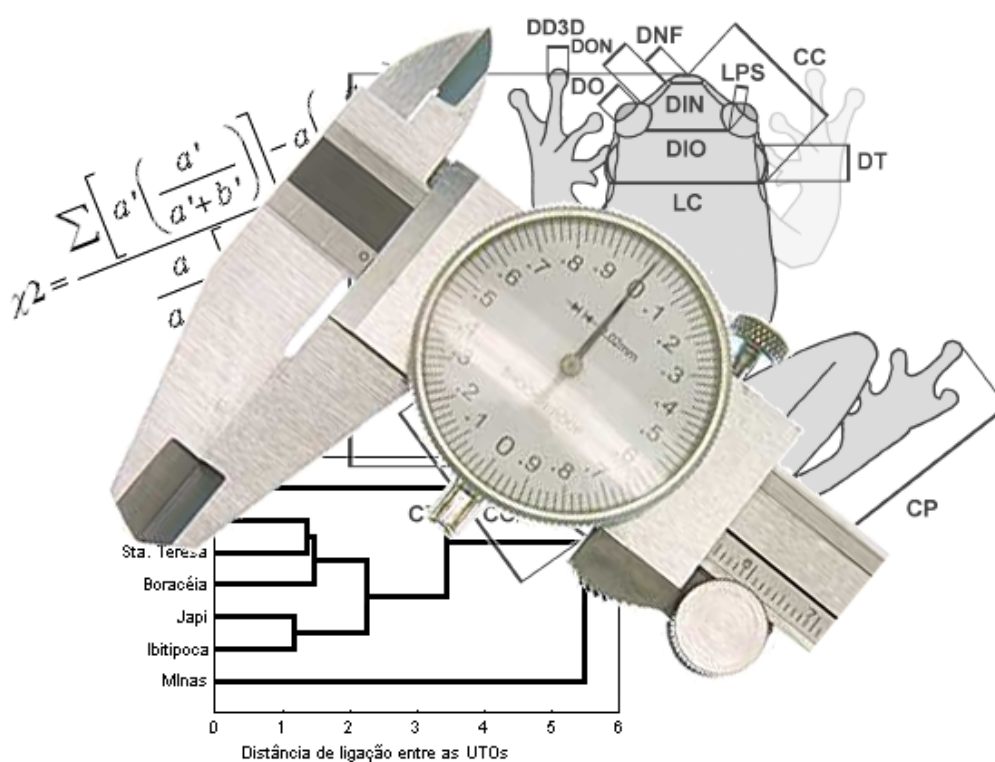


Fundamentos da Prática Estatística em Taxonomia Zoológica

Dr. *Marcelo Felgueiras Napoli*



Apostila destinada ao Mini-curso oferecido no XIV Encontro de Zoologia do Nordeste, sediado na Universidade Federal de Alagoas, Maceió, 12 a 16 de setembro de 2003.

INTRODUÇÃO

Este manual foi concebido com o objetivo de servir de material de apoio para os alunos do Mini-curso intitulado “Fundamentos da prática estatística em taxonomia zoológica”, a ser oferecido durante o XIV Encontro de Zoologia do Nordeste, sediado pela Universidade Federal de Alagoas.

O Curso tem duração de aproximadamente 7h30min, e para um período tão curto, tentei ilustrar alguns aspectos comuns, e outros nem tanto, da aplicação da estatística como apoio à taxonomia zoológica. Embora o tempo de duração do mesmo seja escasso, além da exemplificação de testes e métodos de análise, forneço as computações que têm sido mais apropriadas em meus estudos em taxonomia de anuros. Certamente não estarei propondo novos caminhos, mas apenas indicando para aqueles que desejam iniciar nesta área, não como estatísticos, mas como biólogos, alguns atalhos que demorei alguns anos para descobrir.

Aproveitando o parágrafo anterior, saliento que não sou estatístico e tão pouco planejo me tornar. Sou biólogo de formação e vejo na estatística um instrumento muito útil no auxílio à ordenação, visualização e compreensão dos dados biológicos com que lidamos cotidianamente.

Espero que esta breve compilação de dados, recheadas de experiências pessoais, possa ser útil e assim alcançar meu objetivo principal, que é o de incentivar o uso da estatística, de maneira sóbria e objetiva, no campo da taxonomia zoológica.

EMENTA

Utilização de métodos estatísticos em taxonomia zoológica, obedecendo a uma visão comentada e direta de aspectos tratados rotineiramente nas descrições da diversidade zoológica, desde a organização metodológica do material a ser examinado ao exame das variações intra e interpopulacionais, como variáveis morfológicas, morfométricas e padrões de colorido e desenho.

I ANTES DA ESTATÍSTICA...

1.1 – PROCEDIMENTOS DE ANÁLISE

Os procedimentos de análise da morfologia externa dos espécimes zoológicos aqui adotados assemelham-se ao descrito por VANZOLINI (1970) e HEYER (1984), já tendo sido utilizado e modificado em minha Dissertação de Mestrado (NAPOLI, 1995) e Tese de Doutorado (NAPOLI, 2000), as quais serão em grande parte adotadas como exemplo neste manual. As variações amostrais serão examinadas em cinco etapas consecutivas como linhas gerais da metodologia de trabalho. Tais subdivisões serão adotadas por acreditar serem didáticas, acrescentando-se às mesmas etapas e informações complementares, conforme estabelecido a seguir.

Procedimento 1: *“levantamento dos espécimes disponíveis em coleções zoológicas relevantes”.*

Recomendo que solicitações de empréstimo de numerosos espécimes oriundas de diversas coleções somente sejam efetivadas após o conhecimento primário da diversidade do grupo em questão. Nem sempre é possível e/ou desejável às instituições que abrigam coleções zoológicas emprestarem por longos períodos o material a ser por nós examinado. Um bom conhecimento prévio dos métodos a serem empregados e dos caracteres a serem examinados é fundamental para uma análise complementar mais ágil. Uma sugestão é começar com espécimes recém coletados pelo pesquisador ou oriundos de uma coleção representativa que forneça maior disponibilidade de tempo de empréstimo para esta análise preliminar. Se assim feito, o “levantamento” das coleções que abrigam os exemplares desejados não implicará em empréstimo e análise simultâneos, mas sim no conhecimento prévio de “onde” possam estar os espécimes objetos do estudo, sendo então paulatinamente emprestados e examinados no decorrer do trabalho. Vale ainda ressaltar que este procedimento ainda torna menos dispendiosa a pesquisa, tanto para as instituições, que têm gastos com o envio deste material, quanto para o pesquisador, o qual tem reduzida a quantidade de visitas a instituições que não se prontificaram a emprestá-lo (não é rara a necessidade de vermos espécimes chaves, como holótipos e parátipos, mais de uma vez no decorrer do trabalho).

Procedimento 2: *“listagem das amostras populacionais disponíveis por localidade geográfica e mapeamento das mesmas”.* Para cada localidade registrada deve-se obter valores aproximados de latitude e longitude (até minutos), sendo estas plotadas em mapas regionais. O acesso a estes dados, principalmente às localidades de coleta mais recentes, é fácil e gratuita, como a lista de “Cidades e Vilas” fornecida pelo IBGE para o ano de 1996 ou Gazetteers internacionais (p.ex., pelo NIMA). Este procedimento nos permite, em geral, analisar inicialmente amostras populacionais morfologicamente mais similares entre si para que gradativamente analisemos amostras oriundas de populações mais distantes e de morfologia mais distinta. Possivelmente, hiatos morfológicos poderão ser observados ou mesmo variações morfológicas em gradientes geográficos (“clines” direcionais). Esta noção da variação geográfica é fundamental, já que populações podem ou não manter fluxo gênico entre si, o que pode resultar em isolamentos geográficos e processos de especiação. Contudo, o uso de mapas desprovidos de informação

altimétrica, vegetacional ou mesmo de divisões em domínios morfoclimáticos (AB’SÁBER, 1977), mostram-se muitas vezes insuficientes no auxílio à compreensão da distribuição dos táxons envolvidos, como aqueles restritos a áreas montanhosas. Atualmente, este “problema” não é impedimento para uma boa análise, visto a generosa quantidade de programas computacionais, muitos destes gratuitos, que permitem a visualização e plotagem destes pontos geográficos em mapas topográficos (GIS). Um bom exemplo é o modelo de elevação digital global (DEM) conhecido como GTOPO30, disponibilizado pelo U. S. Geological Survey (USGS) que pode ser trabalhado por vários destes programas computacionais (p.ex., ARCVIEW e KASHMIR). As elevações neste modelo são regularmente espaçadas em aproximadamente 1 (um) km. A precisão deste modelo varia com as fontes de informação utilizadas para cada área, mas em geral é bastante satisfatória para estudos gerais de distribuição geográfica. A América do Sul pertence à fonte conhecida como “Digital Chart of the World” (DCW), cuja precisão vertical varia em aproximadamente 97 m. Uma discussão completa de como se alcançaram estes índices de precisão e outros detalhes técnicos podem ser encontrados em <http://edcwww.cr.usgs.gov/landdaac/gtopo30/readme.html>.

Procedimento 3: *“análise das localidades de maior representatividade em número de espécimes (“amostras básicas”, segundo VANZOLINI, 1970), a fim de se determinar a natureza das variações para machos e fêmeas, separadamente”*. Antes de compararmos as populações entre si quanto a caracteres de forma, desenho e colorido devemos tentar definir os padrões de variação inerentes a uma dada população e somente a partir daí compararmos estas variações com as demais amostras. Será mais proveitoso se começarmos a análise por amostras populacionais mais bem representadas em número de espécimes coligidos, já que estas “deverão” conter uma maior quantidade de variações dos padrões considerados básicos para os caracteres estudados. Quando diferenças morfológicas forem observadas entre amostras geográficas de um determinado táxon, recorre-se à organização destas em Unidades Taxonômicas Operacionais (UTOs), a fim de contrapor-las em análises interpopulacionais subseqüentes.

Procedimento 4: *“análise de localidades com amostragem de menor representatividade em número de espécimes (“amostras pequenas”, de um a poucos exemplares, segundo VANZOLINI, 1970)”*. Uma vez tendo sido analisado as principais amostras partimos para “tentar” encaixar as amostras detentoras de um número de espécimes muito limitado a alguma UTO já definida, tomando o cuidado para não “forçar” tal associação. Não é raro termos em mãos poucos espécimes oriundos de localidades mais distantes geograficamente que não conferem exatamente com nenhuma UTO previamente estabelecida. Neste caso, podemos tentar (1) coletar mais espécimes desta localidade, (2) unir esta amostra a outras igualmente pequenas e de morfologia similar a fim de aumentar o número amostral ou (3), caso as opções 1 e 2 não sejam possíveis, retirar este(s) exemplar(es) da análise preliminar e discutir estas variações no item de discussão do trabalho.

Procedimento 5: *“exame dos padrões de variações entre as UTOs, a fim de se definir os limites de cada população dentro do grupo estudado”*. Uma vez tendo “arrumado” os espécimes disponíveis em UTOs, determinado diferentes padrões para os caracteres de desenho e colorido e tomado as mensurações dos espécimes (caracteres mensuráveis), iremos às análises estatísticas mais pertinentes para tentar **ordenar** este conhecimento, o que facilitará em muito uma visualização dos resultados e sua interpretação. Ressalto que, o exame das variações **dentro das UTOs**, a fim de se verificar variações em **grau** e/ou **tipo** para cada sexo e/ou estágio ontogenético, pode compor duas situações diferentes: (1) onde a UTO é composta por uma ou mais amostras “básicas” e outras “pequenas”; (2) onde a UTO é representada por apenas uma amostra geográfica. Para a **situação 1** faz-se necessário primeiramente um estudo de como os caracteres em estudo (mensuráveis ou não) variam entre as populações, para somente então se analisar possíveis variações ontogenéticas dentro das amostras “básicas” que expliquem

diferentes formas, padrões de desenho ou colorido encontrados dentro e entre as amostras analisadas. A **situação 2** somente nos permite fornecer os dados brutos referentes aos padrões de forma e desenho para cada UTO em separado. Cabe ainda enfatizar que, os critérios para estabelecimento de combinações de padrões e subdivisão de áreas geográficas ao longo de transectos para cada UTO podem variar consideravelmente entre elas, no que recomendo explicá-los junto aos resultados e não no item “Materiais e Métodos”.

1.1.1 – Exemplo comentado 1: organização em morfoespécies (NAPOLI, 1995)

NAPOLI (1995) agrupou os espécimes disponíveis inicialmente em treze **morfoespécies**, levando-se em consideração a semelhança morfológica e a proximidade geográfica.

Morfoespécies: categorias que, se acredita, representam diferentes espécies biológicas (*sensu* MAYR, 1942).

A identificação das morfoespécies foi determinada pela utilização de siglas que representam características morfológicas ou geográficas para as mesmas. Notou-se que tais morfoespécies se aproximavam umas das outras em graus de semelhança diferentes, o que possibilitou um subagrupamento destas, tornando análises *a posteriori* mais fáceis e didáticas. A seguir, são listados dois dos quatro subgrupos propostos pelo autor, onde consta o número de espécimes disponíveis em cada um e suas respectivas localidades amostrais. Note que somente a **Tabela 1** original é aqui fornecida e por motivos ilustrativos. Nesta tabela os padrões foram unidos em categorias e representados em porcentagens, pelos motivos já expostos anteriormente. A **Fig. 1** apresenta os padrões de desenho dorsal.

Subgrupo A

- *Morfoespécie RU, Hyla rubicundula*, n = 144: Alfenas (MG), Andrequicé (MG), Arinos (MG), Barão de Cocais (MG), Barreiras (BA), Belo Horizonte (MG), Buritizeiro (MG), Cristalina (GO), Esmeraldas (MG), Jaboticatubas (MG), Januária (MG), Jupaguá (BA), Buritis (MG), Lagoa Formosa (MG), Lagoa Santa (MG), Manga (MG), Pirapora (MG), Três Marias (MG), Unaí (MG), Vespasiano (MG).
- *Morfoespécie CBO*, “Cachimbo”, n = 15: Cachimbo (PA).
- *Morfoespécie PRU*, “pseudorubicundula”, n = 54: Aragarças (GO), Cavalcante (GO), Goiânia (GO), Iaciara (GO), Monte Alegre de Goiás (GO), Nova Roma (GO), Uruçuí (PI), Porangatu (GO), Santa Rita do Araguaia (GO), São Domingos (GO), escarpa da Serra Dourada (GO), Uberlândia (MG).

• *Diagnose.*— distingue-se do subgrupo B por apresentar os seguintes padrões de caracteres: desenho dorsal do corpo – categorias I, III e IV (Tabela 1); faixa lateral – categoria III; desenho dorsal da tibia – categoria II; contorno da prega supra timpânica / faixa lateral – padrão B; faixas cantal e loreal – categoria III; formato dorsal da cabeça – padrões A1 ao A7 e E. Distingue-se do **subgrupo C** por apresentar os padrões de caracteres: *faixa vertebral – categoria I; faixa lateral – categoria I; desenho dorsal da tibia – categorias I e II ; contorno da prega supra-timpânica / faixa lateral – padrão A; faixas cantal e loreal – categoria I; formato dorsal da cabeça – padrões A1 ao A6 e E.* Distingue-se do subgrupo D por apresentar os padrões de caracteres: desenho dorsal do corpo – categorias I à IV; faixa lateral – categorias II e IV; desenho dorsal da tibia – categoria III; faixas cantal e loreal – categoria II à IV; formato dorsal da cabeça – padrões A1 ao A6 e E.

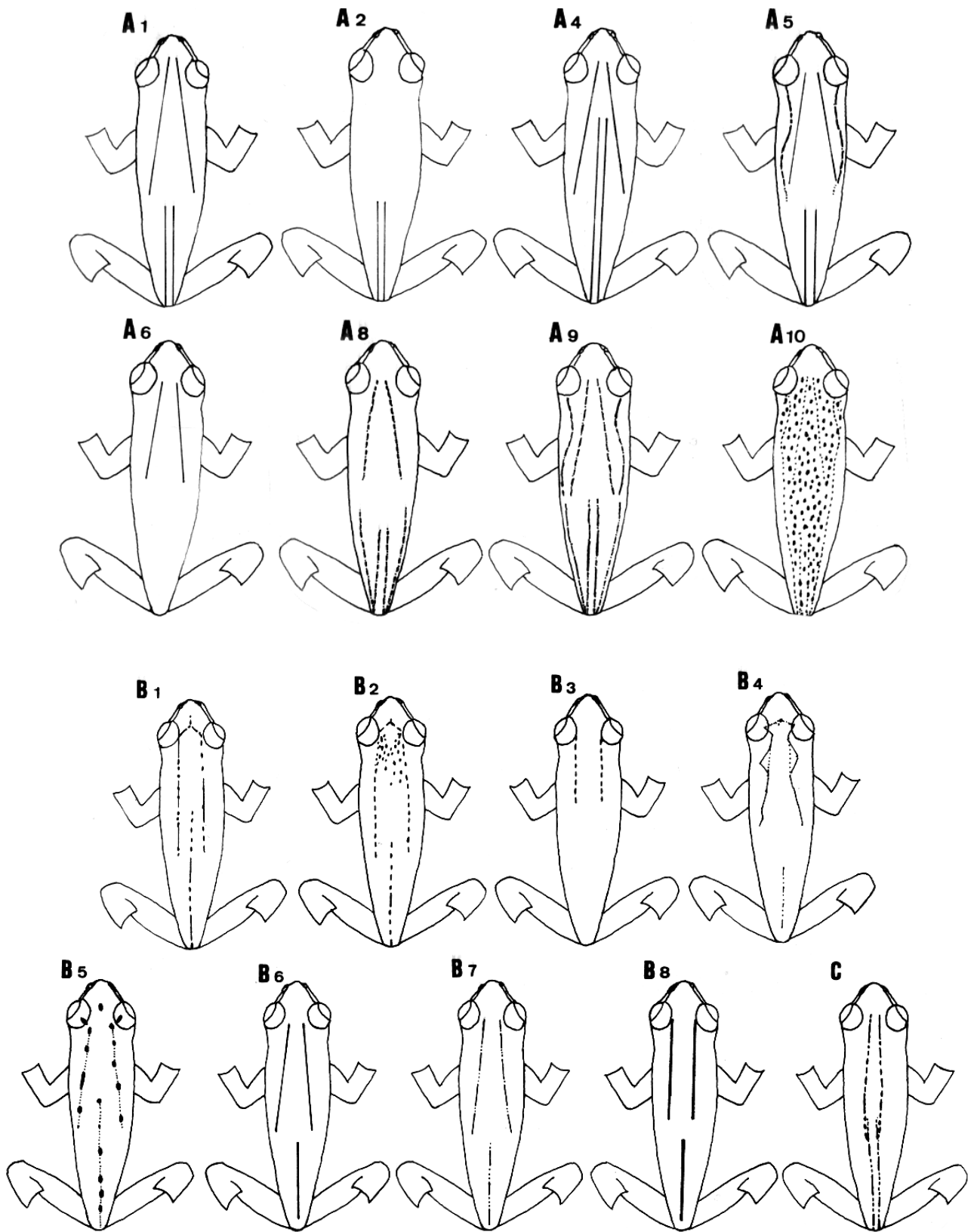


Fig. 1 – Padrões de desenho dorsal nas morfoespécies do grupo de *Hyla rubicundula* Reinhardt & Lütken, 1862 (NAPOLI, 1995).

Observe na **Tabela 1** que as categorias I a IV diagnosticam os subgrupos A ao C de D; a categoria XI (padrão C) diagnostica o subgrupo B (*Hyla anataliasiasi*); as categorias V a X diagnosticam o subgrupo D; ainda, as categorias I, III, e IV diagnosticam os subgrupos A e C de B. Outras conclusões podem ser tomadas para as morfoespécies dentro do subgrupo D. Todavia, as **morfoespécies RU e PRU no subgrupo A** não apresentam padrões exclusivos quando comparadas entre si. Logo, variam em **grau** (porcentagem de cada padrão) mas não em **tipo**. NAPOLI (1995) notou que o mesmo acontece para os demais caracteres analisados.

Note que uma tabela com dados ordenados, como a em questão, nos permite interpretar preliminarmente os dados disponíveis e já tirar certas conclusões, mesmo antes de realizar qualquer teste estatístico.

Tabela 1. Distribuição dos padrões de desenho dorsal do corpo (**em porcentagem**) entre as treze morfoespécies; os padrões estão definidos na Figura 2 [Fig. 1 deste manual] [combinados aqui em categorias] (n = número de espécimes examinados; um espaço vazio indica que nenhum espécime apresentou tal padrão; os algarismos em romano indicam categorias formadas pela reunião de padrões: I = A1 + A4; II = A2 + A6; III = A5 + A8 + A9; IV = A10; V = B1 + B2 + B3; VI = B4; VII = B5; VIII = B6 + B7; IX = B8; X = B9; XI = C; XII = A7 + A3).

Subgrupo	Morfoespécie	n	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
A	RU	130	53,0	12,3	22,3	0,7								11,5
	PRU	48	37,5	10,4	25,0	2,1								25
	CBO	15												100
B	ANA	81		14,8									63,4	61,7
C	SPA	42	2,3	23,8	4,7	2,3								66,6
	RUM	34	5,8	35,9	2,9	8,8								47
	RDI	65	1,5		1,5									96,9
D	TT	3					57,7				100			
	TTBotu	45								79		6,6		35,5
	TTPira	43						9,3		50				11,6
	TTMG	4					25,0							25,0
	TTRRP	6							71,4					28,5
	RRARA	16					43,7			37,5	6,2	6,2		6,25

1.1.2 – Exemplo comentado 2: Unidades Taxonômicas Operacionais (UTOs)

NAPOLI (2000) utilizou unidades taxonômicas operacionais ao invés de morfoespécies. Em termos práticos não houve diferença. Contudo, as UTOs não trazem implícita a idéia de que as amostras “são categorias que, se acredita, representam diferentes espécies biológicas”. Por esse motivo em particular, julgo mais pertinente a utilização de UTOs, principalmente naquelas análises onde comparamos amostras geograficamente distintas mas que acreditamos pertencerem à mesma espécie. O trecho abaixo foi retirado de minha tese de doutorado e resume a idéia aqui exposta:

“[...] A observação dos espécimes de *Hyla circumdata* provenientes de amostras geográficas distintas ao longo das serras do Mar, Mantiqueira, Quadrilátero Ferrífero e Espinhaço central (Fig. 11) [Fig. 2, neste trabalho], revelou serem altas as variações intra e interpopulacionais deste táxon. Além disso, a aquisição e análise dos dados morfológicos que diagnosticam *H. circumdata* mostrou que os caracteres correntemente utilizados na distinção destes táxons são altamente variáveis, principalmente quando a análise obedece a certos gradientes geográficos. A fim de analisar as variações dentro e entre as amostras geográficas de *H. circumdata*, organizei-as em **Unidades Taxonômicas Operacionais (UTOs)**, quando compostas por três ou mais espécimes adultos em bom estado de conservação, sendo as análises realizadas separadamente em machos e fêmeas. O número de UTOs organizadas para as fêmeas foi inferior ao de machos, mas seus posicionamentos geográficos aproximadamente coincidentes. São elas:

- (1) UTO Rio (Fig. 2C) = Petrópolis e Teresópolis (Serra dos Órgãos) e Tijuca (Maciços Costeiros), Rio de Janeiro, no Estado do Rio de Janeiro. Planalto da Serra do Mar propriamente dito.
- (2) UTO Parati = Mangaratiba e Parati, no Estado do Rio de Janeiro, e Picinguaba, no Estado de São Paulo. Planalto da Serra do Mar propriamente dito.
- (3) UTO Bocaina = Campo de Fruticultura, São José do Barreiro, São Paulo. Planalto da Bocaina, complexo Serra do Mar.
- (4) UTO Boracéia = Estação Biológica de Boracéia, São Paulo. Entre os planaltos da Bocaina e de Paranapiacaba, no complexo Serra do Mar.
- (5) UTO Japi (Fig. 2B) = Jundiaí (Serra do Japi), Capão Bonito (fazenda Intervales), Ribeirão Branco e São Miguel Arcanjo, no Estado de São Paulo. Contrafortes da Serra da Mantiqueira e Planalto de Paranapiacaba, complexo Serra do Mar.
- (6) UTO Itatiaia = Itatiaia, Rio de Janeiro e Piquete, São Paulo. Setor meridional da Serra da Mantiqueira, Planalto de Itatiaia.
- (7) UTO Ibitipoca = Conceição do Ibitipoca, Minas Gerais. Setor meridional da Serra da Mantiqueira, Planalto de Itatiaia.
- (8) UTO Espinhaço = Belo Horizonte, Nova Lima, Ouro Preto e Barão de Cocais, no Estado de Minas Gerais. Setor meridional da Serra do Espinhaço e Planalto e serras do Quadrilátero Ferrífero.
- (9) UTO Sta. Teresa (Fig. 2A) = Araponga (Serra do Brigadeiro), Estado de Minas Gerais; Domingos Martins, Serra do Caparaó e Santa Teresa, no Estado de Espírito Santo. Setor setentrional da Serra da Mantiqueira.

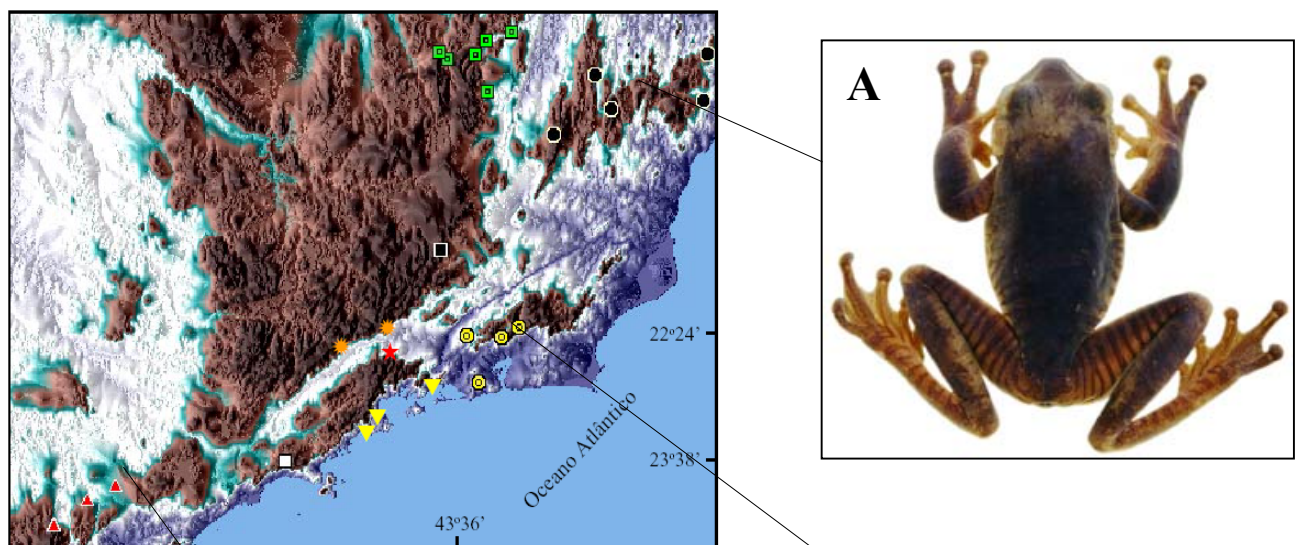
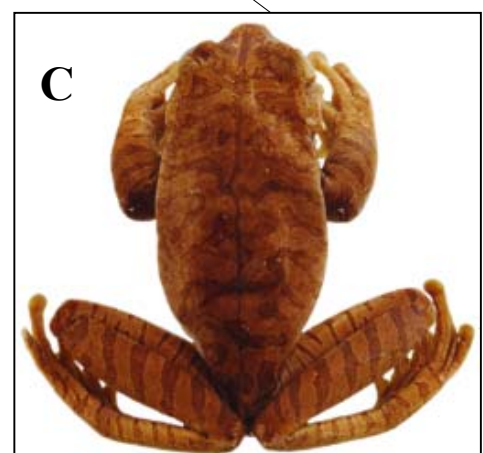
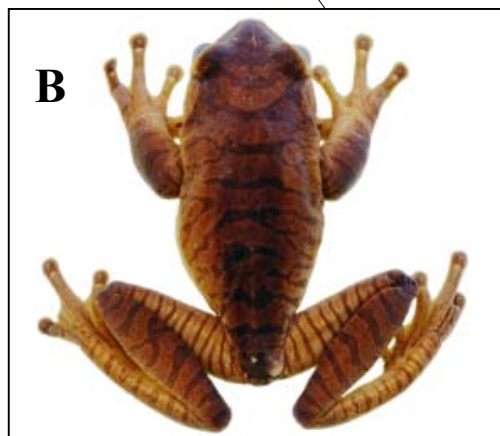


Fig. 2 . Distribuição geográfica em mapa topográfico (GTOPO30) das UTOs Rio (⊙), Parati (▼), Bocaina (★), Boracéia (□), Japi (▲), Itatiaia (●), Ibitipoca (■), Sta. Teresa (●) e Espinhaço (▣), complexo de espécies de *H. circumdata*/*H. luctuosa*. A altimetria está representada pelo seguinte padrão de cores: azul escuro = 0 m; branco = 600 m; verde = 700 m; marrom = 800 m; preto = 1.000 m; tonalidades de cores intermediárias representam variação altimétrica gradual.



