RENY REIS GATTÁS

TÉCNICA "PROBIT ANALYSIS"

E SUA

APLICAÇÃO AOS TESTES PSICOLÓGICOS

TRABALHO APRESENTADO AO IME DA USP PARA OBTENÇÃO DO GRAU DE MESTRE EM ESTATÍSTICA APLICADA EM JUNHO DE 1971

JUNHO DE 1971

- SÃO PAULO -

RENY REIS GATTÁS

TÉCNICA "PROBIT ANALYSIS"
E SUA
APLICAÇÃO AOS TESTES PSICOLÓGICOS

Trabalho apresentado ao IME da USP
para obtenção do grau de Mestre em
Estatística Aplicada em Junho de
1971

JUNHO DE 1971
← SÃO PAULO →

INTRODUÇÃO

A ciência tem por objetivo o estabelecimento de relações constantes e gerais entre fenômenos, ou seja, leis,e,a coordenação destas leis formando princípios e teorias através dos quais os fatos empíricos podem ser justificados e preditos. Assim, diante da observação de um fenômeno queremos explicá-lo em função de fatos conhecidos, relacioná-los de forma rigorosa ao que já está cientificamente estabelecido. Várias hipóteses vão surgir para explicá-lo, havendo então, necessidade de execução de experimentos ou séries de observações, por meio dos quais chegaremos a um certo resultado.

Com o fim de chegar a resultados mais precisos surgem as medidas e as relações quantitativas. Nas ciências menos evoluídas já achamos satisfatório o estabelecimento de correlações entre os fenômenos, ao passo que para as mais evoluídas é possível ir além, chegando a uma racionalização.

Considerando uma ciência como a física, por exemplo, vamos encontrar uma série de conceitos já estabelecidos como tempo, massa, velocidade, etc. ..., aos quais podemos chamar constructos. Sendo uma ciência bem desenvolvida existem relações entre vários constructos, estabelecidas através de deduções lógicas. Para que os constructos tenham valor científico é preciso que sejam operacionais, ou seja, que estejam ligados aos dados observados direta ou indiretamente. Daí a necessidade de regras de correspondência que definirão constructos teóricos em função de dados observáveis. Em geral, são regras que permitem a determinação de números para representar magnitudes. Assim, por exemplo, existem certas relações entre os números. Se verificarmos que relações equivalentes são satisfeitas pelas magnitudes da grandeza em questão podemos atribuir números a estas magnitudes, possibilitando-nos exprimir as relações entre os constructos em termos de equações matemáticas, o que nos permitirá caminhar com maior rapidez através do campo teórico.

Ligando, então, o campo teórico aos dados observados temos regras de correspondência que em geral consistem, como dissemos, na atribuição de números para representar as variações e oscilações dos fatos observados. Esta atribuição é feita,então, de de modo que as relações entre os números correspondam às às relações entre as magnitudes da grandeza estudada. Sabemos que os números reais, entre outras características, possuem as características de (a) ordem, (b) distância e (c) origem. Nem sempre é possível o estabelecimento de uma escala de sorte que as relações entre as magnitudes satisfaçam as relações gozadas pelos números reais. Segundo as características acima satisfeitas pelas escalas,podemos classificálas em:

1. Escala ordinal - satisfaz a característica (a).
2. Escala de intervalo - satisfaz as características (a) e (b).
3. Escala razão - satisfaz as três características.

Através da primeira podemos apenas ordenar os elementos quanto às magnitudes da grandeza estudada. A segunda, além da ordenação permite à atribuição de números às magnitudes de sorte que a diferença entre estes números refletirão a real diferença entre as magnitudes. A terceira, além das duas possibilidades acima, possibilita nos chegar às razões entre as magnitudes.

É claro que segundo as escalas usadas, nessas conclusões poderão chegar a um maior ou menor grau de precisão e um maior ou menor refinamento.

Costuma-se também, classificar as escalas segundo a maneira pela qual as observações são feitas, segundo o meio pelo qual elas foram contraídas. É aqui que existe a real diferença entre escalas psicológicas, biológicas e,escalas físicas e químicas. Podemos construir escalas para medir uma certa grandeza através de um processo de observação direta, por meio de um processo dedutivo, ou ainda através de definição. Consideramos a grandeza extensão. Existem relações entre as várias magnitudes deste constructo que podem ser observadas diretamente. É possível então, a atribuição de números a essas magnitudes de modo a refletir as relações existentes entre elas. Esta atribuição é feita diretamente às magnit... do constructo em questão, não através de outras variáveis. Ela se baseia nas relações entre as magnitudes do próprio constructo a ser medido. Campbell (28) chama esta espécie, de medida fundamental, ou medida de magnitude A. Algumas escalas físicas são construídas através de processo dedutivo. Quanto mais progridem as ciências mais relações exatas são estabelecidas entre os fenômenos. Esca la dedutiva de certa grandeza é construída através de leis relacionando esta grandeza a outras, sendo as últimas avaliadas por meio de escalas fundamentais. Esta espécie de medida é chamada medida deduzida ou medida de magnitude B. Finalmente temos uma terceira espécie de medida que é a medida por definição ou medida por *fiat*. Constructos como inteligência, status sócio-econômico, por exemplo, não podem ser observados diretamente. É preciso então definir estes constructos em termos de constructos passíveis de observação Assim por exemplo, inteligência definida teoricamente como *habilidade de adaptação às novas situações*, será definida operacionalmente como um score em dado teste. A maior parte das escalas psicológicas, biológicas, sociológicas, é estabelecida através deste meio.

O objetivo de todo pesquisador é chegar à elaboração de escalas de reconhecimento e compreensão universais. Considerando uma fita métrica, rigorosamente estabelecida, com mínimas subdivisões, vemos que mede extensões de qualquer objeto, chegando a resultados que podem ser comparados e interpretados universalmente. Não só é indiscutível o fato que esta fita preenche suas finalidades, ou seja, que realmente mede extensão, como também que os dados obtidos são altamente precisos. Isto é decorrente do fato de escala para medir extensão ser construída diretamente. Em geral, para instrumentos cuja elaboração repousa nas observações diretas das magnitudes da propriedade em questão, ou através de processo dedutivo, o principal objetivo é conseguir a mais perfeita precisão, pois é plenamente aceita a adequabilidade destes para medir o que desejamos. No campo psicológico, principalmente no que diz respeito aos testes, esta última questão é de grande importância. Construindo escalas através de definição surge o problema de se saber até que ponto a definição operacional do constructo em questão, é condizente com a definição teórica. Medida por definição exige muita intuição, habilidade, visão do pesquisador. De certa forma é algo subjetiva. Uma consequência deste fato é encontrarmos grande variedade de escalas, todas pretendendo medir a mesma propriedade, às vezes aparentemente eficientes, levando-nos, contudo, a resultados completamente diversos. É difícil imaginarmos um teste com valor universal. Sabemos que, ao nascer, todo indivíduo possue habilidades em forma latente que se desenvolverão com o tempo e, de acordo com o meio em que vive. Assim, a inteligência de uma certa criança vai depender não somente de fatores hereditários, mas também de seu ambiente. Esta é uma das causas pelas quais dificilmente podemos a dotar um teste em âmbito universal. Nível cultural de uma pessoa, costumes,hábitos,influem grandemente sobre os resultados. Muitos pesquisadores têm se esforçado pela elaboração de escalas de inteligência com a finalidade de obter resultados independentes do meio,da educação,instrução,mas,sim apenas da real inteligência do indivíduo. Este objetivo,contudo,até agora não foi atingido. Assim,uma escala de grande valor seletivo em um país pode não o ser em outro,questões com alto poder discriminativo para o primeiro poderão não ser eficientes para o segundo. É preciso, então estudá-la novamente a fim de obter uma melhor adaptação.

Do que ligeiramente foi exposto, um fato podemos constatar: validade é um problema fundamental para os que se dedicam à elaboração de testes. Primeiramente é preciso saber se o teste realmente mede o que pretendemos medir. Neste ponto devemos lançar mão de algo que seja indicador deste fato. Assim,constuído um teste para medir vocação para matemática,podemos aplicá-lo a uma amostra de crianças,e comparar os resultados do teste às suas notas de aproveitamento nesta matéria,uma vez garantida a igualdade de condições durante o curso. As notas de aproveitamento seriam então, o critério com base no qual,vamos decidir se o teste é ou não válido. Um indicador desta validade poderia ser o coeficiente de correlação entre o teste e o critério.

Outro problema que surge também ao elaborarmos testes,e instrumentos de uma forma geral,é a precisão. A qualidade necessária e suficiente de um teste é a validade. Mas a validade depende em parte da precisão. Se através de um grande número de mensurações de um mesmo segmento,sempre com a mesma escala obtivermos resultados os mais diversos possíveis,apesar de manter as mesmas condições de mensuração,não podemos confiar em nosso instrumento. Nossos resultados não serão válidos. Logo a alta precisão é qualidade indispensável para existência de alta validade. É necessário então,ao elaborarmos um teste,ter algo que indique sua precisão. A primeira idéia que surge é medir precisão por meio de correlação entre os resultados obtidos através de suas aplicações do teste ao mesmo grupo de pessoas,em ocasiões diferentes. Quanto mais preciso o teste,maior aproximação haverá entre as notas referentes a uma e outra aplicação e,portanto,maior será o coeficiente de correlação. Contudo no campo dos testes,nos deparamos com mais uma dificuldade que,em geral,não existe em problemas físicos,químicos,etc.. Não podemos aplicar um dado teste,repetidas vezes a uma mesma pessoa, ou a um mesmo grupo de pessoas,pelo fato de que as aplicações não serem independentes. Ao aplicarmos um teste pela segunda vez a um grupo de indivíduos,não podemos garantir que suas condições sejam idênticas às existentes quando o aplicamos pela primeira vez. Uma solução seria considerar formas paralelas de um mesmo teste. Dizemos,intuitivamente,que estamos diante de formas paralelas de um teste,se para nossos objetivos for indiferente aplicar uma ou outra. Estatisticamente falando, dizemos que estamos diante de formas paralelas,se as médias e as variâncias obtidas através da aplicação de cada uma forem respectivamente iguais,considerando o mesmo grupo de indivíduos. O inconveniente de formas paralelas aplicadas em ocasiões diferentes ao mesmo grupo de indivíduos é a experiência adquirida durante a primeira aplicação. Um modo mais eficiente de abordar esta questão é considerar partes equivalentes de um mesmo teste. Assim,em um teste cujos ítens estão dispostos em ordem de dificuldade crescente podemos considerar,por exemplo, ítens pares constituindo uma forma e,ímpares outra. Deste modo as duas formas poderiam ser apresentadas concomitantemente, evitando a influência da experiência adquirida na primeira aplicação,ou através do tempo decorrente entre ambas.

Vimos que precisão é um requisito necessário para a validade de um teste,mas,não é suficiente. Assim,se duas formas paralelas de um teste forem constituídas por questões tão difíceis que nenhum indivíduo da amostra consiguirá resolvê-las, ou tão fáceis que todos as acertarão,a precisão será perfeita,porém,o teste não será válido. A dificuldade dos ítens tem grande influência sobre a validade e deve ser considerada com rigor. A primeira idéia que surge é associar a dificuldade do ítem à proporção de acertos, ou seja,se N indivíduos tentarem responder o ítem,e n acertarem,podemos definir a dificuldade do ítem como n/N. Outra definição será possível se conhecermos a forma da distribuição da habilidade em estudo. Assim,sendo F(X) sua função de distribuição e, P_0 a porporção de respostas certas do ítem em questão, sua dificuldade será solução de:

$$\int_{-\infty}^{X} d\,F(X) = P_0$$

É possível ainda,como veremos adiante,definir dificuldade em outros termos, sob outros pontos de vista.

O estudo da eficiência de cada ítem ao elaborarmos um teste, é indispensável, ítens sem qualquer valor discriminativo devem ser rejeitados. Assim como falamos em validade do teste, podemos falar em <u>validade do ítem</u>. Há uma variedade de meios que nos levam ao seu cálculo, dependendo das condições em que nos encontramos. Podemos, conforme o caso, lançar mão do coeficiente de correlação bi-serial; coeficiente de correlação tetracórica; coeficiente *phi*, etc..

Sob todas as formas tem-se procurado aperfeiçoar os testes psicológicos, de forma a obter maior aproximação das escalas elaboradas pelas ciências mais evoluídas.

Um dos campos da psicologia onde procurou-se um maior rigor nas mensurações, tentanto exprimir a relação entre estímulo físico e sensação através de equação matemática, é a psicofísica. Ferguson (12), lançou mão do Método Constante, empregado em psicofísica para o estudo de <u>dificuldades do ítem</u> do teste. Finney (5) conhecedor do trabalho de Ferguson resolveu o mesmo problema através da técnica *Probit Analysis*, dizendo que poder-se-ia fazer extensão da aplicação desta técnica a outros problemas referentes aos ítens. Mais recentemente encontramos, como veremos adiante, um artigo de Rhea S. Das (4), onde algumas considerações são feitas com relação a sua aplicação <u>à validade do ítem</u>, ítens paralelos, etc..

Este trabalho tem por finalidade apresentar a técnica *Probit Analysis* e a seguir a sua aplicação no campo dos testes psicológicos, educacionais.

EXPOSIÇÃO DA TÉCNICA "PROBIT ANALYSIS"

Introdução

Suponhamos que um certo estímulo E, apresentado a uma população de indivíduos, admita apenas duas respostas alternativas E_1, E_2, mutuamente exclusivas, dependendo:

a) da intensidade do estímulo aplicado.

b) da intensidade do atributo do indivíduo responsável pela resposta ao estímulo, a qual designaremos por α. Representando por X a intensidade do estímulo, vamos supor que para $X = X_0$, todos os indivíduos cuja intensidade de atributo seja inferior a um certo nível α_0 darão resposta E_1, e os restantes darão resposta E_2. Fazendo X assumir outros valores possíveis, vamos, em correspondência, obter outros valores para α. A proporção de indivíduos cuja resposta é E_1 será portanto, função de X. Isto nos induz a avaliar a intensidade do atributo de um indivíduo através da intensidade do estímulo. Definimos então, grau de intensidade do atributo α do indivíduo I como a intensidade X_1 do estímulo necessária para produzir a resposta desejada, que indicaremos por E_1. Será uma intensidade de estímulo tal que, qualquer que seja a intensidade X apresentada ao indivíduo se $X > X_1$ ocorrerá E_1 e, para $X < X_1$ não ocorrerá a resposta E_1. |(Ou, segundo nossos objetivos, podemos dizer que se $X > X_1$ a resposta não ocorrerá, mas, ocorrerá sempre para $X < X_1$).|

Podemos, então, pensar na função de distribuição do atributo sobre a população dada, baseados na definição acima. Seja então, a função de distribuição do atributo em questão:

$$F(X_0) = \int_{-\infty}^{X_0} f(X)\, dX = P(-\infty < X < X_0)$$

ou seja, é a proporção de indivíduos cuja intensidade do atributo é inferior a uma certa intensidade dada. O fato de considerarmos menos infinito como limite inferior de integração, sendo o valor mínimo de X finito, não altera a probabilidade de ocorrência da resposta E_1, para um valor qualquer X_0 de X, pois neste caso

$$\int_{-\infty}^{\min X} f(X)\, dX = 0 \qquad \text{e portanto:}$$

$$\int_{-\infty}^{X_0} f(X)\,dX = \int_{-\infty}^{\min X} f(X)\,dX + \int_{\min X}^{X_0} f(X)\,dX = \int_{\min X}^{X_0} f(X)\,dX = F(X_0)$$

O estudo de tóxicos, em biologia, serve para exemplificar a situação acima. O estímulo será o tóxico, e as possíveis respostas serão morte ou sobrevivência do inseto. O inseto possue uma certa tolerância com relação ao tóxico, e portanto, sua morte dependerá da intensidade do tóxico apresentada, intensidade que poderá ser expressa em termos de quantidade de tóxico, unidades de concentração, etc.. Podemos, então, definir a tolerância do inseto, como a dose de tóxico necessária para matá-lo.

