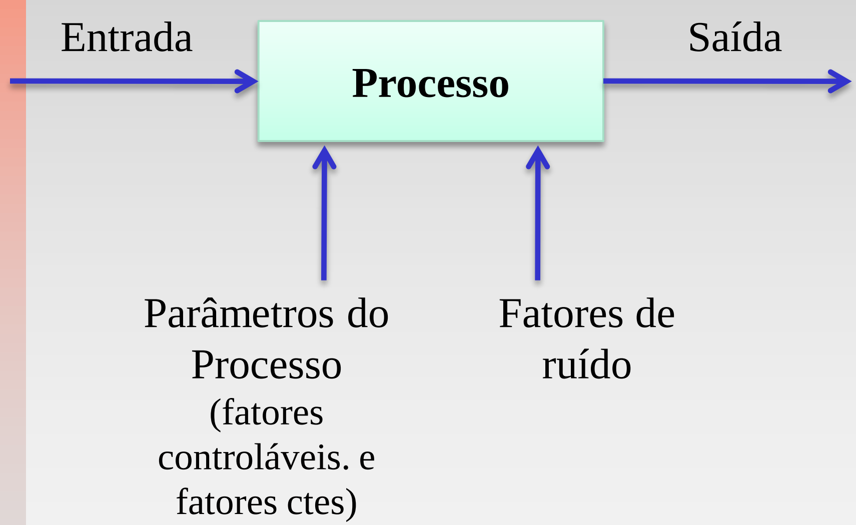
Projeto de Experimentos

– **O uso de Projeto de Experimentos conduz a uma seqüência estruturada de ensaios, que assegura o máximo de informação com um gasto mínimo de tempo/dinheiro.**



Testes onde são feitos mudanças propositais nos fatores controláveis

LOPP

Análise de Variância - ANOVA

– A análise de variância é a metodologia estatística que avalia a significância dos diversos fatores e interações.

– Há suposições básicas para validar a análise de variância:

- Distribuição normal dos dados.
- Homogeneidade das variâncias (em cada grupo) - aleatoriedade dos erros.
- Aditividade dos efeitos.
- Independência estatística dos valores observados (sem correlação).

LOPP

Análise de Variância

– Se as suposições de normalidade e homogeneidade não forem satisfeitas, o resultado da análise de variância deixa de ser exato, e passa a ser aproximado.

– Em raras situações a suposição de aditividade dos efeitos não é satisfeita. Nesse caso, uma transformação dos dados (log, √, etc.) pode recuperar a aditividade e permitir uma análise mais precisa.

– A independência estatística dos valores observados é obtida com o uso da aleatorização.

LOPP

One Way ANOVA

Experimentos que envolvem:

- **1 Variável de Resposta**
- **1 Fator Controlável a vários níveis**

Objetivo:

Identificar se os valores da variável de resposta medidos nos diversos níveis diferem entre si devido ao efeito do fator controlável.

One-Way ANOVA

2 tipos de experimentos:

–Fatores Controláveis a níveis fixos

É possível repetir o ensaio tempos depois, basta utilizar os níveis dos FC escolhidos

(Por ex., 5 valores de temperatura)

–Fatores Controláveis a níveis aleatórios

(Por ex., 3 lotes escolhidos ao acaso)

Nunca mais será possível ter os mesmos fatores controláveis

LOPP

A ANÁLISE DE VARIÂNCIA (ANOVA)

Formulação matemática do problema:

Modelo Estatístico:

onde: μ **é a média geral;**

τ_j **é o efeito do grupo j;**

ε_{ij} **é um erro aleatório.**

Hipóteses:

Ho: não há diferenças significativas entre os grupos;

H1: há diferenças significativas entre os grupos.

One-Way ANOVA

Disposição dos dados:

Fator A		A1	A2	...	Ak	
		y11	y12	...	y1k	
		y21	y22	...	y2k	
		:	:	yij	:	
		:	:	...	:	
		yn1	yn2	...	ynk	
Totais	T.j	T.1	T.2	...	T.k	T.. =
No.Obs.	nj	n1	n2	...	nk	N =
Médias	$\bar{y}_{.j}$	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.k}$	$\bar{\bar{y}}_{..}$ =

One-Way ANOVA

Exemplo: Um profissional deseja estudar se a temperatura ambiente influencia na produtividade dos funcionários. Para isso realizou três medidas de produtividade (peças/hora) em três temperaturas diferentes.

Temperatura		
15°C	25°C	35°C
12	20	17
13	19	16
11	18	18

Fator controlável

Níveis do fator controlável

Repetições

Medição da variável de resposta

LOPP

Exemplo: One-Way ANOVA

	Temperatura			
	15°C	25°C	35°C	
	12	20	17	
	13	19	16	
	11	18	18	
$T_{.j} =$	36	57	51	$T_{..} = 144$
$n_j =$	3	3	3	$N = 9$
$\bar{Y}_{.j} =$	12	19	17	$\bar{\bar{Y}}_{..} = 16$