1.2 – TESTES ESTATÍSTICOS

Embora nosso curso não objetive ser um curso de bioestatística, e sim o de como utilizar certos testes estatísticos no auxílio à resolução de determinadas perguntas no estudo da taxonomia zoológica, teremos de entender alguns conceitos elementares em estatística.

Os **caracteres** ou **variáveis** são classificados de acordo com as operações a que podem ser submetidos (VANZOLINI, 1993). As seguintes variáveis são as mais utilizadas no tipo de estudo em questão:

- (A) **Variáveis de escala nominais** ou **variáveis categóricas**. Por estas variáveis, os espécimes zoológicos são inseridos em categorias mutuamente exclusivas e não podem ser objeto de mensurações ou ordenações. É onde usaremos testes estatísticos não paramétricos (explicados no Capítulo II). O exemplo anterior envolvendo “padrões de desenho” se enquadra neste tipo de variável.
- (B) **Variáveis de escala geométrica** (ou **variáveis quantitativas**). Podem ser subdivididas em: (B.1) **variáveis discretas**, onde somente valores positivos inteiros são possíveis (p.ex., contagem de escamas de uma cobra) e (B.2) **variáveis contínuas**, valores positivos quaisquer (p.ex., medições). É onde usaremos testes estatísticos paramétricos (explicados no Capítulo III).
- (C) **Variáveis de escala ordinal**. São as variáveis que podem ser “ranqueadas”, isto é, ordenadas em valores crescentes ou decrescentes. Um bom exemplo é a variação de tonalidade de cor e forma de uma estrutura. É onde usaremos testes estatísticos não paramétricos (citados nos Capítulos II e III).

II ANÁLISE DOS DADOS NÃO MORFOMÉTRICOS

Os caracteres de morfologia externa e os padrões de desenho e colorido são preferencialmente avaliados individualmente, **um por vez**, antes mesmo de serem estabelecidas as UTOs, sendo então configurados padrões de variação, imediatamente esquematizados. Se assim feito, certamente teremos maior critério na escolha dos padrões (evitando padrões muito similares) e na inclusão dos espécimes observados dentro dos mesmos. Recomendo ainda que cada padrão seja caracterizado também na forma escrita, com o propósito de se descrever as características mais diagnósticas entre eles. Nos espécimes que se encontrarem parcialmente descolorados ou danificados somente devem ser considerados os caracteres que se apresentarem bem nítidos. Este procedimento poderá levar a tamanhos amostrais diferentes entre os caracteres examinados, o que não representa qualquer impedimento à análise.

Após este procedimento, os espécimes serão reunidos em grupos, pela forma e/ou desenho e colorido semelhantes. Os resultados devem ser então apresentados em **tabelas e gráficos** comparativos. Quando não há limitação de número de páginas (p.ex., monografias, dissertações de mestrado e teses de doutorado) podemos apresentar estas tabelas de duas maneiras diferentes: (1) no corpo do trabalho, onde forneceremos os dados em porcentagem, a fim de se eliminar o efeito do tamanho das amostras para cada caráter em questão, (1.1) para cada padrão individualmente ou, quando necessário, (1.2) através de combinações entre os padrões muito semelhantes, a fim de tornar a interpretação dos dados menos confusa e/ou aumentar o número amostral, como exigem alguns testes estatísticos (**Tabela 1**); (2) nos apêndices, os resultados devem ser fornecidos sem quaisquer alterações, em valores brutos, a fim de possibilitar uma análise posterior independente de considerações prévias.

2.1 – Exemplo comentado 3: NAPOLI (1995), modificado em NAPOLI & CARAMASCHI (1999)

Na análise do caráter “faixa dorsolateral” em *Hyla rubicundula*, seis padrões foram encontrados (**Fig. 3**). Este caráter consiste de duas faixas, uma clara superior e outra escura inferior, que se estendem desde o bordo posterior do olho até a virilha, podendo ou não ocorrer simultaneamente, além de variarem em intensidade e coloração.

- **Padrão C1** – faixa fina, de coloração castanha, contínua, que não sugere pontilhamento; inicia-se no bordo posterior do olho, ultrapassa superiormente o tímpano, sem contorná-lo, e segue até próximo à cloaca; uma faixa mais clara que a cor de fundo e esbranquiçada margeia superiormente a faixa escura.
- **Padrão C2** – semelhante ao padrão C1, sendo que a faixa clara é vestigial ou ausente.
- **Padrão C3** – faixa escura mais espessa e fragmentada que a dos padrões anteriores (sugere pontilhamento); inicia-se no bordo posterior do olho, ultrapassa superiormente o

tímpano, podendo ou não contorná-lo, e segue até próximo à cloaca. Presença de faixa clara marginando superiormente a faixa marrom escura, mas pouco discernível da cor de fundo dorsal, com tonalidade rosa-claro.

- **Padrão C4** – semelhante ao padrão C3, sendo que **não** apresenta a faixa clara marginando superiormente a faixa escura.
- **Padrão C5** – faixa escura vestigial, delineada por fina granulação, que se estende desde o bordo posterior do olho até próximo à cloaca, contornando ou não o tímpano; a faixa clara superior está ausente.
- **Padrão C6** – ausência completa de faixas escura e clara.
- **Padrão C7** – faixa escura dorsolateral quase preta, bem marcada e faixa sobrejacente bem marcada, quase branca.

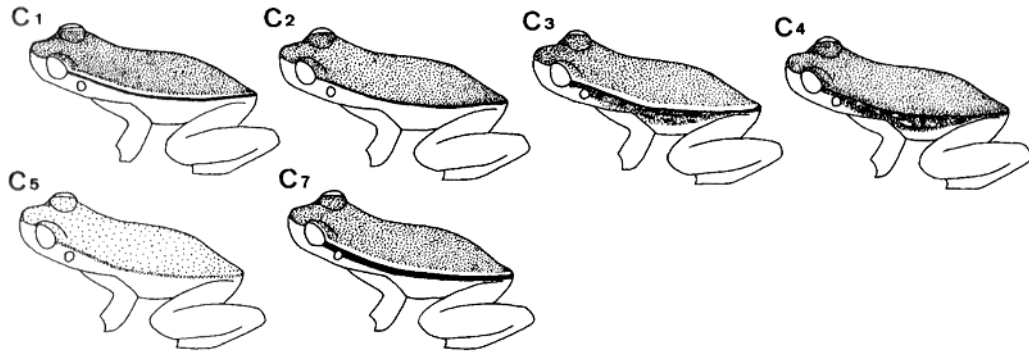


Fig. 3 – Padrões de faixa dorsolateral em morfoespécies do grupo de *H. rubicundula* (NAPOLI, 1995, modificado em NAPOLI & CARAMASCHI, 1999).

➤ **QUESTÃO:** Analisando-se todas as amostras populacionais da morfoespécie PRU (Pág. 5), simultaneamente, no Estado de Goiás, teremos estes padrões variando em **mosaico** ou variando **ao longo de gradientes geográficos** (“clines” direcionais)?

Procedimento 1: NAPOLI (1995) tomou as localidades amostrais São Domingos e Santa Rita do Araguaia, pontos limites de ocorrência desta morfoespécie, para o traçado de um **transecto** de direção Norte-Sul (Fig. 4).

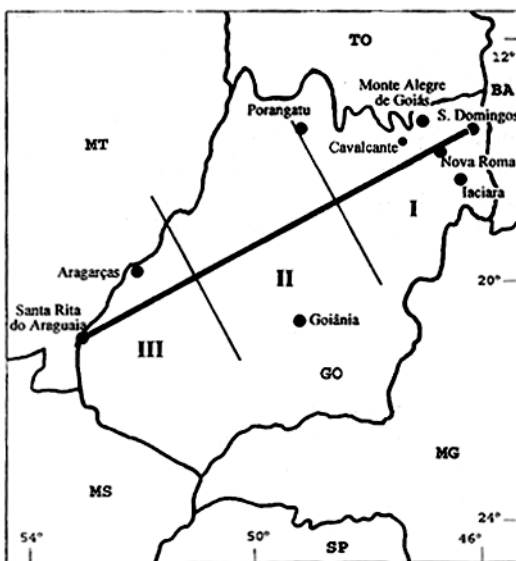


Fig. 4 – Distribuição da morfoespécie PRU no Estado de Goiás.

Procedimento 2: ao longo deste transecto foram **delimitadas arbitrariamente três áreas**, aproximadamente equidistantes, contendo todas as amostras disponíveis. Os padrões de distribuição das variações padronizadas para cada caráter foram então examinados em cada área e depois comparados.

Procedimento 3: a fim de diminuir a quantidade de padrões de faixa dorsolateral para não violar certas exigências de testes estatísticos e facilitar a compreensão das variações observadas, o autor uniu padrões de desenho, seguindo os seguintes critérios: C1 e C2, semelhantes e típicos da população topotípica de *H. rubicundula* (Lagoa Santa, MG); C3 e C4, semelhantes e típicos das amostras de Goiás; C5 e C6 vestigiais ou ausência de faixas.

Procedimento 4: a fim de eliminar o efeito do tamanho das amostras, o autor transformou os valores originais de cada padrão encontrado em porcentagens, o que permite comprar as distribuições de cada padrão entre as áreas a serem comparadas (I a III).

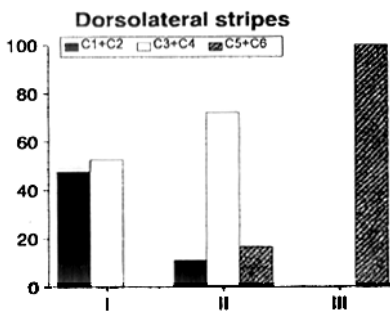


Fig. 5 – Frequência (em porcentagem) dos padrões obtidos na morfoespécie PRU para o caráter Faixa dorsolateral.

Procedimento 5: os valores em porcentagem por cada área foram então representados em um **gráfico de barras** (Fig. 5). A visualização dos dados em gráficos é muito mais facilmente interpretada do que nas tabelas originais e devem ser preferidos a estas.

Procedimento 6: a fim de se verificar se havia diferença estatística significativa entre as áreas analisadas, o autor recorreu ao teste estatístico do **qui-quadrado** (χ^2).

O gráfico da **Figura 5** mostra claramente que este caráter tem sua variação ao longo de um gradiente geográfico, e não em mosaico, da área I à área III. Um “cline” direcional é marcado pelo progressivo desaparecimento da faixa branca dorsolateral em Goiás, do norte para o sul. Note que a diferença encontrada entre as áreas foi principalmente em **GRAU** e não em **TIPO**, o que não nos permite discriminar de maneira plausível estas amostras.

Os testes estatísticos de associação denotaram diferença significativa ($P < 0,05$) da área I em relação às demais, mas não da área II com a III. Para considerarmos os resultados como “ótimos” quanto à existência de um “cline”, é desejável a discriminação das três áreas. Podemos assim inferir, ao menos preliminarmente, que a variação ocorre ao longo de um gradiente geográfico norte sul, e estas populações podem pertencer à mesma espécie biológica, já que “parece haver fluxo gênico” entre elas.

No exemplo, utilizamos testes estatísticos de associação para comparar os padrões encontrados entre as áreas I, II e III e tentar rejeitar a hipótese nula (de que todas pertencem à mesma população). **Mas, quais os testes que poderiam ser usados?**

Já que nosso problema, por enquanto, se limita ao tipo encontrado no **exemplo 3**, vamos entender então como trabalhamos, em geral, com **variáveis categóricas**. Neste caso especificamente, o que mais nos importa é entender o que são **tabelas de contingência** e por conseguinte, começamos por certos tipos de **testes não paramétricos**.

2.2 – Variáveis de escala nominais ou variáveis categóricas (testes não paramétricos)

2.2.1 – Tabelas de contingência:

São tabelas que permitem a análise estatística da **associação** entre **variáveis categóricas** e são freqüentemente usadas em testes não paramétricos.

A situação mais simples é quando temos uma tabela, de dupla entrada, com duas variáveis categóricas, cada uma com duas classes, lembrando que as classes são mutuamente exclusivas (independentes).

No exemplo anterior, temos três categorias de padrões (C1+C2; C3+C4; C5+C6) e três áreas (I, II e III), isto é, três variáveis categóricas, cada uma com três classes.

Desejamos verificar se existe variação geográfica neste caráter entre as três áreas.

Antes de resolvermos este problema, vamos treinar um pouco com outro grupo de dados do mesmo trabalho, descobrindo se há diferença estatística significativa entre as duas morfoespécies (RU e PRU) para o caráter “faixa dorsolateral”. Deste modo, **você** poderá resolver sozinho este problema.

2.2.1.1 – Tabela de frequências observadas

A **Tabela 2** mostra os valores originais obtidos de cada padrão para o caráter “faixa dorsolateral” nas morfoespécies RU e PRU.

Tabela 2. Faixa dorsolateral. <i>n</i> , número de espécimes. C1 a C7, padrões de faixa dorsolateral; valores entre parêntesis, porcentagem de ocorrência.							
Morfoespécies	<i>n</i>	C1	C2	C3	C4	C5	C6
RU	132	83 (62)	49 (37)				
PRU	46	5 (10)	8 (17)	15 (32)	10 (21)	3 (6)	5 (10)

Em uma tabela 2x2, montamos a tabela das frequências observadas da seguinte maneira (Tabela 3):

Tabela 3. Modelo de tabela 2 x 2.			
	1ª categoria	2ª categoria	Totais
1ª categoria	a	b	a + b
2ª categoria	c	d	c + d
Totais	a + c	b + d	a + b + c + d = n

Neste caso, se comparamos as morfoespécies RU e PRU, com somente os quatro primeiros padrões (neste caso duas categorias apenas), teremos a seguinte tabela 2x2 com as **frequências obtidas**:

Tabela 4			
	C1+C2	C3+C4	Totais
RU	132	0	132
PRU	13	25	38
Totais	145	25	170

Para testarmos se as frequências obtidas em RU e PRU têm alguma associação, é necessário determinar primeiro quais seriam as frequências quando não há associação, isto é, quando são **independentes**. Em outras palavras, precisamos determinar quais são as **frequências esperadas**. A fórmula para se achar as frequências esperadas é fácil.

Para cada célula, basta multiplicar os totais da linha (a+b ou c+d) pela coluna (a+c ou b+d) que esta ocupa e dividir pela frequência total (a+b+c+d).

Cálculo: $132 \times 145 / 170 = 112,58$; $38 \times 145 / 170 = 32,41$; $132 \times 25 / 170 = 19,41$; $38 \times 25 / 170 = 5,58$.

Assim sendo, temos a seguinte tabela com os valores obtidos e esperados (entre parêntesis):

Tabela 5			
	C1+C2	C3+C4	Totais
RU	132 (112,58)	0 (19,41)	132
PRU	13 (32,41)	25 (5,58)	38
Totais	145	25	170

É pouco provável, como vimos acima, que os valores obtidos e esperados sejam exatamente iguais, ou mesmo muito próximos. O que precisamos determinar é a probabilidade dos valores obtidos terem sido extraídos, ao acaso, da mesma população, onde os valores “corretos” seriam os representados pelas frequências esperadas, ou seja, julgar a significância das discrepâncias entre as frequências observadas (f_o) e esperadas (f_e).

Via de regra, o próximo passo consiste em testarmos a hipótese de nulidade.

Hipótese nula (H_0): as amostras foram retiradas de uma mesma população estatística, onde as diferenças observadas devem-se meramente a variações intrapopulacionais (flutuações de amostra).

Cabe rejeitar ou não a hipótese nula. Se os valores obtidos tiverem uma probabilidade (P) baixa (for pequena a probabilidade de serem retirados de uma mesma população; p.ex., $P < 0,001$), **recusamos a hipótese nula**. Se a probabilidade for alta (p.ex., $P < 0,9$), **não rejeitamos a hipótese nula**. O **nível de significância** tradicionalmente aceito é de 5% ($P \leq 0,05$), isto é, a porcentagem de erro máxima aceita, o que significa uma probabilidade igual ou maior que 95% da decisão estar correta. Em geral, uma **$P \leq 0,05$** é significativa e **$P \leq 0,01$** é altamente significativa.

O teste mais tradicionalmente utilizado quando trabalhamos com tabelas de contingência é o **teste do qui-quadrado** (χ^2), mas que por vezes deve ser substituído por outros, como o **teste exato de Fisher**. SOKAL & ROHLF (1981: 735) ressaltam que os métodos utilizados para se trabalhar com estes tipos de dados, como o **G-teste**, **teste exato de Fisher** e **qui-quadrado**, acabam por levar a resultados similares, o que, felizmente, acaba por não invalidar vários trabalhos já publicados com os testes indevidos. Além disso, atentam que o critério de escolha deve ser o **tipo de modelo** (I, II ou III, segundo os autores: ver mais adiante), baseado no “*design*” do experimento ou procedimento de coleta de dados, e não no n amostral ou a facilidade destes testes. O **tipo de modelo** se baseia na questão se os totais marginais em uma tabela de dupla entrada (contingência) foram fixados pelo investigador ou são livres para variar e refletem parâmetros populacionais. A **Tabela 6** mostra as recomendações destes autores para tabelas 2x2. Nos exemplos aqui trabalhados, o número total de espécimes amostrados é estabelecido a priori pela disponibilidade total das amostras, mas são livres da vontade do pesquisador quanto a pertencerem a uma determinada morfoespécie e qual(ais) padrão(ões) a eles correspondentes, o que nos leva ao Modelo I de SOKAL & ROHLF (1981). Contudo, o **G-teste** (Modelo I) e o **teste exato de Fisher** (Modelo III) foram designados originalmente para atender a modelos específicos. Já o **qui-quadrado**, não foi designado para atender a nenhum modelo específico, sendo o mais antigo para tabelas 2x2, o mais genérico e também o mais usado. Outros autores bem populares na biologia, como CENTENO (1982) e VANZOLINI (1993), definem parâmetros mais precisos para o uso do χ^2 e **teste exato de Fisher**, justamente se baseando no tamanho amostral e frequências esperadas, o que é justamente contestado por SOKAL & ROHLF (1981), onde o tipo de modelo, e não o n amostral, deve ser considerado.

Tabela 6. Modelos e testes estatísticos propostos por SOKAL & ROHLF (1981: 735).		
Modelos	Totais marginais	Teste recomendado
I	Não fixados	G-teste de independência
II	Fixado para um critério	G-teste de independência
III	Fixado para ambos os critérios	Teste exato de Fisher

2.2.1.2 – Teste de independência G (Wilks)

É o teste mais recomendado por SOKAL & ROHLF (1981: 753) (para os modelos I e II) e especialmente útil na computação de tabelas com várias linhas e colunas (VANZOLINI, 1993). Os primeiros autores recomendam ainda a utilização da correção de Williams.

O modelo de tabela de contingência é o já comentado (Tabela 3) e a fórmula para uma tabela 2x2 é a seguinte (SOKAL & ROHLF, 1981: 744):

$$G = 2[(\sum f \ln f \text{ para as freqüências das células}) - (\sum f \ln f \text{ para os totais das linhas e colunas}) + n \ln n]$$

A correção de Williams para uma tabela 2x2 é a seguinte (SOKAL & ROHLF, 1981: 737):

$$q = 1 + \frac{\left(\frac{n}{a+b} + \frac{n}{c+d} - 1\right) \left(\frac{n}{a+c} + \frac{n}{b+d} - 1\right)}{6n}$$

Em tabelas com n linhas e n colunas (testes de independência $R \times C$) usaremos a mesma computação acima descrita, aqui desmembrada a fim de facilitar a demonstração:

$$G = 2 (\sum (f \ln f) - \sum (C \ln C) - \sum (R \ln R) + n \ln n)$$

onde,

$\sum (f \ln f)$ é o somatório dos produtos da freqüência de cada célula por seu logaritmo natural;

C é cada total de coluna;

R é cada total de linha;

n é o número total de observações.

A correção de Williams também pode ser empregada, mas é mais difícil de computar manualmente. Assim sendo, SOKAL & ROHLF (1981: 744) fornecem uma fórmula simplificada e a recomendam quando a amostra for pequena e o valor de G estiver à margem do nível de significância.

$$q_{\min} = 1 + \frac{(a^2 - 1)(b^2 - 1)}{6n(a-1)(b-1)}$$

onde, a é o número de colunas, b é o número de linhas, n é o número total de observações.

Tanto nas tabelas 2x2 quanto em tabelas com n linhas e n colunas ($R \times C$) o valor original de G será dividido pelo valor ajustado q e o quociente conferido em uma tábua de χ^2 (Anexo, **Tabela 1**), onde os graus de liberdade são calculados com o seguinte critério: $GL = (l - 1)(c - 1)$.

No exemplo da **Tabela 2** ($R \times C$) temos a seguinte situação quando unimos os padrões em categorias, como se segue:

Tabela 7. Distribuição dos padrões de faixa dorsolateral na morfoespécie PRU. n , número de espécimes. C1 a C6, padrões de faixa dorsolateral, já agrupados 2 a 2 em categorias por critério de semelhança. Entre parênteses, freqüências esperadas.				
Morfoespécies	C1+C2	C3+C4	C4+C5	Totais
RU	132 (107,52)	0 (18,53)	0 (5,93)	132
PRU	13 (37,47)	25 (6,46)	8 (2,06)	46
Totais	145	25	8	178

(A) $\sum f \ln f$ para as frequências das células:

$$= 132 \ln 132 + 0 \ln 0 + 0 \ln 0 + 13 \ln 13 + 25 \ln 25 + 8 \ln 8$$

$$= 774,96$$

(B) $\sum f \ln f$ para os totais de colunas:

$$= 145 \ln 145 + 25 \ln 25 + 8 \ln 8$$

$$= 721,62 + 80,47 + 16,63$$

$$= 818,72$$

(C) $\sum f \ln f$ para os totais de linhas:

$$= 132 \ln 132 + 46 \ln 46$$

$$= 644,52 + 176,11$$

$$= 820,63$$

(D) $n \ln n = 178 \ln 178 = 922,35$

$$G = 2 (A - B - C + D)$$

$$G = 2 (774,96 - 818,72 - 820,63 + 922,35)$$

$$G = 2 (57,96)$$

$$G = \mathbf{115,92}$$

Correção de Williams:

$$q_{\min} = 1 + \frac{(9-1)(4-1)}{6(178)(2)(1)} = 1,0112$$

G ajustado = $G/q = 115,92 / 1,0112 = \mathbf{114,63}$ (pela elevada significância do valor G , o ajuste não era necessário, tendo sido sendo feito como exemplo)

Através da tábua de χ^2 (Anexo, **Tabela 1**), verificamos que, para um $Gl = 2$, temos um $P < 0,001$, ou seja, é altamente significativo e rejeitamos a hipótese nula (as amostras têm uma probabilidade muito baixa de terem sido retiradas da mesma população).

2.2.1.3 – Teste do qui-quadrado (χ^2)

$$\chi^2 = \sum ((f_o - f_e)^2 / 2)$$

No exemplo da **Tabela 5** (2 x 2), teremos então:

$$\chi^2 = ((19,42)^2 / 112,58) + ((-19,41)^2 / 19,41) + ((-19,41)^2 / 32,41) + ((19,42)^2 / 5,58) / 2$$

$$\chi^2 = 3,34 + 19,40 + 11,62 + 67,58$$

$$\chi^2 = \mathbf{101,94}$$

O próximo passo é consultar a tábua de χ^2 a fim de observarmos a respectiva probabilidade, o que nos permitirá rejeitar ou não a hipótese de nulidade. Para tal necessitamos definir

primeiramente com quantos **graus de liberdade (GL)** estamos trabalhando. Use para tal a seguinte fórmula:

$$n = (l - 1)(c - 1)$$

onde, n = o número de graus de liberdade,

l = o número de linhas da tabela de contingência,

c = número de colunas.

Observando-se a **tábua de χ^2** (Anexo, **Tabela 1**) obtemos para um $\chi^2 = 101,94$, com **GL = 1**, a probabilidade **P<0,001**, o que significa que é muito pequena a probabilidade das amostras terem sido retiradas de uma mesma população, o que **rejeita a hipótese nula**.

Uma maneira de se calcular o qui-quadrado sem calcular os valores esperados (f_e) foi fornecido em SIMPSON (1939), SOKAL & ROLHF (1981) e VANZOLINI (1993), como se segue:

$$\chi^2 = (ad - bc)^2 n / (a+b)(c+d)(a+c)(b+d)$$

Contudo, a tabela original dos dados de “faixa dorsolateral” contém três valores categóricos de padrões de desenho. Isto fará nossa tabela de contingência ser 2x3. Felizmente, o processo de cálculo será o mesmo, embora mais trabalhoso. SIMPSON (1939: 295) nos fornece o “Ratio Method” mais facilmente computado. Na **Tabela 7**, reproduzida abaixo já com os valores a serem calculados, teremos então:

	C1+C2	C3+C4	C5+C6	Totais
RU (a')	132 (107,52)	0 (18,53)	0 (5,93)	$a = 132$
PRU (b')	13 (37,47)	25 (6,46)	8 (2,06)	$b = 46$
Totais ($a'+b'$)	145	25	8	$a + b = 178$
$\left(\frac{a'}{a'+b'} \right)$	0,91	0	0	$\frac{a}{a+b} = 0,74$
$a' \left(\frac{a'}{a'+b'} \right)$	120,21	0	0	$a \left(\frac{a}{a+b} \right) = 97,68$

onde, as incógnitas estão explicadas na tabela acima, nas próprias células.

Aplicando-se esta fórmula teremos:

$$\chi^2 = \frac{120,21 - 97,68}{0,74 \times 0,26} = \frac{22,53}{0,19} = 118,57$$

Observando-se a **tábua de χ^2** , para um **GL = 2**, teremos um valor **P<0,001**, ou seja, altamente significativo, o que denota que as amostras não foram tomadas ao acaso da mesma população, rejeitando-se a hipótese nula. **Note que o valor do χ^2 é muito próximo ao de G, como já salientado por SOKAL & ROHLF (1981: 735).**

IMPORTANTE: Não existe um trabalho extenso sobre qual é o tamanho mínimo de uma amostra para que esta possa ser usada no teste de **G** (SOKAL & ROHLF, 1981: 709). Contudo, existem procedimentos tradicionais, recomendações, em relação às frequências esperadas (f_e) e número amostral. Além disso, o baixo número de exemplares em algumas amostras pode exigir uso de correções, como já visto anteriormente. Abaixo, seguem algumas recomendações tradicionais.

- (a) No caso do teste G , a menor f_e deve ser igual ou maior do que cinco. Quando esta regra for violada, geralmente se unem classes a fim de que a menor f_e seja maior do que cinco (SOKAL & ROHLF, 1981: 709). Contudo, veja a recomendação **b**, abaixo.
- (b) Em tabelas de contingência com mais de 1 grau de liberdade, onde 20% ou mais das frequências esperadas forem maiores do que 5 (ou menor do que o número de graus de liberdade), não haverá necessidade de se unir variáveis adjacentes para a realização do teste do χ^2 . Contudo, caso esta regra seja violada, as variáveis adjacentes deverão ser unidas de modo que a menor f_e seja maior do que cinco (SIMPSON, 1939: 296; CENTENO, 1982: 164). Todavia, para qualquer caso, o χ^2 só poderá ser usado se a frequência menor (f_e) não for menor do que 1 (CENTENO, 1982: 164).
- (c) Em tabelas de contingência com apenas 1 grau de liberdade (2x2) onde a frequência total n for $20 < n \leq 40$ e todas as frequências esperadas foram iguais ou superiores a 5, aplicar-se-á a *correção de Yates* ao teste do χ^2 (CENTENO, 1982: 164). **Contudo, GRIZZLE (1967), apud SOKAL & ROHLF (1981: 743), demonstrou que a correção de Yates não é necessária, mesmo com amostras muito pequenas, como $n = 20$.**
- (d) Em tabelas de contingência com apenas 1 grau de liberdade (2x2) onde $n < 20$, ou $20 \leq n \leq 40$ e a menor $f_e < 5$, deverá ser utilizado o *teste exato de Fischer* (CENTENO, 1982: 164).
- (e) Ressalta-se que, a correção de Yates tende a subestimar a significância quanto à dependência das amostras, e em menor grau, o próprio teste exato de Fischer, enquanto que o teste do χ^2 superestima tais diferenças SOKAL & ROHLF (1981).

Correção de Yates para o teste do χ^2 : consiste na subtração de 0,5 de cada contribuição.

$$\chi^2 = \left(|fe - fo| - 0,5 \right)^2 / fe$$

ou

$$\chi^2 = \frac{\left(|ad - bc| - \frac{n}{2} \right)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

2.2.1.4 – Teste exato de Fisher: é de difícil cálculo manual, já que exige valores de logaritmos de fatoriais (disponível em tábuas, ROHLF & SOKAL, 1981: 3). Contudo, calculadoras virtuais, como a calculadora científica do Microsoft Windows, apresenta o cálculo fatorial, o que descarta a necessidade destas tábuas, principalmente porque usaremos este teste basicamente quando as amostras forem muito pequenas e não permitirem o uso do teste G ou χ^2 . Todavia, com amostras muito grandes, o cálculo manual torna-se impraticável, mesmo com calculadoras virtuais. Felizmente, o teste exato de Fisher é encontrado em vários pacotes estatísticos disponíveis no mercado. Ressalta-se que, como geralmente não conhecemos a fundo a distribuição dos caracteres nas amostras populacionais que estamos trabalhando e, portanto, não possuímos hipótese alternativas bem fundamentadas, consideraremos as duas “caudas” da distribuição (“two-tailed test”), isto é, uma distribuição simétrica. Para maiores esclarecimentos, ver distribuição binomial (SOKAL & ROHLF, 1981: 79) e CENTENO (1982: 51).