A resposta dos indivíduos aos ítens de um teste constitue uma situação análoga. Os indivíduos são portadores de uma certa habilidade Y para responder a um dado ítem. O fato de um indivíduo com certa habilidade acertar o ítem (supondo que o acerto não seja casual) vai depender do grau de dificuladade deste ítem. Podemos, portanto dizer que o nível de habilidade Y do indivíduo é o máximo grau de dificuldade dos ítens que foi possível acertar.

Podemos então, falar em distribuição de tolerância dos insetos, distribuição de habilidade Y dos indivíduos.

Suponhamos que estamos interessados em resolver um dos dois seguintes tipos de problemas:

1- Estudar o estímulo, compará-lo a outros, a fim de verificar sua eficiência, baseados nos resultados que os estímulos produzem sobre os indivíduos.

2- Estudar os indivíduos, classificá-los, baseados no estímulo a eles aplicado.

No primeiro caso estaríamos, por exemplo, diante da verificação da eficiência de um tóxico para matar insetos, ou da eficiência de um teste para classificar indivíduos. No segundo caso sabemos que nossos estímulos preenchem suas finalidades, o que desejamos agora é, baseados neles, estudar o indivíduo. Assim, por exemplo, sabendo que um teste classifica realmente os indivíduos, desejamos aplicá-lo a uma certa pessoa a fim de ver o nível de habilidade que possue.

Suponhamos que desejamos comparar dois estímulos, ou seja, verificar qual dos dois possue maior eficácia. Poderemos verificar isto, por exemplo, através das proporções de mortos devidas a um e a outro mediante a aplicação de iguais doses, ou podemos verificar para ambos quais as doses necessárias para ocasionarem respostas iguais. Se quizermos avaliar a situação de uma criança com relação à dada habilidade medida pelo teste X, podemos nos basear na proporção de crianças nas mesmas condições, cujas intensidades de habilidade são inferiores à da criança em questão. Para os dois casos temos a necessidade de conhecer completamente a distribuição de frequência da variável considerada a fim de calcularmos as proporções. É claro que se um tóxico já foi estuado e adotado, ou se um teste já foi elaborado e estandardizado, não haverá problema quanto ao conhecimento completo da distribuição de frequência da variável. Seria o caso do último problema proposto. Contudo, podemos estar diante da situação (1), ou seja, podemos estar na fase de elaboração de um teste, ou preparação de um tóxico, e portanto não temos conhecimento completo, ou mesmo nenhum conhecimento, das distribuições de frequências a eles associadas. Assim, dependendo do fato de o teste ser muito fácil, ou muito difícil para a população ao qual se destina, vamos ter uma média muito elevada ou muito baixa. O mesmo para o tóxico, dependendo de ser muito poderoso ou não, vamos ter necessidade de pequena dose ou de elvada dose para matar, por exemplo 50% dos insetos. Além do problema dos parâmetros desconhecidos, temos também que pensar na forma de distribuição de frequência.

Nosso principal objetivo, aqui, é o estudo dos testes no campo educacional. A maioria de trabalhos já apresentados neste setor baseia-se sempre na suposição de que a habilidade em questão possue distribuição normal, e em geral, os resultados são compatíveis com a realidade, confirmando assim a hipótese feita. Além disso muitas vezes, embora a distribuição da variável não seja normal, é possível encontrar uma transformação desta variável, que leva a uma distribuição normal. Assim, em biologia, em estudos sobre inseticidas, mesmo que a distribuição de tolerância, baseada por exemplo na concentração de tóxico, seja assimétrica, pode-se chegar a uma distribuição normal, considerando-se em lugar da concentração da dose, o logaritmo da concentração, ou ainda a dose elevada a uma certa potência menor que a unidade. Isto é, sendo X a concentração podemos considerar: $x = \log_{10} X$, ou $x = X^k$, $k < 1$. Vamos então, considerar daqui em diante, a forma da distribuição da variável como sendo normal.

Um dos campos ao qual se dedica a biologia são os ensáios, preparação de drogas, tóxicos, inseticidas, etc.. Surgem então, constantemente, questões do tipo (1) acima mencionado. Para por à prova a eficiência de um inseticida, por exemplo, é preciso fazer observações e ver os resultados. Sendo impossível lidar com toda a população de insetos aos quais o inseticida se destina, o estuod é baseado em amostras. Poderemos então, selecionar casualmente grupos de insetos e, aplicar a cada um, uma certa dose de inseticida em questão. A cada dose associaremos, então a proporção de insetos mortos do grupo ao qual foi aplicado. Nosso objetivo é caracterizar o inseticida, é associar-lhe algo que o caracterize em relação aos demais, algo que sirva de termo de comparação. O ideal seria, por exemplo, verificar qual foi a dose necessária para matar 100% dos insetos de um grupo, e, compará-lo às doses de outros inseticidas nas mesmas condições. Contudo, tal técnica é impossível, pelo simples fato de estarmos lidando com amostras. Sempre há a possibilidade de existência de um inseto cuja tolerância seja superior a de qualquer inseto das amostras escolhidas. Logo, o ma conveniente será a comparação das doses cuja proporção é 50%, pois as variações casuais das proporções correspondentes aos valores centrais, não alterarão tanto as doses quanto as variações das proporções correspondentes aos valores extremos.

Se supusermos, como anteriormente mencionamos, a população normal distribuída com relação às tolerâncias, ou se necessário, com relação ao logaritmo das tolerâncias, a dose que corresponderá a 50% será a solução de:

$$\int_{-\infty}^{X} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dX = 0,5$$

onde μ e σ^2 são respectivamente, tolerância média e variância populacional.

Então, baseados nos dados da observação devemos esti- μ e σ^2 a fim de solucionar nosso problema. É aqui que surge a técnica do Probit Analysis. Havendo semelhança de situações entre o estudo de ítens de um teste em psicologia, e o dos inseticidas em biologia, Finney sugeriu que se estendesse a aplicação desta técnica ao campo educacional, (5). Tentaremos expor o que foi feito neste sentido. Antes, porém, veremos a justificativa teórica do que será feito.

Considerações Teóricas

a) Princípio de Máxima Verossimilhança:

Seja $f(x, \theta)$ a função de frequência de uma população, sendo θ parâmetro desconhecido. Nosso problema, então, é estimar θ. Para isso selecionamos uma amostra A: $x_1, x_2, \ldots, x_n$, da população em questão, e baseados nestes elementos é que vamos chegar ao nosso objetivo.

O parâmetro θ é certa função dos elementos da população, ou seja, $\theta = h(X_1, X_2, \ldots)$. Embora a primeira idéia seja considerada a restrição desta função aos elementos da amostra, é possível que existam outras funções destes elementos que estimarão tão bem, ou melhor ainda o parâmetro em questão. A função dos valores amostrais por meio da qual estimamos θ é chamada estimador. Nosso problema é escolher o melhor estimador de θ. Pelo fato de não conhecermos θ, podemos considerá-lo como uma variável.

Seja $f(x_1, x_2, \ldots, x_n, \theta)$ a função de probabilidade para uma amostra aleatória de tamanho n. Esta função depende do parâmetro desconhecido θ. Vamos então adotar como estimador de θ a função $\hat{\theta}_0$, tal que $f(x_1, x_2, \ldots, x_n, \theta) < f(x_1, x_2, \ldots, x_n, \hat{\theta}_0)$ onde $\hat{\theta}$ é qualquer outro estimador de θ. Ou seja, sendo $L = f(x_1, x_2, \ldots, x_n, \theta)$ vamos escolher como estimador de θ, a função $g(x_1, x_2, \ldots, x_n)$ que torna máxima L. Este é o princípio da máxima verossimilhança, e o estimador obtido é chamado de máxima verossimilhança. Este estimador deverá ser pois, solução de:

$$\frac{\delta L}{\delta \theta} = 0 \quad , \quad \frac{\delta^2 L}{\delta \theta^2} < 0$$

como estamos considerando função de probabilidade $L > 0$, e podemos então, usar a seguinte forma equivalente à anterior:

$$\frac{1}{L}\,\frac{\delta L}{\delta \theta} = \frac{\delta}{\delta \theta} \log L = 0$$

Os estimadores assim obtidos gozam de certas propriedades importantes ou seja, tendem à normalidade à medida que aumenta o tamanho da amostra; são coerentes, isto é, $\lim_{n\to\infty} P(|\theta - \hat{\theta}_0| < \varepsilon) = 1$, onde $\varepsilon > 0$, arbitrário; são eficientes isto é, sua variância para $n \to \infty$ é menor que a de outro qualquer estimador; havendo possibilidade de estimadores suficientes para θ, o estimador de máxima verossimilhança será um deles, ou seja, a distribuição condicionada $(X_1, X_2, \ldots, X_n)$, dado $\hat{\theta}$ não depender de θ.

De um modo geral, podemos estar interessados em estimar mais de um parâmetro populacional. Assim para uma população cuja função de probabilidade é $f(x; \theta_1, \theta_2, \ldots, \theta_k)$ podemos obter os estimadores de máxima verossimilhança $(\hat{\theta}_{01}, \hat{\theta}_{02}, \ldots, \hat{\theta}_{0k})$. Para n tendendo ao infinito a distribuição de $(\hat{\theta}_{o1}, \hat{\theta}_{o2}, \ldots, \hat{\theta}_{ok})$ tende à distribuição normal multidimencional, e as variâncias e covariâncias destes estimadores são dadas pelos elementos da matriz inversa de uma matriz cujo elemento de posição (ij) é dado por:

$$\sigma_{ij} = -E\left| \frac{\delta^2}{\delta\theta_i \delta\theta_j} \log f(x; \theta_1, \theta_2, \ldots, \theta_k) \right|$$

$$i,j = 1, 2, \ldots, k.$$

b) Técnica Probit Analysis

Seja que estamos interessados no estudo de um determinado fator P, dentre outros, destinado a influir de certa forma sobre uma população caracterizada por um dado atributo. Suponhamos que este fator agindo sobre a população possa levantar uma de duas respostas E_1 ou E_2. Vamos, como já mencionamos, medir a intensidade do atributo, em cada indivíduo, pela intensidade X do fator P necessária para produzir uma das respostas, digamos E_1. (Intensidades menores que a considerada não produzirão a resposta e, maiores sempre produzirão). Supondo então, X normalmente distribuída, nosso objetivo é estimar sua média, μ, e sua variância σ^2. Para sito, vamos lançar mão da variável:

$$Z = \frac{X - \mu}{\sigma} + 5$$

À variável Z assim definida foi dado, por Bliss, o nome de probit. O probit de uma probabilidade P é definido como o valor de abcissa que corresponde a P numa distribuição normal com média 5 e variância unitária.

Estimando a reta que exprime a relação entre Z e X, chegamos às estimativas de μ e σ.

Sejam $\alpha = \dfrac{5\sigma - \mu}{\sigma}$

$$\beta = \frac{1}{\sigma} \qquad (1)$$

Podemos então, exprimir a reta a ser estimada por $Z = \beta X + \alpha$.

Colhamos então, m amostras casuais independentes, A_1, $A_2, \ldots, A_m$, da população em questão, com respectivamente n_1, n_2, $\ldots$, n_m elementos. A cada amostra apliquemos uma certa intensidade do fator F que será designada por X_i, $i = 1, 2, \ldots, m$. Nosso primeiro passo é estimar as probabilidades populacionais P_i, associadas a X_i, $i = 1, 2, \ldots, m$, para a seguir chegarmos aos Z_i. A probabilidade de r_i indivíduos da amostra A_i responderem E_i, mediante a dose X_i é dada por:

$$P_i(r_i) = \begin{pmatrix} n_i \\ r_i \end{pmatrix} P_i^{r_i}(1-P_i)^{n_i-r_i} = \begin{pmatrix} n_i \\ r_i \end{pmatrix} P_i^{r_i} Q_i^{n_i-r_i}$$

pois, estamos dentro dos moldes da distribuição binomial. Considerando, agora as m amostras independentes, a probabilidade de obtermos $r_1, r_2, \ldots, r_m$ respostas favoráveis, para $A_1, A_2, \ldots, A_m$ respectivamente, sera dada por:

$$L = \prod_{i=1}^{m} P_i(r_i) = \prod_{i=1}^{m} \begin{pmatrix} n_i \\ r_i \end{pmatrix} P_i^{r_i} Q_i^{n_i-r_i}$$

pois, as amostras são independentes.

Usemos para estimar α e β o princípio de máxima verossimilhança. Vamos então, estimar α e β de sorte que sejam simultaneamente satisfeitas.

$$(S_1) \quad \begin{cases} \dfrac{\delta L}{\delta \alpha} = 0 \\[2mm] \dfrac{\delta L}{\delta \beta} = 0 \end{cases}$$

Vamos obter então:

$$(S_1') \quad \begin{cases} \dfrac{\delta L}{\delta \alpha} = \dfrac{\delta}{\delta \alpha} \left[\sum_{i=1}^{m} r_i \log P_i + \sum_{i=1}^{m} (n_i - r_i) \log (1-P_i) \right] = 0 \\[2mm] \dfrac{\delta L}{\delta \beta} = \dfrac{\delta}{\delta \beta} \left[\sum_{i=1}^{m} r_i \log P_i + \sum_{i=1}^{m} (n_i - r_i) \log (1-P_i) \right] = 0 \end{cases}$$

que simplificando torna-se

$$(S_2) \quad \begin{cases} \dfrac{\delta L}{\delta \alpha} = \sum_{i=1}^{m} \dfrac{n_i (p_i - P_i)}{P_i Q_i} \cdot \dfrac{\delta P_i}{\delta \alpha} = 0 \\[2mm] \dfrac{\delta L}{\delta \beta} = \sum_{i=1}^{n} \dfrac{n_i (p_i - P_i)}{P_i Q_i} \cdot \dfrac{\delta P_i}{\delta \beta} = 0 \quad \text{onde } p_i = \dfrac{r_i}{n_i} \end{cases}$$

Sendo praticamente impossível a resolução direta do sistema (S_2), vamos lançar mão, de acordo com Finney e outros, do processo iteração. Nossas funções, no caso, são $\frac{\delta L}{\delta \alpha}$ e $\frac{\delta L}{\delta \beta}$. Considerando apenas as primeiras derivadas, nossas funções serão aproximadas por:

$$\frac{\delta L}{\delta\alpha} + h\,\frac{\delta^2 L}{\delta\alpha^2} + k\,\frac{\delta^2 L}{\delta\alpha\delta\beta} \quad e \quad \frac{\delta L}{\delta\beta} + h\,\frac{\delta^2 L}{\delta\alpha\delta\beta} + k\,\frac{\delta^2 L}{\delta\beta^2}$$

E o sistema (S_2) será aproximado por:

$$(S_3)\quad \begin{cases} \dfrac{\delta L}{\delta\alpha} + h\,\dfrac{\delta^2 L}{\delta\alpha^2} + k\,\dfrac{\delta^2 L}{\delta\alpha\delta\beta} = 0 \\[2ex] \dfrac{\delta L}{\delta\beta} + h\,\dfrac{\delta^2 L}{\delta\alpha\delta\beta} + k\,\dfrac{\delta^2 L}{\delta\beta^2} = 0 \end{cases}$$

