

Teorias sobre os testes discriminativos

Profª Silvia Deboni Dutcosky



Testes de diferença

Podem ser classificados em:

- Testes de diferença
- Testes de similaridade

A Potência (ou Poder) de um Teste:

$$P = 1 - \beta$$

Ou seja, é a probabilidade de fazer a decisão correta quando duas amostras são perceptivelmente diferentes.

Os erros tipo I e II são inversamente relacionados e, à medida que o erro tipo I se torna mais restritivo (se aproxima de zero), o erro tipo II aumenta.

Assim o pesquisador deve jogar com o equilíbrio entre o nível alfa e a potência do teste.

		Realidade	
		H_0 :sem diferença	H_1 :com diferença
Decisão Estatística	H_0 : sem diferença	$1 - \alpha$	β Erro tipo II
	H_1 : com diferença	α Erro tipo I	$1 - \beta$ Potência do teste

Nº de avaliadores para testes de diferença e de similaridade

O número de avaliadores requeridos para um teste de diferença não é crítico como para um teste de similaridade.

- Se você quer demonstrar que uma nova formulação é mais macia que a velha, por exemplo, você pode usar um teste de diferença com o objetivo de minimizar alfa. O erro do tipo I é minimizado neste caso porque os analistas sensoriais querem ter a certeza de que a amostra nova é diferente da velha, portanto o poder do teste não é crítico.
- Quando o objetivo do teste é que a diferença não seja perceptível, nós queremos minimizar o erro beta. Este erro é o de decidirmos incorretamente que as amostras são similares, quando na realidade elas são diferentes. A potência ($1 - \beta$) é crítica e deve ser maximizada, o que significa que o número de avaliadores deverá ser suficientemente grande para este fim.

Outra distinção entre testes de diferença e de similaridade

- Em testes de diferença nós utilizamos estatística binomial para desenvolver as tabelas que determinam o número mínimo de respostas corretas para estabelecer diferença e testamos contra as chances de acerto ao acaso.
- Em testes de similaridade, nós iremos testar contra algum nível mais alto de proporção esperada de respostas corretas e ver se nós estamos significativamente abaixo desse, a fim de chegarmos à conclusão de que as duas amostras são similares. Nós chamamos isto de proporção de discriminação permitida.

Como nós descobrimos qual é a proporção que deveria ser? Nós podemos optar por uma situação conservadora, na qual nós permitimos apenas 10% de avaliadores discriminadores, ou por uma situação menos crítica e permitir 30% de avaliadores discriminadores.

1. Teoria "Thurstoniana"
Critério d' (distância discriminativa)
2. Teoria da Detecção do Sinal
3. Abordagem R-Index

Estas teorias permitem o entendimento dos fatores que contribuem na sensibilidade, confiabilidade e na validade dos testes de diferença/ similaridade.



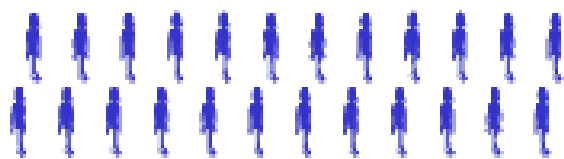
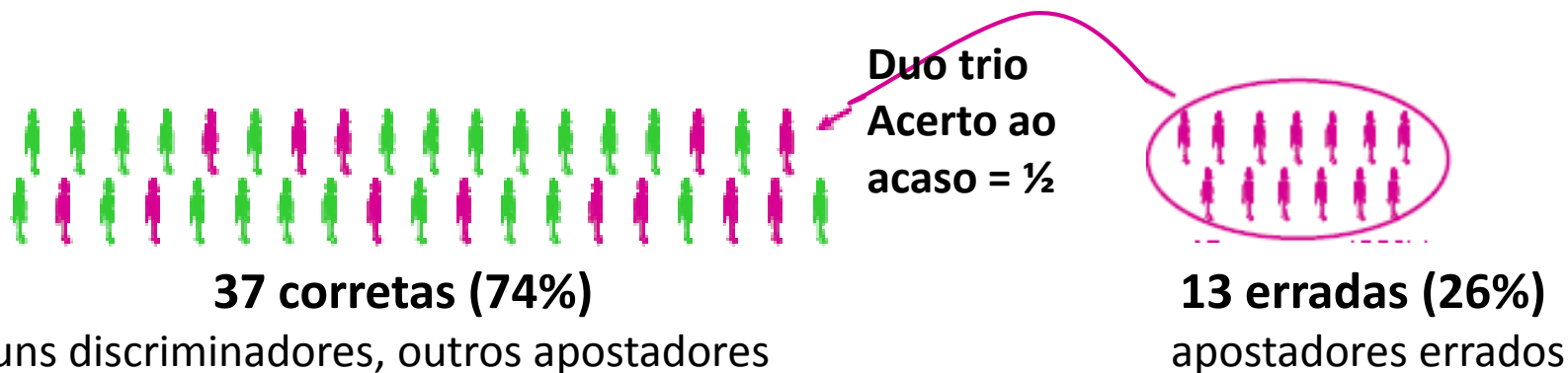
Cenário comum

- ❖ Redução de custos ou substituição de ingredientes em um produto com considerável *market share* e uma boa base de consumidores fiéis.
 - Objetivo: o produto em questão deve manter seu padrão de qualidade sensorial.
- O QUE ACONTECE SE NÓS NÃO IDENTIFICARMOS A DIFERENÇA?
 - RISCO: Decepção do consumidor fiel – *heavy users*

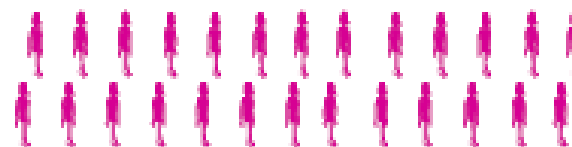
Critério nas normas atuais ISO e ASTM:

Proporção de discriminadores

Ex. 50 consumidores realizaram um teste duo-trio, obtendo-se 37 respostas corretas



24 discriminadores (48%)



26 apostadores (52%)

Teoria Thurstoniana

Thurston (1927)

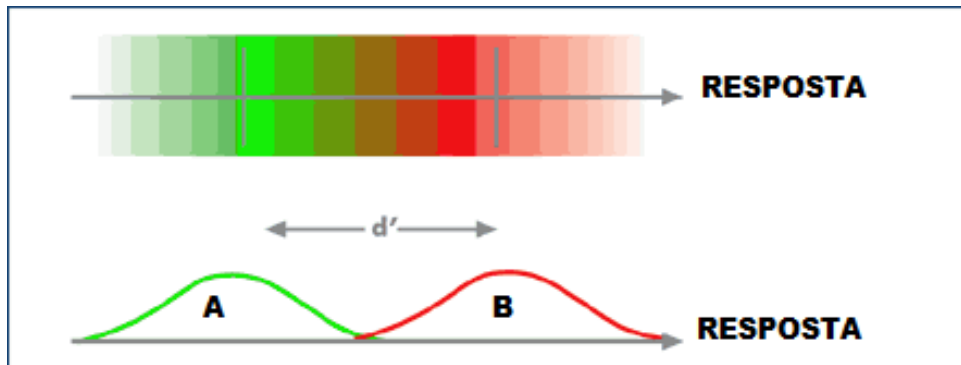
Introduzindo o conceito d' pág.156

Ao comermos o mesmo alimento várias vezes, a resposta poderia ser medida pelo número de impulsos por segundo do sistema nervoso que alcançam o cérebro. Quando degustamos um alimento, a resposta atual irá variar ligeiramente da próxima em que o alimento será degustado, podendo acontecer o efeito de adaptação ou alguns ruídos nas sinapses do sistema nervoso, pois todos os sinais dos estímulos que alcançam o cérebro podem ser acompanhados de ruídos.

