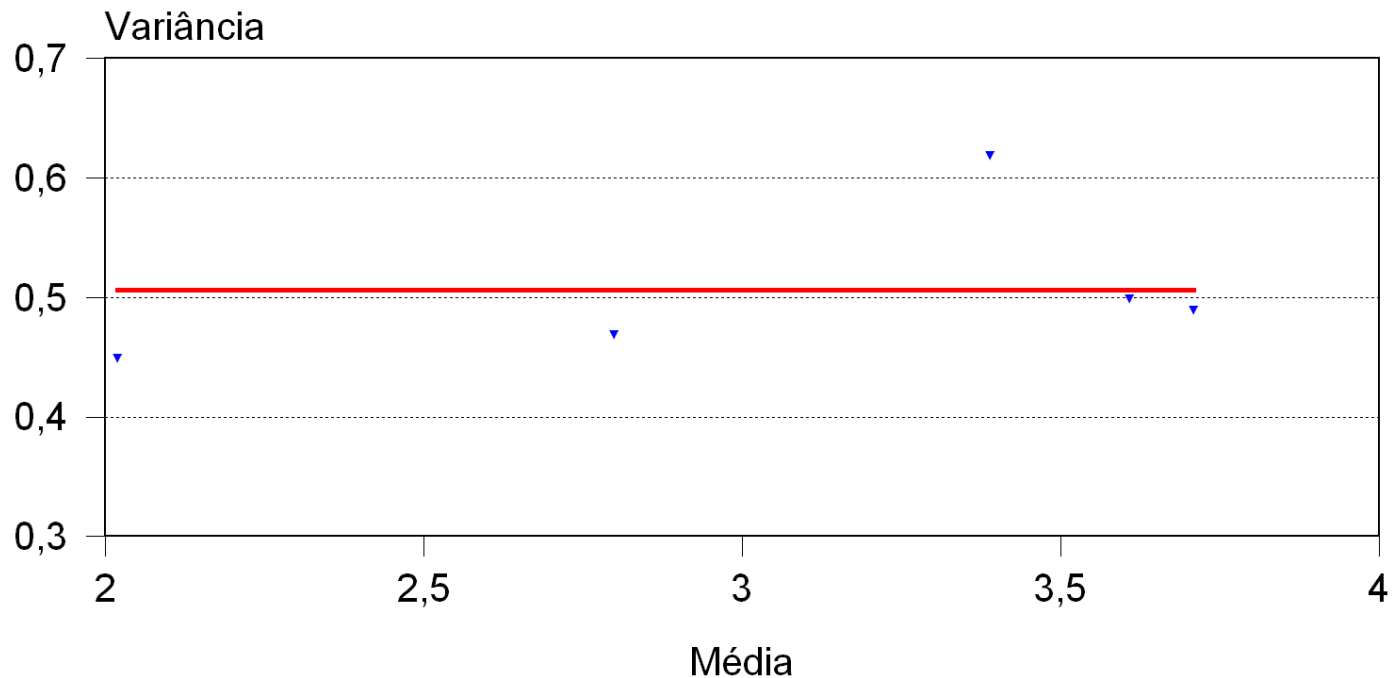




Transformação logarítmica

Produto A	Produto B	Produto C	Produto D	Produto E
3,91	2,02	2,87	4,41	3,73
2,01	2,09	3,21	3,34	3,33
3,58	2,41	4,44	2,41	2,15
3,82	3,02	2,70	4,19	2,82
4,11	1,34	4,02	3,52	2,84
2,92	1,44	4,60	3,76	1,91
3,39	2,02	3,71	3,61	2,80
0,62	0,45	0,49	0,50	0,47

Relação entre média e variância: dados transformados





PROGRAMA SAS (1)

DATA A;

INPUT TRATA TRATB TRATC TRATD

TRATE;

DATALINES;

49.7	7.6	17.7	82.4	41.9
7.5	8.1	24.8	28.3	27.9
35.9	11.1	84.7	11.2	8.6
45.5	20.4	14.8	66.3	16.7
60.9	3.8	55.5	33.8	17.2
18.6	4.2	99.5	42.9	6.8
;;;				



PROGRAMA SAS (2)

```
DATA B;  
SET A;  
DROP TRATA TRATB TRATC TRATD TRATE;  
TRAT='A'; RESID=TRATA;OUTPUT;  
TRAT='B'; RESID=TRATB;OUTPUT;  
TRAT='C'; RESID=TRATC;OUTPUT;  
TRAT='D'; RESID=TRATD;OUTPUT;  
TRAT='E'; RESID=TRATE;OUTPUT;  
RUN;
```



PROGRAMA SAS (3)

```
DATA C;  
SET B;  
TRESID=LOG(RESID);  
PROC SORT DATA=C;  
BY TRAT;  
PROC MEANS DATA=C NOPRINT;  
BY TRAT;  
VAR RESID TRESID;  
OUTPUT OUT=D MEAN=MEDIA MEDIAT  
VAR=VARIAN VARIANT;  
RUN;  
PROC PRINT DATA=D;  
RUN;
```



PROGRAMA SAS (4)

```
PROC REG DATA=D
PLOTS(ONLY)=PREDICTIONS(X=MEDIA
);
MODEL VARIAN = MEDIA;
RUN;
PROC REG DATA=D
PLOTS(ONLY)=PREDICTIONS(X=MEDIA
T);
MODEL VARIANT = MEDIAT;
RUN;
PROC UNIVARIATE DATA=C NORMAL;
VAR RESID TRESID;
RUN;
PROC TRANSREG DATA=C PLOTS=NONE;
MODEL BOXCOX(RESID
TRESID)=CLASS (TRAT);
RUN;
```



PROGRAMA SAS (5)

- Coloque no programa SAS uma saída ODS (Output Delivery System) em RTF.
- Vamos discutir o resultado das análises?



***OBRIGADO E
ATÉ A
PRÓXIMA !!!***