Em particular, para cada derivada obteremos:

$$\frac{\delta L}{\delta\alpha} = \sum_{i=1}^{m} \frac{n_i(p_i - P_i)}{P_i\,Q_i} \cdot \frac{\delta P_i}{\delta\alpha} \qquad e,$$

$$\frac{\delta L}{\delta\beta} = \sum_{i=1}^{m} \frac{n_i(p_i - P_i)}{P_i\,Q_i} \cdot \frac{\delta P_i}{\delta\beta}$$

$$\frac{\delta^2 L}{\delta\alpha^2} = \sum_{i=1}^{m} \frac{n_i\left[(-P_iQ_i) - (1 - 2P_i)(p_i - P_i)\right]}{P^2_{\ i}Q^2_{\ i}} \left(\frac{\delta P_i}{\delta\alpha}\right)^2 + \sum_{i=1}^{m} \frac{n_i(p_i - P_i)}{P_i\,Q_i} \cdot \frac{\delta^2 P_i}{\delta\alpha^2}$$

$$\frac{\delta^2 L}{\delta\alpha\delta\beta} = \sum_{i=1}^{m} \frac{n_i\left[(-P_iQ_i) - (1 - 2P_i)(p_i - P_i)\right]}{P^2_{\ i}Q^2_{\ i}} \left(\frac{\delta P_i}{\delta\alpha}\right) \cdot \left(\frac{\delta P_i}{\delta\beta}\right) + \sum_{i=1}^{m} \frac{n_i(p_i - P_i)}{P_i\,Q_i} \cdot \frac{\delta}{\delta\beta}\left(\frac{\delta P_i}{\delta\alpha}\right)$$

Se substituirmos as derivadas de segunda ordem por suas esperanças temos:

$$\frac{\delta^2 L}{\delta\alpha^2} = \sum_{i=1}^{m} \frac{-n_i}{P_iQ_i} \left(\frac{\delta P_i}{\delta\alpha}\right)^2$$

$$\frac{\delta^2 L}{\delta\beta^2} = \sum_{i=1}^{m} \frac{-n_i}{P_iQ_i} \left(\frac{\delta P_i}{\delta\beta}\right)^2$$

$$\frac{\delta^2 L}{\delta\alpha\delta\beta} = \sum_{i=1}^{m} \frac{-n_i}{P_i Q_i} \left[\frac{\delta P_i}{\delta\alpha}\right]\left[\frac{\delta P_i}{\delta\beta}\right]$$

Considerando (S_3), os valores das derivadas acima, e como primeira aproximação $Z_1 = \alpha_1 + \beta_1 X$, podemos escrever:

$$(S_4) \begin{cases} h \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} \left[\frac{\delta P_{1i}}{\delta\alpha_1}\right]^2 + k \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} \left[\frac{\delta P_{1i}}{\delta\alpha_1}\right]\left[\frac{\delta P_{1i}}{\delta\beta_1}\right] = \\ \qquad = \sum_{i=1}^{m} \frac{n_i(p_i - P_{1i})}{P_{1i}Q_{1i}} \frac{\delta P_{1i}}{\delta\alpha_1} \\ h \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} \left[\frac{\delta P_{1i}}{\delta\alpha_1}\right]\left[\frac{\delta P_{1i}}{\delta\beta_1}\right] + k \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} \left[\frac{\delta P_{1i}}{\delta\beta_1}\right]^2 = \\ \qquad = \sum_{i=1}^{m} \frac{n_i(p_i - P_{1i})}{P_{1i}Q_{1i}} \frac{\delta P_{1i}}{\delta\beta_1} \end{cases}$$

A resolução do sistema (S_4) nos fornecerá os valores de h e k que serão os ajustamentos que devemos fazer a α_1 e β_1 a fim de obter melhor aproximação de α e β. Obteremos então:

$$\alpha_2 = \alpha_1 + h \quad e \quad \beta_2 = \beta_1 + k$$

Podemos, se achar necessário, repetir o processo obtendo correções para α_2 e β_2, e assim sucessivamente.

De acordo com nossas hipóteses, temos:

$$P_{1i} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_{1i}} e^{-\frac{1}{2}(Z_{1i}-5)^2}\, dZ, \text{ pois } Z \text{ é } N(5, 1).$$

A reta estimada por simples inspeção é $Z_1 = \alpha_1 + \beta_1 X$. Logo:

$$\frac{\delta P_i}{\delta\alpha} = \frac{\delta P_i}{\delta Z_i} \cdot \frac{\delta Z_i}{\delta\alpha}\bigg|_{\alpha=\alpha_1} = \frac{1}{\sqrt{2\pi}} e^{\frac{-(Z_{1i}-5)^2}{2}} = Y_{1i}$$

$$\frac{\delta P_i}{\delta\beta} = \frac{\delta P_i}{\delta Z_i} \cdot \frac{\delta Z_i}{\delta\beta}\bigg|_{\beta=\beta_1} = X_i Y_{1i}$$

Logo (S_4) poderá ser escrito:

$$(S_5)\begin{cases} h \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} Y_{1i}^2 + k \sum_{i=1}^{m} \frac{n_i X_i Y_{1i}^2}{P_{1i}Q_{1i}} = \sum_{i=1}^{m} \frac{n_i(p_i - P_{1i})Y_{1i}}{P_{1i}Q_{1i}} \\ h \sum_{i=1}^{m} \frac{n_i}{P_{1i}Q_{1i}} Y_{1i}X_i + \sum_{i=1}^{m} \frac{X_i^2 Y_{1i}^2 n_i}{P_{1i}\ Q_{1i}} = \sum_{i=1}^{m} \frac{n_i(p_i - P_{1i})}{P_{1i}\ Q_{1i}} X_i Y_{1i} \end{cases}$$

A fim de introduzirmos simplificações, chamaremos, de acordo com Finney:

$$W_i = \frac{Y_{1i}^2}{P_{1i}\ Q_{1i}} \quad , \qquad i = 1, 2, \ldots, m.$$

Nosso sistema (S_5) passará a:

$$(S_6)\begin{cases} h \sum_{i=1}^{m} n_i W_i + k \sum_{i=1}^{m} n_i W_i X_i = \sum_{i=1}^{m} n_i W_i \frac{p_i - P_{1i}}{Y_{1i}} \\ h \sum_{i=1}^{m} n_i W_i X_i + k \sum_{i=1}^{m} n_i W_i X_i^2 = \sum_{i=1}^{m} n_i W_i X_i \frac{p_i - P_{1i}}{Y_{1i}} \end{cases}$$

Como vemos representam as equações de regressão ponderada de:

$$\frac{p_i - P_{1i}}{Y_{1i}} \text{ sobre } X_i.$$

Com o fim de facilitar os cálculos definamos probit operacional pe la relação:

$$Z_i' = Z_{1i} + \frac{p_i - P_{1i}}{Y_{1i}}$$

Somando $\sum_{i=1}^{m} n_i W_i Z_{1i}$ a ambos os membros da 1ª equação e $\sum_{i=1}^{m} n_i W_i X_i Z_{1i}$, a ambos os membros da segunda, teremos:

$$(S_7)\begin{cases} h \sum_{i=1}^{m} n_i W_i + k \sum_{i=1}^{m} n_i W_i X_i + \sum_{i=1}^{m} n_i W_i Z_{1i} = \sum_{i=1}^{m} n_i W_i Z_i' \\ h \sum_{i=1}^{m} n_i W_i X_i + k \sum_{i=1}^{m} n_i W_i X_i^2 + \sum_{i=1}^{m} n_i W_i X_i Z_{1i} = \sum_{i=1}^{m} n_i W_i X_i Z_i' \end{cases}$$

Dividindo ambos os membros das duas equações por

$$\sum_{i=1}^{m} n_i W_i$$

teremos:

$$\begin{cases} h + k\bar{X} + \bar{Z}_1 = \bar{Z}'. \\[2ex] h\bar{X} + k\dfrac{\sum_{i=1}^{m} n_i W_i X_i^2}{\sum_{i=1}^{m} n_i W_i} + \dfrac{\sum_{i=1}^{m} n_i W_i Z_{1i}}{\sum_{i=1}^{m} n_i W_i} = \dfrac{\sum_{i=1}^{m} n_i W_i X_i Z_i'}{\sum_{i=1}^{m} n_i W_i} \end{cases}$$

onde $\bar{X}$, $\bar{Z}_1$ e $\bar{Z}'$ são respectivamente as médias ponderadas de X_i, Z_{1i} e Z_i'.

Substituindo na segunda equação: h por $\bar{Z}' - k\bar{X} - \bar{Z}_1$;

$\bar{Z}_1$ por $\beta_1\bar{X} + \alpha_1$;

Z_{1i} por $\beta_1 X_i + \alpha_1$, teremos:

$$\bar{X}\bar{Z}' - k\bar{X}^2 - \beta_1\bar{X}^2 + k\frac{\sum_{i=1}^{m} n_i W_i X_i^2}{\sum_{i=1}^{m} n_i W_i} + \beta_1\frac{\sum_{i=1}^{m} n_i W_i X_i^2}{\sum_{n=1}^{m} n_i W_i} = \frac{\sum_{i=1}^{m} n_i W_i Z_i' X_i}{\sum_{i=1}^{m} n_i W_i}$$

indicando: $\dfrac{\sum_{i=1}^{m} n_i W_i Z_i' X_i}{\sum_{i=1}^{m} n_i W_i} - \bar{X}\bar{Z}'$ por $S_{xz'}$, e $\dfrac{\sum_{i=1}^{m} n_i W_i X_i^2}{\sum_{i=1}^{m} n_i W_i} - \bar{X}^2$ por S_x^2

vamos ter: $kS_x^2 + \beta_1 S_x^2 = S_{xz'}$, e portanto:

$$(k + \beta_1) = \frac{S_{xz'}}{S_x^2} \quad \text{e} \quad (h + \alpha_1) = \bar{Z}' - (k + \beta_1)\bar{X}.$$

Vamos pois, ter como melhor aproximação de Z:

$$Z_2 = (k + \beta_1)X + (h + \alpha_1) = \bar{Z}' + \frac{S_{xz'}}{S_x^2}(X - \bar{X}).$$

O processo poderá ser reptedido a fim de obter, se necessário, melhor aproximação. Suponhamos que Z_2 já seja uma boa aproximação de Z. Então, nossa reta teórica $Z = \alpha + \beta X$ será estimada por $Z_2 = \beta_2 X + \alpha_2$, onde $\beta_2 = k + \beta_1$ e $\alpha_2 = h + \alpha_1$. Lembrando que $\sigma = \frac{1}{\beta}$ e $\mu = \sigma(5 - Z)$ estimaremos $\sigma^2 = \frac{1}{\beta_2^2}$ e μ pelo valor de X correspondente à $Z_2 = 5$.

Como vimos foi usado o processo de máxima verossimilhança para a estimação, logo as variâncias e covariâncias serão os elementos da matriz:

$$\begin{bmatrix} \sum\limits_{i=1}^{m} n_iW_i & \sum\limits_{i=1}^{m} n_iW_iX_i \\ \sum\limits_{i=1}^{m} n_iW_iX_i & \sum\limits_{i=1}^{m} n_iW_iX_i^2 \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{\sum\limits_{i=1}^{m} n_iW_iX_i^2}{SS_x^2 \sum\limits_{i=1}^{m} n_iW_i} & \dfrac{-\bar{\bar{X}}}{SS_x^2} \\ \dfrac{-\bar{\bar{X}}}{SS_x^2} & \dfrac{1}{SS_x^2} \end{bmatrix}$$

onde, $SS_x^2 = \left[\sum\limits_{i=1}^{m} n_iW_i\right] S_x^2$

Para a variância de $\hat{Z}$, estimador de Z, basta lembrar que $\hat{Z} = \hat{\beta}X + \hat{\alpha}$, onde $\hat{\beta}$ e $\hat{\alpha}$ são estimadores de α e β. Logo, para qualquer valor de X considerado, vamos ter:

$$\sigma_{\hat{Z}}^2 = X^2\sigma_{\hat{\beta}}^2 + \sigma_{\hat{\alpha}}^2 + 2X\sigma_{\hat{\beta}\hat{\alpha}} = \frac{1}{\sum\limits_{i=1}^{m} n_iW_i} + \frac{(X - \bar{\bar{X}})^2}{SS_x^2}$$

De $\bar{\hat{Z}} = \hat{\alpha} + \hat{\beta}\bar{X}$, segue que $\sigma_{\bar{\hat{Z}}}^2 = \dfrac{1}{\sum\limits_{i=1}^{m} n_iW_i}$

Através da reta estimada, podemos então, estimar qual a intensidade do fator X que levanta cinquenta por cento das respostas, exatamente, ou seja, a mediana. Para isto devemos encontrar o valor de Z que corresponde a cinquenta por cento. Como Z é variável aleatória, a mediana também será, pois:

$$\hat{m} = \frac{\hat{Z}}{\hat{\beta}} - \frac{\hat{\alpha}}{\hat{\beta}} ;$$

onde $\hat{m}$ representa a mediana estimada, ou melhor o estimador de m. Desta relação segue que:

$$\sigma_{\hat{m}}^2 = \frac{1}{\hat{\beta}^2}\sigma_{\hat{Z}^2}^2 = \frac{1}{\beta^2}\left[\frac{1}{\sum\limits_{i=1}^{m} n_iW_i} + \frac{(\hat{m} - \bar{\bar{X}})^2}{SS_x^2}\right]$$

Uma das nossas hipóteses era a normalidade da variável X. Pode acontecer que tal hipótese não seja verdadeira, de sorte que a relação entre X e Z não será linear. Daí a necessidade de um teste de significância a fim de verificar o acordo entre a reta obtida e os dados observados. Isto pode ser verificado através do calculo de:

$$\chi^2_{m-2} = SS^2_{z'} - \frac{(SS_{xz'})^2}{SS^2_x} \quad \text{(I) , ou}$$

$$\chi^2_{m-2} = \sum_{i=1}^{m} \frac{(r_i - n_i P_i^*)^2}{n_i P_i^* Q_i^*} \quad \text{(II)}$$

(Onde o asterisco indica estimativa do parâmetro), ambos com (m-2) graus de liberdade, pois foram estimados dois parâmetros. Se algumas das frequências esperadas ($n_i P_i^*$) forem inferiores a 5, χ^2 não será aconselhável, pois poderá nos levar a resultados desnorteadores, a menos que combinemos um ou mais grupos de sorte a obter frequências esperadas maiores. Neste caso, a fórmula (I) não poderá ser empregada. Deve ser notado também que a fórmula (I) pode ser usada apenas quando as aproximações obtidas pelo processo iterativo, das soluções a que deveríamos chegar resolvendo o sistema S_1, forem bastante grandes.

Supondo que a relação seja linear, que as amostras e o processo de estimação sejam escolhidos com rigor, as diferenças entre Z^* e Z serão puramente casuais. Logo se o nosso χ^2 observado for menor que χ^2 crítico não rejeitamos a hipótese de linearidade entre os valores de X e Z.