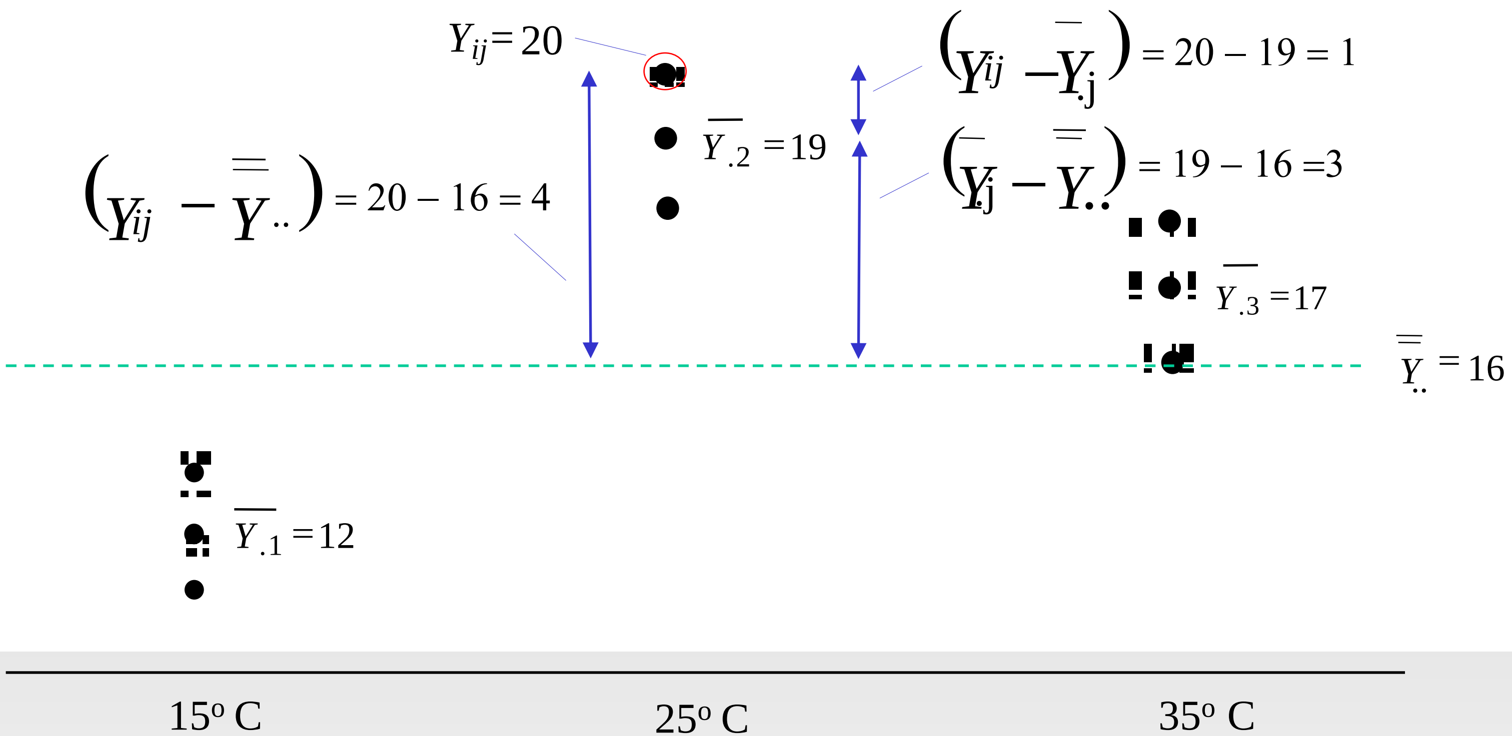
SQR dentro do grupo
Variabilidade residual (erro)

SQG= Variabilidade entre grupos (fator controlável)

SQT = SQG grupo + SQR resíduo

LOPP

Decomposição da variabilidade



$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

20 = 16 + 3 + 1

SQT = SQG + SQR

$$\left(Y_{ij} - \bar{\bar{Y}}_{..}\right) = \left(\bar{Y}_{.j} - \bar{\bar{Y}}_{..}\right) + \left(Y_{ij} - \bar{Y}_{.j}\right)$$

4 = 3 + 1

Decomposição dos resíduos:

$$(Y_{ij} - \overline{\overline{Y}}_{..}) = (\overline{Y}_{.j} - \overline{\overline{Y}}_{..}) + (Y_{ij} - \overline{Y}_{.j})$$

Elevando ao quadrado e somando:

$$\Sigma (Y_{ij} - \overline{\overline{Y}}_{..})^2 = \Sigma\, n\, (\overline{Y}_{.j} - \overline{\overline{Y}}_{..})^2 + \Sigma (Y_{ij} - \overline{Y}_{.j})^2$$

SQT = SQG + SQR

Graus de Liberdade:

(N - 1) = (K - 1) + (N - K)

Médias quadradas ou variâncias:

MQG (entre grupos-fator) = SQG / (K - 1)

MQR (dentro do grupo-erro) = SQR / (N - K)

LOPP

O Teste F compara as duas variâncias:

$$F_{calc} = \frac{Variância \;\; entre\ grupos}{Variância \;\; dentro\ do\ grupo} = \frac{MQG(fator)}{MQR(erro)}$$

Se o Fcalc = 1, conclui-se que não há diferenças significativas entre os grupos pois
E [MQG] = E [MQR]

logo não rejeita-se H_o pois não há evidências suficientes para comprovar que as médias diferem entre si, logo não tem evidências suficientes para afirmar que a temperatura tenha efeito significativo

$$F_{calc} = \frac{Variância \ \ entre\ grupos}{Variância \ \ dentro\ do\ grupo} = \frac{MQG(\,fator\,)}{MQR(erro)}$$

Se o $$F_{calc} > F_{tab} = F_{\alpha, k-1, N-k}$$

ou valor- p < 0,05 (α = 5%)

Sendo **α** o nível de significância (geralmente 5%), que é um valor aceitável de se cometer o erro do tipo I que é rejeitar H_0 sendo que a hipótese é verdadeira, ou seja, errar na conclusão. O intervalo de confiança na decisão é (1- **α)**

