UNIVERSIDADE DE SÃO PAULO
ESCOLA DE EDUCAÇÃO FÍSICA E ESPORTE DE RIBEIRÃO PRETO
EEFERP - USP

Critérios de autenticidade científica: análise e escolha do método

Prof. Dr. Dalmo Roberto Lopes Machado

Regras básicas na seleção dos testes e medidas

População avaliada

- Para quem o instrumento (testes) foi desenvolvido?

Objetivos da avaliação

- Qual o objetivo do instrumento? O que se espera da avaliação?

Tamanho do grupo de avaliados

- Instrumento: teste, escala, lista, inventário... em massa ou individual? Tem equipe?

Treinamento do administrador

- Requer especialização? E.C.G., psicometria, nutricional?

Tempo de administração

- Com instrução, aquecimento? Quantos dias?

Ambiente adequado para a administração

- Laboratório, campo? Área aberta, privada? Barulho, temperatura, umidade...

Custos operacionais

- Despesas: equipamentos, transportes, incentivos, infra-estrutura...

(Tritschler, 2003)

Qualidade das medidas Psicométricas

Psicometria (do grego *psyké*, alma e *metron*, medida, medição) é uma área da Psicologia que faz a ponte entre as ciências exatas, principalmente a matemática aplicada e a Estatística.

Sua definição consiste no conjunto de técnicas utilizadas para mensurar, de forma adequada e comprovada experimentalmente, um conjunto ou uma gama de comportamentos que se deseja conhecer melhor.

Fatores determinantes da eficiência em medidas

- permite replicação de um modelo em outra situação

Validade – capacidade de medir o que ... opõe medir

Confiabilidade – funcionamento em circunstâncias de rotina

- como o instrumento repr... ...casos (avaliadores)

Variabilidade – amplitude da dispersão dos escores nu...

- medidas ...

Reprodutibilidade

- confiança no instrumento de medida – probabilidade de não falhar na medição

Exatidão – expressão a medida ve...

- com...

Precisão – consistên...

- acertar o alvo, A medida obtida ... ...dadeira do atributo em questão,

Fidedignidade – variação ...

Objetividade – externo à consciência. Observação imparcial, sem preferências

...dadeira – quanto se acredita na medida

...(conhecido)

...resultados semelhantes

...ática

...ssão de ...

...ida – manter o padrão

...a – real

...ia entre os valores medido e real

...Grau de consistência nos resultados de um teste quando aplicado por dif...

Qualidade psicométrica

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education: Standards for Educational and Psychological Testing (1985).

Validade – capacidade do instrumento medir o que se propõe medir. Pode ser de conteúdo, relacionada com um critério ou de constructo

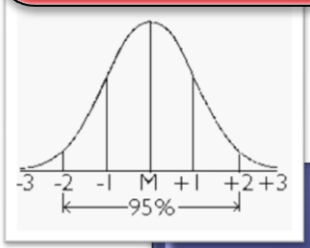
Fidedignidade – característica de um teste ou instrumento que apresenta resultados consistentes. (Reprodutibilidade, Confiabilidade)

Objetividade – consistência da medida quando realizada por diferentes testadores

Liberdade de Tendência em Avaliação – experiência diferente da população para quem o teste foi desenvolvido – etnia, sexo, idade, capacidade, habilidade – explicar as eventuais diferenças

Fatores determinantes da eficiência em medidas e avaliação

Validade – capacidade do instrumento medir o que se propõe medir. Pode ser de conteúdo, relacionada com um critério ou de constructo



Referenciada por NORMA

- Validade de Conteúdo
- Validade de Critério
  - Concorrente
  - Preditiva
- Validade de Constructo

Referenciada por CRITÉRIO

- Validade Referenciada ao Domínio
- Validade Referenciada por Decisão

Fatores determinantes da eficiência em medidas e avaliação

NORMA

Validade

De Conteúdo (lógica)
Grau de representatividade de um universo ou domínio. Subjetividade esperada da avaliação.