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)}{n!a!b!c!d!}$$

Exemplo: o procedimento segue VANZOLINI (1993) e SOKAL & ROHLF (1981: 740).

Tabela 8. *Phyllopezus pollicaris*, presença de tubérculos ad-anais em machos e fêmeas (VANZOLINI, 1993: 36).

	Masculino	Feminino	Totais
Tubérculos presentes	6	3	9
Tubérculos ausentes	4	1	5
Totais	10	4	14

$$p = \frac{(9)!(5)!(10)!(4)}{14!6!3!4!1!}$$

(A) Caso tenhamos a possibilidade de obter os fatoriais diretamente, que são números muito grandes, o processo é simplificado (**a calculadora do windows, no modo científico, o faz sem maiores problemas**).

Neste caso,

$$p_1 = \frac{362.880 \times 120 \times 3.628.800 \times 24}{87.178.291.200 \times 720 \times 6 \times 24 \times 1} = \frac{3.792.438.558.720.000}{9.038.645.231.616.000} = 0,4195$$

(B) Em seguida, faça o mesmo procedimento para a **tabela extrema**. Nesta, se $ad - bc$ for negativo, subtraia **1** das células a e d e some **1** às células b e c . Nesta tabela extrema as f_o destoam ainda mais das f_e . Contudo, se $ad - bc$ for positivo, iremos decrescer **1** das células b e c , e somar **1** às células a e d . **Note ainda que os valores totais marginais continuam os mesmos**. Repare que este procedimento deve ser repetido até que uma das células tenha frequência = **0**.

	Masculino	Feminino	Totais
Tubérculos presentes	5	4	9
Tubérculos ausentes	5	0	5
Totais	10	4	14

Assim sendo,

$$p_0 = \frac{(9)!(5)!(10)!(4)}{14!5!4!5!0!}$$

$$p_0 = \frac{362.880 \times 120 \times 3.628.800 \times 24}{87.178.291.200 \times 120 \times 24 \times 120 \times 1} = \frac{3.792.438.558.720.000}{30.128.817.438.720.000} = 0,1258$$

(C) O terceiro passo é calcular o valor assimétrico de p (“one-tailed distribution”).

$$p = p_1 + p_0 = 0,4195 + 0,1258 = \mathbf{0,5453}$$

(D) Como nos interessa o valor simétrico (“two-tailed”), já que não temos informações contundentes sobre as distribuições destes caracteres nessas populações (informações prévias consistentes das distribuições), então basta que multipliquemos este valor por **2**.

$$\mathbf{p = 0,5453 \times 2 = 1,0.}$$

Este valor não rejeita a hipótese nula (aliás, tão pouco o valor de assimetria, 0,5453).



Exercício: verifique se existe variação geográfica no caráter “faixa dorsolateral” entre as três áreas amostrais, norte-sul, para a morfoespécie PRU (Fig. 3-5). (1) **Procure fazer análises 2x3, isto é, áreas I x II, I x III e II x III.** (2) **Lembre-se de verificar, primeiro, as recomendações para o uso dos testes e escolha o mais pertinente.** A resolução seguirá com o teste to qui-quadrado, originalmente escolhido.

Tabela 8. Distribuição dos padrões de faixa dorsolateral na morfoespécie PRU. *n*, número de espécimes. C1 a C6, padrões de faixa dorsolateral, já agrupados 2 a 2 em categorias por critério de semelhança.

Áreas	n	C1+C2	C3+C4	C5+C6	Totais
I	22	11	11	0	22
II	18	2	13	3	18
III	5	0	0	5	5
Totais		13	24	8	45

Proposta de resolução:

Áreas	C1+C2	C3+C4	C5+C6
I	11	11	0
II	2	13	3
III	0	0	5

(A) Áreas I x II

Áreas	C1+C2	C3+C4	C5+C6	Totais
I	11 (132)	11 (132)	0 (1,65)	22
II	2 (9,9)	13 (10,8)	3 (1,35)	15
Totais	13	24	3	40

Note que há duas $f_e < 5$ em um $GL = 2$, o que viola o uso do χ^2 e teste G . Para prosseguirmos, vamos unir as duas últimas classes e aumentar os f_e .

Áreas	C1+C2	C3 a C6	Totais
I	11 (7,15)	11 (14,8)	22
II	2 (9,9)	16 (12,1)	18
Totais	13	27	40

Agora, temos todos os $f_e > 5$. Contudo, como o total é de $n=40$, podemos aplicar a correção de Yates no caso do χ^2 (caso sigamos CENTENO, 1982), e correção de Williams no teste G . Contudo, se considerarmos SOKAL & ROHLF (1981: 743) então calcularemos o χ^2 sem correções. Aqui, neste exemplo, aplicaremos a correção de Yates com a finalidade de exercitá-la.

$$\chi^2 = \frac{\left(\left| 176 - 22 \right| - \frac{40}{2} \right)^2 40}{22 \times 18 \times 13 \times 27} = \frac{(154 - 20)^2 40}{138996} = \frac{718240}{138996} = 5,16$$

Para um valor de $\chi^2 = 5,15$ e $GL=1$ temos um $P < 0,05$, o que nos permite rejeitar a H_0 e inferir que as amostras das áreas I e II não eram da mesma população estatística.

(B) Áreas I x III

Áreas	C1+C2	C3 a C6	Totais
I	11 (8,96)	11 (13,0)	22
III	0 (2,03)	5 (2,96)	5
Totais	11	16	27

Note que violamos uma premissa para tabelas 2x2, já que temos duas $f_e < 5$. Neste caso, somente nos resta utilizar o teste exato de Fisher.

$$p = \frac{(22)!(5)!(11)!(16)}{27!11!11!0!5!} = 0,0541$$

Como o valor de uma das células é 0, então não precisamos computar uma tabela extrema, sendo $p = 0,0541$ (valor simétrico, ou “one-tailed”). O valor assimétrico (“two-tailed”) seria $0,0541 \times 2 = 0,1082$ (repare que o valor dobrado não é exatamente o valor assimétrico corretamente derivado do cálculo da segunda “cauda”, que seria de $p = 0,0598$). Assim sendo, temos valores p acima de 0,05, o que não nos permite rejeitar a hipótese nula,

mas beiram esta rejeição, o que nos obriga a aceita-la, ou não, com cautela. Lembre ainda que, com um pacote estatístico, o cálculo das duas caudas é obtido mais facilmente e é mais preciso.

(C) Áreas II x III

Áreas	C1+C2	C3 a C6	Totais
II	2 (1,56)	16 (16,43)	18
III	0 (0,43)	5 (4,56)	5
Totais	2	21	23

Note que, entre outras violações, uma das f_e é < 1 , o que inviabiliza os teste do χ^2 . Assim sendo, vamos direto para o teste exato de Fisher.

$$p = \frac{(18)!(5)!(2)!(21)}{23!2!16!0!5!} = 0,6047$$

O valor assimétrico (“two-tailed”) seria de $0,6047 \times 2 = 1$, o que denota a não rejeição da hipótese nula.



III ANÁLISE DOS DADOS MORFOMÉTRICOS

3.1 – Organização e aquisição de dados

Os dados derivados de mensurações devem ser transferidos para uma planilha de formato compatível com as matrizes de dados a serem usadas em pacotes de programas estatísticos. O padrão, é arrumar as variáveis em colunas e os espécimes nas linhas. A identificação de localidades, morfoespécies, UTOs, sexo e outras variáveis de identificação, entram igualmente como colunas. Veja o exemplo abaixo:

Tombo	Local	Sexo	UTO	CRC	LC	CC	DO	LPS	DIO	DON	DIN	DT	DNF	CCX	TB	PÉ	DD3	D4A
MNRJ 23336	1	1	1	50,9	18,2	18,3	5,8	4,9	4,6	4,5	6,2	3,6	2,7	27,1	26,1	37,3	2,7	2,6
MNRJ 23335	1	2	1	52,5	18,0	18,5	5,4	4,7	4,4	4,1	5,0	3,2	2,9	27,0	26,8	36,9	2,4	2,3
CFBH 754	2	1	2	48,0	17,5	17,6	5,9	4,3	4,1	5,5	2,9	2,6	2,5	24,5	28,8	39,0	2,6	2,5

Dados aferidos com paquímetro devem ser repetidos n vezes, a fim de minimizarmos o erro de medidas, e sempre conferidos pelo mesmo pesquisador. Existem opiniões diversas de como se deve mensurar um caráter. Por exemplo, há autores que medem a mesma medida 20 vezes e daí tiram a média aritmética: imagine medir 1000 espécimes com 18 medidas cada um! Embora menos preciso, podemos medir uma mesma medida 3 a 5 vezes e daí optar pelo valor modal (que mais se repete). Caso utilizemos uma ocular micrométrica, a precisão é muito maior e acredito ser desnecessário repetir a tomada da medida mais de uma vez. Outro aspecto, é que por mais que sejamos criteriosos, há um grau de erro na mensuração. Assim sendo, somente registraremos um valor até a primeira casa decimal (p.ex., 34,5 mm e não 34,55 mm). O esquema (Fig. 6) abaixo ilustra 15 medidas usadas em anuros arborícolas.

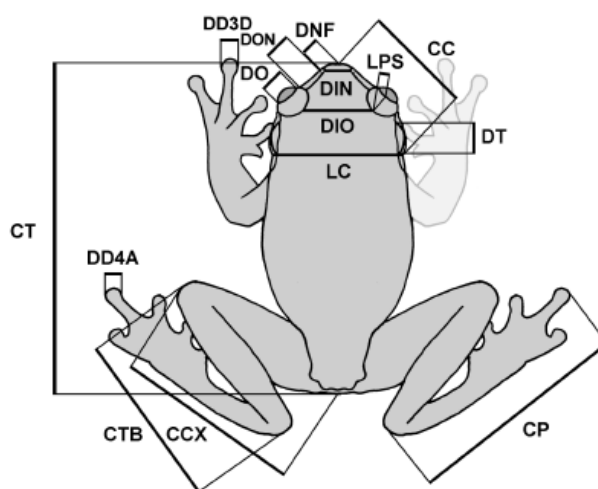


Figura 6. Caracteres morfométricos. CT, comprimento total; CC, comprimento da cabeça; LC, largura da cabeça; DO, diâmetro do olho; DIO, distância interorbital; LPS, largura da pálpebra superior; DON, distância olho-narina; DIN, distância internasal; DNF, distância narina-ponta do focinho; DT, diâmetro do tímpano; DD3D, diâmetro do disco do terceiro dedo; CCX, comprimento da coxa; CTB, comprimento da tíbia; CP, comprimento do pé; DD4A, diâmetro do disco do quarto artelho (NAPOLI, 2000).

3.2 – Variáveis de escala geométrica (ou variáveis quantitativas)

Como já definido anteriormente, podem ser subdivididas em: (A.1) **variáveis discretas**, onde somente valores positivos inteiros são possíveis (p.ex., contagem de escamas de uma cobra) e (A.2) **variáveis contínuas**, valores positivos quaisquer (p.ex., medições). É onde usaremos testes estatísticos paramétricos, isto é, que se baseiam na premissa de que os dados (variáveis) obedecem a certas distribuições estatísticas. Os testes paramétricos têm mais “poder” para rejeitar a hipótese nula do que os não paramétricos, mas igualmente têm uso mais restrito, justamente pela necessidade de obedecerem a certas premissas.

3.2.1 – Estatística descritiva

É o primeiro passo a seguir quando nos deparamos com mensurações e contagens. Objetiva-se descrever os **parâmetros** da amostra populacional analisada. A seguir, comentarei sucintamente os principais parâmetros populacionais usados em trabalhos de cunho taxonômico, sendo eles **média aritmética, moda, amplitude, variância, desvio padrão e coeficiente de variação**. Exemplificarei estes parâmetros mais adiante em um exercício geral, sendo que a seguir discutirei brevemente cada um dos mesmos.

3.2.1.1 – Média aritmética

É o parâmetro de tendência central mais usado.

$$\bar{X} = \frac{\sum x}{n}$$

onde, $\sum x$ é a soma dos n espécimes da amostra.

3.2.1.2 – Moda

É um parâmetro de tendência central menos usado que a média. É o valor que representa a maior frequência numa distribuição e é representado por M_0 (o que mais se repete).

Exemplos:

amodal: 1, 2, 4, 7, 11;

unimodal: 3, 8, 8, 10, 12 ($M_0 = 8$);

bimodal: 3, 3, 5, 6, 8, 8, 9 ($M_0 = 3$ e 8).

3.2.1.3 – Amplitude

É o parâmetro de dispersão mais usado em tabelas descritivas em taxonomia. É a diferença entre os valores do maior e do menor elemento do conjunto. Contudo, nas tabelas descritivas usadas para mostrar os parâmetros amostrais, não apresentamos esta diferença, mas apenas o *menor valor seguido do maior valor*. Exemplo: 34,7-89,8.

3.2.1.4 – Variância

É o parâmetro de dispersão mais usado.

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

onde, $\sum_{i=1}^n (X_i - \bar{X})^2$ é a soma dos quadrados das diferenças entre cada observação e a média; $n - 1$ é o número de graus de liberdade. Se já tivermos o desvio padrão (abaixo), então basta elevarmos o mesmo ao quadrado para obtermos a variância.

3.2.1.5 – Desvio padrão

Junto com a variância, é o parâmetro de dispersão mais usado. O desvio padrão expressa o desvio de cada um dos elementos, x_i , em relação à média aritmética, sendo a **raiz quadrada da variância**.

$$s = \sqrt{s^2}$$

3.2.1.6 – Coeficiente de variação

O coeficiente de variação é uma medida abstrata que independe de unidades em que foram medidos os dados, sendo então bastante útil para comparar a variação entre duas séries de dados medidos em unidades distintas (CENTENO, 1982). Expressa o desvio padrão se a média representasse o índice 100.

$$CV = \frac{s}{\bar{X}} \cdot 100$$

3.2.2 – Comparação entre médias

Dois testes são de uso geral e fundamentais na comparação entre médias de variáveis morfométricas: **teste t de Student** e **análise de variância segundo um único critério (“One-way ANOVA”)**. Todavia, estes testes são paramétricos e portanto se baseiam em certas premissas, distribuições, exigidas para o uso dos mesmos. No caso, seria a exigência de que as distribuições fossem normais, isto é, obedecessem a uma curva de distribuição normal, como veremos a seguir.

3.2.2.1 – Distribuição normal

Na maioria dos fenômenos biológicos mensuráveis os valores extremos são raros e, gradativamente, à medida que se aproximam da média tornam-se mais frequentes (CENTENO, 1982) (Fig. 7). Diferenças na média da população (μ) deslocam a curva para a direita ou esquerda; valores grandes no desvio padrão (σ) achatam a curva, enquanto valores pequenos a tornam mais aguda, o que pode afastar a distribuição de nossas variáveis da distribuição normal (Fig. 8). Via de regra, o aumento considerável da amostra resolve este problema.

Em termos práticos, não interessa saber se nossas variáveis têm, ou não, uma distribuição normal, ou ao menos próxima da mesma. Caso nossas amostras fossem sempre muito grandes, não teríamos tal preocupação; contudo, geralmente são pequenas e tais desvios podem dificultar a rejeição da hipótese nula (de que a distribuição é normal). A computação para tal é tediosa, mas podemos utilizar testes disponíveis em pacotes estatísticos como o teste de Kolmogorov-Smirnov e **teste W de Shapiro-Wilks**, somados a superimposição da curva normal esperada em histogramas (gráficos de frequência) (Fig. 9), o que facilita em muito esta tarefa. No pacote estatístico STATISTICA for Windows (ver. 5.1), pacote este bem considerado no meio científico, encontramos a recomendação para o uso do teste W de Shapiro-Wilks para normalidade, opinião esta baseada em SHAPIRO, WILK & CHEN (1968). Neste caso, se o valor W for significativo, então a hipótese (H_0) de que a distribuição é normal deve ser rejeitada.

Em variáveis morfométricas frequentemente usamos a transformação dos valores originais em logaritmos a fim de tornar mais linear a distribuição e torná-la mais próxima à curva de normalidade. Assim, distribuições de frequência que tendem para a direita quase sempre se tornam mais simétricas em escala logarítmica (SOKAL & ROHLF, 1981: 419).

Note que, dependendo dos dados, outros tipos de transformação devem ser empregados, como transformação em raiz quadrada (para contagens) e transformação angular (arcseno, para porcentagens e razões) (SOKAL & ROHLF, 1981: 417-428).

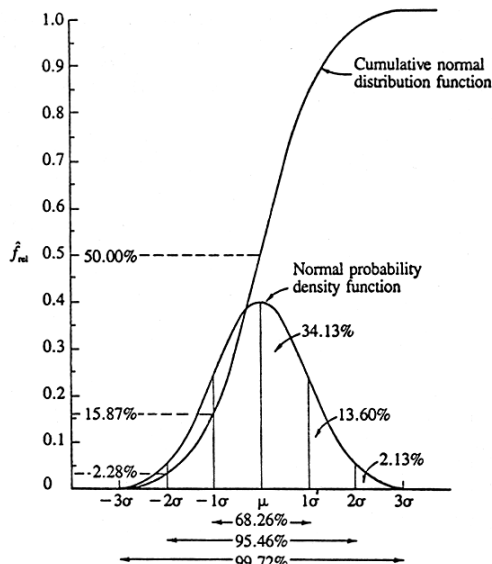


Fig. 7 – Áreas sobre a curva de distribuição normal e curva de distribuição normal acumulada (SOKAL & ROHLF, 1981: 103).

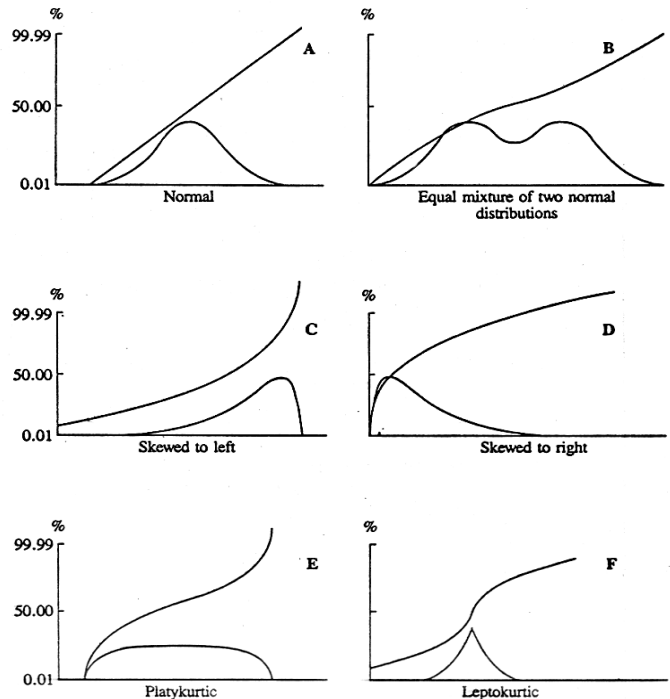


Fig. 8 – Exemplos de certas distribuições de frequência com suas distribuições cumulativas plotadas com a ordenada em escala de probabilidade normal (SOKAL & ROHLF, 1981: 119). Note os desvios de A a F.

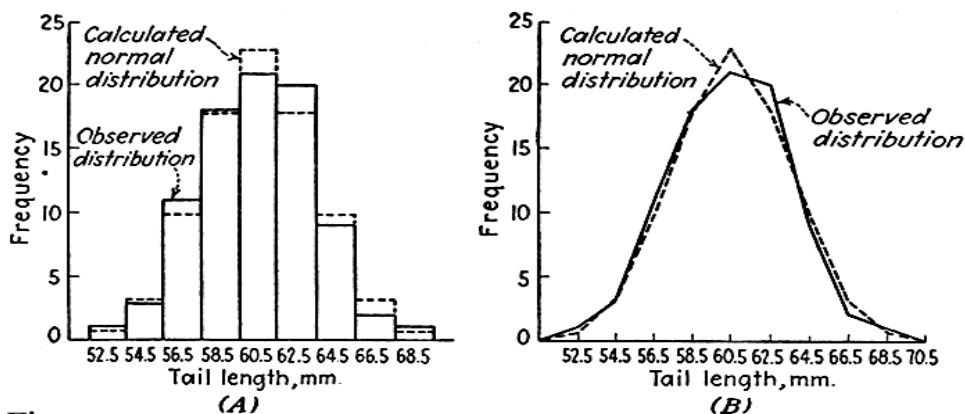


Fig. 9 — Approximation of a theoretical normal distribution to an observed distribution. Tail length in the mammal *Peromyscus maniculatus bairdii*. The continuous lines represent the observed facts, as given in Example 36, and the broken lines an approximately equivalent normal distribution, as calculated in Example 46. A, histogram. B, frequency polygon. The histogram permits clearer visual comparison of the observed and theoretical frequencies. The polygon more clearly shows the approximation to a normal curve. (Simpson & Roe, 1939: 140)

3.2.2.2 – Comparação entre duas médias (com distribuição normal): teste *t* de Student

O teste *t* de Student é o mais usado. O valor obtido deve ser contraposto em uma tábua para o teste *t* (Anexo, Tabela 2). Seu uso depende que (1) as distribuições das amostras sejam normais e (2) que tenham a mesma variância. A verificação quanto à normalidade da amostra já foi discutida

no item anterior. Para sabermos se as amostras têm a mesma variância, usaremos a seguinte fórmula (SOKAL & ROHLF, 1981: 190).

$$F = \frac{s_1^2}{s_2^2} \text{ ou } \frac{s_2^2}{s_1^2}$$

onde, F representa a distribuição F (em homenagem a R. A. Fisher), s_1^2 é a variância da amostra 1 e s_2^2 a variância da amostra 2. *Note que, o numerador deve ser aquele da amostra de maior variância.*

O valor F deve ser verificado em uma tabela de duas entradas (**Anexo, Tabela 3**), onde cruzamos dados do numerador (graus de liberdade da soma dos quadrados) com dados do denominador (graus de liberdade da soma dos quadrados). Em taxonomia, e em outras áreas biológicas, o teste geralmente é assimétrico (“two-tailed”), já que não temos razão para acreditar que os resultados em uma amostra devam ser mais variáveis do que na outra. Logo, a hipótese nula é $H_0 : \sigma_1^2 = \sigma_2^2$ e a hipótese alternativa é $H_1 : \sigma_1^2 \neq \sigma_2^2$. Note ainda que, como o resultado será simétrico (“one-tailed”), precisamos dobrar os valores de α para que ele seja assimétrico (“two-tailed”), isto é, se $F_{0,025[9,9]} = 4,03$, então o valor F para 4,03 representará uma probabilidade de $\alpha = 0,05$ ($F_{\alpha[v1,v2]}$ ou $F_{\alpha[v2,v1]}$, onde α é o erro do tipo I (quando rejeitamos a H_0 e ela é verdadeira), $v1 = n_1 - 1$ e $v2 = n_2 - 1$ graus de liberdade para as variâncias do numerador e denominador).

Caso as duas premissas acima mencionadas não tenham sido violadas, então poderemos usar a seguinte fórmula para o teste t (SOKAL & ROHLF, 1981: 226), **lembrando que este igualmente é um teste assimétrico (“two-tailed”)**, e os valores críticos tabelados **já** correspondem à assimetria (caso desejemos um valor simétrico, temos de dividir o valor crítico de P por dois) já que nossa hipótese alternativa é que $H_1 : \mu_1 \neq \mu_2$ ($H_0 : \mu_1 = \mu_2$):

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right] \left(\frac{n_1 + n_2}{n_1 n_2} \right)}}$$

onde, $\bar{Y}_1 - \bar{Y}_2$ representa a diferença das médias aritméticas das duas amostras em questão; $(n_1 - 1)$ e $(n_2 - 1)$ são os graus de liberdade das amostras; s_1^2 e s_2^2 são as variâncias das amostras.

Quando as amostras tiverem o mesmo n , então podemos usar a seguinte fórmula simplificada:

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\frac{1}{n}(s_1^2 + s_2^2)}}$$

Quando ambas as amostras não obedecerem à distribuição normal, teremos que utilizar testes não paramétricos. Contudo, caso tenhamos distribuições normais mas **variâncias distintas**, então utilizaremos a seguinte correção (SOKAL & ROHLF, 1981: 411):

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}}$$

Contudo, o que distingue esta fórmula do teste t para grandes amostras com igual variância (idêntica à acima apresentada) é o cálculo do valor crítico de t_α para o erro do Tipo I (quando rejeitamos a H_0 e ela é verdadeira) que é computado da seguinte maneira:

$$t_\alpha = \frac{\frac{t}{\alpha[v_1]} \frac{s_1^2}{n_1} + \frac{t}{\alpha[v_2]} \frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Como exemplo, se $t_{0,05[36]}=2,028$ e $t_{0,05[5]}=2,571$ e $\frac{s_1^2}{n_1} = 0,000289$ e $\frac{s_2^2}{n_2} = 0,020736$ (onde, $[36] = n_1-1$; $[5] = n_2-1$), então $t' = \frac{2,028(0,000289) + 2,571(0,020736)}{0,021025} = 2,564$. Logo, as duas médias são significativamente diferentes.

Exercício: A tabela ao lado apresenta os valores do diâmetro do comprimento da cabeça, em milímetros, para *Hyla cachimbo* (morfoespécie CBO) e *H. rubicundula* (morfoespécie RU). Sabe-se que, pela diagnose de *H. cachimbo*, esta espécie tem o focinho acuminado, quando comparada às demais espécies do grupo de *H. rubicundula*. Queremos saber se há diferença entre estas duas espécies (amostras) quanto ao comprimento da cabeça, o qual estaria influenciado pelo comprimento do focinho, que é mais ou menos acuminado em *H. cachimbo* (morfoespécie CBO). A hipótese nula para nossa comparação é a de que não há diferença entre as amostras e ambas decorrem da mesma população estatística. A hipótese alternativa é a de que *H. cachimbo* apresenta a cabeça mais comprida do que *H. rubicundula*. Assim sendo, teremos um teste monocaudal. Contudo, se partirmos da premissa de que temos morfoespécies ainda desconhecidas e desejamos descrever justamente a variação, então a hipótese alternativa é simplesmente a de que as amostras diferem entre si, o que nos leva a um teste bi-caudal. Leve ainda em consideração que não existe diferença no comprimento rostro-cloacal entre as duas amostras, isto é, apresentam o mesmo “tamanho” e somente machos adultos foram considerados (dados extraídos de NAPOLI, 1995). **Procure usar 4 casas decimais para maior precisão dos resultados.**

<i>Hyla cachimbo</i>	<i>Hyla rubicundula</i>
6,4	6,5
6,8	6,4
6,4	6,6
6,3	6,5
6,2	6,5
6,8	6,3
6,3	6,1
6,7	6,1
6,4	6,1
6,4	6,5
6,5	6,2
6,2	6,2
6,6	6,4
6,8	6,5
6,5	6,6
	6,1
	6,3
	6,7
	6,3
	6,7

Resolução:

$$\bar{X}_1 = 6,4866; \bar{X}_2 = 6,3800; s_1^2 = 0,0441; s_2^2 = 0,0406; n_1 = 15; n_2 = 20.$$

$$F = \frac{s_1^2}{s_2^2} = \frac{0,04}{0,04} = 1; v_1 = n - 1 = 15 - 1 = 14; v_2 = 20 - 1 = 19. F_{0,05[14,20]} = 2,39 \text{ (não temos } v_2 = 19 \text{ na tabela). Então, no}$$

teste monocaudal, não rejeitamos a hipótese nula, isto é, as variâncias não diferem significativamente. Caso fosse um teste bi-caudal, então dobraríamos o valor crítico, isto é $F_{0,025[14,20]} = 2,84$, valor este que representaria o $F_{0,05}$. Assim, podemos prosseguir a análise, já que não violamos as premissas quanto à normalidade e igualdade das variâncias. Caso recorrêssemos a um programa estatístico (p.ex., Statistix for Windows, teríamos, $F=1,09$, $GL = 14,19$, $P=0,4257$).

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right] \left(\frac{n_1 + n_2}{n_1 n_2} \right)}} = \frac{6,4866 - 6,3800}{\sqrt{\left[\frac{(14)0,0441 + (19)0,0406}{15 + 20 - 2} \right] \left(\frac{15 + 20}{15 \cdot 20} \right)}} = \frac{0,1066}{\sqrt{\frac{1,3888 \left(\frac{35}{300} \right)}{0,0700}}} = 1,5228$$

$$GL = n_1 + n_2 - 2 = 15 + 20 - 2 = 33$$

$t_{0,05[33]}$ = não tem na tabela: teremos de interpolar entre $t_{0,05[30]}=2,042$ e $t_{0,05[40]}=2,021$ ou usar o maior e mais próximo valor de α . Lembre-se que, quando usamos programas estatísticos a probabilidade exata já é fornecida, o que dispensa na prática a interpolação, sendo por isto, dispensável este tipo de computação para nossos objetivos. Indubitavelmente, o t obtido indica uma $P < 0,1$, o que não nos permite rejeitar a H_0 : não há diferença significativa entre as amostras. O valor exato é $P = 0,1375$. Morfologicamente falando, lembre-se que o comprimento total da cabeça é influenciado por outras distâncias e a diferença no focinho poderia ser mais bem aferida usando-se somente a distância narina-focinho.