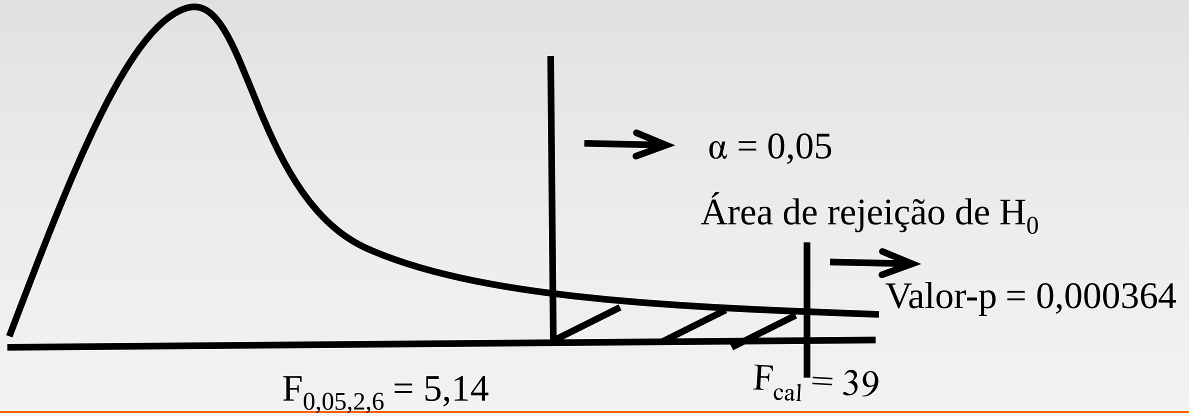
Rejeita-se H_o e conclui-se que existem diferenças significativas entre os grupos (níveis) provocadas pelo fator controlável, a um nível de significância de 5%.

O valor de F_{tab} é estabelecido usando os valores tabelados da distribuição de Fischer:

$$F_{tab} = F_{\alpha, GL_{numerador}, GL_{denomnador}} = F_{\alpha, K-1, N-K}$$

α : nível de significância (usualmente 0,05 ou 5%)
k-1: graus de liberdade do numerador (MQG):
N-k : graus de liberdade do denominador (MQR)

Para o exemplo, $F_{0,05,2,6}$=5,14



LOPP

One-Way ANOVA

Exemplo no Excel:

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
15°C	3	36	12	1
25°C	3	57	19	1
35°C	3	51	17	1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	78	2	39	39	0,000364	5,143253
Within Groups	6	6	1			

Como Fcalc > Ftab (valor-p=0,000364<0,05) existe efeito significativo da temperatura sobre a produtividade a um nível de significância de 5%

Fórmulas para os cálculos (simplificadas)

$TC = T_{..}^2 / N$

TC = termo de correção

$SQT = \sum(Y_{ij}^2) - TC$

$SQG = \sum(T_{.j}^2 / n_j) - TC$

$SQR = SQT - SQG$

Tabela ANOVA:

Fonte	SQ	GDL	MQ	Teste F
Entre Grupos	SQG	K - 1	MQG	F = MQG / MQR
Dentro Grupos	SQR	N - K	MQR	
Total	SQT	N - 1		

Exemplo a níveis fixos:

Um pesquisador deseja investigar o efeito da temperatura do forno sobre o número de bactérias contadas após o processo de esterelização. Os dados revelaram o seguinte:

Hipóteses:

Ho: não há diferenças significativas entre os grupos, ou seja, não há efeito da temperatura do forno;

H1: há diferenças significativas entre os grupos provocada pela temperatura do forno.

Outro exemplo

Tempera-tura	70	80	90	100	110	
	15,0	13,1	12,4	10,4	13,1	
	15,9	14,1	11,2	13,4	10,0	
	18,4	18,2	15,9	11,5	13,9	
	17,2	11,1	13,4	14,2	11,1	
	18,6	15,5	9,00	12,7	13,6	
	18,7	12,2	10,3	13,8	12,4	
	16,0	12,3	10,0	12,6	11,2	
	17,1	13,0	13,2	11,4	12,3	
	21,5	15,5	11,0	16,1	13,4	
	14,2	14,3	13,8	13,7	15,9	
	18,4	15,9	12,4	9,20	9,10	K=5
	15,1	15,6	13,4	10,6	10,2	n=12
Totais	206,10	170,80	146,00	149,60	146,20	T..=818,7
No.Obs.	12	12	12	12	12	N = 60
Médias	17,18	14,23	12,17	12,47	12,18	$\bar{\bar{Y}}_{..} = 13,65$

Cálculos iniciais:

$$TC = T_{..}^2 / N = (818,7)^2 / 60 = 11.171,1$$

$$SQT = \Sigma (Y_{ij}^2) - TC = 11.608,2 - 11.171,1 = 437,1$$

$$SQG = \Sigma (T_{.j}^2 / n_j) - TC$$
$$= [(206,1)^2 / 12] + \dots + [(146,2)^2 / 12] - 11.171,1 = 222,3$$

$$SQR = SQT - SQG = 437,1 - 222,3 = 214,8$$

Tabela Anova:

Fonte	SQ	GDL	MQ	Fcalc	Ftab
Entre Grupos –MQG (Temperatura)	222,3	4	55,6	14,2	2,54
Dentro Grupos-MQR (Residual ou erro)	214,8	55	3,9		
Total	437,1	59			

F calculado	>	F tabelado
14,2	>	2,54

Há diferenças significativas entre os grupos provocada pelo fator controlável temp. do forno ou existe efeito da temp. do forno sobre o número de bactérias a um nível de significância de 5% (Fcal=14,2>F0,05=2,54)

LOPP

Otimização: Comparação múltipla de médias-CMM

Se na Tabela ANOVA, a hipótese nula for rejeitada

($F_{calc} > F_{tab}$ ou valor-p$<0,05$), é necessário realizar

Uma comparação múltipla de médias (CMM) para os

fatores que possuam mais do que dois níveis a fim de identificar quais deles diferem entre si

São diversos métodos de CMM: Duncan,Tukey, Scheffé,

Bartlett.