Suponhamos que desejamos comparar os fatores F_1, F_2, ..., F_s, com respeito à variabilidade ou melhor com respeito às variâncias, isto poderá ser feito da seguinte maneira:

Como vimos, a intensidade do atributo foi definida como a intensidade do estímulo, a qual indicamos com X, extritamente necessária para produzir a resposta. Diante dos fatores $F_1, F_2, \ldots, F_s$, todos com a mesma finalidade, vemos que determinada intensidade do atributo receberá valores diferentes segundo o fator considerado. Para cada fator vamos possuir certa média e certa variância. Pode ocorrer que embora as médias variem, as variâncias permaneçam as mesmas. É o que em geral acontece aos tóxicos. A igualdade de variâncias será evidenciada pelo paralelismo das retas estimadas, pelo processo anteriormente descrito. Um de nossos problemas então consiste na verificação deste paralelismo e, consequentemente, na igualdade das variâncias. Aqui, poderemos processar da seguinte maneira: Para cada fator F_j, j = 1, 2, ,,,, s, consideremos os probits obtidos diretamente, através das proporções observadas, correspondentes à resposta E_1, para os diversos valores de X_{ji}, i=1, 2, ..., m_j; j = 1, 2, ..., s. Indiquemos tais probits por Zo. Consideremos, então, no plano (XZ), os pontos (X_{ji}, Zo_{ji}), j = 1, 2, ..., s; i = 1, 2, ..., m_j. Tendo s fatores, vamos ter s conjuntos de pontos. Devemos como vimos anteriormente, traçar arbitrariamente, para cada fator uma reta. Como estamos supondo a igualdade de variâncias vamos traçar retas paralelas. A seguir, cada fator será estudado independentemente segundo o processo já exposto. Chegamos então, aos valores β_1, β_2, ..., β_s, onde

$$\beta_j = \frac{S_{z'_j x_j}}{S^2_{x_j}}, \quad j = 1, 2, \ldots, s.$$

Vamos considerar então:

$$\beta^* = \frac{\sum_{j=1}^{s} (SS_{x_j z'_j})}{\sum_{j=1}^{s} (SS^2_{x_j})}$$

Como coeficiente comum para todas as retas, é necessário agora, testar a hipótese de paralelismo de retas. Ao todo foram estimados s+ +1 parâmetros, a saber: α_1, α_2, ..., α_s, e β. Logo temos:

$\gamma = \sum_{j=1}^{s} m_j - (s+1)$ graus de liberdade. Podemos, então, considerar:

$$\chi^2_\phi = \sum_{j=1}^{s} (SS^2_{z'_j}) - \frac{\sum_{j=1}^{s} (SS_{z'_j x_j})^2}{\sum_{j=1}^{s} (SS^2_{x_j})}$$

com ϕ graus de liberdade. Tal χ^2_ϕ pode ser considerado como uma soma de dois χ^2 que indicamos χ^2_γ e χ^2_δ. Através de χ^2_γ poremos à prova a linearidade das s relações. Como os fatores são independentes temos:

$$\chi^2_\phi = \sum_{j=1}^{s} \left[SS^2_{z'_j} - \frac{(SS_{z'_j x_j})^2}{SS^2_{x_j}} \right]$$

com $\gamma = \sum_{j=1}^{s} m_j - 2s$ graus de liberdade. A diferença entre χ^2_ϕ e χ^2_γ será a contribuição da soma de quadrados referentes ao paralelismo.

Então χ^2_δ para o teste de paralelismo será:

$$\chi^2_\delta = \chi^2_\phi - \chi^2_\gamma = \sum_{j=1}^{s} \frac{(SS_{z_j^{\prime}x_j})^2}{SS^2_{x_j}} - \frac{\left[\sum_{j=1}^{s} SS_{z_j^{\prime}x_j}\right]^2}{\sum_{j=1}^{s} SS^2_{x_j}}$$

com $\delta = \phi - \gamma = s-1$ graus de liberdade.

Sendo comprovada a igualdade de variâncias, podemos estar interessados em saber se não diferem também quanto às médias. Vamos considerar aqui apenas, dois fatores, F_1 e F_2. Como as linhas são paralelas, dados Z_i e Z_h quaisquer, vamos ter para F_1 e F_2, respectivamente, X_{1i}, X_{2i}, correspondentes a Z_i, e X_{1h}, X_{2h}, correspondentes a Z_h, satisfazendo:

$$X_{1i} - X_{2i} = X_{1h} - X_{2h}$$

Nosso problema é saber se esta diferença é significativa. Para isso vamos aplicar o teorema de Fieller:

<u>TEOREMA</u>: Sejam $\hat{\theta}$, $\hat{\lambda}$ estimadores não viciados de θ, λ, normalmente distribuídos. Seja S^2 variância dos valores amostrais com f graus de liberdade. Suponhamos que as variâncias e covariâncias de $\hat{\theta}$ e $\hat{\lambda}$ possam ser expressas por $S^2K_{\hat{\theta}}$, $S^2K_{\hat{\lambda}}$ e $S\,K_{\hat{\theta},\hat{\lambda}}$, onde $K_{\hat{\theta}}$, $K_{\hat{\lambda}}$ e $K_{\hat{\theta},\hat{\lambda}}$ são funções apenas dos dados amostrais.

Seja $R = \frac{\hat{\theta}}{\hat{\lambda}}$ estimador de $\rho = \frac{\theta}{\lambda}$

Então, limites de confiança para ρ são obtidos através de

$$R_s, R_I = \left[R - \frac{gK_{\hat{\lambda},\hat{\theta}}}{K_{\hat{\lambda}}} \pm \frac{tS}{\hat{\lambda}} \left\{ K_{\hat{\theta}} - 2RK_{\hat{\lambda},\hat{\theta}} + R^2K_{\hat{\lambda}} - g\left(K_{\hat{\theta}} - \frac{K^2_{\hat{\lambda},\hat{\theta}}}{K_{\hat{\lambda}}}\right) \right\}^{1/2} \right] : (1-g)$$

onde $g = \frac{t^2S^2K}{\hat{\lambda}}$, e t deve ser obtido da distribuição de Student com f graus de liberdade. Se g for negligenciável teremos:

$$R_s, R_I = R \pm \frac{tS}{\hat{\lambda}} \left\{ K_{\hat{\theta}} - 2RK_{\hat{\lambda},\hat{\theta}} + R^2K_{\hat{\lambda}} \right\}^{1/2}$$

Além de estimar a diferença entre as intensidades dos fatores F_1 e F_2, que produzem iguais respostas considerando amostras independentes, queremos também associar a esta diferença de intervalo, digamos I, e uma probabilidade P_c, tal que a probabilidade de I conter a real diferença é P_c. Sejam, então, as retas paralelas estimadas por:

$$Z_1^* = \beta^* X_1 + \alpha_1^*$$
$$Z_2^* = \beta^* X_2 + \alpha_2^*$$

A um dado Z, qualquer, vão corresponder dois valores de abcissa, X_1 e X_2, correspondentes a F_1 e F_2 respectivamente. Temos então:

$$D^* = X_2^* - X_1^* = \frac{\alpha_1^* - \alpha_2^*}{\beta^*} = \bar{X}_1 - \bar{X}_2 - \frac{\bar{Z}_2^* - \bar{Z}_1^*}{\beta^*}$$

Queremos limites de confiança para D. Como $\bar{X}_2 - \bar{X}_1$ é constante, podemos considerar:

$$D^* - (\bar{X}_2 - \bar{X}_1) = \frac{\bar{Z}_1^* - \bar{Z}_2^*}{\beta^*} = d^*,$$ encontrar limites para d e, a seguir, somar o que subtraímos, ou seja $(\bar{X}_2 - \bar{X}_1)$.

Cabe aqui observar que os estimadores obtidos pelo princípio da máxima verossimilhança são assintoticamente não viciados e além disso suas distribuições tendem à normal para $n \to \infty$ o que nos permite aplicar o teorema. Considerando amostras independentes, temos:

$$\sigma^2_{\hat{z}_1 - \hat{z}_2} = \frac{1}{\sum_{i=1}^{m_1} n_{1i} W_{1i}} + \frac{1}{\sum_{i=1}^{m_2} n_{2i} W_{2i}}.$$ Logo:

$$\left[D^* - (\bar{X}_2 - \bar{X}_1) \pm \frac{t}{\beta^*} \left\{ \left(\frac{1}{\sum_{i=1}^{m_1} n_{1i} W_{1i}} + \frac{1}{\sum_{i=1}^{m_2} n_{2i} W_{2i}} \right) (1-g) + \frac{(D - \bar{X}_2 + \bar{X}_1)^2}{\sum_{j=1}^{2} SS^2_{x_j}} \right\}^{\frac{1}{2}} \right] (1-g)$$

sendo $g = \dfrac{t^2}{\beta^* \sum_{j=1}^{2} SS^2_{x_j}}$.

Então os limites superior e inferior para $X_1 - X_2$ serão obtidos somando $X_2 - X_1$ a expressão acima.

Algumas extensões desta teoria são apresentadas por Garwood (*8*) e Maritz (*21*). A técnica foi desenvolvida supondo a distribuição de X normal. Isto contudo, pode não só ocorrer como também não haver transformação de X que leve a esta distribuição. Garwood, então desenvolve a técnica para o caso em que a distribuição é conhecida, digamos f, dependendo de k parâmetros. Maritz considera o caso quando não conhecemos a forma de distribuição. Procurando a relação entre probits e X observamos que a regressão de Z sobre X é não linear. O problema é a determinação desta função. O autor apresenta solução para o caso em que $Z = \alpha_0 + \alpha_1 X + \ldots + \alpha_p X^p$, onde p é inteiro, impar. Tanto Garwood como Maritz seguem a mesma linha apresentada anteriormente. Para ambos os casos é usado o princípio de máxima verossimilhança e o processo iterativo.

Podemos estar interessados na comparação entre o processo dos mínimos quadrados ponderados e, o princípio da máxima verossimilhança com soluções das equações de verossimilhança obtidas pelo processo iterativo, quando resolvemos problemas com base em respostas quantais. J.A.Nelder (*14*) e, R.H.Moore (*26*) e R. K. Zeigler (*26*), analisaram esta questão. A semelhança entre ambos é observada quando comparamos o sistema (S_7) com as equações normais obtidas pelo processo dos mínimos quadrados ponderados: Sendo:

$$Z_i = \beta X_i + \alpha + e_i$$

$$SR = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} (Z_i - \beta X_i - \alpha)^2$$

as equações normais serão obtidas por:

$$\frac{\delta SR}{\delta \beta} = -2 \sum_{i=1}^{m} W_i'(Z_i - \beta X_i - \alpha)^2 \, \frac{\delta(\beta X_i + \alpha)}{\delta \beta} = 0$$

$$\frac{\delta SR}{\delta \alpha} = -2 \sum_{i=1}^{m} W_i'(Z_i - \beta X_i - \alpha)^2 \, \frac{(\beta X_i + \alpha)}{\delta \alpha} = 0$$

e portanto:

$$\hat{\alpha} \sum_{i=1}^{m} W_i' + \hat{\beta} \sum_{i=1}^{m} W_i' X_i = \sum_{i=1}^{m} W_i' Z_i$$

$$\hat{\alpha} \sum_{i=1}^{m} W_i' X_i + \hat{\beta} \sum_{i=1}^{m} W_i' X_i^2 = \sum_{i=1}^{m} W_i' Z_i X_i$$

A cada X_i, $i=1,2,\ldots,m$, corresponde uma distribuição dos Z. Então, com relação aos Z temos m distribuições independentes, cada qual com variância $\sigma^2_{Z_i}$, $i=1,2,\ldots,m$. Em geral, considera-se W_i' como recíproco de $\sigma^2_{Z_i}$. Como vimos:

$$P_0 = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{Z_0} e^{\frac{-(Z-5)^2}{2}} \, dZ \quad \text{e, portanto } \frac{dP}{dZ} = Y.$$

Logo $\sigma^2_{P_i} \sim Y^2\sigma^2_{Z_i}$, onde $\sigma^2_{Z_i} \sim \dfrac{P_iQ_i}{n_iY_i^2}$. Além disso as matrizes das variâncias e covariâncias referentes ao processo dos mínimos quadrados e o princípio da máxima verossimilhança, são idênticas. Esta equivalência de resultados foi possível pelo fato de as primeiras derivadas do princípio de máxima verossimilhança poderem ser escritas como:

$$\frac{\delta L}{\delta \beta} = \sum_{i=1}^{m} \frac{p_i - P_i}{\sigma^2_{P_i}} \frac{\delta P_i}{\delta Z_i} X_i$$

$$\frac{\delta L}{\delta \alpha} = \sum_{i=1}^{m} \frac{p_i - P_i}{\sigma^2_{P_i}} \frac{\delta P_i}{\delta Z_i}$$

sendo $Z = \beta X + \alpha$ e P função de Z. Pelo fato acima Nelder concluiu que sempre que as primeiras derivadas do princípio de máxima verossimilhança podem ser escritas sob a forma acima, sendo Z linear em α e β, e P função de Z, os dois processos são equivalentes, se considerarmos Z' em lugar de Z nas equações normais, e além disso, a solução de máxima verossimilhança convergir.

APLICAÇÕES DA TÉCNICA "PROBIT ANALYSIS" AOS TESTES PSICOLÓGICOS

a) Antecedentes e contribuições feitas neste setor

A técnica *probit analysis* tem suas raízes no campo psicofísico com os trabalhos de Fechner (1860). Os métodos psicofísicos procuraram relações entre os estímulos físicos e as sensações que lhes correspondem com o objetivo de elaborar escalas para suas mensurações. Vários tipos de experimentos surgiram com a finalidade de medir o poder discriminativo dos indivíduos em relação a certos estímulos. Assim, por exemplo, um dos experimentos neste setor consistia em medir a capacidade de uma pessoa em sentir as duas pontas de um compasso aplicadas, a uma dada distância, na região interna do ante-braço. Isso porque, embora duas, a proximidade entre as pontas poderia dar a sensação de uma só. Podemos também estar interessados na discriminação de pesos, comprimentos de segmentos, tonalidades de cores, sons, etc..

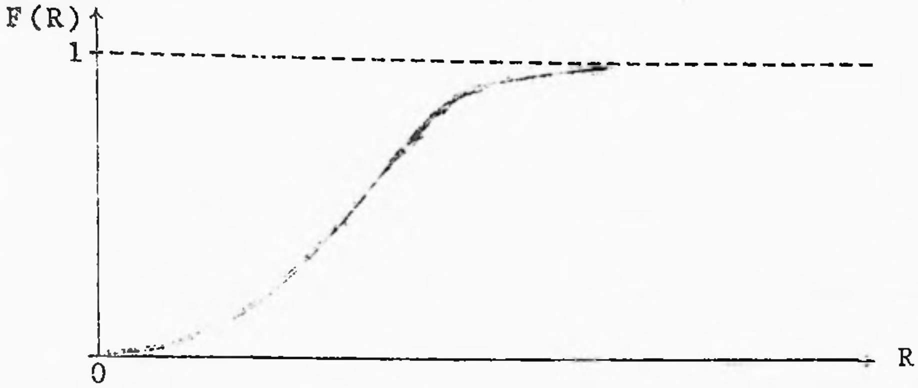
Para tais experimentos o pesquisador organiza séries de intensidades do estímulo, que deverão ser aplicadas a cada pessoa. Assim, para estudar a sensibilidade do indivíduo às duas pontas em relação a distância, poder-se-ia considerar várias distâncias, d_1, d_2, ..., d_n, e aplicar cada uma, um certo número de vezes a cada indivíduo. Todas as precauções devem ser tomadas com relação à ordem das distâncias aplicadas, à influência de sensações anteriores, etc..

Se considerarmos as distâncias entre as duas pontas variando continuamente, desde a mais ínfima possível e em correspondência a sensibilidade do indivíduo em reconhecer a existência de realmente duas, em lugar de uma, vamos ver que, de início, para distâncias muito pequenas, há a impressão de que apenas uma ponta está sendo aplicada. Com o aumento gradativo das distâncias, chegamos a uma região delas para as quais o indivíduo terá a sensação ora de uma, ora de duas pontas, e é claro que poderá sentir um número maior de vezes uma que duas, ou vice-versa, conforme a longitude do intervalo. Haverá uma distância, acima da qual não há praticamente dúvida em reconhecer duas.