Ex. Questões de conteúdos não ensinados (não há); Um exame físico abrangente sem um teste cardior-respiratório (não há)

De Critério
Os resultados estão verdadeira e sistematicamente relacionados a um ou mais critérios do avaliado

Concorrente (presente)
Parte de uma referência. Mede a intensidade da relação entre instrumento-referência
Ex. Correlação entre 2 instrumentos (ref.-válido)

Preditiva (futuro)
Correlação de desempenho (resultado) estimado. Envolve a variável *Preditora* e a variável *Predita*.

De Constructo (de conceito)
O atributo de interesse não pode ser diretamente medido - o teste reflete quanto do atributo tem o avaliado. Validade lógica ou estatística

Ex. Golfistas com
Baixas deficiências (0-12 erros)
Altas deficiências (13-18 erros)
Diferença entre as médias dos erros

Fatores determinantes da eficiência em medidas e avaliação

NORMA

Validade

De Conteúdo (lógica)

Grau de representatividade de um universo ou domínio. Subjetividade esperada da avaliação.

Ex. Questões de conteúdos não ensinados (não há); Um exame físico abrangente sem um teste cardior-respiratório (não há)

De Critério

Os resultados estão verdadeira e sistematicamente relacionados a um ou mais critérios do avaliado

Concorrente (presente)

Parte de uma referência. Mede a intensidade da relação entre instrumento-referência

Ex. Correlação entre 2 instrumentos (ref.-válido)

Preditiva (futuro)

Correlação de desempenho (resultado) estimado. Envolve a variável *Preditora* e a variável *Predita*.

De Constructo (de conceito)

O atributo de interesse não pode ser diretamente medido - o teste reflete quanto do atributo tem o avaliado. Validade lógica ou estatística

Ex. Golfistas com
Baixas deficiências (0-12 erros)
Altas deficiências (13-18 erros)
Diferença entre as médias dos erros

Fatores determinantes da eficiência em medidas e avaliação

CRITÉRIO

Validade

Referenciada ao Domínio

Os itens do instrumento devem representar bem o comportamento critério – "Domínio"

Validade Lógica

Ex. Teste de arbitragem de Voleibol;
1. Identificar os domínios;
2. Classificar as regras (categorias);
3. Retirar exemplos sistemáticos a partir do domínio – de forma bem representada.

(Parece com validade de conteúdo?)

Referenciada por Decisão

Classificar adequadamente grupos contrastantes. Com base nos seus dados estarem acima ou abaixo de um valor limite – *Coeficiente de Contingência ="C"*

Validade Estatística

Ex. classificados corretamente:

$$C = \frac{(\text{proficientes} + \text{não-profecientes})}{\text{Total dos 2 grupos}}$$

Valores acima de 0,80 são ideais

(Krippendorff, 1980)

Fatores determinantes da eficiência em medidas e avaliação

CRITÉRIO

Ex. Um grupo de praticantes de arco-e-flecha. Alguns fizeram 8 aulas (instruídos).

Total (n=78)	Instruídos (n=47)	Não-instruídos (n=31)
Proficientes	42	5
Não-proficientes	5	26

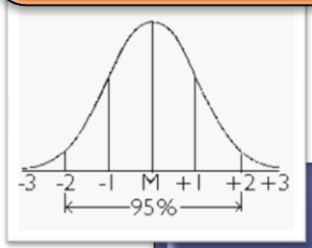
$$C = \frac{(42 + 26)}{(47 + 31)} = \frac{68}{78} = 0{,}87$$

C = Coeficiente de contingência

Valores acima de 0,80 são ideais
(Krippendorff, 1980)

Fatores determinantes da eficiência em medidas e avaliação

Fidedignidade – característica de um teste ou instrumento que apresenta resultados consistentes. (Reprodutibilidade, Confiabilidade)



Referenciada por NORMA

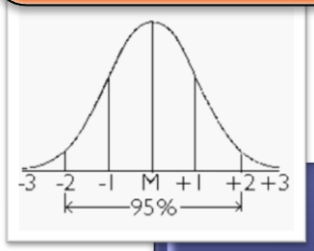
- Teste-reteste
- Fidedignidade de Consistência Interna
  - Fidedignidade dividida ao meio
- Fidedignidade de Formas Paralelas