Otimização: Comparação múltipla de médias-CMM

1. **Calcular o desvio padrão das médias**

O valor esperado da MQR é igual a variância: E(MQR) = σ^2

Pelo teorema do limite central

$$S_{\bar{y}} = \frac{\sigma}{\sqrt{n_c}} = \frac{\sqrt{MQR}}{\sqrt{n_c}} = \frac{\sqrt{3,9}}{\sqrt{12}} = \frac{1,97}{3,46} = 0,57$$

onde $n_c = (n_1 + n_2 + ... + n_k) / k$

2. Calcular o limite de decisão

$$L_d = 3 \times S_{\bar{Y}} = 3 \times 0,57 = 1,71$$

3. Escrever as médias em ordem crescente ou decrescente e compará-las duas a duas. A diferença será significativa se for maior que o L_d

LOPP

$\bar{Y}_{70} = 17{,}18 \quad \bar{Y}_{80} = 14{,}23 \quad \bar{Y}_{90} = 12{,}17 \quad \bar{Y}_{100} = 12{,}47 \quad \bar{Y}_{110} = 12{,}18$

$\bar{Y}_{70}(17{,}18) - \bar{Y}_{80}(14{,}23) = 2{,}95 > Ld = 1{,}71 \; Dif \; Signif.$

$\bar{Y}_{80}(14{,}23) - \bar{Y}_{90}(12{,}17) = 2{,}06 > Ld = 1{,}71 \; Dif \; Signif.$

$\bar{Y}_{90}(12{,}17) - \bar{Y}_{100}(12{,}47) = 0{,}3 < Ld = 1{,}71 \; Dif \; Não \; Signif.$

$\bar{Y}_{100}(12{,}47) - \bar{Y}_{110}(12{,}18) = 0{,}29 < Ld = 1{,}71 \; Dif \; Não \; Signif.$

4. Usar barras contínuas sobre as médias que não diferem entre si

$\bar{Y}_{70} = 17{,}18 \quad \bar{Y}_{80} = 14{,}23 \quad \bar{Y}_{90} = 12{,}17 \quad \bar{Y}_{100} = 12{,}47 \quad \bar{Y}_{110} = 12{,}18$

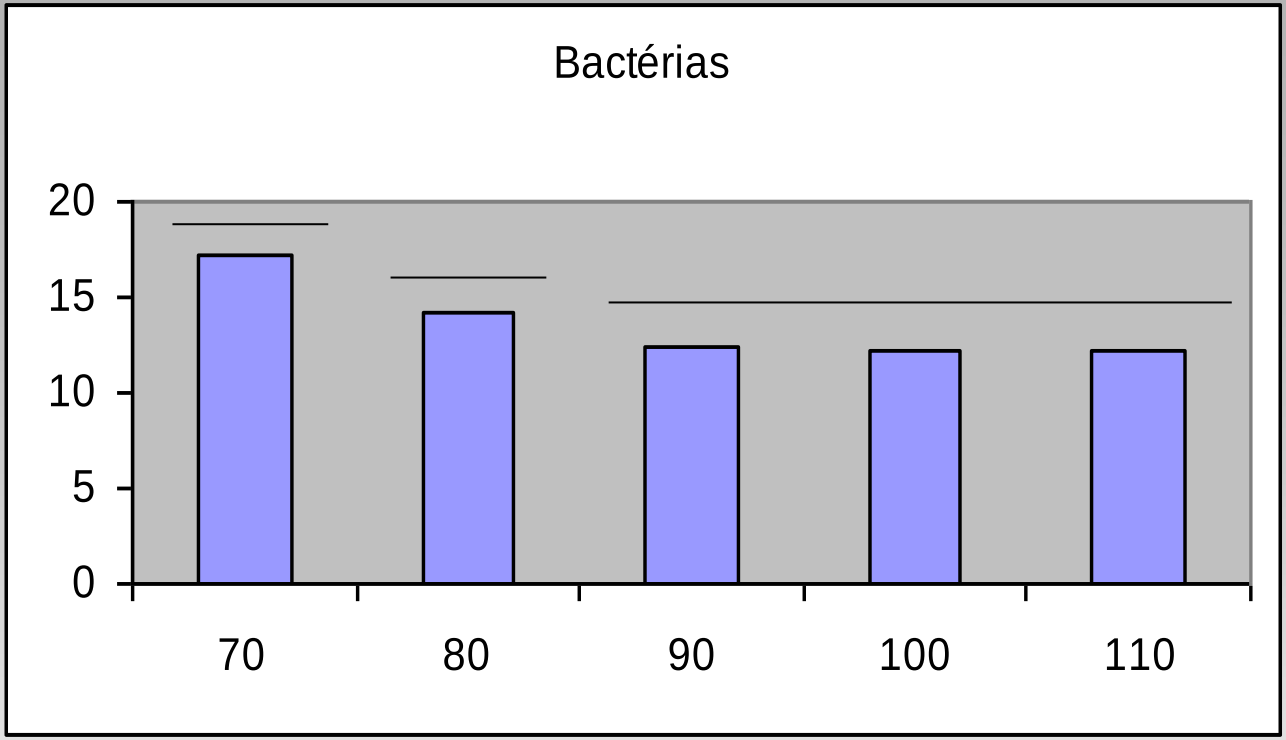
Otimização

– Para realizar a otimização do produto ou processo é feita uma análise técnica.