O mesmo raciocínio poderá ser feito para o estudo da sensibilidade em reconhecer diferenças entre intensidades de um mesmo estímulo. Poderemos ter um peso padrão, e uma série de pesos variando gradativamente. Pedimos então, para o indivíduo comparar cada peso da série com o padrão a fim de verificar se são iguais ou diferentes. A apresentação poderá variar começando por pesos iguais ao padrão e ir considerando pesos gradativamente maiores, ou gradativamente menores, como aconselhava Weber. Em ambos os casos as séries são ascendentes, uma vez que as diferenças vão se tornando cada vez maiores. Poderemos também começar, segundo Fechner, considerando pesos da série bem discrepantes em relação ao padrão mais leves ou mais pesados, e ir-nos aproximando gradativamente do padrão, até ultrapassá-lo. Neste caso, a ordem será descendente. Os problemas de apresentação, ordenação, devem ser considerados com rigor. Wundt, Nirschmann, por exemplo, começam com estímulos bem diferentes, e o indivíduo sabe se a ordem é descendente ou ascendente.

Considerando os dois exemplos acima, podemos então, compreender o que os psicofísicos chamam de limiar de sensação, e limiar de diferença de sensações, ou seja, haverá um ponto onde o indivíduo começa a sentir a situação real, embora sem muita certeza. Suas respostas não tendem mais ao lado errado, isto é, sensação de uma ponta apenas, ou os estímulos são iguais (no caso de serem diferentes), nem tão pouco vão tender para o lado correto. Definimos então, como *stimulus limen* a intensidade de estímulo que levanta exatamente 50% de respostas certas. Assim, considerando o primeiro exemplo acima, se fizermos a distância variar com continuidade, imaginando infinitas aplicações para cada distância, haverá uma distância tal que o indivíduo sentirá duas pontas em 50% das aplicações. Esta distância variará de indivíduo a indivíduo, e servirá portanto para caracterizá-los quanto à sensibilidade e este estimá-lo. É claro que, em situações experimentais temos que nos satisfazer com um número limitado de distâncias e para cada uma, um número limitado de aplicações. Surge então o problema de estimação de *stimulus limen* e da variabilidade das respostas de um dado indivíduo, através dos dados obtidos.

Uma hipótese em geral feita por Fechner, Urban, Muller e outros, é a seguinte: à medida que a intensidade do estímulo cresce, em unidades constantes, o indivíduo vai mostrando um progresso em suas respostas, mas, não no mesmo ritmo. As proporções de respostas corretas vão sofrendo aumentos cada vez maiores ao irmos nos aproximando do limen, para depois sofrerem aumentos de ordem decrescente ao nos afastarmos dele. Então, por hipótese, a curva representando a habilidade da pessoa vai ser a curva correpondente à distribuição normal. Sejam α e σ^2 a média e a variância da distribuição de certa habilidade, de um dado indivíduo. Segundo nossa hipótese, teremos então:



$$F(R_0) = \int_{-\infty}^{R_0} \frac{1}{\sigma\sqrt{2\pi}} \, \frac{(R_i-\alpha)^2}{2\sigma^2} \, dR = P_0 \qquad (I)$$

É a probabilidade de respostas certas a uma dada intensidade do estímulo. Em nossa anotação, de acordo com os psicofísicos, R_i representa intensidade de estímulo. Em psicofísica costuma-se fazer uma modificação na apresentação da curva:

Em lugar de $(R_i-\alpha)$ consideramos (R_i-L) onde L é o *stimulus limen* e substituimos σ^2 por $1/2h^2$. Como nossa hipótese é de normalidade, temos $L=\alpha$. Nossa distribuição será, então, representada por:

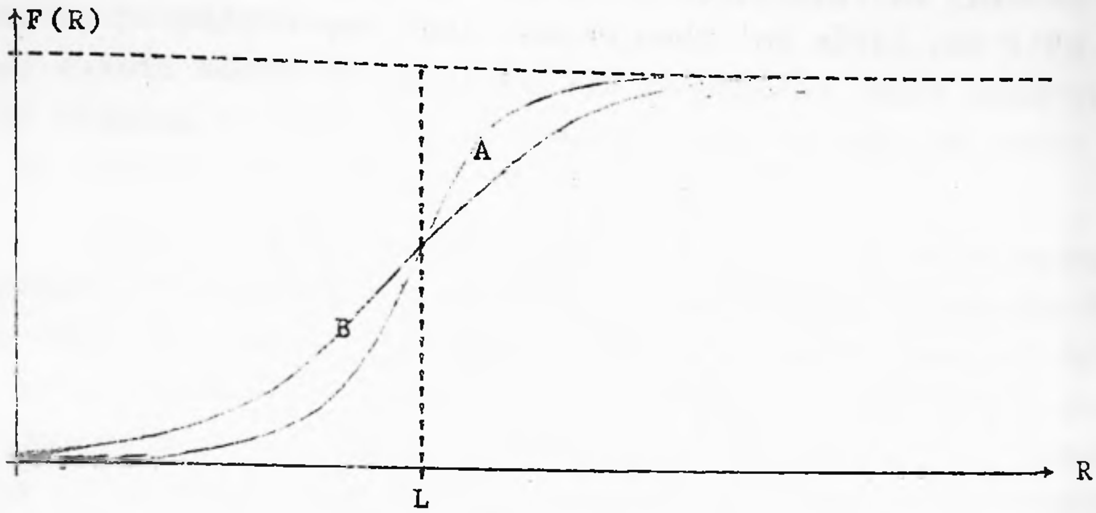
$$F(R_0) = \int_{-\infty}^{R_0} \frac{h \, e^{-h^2(R_i-L)^2}}{\sqrt{\pi}} \, dR = P_0 \qquad (II)$$

Se introduzirmos a transformação $\gamma_i = h(R_i-L)$ e portanto $dR = d\gamma_i/h$, vemos ter:

$$\phi(\gamma_0) = \int_{-\infty}^{\gamma_0} \frac{h \, e^{-\gamma_i^2}}{\sqrt{\pi}} \, d\gamma = P_0 \qquad (III)$$

A função assim representada passa a denominar-se *phigamma*.

Sabemos que σ^2 mede a dispersão das intensidades do estímulo em torno da média. Quanto mais preciso for o indivíduo em suas respostas, menor σ^2. Sendo $\sigma^2=1/2h^2$ podemos então considerar h como medida de precisão. Quanto menor a oscilação das respostas, as intensidades de estímulo se agruparão em torno da média do *stimulus* *limen*, em grandes proporções. Associemos então à precisão o valor de h, que como podemos ver é responsável pela inclinção da ogiva normal. A figura abaixo dá uma idéia intuitiva deste fato.



As curvas A e B representam as distribuições de certa habilidade dos indivíduos I_A e I_B respectivamente. Ambos possuem o mesmo *stimulus limen*, contudo I_A possue maior precisão em suas respostas que I_B. Vamos cometer menor erro ao predizer uma resposta de I_A que ao predizer uma resposta de I_B.

Nosso problema é estimação de L e h. Fechner (10), estudando discriminação de pesos, imaginou o seguinte meio de chegar às estimativas de L e h: Considerar a proporção de respostas certas para cada diferença de pesos, e a seguir procurar o valor γ, através a equação (III), correspondente a esta proporção. Poder-se-ia então, estabelecer uma relação linear entre diferenças de pesos e os valores obtidos através de ϕ(γ) que corresponde, como vimos, a N(0, 1). Obtida a reta estimaríamos L e h. Fechner foi quem primeiro teve a idéia de transformar a sigmóide normal a uma reta. Contudo, a ponderação das proporções não foi levada devidamente em conta.

Com Müller (1789) e Urban (1910) este problema já é considerado. Analisemos de uma forma geral, em que consiste o *método constante* devido principalmente aos dois últimos pesquizadores.

Consideramos para cada intensidade de estímulo a proporção de acertos e, a seguir, através da equação (III), chegamos ao valor correspondente a esta pròporção. Como $\gamma_i = h(R_i - L)$, procuramos então, estimar a reta que relaciona γ a R. Mas, para isto não devemos considerar igualmente todos os valores obtidos. É interessante considerar aquí a precisão das observações, e as alterações que sofrem devido às ligeiras modificações casuais das proporções. Nestes dois fatos é que se baseiam Urban e Müller.

Urban aconselha a ponderar as observações feitas, bem como os valores Y_i, por uma quantidade proporcional ao inverso da variância da proporção que lhes corresponde, ou seja, por 1/PQ. Isto porque, quanto menor a variância das proporções, mais estáveis os valores obtidos através delas. Estamos aqui levando em conta a precisão de nossas observações.

Müller baseia-se no seguinte fato para a consideração de seus pesos: Considerando a ogiva normal, vemos que a porção da curva que apresenta maior inclinação corresponde ao intervalo dos valores próximos ao *stimulus limen*. Vemos que as flutuações casuais das proporções correspondentes a esta porção, não vão ocasionar grandes diferenças aos valores de R_i ou Y_i, que lhes correspondem, o que aconteceria para proporções correspondentes aos pontos mais extremos. Para proporções extremas pequenos valores em P acarretam grandes diferenças para Y e R. Logo os valores centrais deverão ser levados mais em conta. Müller, então, aconselha a ponderar as observações por Y^2, onde Y é a ordenada da curva de densidade de freqüência.

Considerando agora uma combinação dos pesos de Urban e Müller, chegamos ao peso Y^2/PQ, que como vimos, se aproxima aos obtidos considerando-se a técnica do *probit analysis*. Contudo, como podemos ver, os pesos para o *probit analysis* são calculados com base na linha reta arbitrariamente traçada, e não diretamente com base nas proporções observadas. Se cuidados forem tomados ao se traçar esta linha provisória, isto levará a uma grande redução da influência dos erros casuais.

Sendo de grande utilidade prática para os problemas no terreno biológico, a técnica adotada pelos psicofísicos passou também a ser utilizada pelos biólogos e com estes sofreu algumas alterações, chegando a uma maior perfeição.

Apesar disso, como diz Finney (5), por muito tempo os psicólogos desconheceram as melhorias alcançadas por sua técnica pelos biólogos. É o que podemos ver, por exemplo, através do artigo de Ferguson publicado em 1942 - *ITEM SELECTION BY THE CONSTANT PROCESS*. Este artigo traz uma grande contribuição para o estudo dos testes mentais, permitindo a estimação da dificuldade do item e da precisão, por meio das técnicas usadas pelos psicofísicos. Cada item, diz Ferguson, pode ser escrito em termos de um limen que é um índice do ponto em que o item discrimina, e de um desvio padrão que é um índice da eficiência da discriminação.

Para aplicar tais técnicas ao campo de itens, Ferguson colocou o experimento nos seguintes moldes: Utilizou os resultados de um teste aplicado a uma amostra de crianças[1], aos quais o teste se destinava, cujas idades não diferiam mais de um ano, uma vez que a variância da habilidade varia com a idade. A seguir, supondo a habilidade em questão normalmente distribuída, considerou k categorias de indivíduos, conforme o grau de habilidade, sendo que estas categorias são representadas por intervalos de amplitude constantes, com exceção dos extremos. Foram estudados vários itens. Para cada item foi calculada a proporção dos indivíduos de cada categoria que responderam corretamente. Calculadas as proporções p_i, através da normal chegou-se aos valores de z. O problema é, então,

$$p_i = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du$$

adaptar uma reta aos pontos que têm X_i por abcissa e Z_i por ordenada, de modo a estimar a real relação entre Z e X. Neste estágio, Ferguson, embora conhecedor da superioridade do processo constante, utiliza além deste, outros dois meios, a saber, processo dos mínimos quadrados e, simples interpolação por dois pontos, a fim de ver a eficiência dos dois últimos em relação ao primeiro. Conclue que, pelo fato de não considerarmos pesos para o processo dos mínimos quadrados, os resultados diferem um tanto dos obtidos através do processo constante, devido, principalmente, à influência dos valores extremos (sobre isto já tivemos a oportunidade de falar). Se, por questões práticas, o processo dos mínimos quadrados for usado, para a estimação da reta, seria aconselhável , então, a abandonar os casos extremos. Devemos observar aqui um outro inconveniente: sabemos que a cada X está associada uma distribuição de Z. Se as variâncias das distribuições dos Z não forem iguais, o método dos mínimos quadrados, não nos permite ir além no sentido de fazer inferências sobre $\hat{\beta}$ e $\hat{\alpha}$. Uma das hipóteses exigidas para isso é a igualdade dessas variâncias. Os resultados obtidos através de simples interpolação foram tão bons quanto os obtidos pelos mínimos quadrados, podendo ser empregado, segundo Fer-

---

(1) Dos resultados da aplicação do *MORAY HOUSE TEST*, as crianças de 11 anos, foi selecionada uma amostra de 216 crianças, como representativa de todo o grupo.

guson, sempre que muito rigor não for exigido.

Através da reta poderíamos chegar ao limen e a variância do item, ou seja, o limen seria o valor da abcissa que corresponderia ao Z relativo a P=0,5, e a variância seria o inverso do coeficiente angular da reta. Estes dois elementos caracterizariam o item, pois quanto mais difícil, maior o valor do limen, e quanto mais preciso, menor a variância, ou seja, maior a inclinação da reta.

Fergurson chama a atenção para o fato, de que o método acima desenvolvido, poderá também servir para descrever pessoas. A habilidade de um indivíduo poderá ser descrita em termos de um limen e de seu desvio padrão. Basta para isto considerar os itens de um teste como representativos de uma população definida de itens. Consideraríamos agora k sub-testes de ordem de dificuldades crescentes, dificuldades estabelecidas em relação a uma certa população. Aplicado o teste a uma dada pessoa, podemos então, verificar as proporções de respostas certas para cada sub-teste, transformá-las a valores de Z, e considerar então, a reta ligando estes valores aos pontos médios das dificuldades de cada categoria. Poderíamos pois, para cada pessoa, encontrar um limen e um desvio padrão, que servirão para caracterizá-la.

Uma questão deve ser mencionada: Para o estudo do item Ferguson baseou-se em K amostras independentes. Para cada uma das K categorias de habilidade considerou uma amostra. Considerando agora este caso, podemos considerar as K situações independentes? É muito comum, por exemplo, nos testes poder, a apresentação dos itens numa ordenação crescente de dificuldade. Tal procedimento não levaria a uma certa dependência no sentido de diminuir a real diferença entre as porcentagens de acertos das diversas categorias, de sorte a obter uma super estimativa do limiar do indivíduo? E a precisão seria também afetada? Este fato afetaria igualmente a todos os indivíduos Os itens devem ser dispostos de maneira a diminuir o mais possível essa influência. O aspecto interessante da observação de Ferguson é podermos associar a cada indivíduo uma precisão. É claro que existem fatores momentâneos, temporários, influindo sobre a variabilidade das respostas, mas, mesmo assim deve ser levada em conta para o estudo do indivíduo, principalmente quando estamos diante de indivíduos problemas.

Em 1944, são publicados, na Psychometrika, dois artigos relacionados ao problema de testes e métodos psicofísicos. Um, devido a Maurice Lorr (20), que estabelece a relação matemática entre o score do indivíduo e, o score limen estimado pelo processo constante, baseado na hipótese de a habilidade ser normalmente distribuída, para testes homogêneos quanto ao tipo de questões, porém, com dificuldade variável. Outro foi escrito por Finney (5).

Tomando conhecimento do artigo escrito por Ferguson, Finney compara os problemas por ele abordados aos existentes no terreno toxicológico, mostrando a grande analogia entre eles. Em ambos estamos diante de resposta quantal, ou seja, o inseto morre ou sobrevive, o indivíduo acerta ou erra o item, e para ambos os casos, vamos registrar o número de elementos dando um certo tipo de resposta. No terreno dos inseticidas temos o veneno, os insetos e a tolerarcia. Passando agora às medidas psicológicas, vamos ter em correspondência o item, os indivíduos, e a habilidade que o item pretende medir. Assim, o nível de habilidade necessário para responder corretamente ao item, vai corresponder à dose necessária de tóxico para matar o inseto. Uma vez que a técnica *probit analysis* dá estimativas máximas verossímeis, e como existem tábuas publicadas de modo a facilitar grandemente os cálculos, sua aplicação ao estudo dos testes mentais traria grandes vantagens.