Referenciada por CRITÉRIO

- Fidedignidade de Proficiência Repetida

Fatores determinantes da eficiência em medidas e avaliação

Fidedignidade – característica de um teste ou instrumento que apresenta resultados consistentes. (Reprodutibilidade, Confiabilidade)



Referenciada por NORMA

- Teste-reteste
- Fidedignidade de Consistência Interna
  - Fidedignidade dividida ao meio
- Fidedignidade de Formas Paralelas

Referenciada por CRITÉRIO

- Fidedignidade de Proficiência Repetida

Fatores determinantes da eficiência em medidas e avaliação

Teste-reteste – estabilidade/consistência com o passar do tempo

Mesmo sujeito – avaliado 2 vezes

Salto sêxtuplo em 5 jogadores, realizados em 2 momentos: X e Y

	X	Y	XY	X^2	Y^2
1	12,3	12,8	157,4	151,3	163,8
2	13,7	14,5	198,7	187,7	210,3
3	13,2	13,7	180,8	174,2	187,7
4	14,0	14,5	203,0	196,0	210,3
5	13,6	13,2	179,5	185,0	174,2
$\sum$	66,8	68,7	919,5	894,2	946,3

$$r_{xy} = \frac{n \,.\, \Sigma xy - (\Sigma x)\,(\Sigma y)}{\sqrt{(n \,.\, \Sigma x^2 - (\Sigma x)^2)\,(n \,.\, \Sigma y^2 - (\Sigma y)^2)}}$$

$$r = \frac{5 \,.\, 919{,}5 - (66{,}8).(68{,}7)}{\sqrt{(5 \,.\, 894{,}2 - (66{,}8)^2)\,.\,(5 \,.\, 946{,}3 - (68{,}7)^2)}}$$

$$r = 0{,}81$$

Fatores determinantes da eficiência em medidas e avaliação

Consistência interna –a fidedignidade pode ser estimada em uma única administração. Equação de Spearman–Brown.

$$r\ do\ teste\ total = \frac{2(r\ das\ metades)}{1 + (r\ das\ metades)}$$

Fidedignidade = n (r) \ 1+(n-1)r

	X
1	12,3
2	13,7
3	13,2
4	14,0
5	13,6
6	12,5

	X	Ímp.	Par.
1	12,3	12,3	13,7
2	13,7	13,2	14
3	13,2	13,6	12,5
4	14		
5	13,6		
6	12,5		

$r = 0{,}60$

Primeiro calcular o r

$$= \frac{2(0{,}60)}{1 + (0{,}60)} = \frac{1{,}20}{1{,}60} = 0{,}75$$

Em seguida o r do teste total

$r = 0{,}75$

Fatores determinantes da eficiência em medidas e avaliação

De formas paralelas –é calculado o "Coeficiente de Equivalência" entre 2 momentos: Pré e Pós teste – (efeito de um tratamento)

Pearson

Fonte de variação	Soma dos quadrados (SQ)	Graus de liberdade	Quadrado médio (QM)
Entre sujeitos	$\sum \frac{(\sum x)^2}{n} - \frac{(\sum x_{total})^2}{n_{total}}$	$K-1$	$\frac{SQ_{entre\ sujeitos}}{K-1}$
Resíduo	$SQ_{total} - SQ_{entre\ sujeitos}$	$K(n-1)$	$\frac{SQ_{resíduo}}{K(n-1)}$
Total	$\sum x^2_{total} - \frac{(\sum x_{total})^2}{n_{total}}$	$N_{total}-1$	

Em que:

$\sum \frac{(\sum x)^2}{n}$ ∴ soma dos valores de cada estrato, elevado ao quadrado, dividido pelo número de elementos em cada estrato;

$\frac{(\sum x_{total})^2}{n_{total}}$ ∴ soma dos valores combinando todos os estratos envolvidos no delineamento, elevado ao quadrado, dividido pelo número de elementos também combinando todos os estratos;

$\sum x^2_{total}$ ∴ soma dos valores elevado ao quadrado, combinando todos os estratos envolvidos no delineamento;

K ∴ número de réplicas realizadas;

N ∴ número de avaliados envolvidos no delineamento; e

n_{total} ∴ número de elementos combinando todos os estratos.