NOTA: Quando as premissas para o uso do teste t , neste caso a de normalidade, for violada, então poderemos usar testes não paramétricos, isto é, aqueles que não dependem de hipóteses sobre a natureza das distribuições. Três testes são bem eficientes: o teste de **Mann-Whitney**, **Wilcoxon para duas amostras** e o teste de **Kolmogorov-Smirnoff**. Estes testes não serão aqui discutidos, mas podem ser encontrados em SOKAL & ROHLF (1981) e VANZOLINI (1993).

3.2.2.3 – Comparação entre mais de duas médias (com distribuição normal): análise de variância, segundo um único critério (“One-way ANOVA”)

Em termos práticos, a ANOVA é mais poderosa do que o teste t , já que não compara apenas médias, mas também as distribuições dos espécimes dentre as amostras, embora sua computação seja mais trabalhosa. Contudo, o resultado obtido em um teste t para duas amostras é rigorosamente o mesmo na análise de variância. A ANOVA procura contrapor a variação dentro das amostras com a variação entre as amostras, o que nos permite determinar se as variações tendem a ser classificadas como intrapopulacionais ou interpopulacionais, ou ainda, intraespecíficas ou interespecíficas. Em outras palavras, “a análise de variância consiste na decomposição da variação total de m amostras em duas partes: (1) uma parte relativa à variação *dentro das amostras* (erro) e (2) uma parte relativa à variação *entre as amostras* (efeito)”. Este “erro” ou “resíduo” pode ser de dois tipos: (a) inerente à própria variabilidade dos espécimes analisados e (b) falta de critérios bem estabelecidos para aquisição dos dados, no nosso caso, pouca precisão da técnica de medição. Ressalto ainda que podemos ter dois modelos de ANOVA: Modo I e Modo II. Em taxonomia, queremos contrapor populações onde os espécimes foram supostamente extraídos da população ao acaso, o que nos inclui no Modo II.

Como já dito, sua computação (decomposição) é mais trabalhosa do que o teste t . A computação manual segue o exemplo de SOKAL & ROHLF (1981: 211) para o Modo II, amostras com n desigual.

Amostras ($a = 4$) [p.ex., amostras populacionais]				
[Espécimes]	1	2	3	4
[1]	380	350	354	376
[2]	376	356	360	344
[3]	360	358	362	342
[4]	368	376	352	372
[5]	372	338	366	374
[6]	366	342	372	360
[7]	374	366	362	
[8]	382	350	344	
[9]		344	342	
[10]		364	358	
[11]			351	
[12]			348	
[13]			348	
$\sum_{n_i} Y$	2978	3544	4619	2168
n_i	8	10	13	6
$\sum_{n_i} Y^2$	1.108.940	1.257.272	1.642.121	784.536
S^2	54.21	142.04	79.56	233.07

1. Somatório dos totais das amostras (colunas):

$$\sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij} = 2978 + 3544 + 4619 + 2168 = 13.309$$

2. Somatório dos valores individuais ao quadrado para cada coluna:

$$\sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}^2 = 1.108.940 + 1.257.272 + 1.642.121 + 784.536 = 4.792.869$$

3. Somatório dos totais, ao quadrado, de cada grupo, cada um dividido pela seu respectivo n :

$$\sum_{i=1}^a \frac{\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2}{n_i} = \frac{(2978)^2}{8} + \frac{(3544)^2}{10} + \frac{(4619)^2}{13} + \frac{(2168)^2}{6} = 4.789.091$$

4. [Passo 1] ao quadrado dividido pelo número amostral total = Fator de Correção (FC):

$$FC = \frac{\left(\sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}\right)^2}{\sum_{i=1}^a n_i} = \frac{(13.309)^2}{37} = 4.787.283,3$$

$$\begin{aligned} \text{5. Soma dos Quadrados totais} &= \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}^2 - FC \\ &= [\text{Passo 2}] - [\text{Passo 4}] = 4.792.869 - 4.787.283,3 = 5585,7 \end{aligned}$$

$$\begin{aligned} \text{6. Soma dos Quadrados entre as amostras (SQ}_1\text{): } &\sum_{i=1}^a \frac{\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2}{n_i} - FC \\ \text{SQ}_1 &= [\text{Passo 3}] - [\text{Passo 4}] = 4.789.091 - 4.787.283,3 = 1807,7 \end{aligned}$$

7. Soma dos quadrados dentro das amostras (“resíduo” ou “erro”) (SQ₂):

$$\text{SQ}_2 = [\text{Passo 5}] - [\text{Passo 6}] = 5585,7 - 1807,7 = \mathbf{3778,0}$$

Dica: Passo 3 $\leq$ Passo 2; Passo 4 $\leq$ Passo 2 e 3; Passo 6 $\leq$ Passo 5.

8. Construção da Tabela da ANOVA:

Fonte da variação	Graus de Liberdade (GL)	Soma dos Quadrados (SQ)	Quadrados Médios (QM)	F
Entre amostras	$a - 1$	SQ ₁	$QM_1 = \frac{SQ_1}{a - 1}$	$\frac{QM_1}{QM_2}$
Dentro das amostras	$\sum_{i=1}^a n_i - a$	SQ ₂	$QM_2 = \frac{SQ_2}{\sum_{i=1}^a n_i - a}$	
Total	$\sum_{i=1}^a n_i - 1$	SQ ₁ +SQ ₂		
Fonte da variação	Graus de Liberdade (GL)	Soma dos Quadrados (SQ)	Quadrados Médios (QM)	F
Entre amostras	3	1807,7	602,6	5,26
Dentro das amostras	33	3778,0	114,5	
Total	36	5585,7		
$F_{,05[3,33]}=2,89$ $F_{,01[3,33]}=4,44$ $F_{,001[3,33]}=6,89$				
Conclusão: Existe uma diferença significativa de variância entre as amostras.				

9. Comparação entre médias: comparações não planejadas (“unplanned comparisons”)

Caso haja variância significativa entre as amostras é de nosso interesse saber como está distribuída esta variação, isto é, quais amostras distam significativamente uma das outras. A quantidade de comparações necessárias é estimada por $a(a-1)/2$. Como o resultado não era sabido até agora (de que as amostras têm variância significativa), o modelo de comparações entre médias deve ser do tipo “não planejado” (SOKAL & ROHLF, 1981: 242).

SOKAL & ROHLF (1981: 245; 261) recomendam três métodos principais para comparar as médias entre si quando os n são desiguais (o que quase na totalidade das vezes acontece):

- Método GT2 (semelhante ao teste de Gabriel, SOKAL & ROHLF, 1981: 249), aplicado principalmente quando os n das amostras diferem muito;
- Método T' (Spjotvoll & Stoline, 1973), aplicado principalmente quando os n das amostras não diferem tanto;
- e o teste de *Tukey-Kramer*, aplicado principalmente quando os n das amostras diferem muito.

Quanto à escolha do método, os referidos autores recomendam que se faça os três testes e escolha-se o melhor resultado (a menor *diferença mínima significativa* - DMS). Ressalto que, caso venhamos a utilizar um pacote estatístico conhecido no mercado, o método de Spjotvoll/Stoline é encontrado no Programa para computador STATISTICA for Windows e o programa MCPAIR, pertencente ao pacote dos autores, apresenta os três testes.

9.1 – DMS: *Diferença Mínima Significante*

Os testes de comparações entre médias se baseiam no tradicional DMS (*diferença mínima significativa*) (“least significant difference-LSD”) (“planned comparisons”). Assim sendo, irei explicar sucintamente este teste. O DMS é um intervalo de confiança entre duas médias:

$$(1) \text{ } n \text{ diferentes: } DMS = t_{\alpha(GL)} \sqrt{\frac{s^2(n_1+n_2)}{n_1 \cdot n_2}} \quad (2) \text{ } n \text{ iguais: } DMS = t_{\alpha(GL)} \sqrt{2 \frac{s^2}{n}}$$

onde, $t_{\alpha(GL)}$ é o valor tabelado de t ao nível de significância α para o GL do resíduo (dentro das amostras); s^2 é a variância do resíduo (QM do resíduo); n é o número de observações por tratamento. Qualquer diferença entre duas médias maior que o DMS calculado é considerada significativa ao nível especificado. Vejamos o exemplo abaixo retirado de CENTENO (1982: 118), onde o autor pretende comparar médias de tratamentos (amostras): $n = 4$, $GL \text{ res.} = 15$ e $s^2 = 28,18$. Médias de 6 amostras: $\bar{X}_A = 41,75$, $\bar{X}_B = 51,00$, $\bar{X}_C = 58,25$, $\bar{X}_D = 30,00$, $\bar{X}_E = 33,00$ e $\bar{X}_F = 45,00$. $t_{0,05[15]} = 2,13$.

$$DMS = 2,13 \sqrt{2 \cdot \frac{28,18}{4}} = 8,01$$

$$\begin{aligned} |\bar{X}_A - \bar{X}_B| &= 9,25; & |\bar{X}_A - \bar{X}_C| &= 16,50; & |\bar{X}_A - \bar{X}_D| &= 11,75; & |\bar{X}_A - \bar{X}_E| &= 8,75; & |\bar{X}_A - \bar{X}_F| &= 3,75*; \\ |\bar{X}_B - \bar{X}_C| &= 7,25*; & |\bar{X}_B - \bar{X}_D| &= 21,00; & |\bar{X}_B - \bar{X}_E| &= 18,00; & |\bar{X}_B - \bar{X}_F| &= 5,50*; & |\bar{X}_C - \bar{X}_D| &= 28,25; \\ |\bar{X}_C - \bar{X}_E| &= 25,25; & |\bar{X}_C - \bar{X}_F| &= 12,75; & |\bar{X}_D - \bar{X}_E| &= 3,00*; & |\bar{X}_D - \bar{X}_F| &= 15,50; & |\bar{X}_E - \bar{X}_F| &= 12,50 \end{aligned}$$

Todas as diferenças maiores do que o DMS de 8,01 são significantes. Foram assinaladas com asterisco aquelas **não** significantes.

Os três outros testes serão explicados baseando-se no exemplo extraído de SOKAL & ROHLF (1981: 244).

	Localidades (ordenadas pela magnitude das médias) (a)							
	1	2	3	4	5	6	7	8
$\bar{Y}$	3,5123	3,6940	3,7223	3,7229	3,7376	3,7479	3,7843	3,8451
n_i	20	20	20	14	10	15	11	20
s^2	0,0357	0,0445	0,0135	0,0284	0,0204	0,0184	0,0338	0,0157

Computação da variância combinada = Quadrados médios dentro das amostras (QM_2)

$$QM_2 = \frac{\sum_{i=1}^a (n_i - 1) s_i^2}{\sum_{i=1}^a (n_i - 1)} = \frac{(20-1)0,0357 + (20-1)0,0445 + \dots + (20-1)0,0157}{19+19+\dots+10+19} = \frac{3,2270}{122} = 0,02645$$

$$\text{Graus de liberdade } (v) = \sum (n_i - 1) = 122$$

9.2 – Método *GT2* de Hochberg

Tomemos o *MSD* como base para o cálculo do *GT2*:

$$MSD = (\text{valor crítico de significância}) \times SE (\text{erro padrão})$$

onde, o valor crítico padronizado é determinado pelo valor $m_{\alpha[k^*, v]}$, encontrado a partir da **Tabela 4 (Anexo)**. Nesta tabela, substitua o valor k pelo a ; $k^* = k(k-1)/2 = a(a-1)/2 = 8(8-1)/2 = 28$. Como nesta tabela não temos a entrada [8, 122] então usaremos a aproximada para $0,05[8,120] = 3,183$. Observe que quando procuramos pelo valor k na tabela estamos procurando o valor a , que é o número de amostras ou colunas de sua tabela original. Caso você use o valor k^* , então estaríamos usando um valor $k = 22$ e certamente os resultados seriam outros. Logo, $m_{\alpha 0,05[28,120]} = 3,183$. O cálculo para o **Erro Padrão** para a comparação das médias $\bar{Y}_1$ e $\bar{Y}_2$ será:

$$SE_{ij} = \sqrt{s_{\bar{Y}_i}^2 + s_{\bar{Y}_j}^2} = \sqrt{\frac{QM_2}{n_i} + \frac{QM_2}{n_j}}$$

onde, $s_{\bar{Y}_i}^2 = \frac{QM_2}{n_i}$. Logo, aplicando-se a última computação, obteremos os erros padrões das oito amostras, que serão, respectivamente: 0,001323, 0,001323, 0,001323, 0,001889, 0,002645, 0,001763, 0,002405 e 0,001323.

O passo seguinte é, como no cálculo do *DMS* (item 8.1.1), computar as diferenças entre todos os pares de médias disponíveis. Por exemplo, para as localidades 1 e 2,

$$DMS_{1,2} = 3,183 \sqrt{0,001323 + 0,001323} = 3,183.(0,051439) = 0,16370$$

Como já vimos, em um par de médias $(\bar{Y}_i - \bar{Y}_j)$ será declarado significativamente diferente quando o valor desta diferença for igual ou maior que o valor crítico da diferença (DMS_{ij}) que, neste caso, será $|\bar{Y}_j - \bar{Y}_i| \geq DMS_{ij}$. Assim, $3,6940 - 3,5123 = 0,1817 > 0,16370$, isto é, as localidades **1** e **2** distam significativamente em suas médias. Uma boa sugestão para todas as comparações entre as médias é montar uma tabela onde $|\bar{Y}_j - \bar{Y}_i|$ será situada abaixo da diagonal e a DMS acima da diagonal, como no exemplo abaixo:

Amostras								
Amostras	1	2	3	4	5	6	7	8
1	—	0,1637	0,1637	0,1804	0,2005	0,1768	0,1943	0,1637
2	0,1817*	—	0,1637	0,1804	0,2005	0,1768	0,1943	0,1637
3	0,2100*	0,0283	—	0,1804	0,2005	0,1768	0,1943	0,1637
4	0,2106*	0,0289	0,0006	—	0,2143	0,1924	0,2086	0,1804
5	0,2253*	0,0436	0,0153	0,0147	—	0,2113	0,2262	0,2005
6	0,2356*	0,0539	0,0256	0,0250	0,0103	—	0,2055	0,1768
7	0,2720*	0,0903	0,0620	0,0614	0,0467	0,0364	—	0,1943
8	0,3328*	0,1511	0,1228	0,1222	0,1075	0,0972	0,0608	—

As diferenças maiores ou iguais ao valor de DMS serão significativas. *, $P < 0,05$.

No exemplo ilustrado, a localidade amostral **1** dista significativamente de todas as demais.

9.3 – Método T' (Spjøtvoll/Stoline)

Usaremos a mesma expressão para o $DMS = (\text{valor crítico de significância}) \times SE$. O valor crítico $Q'_{\alpha[k,v]}$ será capturado da **Tabela 5 (Anexo)**, onde $k = a$, isto é, número de colunas (amostras); v será o GL de QM_2 (quadrados médios dentro das amostras). Logo, $k = 8$ e $v = 122$. Consultando-se a tabela teremos $Q'_{0,05[8,122]} = 4,363$. O Erro Padrão para amostras com n desiguais será:

$$SE_{ij} = \sqrt{\frac{QM_2}{\min(n_i, n_j)}}$$

onde $\min$ é o menor valor n das duas médias analisadas. Conseqüentemente,

$$SE_{3,4} = \sqrt{\frac{0,02645}{14}} = \sqrt{0,001889} = 0,04347$$

e o DMS

$$DMS_{3,4} = 4,363(0,04347) = 0,1896$$

Como já exposto para o cálculo de $GT2$, um par de médias $(\bar{Y}_i - \bar{Y}_j)$ será declarado significativamente diferente quando o valor desta diferença for igual ou maior que o valor crítico da diferença (DMS_{ij}) que, neste caso, será $|\bar{Y}_j - \bar{Y}_i| \geq DMS_{ij}$. Note que o $DMS_{3,4}$ (0,1896) é maior do que o encontrado para o $GT2$ (0,1804), o que indica que, para este grupo de dados, o $GT2$ é mais poderoso e deve ser usado, ao invés do método de T' . Lembre que no T' faremos uma tabela igual a do $GT2$.

9.4 – Teste de *Tukey-Kramer*

Usaremos a mesma expressão para o $DMS = (\text{valor crítico de significância}) \times SE$. O valor crítico $Q'_{\alpha[k,v]}$ será capturado da **Tabela 6 (Anexo)**, onde $k = a$, isto é, número de colunas (amostras); v será o GL de QM_2 (quadrados médios dentro das amostras). Logo, $k = 8$ e $v = 122$. Consultando-se a tabela teremos $Q'_{0,05[8,122]} = 4,362$. O Erro Padrão para amostras com n desiguais será:

$$SE_{ij} = \sqrt{\frac{QM_2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}{2}}$$

Logo,

$$SE_{3,4} = \sqrt{\frac{0,02645 \left(\frac{1}{20} + \frac{1}{14} \right)}{2}} = \sqrt{0,001606} = 0,040074$$

O DMS será:

$$DMS_{3,4} = 4,362(0,040074) = 0,17480$$

Novamente, um par de médias $(\bar{Y}_i - \bar{Y}_j)$ será declarado significativamente diferente quando o valor desta diferença for igual ou maior que o valor crítico da diferença (DMS_{ij}) que, neste caso, será $|\bar{Y}_j - \bar{Y}_i| \geq DMS_{ij}$. Todavia, note que o $DMS_{3,4}$ (0,17480) é menor do que o encontrado para o $GT2$ (0,1804), o que indica que, para este grupo de dados, o teste de *Tukey-Kramer* é mais poderoso e deve ser usado, ao invés do $GT2$. Lembre que no T' faremos uma tabela igual a do $GT2$. Neste grupo de dados, não houve diferença entre o *Tukey-Kramer* e o $GT2$, onde a localidade amostral 1 dista significativamente de todas as demais, e estas, não diferem significativamente entre si.

Uma maneira bastante útil de demonstrar os resultados é a seguinte: suponhamos que as seguintes amostras, dois a dois, distem significativamente uma das outras: A e B, C e D e E e F; contudo, B e C não distem significativamente uma da outra.

A B C D E F

Exercício: Com os dados abaixo, verifique se há diferença para o caráter “diâmetro do tímpano” (DT) entre espécimes machos do complexo de espécies de *Hyla circumdata* / *H. luctuosa* (UTOs 1-3) e de (4) *H. hylax*.

Espécimes	UTOs			
	1 (Rio)	2 (Japi)	3 (Sta. Teresa)	4 (<i>H. hylax</i>)
1	4,4	6,1	5,6	3,12
2	5,2	5,4	5,2	3,2
3	4,8	5,4	5,2	3,3
4	5,9	5,4	5,7	3,6
5	5,6	4,9	5,2	3,2
6	5,9	4,8	5,2	3,0
7	4,4		5,6	2,9
8	4,8		5,8	3,0
9	4,8		6,2	3,4
10	5,4		5,3	3,0
11	5,6		6,2	2,9
12	5,6		5,0	2,8
13	4,9		4,6	3,1
14			5,2	3,2

Resolução:

	$\bar{X}$	n	Somatório	GL	s^2
1 – Rio	5,1769	13	67,3000	0,5278	0,2785
2 – Japi	5,3333	6	32,0000	0,4633	0,2146
3 – Santa Teresa	5,4285	14	76,0000	0,4479	0,2006
4 – <i>Hyla hylax</i>	3,1228	14	43,7200	0,2153	0,0463

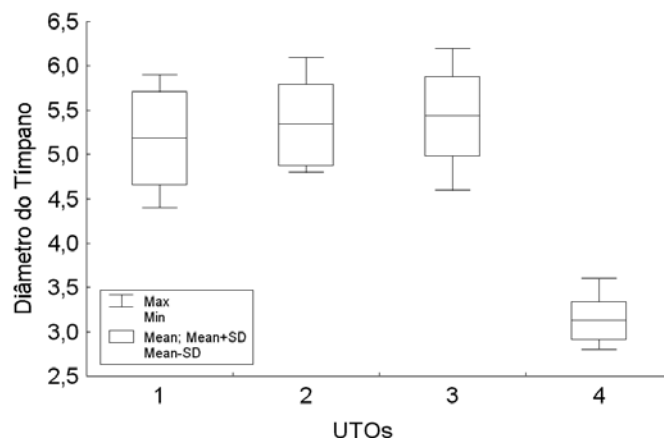
Fonte da variação	Graus de Liberdade (GL)	Soma dos Quadrados (SQ)	Quadrados Médios (QM)	F
Entre amostras	$a - 1$	SQ_1	$QM_1 = \frac{SQ_1}{a - 1}$	$\frac{QM_1}{QM_2}$
Dentro das amostras	$\sum_{i=1}^a n_i - a$	SQ_2	$QM_2 = \frac{SQ_2}{\sum_{i=1}^a n_i - a}$	
Total	$\sum_{i=1}^a n_i - 1$	$SQ_1 + SQ_2$		
Fonte da variação	Graus de Liberdade (GL)	Soma dos Quadrados (SQ)	Quadrados Médios (QM)	F
Entre amostras	3	47,5431	15,8477	89,3835
Dentro das amostras	43	7,6280	0,1773	
Total				
$F_{,05[3,40]}=2,84$ $F_{,01[3,40]}=4,31$ $F_{,001[3,40]}=6,60$				
Conclusão: Existe uma diferença altamente significativa de variância entre as amostras.				

Método T' (Spjotvoll/Stoline): comparação entre médias *a posteriori*

	1 ($\bar{Y} = 5,1769$)	2 ($\bar{Y} = 5,1769$)	3 ($\bar{Y} = 5,1769$)	4 ($\bar{Y} = 5,1769$)
1				
2	0,917403			
3	0,432886	0,979419		
4	0,000170*	0,000170*	0,000170*	
* $P < 0,05$				

Conclusão: O diâmetro do tímpano em machos de *Hyla hylax* (4) difere significativamente de todas as demais UTOs, que não diferem significativamente entre si.

Uma maneira interessante de representar este resultado segue no gráfico seguinte (“Box and Wisker plots”). Neste caso, o hiato entre a UTO 4 e as demais UTOs isentaria a necessidade de uma análise de variância, além de ser mais didaticamente visualizável pelo leitor.



NOTA: Quando as premissas para o uso da análise de variância segundo um único critério, neste caso a de normalidade, for violada, então poderemos usar testes não paramétricos, isto é, aqueles que não dependem de hipóteses sobre a natureza das distribuições. Um teste é bem utilizado: o teste de **Kruskal-Wallis**. Para os testes de comparações múltiplas podemos usar o teste de **Kolmogorov-Smirnoff para duas amostras**. Estes, e outros testes, não serão aqui discutidos, mas podem ser encontrados em SOKAL & ROHLF (1981: 429-453) e VANZOLINI (1993: 77-84).

3.2.3 – Análises de Regressão e Correlação (bivariadas)

Existe muita confusão quanto ao uso da regressão ou da correlação. Uma das principais causas é que matematicamente são dois métodos de análise muito parecidos e muitos textos acabam por tratar dos dois métodos de maneira pouco distinta.

Usamos a **análise de correlação** a fim de determinar o grau de associação entre duas variáveis, isto é, se duas variáveis são interdependentes ou covariam. Por exemplo, em uma determinada amostra populacional existe relação entre o comprimento rostro-cloacal (CRC) e o comprimento da cabeça (CC), isto é, o CC e CRC crescem proporcionalmente? Esta é a pergunta mais freqüente com que nos deparamos. Aspectos como se há relação entre um determinado padrão de desenho e o sexo, também. Contudo, neste último caso não podemos usar a estatística paramétrica, deixando este caso para outro momento.

Tendo esta idéia clara, avancemos mais um pouco. Na análise de correlação não pretendemos verificar se uma variável varia em função de outra, ou vice-versa, isto é, não estamos perguntando se CC aumenta proporcionalmente em função de CRC. Na verdade, partimos da premissa de que não sabemos se estas duas variáveis têm influência uma sobre a outra, ou se estão sob a influência de outra variável não conhecida. Resumindo, usamos a **análise de correlação** quando não podemos supor a existência de uma variável dependente e outra independente.

Por outro lado, se estabelecemos *a priori* que uma variável é independente e que outra variável varia em função da primeira (dependente), então usaremos a **análise de regressão**, ou seja, temos por objetivo descrever a dependência de Y em X , isto é, no exemplo, se CC varia em função de CRC. SOKAL & ROHLF (1981: 564) resumiram estas relações na seguinte tabela (modificada):

Objetivos do pesquisador	Natureza das duas variáveis	
	Y ao acaso, X fixa	Y_1, Y_2 ambas ao acaso
Estabelecer e estimar a dependência de uma variável sobre outra (descrever relação funcional e/ou prever uma em função da outra.)	Regressão do Modo I	Regressão do Modo II
Estabelecer e estimar associação (interdependência) entre duas variáveis.	Sem sentido para este caso	Coefficiente de Correlação
Os Modos I e II de regressão serão explicados mais adiante.		

3.2.3.1 – Análise de Correlação (Pearson; “product-moment correlation coefficient”)

O seguinte exemplo foi extraído de SOKAL & ROHLF (1981: 576) e investiga a relação entre a massa das brânquias de um caranguejo e a sua respectiva massa corporal.

(1) Y_1 Massa da brânquia em miligramas	(2) Y_2 Massa corporal em gramas
159	14,40
179	15,20
100	11,30
45	2,50
384	22,70
230	14,90
100	1,41
320	15,81
80	4,19
220	15,39
320	17,25
210	9,52

Computação:

$$1. \sum Y_1 = 159 + \dots + 210 = 2347$$

$$2. \sum Y_1^2 = 159^2 + \dots + 210^2 = 583.403$$

$$3. \sum Y_2 = 14,40 + \dots + 9,52 = 144,57$$

$$4. \sum Y_2^2 = (14,40)^2 + \dots + (9,52)^2 = 2204,1853$$

$$5. \sum Y_1 Y_2 = 14,40(159) + \dots + 9,52(210) = 34.837,10$$

$$6. \text{Soma dos quadrados de } Y_1 = \sum y_1^2 = \sum Y_1^2 - \frac{\left(\sum Y_1\right)^2}{n}$$

$$= \text{quantidade } 2 - \frac{(\text{quantidade1})^2}{n} = 583.403 - \frac{(2347)^2}{12} = 124.368,9167$$

$$7. \text{Soma dos quadrados de } Y_2 = \sum y_2^2 = \sum Y_2^2 - \frac{\left(\sum Y_2\right)^2}{n}$$

$$= \text{quantidade } 4 - \frac{(\text{quantidade3})^2}{n} = 2204,1853 - \frac{(144,57)^2}{12} = 462,4782$$

$$8. \text{Soma dos produtos} = \sum y_1 y_2 = \sum Y_1 Y_2 - \frac{\left(\sum Y_1\right)\left(\sum Y_2\right)}{n}$$

$$= \text{quantidade } 5 - \frac{\text{quantidade1} \times \text{quantidade3}}{n} = 34.837,10 - \frac{2347 \times 144,57}{12} = 6561,6175$$

$$9. \text{ Cálculo do coeficiente de correlação } r: r_{1,2} = \frac{\sum y_1 y_2}{\sqrt{\sum y_1^2 \sum y_2^2}} = \frac{\text{quantidade8}}{\sqrt{\text{quantidade6} \times \text{quantidade7}}}$$

$$= \frac{6561,6175}{\sqrt{(124.368,9167)(462,4782)}} = \frac{6561,6175}{\sqrt{57.517.912,7314}} = \frac{6561,6175}{7584,0565} = 0,8652 \approx 0,87$$

Cabe agora verificar se o valor obtido é ou não significativo para $P < 0,05$, através de um teste t , felizmente com os valores já tabelados. Levamos em consideração como hipótese nula $H_0: \rho = 0$ (que o coeficiente de correlação obtido é derivado de uma população com coeficiente de correlação paramétrico igual a 0; logo, que as variáveis não são correlacionadas). Quando não rejeitamos a hipótese nula, significa que as duas variáveis são independentes, enquanto que a rejeição da H_0 significa que as variáveis estão associadas. Para tal, vamos observar a **Tabela 7 (Anexo)**. Nesta tabela somente nos interessa $k = 1$, isto é, a primeira coluna. O GL será calculado como $GL = v = n - 2$. Assim sendo, para o exemplo anterior, $GL = 12 - 2 = 10$, onde $P_{0,05[10]} = 0,576$ e $P_{0,01[10]} = 0,708$. Em ambos os casos o valor $r = 0,87$ é maior e assim rejeitamos a hipótese nula: há correlação significativa entre as variáveis.

NOTA: Quando as premissas para o uso da análise de variância segundo um único critério, neste caso a de normalidade, foram violadas, então poderemos usar testes não paramétricos, isto é, aqueles que não dependem de hipóteses sobre a natureza das distribuições. Um teste é bem utilizado: o **coeficiente de correlação de Spearman**. Estes, e outros testes, não serão aqui discutidos, mas podem ser encontrados em SOKAL & ROHLF (1981) e VANZOLINI (1993).