– A análise técnica deve acompanhar e completar a análise estatística.

– Para isso é recomendável representar graficamente os dados.

– Para os dados do experimento anterior, poderia se usar, por exemplo, um boxplot ou um gráfico de barras.



- **O ajuste ótimo considerando qualidade (bactérias) é temperatura 90, 100 ou 110**
- **O ajuste ótimo considerando qualidade (bactérias) e custo é temperatura 90 (mais barato)**

Os dados a seguir representam o alongamento (maior é melhor) medido sobre um composto de borracha, em função da quant. de agente adicionado durante a mistura (LIE = 52).

Agente	0	5	10	15	20	
	43	47	55	50	52	
	47	53	50	54	49	
	46	52	54	54	54	
	45	50	55	55	55	
	45	49	52	56	55	
	46	51	53	52	56	
	47	55	55	57	56	
	44	48	56	57	53	
	42	49	59	55	57	
	48	50	56	60	60	
	49	47	57	56	57	
	44	49	54	58	55	
Totais	546	600	656	664	659	**T..= 3125**
No.Obs.	12	12	12	12	12	**N = 60**
Médias	45,5	50,0	54,7	55,3	54,9	**Y.. = 52,08**

LOPP

Exemplo

Cálculos iniciais:

$TC = T_{..}^2 / N = (3125)^2 / 60 = 162.760{,}42$

$SQT = \Sigma (Y_{ij})^2 - TC = 163.971{,}00 - 162.760{,}42 = 1210{,}58$

$SQG = \Sigma (T_{.j}^2 / nj) - TC = [(546)^2 / 12] + \ldots + [(659)^2 / 12] - 162.760{,}42 = 875{,}33$

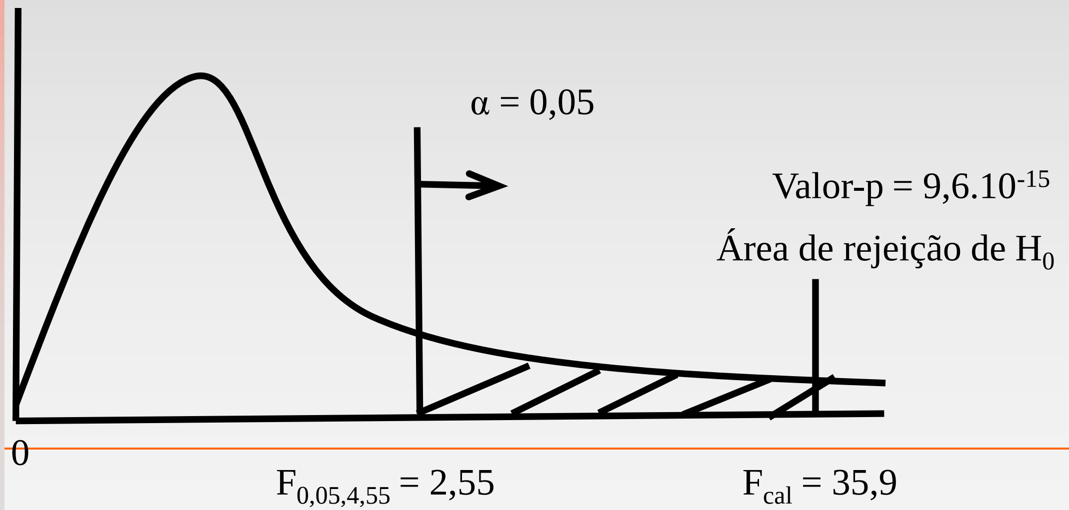
$SQR = SQT - SQG = 1210{,}58 - 875{,}33 = 335{,}25$

LOPP

Exemplo

Tabela ANOVA:

Fonte	SQ	GDL	MQ	Teste F
Entre Grupos (Agente de processo)	875,33	4	218,83	35,9
Dentro Grupos (Residual)	335,25	55	6,09	
Total	1210,58	59		



Há diferenças significativas entre os grupos

$F_{cal} > F_{tab}$

Valor-p < α

Comparação múltipla de médias (CMM)

– **Existem diversos métodos de CMM: Duncan, Tukey, Scheffé, Bartlett, etc**

– **O método apresentado a seguir é o de Duncan**

1. Calcular o desvio padrão das médias pelo Teorema do Limite Central

$$S_{\overline{Y}_{.j}} = \frac{\sqrt{MQR}}{\sqrt{12}} = \sqrt{\frac{6{,}09}{12}} = 0{,}71$$

onde $n_c = (n_1 + n_2 + \dots + n_k) / k$

2. Calcular o limite de decisão

$$L_d = 3 \times S_{\overline{Y}_{.j}} = 3 \times 0{,}71 = 2{,}13$$

LOPP

Comparação múltipla de médias (CMM)

3. **Escrever as médias em ordem crescente ou decrescente e compará-las duas a duas (todas combinações). A diferença será significativa se for maior que o Ld (=2,13)**