A fim de estabelecer uma comparação concreta entre ambos os métodos, isto é, entre o *Probit Analysis* e o *Método Constante*, Finney utilizou-se dos mesmos dados de Ferguson. Baseou-se também nos mesmos princípios, a saber, como citério para a classificação dos indivíduos utilizou o próprio teste, considerou o mesmo número de classes, com intervalos possuindo a mesma amplitude adotada por Ferguson. Contudo ao considerar o valor para representar os elementos de uma classe, não adotou a média aritmética dos limites de classe, o que supunha sobre cada intervalo uma distribuição retangular, e sim levou em conta a hipótese de normalidade, calculando a média por:

$$\bar{x}_i = \frac{Y_{if} - P_{is}}{P_{is} - P_{if}}$$

onde Y_{if} é a ordenada correspondente ao limite inferior x_{if} da iésima classe. Y_{is} é a ordenada correspondente ao limiye superior x_{is}, da iésima classe. P_{if} e P_{is}, são respectivamente:

$$P_{if} = \int_{-\infty}^{X_{if}} \frac{1}{\sqrt{2\pi}}\, e^{\frac{-x^2}{2}}\, dx$$

$$P_{is} = \int_{-\infty}^{X_{is}} \frac{1}{\sqrt{2\pi}}\, e^{\frac{-x^2}{2}}\, dx$$

Estamos considerando os valores em σ - unidades.

Os resultados experimentais obtidos por Finney, pelos 2 métodos, foram semelhantes. Porém através do *probit analysis* foi possível associar ao limen e ao desvio padrão dos resultados os respectivos erros padrões, outra vantagem além da teórica.

Finney, então, observa que muitos problemas existentes no campo dos testes mentais poderiam ser resolvidos satisfatoriamente pela técnica *probit analysis*, e como exemplo menciona a comparação de itens, quanto à dificuldade, supondo o paralelismo das retas obtidas.

Duas décadas após a publicação do artigo de Finney, Rhea S. Das (1964), retomando este assunto, fez algumas considerações interessantes e práticas, mostrando que através do *probit analysis* poderíamos obter ao mesmo tempo várias informações sobre o item, a saber:

1 - a dificuldade do item seria interpretada, de acordo com Finney como o limen, ou seja, o valor da habilidade para o qual 50% das pessoas possuidoras deste valor acertam o item e 50% falham.

2 - o poder discriminativo do item seria dado pelo coeficiente da reta obtida através desta técnica.

3 - A relação entre item e habilidade seria estimada por meio de equação da reta obtida. Nos experimentos anteriormente mencionados foi adotado o critério interno, ou seja, o próprio score do teste total, como indicador da habilidade. Designando agora por X a variável representando o critério, seja interno ou externo, a relação será estimada por:

$$\overset{*}{Z} = \overset{*}{\beta} X + \overset{*}{\alpha}$$

4 - A hipótese inicial é que a habilidade para responder ao item é normalmente distribuída. Sendo assim, a relação entre

a variável probit e a variável independente representando a habilidade, deve ser linear. Obtemos então, por meio da reta estimada os valores teóricos para o probit e, por meio destes as proporções que lhes correspondem. Para um dado X vamos ter então, as proporções teóricas e as observadas. É preciso ver o acordo entre elas e, isto é feito por meio de um teste χ^2. Se as hipóteses de perfeito acordo (não diferença entre os valores teóricos observados) for rejeitada, concluimos que a relação não é linear e portanto, a habilidade em responder ao item não é normalmente distribuída. Tendo razões para supor que a habilidade obedece à distribuição normal, conclui-se que através da prova χ^2, poderíamos selecionar os itens para o teste final, pois todos os itens cuja relação linear for rejeitada não satisfazem a hipótese inicial, ou seja, a habilidade em respondê-los não é normalmente distribuída, e portanto devem ser rejeitados.

5 - Poder-se-ia também, com facilidade, selecionar itens para formas paralelas. Medindo a dificuldade dos itens por meio do limen, m, e o poder discriminativo através do coeficiente angular β, poderíamos obter vários conjuntos de itens com mesmo limen e mesmo poder discriminativo. Seria possível então construir formas paralelas distribuindo os itens dos vários grupos por meio das diversas formas.

A definição de Finney, Ferguson, Lorr, e outros de dificuldade do item como o limen, em lugar de considerar a proporção de indivíduos que o acertaram, traz a vantagem de defini-la em termos de habilidade em estudo, ao passo que a definição considerando a proporção de indivíduos que o acertaram é dada em termos de população. É claro que tendo a definição em termos de limen, podemos facilmente passar às proporções e ainda mais, proporções obtidas através um processo mais definido.

Outro fato que deve ser notado é referente ao poder discriminativo do item, que como dissemos é indicado pelo coeficiente da linha de regressão. Em condições ideais, um item faria perfeita discriminação se todos os indivíduos que possuissem um nível de habilidade inferior a um certo valor, digamos X_0, errassem o item, e os que possuissem valor igual ou maior acertassem. Neste caso iríamos obter uma reta paralela ao eixo das ordenadas, e portanto $\overset{*}{\beta}$ tendendo ao infinito. Por outro lado, se o item não tiver poder discriminativo, indivíduos com qualquer grau de habilidade poderão acertá-lo indistintamente, e a nossa reta será paralela ao eixo das abcissas, tendo agora $\overset{*}{\beta} = 0$. Logo quanto maior o poder discriminativo do item, maior a inclinação da reta. É portanto um meio através do qual podemos verificar a validade do item. Itens cujas retas de regressão possuem muito baixo coeficiente de regressão não estão medindo de forma alguma o que desejamos, e devem ser rejeitados. Como vemos é outra forma de verificação da validade do item, além das mencionadas na introdução, como coeficiente de correlação bi-serial, coeficiente tetracórico, etc.. A técnica probit, então, fornecendo nos um índice de dificuldade e de validade para cada item, nos fornece dois elementos indispensáveis para a construção do teste. Não basta escolhermos os itens com maior poder discriminativo. É preciso que estas discriminações sejam feitas através das várias categorias de habilidade. Devemos considerar várias classes de dificuldades. Os itens serão classificados segundo as dificuldades. De cada classe será escolhido um certo número de itens; serão escolhidos os de maior validade.

Até aqui consideramos cada item com apenas uma alternativa certa. Frank B. Baker e Jonh Gurland (9) estendem a técnica *probit* ao caso em que para cada item, possuindo múltiplas alternativas não ordenadas, uma apenas é errada sendo as restantes certas, e a distribuição da habilidade responsável pela resposta a cada alternativa é normal. Admitindo a independência entre as alternativas, para um item possuindo t alternativa teremos que estimar 2(t-1) parâmetros, pois a cada alternativa vamos associar uma reta, e portanto a cada uma devemos associar um α e um β. O raciocínio é análogo ao já exposto para o caso de uma alternativa correta apenas: em lugar de binomial partimos da multinomial e a seguir todas as passagens consideradas para o caso simples são efetuadas. Assim chegaríamos ao seguinte:

Supondo o item com t alternativas, sendo a t a errada, os indivíduos classificados em m categorias, e indicando:

n_i = número de elementos da categoria i respondendo o item.

r_{ij} = número de elementos da categoria i respondendo a alternativa j.

$$E(p_{ij}) = E(r_{ij}/n_i) = P_{ij} \quad ; \; i=1,2,\ldots,m \quad e \quad j=1,2,\ldots,t.$$

$Z_j = \alpha_j + \beta_j x_j$, a reta a ser estimada para a alternativa j; $j=1,2,\ldots,t-1$.

$a_1, a_2, \ldots, a_{t-1}$; $b_1, b_2, \ldots, b_{t-1}$ primeiras aproximações a $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}$; $\beta_1, \beta_2, \ldots, \beta_{t-1}$ e $\delta a_1, \delta a_{21}, \ldots, \delta a_{t-1}$; $\delta b_1, \delta b_{21}, \ldots, \delta b_{t-1}$ acréscimos, obtidos após o primeiro ciclo de iteração aos a_j e b_j ; $j=1,2,\ldots,t-1$.

$$Z_{ij} = \frac{P_{ij} \cdot P_{Iit} - P_{Iij} \cdot P_{it}}{Y_{Iij}} \quad \text{e} \quad W_{ij} = \frac{Y^2_{Iij}}{P_{Iij} \cdot P_{Iit}}$$

onde P_{Iij}, Y_{Iit}, P_{Iit} são obtidos a partir da reta $Z_{Ij} = a_j + b_j x_i$, conservando o mesmo significado anteriormente adotado.

$$\lambda a_j = \sum_{i=1}^{m} n_i W_{ij} Z'_{ij} \quad \text{e} \quad \lambda b_j = \sum_{i=1}^{m} n_i W_{ij} x_i Z'_{ij} \qquad j=1,2,\ldots,t-1.$$

$$\delta' = \left| \delta a_1, \ \delta b_2, \ \ldots, \ \delta b_{t-1} \right| \quad \text{e} \quad Y' = \left| \lambda a_1, \ \lambda b_2, \ \ldots, \ \lambda b_{t-1} \right|$$

Λ indicando as expressões obtidas quando substituímos os verdadeiros valores de P e Y por P_I e Y_I nas esperanças das segundas derivadas. Assim:

$$\left.\begin{aligned}
\Lambda a_j, a_j &= - \sum_{i=1}^{m} \frac{n_i Y^2_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij} + P_{Iit}) \\
\Lambda b_j, b_j &= - \sum_{i=1}^{m} \frac{n_i x_i^2 Y^2_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij} + P_{Iit}) \\
\Lambda a_j, b_j &= - \sum_{i=1}^{m} \frac{n_i x_i Y^2_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij} + P_{Iit})
\end{aligned}\right| \quad \text{para } j = j'$$

$$\left.\begin{aligned}
\Lambda a_j, a_j &= - \sum_{i=1}^{m} \frac{n_i Y_{Iij}, Y_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij}) \\
\Lambda b_j, b_j &= - \sum_{i=1}^{m} \frac{n_i x_i^2 Y_{Iij}, Y_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij}) \\
\Lambda a_j, b_j &= - \sum_{i=1}^{m} \frac{n_i x_i Y_{Iij}, Y_{Iij}}{P_{Iij} \cdot P_{Iit}} (P_{Iij})
\end{aligned}\right| \quad \text{para } j \neq j'$$

$$C = - \begin{bmatrix} \Lambda a_1 a_1 & \cdots & \Lambda a_1 b_{t-1} \\ \vdots & & \vdots \\ \Lambda b_{t-1}, a_1 & \cdots & \Lambda b_{t-1}, b_{t-1} \end{bmatrix}$$

Então os valores $\delta a_1, \delta a_2, \ldots, \delta b_{t-1}$ são obtidos por:

$$\delta = C^{-1}Y$$

Nossas novas estimativas serão $(a_j + \delta a_j)$ e $(b_j + \delta b_j)$; $j=1,2,\ldots t-1$

O processo pode ser repetido enquanto for necessário.

As variâncias e covariâncias assíntotas serão estimadas por:

$$\begin{bmatrix} -\Lambda a_1 a_1 & \cdots & \Lambda a_1 b_{t-1} \\ \vdots & & \vdots \\ -\Lambda b_{t-1}, a_1 & \cdots & -\Lambda b_{t-1}, b_{t-1} \end{bmatrix}^{-1}$$

Um teste total de linearidade será dado por:

$$\chi^2 = \sum_{j=1}^{t} \sum_{i=1}^{m} n_i \frac{(P_{ij} - P_{Iij})^2}{P_{Iij}}$$ com $(t-1)(m-2)$ graus de liberdade.

Para um teste assim formulado, vamos encontrar um problema sério que é posição da alternativa. Muitos ao lerem a primeira alternativa correta, reconhecendo-a como tal, poderão escolhê-la imediatamente e não irem além. Isto contribuirá para uma maior proporção de escolhas desta alternativa, o que não indica a preferência, nem o fato de ser mais fácil do que as outras. Além, disso, a técnica aplicada a situações como esta se torna bastante laboriosa. É preciso ver até que ponto o fator posição da alternativa interfere sobre os reais resultados, se a influência for grande, não se justifica a aplicação desta técnica.

Exemplo: Passemos a um exemplo prático do que foi anteriormente dito: Do total de resultados obtidos através da aplicação do teste Dominós aos alunos dos primeiros e segundos anos, do segundo ciclo, de determinado colégio, foi selecionada uma amostra de 154 indivíduos. Nosso objetivo será o estudo de alguns itens desse teste através da técnica *probit analysis*. Para o estudo da validade do item, devemos primeiramente classificar os indivíduos de acordo com o nível de habilidade que o teste pretende medir. Para isso, devemos lançar mão de um meio, já reconhecido como válido e que não seja o próprio teste em estudo, ou seja, devemos usar um critério externo. Usando o critério interno, temos o problema da não independência entre item e critério. Contudo se o teste possuir um número razoável de itens, o fato de o aluno ter acertado ou errado o item terá pouca influência sobre o seu total. Outra observação a ser feita é que usando o critério interno, ao concluirmos que um certo item tem alta validade, não podemos interpretar esta validade com relação à habilidade que o teste pretende medir, mas, sim com relação ao que o teste realmente mede. Assi, segundo Richardson: *A correlação item-teste dá uma indicação da extensão com que o item mede, o que o teste mede como um todo (27)*. Na realidade a correlação item-teste nos dará um índice de homogeneidade.

Como nosso objetivo é apenas mostrar como aplicar a técnica, dar um exemplo palpável do que foi exposto, vamos usar o próprio teste Dominós para classificar indivíduos, portanto um critério interno. Embora estudos anteriores já tenham indicado que a habilidade medida por este teste se distribue normalmente, nosso primeiro passo será verificar esta afirmação:

A média e a variância observadas foram $\bar{X} = 34{,}04$ e $s^2 = 26{,}66$. Para nossa hipótese vamos adotar um nível de 5%.

Os dados foram agrupados inicialmente, em 7 classes e as proporções observadas e as teóricas, obtidas sob a hipótese de normalidade, foram calculadas. O teste de χ^2 (TABELA I) nos leva a não rejeição da hipótese de normalidade, pois, ao nível de 5% e 3 graus de liberdade temos $\chi^2_c = 7{,}815$. Como foram estimados dois parâmetros a saber α e σ e as duas últimas classes combinadas, pois $n\tilde{P}_i < 5$, temos apenas 3 graus de liberdade.