3.2.3.2 – Análise de Regressão Linear

Antes de iniciarmos a computação da análise de regressão, esclareço que existem dois Modos de regressão: o I e o II. A diferença reside em se há ou não “erro” (variação) em ambas as variáveis, o que nos permitirá dizer se há uma variável dependente e outra independente. Caso uma das variáveis seja “fixa” (independente- X), isto é, desprovida de “erro” (inerente ao erro de amostragem pelo pesquisador ou se de variação intrapopulacional), então temos uma variável independente (X) e outra dependente (Y) – Modo I. Caso ambas estejam sujeitas a “erro”, então temos o Modo II, isto é, não temos uma variável independente (X), sendo então denominadas Y^1 e Y^2 . SOKAL & ROHLF (1981: 549) esclarecem que o uso do Modo II é controverso e que é difícil fornecer recomendações definitivas para sua utilização. **Felizmente, caso o uso da regressão linear tenha o objetivo de basicamente prever, estimar, os valores de y em x , então podemos usar o Modo I, mesmo quando não apresentamos uma variável independente.** Todavia, caso tenhamos igualmente o objetivo de estabelecer relações funcionais entre as variáveis, então teremos de usar as correções do Modo II. Como vamos utilizar quase sempre em taxonomia a contraposição de variáveis morfométricas, logo sujeitas a “erro”, cairemos fatalmente no Modo II caso desejemos fazer inferências funcionais entre as variáveis. Seria então providencial termos noções de cálculo do *componente principal*, o que extrapola, em muito, nossos objetivos para este momento. SOKAL & ROHLF (1981: 551-554) fornecem o *teste de Bartlett* como alternativa, porém desprovido de teste de significância da reta. Lembro ainda que, caso desejemos verificar se há associação entre duas variáveis “independentes” podemos simplesmente usar o **teste de correlação** e evitar assim maiores problemas, e ainda, caso estejamos usando pacotes estatísticos que já disponibilizem a análise de regressão linear para amostras independentes, então nosso problema se limitará a saber quando usar o Modo I e o Modo II.

3.2.3.2.1 – Computação (Modo I)

A equação que resume a relação de dos valores de y (variável dependente) em x é (variável independente) a seguinte:

$$y = a + bx$$

onde, b é o coeficiente angular da reta (define o ângulo de inclinação da reta, isto é, o ângulo formado pela reta ao cortar o eixo dos x) e a expressa o valor de y para $x = 0$. Assim, conhecendo-se os valores de a e b podemos traçar a linha que melhor representa a dispersão de pontos. Contudo, queremos saber os valores individuais (pontos) de y . Logo, faremos a seguinte substituição:

$$\hat{Y} = a + bx$$

onde, $\hat{Y}$ representa os pontos de y por onde a reta passa.

Os parâmetros a e b serão determinados pelo *método dos mínimos quadrados*, como se segue:

1. Computação de b :

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{SP}{SQ_x}$$

onde, SP indica a *soma dos produtos* e SQ_x indica a soma dos quadrados de x .

2. Computação de a :

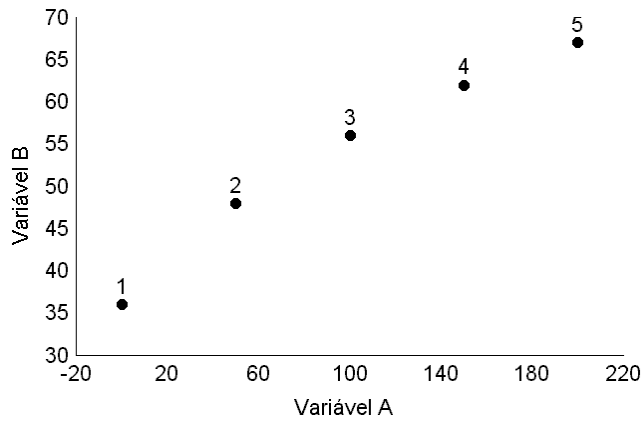
O valor de a será isolado da equação da reta:

$$a = \bar{Y} - b\bar{X}$$

Assim sendo, a reta calculada será a que apresentar o menor desvio em relação aos valores experimentais de y , isto é, aquela em que $\sum (y - \bar{Y})^2$ é mínima = método dos mínimos quadrados.

O seguinte exemplo foi extraído de CENTENO (1982: 134, modificado). Abaixo da tabela se encontra a plotagem dos valores de y em x .

Variáveis	Espécimes	1	2	3	4	5
Variável A (x)		0	50	100	150	200
Variável B (y)		36	58	56	62	67

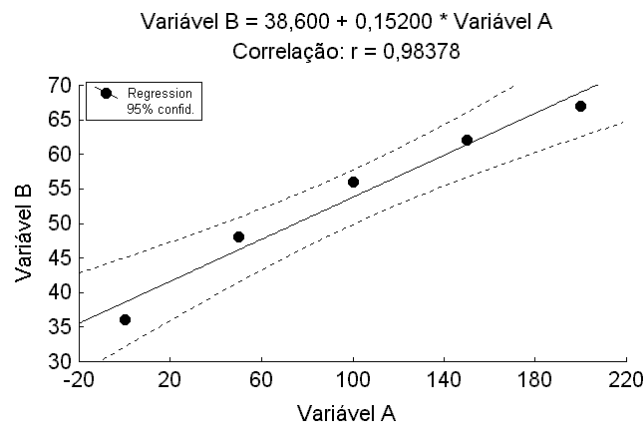


$$\sum x = 500; \sum y = 269; n = 5; \bar{X} = 100; \bar{Y} = 53,8$$

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{SP}{SQ_x} = \frac{(0 \times 36) + (50 \times 48) + \dots + (200 \times 67) - \frac{500 \times 269}{5}}{0^2 + 50^2 + 100^2 + 150^2 + 200^2 - \frac{500^2}{5}} = \frac{3.800}{25.000} = 0,1520$$

$$a = \bar{Y} - b\bar{X} = 38,60$$

$$\hat{Y} = 38,60 + 0,1520X$$



3.2.3.2.2 – Teste de hipótese sobre o ângulo de inclinação (β)

Destina-se a verificar se há regressão de y em x , ou não, através de uma análise de variância, isto é, se a reta traçada é significativa em termos estatísticos. A hipótese nula é a de que não há regressão de y em x .

Causas da variação	GL	SQ	QM	F
Regressão (desvio esperado, entre amostras)	1	$SQ_R = \sum (\hat{y} - \bar{y})^2$	$QM_R = \sum (\hat{y} - \bar{y})$	$F = \frac{QM_R}{QM_{DR}}$
Desvio da regressão (desvio não esperado, não controlável, resíduo, erro)	$n - 2$	$SQ_{DR} = \sum (y - \hat{y})^2$	$QM_{DR} = \frac{\sum (y - \hat{y})^2}{n - 2}$	
Total	$n - 1$	$SQ_y = \sum (y - \bar{y})^2$		

*A computação pode ser mais facilmente obtida:

- $SQ_y = \sum y^2 - \frac{(\sum y)^2}{n} = (36)^2 + (48)^2 + \dots + (67)^2 - (269)^2 / 5 = 15.472,2 = 596,8$
- $SQ_R = \frac{SP^2}{SQ_x} = \frac{3.800^2}{25.000} = \frac{14.440.000}{25.000} = 577,6$ [já calculamos SP e SQ na computação anterior da equação da reta, bastando substituir os valores]
- $SQ_D = SQ_y - SQ_R = 596,800 - 577,6 = 19,2$
- Os QM são obtidos dividindo-se os SQ pelos respectivos graus de liberdade, isto é, **1 e 3**.
- $F = \frac{577,6}{6,4} = 90,25$

Causas da variação	GL	SQ	QM	F
Regressão (desvio esperado, entre amostras)	1	577,6	577,6	90,25
Desvio da regressão (desvio não esperado, não controlável, resíduo, erro)	3	19,2	6,4	
Total	4	596,80		

$$F = 90,25 > F_{0,05 [1, 3]} = 10,138$$

Logo, concluímos que a hipótese nula foi rejeitada e que há regressão de y em x , ou de que a variável B é determinada pela variável A , ou ainda, que “a reta traçada é significativa em termos estatísticos”.

3.2.3.2.3 – Comparação entre n retas de regressão

Há casos onde desejamos saber se duas ou mais retas de regressão diferem significativamente uma da outra. Esta dúvida é pertinente e sua resolução bastante útil. Precisamos saber se (1) há diferença entre os coeficientes de inclinação da reta e (2) se o afastamento entre as duas retas pode ser atribuído a flutuações de amostra (H_0).

A diferença entre os coeficientes de inclinação é dada por uma análise de variância ou pelo teste t no caso de apenas duas amostras. SOKAL & ROHLF (1981: 499-509) fornecem exemplificação de n retas de regressão pela análise de variância e VANZOLINI (1993: 99-104) pelo teste t . O Programa Statistix for Windows ver. 8 nos permite calcular facilmente a diferença entre n retas de regressão pela análise de variância, evitando assim um cálculo tedioso. Mais adiante, no item 3.3.1.4, irei ilustrar a utilização da comparação entre retas de regressão em trajetórias ontogenéticas.

3.3 – Métodos morfométricos multivariados

São métodos que procuram analisar simultaneamente um conjunto n de variáveis e assim reduzir o número de análises a uma única, o que determina uma maior sumarização das relações de variância e covariância entre as variáveis e conseqüentemente maior poder de interpretação dos dados. Dentre vários métodos multivariados, dois são de uso rotineiro em taxonomia zoológica: **análise do componente principal (PCA)** e **análise discriminante múltipla** (= análise das variáveis canônicas). Embora sejam métodos bastante eficazes naquilo a que se destinam, trabalham rotineiramente, como no nosso caso, com distâncias morfométricas (p.ex., distância narina-olho) e, portanto, não refletem propriamente a variação na forma do organismo, mesmo que utilizemos métodos como o das “treliças”, bastante utilizado em taxonomia de peixes (STRAUSS & BOOKSTEIN, 1982), onde somente os comprimentos das linhas (externas) que ligam os pontos são computados, e não informação dos padrões de conexão entre estas linhas externas. Assim sendo, os métodos aqui

discutidos não têm por objetivo levar em consideração os padrões geométricos dos organismos: somente descrevem as covariâncias entre pares de variáveis (BOOKSTEIN, 1990: 75).

Outro aspecto importante é que tais métodos se baseiam em que a distribuição da amostra multivariada obedece a uma curva de distribuição normal. Contudo, tal exigência é discutível e, dependendo de nossos objetivos, pode ser descartada. A exigência de que haja normalidade em uma amostra multivariada somente se faz necessária quando estivermos testando hipóteses, já que a teoria estatística por trás baseou-se em amostras de distribuição normal. Contudo, como é no nosso caso, se o teste de hipótese não for o objeto principal da investigação (p.ex., uma MANOVA), então esta exigência pode ser descartada. Em outras palavras, se nosso interesse for **ordenar** nossos dados, isto é, torná-los graficamente orientados com o fim de representar as propriedades de uma amostra multivariada em algumas poucas dimensões, então a violação da normalidade desta amostra não é considerável (REYMENT, 1990: 123-124).

Cabe ressaltar que não faremos a computação manual neste tipo de análise, simplesmente inconcebível pela quantidade de variáveis, espécimes e dificuldades inerentes ao método. Teremos que procurar identificar quando usar estes métodos, as premissas a serem respeitadas e como interpretar os resultados.

3.3.1 – Análise dos componentes principais (PCA)

3.3.1.1 – **Objetivos:** destaco três objetivos principais para a análise dos componentes principais em taxonomia zoológica: investigar (1) o grau de variação morfométrica entre unidades taxonômicas operacionais (UTOs), ou morfoespécies, (2) o dimorfismo morfométrico entre os sexos, em caráter **exploratório**, já que neste método as amostras não são distribuídas em grupo de espécimes definidos *a priori*, o que o faz muito útil na discriminação de grupos a serem formados com base nas mensurações, e não em agrupamentos anteriores feitos com base na morfologia, padrões de desenho e colorido e (3) estudar os padrões de alometria multivariada.

3.3.1.2 – **O Método:** consiste em uma análise multivariada, amplamente utilizada, que sumariza as principais linhas de variação dentro de um conjunto de dados, reduzindo estas linhas a um pequeno número de variáveis independentes que incorporam a maior parte da informação original (STRAUSS & FUIMAN, 1985). As variáveis originais são transformadas em seus logaritmos decimais e calculada a matriz de variância e covariância. Este procedimento preserva as alometrias originais (JOLICOEUR, 1963) e não distorce o espaço multivariado (STRAUSS, 1987). Desta matriz, são obtidos os autovalores e seus autovetores associados que, combinados às variáveis originais, produzirão as combinações lineares que definem os componentes principais. Os escores dos indivíduos são então projetados no espaço reduzido dos componentes principais de maiores contribuições (CAVALCANTI & LOPES, 1993).

Quando os coeficientes obtidos no primeiro componente principal se mostrarem todos positivos ou negativos, este estará fortemente influenciado por variações no “tamanho” e os demais autovetores, alternando sinais positivos e negativos, indicando variações da “forma” (HUMPHRIES et al., 1981; REIS et al., 1988 e 1990; CAVALCANTI & LOPES, 1993). Esta interpretação reside no fato de que o primeiro componente (CP1) é o que captura a maior variação possível dos dados originais. O segundo componente (CP2) é ortogonal ao primeiro (independente) e fornece a variação máxima restante, e assim por diante. Partindo-se da premissa que o principal tipo de variação entre indivíduos é geralmente o tamanho, o CP1 pode ser interpretado como uma variável geral de tamanho e/ou mudanças relacionadas ao tamanho (alometrias), onde CP2 e os demais componentes são considerados como representantes da “forma”, desde que sejam independentes de CP1 (PERES-NETO & BIZERRIL, 1994).

HUMPHRIES *et al.* (1981), objetivando tornar os demais componentes independentes de CP1 (vetor “tamanho”), propuseram o método de distorção ("shearing") para os componentes principais acima de CP1. Todavia, este método é contestado por vários autores, pois não traria benefícios significantes, ou ainda, não elimina ou reduz a influência do tamanho (CP1) sobre os demais

vetores (CP2 e demais), fornecendo resultados idênticos. Entretanto, minha experiência com vários grupos de distâncias morfométricas em anuros mostra que, dependendo dos dados existentes na matriz original, os resultados podem apresentar discriminações superiores ao do PCA tradicional ($\neq$ “Common PCA”).

Outro aspecto considerável é que, quando usamos amostras populacionais derivadas de populações supostamente distintas, o método implica na premissa de que a direção de variação máxima (CP1) nos diferentes grupos é idêntica (NEFF & MARCUS, 1980). Esta premissa será testada calculando-se o coeficiente de correlação vetorial (MORRISON, 1976; REIS *et al.*, 1988) entre os coeficientes do primeiro componente principal, calculado para cada grupo em separado. Estes coeficientes serão extraídos do produto interno dos coeficientes dos caracteres morfométricos no primeiro componente principal. Os coeficientes de correlação vetorial irão indicar o grau de semelhança entre os componentes principais (STRAUSS & FUIMAN, 1985; REIS *et al.*, 1988; CAVALCANTI & LOPES, 1991). Os valores dos coeficientes de correlação vetorial próximos a zero indicam direções de variação diferentes (ortogonalidade dos autovetores), enquanto que valores próximos à unidade, direções de variação idêntica (coincidência dos vetores) (REIS *et al.*, 1988).

3.3.1.3 – Exemplo em discriminação de amostras: o exemplo seguinte ilustra a utilização do PCA (método de “shearing”) para verificar o grau de dimorfismo sexual na UTO Santa Teresa (NAPOLI, 2000). Foram combinadas todas as amostras de machos e fêmeas, utilizando-se quinze caracteres morfométricos. A correlação vetorial entre as duas amostras foi alta ($P>0,99$), satisfazendo a premissa de que as direções de variação são semelhantes. O primeiro componente principal obteve todos os valores positivos e o segundo alternou valores positivos e negativos, indicando, respectivamente, “tamanho” e “forma” (HUMPHRIES *et al.*, 1981). A projeção dos escores individuais resultantes da análise (**Fig. 8**) denotou discriminação completa entre machos e fêmeas ao longo do segundo componente (“forma”), sendo os caracteres que mais contribuíram neste vetor o diâmetro do olho e a largura da cabeça (**Tabela 9**). A observação da estatística descritiva para estes dois caracteres, revelou que o diâmetro do olho pouco difere entre os dois sexos ($t=0,10$, $GL=65$, $P<0,9170$), mas a largura da cabeça apresenta-se consideravelmente maior em machos do que em fêmeas ($t=4,68$, $GL=65$, $P<0,0001$). Além disso, machos e fêmeas não apresentam comprimentos totais estatisticamente diferentes ($t=1,77$, $GL=65$, $P<0,0808$) e as fêmeas tendem a possuir o diâmetro do tímpano pouco maior que o dos machos ($t=2,64$, $GL=65$, $P<0,0103$).

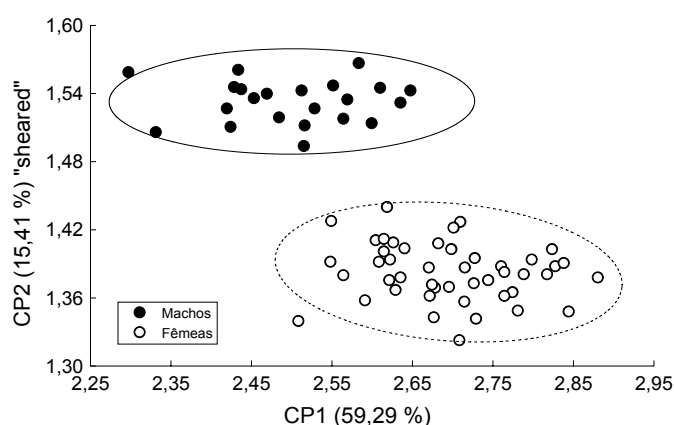


Figura 8. Projeção dos escores individuais resultantes da análise dos componentes principais em machos e fêmeas adultas UTO Santa Teresa, localizados em Santa Teresa, Espírito Santo, utilizando-se quinze caracteres morfométricos. O primeiro componente apresentou todos os coeficientes positivos (“tamanho”), enquanto o segundo alternou valores positivos e negativos (“forma”). Notar a discriminação das amostras ao longo do segundo eixo, indicando “formas” diferentes entre os sexos. (NAPOLI, 2000).

Tabela 9. Coeficientes padronizados derivados da análise dos componentes principais (**Fig. 8**) dos espécimes machos e fêmeas combinados da UTO Santa Teresa, para quinze caracteres morfométricos (NAPOLI, 2000).

Caracteres (variáveis)	CP1	CP2
CRC	0,0932	0,2393
CC	0,0800	0,2728
LC	0,0436	0,3462
DO	0,3846	-0,3531
DON	0,3369	-0,2005
DT	0,3626	-0,0830
LPS	0,4216	-0,2980
DIO	0,2522	-0,1360
DIN	0,2908	-0,1696
DNF	0,2546	-0,0481
CCX	0,0543	0,2906
CTB	0,0551	0,2743
CP	0,0411	0,2817
D3D	0,3142	-0,1168
D4A	0,3050	-0,0758

3.3.1.4 – Exemplo no traçado de trajetórias ontogenéticas e cálculo de alometrias multivariadas:

A análise dos componentes principais também pode ser empregada para se estudar os padrões de alometria multivariada (JOLICOEUR, 1963), úteis na determinação de diferenças morfométricas ao longo do crescimento, em tamanho do corpo, dos espécimes da amostra (causas ontogenéticas). Tal procedimento é equivalente, no caso bivariado, à suposição de que o coeficiente angular da reta de regressão é semelhante nos diferentes grupos em estudo. Estes coeficientes podem ser calculados a partir do produto interno dos coeficientes do primeiro componente principal, computado separadamente para cada grupo, ou pela seguinte computação:

$$Alom_i = \frac{Coef.CPI_i}{\sqrt{1/V}}$$

onde, $Alom_i$ é o coeficiente alométrico multivariado da variável i ; $Coef. CPI_i$ é o coeficiente da variável i no primeiro componente; $\sqrt{1/V}$ é o valor hipotético para um caráter isométrico. Assim, dividindo-se o $Coef. CPI_i$ por $\sqrt{1/V}$ produzimos um rescalonamento das variáveis, onde valores abaixo de um (01) indicam alometria negativa, acima de um (01) alometria positiva e iguais a um (01) isometria.

As trajetórias ontogenéticas consistem em retas de regressão linear dos escores de n amostras no espaço reduzido da projeção. Pelas retas de regressão tem-se a previsão de quando as amostras se aproximam ou se distanciam morfometricamente ao longo do crescimento, desde que o primeiro componente represente o tamanho. Tanto as trajetórias ontogenéticas, quanto as alometrias podem ser apresentadas em conjunto, o que facilita bastante a interpretação das variação da forma em função do tamanho (crescimento).

A **Fig. 9** (NAPOLI & CARAMASCHI, 2000) ilustra este tipo de utilização. Neste resultado, duas amostras, representadas por grupos de sub-amostras geográficas, morfologicamente diferentes, mas com alguns indivíduos similares, são contrapostas. Os indivíduos que mais se sobrepõem são justamente aqueles que apresentam os tamanhos mais similares (**Fig. 10A**). Tal resultado mostra que a diferença entre as localidades geográficas (Ia, IIa e IIIa) para a morfoespécie SSP ocorre principalmente pela diferença no tamanho dos espécimes, ditado por variação ontogenética, isto é, se todos fossem do mesmo tamanho seriam morfometricamente, e portanto morfologicamente, similares. Assim, foram considerados pertencentes a mesma espécie biológica. As alometrias são fornecidas na **Fig. 10B**.

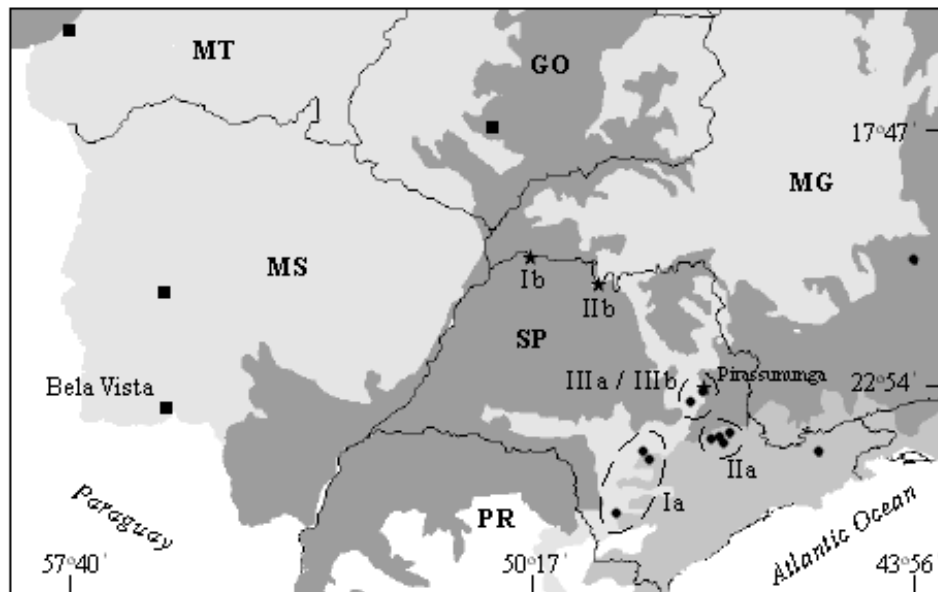


Figura 9. Geographic distribution of morphotypes SSP (circles), NSP (stars) and CBR (squares) of *Hyla elianeae* in Brazil. Nine geographic samples of morphotype SSP throughout the “cerrado” corridor of São Paulo were grouped into three areas (Ia, IIa, and IIIa), based on geographic proximity and morphological similarity, in order to analyze morphological variation. The samples of morphotype NSP were named Ib, IIb, and IIIb following the same criteria for morphotype SSP. Shadings represent, from darker to lighter, respectively: deciduous tropical forest, Atlantic forest, and “cerrado”. GO, State of Goiás; MG, State of Minas Gerais; MS, State of Mato Grosso do Sul; MT, State of Mato Grosso; PR, State of Paraná; SP, State of São Paulo. (NAPOLI & CARAMASCHI, 2000).

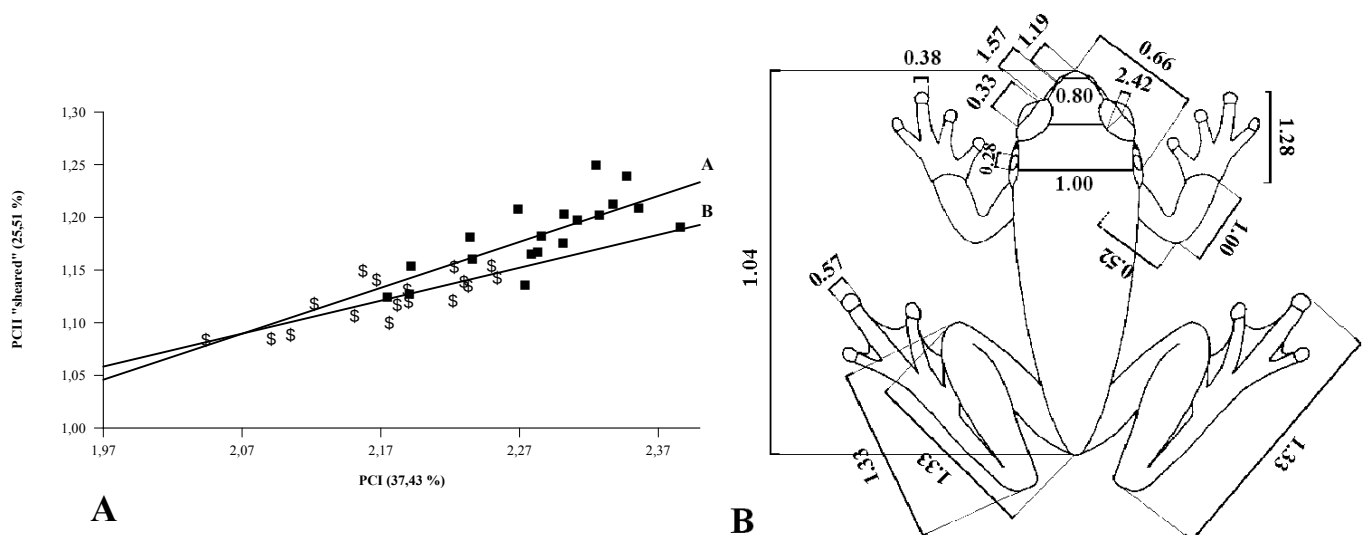


Figura 10. (A) Projection of individual scores resulted from the principal component analysis, using the “shearing method” (HUMPHRIES *et al.*, 1981), of 17 morphometric characters for the combined sample areas Ia (close squares) and IIa / IIIa (open squares) from morphotype SSP. A and B are regression lines of the individual scores of the second component by the first, respectively, for sample areas Ia and IIa / IIIa (ontogenetic trajectories). Both regressions are statistically significant for $P < 0.01$. (B) Allometric multivariate coefficients for combined sample areas Ia, IIa, and IIIa of morphotype SSP. A value lower than 1 indicates negative allometry; higher than 1, positive allometry; equal to 1, an isometric character (NAPOLI & CARAMASCHI, 2000).

3.3.2 – Análise das variáveis canônicas ou análise discriminante múltipla

É um procedimento estatístico multivariado que permite comparar o grau de variação existente dentro de cada amostra com o grau de variação entre as amostras. Esta análise tem a propriedade de maximizar as diferenças entre os grupos (estabelecidos *a priori*) e indicar quais os caracteres que mais contribuíram para tal discriminação ao longo de cada variável canônica.

Os grupos são previamente definidos, sendo então realizado o produto da matriz de covariância entre os grupos pela inversa da matriz de covariância combinada dentro dos grupos. Os eixos canônicos são obtidos desta matriz pela extração de autovalores e seus autovetores associados, sendo os escores dos indivíduos projetados no espaço reduzido dos eixos canônicos, permitindo-se uma análise gráfica das discriminações entre as espécies (CAVALCANTI & LOPES, 1993).

Os coeficientes obtidos para cada autovetor fornecem o grau de contribuição de cada um destes na referida análise. Contudo, se dois caracteres são correlacionáveis, o coeficiente do caráter que tiver o menor F será próximo de zero, mesmo que este caráter seja tão importante quanto os demais (REIS *et al.*, 1990). Assim sendo, evitamos tal infortúnio computando os “loadings”, isto é, transformamos os coeficientes em vetores de correlação, calculando-se a correlação (Pearson) entre os escores individuais dos componentes principais e os valores originais dos caracteres para cada indivíduo (STRAUSS, 1985; REIS *et al.*, 1990). A significância de cada coeficiente ($P < 0,05$) será fornecida por uma tabela de Correlação (Pearson). A significância dos autovetores segue o critério de Wilks ($P < 0,05$). Adicionalmente, quando se julgar oportuno, podem ser fornecidas Tabelas de Classificação, onde são relacionados para cada grupo o número de indivíduos inclusos nos mesmos. A análise discriminante múltipla tem sido empregada por mim com dois objetivos principais: (1) testar a possibilidade de discriminação por morfometria de morfoespécies (ou UTOs) envolvidas e, (2) quando uma morfoespécie apresentar uma quantidade considerável de amostras populacionais e de espécimes contidos nestas, verificar o grau de semelhança entre suas populações. Ressalto que, quando uma morfoespécie é composta por várias populações distintas e morfologicamente heterogênea, e menos de 95 % de uma ou mais amostras são discriminadas dos demais grupos, utilizei a análise discriminante múltipla entre as amostras de localidades de maior representatividade (“amostras básicas”), avaliando-se a morfoespécie com uma maior homogeneidade e uma menor interferência por parte de possíveis espécimes híbridos, representantes de populações locais mal representadas e ainda de outros possivelmente associados de forma errônea às amostras principais.