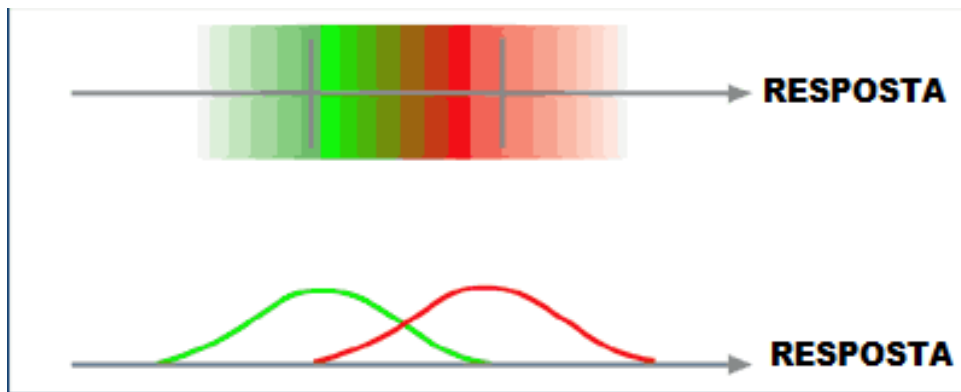
Estas variações das intensidades percebidas dos sinais podem ser representadas em uma curva chamada distribuição das freqüências. A maioria das vezes o sinal do estímulo irá ocorrer num nível médio, mas algumas vezes o sinal será percebido mais forte e outras vezes mais fraco.

Teoria Thurstoniana

Respostas não confundíveis X confundíveis

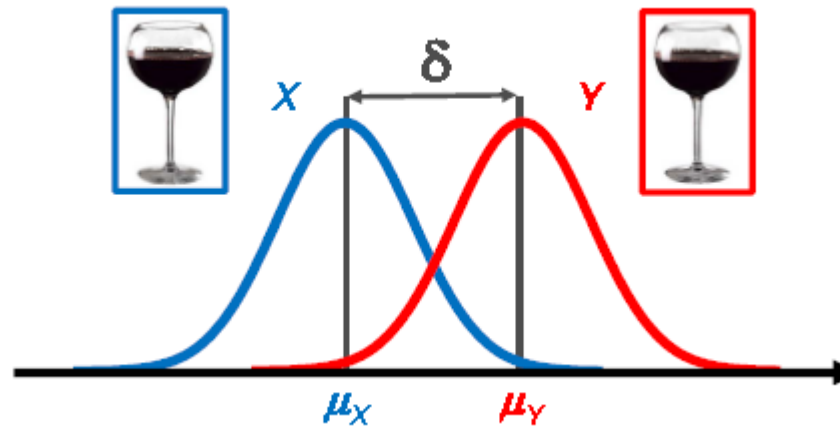


Para dois alimentos A e B não confundíveis



Para dois alimentos A e B confundíveis

Introduzindo o conceito δ e d'



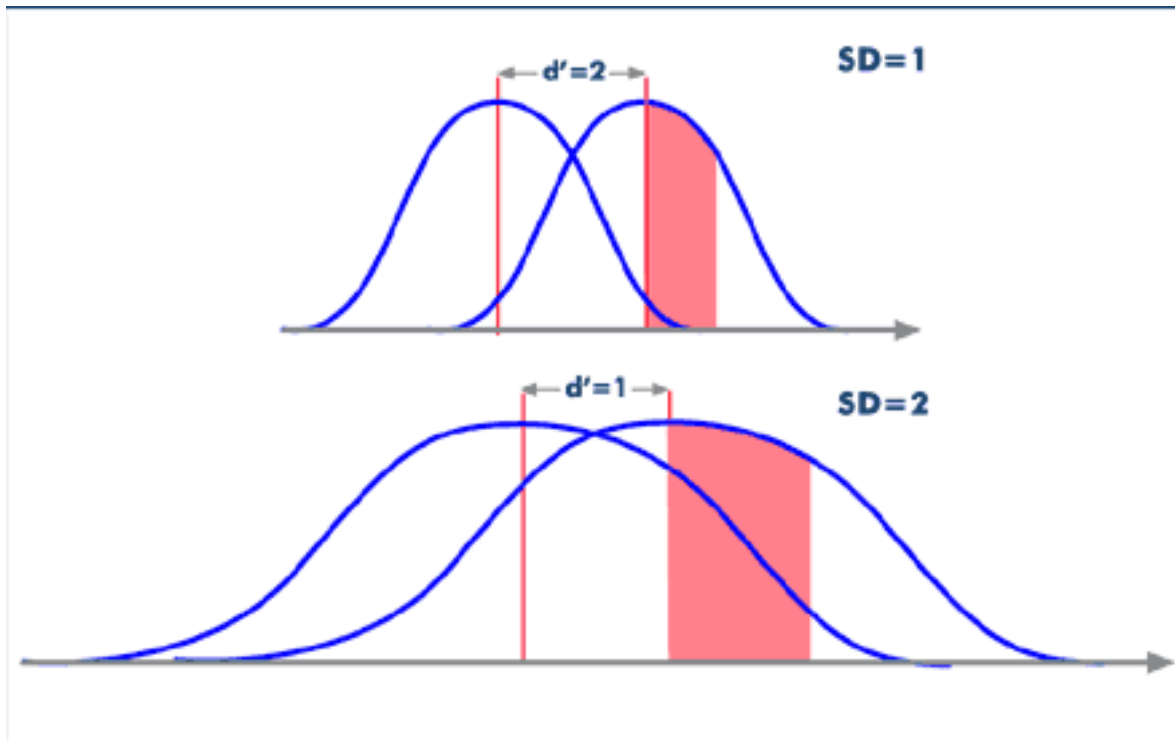
δ = distância entre as médias (μ_X e μ_Y) calculada em termos do número de desvios padrões pelos quais as duas distribuições são separadas.

d' = valor experimental de δ

Assume-se que ambos os produtos (X e Y) têm o mesmo desvio padrão. Como o estímulo avaliado é o mesmo, assume-se que as distribuições serão similares.

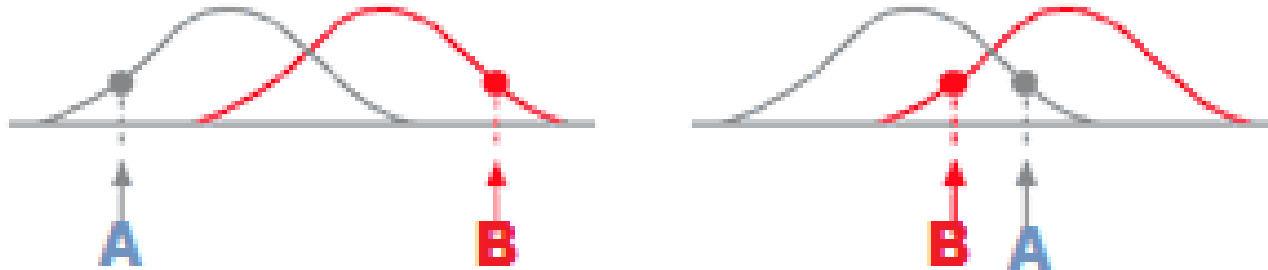
Quando δ é menor do que 1, significa que os dois estímulos podem ser confundíveis e quando δ é maior do que 1 será mais fácil distinguir os dois estímulos. O $\delta = 2$ apresentará uma probabilidade de 92% de resposta correta. O $\delta = 3$ dará uma probabilidade de 98% de resposta correta.

Relação quanto à variância



Quanto à variância, se a distribuição tiver variância maior, o d' será menor, ou seja, suas distribuições de frequência terão mais sobreposição. Podemos dizer que as curvas de distribuição acima são de um julgador mais consistente, com menor variação, indicado por um d' maior.