TABELA I - COMPARAÇÃO ENTRE AS FREQUÊNCIAS ESPERADAS, SOB A HIPÓTESE DE NORMALIDADE, E AS OBSERVADAS

X	n_i	$\tilde{P}_i$	$n\tilde{P}_i$	$D_i = \lvert n_i - n_i\tilde{P}_i \rvert$	D_i^2/nF_i
abaix.25	5	0,0401	1,1754	1,1754	0,2237
25 — 29	15	0,1235	19,0190	4,0190	0,8492
29 — 33	41	0,2572	39,6088	1,3912	0,0489
33 — 37	40	0,2949	45,4146	5,4146	0,6455 $\chi^2 = 3{,}9475$
37 — 41	35	0,1958	30,1532	4,8468	0,7796
41 — 45	17	0,0715	11,0125	4,3695	1,4006
45 —	1	0,0170	2,1680		

Podemos então partir da hipótese que a habilidade medida pelo teste é normalmente distribuída. Vamos exprimir nossas medidas em termos de valores reduzidos, ou seja σ-unidades. A fim de evitar um número pequeno de observações nas classes extremas, distribuiremos nossas observações em apenas 5 classes, sendo as extremas de $-\infty$ — $-1{,}2$; $1{,}2$ — ∞ e as restantes com amplitudes iguais a 0,8, ou seja. $-1,2 \vdash -0,4$; $-0,4 \vdash 0,4$ e $0,4 \vdash 1,2$. Supondo a distribuição $N(34,04;\ 5,16)$ em correspondência aos pontos $-1,2$; $-0,4$; $0,4$ e $1,2$ temos os valores: 27,65; 31,93; 36,104; 40,23, o que nos permite classificar os indivíduos. Temos então:

$-\infty \vdash -1,2$	18
$-1,2 \vdash -0,4$	35
$-0,4 \vdash 0,4$	48
$0,4 \vdash 1,2$	35
$1,2 \vdash \infty$	18

É claro que a cada categoria devemos associar um único valor e não uma classe de valores. Considerar o ponto médio de classe como fez Ferguson, para representar o nível de cada categoria tem como vimos dois incovenientes: o primeiro com relação às classes extremas, pois a primeira tem limite inferior $(-\infty)$ e a última limite superior $(+\infty)$; o segundo é que ao fazermos isto, estamos admitindo que sobre cada intervalo a distribuição é uniforme, o que não é verdade. É preferível de acordo com Finney calcularmos para cada intervalo a média, considerando para cada um a distribuição normal. Sendo assim a média dos valores de cada classe será calculada, como vimos, por:

$$\bar{x}_i = \frac{Y_{if} - Y_{is}}{P_{is} - P_{if}}$$

Assim temos para a primeira classe: $Y_{if}=0$; $Y_{is}=0,19419$; $P_{if} = 0$ e $P_{is} = 0,11507$. E portanto:

$$\bar{x}_1 = \frac{-0,19419}{0,11507} = -1,687 \simeq -1,69$$

Para a segunda classe chegamos a $\bar{x}_2 = -0,76$ e pela simetria da distribuição normal, $\bar{x}_3 = 0$; $\bar{x}_4 = +0,76$ e $\bar{x}_5 = +1,69$. Vamos indicar os valores das médias por x.

Para nosso exemplo, escolhemos os itens 29, 33 e 34. Nosso objetivo é associar a cada item a reta mostrando a relação existente entre o item e o critério. O coeficiente angular da reta será um índice, como dissemos, da validade do item. Se o item e o critério forem independentes no sentido de medirem habilidades independentes, vamos ter um coeficiente angular igual a zero. Queremos também associar a cada item seu limiar que será um índice da dificuldade. Como podemos estar interessados no problema de paralelismo dos itens, vamos considerar este caso para os itens 33 e 34. Para cada item devemos inicialmente calcular a proporção dos indivíduos de cada categoria que o acertaram, e a seguir, através da tábua ja preparada (31) chegar aos probits correspondentes.

Nosso resultados estão na tabela II:

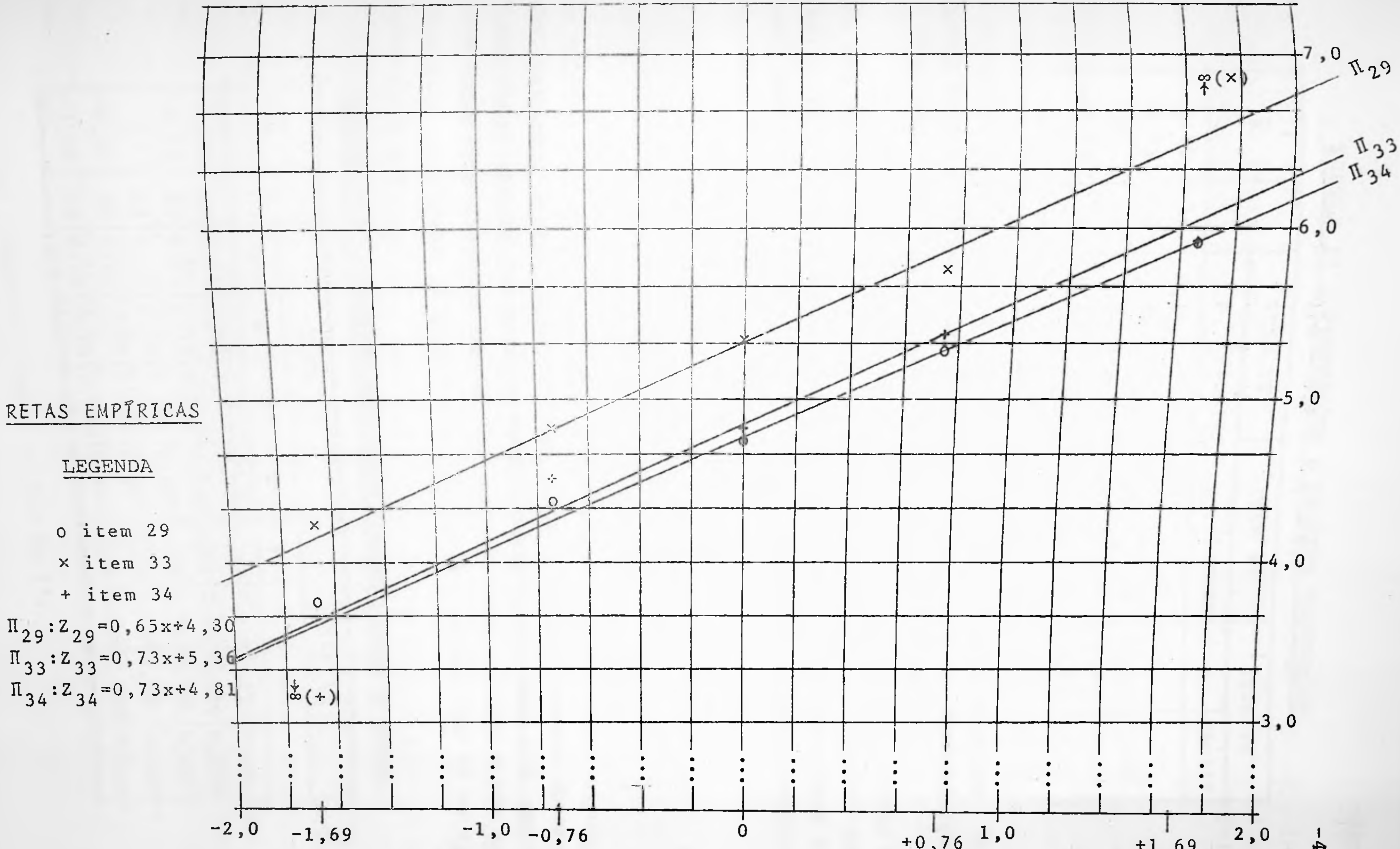

RETAS EMPÍRICAS
LEGENDA
o item 29
× item 33
+ item 34
Π29:Z29=0,65x+4,30
Π33:Z33=0,73x+5,36
Π34:Z34=0,73x+4,81
Π29
Π33
Π34
∞(×)
∞(+)
7,0
6,0
5,0
4,0
3,0
-2,0
-1,69
-1,0
-0,76
0
+0,76
1,0
+1,69
2,0

TABELA II - PROPORÇÕES E PROBITS OBSERVADOS

$\bar{x}$	n	ITEM 29		ITEM 33		ITEM 34	
		p	PROBIT	p	PROBIT	p	PROBIT
-1,69	18	0,11	3,77	0,22	4,23	0	$-\infty$
-0,76	35	0,26	4,36	0,43	4,82	0,31	4,50
0	48	0,39	4,72	0,64	5,36	0,40	4,75
+0,76	35	0,63	5,33	0,77	5,74	0,66	5,41
+1,69	18	0,83	5,95	1,00	$+\infty$	0,83	5,95

O processo iterativo será usado para a resolução das equações de máxima verossimilhança. Nosso primeiro passo será a determinação da reta empírica, relacionando os probits aos x (Fig.1).

Como desejamos saber se os itens 33 e 34 são paralelos e uma das condições é a igualdade de variâncias, sendo o desvio padrão o inverso do coeficiente angular, devemos, para estes dois itens, partir de um coeficiente angular comum. Nossas retas empíricas serão então:

$$Z_{29} = 0,65x + 4,80$$

$$Z_{33} = 0,73x + 5,36$$

$$Z_{34} = 0,73x + 4,81$$

Os valores de Z dados por estas retas são os nosso probits esperados. Baseados nestes probits e nas proporções inicialmente observadas vamos chegar aos probits operacionais z. Com base nos probits esperados chegamos também aos pesos w. Tanto z como w são encontrados em tábuas já preparadas. A tabela a seguir nos dá os valores de z e w.

TABELA III - PROBITS ESPERADOS, OPERACIONAIS E PESOS

$\bar{x}$	n	ITEM 29				ITEM 33			
		p	Z	z	w	p	Z	z	w
-1,69	18	0,11	3,70	3,777	0,336	0,22	4,13	4,233	0,481
-0,76	35	0,26	4,31	4,357	0,532	0,43	4,81	4,824	0,628
0	48	0,39	4,80	4,721	0,627	0,64	5,36	5,358	0,607
+0,76	35	0,63	5,29	5,332	0,617	0,77	5,92	5,723	0,465
+1,69	18	0,83	5,90	5,953	0,455	1,00	6,60	7,094	0,238

$\bar{x}$	n	ITEM 34			
		p	Z	z	w
-1,69	18	0	3,58	2,968	0,295
-0,76	35	0,31	4,36	4,526	0,520
0	48	0,40	4,81	4,747	0,628
+0,76	35	0,66	5,37	5,410	0,605
+1,69	18	0,83	6,04	5,948	0,425

Consideremos primeiramente o item 29 como resultado do primeiro ciclo de operações, temos:

$$\sum nw = 84{,}549 \qquad \sum nwxz' = 69{,}6441$$
$$\sum nwx = 5{,}8810 \qquad \sum nwz = 409{,}9534$$
$$\sum nwx^2 = 63{,}8933 \qquad \sum nwz^2 = 2.014{,}7152$$

e portanto:

$$\beta_2 = \frac{SSxz'}{x^2} = \frac{\sum nwxz - \dfrac{(\sum nwx)(\sum nwz')}{\sum nw}}{\sum nwx^2 \quad \dfrac{(\sum nwx')^2}{\sum nx}} = \frac{41{,}1289}{63{,}4842} = 0{,}6479$$

$$\alpha_2 = \bar{z} - \beta_2\bar{x} = 4{,}8487 - (0{,}6479)(0{,}0695) = 4{,}8037$$

Como resultado do primeiro ciclo de iteração obtivemos:

$$Z_2 = 0{,}6479x + 4{,}8037$$

Comparando estes coeficientes aos da reta empírica, temos:

$$h = (0{,}6500 - 0{,}6479) = 0{,}0021 \quad \text{e}$$
$$k = (4{,}8000 - 4{,}8037) = -0{,}0037$$

Considerando o fato que para os cálculos foram consideradas quatro casas decimais, podemos considerar os resultados como próximos, não havendo portanto necessidade de um novo ciclo de iteração. Caso fosse necessário, iríamos partir agora, da reta 0,648 + 4,804. Para cada x obteríamos um probit esperado Z_2, e a partir deste e da proposição inicialmente observada chegaríamos aos w_2 e z_2. Enfim todo o processo seria repetido considerando 0,648x + 4,804 em lugar da empírica.

Vamos então considerar 0,648x + 4,804 como a solução das equações de máxima verossimilhança. É preciso agora verificar se esta é a melhor relação entre teste e item. É preciso verificar o acordo entre Z_2 (teóricos) e os Z observados. Pode ser que os Z observados se afastem sistematicamente da reta, não sendo, portanto, linear a relação verdadeira. Um teste de χ^2 nos levará a não rejeição ou rejeição da equação estimada. Como vimos, podemos chegar ao χ^2 por uma das seguintes equações:

$$\chi^2_{m-2} = SS^2_{z'} - \frac{(SS_{xz'})^2}{SS^2_x} \qquad \text{(I)} \quad \text{ou}$$

$$\chi^2_{m-2} = \sum_{i=1}^{m} \frac{(r_i - n_i P_{2i})^2}{n_i P_{2i} Q_{2i}} \qquad \text{(II)}$$

Onde P_{2i} é a proporção correspondente ao probit teórico Z_{2i}.
A fórmula (I) leva a:

$$\chi^2_3 = \left[\sum nwz'^2 - \frac{(\sum nwz')^2}{\sum nw}\right] - \frac{(SS_{xz'})^2}{SS^2} = 26{,}9408 - 26{,}6458 = 0{,}295$$

Com 3 graus de liberdade e um nível de significância de 5%, temos $\chi^2_3 = 7{,}815$, o que nos leva a não rejeição da equação $Z_2 = 0{,}648x + 4{,}804$ como representando a relação entrem item 29 e o teste.

Um fato deve ser observado aqui: o número esperado de acertos da primeira classe é $n_1 P_{21} = (18)(0{,}097) = 1{,}746$. Para a última classe temos um número esperado de acertos igual a 14,670, o que nos leva a um número esperado de erros igual a 3,330. Frequências teóricas muito baixas, podem nos levar a um resultado de χ^2 desnorteador. Havendo classes com frequências teóricas baixas, digamos, menores que 5, convém agrupar estas classes a outras, aumentando assim a frequência teórica. Agrupando classes, não podemos mais usar a fórmula (I), apenas a (II). Vamos, então, agrupar as duas primeiras classes e as duas últimas conforme a tabela IV:

TABELA IV - COMPARAÇÃO ENTRE AS FREQUÊNCIAS ESPERADAS BASEADAS EM $Z_1 = 0{,}648x + 4{,}804$, E AS OBSERVADAS

x	Z_2	P_2	n	r	nP_2	r_c	nP_c	Q_c	$(r_c - nP_c)$	nP_cQ_c	$\frac{(r_c-nP_c)^2}{nP_cQ_c}$
-1,69	3,705	0,097	18	2	1,746	11	10,286	0,806	0,714	8,2905	0,0615
-0,76	4,308	0,244	35	9	8,540						
0	5,800	0,413	48	19	19,824	19	19,824	0,587	0,824	11,6367	0,0583
+0,76	5,292	0,615	35	22	21,525	37	36,195	0,317	0,805	11,4738	0,0565
+1,69	5,895	0,815	18	15	14,670						
											$\chi^2_1=0{,}1763$

Temos agora um grau de liberdade apenas, pois além dos dois parâmetros estimados fizemos os agrupamentos acima, perdendo assim mais dois graus de liberdade. Com um grau de liberdade temos $\chi^2_c = 3,841$, o que nos leva também, a não rejeição da equação $Z' = 0,648x + 4,804$. Como já foi exposto a variância de β_2 será dada por:

$$\sigma^2_{\beta_2} = \frac{1}{SS^2_x} = 0,01565$$

Sendo β obtido pelo princípio da máxima verossimilhança, sua distribuição tende a normal, a medida que o tamanho da amostra cresce. Sendo assim o intervalo de confiança de 95%, para β tem os limites:

$$0,648 \pm 1,96 \times \sqrt{0,01565} \text{ e } \therefore$$

$$0,403 < \beta < 0,893$$

O limiar será dado pela mediana. A estimativa da mediana é obtida através da equação:

$$m_2 = \frac{Z_2}{\beta_2} - \frac{\alpha_2}{\beta_2}$$

Como Z_2 deve corresponder a 50% temos:

$$m_2 = \frac{1}{\beta_2}\left[5 - \alpha_2\right] = \frac{5,000 - 4,804}{0,648} = 0,3025$$

Será sua variância:

$$\alpha^2_{m_2} = \frac{1}{\beta^2_2}\left[\frac{1}{\Sigma n_i w_i} + \frac{(m-\bar{x})^2}{SS^2_x}\right] = \frac{0,01268}{0,4199} = 0,0302$$

Intervalo de confiança de 95% para mediana, m, será:

$$0,3025 - (1,96)(0,174) < m < 0,3025 + (1,96)(0,174) \therefore$$

$$-0,0385 < m < 0,6435$$

Se $\beta=0$ teríamos $Z=\alpha$, portanto item e teste seriam linearmente independentes. No nosso caso vemos que $0,403 < \beta < 0,893$ não inclue o zero, o que nos leva a considerar a existência de certa validade do item.