Em casos de representação tridimensional é comum ocorrer distorção das relações de proximidade real entre os centróides. Para tal, usamos linhas que interligam um elemento centróide de uma morfoespécie (o sinal + na **Fig. 11**) ao centróide de outra morfoespécie que estiver mais próximo (“minimum spanning tree”). Esta ligação de centróides se baseia no cálculo das distâncias de Mahalanobis entre os mesmos. A **Fig. 11** ilustra a discriminação de três morfoespécies atribuídas por NAPOLI & CARAMASCHI (2000) como pertencentes a *Hyla elianeae*: SSP, Sul de São Paulo; NSP, Norte de São Paulo; CBR, Brasil Central. Note que justamente SSP e NSP, mais próximas geograficamente, têm discriminação acentuada ao longo do segundo vetor canônico. A Tabela 10 fornece os coeficientes e “loadings”. **Procure interpretar os resultados e discuti-los em aula!**

3.3.3 – Análise de agrupamento (“cluster”)

Análises de agrupamento podem ser utilizadas para os caracteres qualitativos e morfométricos, utilizando-se o método de UPGMA (*unweighted pair-group method using arithmetic averages*). Este algoritmo computa a média de similaridade ou dissimilaridade de uma UTO (unidade taxonômica operacional) para com um agrupamento relacionado, pesando cada UTO dentro deste agrupamento de forma idêntica (SNEATH & SOKAL, 1973), isto é, a distância entre dois agrupamentos é calculada como a distância média entre todos os pares de objetos em dois agrupamentos diferentes. Para a confecção do UPGMA pode-se calcular a tradicional *distância euclidiana* entre os objetos da análise, a qual corresponde à distância geométrica entre os objetos da

análise no espaço multidimensional ou, quando dispomos da *distância de Mahalanobis* derivada da análise discriminante, fornecer um UPGMA das distâncias entre os centróides, o que pode incrementar a compreensão da análise discriminante através da definição dos principais grupos formados, útil em taxonomia (**Fig. 12-13**; NAPOLI, 1995: 67).

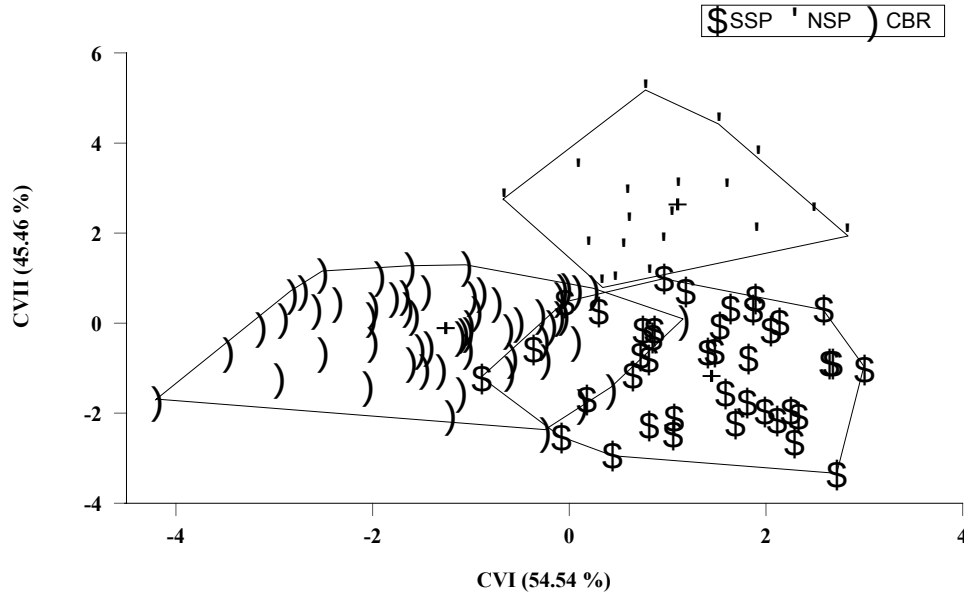


Figura 11. – Projection of individual scores resulted from the multiple discriminant function analysis for 18 morphometric characters of the combined samples of morphotypes SSP, NSP, and CBR in the space of the first with the second canonical axes. NAPOLI & CARAMASCHI, 2000.

Tabela 10. Standardized coefficients and loadings from discriminant function analysis for morphometric characters of combined samples from morphotypes SSP, NSP, and CBR.

Characters	CVI	CVII	<i>r</i> (I)	<i>r</i> (II)
SVL	-0.36	-0.16	-0.35***	-0.03ns
HW	-0.62	0.35	-0.45***	-0.01ns
HL	-0.61	0.60	-0.54***	0.21*
ED	-0.07	0.35	-0.27***	0.57***
UEW	0.44	0.25	-0.03ns	0.11ns
IOD	0.40	0.26	0.30***	0.12ns
END	-0.07	-0.50	-0.20*	-0.51***
IND	-0.13	-0.48	-0.05ns	-0.36***
THL	-0.82	-1.37	0.00ns	-0.37***
TL	1.98	1.12	0.15ns	-0.15ns
TD	-0.40	-0.30	-0.52***	-0.05ns
NSD	0.12	-0.08	-0.15ns	-0.31***
UAR	0.24	0.35	0.01ns	0.20*
FAR	-0.03	-0.19	0.06ns	0.00ns
HAL	-0.06	-0.34	-0.04ns	-0.32***
3FD	-0.39	-0.01	-0.28***	0.29***
FL	0.00	-0.06	0.04ns	-0.24**
4TD	0.37	0.06	-0.00ns	0.14

r, correlation coefficient (Pearson) of the original data with the scores obtained from the discriminant function analysis (=loadings); ns, not significant; * $P < 0.05$; ** $P < 0.02$; *** $P < 0.01$.

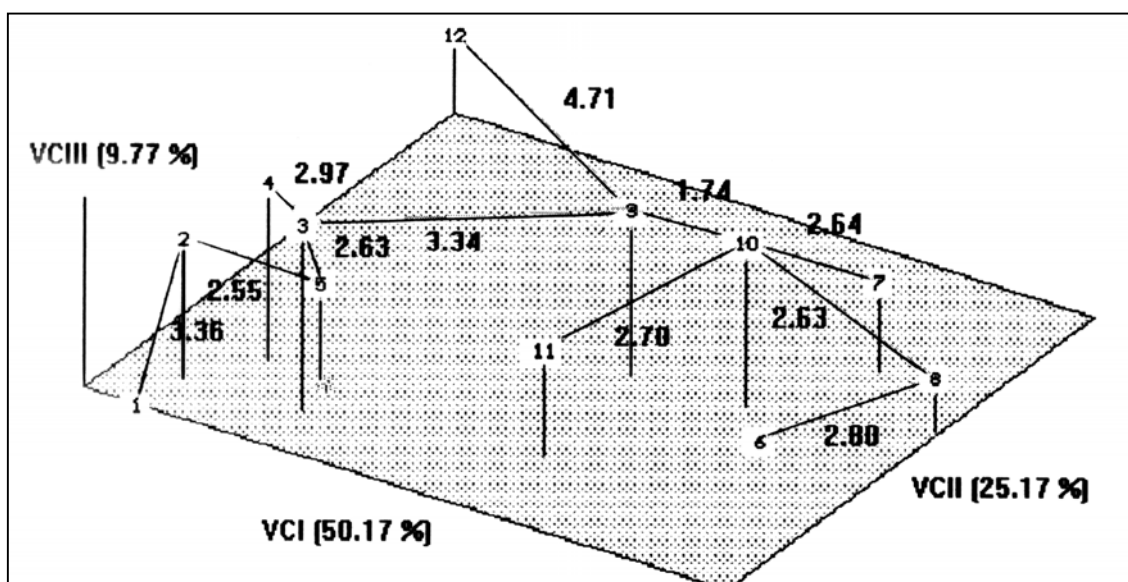


Figura 12 – Projeção dos centróides obtidos na análise discriminante múltipla das amostras combinadas de doze morfoespécies no espaço reduzido das variáveis canônicas I, II e III. Os algarismos arábicos representam as morfoespécies: 1 = TTPira, 2 = TTBotu, 3 = TTArá, 4 = TTRRP, 5 = TTMG, 6 = SPA, 7 = RUM, 8 = RDI, 9 = RU, 10 = PRU, 11 = CBO e 12 = ANA. As linhas que interligam as morfoespécies representam aquelas mais próximas umas das outras ("minimum spanning tree"); estas se baseiam no cálculo das distâncias entre os centróides (Mahalanobis) e têm como objetivo corrigir a distorção provocada pela projeção tridimensional.

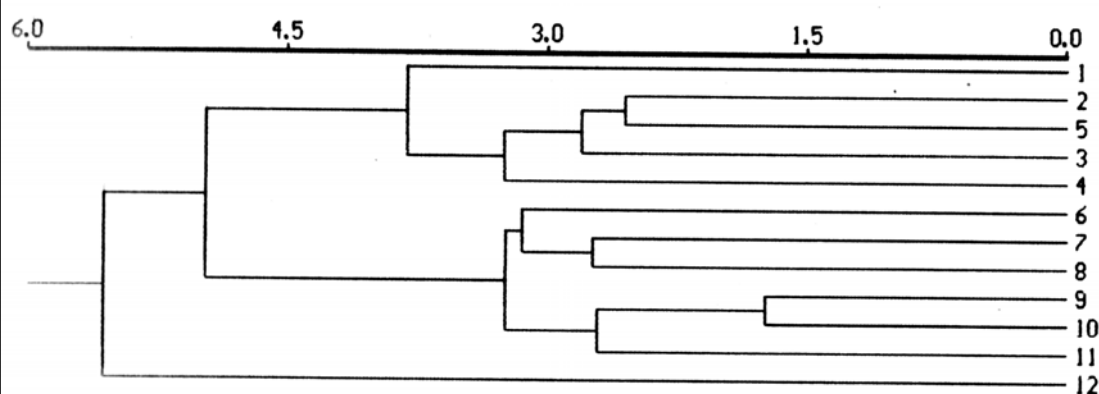


Figura 13 – Análise de agrupamento (UPGMA) para as doze morfoespécies do grupo de *Hyla rubicundula* a partir das distâncias (Mahalanobis) obtidas na análise das variáveis canônicas cujos escores dos três primeiros eixos se encontram projetados na Figura 9. Tal procedimento tem por finalidade única demonstrar mais claramente as afinidades morfométricas entre as morfoespécies envolvidas. Nota-se que os subgrupos A, B, C e D propostos anteriormente com base nos padrões de forma e desenho, se enquadram bem neste resultado. O subgrupo A é constituído no esquema pelas morfoespécies 9, 10 e 11; o subgrupo B pela morfoespécie 12; o subgrupo C pelas morfoespécies 6, 7 e 8; e o subgrupo D pelas morfoespécies 1, 2, 3, 4 e 5. A identificação de cada morfoespécie se encontra na legenda da Figura

IV REFERÊNCIAS BIBLIOGRÁFICAS

- AB'SÁBER, A. N., 1977. Os domínios morfoclimáticos na América do Sul. Primeira aproximação. **Geomorfologia**, 52:1-23.
- BOOKSTEIN, F. L., 1990. Multivariate Methods. Part III. Section A. In: Rohlf, F. J. & Bookstein, F. L. (eds.), **Proceedings of the Michigan Morphometrics Workshop**, The University of the Michigan Museum of Zoology, Special Publication Number 2:75-76.
- CAVALCANTI, M. J. & LOPES, P. R. D., 1991. Morfometria comparada de *Ctenosciaena gracilicirrhus*, *Paralonchurus brasiliensis* e *Micropogonias furnieri* (Teleostei, Scianidae) pela análise multivariada de redes de treliças. **Revta. Bras. Zool.**, 7(4):627-635.
- CAVALCANTI, M. J. & LOPES, P. R. D., 1993. Análise morfométrica multivariada de cinco espécies de Serranidae (Teleostei, Perciformes). **Acta Biologica Leopoldensia**, 15(1):53-64.
- CENTENO, A. J., 1982. **Curso de estatística aplicada à biologia**. Goiânia, Universidade Federal de Goiás, 188p.
- GRIZZLE, J. E., 1967. Continuity correction in the χ^2 -test for 2x2 tables. **Amer. Stat.**, 21(4):28-32.
- HEYER, W. R., 1984. Variation, systematics, and zoogeography of *Eleutherodactylus guentheri* and closely related species (Amphibia: Anura: Leptodactylidae). **Smithsonian Contributions to Zoology**, 402, 42p.
- HUMPHRIES, J. M.; BOOKSTEIN, F. L.; CHERNOFF, B.; SMITH, G. R.; ELDER, R. L. & POSS, S. G., 1981. Multivariate discrimination by shape in relation to size. **Syst. Zool.**, 30(3):291-308.
- JOLICOEUR, P., 1963. The multivariate generalization of the allometry equation. **Biometrics**, 19:497-499.
- MAYR, E., 1942. **Systematics and the origin of species**. Columbia University Press, New York, 334p.
- MORRISON, D. F., 1976. Multivariate statistical methods. New York, McGraw-Hill, ?p.
- NEFF, N. A. & MARCUS, L. F., 1980. A survey of multivariate methods for Systematics. New York, Privately published, ?p.
- NAPOLI, M. F., 1995. **Taxonomia das espécies de *Hyla* pertencentes ao grupo de *H. rubicundula* Reinhardt & Lütken, 1862 (Amphibia, Anura, Hylidae)**. Dissertação de Mestrado. Rio de Janeiro, Museu Nacional/UFRJ, xv + 287p.
- NAPOLI, M. F., 2000. **Taxonomia, variação morfológica e distribuição geográfica das espécies do grupo de *Hyla circumdata* (Cope, 1870) (Amphibia, Anura, Hylidae)**. Tese de Doutorado. Rio de Janeiro, Museu Nacional/UFRJ, xii + 208p.
- NAPOLI, M. F. & CARAMASCHI, U., 1999. Geographic variation of *Hyla rubicundula* and *Hyla anataliasiasi*, with description of a new species (Anura, Hylidae). **Alytes**, 16(3-4):165-189.

- PERES-NETO, P. R. & BIZERRIL, C. R. S. F., 1994. The jackknifing of multivariate allometric coefficient (Jolicoeur, 1963): a case study on allometry and morphometric variation in *Corydoras barbatus* (Quoy & Gaimard, 1824) (Siluriformes, Callichthyidae). **Arq. Biol. Tecnol.**, **37**(2):449-454.
- REIS, S. F.; CRUZ, J. F. DA & VON ZUBEN, C. J., 1988. Análise multivariada da evolução craniana em roedores caviíneos: convergência de trajetórias ontogenéticas. **Rev. Brasil. Genet.**, **11**(3):633-641.
- REIS, S. F.; PESSÔA, L. M. & STRAUSS, R. E., 1990. Application of size-free canonical discriminant analysis to studies of geographic differentiation. **Rev. Brasil. Genet.**, **13**(3):509-520.
- REYMENT, R. A., 1990. Reification of classical multivariate statistical analysis on morphometry. In: Rohlf, F. J. & Bookstein, F. L. (eds.), **Proceedings of the Michigan Morphometrics Workshop**, The University of the Michigan Museum of Zoology, Special Publication Number 2:123-144.
- ROHLF, F. J. & SOKAL, R. R., 1981. **Statistical tables**. 2 ed. New York, W. H. Freeman and Company, xiii+219p.
- SHAPIRO, S. S., WILK, M. B. & CHEN, H. J., 1968. A comparative study of various tests of normality. **Journal of the American Statistical Association**, **63**:1343-1372.
- SIMPSON, G. G. & ROE, A., 1939. **Quantitative Zoology: Numerical Concepts and Methods in the Study of Recent and Fossil Animals**. New York, McGraw-Hill Book Company, Inc., 414p.
- SNEATH, P. H. A. & SOKAL, R. R., 1973. **Numerical taxonomy**. San Francisco, W. H. Freeman & Co., ?p.
- SOKAL, R. R. & ROHLF, F. J., 1981. **Biometry: the principles and practice of statistics in biological research**. 2 ed. New York, W. H. Freeman and Company, 859p.
- STRAUSS, R. E., 1985. Evolutionary allometry and variation in body form in the South American catfish Genus *Corydoras* (Callichthyidae). **Syst. Zool.**, **34**(4):381-396.
- STRAUSS, R.E., 1987. The importance of phylogenetic constraints in comparisons of morphological structure among fish assemblages. In: W. J. Mathews & D. C. Heins (eds.), **Community and evolutionary ecology of North American stream fishes**. University of Oklahoma Press, ?p.
- STRAUSS, R. E. & BOOKSTEIN, F. L., 1982. The truss: body form reconstructions in morphometrics. **Syst. Zool.**, **31**:113-135.
- STRAUSS, R. E. & FUIMAN, L. A., 1985. Quantitative comparisons of body form and allometry in larval and adult Pacific sculpins (Teleostei: Cottidae). **Can. J. Zool.**, **63**:1582-1589.
- VANZOLINI, P. E., 1970. **Zoologia Sistemática, Geográfica e a Origem das Espécies**. Universidade de São Paulo e Instituto de Geografia, Teses e Monografias, 356p.
- VANZOLINI, P. E., 1993. **Métodos estatísticos elementares em sistemática zoológica**. São Paulo, Editora Hucitec, 130p.

ANEXO

Tabela 1

(ROHLF & SOKAL, 1981: 98-99)

TABLE 14 Critical values of the chi-square distribution

This table furnishes critical values for the chi-square distribution for degrees of freedom $\nu = 1$ to 100 in increments of 1. The percentage points given (corresponding to $\alpha = 0.995, 0.975, 0.9, 0.5, 0.1, 0.05, 0.025, 0.01, 0.005$, and 0.001) represent the area to the right of the critical value of χ^2 in one tail of the distribution, as shown in the accompanying figure. The critical values of χ^2 are given to three significant decimal places, except in cases when $\chi^2 > 100$, when they are given to two significant decimal places.

To find the critical value of χ^2 for a given number of degrees of freedom, look up ν in the left (argument) column of the table and read off the desired values of χ^2 in that row. For example, for eight degrees of freedom, $\chi^2_{.05|8} = 15.507$ and $\chi^2_{.01|8} = 20.090$. The last value indicates that 1% of the area of the chi-square distribution (for eight degrees of freedom) is to the right of the value of $\chi^2 = 20.090$. For values of $\nu > 100$, compute approximate critical values of χ^2 by formula as follows: $\chi^2_{\alpha|\nu} = \frac{1}{2}(t_{2\alpha|\nu-1} + \sqrt{2\nu-1})^2$, where $t_{2\alpha|\nu-1}$ can be looked up in Table 12. Thus $\chi^2_{.05|120} = \frac{1}{2}(17.10462)^2 = 146.284$. For $\alpha > 0.5$, employ $t_{1-2\alpha|\nu-1}$ in the above formula. When $\alpha = 0.5$, $t_{2\alpha} = 0$.

There are numerous applications of the chi-square distribution in statistics. The distribution is described in Section 7.6. It is used to set confidence limits to sample variances (Section 7.7), to test the homogeneity of variances (Section 13.3) and of correlation coefficients (Section 15.5), and for tests of goodness of fit (Chapter 17).

Except for $\alpha = 0.001$, the values of chi-square from 1 to 30 degrees of freedom have been taken from a more extensive table by C. M. Thompson (*Biometrika* 32:188-189, 1941) with permission of the publisher. Values between 31 and 100 degrees of freedom were approximated using the Cornish-Fisher asymptotic expansion following the account of M. Zelen and N. C. Severo, section 26.2.49 of M. Abramovitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions* (U.S. National Bureau of Standards, 1964). The values of the Hermite polynomials were computed, instead of using the tables furnished in the source above. The values for $\alpha = 0.001$ were all taken from H. L. Harter, *New Tables of the Incomplete Gamma-function Ratio and of Percentage Points of the Chi-Square and Beta Distribution* (U.S. Government Printing Office, 1964).

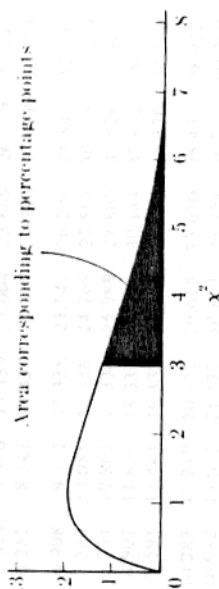


TABLE 14 Critical values of the chi-square distribution

ν	α	.995	.975	.9	.5	.1	.05	.025	.01	.005	.001	α/ν
1	0.000	0.000	0.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.828	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	12.592	13.816	2
3	0.072	0.216	0.584	2.366	7.815	9.348	11.345	13.277	14.860	16.266	17.341	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	16.266	17.341	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	20.515	20.515	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.458	22.458	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	24.453	24.453	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	26.124	26.124	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	27.877	27.877	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	29.588	29.588	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	31.264	31.264	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	32.910	32.910	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	34.528	34.528	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	36.123	36.123	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	37.697	37.697	15
16	5.142	6.908	9.312	15.338	23.542	26.296	28.845	32.000	34.267	39.252	39.252	16
17	5.697	7.564	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790	40.790	17
18	6.265	8.231	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312	42.312	18
19	6.846	8.907	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820	43.820	19
20	7.434	9.591	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315	45.315	20
21	8.034	10.283	13.240	20.337	29.615	32.670	35.479	38.932	41.401	46.797	46.797	21
22	8.643	10.982	14.042	21.337	30.813	33.924	36.781	40.289	42.796	48.280	48.280	22
23	9.260	11.688	14.848	22.337	32.007	35.172	38.076	41.634	44.181	49.728	49.728	23
24	9.886	12.401	15.659	23.337	33.196	36.415	39.364	42.980	45.558	51.179	51.179	24
25	10.526	13.120	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620	52.620	25
26	11.160	13.844	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052	54.052	26
27	11.808	14.573	18.114	26.336	36.741	40.113	43.194	46.963	49.645	55.476	55.476	27
28	12.461	15.308	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892	56.892	28
29	13.121	16.047	19.768	28.336	39.088	42.557	45.722	49.588	52.336	58.301	58.301	29
30	13.787	16.791	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703	59.703	30
31	14.458	17.539	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098	61.098	31
32	15.134	18.291	22.271	31.336	42.585	46.194	49.480	53.486	56.329	62.487	62.487	32
33	15.815	19.047	23.110	32.336	43.745	47.400	50.725	54.776	57.649	63.870	63.870	33
34	16.501	19.806	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247	65.247	34
35	17.192	20.569	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619	66.619	35
36	17.887	21.336	25.643	35.336	47.212	50.998	54.437	58.619	61.582	67.985	67.985	36
37	18.586	22.106	26.492	36.335	48.363	52.192	55.668	59.892	62.884	69.346	69.346	37
38	19.289	22.878	27.343	37.335	49.513	53.384	56.896	61.162	64.182	70.703	70.703	38
39	19.996	23.654	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055	72.055	39
40	20.707	24.433	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402	73.402	40
41	21.421	25.213	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745	74.745	41
42	22.138	25.999	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084	76.084	42
43	22.859	26.785	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419	77.419	43
44	23.584	27.575	32.487	43.335	56.369	60.481	64.202	68.710	71.893	78.750	78.750	44
45	24.311	28.366	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077	80.077	45
46	25.042	29.160	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400	81.400	46
47	25.775	29.956	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720	82.720	47
48	26.511	30.755	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037	84.037	48
49	27.249	31.555	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351	85.351	49
50	27.991	32.357	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661	86.661	50

TABLE 14 Critical values of the chi-square distribution

ν	α	.995	.975	.9	.5	.1	.05	.025	.01	.005	.001	α/ν
51	28.735	33.162	38.560	50.335	64.295	68.669	72.616	77.386	80.747	87.968	87.968	51
52	29.481	33.968	39.433	51.335	65.422	69.832	73.810	78.616	82.001	89.272	89.272	52
53	30.230	34.776	40.308	52.335	66.548	70.993	75.002	79.843	83.253	90.573	90.573	53
54	30.981	35.586	41.183	53.335	67.673	72.153	76.192	81.069	84.502	91.872	91.872	54
55	31.735	36.398	42.060	54.335	68.796	73.311	77.380	82.292	85.749	93.168	93.168	55
56	32.490	37.212	42.937	55.335	69.918	74.468	78.567	83.513	86.994	94.460	94.460	56
57	33.248	38.027	43.816	56.335	71.040	75.752	79.752	84.733	88.237	95.751	95.751	57
58	34.008	38.844	44.696	57.335	72.160	76.778	80.936	85.950	89.477	97.039	97.039	58
59	34.770	39.662	45.577	58.335	73.279	77.931	82.117	87.166	90.715	98.324	98.324	59
60	35.534	40.482	46.459	59.335	74.397	79.082	83.298	88.379	91.952	99.607	99.607	60
61	36.300	41.303	47.342	60.335	75.514	80.232	84.476	89.591	93.186	100.888	100.888	61
62	37.068	42.126	48.226	61.335	76.630	81.381	85.654	90.802	94.419	102.166	102.166	62
63	37.838	42.950	49.111	62.335	77.745	82.529	86.830	92.010	95.649	103.442	103.442	63
64	38.610	43.776	49.996	63.335	78.860	83.675	88.004	93.217	96.878	104.716	104.716	64
65	39.383	44.603	50.883	64.335	79.973	84.821	89.177	94.422	98.105	105.988	105.988	65
66	40.158	45.431	51.770	65.335	81.085	85.965	90.349	95.626	99.331	107.258	107.258	66
67	40.935	46.261	52.659	66.335	82.197	87.108	91.519	96.828	100.55	108.526	108.526	67
68	41.713	47.092	53.548	67.334	83.308	88.250	92.689	98.028	101.78	109.791	109.791	68
69	42.494	47.924	54.438	68.334	84.418	89.391	93.856	99.228	103.00	111.055	111.055	69
70	43.275	48.758	55.329	69.334	85.527	90.531	95.023	100.43	104.21	112.317	112.317	70
71	44.058	49.592	56.221	70.334	86.635	91.670	96.189	101.62	105.43	113.577	113.577	71
72	44.843	50.428	57.113	71.334	87.743	92.808	97.353	102.82	106.65	114.835	114.835	72
73	45.629	51.265	58.006	72.334	88.850	93.945	98.516	104.01	107.86	116.092	116.092	73
74	46.417	52.103	58.900	73.334	89.956	95.081	99.678	105.20	109.07	117.346	117.346	74
75	47.206	52.942	59.795	74.334	91.061	96.217	100.84	106.39	110.29	118.599	118.599	75
76	47.997	53.782	60.690	75.334	92.166	97.351	102.00	107.58	111.50	119.850	119.850	76
77	48.788	54.623	61.586	76.334	93.270	98.484	103.16	108.77	112.70	121.100	121.100	77
78	49.582	55.466	62.483	77.334	94.373	99.617	104.32	109.96	113.91	122.348	122.348	78
79	50.376	56.309	63.380	78.334	95.476	100.75	105.47	111.14	115.12	123.594	123.594	79
80	51.172	57.153	64.278	79.334	96.578	101.88	106.63	112.33	116.32	124.839	124.839	80
81	51.969	57.998	65.176	80.334	97.680	103.01	107.78	113.51	117.52	126.082	126.082	81
82	52.767	58.845	66.076	81.334	98.780	104.14	108.94	114.69	118.73	127.324	127.324	82
83	53.567	59.692	66.976	82.334	99.880	105.27	110.09	115.88	119.93	128.565	128.565	83
84	54.368	60.540	67.876	83.334	100.98	106.39	111.24	117.06	121.13	129.804	129.804	84
85	55.170	61.389	68.777	84.334	102.08	107.52	112.39	118.24	122.32	131.041	131.041	85
86	55.973	62.239	69.679	85.334	103.18	108.65	113.54	119.41	123.52	132.277	132.277	8

Tabela 2: Pontos da distribuição de t
(ROHLF & SOKAL, 1981: 81)

TABLE 12 Critical values of Student's t -distribution.