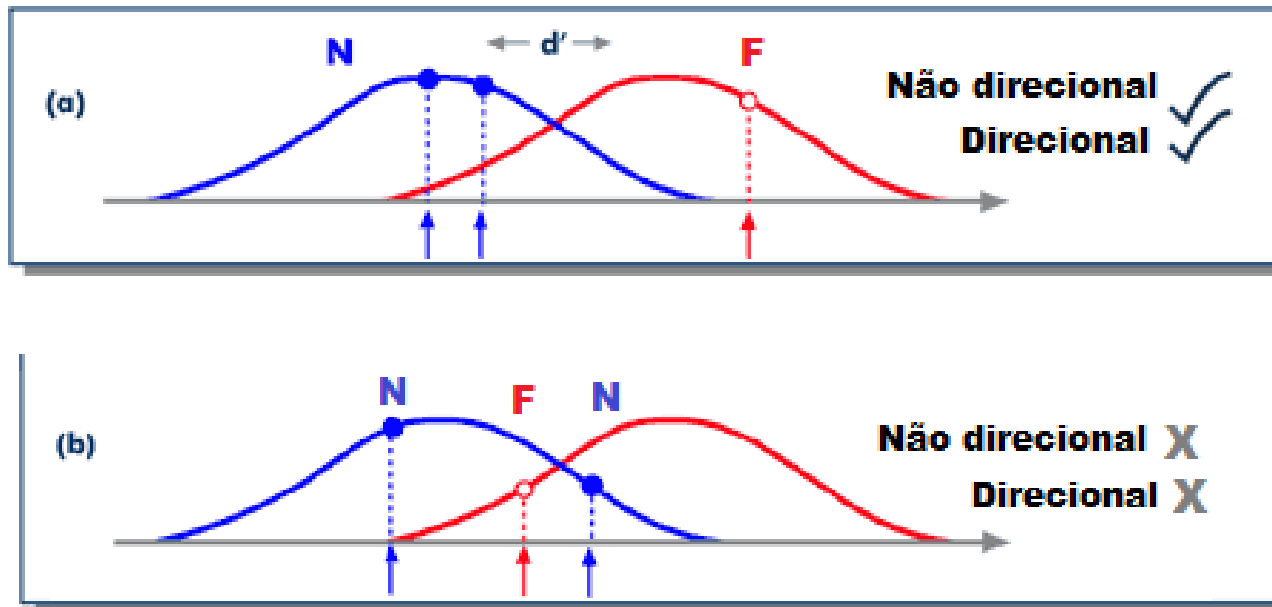
Exemplo Teste Pareado



Diferentes respostas ao estímulo mais fraco (A) e ao estímulo mais forte (B) em um teste pareado

Teste Triangular

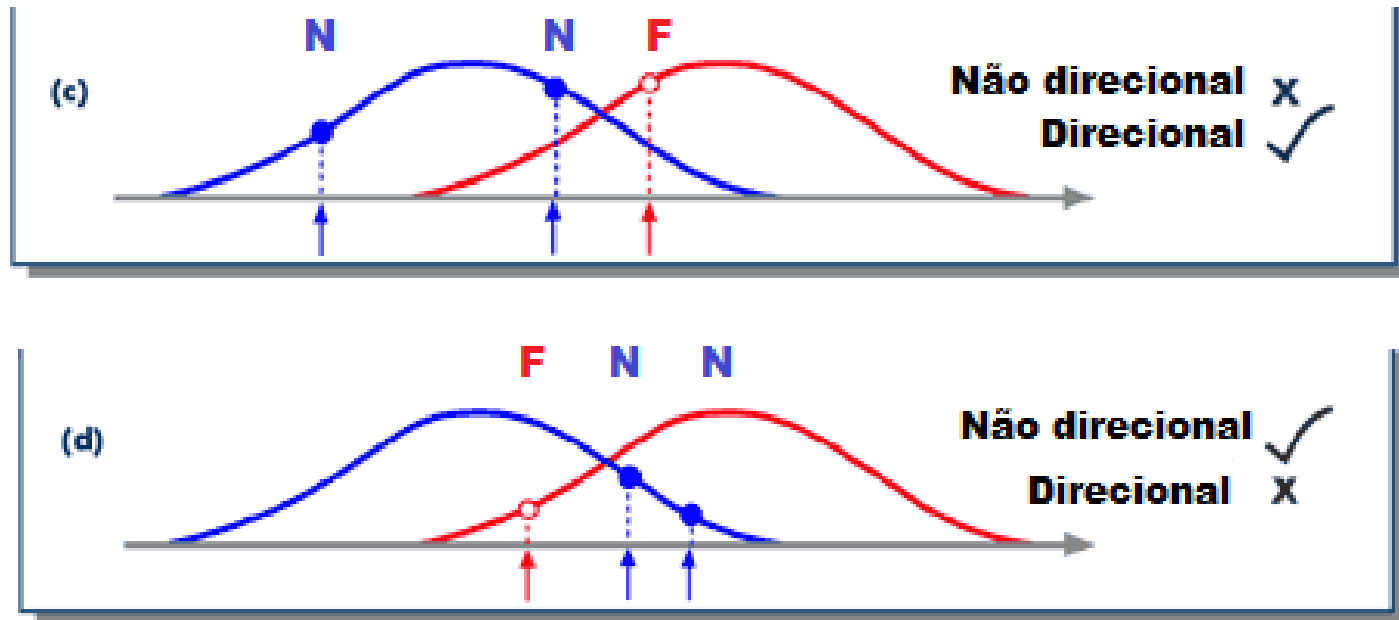
Convencional X Direcional



Diferentes respostas ao estímulo mais fraco (N) e ao estímulo mais forte (F) em um teste triangular

Teste Triangular

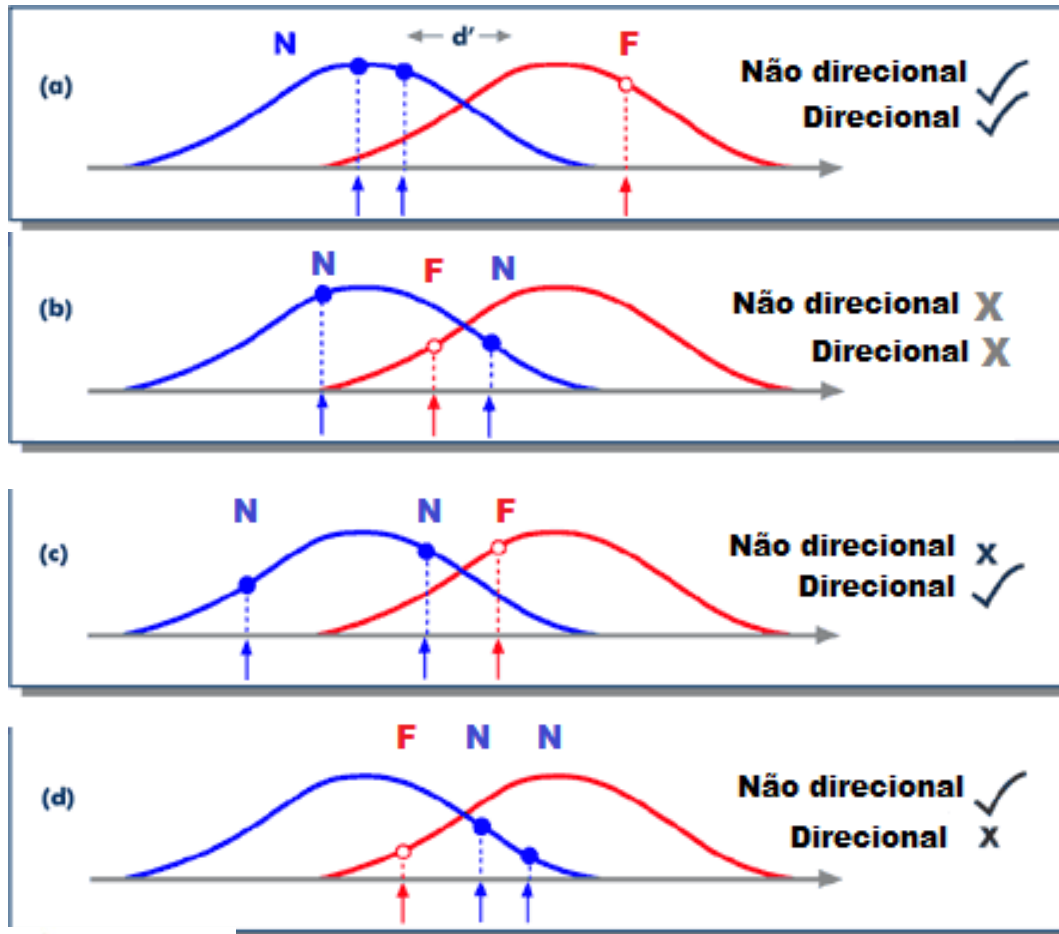
Convencional X Direcional



Diferentes respostas ao estímulo mais fraco (N) e ao estímulo mais forte (F) em um teste triangular