Fatores determinantes da eficiência em medidas e avaliação

De formas paralelas –é calculado o "Coeficiente de Equivalência" entre 2 momentos: Pré e Pós teste – (efeito de um tratamento)

Dois equipamentos foram testados para verificar o salto vertical de uma equipe de voleibol...

X	Y	X.Y
74	76	5624
73	80	5840
81	79	6399
66	66	4356
62	64	3968
79	81	6399
78	85	6630
70	73	5110
80	82	6560
71	70	4970
74	81	5994
808	837	Σ= 61850

$\bar{X}_x = 73{,}45$ cm; $S_x = 5{,}69$

$\bar{X}_y = 76{,}09$ cm; $S_y = 6{,}61$

$$r = \frac{\frac{\Sigma X.Y}{n} - (X).(Y)}{(s_x).(s_y)}$$

$$r = \frac{5622{,}73 - (73{,}45).(76{,}09)}{(5{,}69).(6{,}61)}$$

$r = 0{,}91$

ESCALA DE CORRELAÇÃO			
0,00	~	0,19	⇒ FRACA
0,20	~	0,39	⇒ BAIXA
0,40	~	0,69	⇒ MODERADA
0,70	~	0,89	⇒ ALTA
0,90	~	1,00	⇒ MUITO ALTA

Fatores determinantes da eficiência em medidas e avaliação

Proficiência repetida – consistência da classificação.
Proficiência em 2 ocasiões diferentes, por um mesmo critério

Ex. 80 avaliados foram submetidos a 2 dias de um teste de proficiência.

CRITÉRIO

		Dia 2	
		Proficientes	Não-proficientes
Dia 1	Proficientes	34	9
	Não-proficientes	7	30

$$P = \frac{34 + 30}{34 + 30 + 7 + 9} = \frac{64}{80} = 0{,}80$$

P = Coeficiente de Proficiência

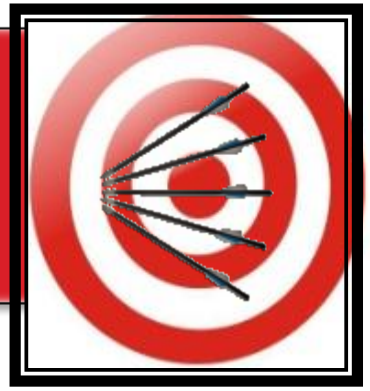
Fatores determinantes da eficiência em medidas e avaliação

Objetividade – consistência da medida quando realizada por diferentes testadores

Inter-resultados	Testes não-paramétricos
• Erro inter-avaliador • Correlação pareada (teste t pareado)	• Consistência da medida entre 2 avaliadores (coeficiente alfa de Cronbach)

Fatores determinantes da eficiência em medidas e avaliação

Precisão – consistência da medida – manter o padrão



- consistência da medida ocorrer com baixa variação

Exatidão – expressão a medida verdadeira – real

- Acertar o alvo. A medida obtida reflete a medida verdadeira do atributo em questão.

Fatores determinantes da eficiência em medidas e avaliação

Liberdade de Tendência em Avaliação – experiência diferente da população para quem o teste foi desenvolvido – etnia, sexo, idade, capacidade, habilidade – explicar as eventuais diferenças