45,5	**50,0**	**54,7**	**54,9**	**55,3**
Y.1	**Y.2**	**Y.3**	**Y.5**	**Y.4**

50,0 – 45,5 = 4,5 > Ld=2,13 DS

54,7 – 50,0 = 4,7 > Ld=2,13 DS

54,9 – 54,7 = 0,2 < Ld=2,13 DNS

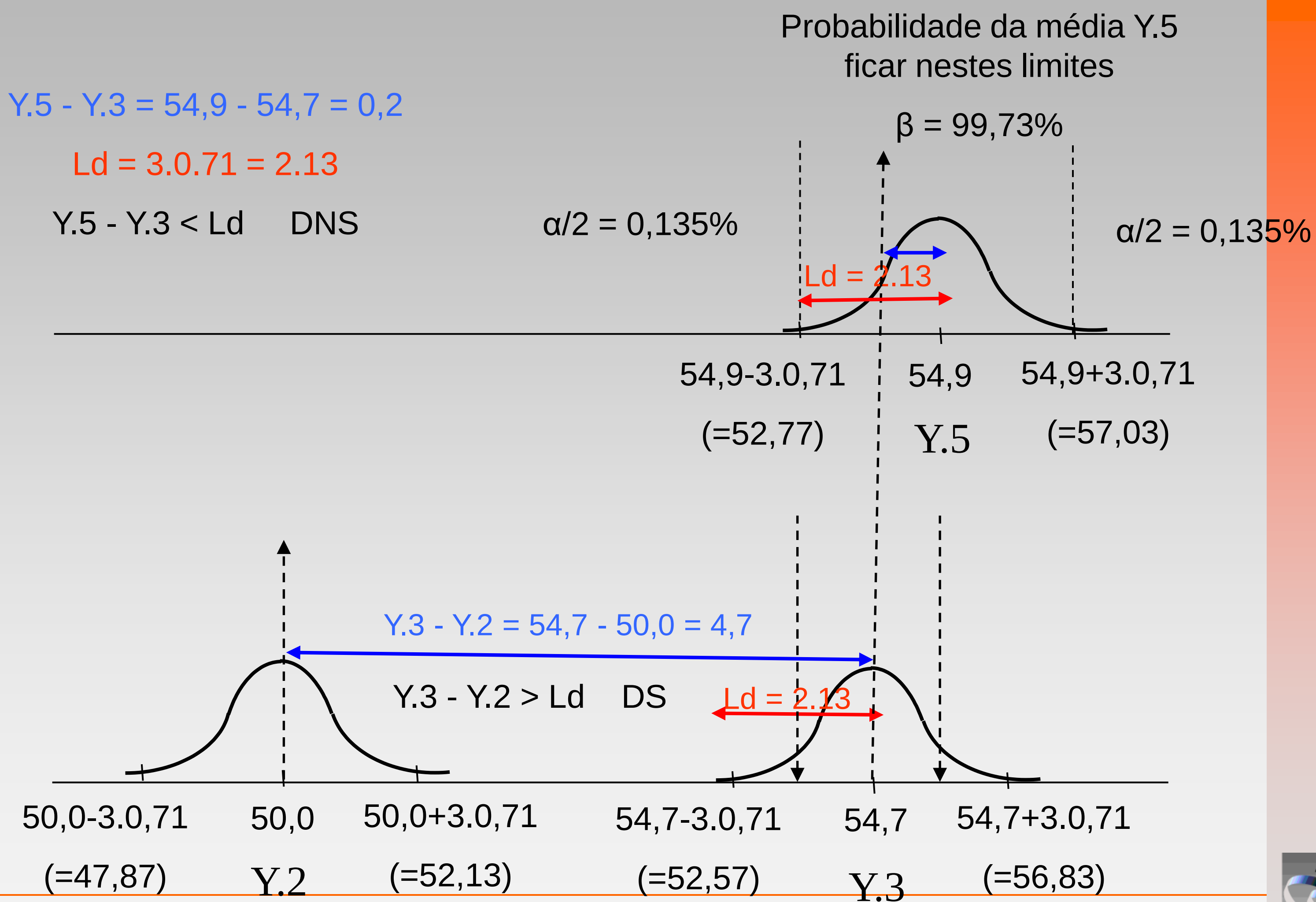
55,3 – 54,9 = 0,4 < Ld=2,13 DNS

4. **Usar barras contínuas sobre as médias que não diferem entre si**

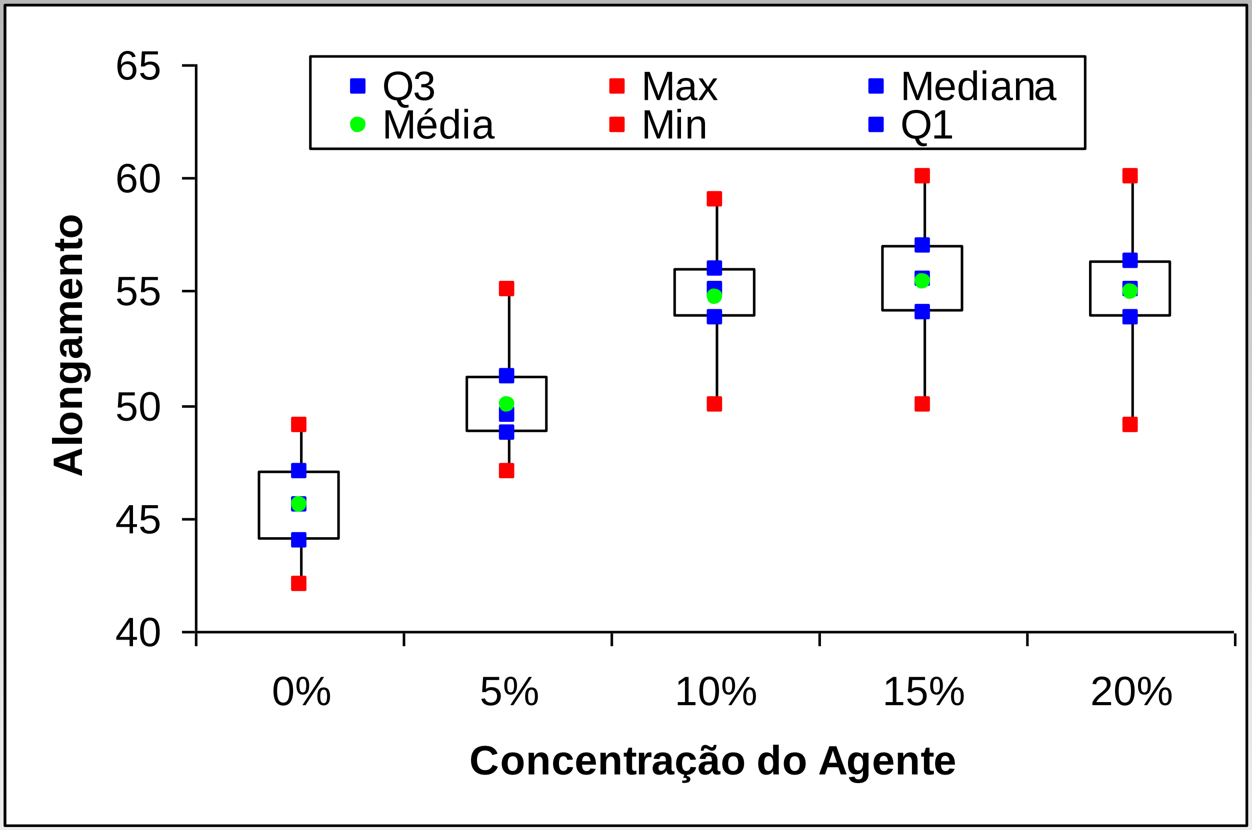
$\overline{\text{Y.1}}$ $\quad$ $\overline{\text{Y.2}}$ $\quad$ $\overline{\text{Y.3 Y.5 Y.4}}$

LOPP

Comparação múltipla de médias (CMM)

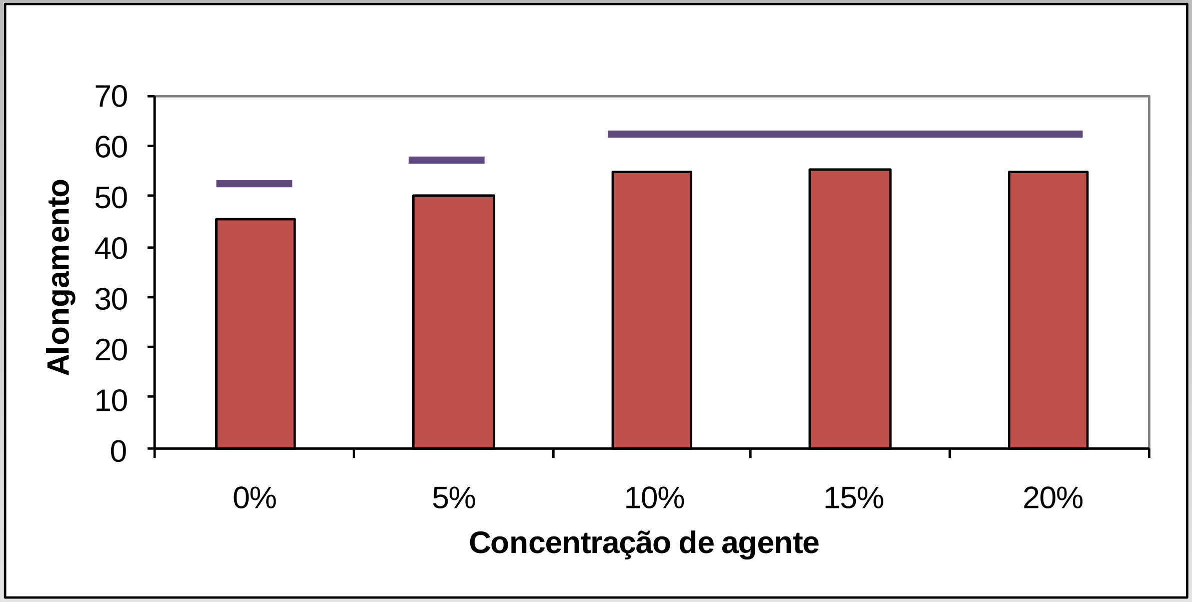


Otimização observando o boxplot



Como o LIE do cliente é 52, então a escolha da concentração de agente pode ser 10, 15 e 20%, pois atendem a especificação.

Otimização observando gráfico de barras



Como a CQ alongamento é do tipo maior-é-melhor, então a escolha da concentração de agente pode ser 10, 15 e 20%, pois produzem a mesma resposta média e ainda superior que 5%.

Contudo, quanto maior a concentração do agente, mas caro é a borracha, então a escolha recomendada é a concentração de 10%.