Simulação: respostas ao estímulo mais fraco (N) e ao estímulo mais forte (F)

Para $\delta=1,5$ e $n=100.000$ teste triangular



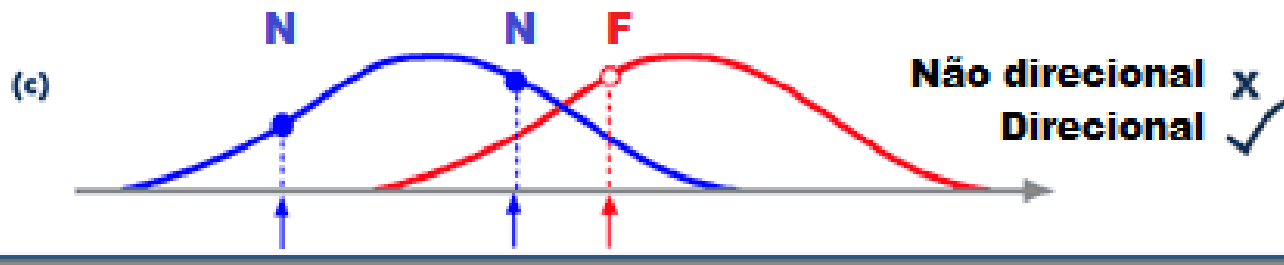
48,0%

21,1%

28,5%

2,3%

Processo de decisão



Triangular (não direcional): Resposta errada

Comparação entre distâncias

3-AFC (direcional): Resposta certa

Magnitudes

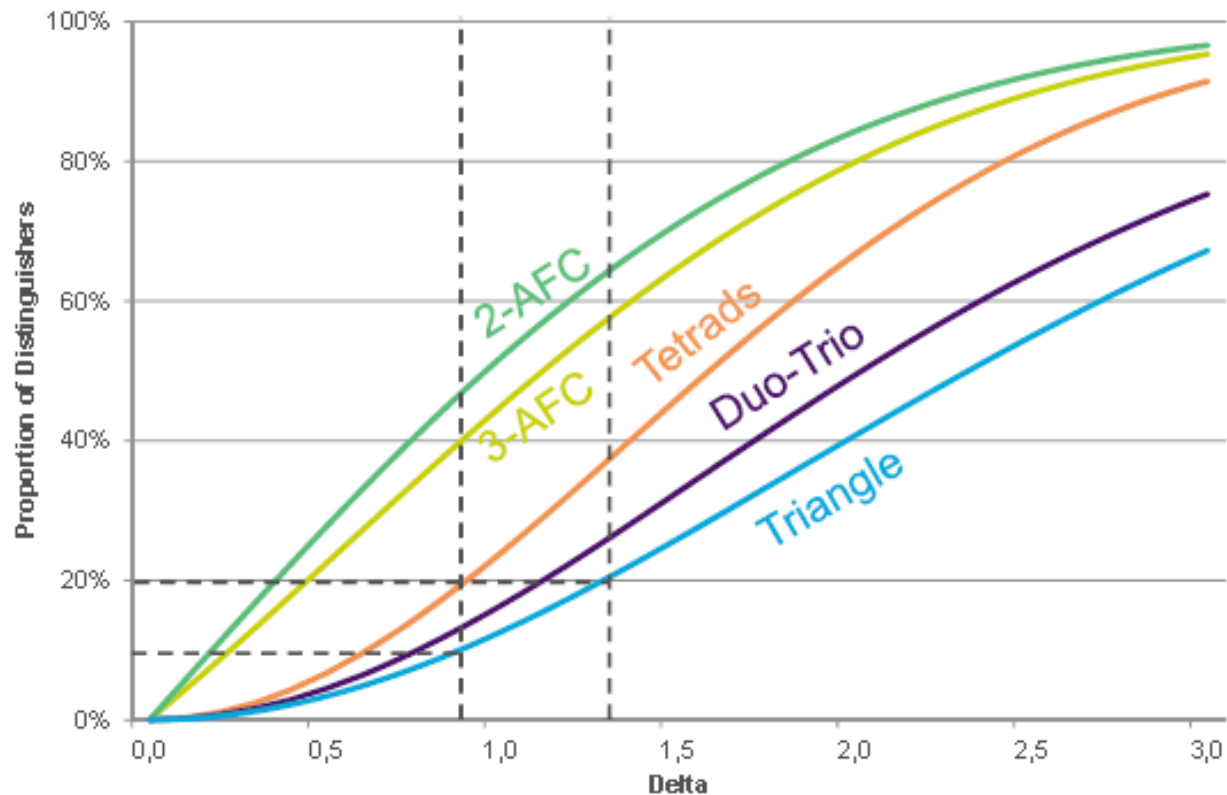
Proporção de respostas corretas e valores de d' para teste triangular direcional X Não direcional

d primo	Proporção de respostas corretas Triangular	
	Não direcional	Direcional
0,00	33,3%	33,3%
0,43	35%	46%
0,88	40%	60%
1,00	42%	64%
1,52	51%	77%
2,03	61%	87%

Proporção respostas corretas	d primo	
	Não direcional	Direcional
33%	0,00	0,00
40%	0,88	0,23
50%	1,47	0,56
60%	1,98	0,89
70%	2,50	1,24
80%	3,13	1,65
90%	4,03	2,23

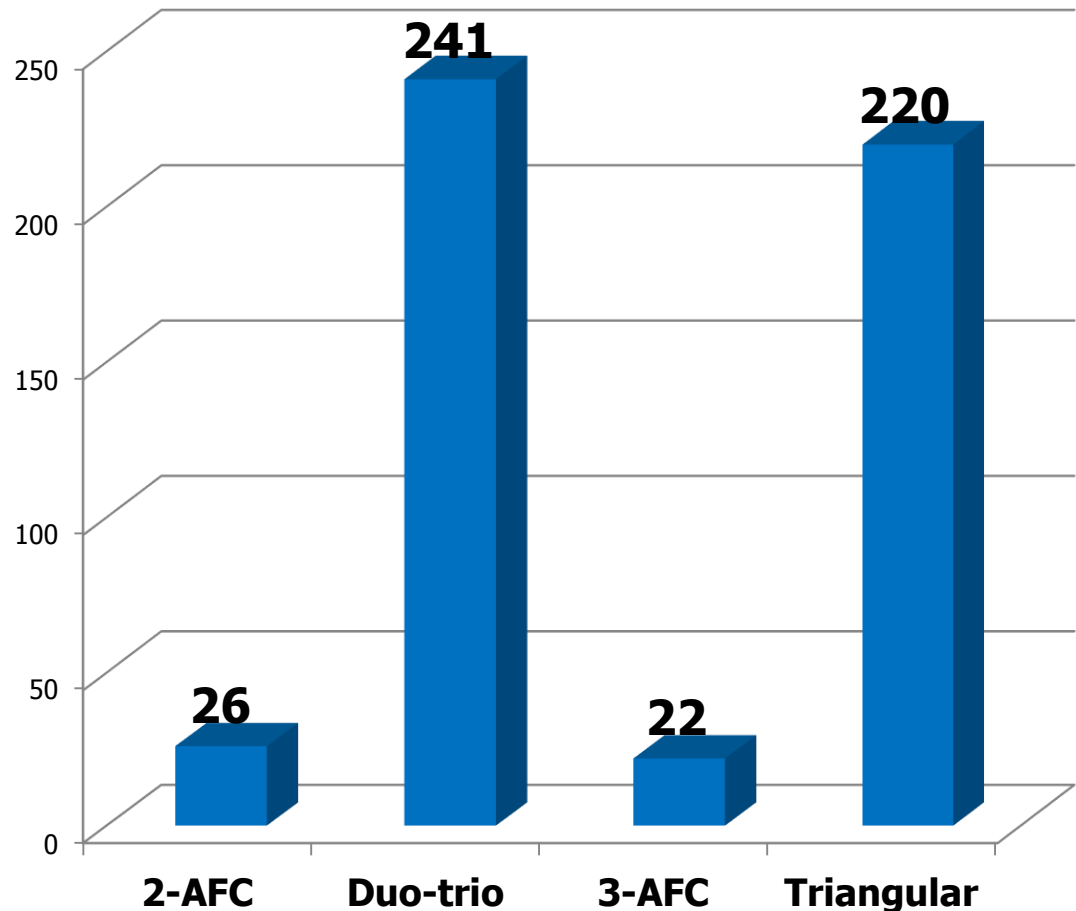
Se você alcança 70% de respostas corretas, o d' será 1,24 para o teste triangular direcional. Nós conseguimos analisar melhor em um teste direcional, pois para o teste não direcional precisaremos que d' seja 2,50 para alcançar os mesmos 70% de respostas corretas.