Qualidade Psicométrica	Objetivo	Sub-tipos	Procedimento estatístico
Validade	Medir o mesmo atributo por diferentes testes/instrumentos	NORMA **Conteúdo**	Lógica
		Critério - Concorrente - Preditiva	Correlação e CCI Regressão
		Constructo	Lógica Índice Kappa
		CRITÉRIO **Proficiência repetida**	Concordância
Fidedignidade	Medidas (repetidas) similares em diferentes pessoas por um mesmo avaliador	NORMA **Teste-reteste** **Consistência interna** - Dividida ao meio **Formas paralelas** - pré-pós teste CRITÉRIO **Proficiência repetida**	Correlação Pearson Spearman-Brown Anova - two way - CCI Concordância
Objetividade	Medidas similares do mesmo avaliado por diferentes avaliadores	**Inter-resultados** **Testes subjetivos** (dança, ginástica...) escalas	Correlação pareada Concordância pelo mesmo critério
Liberdade de Tendência	Validade - bagagem étnica (hábitos, cultura, gênero) tendenciosidade	-	Lógica

Obrigado

Erro Técnico de Medida

$$ETM = \sqrt{[(\Sigma dif^2 / 2n]}$$

n=5	Medidor 1	Medidor 2	Dif(Med1-Med2)	Dif^2
Sujeito 1:	27 mm	29mm	-2	4
Sujeito 2:	13 mm	14 mm	-1	1
Sujeito 3:	17 mm	18 mm	1	1
Sujeito 4:	19 mm	20 mm	-1	1
Sujeito 5:	22 mm	22 mm	0	0

Exemplo para a medição da dobra tricipital em 5 sujeitos (o valor média foi de ~20 mm)

$\Sigma Dif^2 = 7$

$\Sigma Dif^2/2.n = 7/2.5 = 0,7$

$= \sqrt{0,7} = 0,84$ mm

Cálculo do %ETM = (ETM/Média)*100
= (0,84/20)* 100 = 4,2 %

	Iniciante	Experiente
Dc:	10%	7,5%
Outras:	2%	1,5%

N&O (2005)

Validade concorrente

- Mede a intensidade da relação entre instrumento-referência

11 meio-fundistas treinadas – Internacional (fem. entre 18-25 anos)

Teste cardiorrespiratório em esteira (Referência) (Bruce – máximo)

Teste de corrida na pista de 400 mts. (Cooper – corrida de 12 min.)

Determinar o VO_2máx. nos dois tipos de testes (Laboratório e Pista)

Determinar o coeficiente de correlação entre os testes (r)