Passemos aos itens 33 e 34.

Como em geral os estudos em biologia referentes às comparações de tóxicos, inseticidas, baseiam-se em amostras independentes, a técnica probit foi desenvolvida baseada nesta condição. Logo para fazermos diretamente sua aplicação ao estudo do paralelismo de itens teremos que admitir a hipótese de independência. Com os dados obtidos não é possível testar esta hipótese. Logo o que vai a seguir servirá apenas como exemplo, de como aplicar a técnica quando não temos dúvidas sobre a independência. Vamos pois, colocar a hipótese de independência das respostas, no sentido que a resposta dada a um item não tenha influenciado sobre a resposta dada ao seguinte. Nosso objetivo aqui é verificar se são paralelos. As retas empíricas, como vimos, são:

$$Z_{33} = 0,73x + 5,36$$

$$Z_{34} = 0,73x + 4,81$$

e os valores de z e w estão expressos na tabela III. Elaborando separadamente os dados referentes a cada item chegamos a:

ITEM 33:	$\sum nw = 80,3333$	$\sum nwz' = 422,3240$	$SS_{xz'} = 41,2823$
	$\sum nwx = -11,7278$	$\sum nwz'^2 = 2.251,7155$	$SS_x^2 = 57,3477$
	$\sum nwx^2 = 59,0598$	$\sum nwxz' = -20,3727$	$\bar{x} = -0,1459$
			$\bar{z} = 5,2572$
ITEM 34:	$\sum nw = 82,4790$	$\sum nwz' = 401,2858$	$SS_{xz'} = 44,3831$
	$\sum nwx = 6,2156$	$\sum nwz'^2 = 1.989,2612$	$SS_x^2 = 59,2897$
	$\sum nwx^2 = 59,7581$	$\sum nwxz' = 74,7239$	$\bar{x} = 0,0754$
			$\bar{z} = 4,8653$

Segundo o que vimos na exposição teórica o coeficiente angular comum β_{2c} será estimado por:

$$\beta_{2c} = \frac{{}_{33}SS_{xz'} + {}_{34}SS_{xz'}}{{}_{33}SS_x + {}_{34}SS_x^2} = \frac{85,6654}{116,6374} = 0,734$$

$${}_{33}\alpha_2 = 5,2572 - (0,734)(-0,1460) = 5,3644$$

$${}_{34}\alpha_2 = 4,8653 - (0,734)(0,0754) = 4,8100$$

Por razões análogas às mencionadas quanto ao item 29, vamos considerar nossas aproximações satisfatórias. Nossas retas estimadas são pois:

$${}_{33}Z_2 = 0,734x + 5,364 \qquad (I)$$

$${}_{34}Z_2 \quad 0,734x + 4,810$$

É preciso testar a hipótese da linearidade exprimindo a relação entre os itens e o teste. Se esta hipótese não for rejeitada, testaremos, então, o paralelismo das retas. Isto será verificado através de:

$$\chi_7^2 = \sum_{i=1}^{2} SS_{z'}^2 - \frac{\left(\sum_{i=1}^{2} SS_{xz'}\right)^2}{\sum_{i=1}^{2} SS^2} = 68,3696 - 62,9177 = 5,4519$$

Este χ^2_7 deve ser considerado como uma soma dos valores χ^2_3 para cada item separadamente, como teste de linearidade, cada qual com 3 graus de liberdade e um χ^2_1 residual com um grau de liberdade como teste de paralelismo.

$$\chi^2_6 = \sum_{i=1}^{2} SS^2_{z'} - \sum_{i=1}^{2} \frac{(SS_{xz'})^2}{SS^2_x} = {}_{33}\chi^2_3 + {}_{34}\chi^2_3 = 1{,}7702 + 3{,}6578 = 5{,}4280$$

$$\chi^2_1 = \sum_{i=1}^{2} \left[\frac{(SS_{xz'})^2}{S^2_x} \right] - \frac{(\sum SS_{xz'})^2}{\sum_{i=1}^{2} SS^2_x} = \chi^2_7 - \chi^2_6 = 5{,}4519 - 5{,}4280 = 0{,}0239$$

Ao nível de 5% que $\chi^2_6 = 12{,}592$ o que nos leva a não rejeição da hipôtese de linearidade e como ao mesmo nível, $\chi^2_1 = 3{,}841$, não rejeitamos a hipôtese de paralelismo, e consequentemente da igualdade variâncias, das distribuições dos itens 33 e 34.

Nossas estimativas da verdadeira relação entre item e teste são dadas por (I).

Passemos agora às estimativas das medianas:

$${}_{33}m_2 = \frac{5{,}000}{0{,}734} - \frac{5{,}364}{0{,}734} = -0{,}4959$$

$${}_{34}m_2 = \frac{5{,}000}{0{,}734} - \frac{0{,}481}{0{,}734} = 0{,}2588$$

$${}_{33}\sigma^2_{m_2} = 1{,}8563 \left[0{,}01245 + \frac{(0{,}3499)^2}{59{,}3477} \right] = 0{,}02706$$

$${}_{34}\sigma^2_{m_2} = 1{,}8563 \left[0{,}01212 + \frac{(0{,}1834)^2}{59{,}2897} \right] = 0{,}02356$$

Limites de confiança de 95% para as medianas são:

$$-0{,}817 < m_{33} < 0{,}175$$

$$-0{,}276 < m_{34} < 0{,}323$$

Vejamos agora os limites de confiança para a diferença de medianas.

Pelo teoerema de Fieller temos:

$$R_s, R_i = (\bar{x}_2 - \bar{x}_1) + \left[D - (x_2 - x_1) \pm \frac{1{,}96}{\beta_{1c}} \left\{ \left(\frac{1}{\sum_{i=1}^{m_1} n_{1i} w_{1i}} + \frac{1}{\sum_{i=1}^{m_2} n_{2i} w_{2i}} \right) (1-g) + \frac{(D - \bar{x}_2 + \bar{x})^2}{\sum_{j=1}^{2} SS^2_x} \right\}^{1/2} \right] : (1-g).$$

onde $D = {}_{34}m_2 - {}_{33}m_2$ e $g = \dfrac{(1,96)^2}{\beta_{1c}^2 \sum_{i=1}^{2} {}_iSS_x^2}$

$$R_s, R_i = 0,2214 + \left[0,7547 + (-0,2214) \pm \frac{1,96}{0,734} \left\{ (0,02457)(0,9389) + \right.\right.$$

$$\left.\left. + \frac{0,28441}{116,6374} \right\}^{1/2} \right] : 0,9389 = 1,244 \; ; \; 0,335$$

portanto: $0,335 < m_{34} - m_{33} < 1,244$

vemos que há uma diferença significativa entre as medianas, pois, o intervalo para a diferença não inclui o zero.

Passemos a validade:

Vimos que $\sigma^2_{\beta_{2c}} = \dfrac{1}{\sum_{i=33}^{34} S_i S_x^2} = \dfrac{1}{116,6374} = 0,00857$

O intervalo de confiança de 95% para β_c será:

$$I_{95\%} = 0,734 - (1,96)(0,0926) < \beta_c < 0,734 + (1,96)(0,0926)$$

$$I_{95\%} = 0,641 < \beta_c < 0,827$$

o que nos leva a não rejeitar a hipótese da validade dos itens 33 e 34,, para medirem em certa extensão a habilidade medida pelo teste. Concluímos também, sob a hipótese de independência, que não podemos considerar os itens 33 e 34 paralelos. Se a independência não puder ser admitida, o raciocínio acima não será válido.

A necessidade de hipótese de independência para o estudo de paralelismo restringe bastante a possibilidade da aplicação da teoria já elaborada sobre probit ao estudo do paralelismo de itens. Em geral os itens a serem selecionados são aplicados a uma amostra cujo tamanho deverá ser grande a fim de chegarmos a resultados precisos. O número de itens do teste provisório, deverá também ser elevado, bem acima do número desejado para constituir o teste final, a fim de escolher os mais convenientes para em conjunto formarem o teste com a maior validade possível. A independência exige que a amostra dos indivíduos se mantenha sempre a mesma de item para item, de maneira a considerá-la sempre como representativa de uma mesma população. Talvez fosse interessante subdividir a amostra em sub-amostras e a cada uma aplicar disposições diferentes dos mesmos itens, a fim de diminuir a influência da posição pois esta estaria ligada ao problema do cansaço, experiência adquirida ao responder os anteriores, etc., contribuindo para uma alteração na amostra. Colocando de lado o problema do paralelismo a teoria sobre *probit analysis* pode ser aplicada a cada item sem alterações. Cada item será caracterizado por estimativas obtidas pelo princípio da máxima verossimilhança, possuindo assim propriedades muito interessantes. Além disso, sabendo que a variável critério é normalmente distribuida a *técnica probit* permite testar a hipótese de normalidade da habilidade responsábel pela resposta ao item, ao mesmo tempo que nos fornece um índice de validade e dificuldade do item. Se considerarmos os coeficientes de correlação bi-serial, tetracórico, temos que uma das exigências para aplicação destes é que a variável item seja normalmente distribuida. Não sendo normalmente distribuida podemos chegar a resultados absurdos. Se não houver fortes razões para considerarmos a variável item com esta distribuição, nossos resultados serão duvidosos. Temos então outra qualidade interessante da técnica *probit analysis*. Contudo tal técnica é bastante laboriosa. É preciso ver qual a finalidade do teste, qual a extensão de sua aplicação para vermos se realmente se justifica sua aplicação. Mesmo com computadores a determinação das retas empíricas, a obtenção dos probits operacionais e pesos, para cada ciclo de iteração, exigem um certo trabalho.

BIBLIOGRAFIA

1. *Anastasi* - PSYCHOLOGICAL TESTING, The MacMillan Company, New York, 1961.

2. *Anstey* - THE METHOD OF ITEM ANALYSIS, The Britsh Journal of Psychology, 1947, Vol. 1.

3. *C.I.Bliss* - THE CALCULATION OF THE DOSAGE - Mortality Curve, The Annals of Applied Biology, 1935, Vol. XXII, Feb.

4. *Das Rhea, S.* - ITEM ANALYSIS "PROBIT" AND FRACTILE GRAPHICAL METHODS, The Journal of Statistical Psycology 1964, Vol. XVII, May.

5. *D.J.Finney* - THE APPLICATION OF PROBIT ANALYSIS TO THE RESULTS OF MENTAL TEST, Psychometrika, Vol.9 nº1 March.

6. *D.J.Finney* - PROBIT ANALYSIS, 1952, Second Edition, Cambridge at the University Press.

7. *D.J.Finney* - STATISTICAL METHOD IN BIOLOGICAL ASSAY, 1952, Hafner Publishing Company, New York.

8. *F. Garwood* - THE APPLICATION OF MAXIMUM, Likelihood to Dosage Mortality Curve, Biómetrika, 1941, 32(46-58)

9. *Frank B.Baker and John Gurland* - AN EXTENSION OF ITEM ANALYSIS PROCEDURES TO THE CASE OF POLYCHOTOMOUS RESPONSE, Psychometrika, 1968, September, Vol. 33, nº 3.

10. *Gustav Fechner* - ELEMENTS OF PSYCHOPHYSICS, Translated by Helmut H. Adler, Henry Holt Editions in Psychology, 1965.

11. *H. Gulliksen* - THEORY OF MENTAL TESTS, 1950, New York, John Wiley.

12. *C.A.Ferguson* - ITEM SELECTION BY THE CONSTANT PROCESS, Psychometrika, 1942, nº 1, February.

13. *J. Aitchison and S. D. Silvey* - THE GENERALIZATION OF PROBIT ANALYSIS TO THE CASE OF MULTIPLE RESPONSES, Biometrika, 1957, Vol. 44 (131-1400.

14. *J. A. Nelder* - WEIGHTED REGRESSION, QUANTAL RESPONSE, AND INVERSA POLYNOMIALS, Biometrics, 1968, Dec., Vol.24

15. *J.P.Guilford* - FUNDAMENTAL STATISTICS IN PSYCHOLOGY AND EDUCATION, 1942, First Edition, McGraw-Hill.

16. *J.P.Guilford* - PSYCHOMETRIC METHODS, 1954, Second Edition, McGraw Hill.

17. *J.P.Guilford* - THE PSYCHOPHYSICS OF MENTAL TESTS DIFFICULTY, Psychometrika, 1937, Vol. II, June.

18. *J.P.Guilford* - THE PHI COEFFICIENT AND THE CHI SQUARE AS INDICES OF ITEM VALIDITY, Psychometrika.

19. *Lindquist* - EDUCATIONAL MEASUREMENT, 1935, Washington, American Council on Education.

20. *M. Lorr* - INTER-RELATIONSHIPS OF NUMBER - CORRECT AND LIMEN SCORES FOR AN AMOUNT LIMIT TEST, Psycometrika, Vol. IX nº 1, 1944, March.

21. *J.S.Maritz* - NON LINEAR PROBIT ANALYSIS AND ITS APPLICATIONS TO PSYCHONOMETRIC DATA,Psychonometrika, Vol. XXX, nº 1, 1965, March.

22. *C. Moiser and J.McQuitty* - METHODS OF ITEM VALIDATION AND ABACS FOR ITEM TESTS CORRELATION AND CRITICAL RETIC OF UPPER LOWER DIFFERENCE,Psychonometrika,1940,Vol. V, March.

23. *P. Horst* - ITEM SELECTION BY MEANS OF A MAXIMIZING FUNCTION,Psychonometrika, 1936, Vol.I nº 4, Dec.

24. *E. A. Peel* - ITEM DIFFICULTY AS THE MEASURING DEVICE IN OBJECTIVE MENTAL TESTS, The British Journal of Psycology, 1947, Vol. I.

25. *R.Duncan Luce, R.R.Bush and Eugene Galanter* - HANDBOOK OF MATHEMATICAL PSYCHOLOGY, New York Wiley, 1963.

26. *R. H. Moore and R. K. Zeiglers* - THE USE OF NON LINEAR REGRESSION METHODS FOR ANALYSING SENSITIVITY AND QUANTAL RESPONSE DATA, Biometrics, 1967, September, Vol.234 nº 3

27. *M.W.Richardson* - NOTES ON THE RATIONAL OF ITEM ANALYSIS, Psychometrika, 1936, Vol. I, March.

28. *W.S.Torgerson* - THEORY AND METHODS OF SCALING, 1960, New York John Willey & Sons, Inc.

29. *L.L.Thurstone* - THE MEASUREMENT OF VALUES, Chicago, University of Chicago Press, 1959.

30. *P. E. Vernon* - INDICES OF ITEM CONSISTENCY AND VALIDITY, The British Journal of Psychology, 1947, Vol. I.

31. *Ronald A.Fisher and Frank Yates* - TABLAS ESTADISTICAS - Tradução de Juan Ruiz Magan e Juan José Ruiz Rubio, Aguilar, S.A.de Ediciones, Madrid, 1947.

IMPRESSO NO DEPARTAMENTO DE ESTATÍSTICA DO IME EM 1970.