LOPP

Exemplo (níveis aleatórios):

Um pesquisador deseja investigar se a permeabilidade das lentes de uso flexível fabricadas em sua indústria permanece uniforme ou não. Escolhe-se aleatoriamente três lotes de produção e realizam-se ensaios:

Lote	L1	L2	L3	
	61	60	60	
	62	61	63	
	64	58	59	
	62	58	64	
	63	60	62	
	63	59		
		60		
Totais $T_{.j}$	375,0	416,0	308	$T_{..} = 1099$
Num. Obs. n_j	6	7	5	$N = 18$
Médias $x_{.j}$	62,50	59,43	61,60	$\overline{\overline{Y}}_{..} = 61,06$

LOPP

Cálculos iniciais:

$$TC = T..^2 / N = (1099)^2 / 18 = 67.100,06$$

$$SQT = (x_{ij}^2) - TC = 67.163,00 - 67.100,06 = 62,94$$

$$SQG = (T._j^2 / n_j) - TC$$

$$= [(375)^2 / 6] + ... + [(308)^2 / 5] - 67.100,06 = 32,53$$

$$SQR = SQT - SQG = 62,94 - 32,53 = 30,41$$

Tabela Anova:

Fonte	SQ	GDL	MQ	Teste F
Entre Grupos (Lotes)	32,53	2	16,26	8,02
Dentro Grupos (Residual)	30,41	15	2,03	
Total	62,94	17		

F calculado	>	F tabelado
8,02	>	3,68

➔ **Há diferenças significativas entre os grupos, provocada pelo fator controlável em estudo**

LOPP

Estimar componentes de variação

$$E\,[MQR] = \sigma^2$$
$$E\,[MQG] = \sigma^2 + n_c\, \sigma_\alpha^2$$

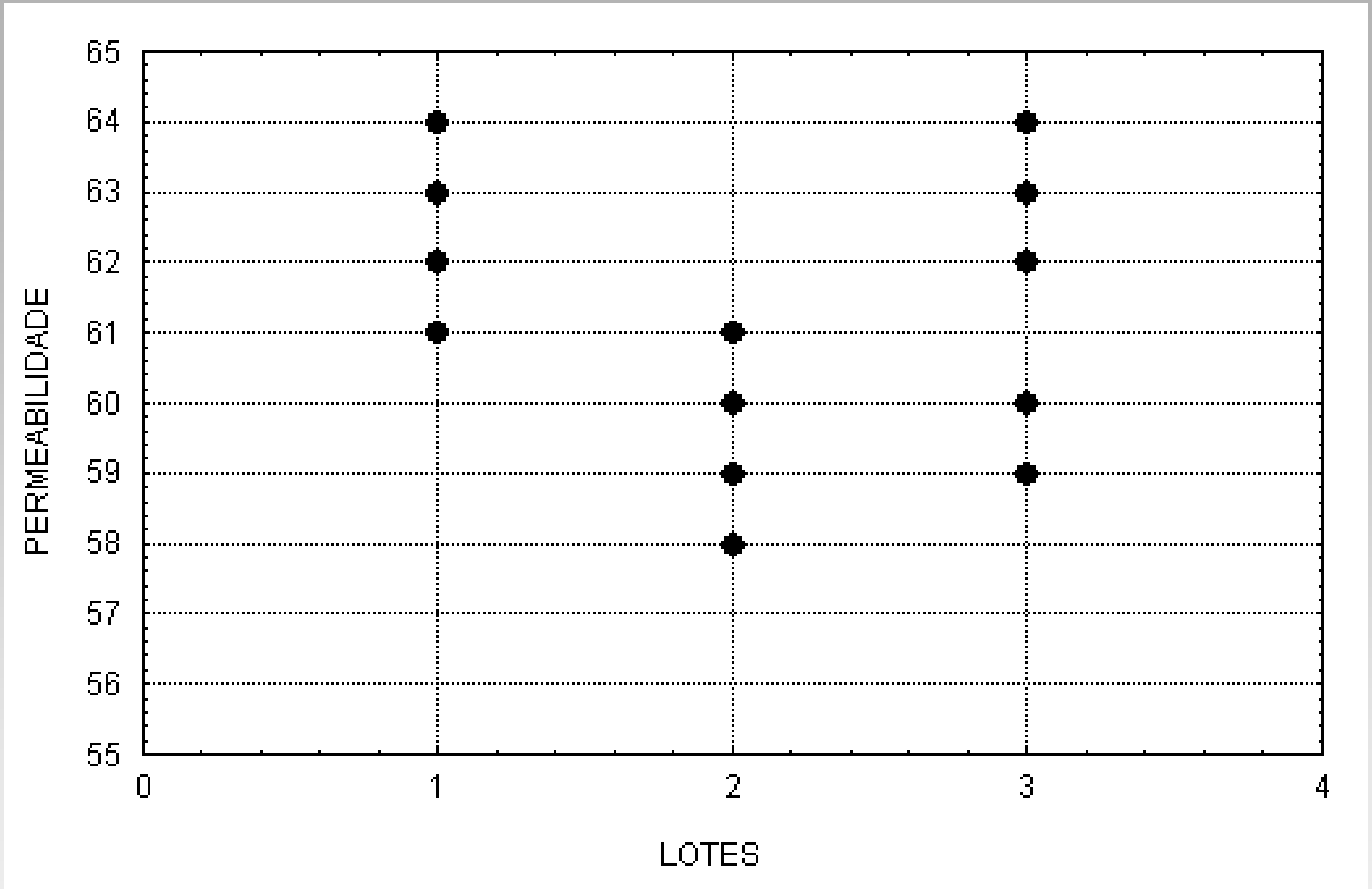
Assim as estimativas são:

$$\sigma^2 = MQR = 2{,}03$$

$$\sigma_\alpha^2 = (MQG - MQR) / n_c = (16{,}26-2{,}03)/6=2{,}37$$

$$\sigma_T^2 = \sigma^2 + \sigma_\alpha^2 = 4{,}40$$

De forma que 2,37 / 4,40 = 54% da variabilidade total observada nos valores de permeabilidade das lentes deve-se a diferenças "entre lotes".



Scatter plot para o exemplo da permeabilidade

LOPP