Capacidade Aeróbia

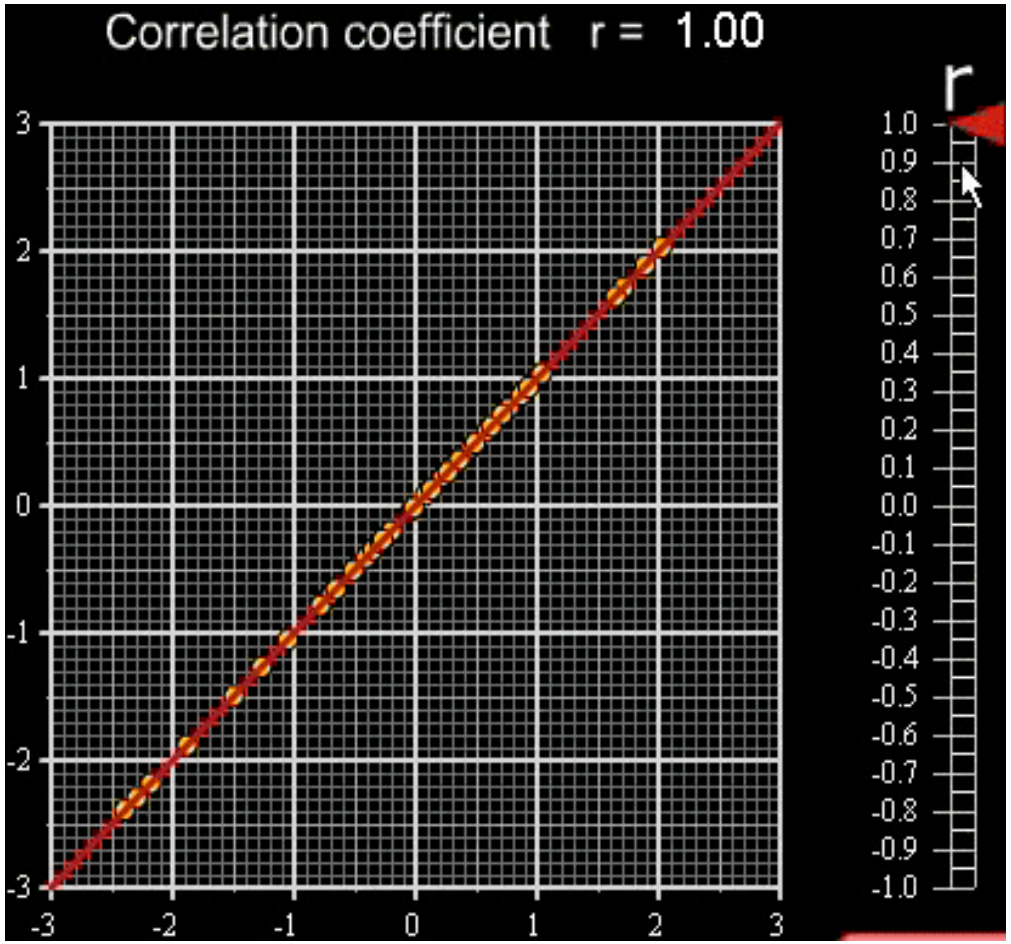
CORRELAÇÃO

Medir o grau de relação existente entre os dois testes

r=1.0: Esteira e Pista estão totalmente relacionados

r=0.0: Esteira e Pista não estão correlacionados

r=-1.0: Esteira e Pista estão inversamente relacionados



CORRELAÇÃO

x=Esteira y=Pista

	X	Y	X-73,45	Y-76,09	$(x-73,45)^2$	$(y-76,09)^2$
	74	76	0,55	-0,09	0,30	0,01
	73	80	-0,45	3,91	0,20	15,29
	81	79	7,55	2,91	57,00	8,47
	66	66	-7,45	-10,09	55,50	101,81
	62	64	-11,45	-12,09	131,10	146,17
	79	81	5,55	4,91	30,80	24,11
	78	85	4,55	8,91	20,70	79,39
	70	73	-3,45	-3,09	11,90	9,55
	80	82	6,55	5,91	42,90	34,93
	71	70	-2,45	-6,09	6,00	37,09
	74	81	0,55	4,91	0,30	24,11
$\Sigma=$	**808**	**837**	**0,05**	**0,01**	**357**	**481**