Relação entre δ e Pc ou Pd pode ser obtida através de curvas psicométricas



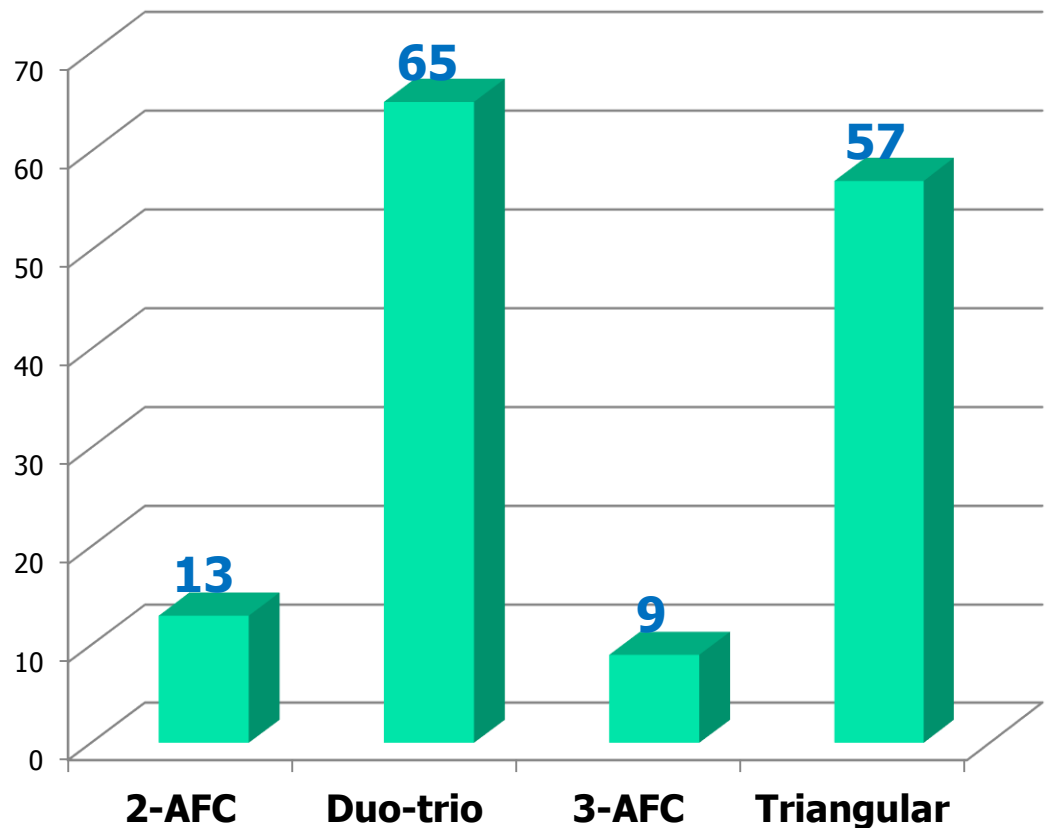
Os custos do processo de decisão

- Tamanho da diferença: 76% resp. corretas no 2-AFC ($\delta = 1,0$)
- Potência: 80%
- Nível $\alpha = 5\%$
- Quantidade necessária de julgamentos →



Os custos do processo de decisão

- Tamanho da diferença: 86% resp. corretas no 2-AFC ($\delta = 1,5$)
- Potência: 80%
- Nível $\alpha = 5\%$
- Quantidade necessária de julgamentos →



O critério P_d é método específico e também não é um critério que descreve acuradamente a habilidade discriminativa de consumidores.

Assumir que alguns consumidores vão sempre discriminar enquanto que outros vão sempre acertar “no chute” é uma hipótese irreal.

A determinação da distância discriminativa d' é dada em função de P_c , em Tabelas na ASTM E2262- 09 com o título “Standard Practice for Estimating Thurstonian Discriminal Distances”.

Proporção de respostas corretas (P_c)x 10^4 em função da distância d' para o teste triangular - pág.490

Encontre na tabela o número de P_c mais próximo do obtido no teste triangular. Para obter o valor estimado de d' : leia o valor da primeira casa após a vírgula na primeira coluna, correspondente à linha do P_c , e a segunda casa após a vírgula leia na própria coluna do P_c .

d'	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	3333	3333	3334	3334	3335	3336	3337	3338	3339	3341
0.1	3343	3344	3347	3349	3351	3354	3357	3360	3363	3366
0.2	3370	3374	3378	3382	3386	3390	3395	3400	3405	3410
0.3	3415	3421	3427	3432	3439	3445	3451	3458	3464	3471
0.4	3478	3486	3493	3501	3508	3516	3524	3533	3541	3550
0.5	3558	3567	3576	3586	3595	3604	3614	3624	3634	3644
0.6	3654	3665	3676	3686	3697	3708	3719	3731	3742	3754
0.7	3766	3778	3790	3802	3814	3827	3839	3852	3865	3878
0.8	3891	3905	3918	3932	3945	3959	3973	3987	4001	4016
0.9	4030	4045	4059	4074	4089	4104	4119	4134	4149	4165
1.0	4180	4196	4212	4228	4244	4260	4276	4292	4309	4325
1.1	4342	4358	4375	4392	4409	4426	4443	4460	4477	4494
1.2	4512	4529	4547	4564	4582	4600	4618	4636	4654	4672
1.3	4690	4708	4726	4745	4763	4782	4800	4819	4837	4856
1.4	4875	4893	4912	4931	4950	4969	4988	5007	5026	5045
1.5	5065	5084	5103	5122	5142	5161	5180	5200	5219	5239

Projeto de alteração de ingredientes nos cookies.

Estudo de Caso Pág. 180

Um dos principais fornecedores de fosfatos de uma grande empresa de biscoitos solicitou a substituição de um tipo de fosfato por outro. Esse ingrediente faz parte da formulação do fermento químico do biscoito. O fornecedor alega que isso não acarretará nenhuma alteração significativa na textura do produto e que essa substituição de ingredientes não será percebida pelo consumidor. Alega, também, que resultaria numa significativa redução de custos do produto.

O gerente da empresa de biscoitos ficou receoso de que essa substituição influenciasse na dureza do produto.

O gerente, então, solicitou que fosse realizada a análise sensorial. Para esse fim, foi feita uma produção-piloto com o novo fosfato *versus* o fosfato usado atualmente. Inicialmente, foi conduzido um teste duo-trio com 100 avaliadores, obtendo-se 55 respostas corretas. Consultando-se o Anexo D – Tabela 26, para $\alpha = 0,05$, o número de respostas corretas necessárias para estabelecer a diferença é 59, concluindo-se que não há diferença significativa entre as amostras.

Estudo de Caso (continuação)

O gerente não ficou convencido desse resultado e decidiu conduzir um teste de comparação pareada, para avaliar diretamente a dureza entre o biscoito atual *versus* biscoito com novo fosfato. Foram recrutados 30 avaliadores. Vinte e um responderam que o novo biscoito era o mais duro e nove responderam que o biscoito atual era o mais duro. Consultando-se o Anexo I – Tabela 31, para $\alpha = 0,05$, o número de respostas necessárias para estabelecer a diferença significativa de dureza é 21, concluindo-se que o novo biscoito é significativamente mais duro que o atual.