$\nu \backslash \alpha$	0.9	0.5	0.4	0.2	0.1	0.05	0.02	0.01	0.001	$\alpha \backslash \nu$
1	.158	1.000	1.376	3.078	6.314	12.706	31.821	63.657	636.619	1
2	.142	.816	1.061	1.886	2.920	4.303	6.965	9.925	31.598	2
3	.137	.765	.978	1.638	2.353	3.182	4.541	5.841	12.924	3
4	.134	.741	.941	1.533	2.132	2.776	3.747	4.604	8.610	4
5	.132	.727	.920	1.476	2.015	2.571	3.365	4.032	6.869	5
6	.131	.718	.906	1.440	1.943	2.447	3.143	3.707	5.959	6
7	.130	.711	.896	1.415	1.895	2.365	2.998	3.499	5.408	7
8	.130	.706	.889	1.397	1.860	2.306	2.896	3.355	5.041	8
9	.129	.703	.883	1.383	1.833	2.262	2.821	3.250	4.781	9
10	.129	.700	.879	1.372	1.812	2.228	2.764	3.169	4.587	10
11	.129	.697	.876	1.363	1.796	2.201	2.718	3.106	4.437	11
12	.128	.695	.873	1.356	1.782	2.179	2.681	3.055	4.318	12
13	.128	.694	.870	1.350	1.771	2.160	2.650	3.012	4.221	13
14	.128	.692	.868	1.345	1.761	2.145	2.624	2.977	4.140	14
15	.128	.691	.866	1.341	1.753	2.131	2.602	2.947	4.073	15
16	.128	.690	.865	1.337	1.746	2.120	2.583	2.921	4.015	16
17	.128	.689	.863	1.333	1.740	2.110	2.567	2.898	3.965	17
18	.127	.688	.862	1.330	1.734	2.101	2.552	2.878	3.922	18
19	.127	.688	.861	1.328	1.729	2.093	2.539	2.861	3.883	19
20	.127	.687	.860	1.325	1.725	2.086	2.528	2.845	3.850	20
21	.127	.686	.859	1.323	1.721	2.080	2.518	2.831	3.819	21
22	.127	.686	.858	1.321	1.717	2.074	2.508	2.819	3.792	22
23	.127	.685	.858	1.319	1.714	2.069	2.500	2.807	3.767	23
24	.127	.685	.857	1.318	1.711	2.064	2.492	2.797	3.745	24
25	.127	.684	.856	1.316	1.708	2.060	2.485	2.787	3.725	25
26	.127	.684	.856	1.315	1.706	2.056	2.479	2.779	3.707	26
27	.127	.684	.855	1.314	1.703	2.052	2.473	2.771	3.690	27
28	.127	.683	.855	1.313	1.701	2.048	2.467	2.763	3.674	28
29	.127	.683	.854	1.311	1.699	2.045	2.462	2.756	3.659	29
30	.127	.683	.854	1.310	1.697	2.042	2.457	2.750	3.646	30
40	.126	.681	.851	1.303	1.684	2.021	2.423	2.704	3.551	40
60	.126	.679	.848	1.296	1.671	2.000	2.390	2.660	3.460	60
120	.126	.677	.845	1.289	1.658	1.980	2.358	2.617	3.373	120
∞	.126	.674	.842	1.282	1.645	1.960	2.326	2.576	3.291	∞

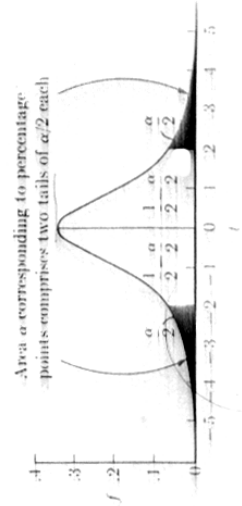


TABLE 12 Critical values of Student's t -distribution

The table furnishes critical values for Student's t -distribution for degrees of freedom $\nu = 1$ to 30 in increments of one, and for $\nu = 40, 60, 120$, and ∞ . The percentage points given (corresponding to $\alpha = 0.9, 0.5, 0.4, 0.2, 0.1, 0.05, 0.02, 0.01$, and 0.001) represent the area beyond the critical values of $\pm t$ in both tails of the distribution, as shown in the accompanying figure. The critical values of t are given to three decimal places.

To look up the critical values of t for a given number of degrees of freedom, look up ν in the left (argument) column of the table and read off the desired values of t in that row. For example, for 22 degrees of freedom, $t_{0.05} = 2.074$ and $t_{0.01} = 2.819$. The last value indicates that 1% of the area of the t -distribution (with 22 degrees of freedom) is beyond $t = \pm 2.819$, with 0.5% in each tail. If a one-tailed test is desired, the probabilities at the head of the table must be halved. Thus, for a one-tailed test with 4 df, the critical value $t = 3.747$ delimits 0.01 of the area of the curve. For degrees of freedom $\nu > 30$, we need to interpolate between the values of the argument ν . The table is designed for harmonic interpolation. Thus, to obtain $t_{0.01433}$ interpolate between $t_{0.01400} = 2.021$ and $t_{0.01600} = 2.000$, which are furnished in the table. Transform the arguments into $120/\nu = 120/43 = 2.791$ and interpolate between $120/60 = 2.000$ and $120/40 = 3.000$ by ordinary linear interpolation:

$$t_{0.01433} = (0.791 \times 2.021) + [(1 - 0.791) \times 2.000] \\ = 2.017$$

When $\nu > 120$, interpolate between $120/\infty = 0$ and $120/120 = 1$. If critical values other than those furnished in this table are desired, E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1 (Cambridge University Press, 1958) quote J. B. Simaika (*Biometrika*, 32:263-276, 1942), who recommends linear interpolation using the logarithms of the two-tailed probability values.

There are numerous applications of the t -distribution in statistics. The distribution is described in Section 7.4. Among the applications is the setting of confidence limits to means of small samples (Section 7.5); tests of difference between two means (Section 9.4); and comparisons of a single specimen with a sample (Section 9.5).

Values in this table have been taken from a more extensive one (table III) in R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed. (Oliver & Boyd, Edinburgh, 1958) with permission of the authors and their publishers.

Tabela 3: Pontos da distribuição de F
(Centeno, 1982: 114-115)

TABELA XII-10

Pontos 5% da distribuição de F $= v_2$

GL (2) do denominador	$= v_1$ GL (1) do numerador										
	1	2	3	4	5	7	9	15	20	50	∞
1	161	200	216	225	230	237	241	246	248	2	254
2	18,51	19,00	19,16	19,25	19,30	19,36	19,38	19,43	19,44	19,47	19,50
3	10,13	9,55	9,28	9,12	9,01	8,88	8,81	8,70	8,66	8,58	8,53
4	7,71	6,94	6,59	6,39	6,27	6,09	6,00	5,86	5,80	5,70	5,63
5	6,61	5,79	5,41	5,19	5,05	4,88	4,78	4,62	4,56	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4,21	4,10	3,94	3,87	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,79	3,68	3,51	3,44	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	3,50	3,39	3,28	3,15	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3,29	3,18	3,01	2,93	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	3,14	3,02	2,84	2,77	2,64	2,54
12	4,75	3,88	3,49	3,26	3,11	2,92	2,80	2,62	2,54	2,40	2,30
15	4,54	3,68	3,29	3,06	2,90	2,70	2,59	2,40	2,33	2,18	2,07
20	4,35	3,49	3,10	2,87	2,71	2,52	2,40	2,20	2,12	1,96	1,84
30	4,17	3,32	2,92	2,69	2,53	2,34	2,21	2,01	1,93	1,76	1,62
60	4,00	3,15	2,76	2,52	2,37	2,17	2,04	1,84	1,75	1,56	1,39
∞	3,48	2,99	2,60	2,37	2,21	2,01	1,88	1,67	1,57	1,35	1,00

TABELA XII-11

Pontos 1% da distribuição de F $= v_2$

GL (2) do denominador.	$= v_1$ GL (1) do numerador										
	1	2	3	4	5	7	9	15	20	50	∞
1	4052	4999	5403	5625	5764	5928	6022	6125	6208	6302	6366
2	94,49	99,00	99,17	99,25	99,30	99,34	99,38	99,43	99,45	99,48	99,50
3	34,12	30,82	29,46	28,71	28,24	27,67	27,34	26,87	26,69	26,35	26,12
4	21,20	18,00	16,79	15,98	15,52	14,98	14,76	14,20	14,02	13,69	13,46
5	16,26	13,27	16,06	11,39	10,97	10,45	10,15	9,72	9,55	9,24	9,02
6	13,74	10,92	9,78	9,15	8,75	8,26	7,98	7,56	7,39	7,09	7,88
7	12,25	9,55	8,45	7,85	7,46	7,00	6,71	6,31	6,15	5,85	5,65
8	11,26	8,65	7,59	7,01	6,53	6,19	5,91	5,51	5,36	5,06	4,86
9	10,56	8,02	6,99	6,42	6,06	5,72	5,35	4,96	4,80	4,51	4,31
10	10,04	7,56	6,55	5,99	5,64	5,21	4,95	4,56	4,41	4,12	3,91
12	9,33	6,93	5,95	5,41	5,06	4,65	4,39	4,01	3,86	3,56	3,36
15	8,68	6,36	5,42	4,89	4,56	4,14	3,89	3,52	3,36	3,07	2,87
20	8,10	5,85	4,94	4,43	4,10	3,71	3,45	3,09	2,94	2,63	2,42
30	7,56	5,39	4,51	4,02	3,70	3,30	3,06	2,70	2,55	2,24	2,01
60	7,08	4,98	4,13	3,65	3,34	2,95	2,72	2,35	2,20	1,87	1,60
∞	6,64	4,60	3,78	3,32	3,02	2,64	2,41	2,04	1,87	1,52	1,00

Tabela 4: "Critical values of the studentized maximum modules distribution"
(ROHLF & SOKAL, 1981:147-153)

k	α	Degrees of freedom ν										∞	k	α
		24	30	40	60	120	24	30	40	60	120			
3	.10	2.231	2.207	2.183	2.160	2.137	2.558	2.522	2.488	2.454	2.420	2.114	3	.10
	.05	2.558	2.522	2.488	2.454	2.420	3.253	3.185	3.119	3.055	2.993	2.388		.05
	.01	3.253	3.185	3.119	3.055	2.993	4.020	3.906	3.796	3.691	3.590	2.984		.01
4	.10	2.534	2.502	2.470	2.439	2.408	2.534	2.502	2.470	2.439	2.408	2.278	4	.10
	.05	2.851	2.805	2.760	2.716	2.673	3.531	3.447	3.367	3.289	3.215	2.631		.05
	.01	3.531	3.447	3.367	3.289	3.215	4.225	4.098	3.975	3.857	3.743	3.634		.01
5	.10	2.747	2.709	2.671	2.633	2.596	2.747	2.709	2.671	2.633	2.596	2.560	5	.10
	.05	3.059	3.005	2.952	2.900	2.849	3.732	3.637	3.545	3.456	3.371	2.800		.05
	.01	3.732	3.637	3.545	3.456	3.371	4.425	4.225	4.098	3.975	3.857	3.743		.01
6	.10	2.911	2.868	2.825	2.782	2.739	2.911	2.868	2.825	2.782	2.739	2.697	6	.10
	.05	3.220	3.160	3.100	3.041	2.984	3.220	3.160	3.100	3.041	2.984	2.928		.05
	.01	3.890	3.785	3.683	3.586	3.492	4.599	4.499	4.399	4.299	4.199	4.099		.01
7	.10	3.044	2.996	2.948	2.901	2.854	3.044	2.996	2.948	2.901	2.854	2.807	7	.10
	.05	3.352	3.285	3.220	3.156	3.093	3.352	3.285	3.220	3.156	3.093	3.031		.05
	.01	4.020	3.906	3.796	3.691	3.590	4.732	4.620	4.509	4.399	4.289	4.179		.01
8	.10	3.156	3.104	3.052	3.001	2.949	3.156	3.104	3.052	3.001	2.949	2.898	8	.10
	.05	3.462	3.391	3.321	3.251	3.183	3.462	3.391	3.321	3.251	3.183	3.117		.05
	.01	4.130	4.009	3.892	3.780	3.672	4.831	4.710	4.593	4.476	4.359	4.242		.01
9	.10	3.251	3.196	3.141	3.086	3.031	3.251	3.196	3.141	3.086	3.031	2.976	9	.10
	.05	3.557	3.482	3.407	3.334	3.261	3.557	3.482	3.407	3.334	3.261	3.190		.05
	.01	4.225	4.098	3.975	3.857	3.743	4.925	4.800	4.677	4.554	4.431	4.308		.01
10	.10	3.335	3.277	3.219	3.160	3.102	3.335	3.277	3.219	3.160	3.102	3.043	10	.10
	.05	3.641	3.562	3.483	3.406	3.329	3.641	3.562	3.483	3.406	3.329	3.254		.05
	.01	4.309	4.176	4.048	3.924	3.805	5.009	4.880	4.753	4.626	4.503	4.380		.01
11	.10	3.409	3.348	3.287	3.226	3.165	3.409	3.348	3.287	3.226	3.165	3.103	11	.10
	.05	3.716	3.633	3.551	3.470	3.389	3.716	3.633	3.551	3.470	3.389	3.310		.05
	.01	4.385	4.247	4.113	3.984	3.860	5.089	4.960	4.833	4.706	4.583	4.460		.01
12	.10	3.476	3.413	3.349	3.286	3.221	3.476	3.413	3.349	3.286	3.221	3.157	12	.10
	.05	3.783	3.697	3.611	3.527	3.443	3.783	3.697	3.611	3.527	3.443	3.361		.05
	.01	4.453	4.310	4.172	4.038	3.910	5.169	5.036	4.903	4.770	4.643	4.516		.01
13	.10	3.536	3.471	3.405	3.339	3.272	3.536	3.471	3.405	3.339	3.272	3.205	13	.10
	.05	3.843	3.755	3.667	3.579	3.492	3.843	3.755	3.667	3.579	3.492	3.407		.05
	.01	4.515	4.368	4.225	4.088	3.955	5.249	5.116	4.983	4.850	4.717	4.584		.01
14	.10	3.591	3.524	3.456	3.388	3.319	3.591	3.524	3.456	3.388	3.319	3.249	14	.10
	.05	3.899	3.808	3.717	3.627	3.537	3.899	3.808	3.717	3.627	3.537	3.449		.05
	.01	4.572	4.421	4.275	4.133	3.997	5.289	5.156	5.023	4.890	4.757	4.624		.01
15	.10	3.642	3.573	3.504	3.433	3.362	3.642	3.573	3.504	3.433	3.362	3.290	15	.10
	.05	3.951	3.857	3.764	3.674	3.578	3.951	3.857	3.764	3.674	3.578	3.487		.05
	.01	4.625	4.470	4.320	4.175	4.035	5.329	5.196	5.063	4.930	4.797	4.664		.01
16	.10	3.689	3.619	3.547	3.475	3.402	3.689	3.619	3.547	3.475	3.402	3.327	16	.10
	.05	3.998	3.902	3.807	3.712	3.617	3.998	3.902	3.807	3.712	3.617	3.523		.05
	.01	4.674	4.516	4.362	4.214	4.070	5.369	5.236	5.103	4.970	4.837	4.704		.01
17	.10	3.733	3.661	3.588	3.514	3.438	3.733	3.661	3.588	3.514	3.438	3.362	17	.10
	.05	4.043	3.945	3.847	3.750	3.652	4.043	3.945	3.847	3.750	3.652	3.556		.05
	.01	4.720	4.559	4.402	4.250	4.103	5.409	5.276	5.143	5.010	4.877	4.744		.01
18	.10	3.774	3.700	3.626	3.550	3.473	3.774	3.700	3.626	3.550	3.473	3.394	18	.10
	.05	4.084	3.985	3.885	3.785	3.685	4.084	3.985	3.885	3.785	3.685	3.587		.05
	.01	4.763	4.599	4.439	4.284	4.134	5.449	5.316	5.183	5.050	4.917	4.784		.01
19	.10	3.812	3.737	3.661	3.584	3.505	3.812	3.737	3.661	3.584	3.505	3.425	19	.10
	.05	4.123	4.022	3.920	3.818	3.717	4.123	4.022	3.920	3.818	3.717	3.615		.05
	.01	4.804	4.636	4.473	4.316	4.163	5.489	5.356	5.223	5.090	4.957	4.824		.01
20	.10	3.848	3.772	3.695	3.616	3.536	3.848	3.772	3.695	3.616	3.536	3.453	20	.10
	.05	4.160	4.057	3.953	3.850	3.746	4.160	4.057	3.953	3.850	3.746	3.643		.05
	.01	4.842	4.672	4.506	4.346	4.191	5.529	5.396	5.263	5.130	4.997	4.864		.01

TABLE 21 Critical values of the studentized maximum modulus distribution

k	α	Degrees of freedom ν										k	α		
		5	7	10	12	16	20								
3	.10	2.769	2.555	2.410	2.357	2.293	2.255	15	.10	4.985	4.479	4.119	3.983	3.813	3.710
	.05	3.399	3.055	2.829	2.747	2.650	2.594		.05	5.942	5.160	4.627	4.429	4.187	4.045
	.01	5.106	4.296	3.801	3.631	3.434	3.323		.01	8.602	6.902	5.841	5.469	5.032	4.785
4	.10	3.239	2.961	2.771	2.701	2.616	2.567	16	.10	5.028	4.542	4.176	4.037	3.863	3.759
	.05	3.928	3.489	3.199	3.095	2.969	2.897		.05	6.057	5.232	4.688	4.487	4.240	4.095
	.01	5.812	4.814	4.205	3.995	3.753	3.617		.01	8.722	6.994	5.914	5.535	5.090	4.837
5	.10	3.576	3.253	3.029	2.946	2.845	2.786	17	.10	5.125	4.601	4.229	4.087	3.911	3.804
	.05	4.312	3.805	3.467	3.345	3.199	3.114		.05	6.106	5.298	4.746	4.540	4.289	4.141
	.01	6.334	5.198	4.503	4.263	3.986	3.831		.01	8.833	7.079	5.982	5.597	5.144	4.886
6	.10	3.837	3.478	3.229	3.136	3.022	2.956	18	.10	5.188	4.657	4.279	4.135	3.955	3.846
	.05	4.610	4.051	3.677	3.540	3.377	3.282		.05	6.179	5.360	4.799	4.591	4.335	4.184
	.01	6.744	5.502	4.739	4.475	4.170	3.999		.01	8.937	7.158	6.045	5.654	5.194	4.932
7	.10	4.048	3.661	3.391	3.290	3.166	3.093	19	.10	5.247	4.708	4.325	4.179	3.996	3.886
	.05	4.853	4.252	3.848	3.700	3.522	3.419		.05	6.248	5.418	4.849	4.638	4.379	4.225
	.01	7.079	5.752	4.933	4.650	4.322	4.137		.01	9.034	7.233	6.105	5.708	5.242	4.975
8	.10	4.224	3.814	3.527	3.419	3.286	3.208	20	.10	5.302	4.757	4.369	4.220	4.035	3.923
	.05	5.057	4.421	3.992	3.835	3.645	3.534		.05	6.313	5.472	4.897	4.682	4.419	4.264
	.01	7.362	5.963	5.099	4.799	4.451	4.255		.01	9.126	7.303	6.161	5.760	5.286	5.016
9	.10	4.375	3.945	3.644	3.530	3.390	3.306								
	.05	5.232	4.566	4.116	3.951	3.751	3.634								
	.01	7.605	6.146	5.242	4.929	4.563	4.357								
10	.10	4.506	4.060	3.746	3.627	3.480	3.393								
	.05	5.384	4.693	4.225	4.053	3.844	3.721								
	.01	7.819	6.307	5.369	5.043	4.662	4.447								
11	.10	4.623	4.162	3.836	3.713	3.560	3.469								
	.05	5.520	4.806	4.322	4.143	3.926	3.799								
	.01	8.008	6.450	5.482	5.145	4.750	4.528								
12	.10	4.727	4.252	3.917	3.790	3.633	3.538								
	.05	5.641	4.907	4.409	4.225	4.001	3.869								
	.01	8.178	6.579	5.584	5.237	4.830	4.601								
13	.10	4.821	4.335	3.991	3.860	3.698	3.601								
	.05	5.750	4.999	4.488	4.299	4.068	3.933								
	.01	8.332	6.696	5.677	5.321	4.903	4.667								
14	.10	4.906	4.410	4.058	3.924	3.758	3.658								
	.05	5.850	5.083	4.560	4.367	4.130	3.991								
	.01	8.473	6.803	5.762	5.398	4.970	4.728								
15	.10	4.985	4.479	4.119	3.983	3.813	3.710								
	.05	5.942	5.160	4.627	4.429	4.187	4.045								
	.01	8.602	6.902	5.841	5.469	5.032	4.785								
16	.10	5.028	4.542	4.176	4.037	3.863	3.759								
	.05	6.057	5.232	4.688	4.487	4.240	4.095								
	.01	8.722	6.994	5.914	5.535	5.090	4.837								
17	.10	5.125	4.601	4.229	4.087	3.911	3.804								
	.05	6.106	5.298	4.746	4.540	4.289	4.141								
	.01	8.833	7.079	5.982	5.597	5.144	4.886								
18	.10	5.188	4.657	4.279	4.135	3.955	3.846								
	.05	6.179	5.360	4.799	4.591	4.335	4.184								
	.01	8.937	7.158	6.045	5.654	5.194	4.932								
19	.10	5.247	4.708	4.325	4.179	3.996	3.886								
	.05	6.248	5.418	4.849	4.638	4.379	4.225								
	.01	9.034	7.233	6.105	5.708	5.242	4.975								
20	.10	5.302	4.757	4.369	4.220	4.035	3.923								
	.05	6.313	5.472	4.897	4.682	4.419	4.264								
	.01	9.126	7.303	6.161	5.760	5.286	5.016								

TABLE 21 Critical values of the studentized maximum modulus distribution

			Degrees of freedom ν												
k	α		20	22	24	26	28	30	32	34	36	38			
20	.10	3.923	3.882	3.848	3.819	3.794	3.772	3.753	3.736	3.721	3.707	3.707	40	.10	4.402
	.05	4.264	4.207	4.160	4.120	4.086	4.057	4.031	4.008	3.988	3.970	3.970		.05	4.760
	.01	5.016	4.921	4.842	4.776	4.720	4.672	4.630	4.593	4.561	4.532	4.532		.01	5.551
22	.10	3.992	3.950	3.914	3.884	3.859	3.836	3.816	3.799	3.783	3.769	3.769	42	.10	4.434
	.05	4.334	4.276	4.228	4.187	4.152	4.121	4.095	4.071	4.050	4.031	4.031		.05	4.793
	.01	5.092	4.993	4.912	4.845	4.787	4.737	4.694	4.657	4.623	4.594	4.594		.01	5.587
24	.10	4.054	4.010	3.974	3.943	3.917	3.894	3.874	3.855	3.839	3.825	3.825	44	.10	4.464
	.05	4.398	4.339	4.289	4.247	4.211	4.180	4.152	4.128	4.107	4.087	4.087		.05	4.825
	.01	5.160	5.059	4.976	4.907	4.848	4.797	4.753	4.714	4.680	4.649	4.649		.01	5.622
26	.10	4.110	4.066	4.029	3.997	3.970	3.946	3.926	3.907	3.891	3.876	3.876	46	.10	4.493
	.05	4.456	4.395	4.345	4.302	4.265	4.233	4.205	4.180	4.158	4.138	4.138		.05	4.855
	.01	5.222	5.119	5.035	4.964	4.903	4.851	4.806	4.766	4.731	4.700	4.700		.01	5.655
28	.10	4.161	4.116	4.079	4.046	4.019	3.995	3.973	3.954	3.938	3.922	3.922	48	.10	4.521
	.05	4.509	4.447	4.396	4.352	4.314	4.282	4.253	4.228	4.205	4.185	4.185		.05	4.884
	.01	5.280	5.175	5.088	5.016	4.954	4.901	4.855	4.815	4.779	4.747	4.747		.01	5.686
30	.10	4.209	4.163	4.124	4.092	4.064	4.039	4.017	3.998	3.981	3.965	3.965	50	.10	4.547
	.05	4.559	4.496	4.443	4.398	4.360	4.327	4.298	4.272	4.249	4.228	4.228		.05	4.911
	.01	5.333	5.226	5.138	5.064	5.002	4.948	4.901	4.859	4.823	4.790	4.790		.01	5.716
32	.10	4.253	4.206	4.167	4.134	4.105	4.080	4.058	4.039	4.021	4.005	4.005	60	.10	4.663
	.05	4.604	4.540	4.487	4.441	4.402	4.369	4.339	4.313	4.289	4.268	4.268		.05	5.033
	.01	5.382	5.274	5.184	5.110	5.046	4.991	4.943	4.901	4.864	4.831	4.831		.01	5.849
34	.10	4.294	4.246	4.207	4.173	4.144	4.119	4.096	4.076	4.058	4.042	4.042	70	.10	4.759
	.05	4.647	4.582	4.528	4.482	4.442	4.408	4.377	4.351	4.327	4.306	4.306		.05	5.133
	.01	5.428	5.318	5.228	5.152	5.087	5.031	4.982	4.940	4.902	4.868	4.868		.01	5.960
36	.10	4.332	4.284	4.244	4.210	4.180	4.154	4.132	4.112	4.094	4.077	4.077	80	.10	4.841
	.05	4.687	4.621	4.566	4.519	4.479	4.444	4.414	4.386	4.362	4.341	4.341		.05	5.219
	.01	5.471	5.360	5.268	5.191	5.125	5.069	5.020	4.976	4.938	4.904	4.904		.01	6.055
38	.10	4.368	4.320	4.279	4.244	4.214	4.188	4.165	4.145	4.126	4.110	4.110	100	.10	4.976
	.05	4.724	4.658	4.602	4.555	4.514	4.479	4.448	4.420	4.396	4.374	4.374		.05	5.361
	.01	5.512	5.400	5.307	5.228	5.162	5.104	5.054	5.011	4.972	4.937	4.937		.01	6.211

TABLE 21 Critical values of the studentized maximum modulus distribution

			Degrees of freedom ν												
k	α		20	22	24	26	28	30	32	34	36	38			
40	.10	4.402	4.353	4.312	4.277	4.247	4.220	4.220	4.197	4.176	4.157	4.141	40	.10	4.402
	.05	4.760	4.692	4.636	4.588	4.547	4.511	4.480	4.452	4.427	4.405	4.405		.05	4.760
	.01	5.551	5.437	5.343	5.264	5.196	5.138	5.087	5.043	5.004	4.968	4.968		.01	5.551
42	.10	4.434	4.385	4.343	4.308	4.277	4.250	4.227	4.227	4.206	4.187	4.170	42	.10	4.434
	.05	4.793	4.725	4.668	4.620	4.578	4.542	4.510	4.482	4.457	4.434	4.434		.05	4.793
	.01	5.587	5.472	5.377	5.297	5.229	5.170	5.119	5.074	5.034	4.998	4.998		.01	5.587
44	.10	4.464	4.415	4.373	4.337	4.306	4.279	4.255	4.234	4.215	4.198	4.198	44	.10	4.464
	.05	4.825	4.756	4.698	4.650	4.607	4.571	4.539	4.510	4.485	4.462	4.462		.05	4.825
	.01	5.622	5.505	5.409	5.329	5.260	5.200	5.149	5.103	5.063	5.027	5.027		.01	5.622
46	.10	4.493	4.443	4.401	4.364	4.333	4.306	4.282	4.260	4.241	4.224	4.224	46	.10	4.493
	.05	4.855	4.786	4.727	4.678	4.636	4.599	4.566	4.537	4.512	4.489	4.489		.05	4.855
	.01	5.655	5.537	5.440	5.359	5.289	5.229	5.177	5.131	5.090	5.054	5.054		.01	5.655
48	.10	4.521	4.470	4.427	4.391	4.359	4.332	4.307	4.286	4.266	4.249	4.249	48	.10	4.521
	.05	4.884	4.814	4.755	4.705	4.662	4.625	4.592	4.563	4.537	4.514	4.514		.05	4.884
	.01	5.686	5.568	5.470	5.387	5.317	5.257	5.204	5.157	5.116	5.080	5.080		.01	5.686
50	.10	4.547	4.496	4.453	4.416	4.384	4.356	4.332	4.310	4.290	4.273	4.273	50	.10	4.547
	.05	4.911	4.841	4.781	4.731	4.688	4.650	4.617	4.588	4.562	4.538	4.538		.05	4.911
	.01	5.716	5.597	5.498	5.415	5.344	5.283	5.230	5.183	5.141	5.104	5.104		.01	5.716
60	.10	4.663	4.610	4.565	4.527	4.494	4.465	4.440	4.417	4.397	4.378	4.378	60	.10	4.663
	.05	5.033	4.959	4.898	4.846	4.801	4.762	4.727	4.697	4.670	4.645	4.645		.05	5.033
	.01	5.849	5.725	5.623	5.537	5.463	5.400	5.344	5.296	5.252	5.214	5.214		.01	5.849
70	.10	4.759	4.705	4.659	4.619	4.585	4.556	4.529	4.506	4.485	4.466	4.466	70	.10	4.759
	.05	5.133	5.058	4.995	4.941	4.895	4.854	4.819	4.787	4.759	4.734	4.734		.05	5.133
	.01	5.960	5.833	5.727	5.639	5.563	5.497	5.440	5.390	5.345	5.305	5.305		.01	5.960
80	.10	4.841	4.786	4.738	4.698	4.663	4.633	4.606	4.582	4.560	4.541	4.541	80	.10	4.841
	.05	5.219	5.142	5.078	5.023	4.975	4.934	4.897	4.865	4.836	4.810	4.810		.05	5.219
	.01	6.055	5.925	5.817	5.726	5.648	5.581	5.522	5.470	5.424	5.383	5.383		.01	6.055
100	.10	4.976	4.918	4.869	4.827	4.791	4.759	4.731	4.706	4.684	4.664	4.664	100	.10	4.976
	.05	5.361	5.281	5.214	5.157	5.107	5.064	5.026	4.993	4.962	4.935	4.935		.05	5.361
	.01	6.211	6.076	5.964	5.870	5.789	5.719	5.658	5.604	5.556	5.513	5.513		.01	6.211