Cálculo do r

$$\bar{x}_x = 73,45\ VO_2$$

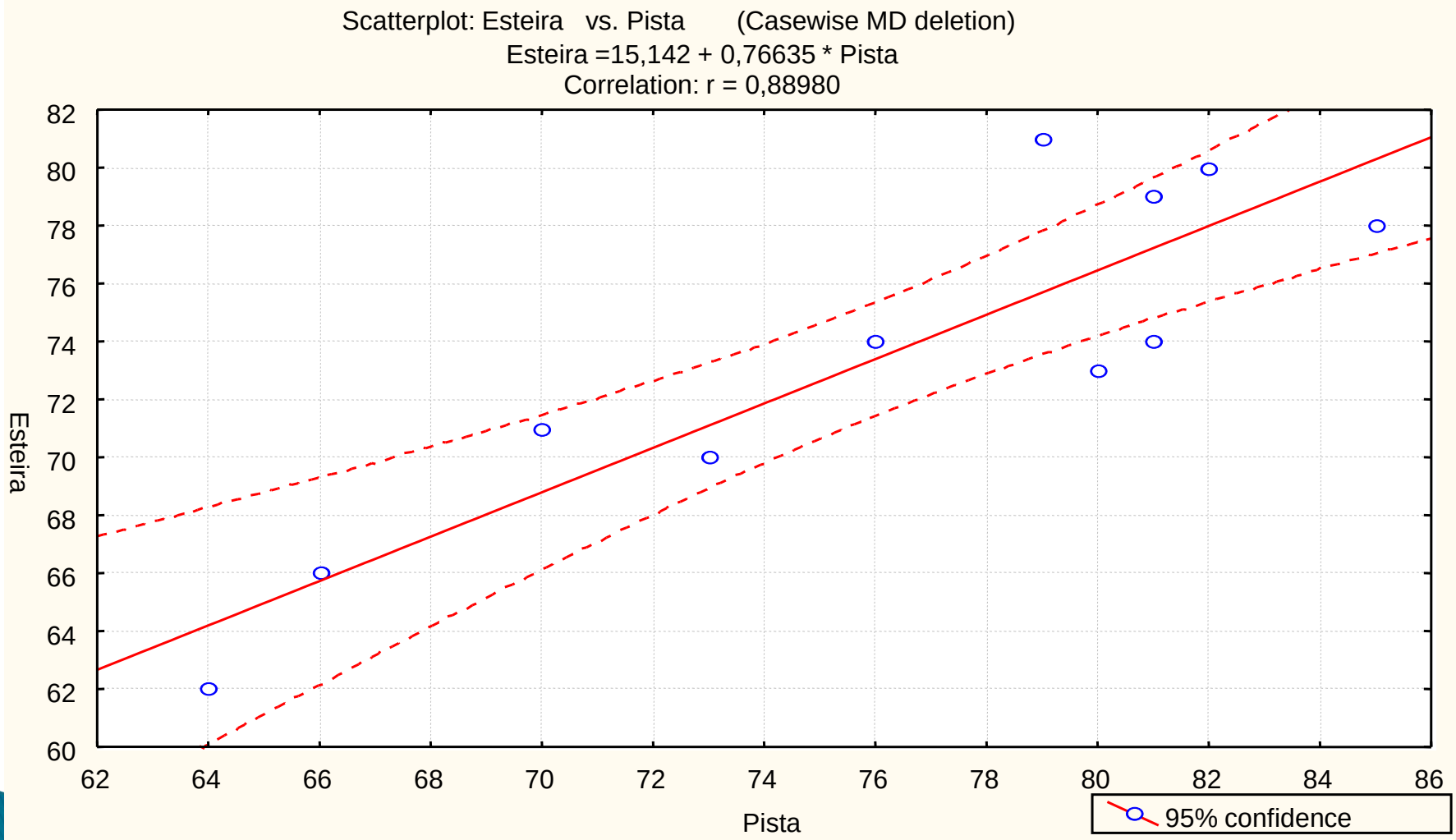
$$\bar{x}_y = 76,09\ VO_2$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum (x_i - \bar{x})^2\right)\left(\sum (y_i - \bar{y})^2\right)}}$$

$$r = \frac{(0,05).(0,01)}{\sqrt{(357).(481)}}$$

$$r = 0,89$$

CORRELAÇÃO


Scatterplot: Esteira vs. Pista (Casewise MD deletion)
Esteira =15,142 + 0,76635 * Pista
Correlation: r = 0,88980
Esteira
Pista
82
80
78
76
74
72
70
68
66
64
62
60
62
64
66
68
70
72
74
76
78
80
82
84
86
95% confidence

CORRELAÇÃO

Coeficiente de correlação (r_{xy})	Validação do Instrumento
0,90 – 0,99	Excelente
0,80 – 0,89	Muito boa
0,70 – 0,79	Aceitável
0,60 – 0,69	Questionável