O gerente apresentou esse último resultado ao fornecedor e este argumentou que ele precisava fazer o teste com maior número de avaliadores e utilizar o teste triangular, que era melhor. Disse, ainda, que eles encontrariam uma proporção de discriminadores inferior a 10%.

Após reunião interna, decidiu-se conduzir um experimento com 240 avaliadores, utilizando o teste triangular, conforme orientação do fornecedor, com delineamento balanceado para apresentação das amostras. Foram obtidas 91 respostas corretas. Consultando-se o Anexo B – Tabela 24, precisaremos utilizar a fórmula para 240 julgamentos, a qual irá resultar no valor de $x = 92$:

$$x = \frac{240}{3} + 1,64 \sqrt{\frac{2 \cdot 240}{9}}$$

Conclui-se que não há diferença significativa entre as amostras.

Estudo de Caso (continuação)

$$P_d = 1,5P_c - 0,5$$

$$P_c = \frac{91}{240} = 38\%$$

$$P_d = 1,5 \cdot 0,38 - 0,5 \rightarrow P_d = 7\%$$

O resultado está de acordo com a argumentação do fornecedor.

O gerente de P&D da fábrica de biscoitos ficou preocupado com a discrepância do conjunto de resultados e decidiu realizar um treinamento para atualização em Análise Sensorial. Nesse curso, ele aprendeu a modelagem thurstoniana e o critério de análise por meio da distância discriminativa thurstoniana, chamada de d' .

Os valores d' para cada teste encontramos nas tabelas ASTM E2262 – 03:

Tabela 43 para o teste triangular. Sendo $P_c = 0,38$, multiplica-se por 10^4 , o que resulta em 3.800. O valor mais próximo encontrado na Tabela é 3.802, equivalente $d' = 0,73$

Tabela 45 para o teste duo-trio. Sendo $P_c = 0,55$, multiplica-se por 10^4 , o que resulta em 5.500. O valor mais próximo encontrado na Tabela é 5.498, equivalente $d' = 0,76$

Tabela 49 para o teste pareado. Sendo $P_c = 0,70$, multiplica-se por 10^4 , o que resulta em 7.000. O valor mais próximo encontrado na Tabela é 6.996, equivalente $d' = 0,74$

Projeto de alteração de ingredientes nos cookies. Teste de similaridade

Teste	Resp. corretas	N	P _c	P _d	Diferença?	d'
Duo-trio	55	100	55%	10%	Não (p = 0,18)	0,76
2-AFC (dureza)	21	30	70%	40%	Sim (p = 0,02)	0,74
Triangular	91	240	38%	7%	Não (p = 0,08)	0,73

Valores d' muito similares
Valor histórico de d' = 0,26

Conclui-se que esta diferença é significativa e será percebida pelos consumidores.

Não foi possível comprovar a similaridade

CONSIDERAÇÕES FINAIS

Para os métodos discriminativos nós podemos dizer que:

- ❖ O critério atual P_d é método específico e também não é um critério que descreve acuradamente a habilidade discriminativa de consumidores;
- ❖ O critério δ se ajusta melhor, porém exige um valor histórico de controle para cada produto.

Teoria da Detecção de Sinal

pág.167

A fundamentação das teorias da modelagem Thurstoniana e da detecção de sinal é a mesma. A principal diferença entre as duas teorias é o método utilizado para cada uma. Com a teoria de detecção de sinal os testes de escolha forçada não são usados para estabilizar o critério de análise.

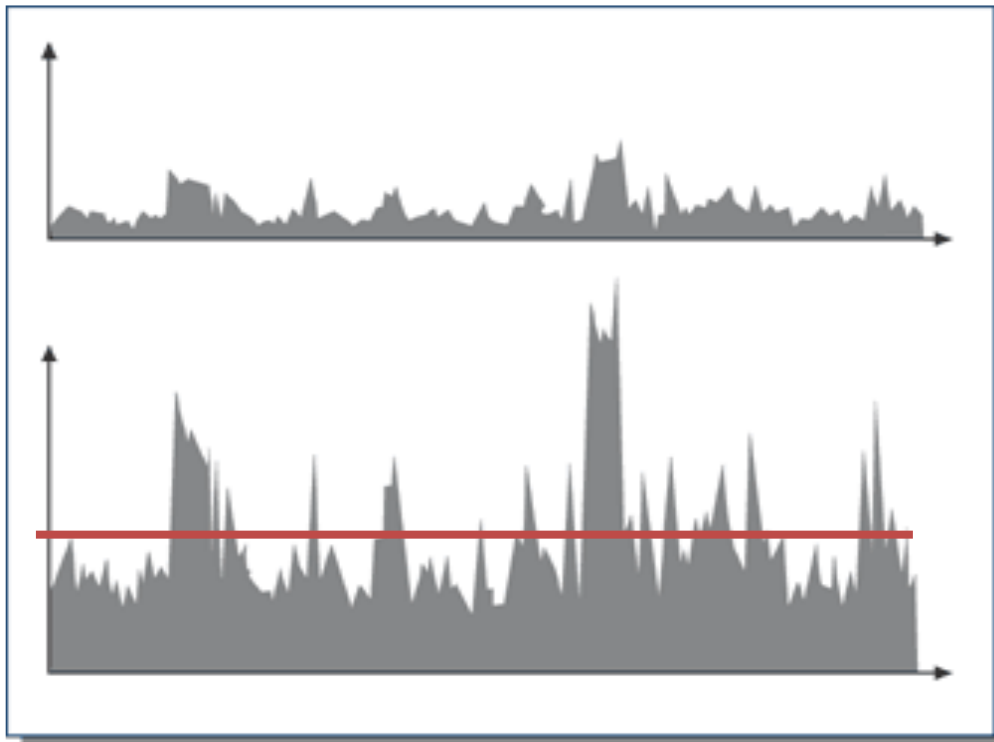
A teoria de detecção de sinais foi inicialmente desenvolvida na área psicofísica da visão e do som e também faz parte do campo de engenharia para aplicações em que os engenheiros estão interessados na diferenciação entre o **Sinal** e o **Ruído** (LEE; HOUT, 2009).

Teoria da Detecção de Sinal

Diagrama representativo dos impulsos nervosos que o cérebro recebe sob uma luz brilhante.

Eixo X = tempo

Eixo Y = intensidade percebida da luz



Escala real

Escala ampliada

Teoria da Detecção de Sinal

Suponha que você esteja em um quarto escuro e olhando para uma parede onde a luz esteja brilhando intermitente, variando a intensidade, e a sua tarefa seja determinar se a luz está brilhando ou não. A sua escolha é “sim” ou “não”. A sua resposta poderá estar dentro de 4 categorias:

Acertos: são as vezes que você diz que existe um sinal e a sua resposta está **correta**;

Alarmes falsos: são as vezes que você diz que existe um sinal, porém não existe e a sua resposta está **errada**;