TABLE 21 Critical values of the studentized maximum modulus distribution

		Degrees of freedom ν											
k	α	40	50	60	80	100	120	240	480	∞			
20	.10	3.695	3.648	3.616	3.576	3.552	3.535	3.494	3.474	3.453			
	.05	3.953	3.891	3.850	3.798	3.767	3.746	3.694	3.668	3.643			
	.01	4.506	4.409	4.346	4.267	4.221	4.191	4.116	4.079	4.042			
22	.10	3.756	3.708	3.675	3.633	3.608	3.591	3.549	3.527	3.505			
	.05	4.015	3.950	3.907	3.854	3.821	3.800	3.746	3.719	3.693			
	.01	4.567	4.467	4.401	4.320	4.273	4.241	4.164	4.126	4.088			
24	.10	3.812	3.762	3.728	3.685	3.659	3.642	3.597	3.575	3.552			
	.05	4.070	4.004	3.960	3.904	3.871	3.849	3.793	3.765	3.738			
	.01	4.622	4.519	4.452	4.368	4.319	4.287	4.207	4.168	4.129			
26	.10	3.862	3.811	3.776	3.732	3.705	3.687	3.642	3.618	3.595			
	.05	4.120	4.053	4.007	3.950	3.916	3.893	3.836	3.807	3.778			
	.01	4.672	4.567	4.497	4.412	4.362	4.328	4.247	4.206	4.167			
28	.10	3.909	3.856	3.820	3.775	3.748	3.729	3.682	3.658	3.634			
	.05	4.167	4.097	4.051	3.992	3.957	3.934	3.875	3.845	3.816			
	.01	4.718	4.611	4.540	4.452	4.401	4.367	4.283	4.242	4.201			
30	.10	3.951	3.898	3.861	3.815	3.787	3.768	3.720	3.695	3.670			
	.05	4.210	4.139	4.091	4.031	3.995	3.971	3.911	3.881	3.850			
	.01	4.761	4.651	4.579	4.490	4.437	4.402	4.316	4.274	4.233			
32	.10	3.991	3.936	3.899	3.852	3.823	3.804	3.754	3.729	3.704			
	.05	4.249	4.177	4.128	4.067	4.031	4.006	3.944	3.913	3.882			
	.01	4.801	4.689	4.615	4.524	4.470	4.435	4.347	4.304	4.262			
34	.10	4.028	3.972	3.934	3.886	3.857	3.837	3.786	3.761	3.735			
	.05	4.286	4.213	4.163	4.101	4.064	4.038	3.975	3.944	3.912			
	.01	4.838	4.724	4.649	4.557	4.502	4.466	4.376	4.333	4.289			
36	.10	4.063	4.006	3.967	3.918	3.888	3.868	3.817	3.790	3.764			
	.05	4.321	4.246	4.196	4.133	4.095	4.069	4.005	3.972	3.940			
	.01	4.873	4.757	4.681	4.587	4.531	4.494	4.404	4.359	4.315			
38	.10	4.095	4.038	3.999	3.949	3.918	3.898	3.845	3.818	3.791			
	.05	4.354	4.278	4.227	4.163	4.124	4.098	4.032	3.999	3.966			
	.01	4.906	4.788	4.711	4.616	4.559	4.521	4.429	4.384	4.339			

Degrees of freedom ν

k	α	40	50	60	80	100	120	240	480	∞
40	.10	4.126	4.067	4.028	3.977	3.946	3.925	3.872	3.845	3.817
	.05	4.385	4.308	4.256	4.191	4.151	4.125	4.058	4.025	3.991
	.01	4.937	4.818	4.739	4.643	4.585	4.547	4.453	4.407	4.362
42	.10	4.155	4.096	4.056	4.004	3.973	3.951	3.897	3.869	3.841
	.05	4.414	4.336	4.283	4.217	4.177	4.150	4.082	4.048	4.014
	.01	4.966	4.846	4.766	4.668	4.610	4.571	4.476	4.430	4.384
44	.10	4.182	4.122	4.082	4.030	3.998	3.976	3.921	3.893	3.864
	.05	4.441	4.363	4.309	4.242	4.202	4.174	4.106	4.071	4.037
	.01	4.994	4.872	4.792	4.692	4.633	4.594	4.498	4.451	4.404
46	.10	4.208	4.148	4.107	4.054	4.022	4.000	3.944	3.915	3.886
	.05	4.468	4.388	4.334	4.266	4.225	4.197	4.128	4.093	4.058
	.01	5.021	4.898	4.816	4.716	4.656	4.616	4.519	4.471	4.424
48	.10	4.233	4.172	4.131	4.077	4.044	4.022	3.966	3.937	3.907
	.05	4.493	4.412	4.358	4.289	4.247	4.219	4.149	4.113	4.078
	.01	5.047	4.922	4.840	4.738	4.677	4.637	4.538	4.490	4.442
50	.10	4.257	4.195	4.153	4.099	4.066	4.044	3.986	3.957	3.927
	.05	4.517	4.435	4.381	4.311	4.269	4.240	4.169	4.133	4.097
	.01	5.071	4.945	4.862	4.759	4.698	4.657	4.557	4.508	4.460
60	.10	4.362	4.297	4.254	4.197	4.162	4.139	4.078	4.046	4.015
	.05	4.623	4.538	4.480	4.407	4.363	4.333	4.258	4.220	4.181
	.01	5.179	5.047	4.960	4.852	4.788	4.746	4.641	4.589	4.538
70	.10	4.449	4.382	4.337	4.278	4.242	4.217	4.154	4.121	4.087
	.05	4.711	4.623	4.564	4.488	4.441	4.410	4.331	4.291	4.251
	.01	5.269	5.133	5.043	4.931	4.864	4.819	4.710	4.656	4.603
80	.10	4.523	4.455	4.408	4.347	4.310	4.284	4.218	4.184	4.149
	.05	4.787	4.696	4.635	4.556	4.509	4.476	4.394	4.353	4.311
	.01	5.347	5.206	5.113	4.998	4.929	4.883	4.769	4.714	4.659
100	.10	4.645	4.574	4.525	4.461	4.422	4.395	4.325	4.288	4.251
	.05	4.911	4.816	4.752	4.670	4.619	4.585	4.498	4.454	4.409
	.01	5.474	5.328	5.230	5.108	5.036	4.987	4.868	4.809	4.750

Tabela 5: “Critical values of the studentized augmented range”
(ROHLF & SOKAL, 1981:145-146)

TABLE 20 Critical values of the studentized augmented range

This table furnishes critical values for the studentized augmented range $Q'_{\alpha[k,r]}$ employed in the Spjøtvoll-Stoline T'-method of multiple comparisons. Values are tabulated for $k = 2$ to 8 means in increments of one and for $r = 5, 7, 10, 12, 16, 20, 24, 30, 40, 60, 120$, and ∞ degrees of freedom. Three experiment-wise error rates $\alpha = 0.10, 0.05$, and 0.01 are given for each combination of k and r . Harmonic interpolation is recommended for degrees of freedom that are not furnished.

To find the 5% critical value of the studentized augmented range over $k = 5$ means at 20 degrees of freedom, enter the table for $k = 5, r = 20$, and $\alpha = .05$ to obtain $Q'_{.05[5,20]} = 4.233$. For $k > 8$, the studentized range furnished in Table 18 may be used since for $k > 8$ these two distributions are very similar.

This table is used for the Spjøtvoll-Stoline T'-test described in Section 9.7. It can also be employed in multiple comparisons of regression coefficients illustrated in Section 14.8. This table has been copied and rearranged from the tables furnished by M. R. Stoline (*J. Amer. Stat. Assoc.*, 73:656-660, 1978) with permission of the publisher.

TABLE 20 Critical values of the studentized augmented range

r	α	k						
		2	3	4	5	6	7	8
5	.10	3.060	3.772	4.282	4.671	4.982	5.239	5.458
	.05	3.832	4.654	5.236	5.680	6.036	6.331	6.583
	.01	5.903	7.030	7.823	8.429	8.916	9.322	9.669
7	.10	2.848	3.491	3.943	4.285	4.556	4.781	4.972
	.05	3.486	4.198	4.692	5.064	5.360	5.606	5.816
	.01	5.063	5.947	6.551	7.008	7.374	7.679	7.939
10	.10	2.704	3.300	3.712	4.021	4.265	4.466	4.636
	.05	3.259	3.899	4.333	4.656	4.913	5.124	5.305
	.01	4.550	5.284	5.773	6.138	6.428	6.669	6.875
12	.10	2.651	3.230	3.628	3.924	4.157	4.349	4.511
	.05	3.177	3.791	4.204	4.509	4.751	4.950	5.119
	.01	4.373	5.056	5.505	5.837	6.101	6.321	6.507
16	.10	2.588	3.146	3.526	3.806	4.027	4.207	4.360
	.05	3.080	3.663	4.050	4.334	4.557	4.741	4.897
	.01	4.169	4.792	5.194	5.489	5.722	5.915	6.079
20	.10	2.551	3.097	3.466	3.738	3.950	4.124	4.271
	.05	3.024	3.590	3.961	4.233	4.446	4.620	4.768
	.01	4.055	4.644	5.019	5.294	5.510	5.688	5.839
24	.10	2.527	3.065	3.427	3.693	3.901	4.071	4.213
	.05	2.988	3.542	3.904	4.167	4.373	4.541	4.684
	.01	3.982	4.549	4.908	5.169	5.374	5.542	5.685
30	.10	2.503	3.034	3.389	3.649	3.851	4.016	4.155
	.05	2.952	3.496	3.847	4.103	4.302	4.464	4.602
	.01	3.912	4.458	4.800	5.048	5.242	5.401	5.536
40	.10	2.480	3.003	3.352	3.605	3.803	3.963	4.099
	.05	2.918	3.450	3.792	4.040	4.232	4.389	4.521
	.01	3.844	4.370	4.696	4.931	5.115	5.265	5.392
60	.10	2.457	2.972	3.315	3.563	3.755	3.911	4.042
	.05	2.884	3.406	3.738	3.978	4.163	4.314	4.441
	.01	3.778	4.284	4.595	4.818	4.991	5.133	5.253
120	.10	2.434	2.943	3.278	3.520	3.707	3.859	3.987
	.05	2.851	3.362	3.686	3.917	4.096	4.241	4.363
	.01	3.714	4.201	4.497	4.709	4.872	5.005	5.118
∞	.10	2.412	2.913	3.243	3.479	3.661	3.808	3.931
	.05	2.819	3.320	3.634	3.858	4.030	4.170	4.286
	.01	3.653	4.121	4.403	4.603	4.757	4.882	4.987

Tabela 6: "Critical values of the studentized augmented range"
(ROHLF & SOKAL, 1981:133-137)

TABLE 18 Critical values of the studentized range $\alpha = 0.05$

$\nu \backslash k$	20	22	24	26	28	30	32	34	36
1	59.56	60.91	62.12	63.22	64.23	65.15	66.01	66.81	67.56
2	16.77	17.13	17.45	17.75	18.02	18.27	18.50	18.72	18.92
3	11.24	11.47	11.68	11.87	12.05	12.21	12.36	12.50	12.63
4	9.233	9.418	9.584	9.736	9.875	10.00	10.12	10.23	10.34
5	8.208	8.368	8.512	8.643	8.764	8.875	8.979	9.075	9.165
6	7.587	7.730	7.861	7.979	8.088	8.189	8.283	8.370	8.452
7	7.170	7.303	7.423	7.533	7.634	7.728	7.814	7.895	7.972
8	6.870	6.995	7.109	7.212	7.307	7.395	7.477	7.554	7.625
9	6.644	6.763	6.871	6.970	7.061	7.145	7.222	7.295	7.363
10	6.467	6.582	6.686	6.781	6.868	6.948	7.023	7.093	7.159
11	6.326	6.436	6.536	6.628	6.712	6.790	6.863	6.930	6.994
12	6.209	6.317	6.414	6.503	6.585	6.660	6.731	6.796	6.858
13	6.112	6.217	6.312	6.398	6.478	6.551	6.620	6.684	6.744
14	6.029	6.132	6.224	6.309	6.387	6.459	6.526	6.588	6.647
15	5.958	6.059	6.149	6.233	6.309	6.379	6.445	6.506	6.564
16	5.897	5.995	6.084	6.166	6.241	6.310	6.374	6.434	6.491
17	5.842	5.940	6.027	6.107	6.181	6.249	6.313	6.372	6.427
18	5.794	5.890	5.977	6.055	6.128	6.195	6.258	6.316	6.371
19	5.752	5.846	5.932	6.009	6.081	6.147	6.209	6.267	6.321
20	5.714	5.807	5.891	5.968	6.039	6.104	6.165	6.222	6.275
24	5.594	5.683	5.764	5.838	5.906	5.968	6.027	6.081	6.132
30	5.475	5.561	5.638	5.709	5.774	5.833	5.889	5.941	5.990
40	5.358	5.439	5.513	5.581	5.642	5.700	5.753	5.803	5.849
60	5.241	5.319	5.389	5.453	5.512	5.566	5.617	5.664	5.708
120	5.126	5.200	5.266	5.327	5.382	5.434	5.481	5.526	5.568
∞	5.012	5.081	5.144	5.201	5.253	5.301	5.346	5.388	5.427

$\nu \backslash k$	38	40	50	60	70	80	90	100
1	68.26	68.92	71.73	73.97	75.82	77.40	78.77	79.98
2	19.11	19.28	20.05	20.66	21.16	21.59	21.96	22.29
3	12.75	12.87	13.36	13.76	14.08	14.36	14.61	14.82
4	10.44	10.53	10.93	11.24	11.51	11.73	11.92	12.09
5	9.250	9.330	9.674	9.949	10.18	10.38	10.54	10.69
6	8.529	8.601	8.913	9.163	9.370	9.548	9.702	9.839
7	8.043	8.110	8.400	8.632	8.824	8.989	9.133	9.261
8	7.693	7.756	8.029	8.248	8.430	8.586	8.722	8.843
9	7.428	7.488	7.749	7.958	8.132	8.281	8.410	8.526
10	7.220	7.279	7.529	7.730	7.897	8.041	8.166	8.276
11	7.053	7.110	7.352	7.546	7.708	7.847	7.968	8.075
12	6.916	6.970	7.205	7.394	7.552	7.687	7.804	7.909
13	6.800	6.854	7.083	7.267	7.421	7.552	7.667	7.769
14	6.702	6.754	6.979	7.159	7.309	7.438	7.550	7.650
15	6.618	6.669	6.888	7.065	7.212	7.339	7.449	7.546
16	6.544	6.594	6.810	6.984	7.128	7.252	7.360	7.457
17	6.479	6.529	6.741	6.912	7.054	7.176	7.283	7.377
18	6.422	6.471	6.680	6.848	6.989	7.109	7.213	7.307
19	6.371	6.419	6.626	6.792	6.930	7.048	7.152	7.244
20	6.325	6.373	6.576	6.740	6.877	6.994	7.097	7.187
24	6.181	6.226	6.421	6.579	6.710	6.822	6.920	7.008
30	6.037	6.080	6.267	6.415	6.543	6.650	6.744	6.827
40	5.893	5.934	6.112	6.255	6.375	6.477	6.566	6.645
60	5.750	5.789	5.958	6.093	6.206	6.303	6.387	6.462
120	5.607	5.644	5.802	5.929	6.035	6.126	6.205	6.275
∞	5.463	5.498	5.646	5.764	5.863	5.947	6.020	6.085

TABLE 18 Critical values of the studentized range $\alpha = 0.05$

$\nu \backslash k$	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.085	8.331	9.798	10.88	11.75	12.44	13.03	13.54	13.99
3	4.501	5.910	6.825	7.502	8.037	8.478	8.853	9.177	9.462
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.802	6.995
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.487
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940	5.077	5.198
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897	5.031	5.150
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824	4.956	5.071
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924	5.038
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.896	5.008
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915
30	2.888	3.486	3.845	4.102	4.302	4.464	4.602	4.720	4.824
40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635	4.735
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560
∞	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474

$\nu \backslash k$	11	12	13	14	15	16	17	18	19
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57
3	9.717	9.946	10.15	10.35	10.53	10.69	10.84	10.98	11.11
4	8.027	8.208	8.373	8.525	8.664	8.794	8.914	9.028	9.134
5	7.168	7.324	7.466	7.596	7.717	7.828	7.932	8.030	8.122
6	6.649	6.789	6.917	7.034	7.143	7.244	7.338	7.426	7.508
7	6.302	6.431	6.550	6.658	6.759	6.852	6.939	7.020	7.097
8	6.034	6.175	6.287	6.389	6.483	6.571	6.653	6.729	6.802
9	5.867	5.983	6.089	6.186	6.276	6.359	6.437	6.510	6.579
10	5.722	5.833	5.935	6.028	6.114	6.194	6.269	6.339	6.405
11	5.605	5.713	5.811	5.901	5.984	6.062	6.134	6.202	6.265
12	5.511	5.615	5.710	5.798	5.878	5.953	6.023	6.089	6.151
13	5.431	5.533	5.625	5.711	5.789	5.862	5.931	5.995	6.055
14	5.364	5.463	5.554	5.637	5.714	5.786	5.852	5.915	5.974
15	5.306	5.404	5.493	5.574	5.649	5.720	5.785	5.846	5.904
16	5.256	5.352	5.439	5.520	5.593	5.662	5.727	5.786	5.843
17	5.212	5.307	5.392	5.471	5.544	5.612	5.675	5.734	5.790
18	5.174	5.267	5.352	5.429	5.501	5.568	5.630	5.688	5.743
19	5.140	5.231	5.315	5.391	5.462	5.528	5.589	5.647	5.701
20	5.108	5.199	5.282	5.357	5.427	5.493	5.553	5.610	5.663
24	5.012	5.099	5.179	5.251	5.319	5.381	5.439	5.494	5.545
30	4.917	5.001	5.077	5.147	5.211	5.271	5.327	5.379	5.429
40	4.824	4.904	4.977	5.044	5.106	5.163	5.216	5.266	5.313
60	4.732	4.808	4.878	4.942	5.001	5.056	5.107	5.154	5.199
120	4.641	4.714	4.781	4.842	4.898	4.950	4.998	5.044	5.086
∞	4.552	4.622	4.685	4.743	4.796	4.845	4.891	4.934	4.974

TABLE 18 Critical values of the studentized range $\alpha = 0.01$

$p \backslash k$	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.88	31.69
3	8.261	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.512	8.120	9.173	9.958	10.58	11.10	11.55	11.93	12.27
5	5.702	6.976	7.804	8.421	8.913	9.321	9.669	9.972	10.24
6	5.243	6.331	7.033	7.556	7.973	8.318	8.613	8.869	9.097
7	4.949	5.919	6.543	7.005	7.373	7.679	7.939	8.166	8.368
8	4.746	5.635	6.204	6.625	6.960	7.237	7.474	7.681	7.863
9	4.596	5.428	5.957	6.348	6.658	6.915	7.134	7.325	7.495
10	4.482	5.270	5.769	6.136	6.428	6.669	6.875	7.055	7.213
11	4.392	5.146	5.621	5.970	6.247	6.476	6.672	6.842	6.992
12	4.320	5.046	5.502	5.836	6.101	6.321	6.507	6.670	6.814
13	4.260	4.964	5.404	5.727	5.981	6.192	6.372	6.528	6.667
14	4.210	4.895	5.322	5.634	5.881	6.085	6.258	6.409	6.543
15	4.168	4.836	5.252	5.556	5.796	5.994	6.162	6.309	6.439
16	4.131	4.786	5.192	5.489	5.722	5.915	6.079	6.222	6.349
17	4.099	4.742	5.140	5.430	5.659	5.847	6.007	6.147	6.270
18	4.071	4.703	5.094	5.379	5.603	5.788	5.944	6.081	6.201
19	4.046	4.670	5.054	5.334	5.554	5.735	5.889	6.022	6.141
20	4.024	4.639	5.018	5.294	5.510	5.688	5.839	5.970	6.087
24	3.956	4.546	4.907	5.168	5.374	5.542	5.685	5.809	5.919
30	3.889	4.455	4.799	5.048	5.242	5.401	5.536	5.653	5.756
40	3.825	4.367	4.696	4.931	5.114	5.265	5.392	5.502	5.599
60	3.762	4.282	4.595	4.818	4.991	5.133	5.253	5.356	5.447
120	3.702	4.200	4.497	4.709	4.872	5.005	5.118	5.214	5.299
∞	3.643	4.120	4.403	4.603	4.757	4.882	4.987	5.078	5.157

TABLE 18 Critical values of the studentized range $\alpha = 0.01$

$p \backslash k$	20	22	24	26	28	30	32	34	36
1	298.0	304.7	310.8	316.3	321.3	326.0	330.3	334.3	338.0
2	37.95	38.76	39.49	40.15	40.76	41.32	41.84	42.33	42.78
3	19.77	20.17	20.53	20.86	21.16	21.44	21.70	21.95	22.17
4	14.40	14.68	14.93	15.16	15.37	15.57	15.75	15.92	16.08
5	11.93	12.16	12.36	12.54	12.71	12.87	13.02	13.15	13.28
6	10.54	10.73	10.91	11.06	11.21	11.34	11.47	11.58	11.69
7	9.646	9.815	9.970	10.11	10.24	10.36	10.47	10.58	10.67
8	9.027	9.182	9.322	9.450	9.569	9.678	9.779	9.874	9.964
9	8.573	8.717	8.847	8.966	9.075	9.177	9.271	9.360	9.443
10	8.226	8.361	8.483	8.595	8.698	8.794	8.883	8.966	9.044
11	7.952	8.080	8.196	8.303	8.400	8.491	8.575	8.654	8.728
12	7.731	7.853	7.964	8.066	8.159	8.246	8.327	8.402	8.473
13	7.548	7.665	7.772	7.870	7.960	8.043	8.121	8.193	8.262
14	7.395	7.508	7.611	7.705	7.792	7.873	7.948	8.018	8.084
15	7.264	7.374	7.474	7.566	7.650	7.728	7.800	7.869	7.932
16	7.152	7.258	7.356	7.445	7.527	7.602	7.673	7.739	7.802
17	7.053	7.158	7.253	7.340	7.420	7.493	7.563	7.627	7.687
18	6.968	7.070	7.163	7.247	7.325	7.398	7.465	7.528	7.587
19	6.891	6.992	7.082	7.166	7.242	7.313	7.379	7.440	7.498
20	6.823	6.922	7.011	7.092	7.168	7.237	7.302	7.362	7.419
24	6.612	6.705	6.789	6.865	6.936	7.001	7.062	7.119	7.173
30	6.407	6.494	6.572	6.644	6.710	6.772	6.828	6.881	6.932
40	6.209	6.289	6.362	6.429	6.490	6.547	6.600	6.650	6.697
60	6.015	6.090	6.158	6.220	6.277	6.330	6.378	6.424	6.467
120	5.827	5.897	5.959	6.016	6.069	6.117	6.162	6.204	6.244
∞	5.645	5.709	5.766	5.818	5.866	5.911	5.952	5.990	6.026

TABLE 18 Critical values of the studentized range $\alpha = 0.01$

$p \backslash k$	38	40	50	60	70	80	90	100
1	341.5	344.8	338.9	370.1	379.4	387.3	394.1	400.1
2	43.21	43.61	45.33	46.73	47.83	48.80	49.64	50.38
3	22.39	22.59	23.45	24.13	24.71	25.19	25.62	25.99
4	16.23	16.37	16.98	17.46	17.86	18.20	18.50	18.77
5	13.40	13.52	14.00	14.39	14.72	14.99	15.23	15.45
6	11.80	11.90	12.31	12.65	12.92	13.16	13.37	13.55
7	10.77	10.85	11.23	11.52	11.77	11.99	12.17	12.34
8	10.05	10.13	10.47	10.75	10.97	11.17	11.34	11.49
9	9.521	9.594	9.912	10.17	10.38	10.57	10.73	10.87
10	9.117	9.187	9.486	9.726	9.927	10.10	10.25	10.39
11	8.798	8.864	9.148	9.377	9.568	9.732	9.875	10.00
12	8.539	8.603	8.875	9.094	9.277	9.434	9.571	9.693
13	8.326	8.387	8.648	8.859	9.035	9.187	9.318	9.436
14	8.146	8.204	8.457	8.661	8.832	8.978	9.106	9.219
15	7.992	8.049	8.295	8.492	8.658	8.800	8.924	9.035
16	7.860	7.916	8.154	8.347	8.507	8.646	8.767	8.874
17	7.745	7.799	8.031	8.219	8.377	8.511	8.630	8.735
18	7.643	7.696	7.924	8.107	8.261	8.393	8.508	8.611
19	7.553	7.605	7.828	8.008	8.159	8.288	8.401	8.502
20	7.473	7.523	7.742	7.919	8.067	8.194	8.305	8.404
24	7.223	7.270	7.476	7.642	7.780	7.900	8.004	8.097
30	6.978	7.023	7.215	7.370	7.500	7.611	7.709	7.796
40	6.740	6.782	6.960	7.104	7.225	7.328	7.419	7.500
60	6.507	6.546	6.710	6.843	6.954	7.050	7.133	7.207
120	6.281	6.316	6.467	6.588	6.689	6.776	6.852	6.919
∞	6.060	6.092	6.228	6.338	6.429	6.507	6.575	6.636

Tabela 7: "Critical values for correlation coefficients"
(ROHLF & SOKAL, 1981:166-168)

TABLE 25 Critical values for correlation coefficients

<i>k</i> Number of independent variables						<i>k</i> Number of independent variables					
<i>v</i>	α	1	2	3	4	<i>v</i>	α	1	2	3	4
1	.05	.997	.999	.999	.999	24	.05	.388	.470	.523	.562
	.01	1.000	1.000	1.000	1.000		.01	.496	.565	.609	.642
2	.05	.950	.975	.983	.987	25	.05	.381	.462	.514	.553
	.01	.990	.995	.997	.998		.01	.487	.555	.600	.633
3	.05	.878	.930	.950	.961	26	.05	.374	.454	.506	.545
	.01	.959	.976	.983	.987		.01	.478	.546	.590	.624
4	.05	.811	.881	.912	.930	27	.05	.367	.446	.498	.536
	.01	.917	.949	.962	.970		.01	.470	.538	.582	.615
5	.05	.754	.836	.874	.898	28	.05	.361	.439	.490	.529
	.01	.874	.917	.937	.949		.01	.463	.530	.573	.606
6	.05	.707	.795	.839	.867	29	.05	.355	.432	.482	.521
	.01	.834	.886	.911	.927		.01	.456	.522	.565	.598
7	.05	.666	.758	.807	.838	30	.05	.349	.426	.476	.514
	.01	.798	.855	.885	.904		.01	.449	.514	.558	.591
8	.05	.632	.726	.777	.811	35	.05	.325	.397	.445	.482
	.01	.765	.827	.860	.882		.01	.418	.481	.523	.556
9	.05	.602	.697	.750	.786	40	.05	.304	.373	.419	.455
	.01	.735	.800	.836	.861		.01	.393	.454	.494	.526
10	.05	.576	.671	.726	.763	45	.05	.288	.353	.397	.432
	.01	.708	.776	.814	.840		.01	.372	.430	.470	.501
11	.05	.553	.648	.703	.741	50	.05	.273	.336	.379	.412
	.01	.684	.753	.793	.821		.01	.354	.410	.449	.479
12	.05	.532	.627	.683	.722	60	.05	.250	.308	.348	.380
	.01	.661	.732	.773	.802		.01	.325	.377	.414	.442
13	.05	.514	.608	.664	.703	70	.05	.232	.286	.324	.354
	.01	.641	.712	.755	.785		.01	.302	.351	.386	.413
14	.05	.497	.590	.646	.686	80	.05	.217	.269	.304	.332
	.01	.623	.694	.737	.768		.01	.283	.330	.362	.389
15	.05	.482	.574	.630	.670	90	.05	.205	.254	.288	.315
	.01	.606	.677	.721	.752		.01	.267	.312	.343	.368
16	.05	.468	.559	.615	.655	100	.05	.195	.241	.274	.300
	.01	.590	.662	.706	.738		.01	.254	.297	.327	.351
17	.05	.456	.545	.601	.641	125	.05	.174	.216	.246	.269
	.01	.575	.647	.691	.724		.01	.228	.266	.294	.316
18	.05	.444	.532	.587	.628	150	.05	.159	.198	.225	.247
	.01	.561	.633	.678	.710		.01	.208	.244	.270	.290
19	.05	.433	.520	.575	.615	200	.05	.138	.172	.196	.215
	.01	.549	.620	.665	.698		.01	.181	.212	.234	.253
20	.05	.423	.509	.563	.604	300	.05	.113	.141	.160	.176
	.01	.537	.608	.652	.685		.01	.148	.174	.192	.208
21	.05	.413	.498	.552	.592	400	.05	.098	.122	.139	.153
	.01	.526	.596	.641	.674		.01	.128	.151	.167	.180
22	.05	.404	.488	.542	.582	500	.05	.088	.109	.124	.137
	.01	.515	.585	.630	.663		.01	.115	.135	.150	.162
23	.05	.396	.479	.532	.572	1,000	.05	.062	.077	.088	.097
	.01	.505	.574	.619	.652		.01	.081	.096	.106	.115