Tritschler, 2003

CONCORDÂNCIA

Exemplo de boa Concordância

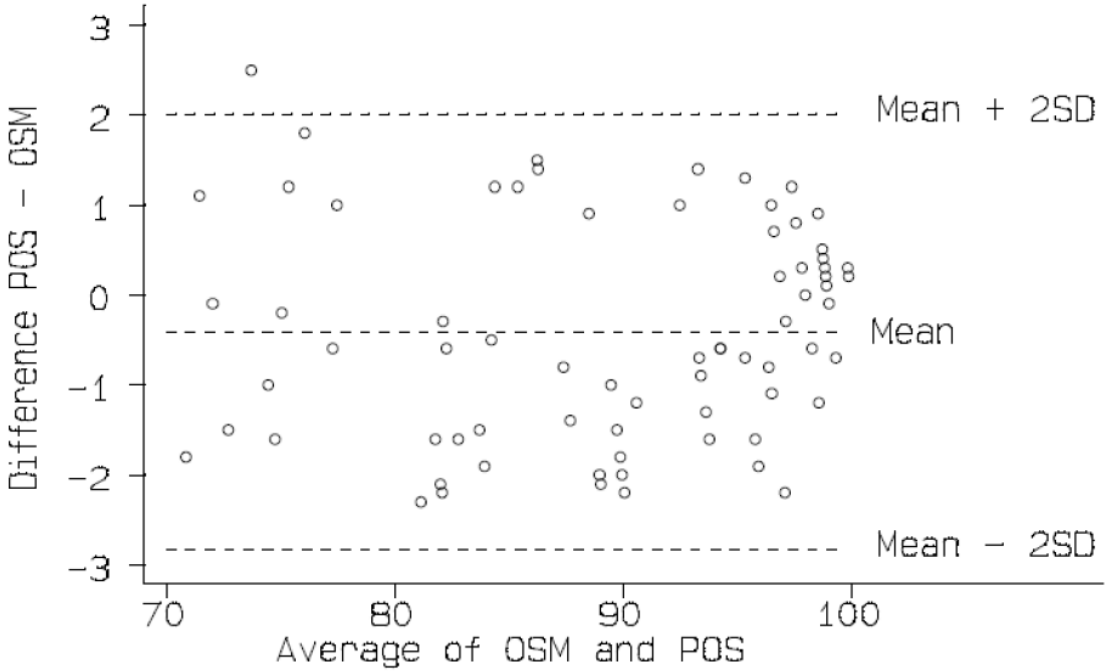
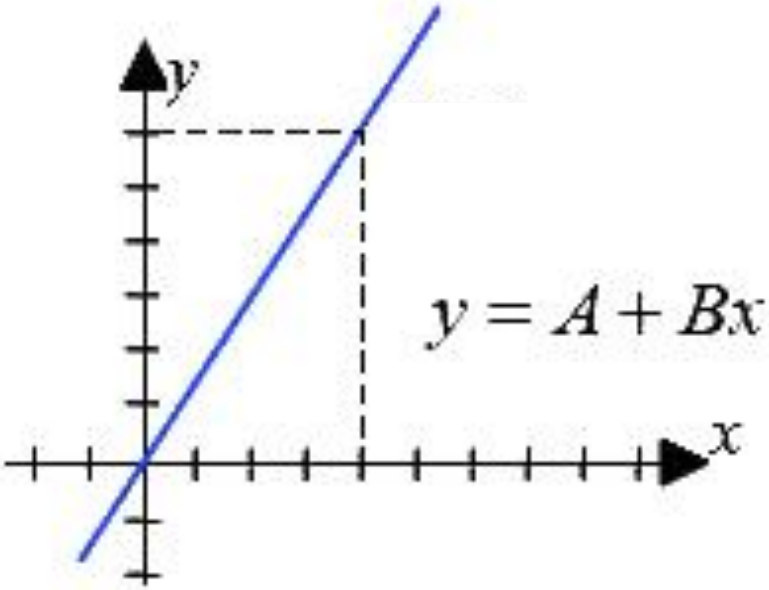


Fig 3. Oxygen saturation monitor and pulsed saturation oximeter

Bland-altman, 2003

Regressão Linear



Equação do percentual de gordura – adultos

Homens (18-61 anos):

$\sum$ 3 dobras = + + *(r=0.89;EPE=0.008g/cm3; ±3.6%MG)*

$$\text{DensCorp}(g/cm^3) = 1.1125025 - 0.0013125*(\sum 3\text{dobras}) + 0.0000055*(\sum 3\text{dobras})^2 - 0.0002440*(\text{idade})$$

$$\%MG = [(4.95/\text{DensCorp}) - 4.50]*100$$

Jackson & Pollock, 1978, Med Sci Sport Exerc