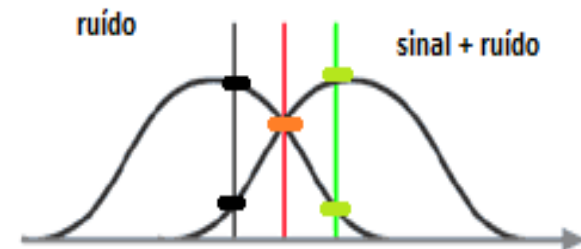
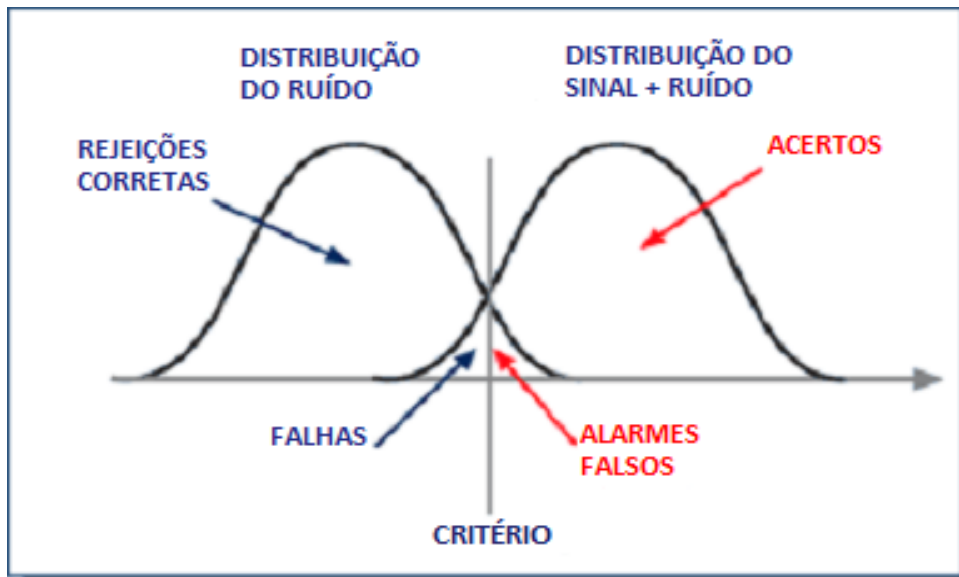
Rejeições corretas: são as vezes que você diz que não existe um sinal e a sua resposta está **correta**;

Falhas (erros): são as vezes que você diz que não existe um sinal, porém existe e a sua resposta está **errada**.

Ao invés de estabilizar o critério em um lugar fixo utilizando os testes de escolha forçada, toma-se um estímulo por vez e faz-se um julgamento. Você pode definir seu critério onde quiser e os dados serão analisados como: acertos, alarmes falsos, rejeições corretas e falhas.

Teoria da Detecção de Sinal

Na modelagem Thurstoniana utiliza-se a distribuição das frequências. Assumindo-se que ocorram estímulos, quando um produto é avaliado repetidamente por um julgador, o sinal ao cérebro pode ser medido como uma distribuição de frequência. Na maioria das vezes é percebido com uma intensidade média, porém algumas vezes este sinal pode ser percebido com intensidade mais fraca ou mais forte.

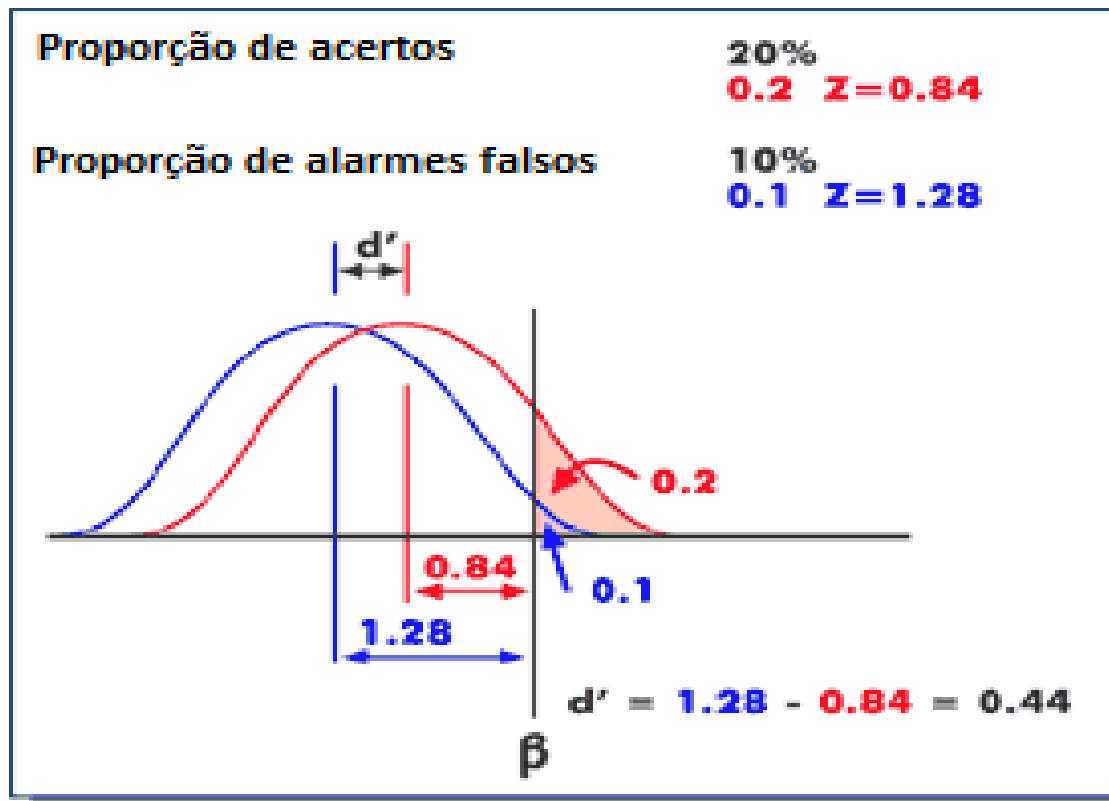


Critério β :

- $\beta = 1,0$ sem viés
- $\beta > 1,0$ critério estrito
- $\beta < 1,0$ critério com folga

Exemplo: Detecção de Sinal

Uma pessoa foi questionada 200 vezes se estava enxergando a luz brilhar na parede. Em 100 vezes o sinal de luz estava sendo emitido na parede e em 100 vezes o sinal de luz não estava sendo emitido. A pessoa conseguiu identificar o sinal correto em 20 vezes e em 10 vezes ela disse que a luz estava sendo emitida quando na realidade não estava.



R-Index

O R-Index foi introduzido por Brown (1974) para investigar os vários mecanismos da memória. O R-Index é uma medida não paramétrica e é definido como a probabilidade de escolha do SINAL em comparações pareadas com o RUÍDO.

O teste é realizado apresentando-se dois produtos em ordem aleatorizada. Por exemplo: um produto é o controle e o outro é reformulado com ingredientes de menor custo. O produto controle é o SINAL e o produto reformulado é o RUÍDO. Quando o produto é analisado, o julgador irá responder se é o Controle ou o Produto Reformulado e responder se ele tem certeza ou não.

	S	S ?	N ?	N	
SIGNAL	a	b	c	d	$N_1 = a + b + c + d$
NOISE	e	f	g	h	$N_2 = e + f + g + h$

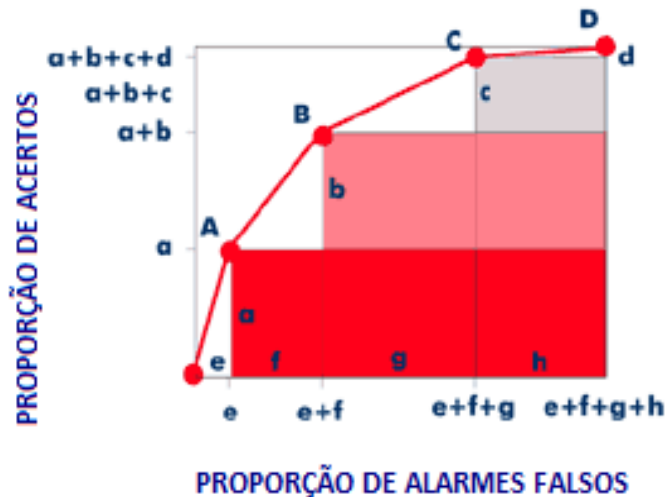
$$R = \frac{a(f+g+h) + b(g+h) + ch + 1/2 (ae+bf+cg+dh)}{N_1 N_2} \times 100$$

$N_1 N_2$
 total number of comparisons

Annotations for the formula:

- Arrows from "correct" point to $a(f+g+h)$, $b(g+h)$, and ch .
- Arrows from "don't know" point to ae , bf , cg , and dh in the $1/2 (ae+bf+cg+dh)$ term.

	S (Controle com certeza)	S ? (Controle, não tenho certeza)	R ? (Reformulado, não tenho certeza)	R (Reformulado com certeza)
SINAL (Controle)	8	1	1	0
RUÍDO (Reformulado)	0	1	2	7



$$R\text{-Index} = \frac{8(1+2+7) + 1(2+7) + 1.7 + \frac{1}{2} (8.0 + 1.1 + 1.2 + 0.7)}{(8+1+1+0) \cdot (0+1+2+7)}$$

$$R\text{-Index} = 0,975 \text{ ou } 97,5\%$$

Respostas do julgador

	S	S?	R?	R	
SINAL	a	b	c	d	TOTAL = a+b+c+d
RUÍDO	e	f	g	h	TOTAL = e+f+g+h

R-index

$$R = \frac{a(f+g+h) + b(g+h) + ch + \frac{1}{2} (ae+bf+cg+dh)}{(a+b+c+d) (e+f+g+h)}$$

Teste Tetraédrico pág.125

Quatro amostras codificadas são apresentadas ao avaliador, duas de um tratamento A e duas de outro tratamento B. O avaliador receberá a instrução de agrupar as amostras duas a duas, em dois conjuntos similares. Esta metodologia é própria para aplicação em casos de diferença sensorial geral, ou seja, detectar qualquer diferença como objetivo da análise ou para casos em que não seja conhecido o atributo específico de controle.

O avaliador é instruído a separar dois grupos de duas amostras, o que é diferente da tarefa de apenas agrupar duas amostras similares. Pois, se for instruído a formar apenas o primeiro grupo de amostras similares ele pode ser conduzido a selecionar as amostras intermediárias como um par, deixando para trás as duas amostras que apresentam maior diferença como o segundo par.

São seis combinações possíveis de ordem de apresentação:

AABB, ABAB, ABBA, BBAA, BABA, BAAB

Teste Tetraédrico

Agrupe as amostras em dois pares similares entre si



Utiliza-se a mesma tabela e os mesmos critérios do teste triangular.

Para entender porque a probabilidade de acerto ao acaso é de $1/3$ você precisa considerar a tarefa do julgador, que ao receber as 4 amostras A1,A2,B1 e B2 ele necessariamente irá experimentar a primeira, por exemplo A1 e terá uma chance em 3 amostras de casar esta amostra com a A2.

Teste Tetraédrico

Ennis (2012) demonstrou, através da modelagem thurstoniana, que o teste tetraédrico é mais sensível, mais potente e ainda requer um menor número de julgadores para identificar diferenças de mesma grandeza entre dois produtos quando comparado ao teste triangular.

O autor também apresentou todas as considerações estatísticas quanto ao acréscimo de um estímulo e seus efeitos sobre a fadiga sensorial, efeitos de memória e de adaptação, demonstrando que esses efeitos são bem menores, mantendo as vantagens do tetraédrico.

Exemplo - Dutcosky (2013)

Ocorreu um problema de manutenção em um equipamento modelador de batom de um fabricante de cosméticos e não foi possível mais alcançar a pressão especificada inicial. O equipamento mantinha a pressão 5% abaixo da especificação. O batom modelado com a pressão 5% menor apresentou alterações na análise instrumental de firmeza, porém questionou-se se essa diferença era perceptível pelo consumidor e se justificava a compra de um novo equipamento.

Após discussão com o analista sensorial, decidiu-se que deveria ser aplicado um teste que detectasse qualquer diferença sensorial, pois esta diferença na pressão da modelagem poderia não estar afetando só a firmeza do batom, mas outras características da textura, aparência ou cor poderiam estar sendo afetadas.

Para o teste tetraédrico, solicita-se ao avaliador que agrupe os dois conjuntos de amostras similares. A metodologia de análise comparativa foi realizada solicitando-se ao julgador que passasse cada uma das 4 amostras de batom sobre círculo previamente demarcado no antebraço direito, com repetição do mesmo procedimento no antebraço esquerdo, para só então, o julgador tomar uma decisão.

Amostras A – 376 e 589 – batom Padrão

Amostras B – 402 e 711 – batom com pressão 5% menor.



Exemplo – Dutcosky (2013)

Amostras A – 376 e 589 – batom Padrão

Amostras B – 402 e 711 – batom com pressão 5% menor.



Número de avaliadores:

Para conhecer o número mínimo de avaliadores no programa V-Power, selecione teste tetraédrico não-direcional (*unspecified tetrads test*) e no novo quadro selecione modelo probabilístico (*Guessing model*). Então, procure o quadro “Antes do Teste” (*Before the test*) para o cálculo do tamanho da amostra.

O objetivo deste teste é provar a similaridade, por isso foi escolhido $\beta = 0,10$ e $\alpha = 0,20$ e $Pd = 20\%$

Coloque a entrada dos dados:

$Pd = 0,20$

risco $\alpha = 0,20$

Potência do teste = 0,90

E o programa recomendará o mínimo de 71 avaliadores.

Exemplo - Dutcosky (2013)

Conseguiu-se 60 avaliadores, com 23 julgamentos corretos.

O delineamento da apresentação das amostras foi completo, repetindo-se 10 vezes as seis combinações possíveis de apresentação.

Decidiu-se avaliar estes dados, antes de continuar o teste.

No quadro “Após o teste” (*After the test*):

$N = 60$ $N_c = 23$ $P_d = 0,20$ risco $\alpha = 0,20$

Obtemos os seguintes valores:

Valor de $p = 0,244$

Potência do teste = 0,878

O valor de p não é significativo, portanto não há diferença perceptível entre os batons analisados.

Podemos concluir com 88% de certeza que o batom modelado com pressão inferior não difere significativamente do batom Padrão. Esta nova especificação para o equipamento de modelagem poderá ser permitida porque os batons são suficientemente similares



Exemplo - Dutcosky (2013)

No quadro “Depois do Teste” (*After the test*), para calcular o intervalo de confiança inserimos os seguintes dados:

$$N = 60$$

$$N_c = 23$$

$$\text{Nível de confiança} = 0,95$$

Resultará que a proporção de discriminadores é 8% e tem 95% de chance de estar no intervalo entre 0% e 26%.

Observe que o critério inicial era que a proporção de discriminadores fosse menor que 20%, porém a análise não foi realizada com o número de avaliadores suficiente para obtenção deste resultado.

Neste caso foi possível obter uma conclusão, mas a análise deveria ter sido realizada com 71 julgadores para que não houvesse o risco da insuficiência de critérios ou de potência para obter-se uma conclusão.



Mesmo Exemplo = resolução com d'

pág. 184

No histórico da garantia de qualidade já é utilizado o teste triangular, porém pretende-se começar a utilizar o teste tetraédrico. O valor médio histórico do d' entre os diferentes lotes de fabricação deste tipo de batom é $\delta = 0,93$

Para conhecer o número mínimo de avaliadores no programa V-Power, selecione: *unspecified tetrads test*.

No novo quadro selecione modelo thurstoniano. Então, procure o quadro *Before the test* para o cálculo do tamanho da amostra.

Coloque a entrada dos dados:

δ histórico = 0,93

risco $\alpha = 0,20$

Potência do teste = 0,90

E o programa recomendará o mínimo de 63 avaliadores.



Mesmo Exemplo = resolução com d'

Conseguiu-se 60 julgadores, obtendo-se 23 julgamentos corretos.

Alimentando-se estes dados no quadro *After the test*:

$N = 60$ $N_c = 23$ δ histórico = 0,93 risco $\alpha = 0,20$

Obtemos os seguintes valores:

Valor de $p = 0,244$

Potência do teste = 0,901

Para sabermos o valor de d' , calculamos a proporção de respostas corretas:

$$P_c = \frac{23}{60} = 0,3833$$



Mesmo Exemplo = resolução com d'

Consultando-se a Tabela de Ennis (2012) para o teste tetraédrico.

$P_c = 0,3833$, multiplica-se por 10^4 , resulta 3833.

O valor mais próximo encontrado na Tabela é 3830

Equivalente $d' = 0,53$

$d' = 0,53$ é inferior ao valor médio entre lotes $\delta = 0,93$

Podemos concluir com 90% de certeza (Potência do teste = 0,901) que o batom modelado com pressão inferior não difere significativamente do batom Padrão.

