

CELSO CURSINO

**RESTRICÇÕES DA CORRELAÇÃO NOS TESTES DE
GERMINAÇÃO DE SEMENTES E EMERGÊNCIA DE
PLÂNTULAS**

Dissertação apresentada à Universidade Federal de Uberlândia, como parte das exigências do Programa de Pós-graduação em Agronomia – Mestrado, área de concentração em Fitotecnia, para obtenção do título de “Mestre”.

Orientadora:

Prof^ª. Dra. Denise Garcia de Santana

Co-orientadora:

Prof^ª. Dra. Marli Ranal

UBERLÂNDIA
MINAS GERAIS – BRASIL
2007

Dados Internacionais de Catalogação na Publicação (CIP)

C977r Cursino, Celso, 1955-

Restrições da correlação nos testes de germinação de sementes e emergência de plântulas / Celso Cursino. - 2007.

51 f. : il.

Orientadora: Denise Garcia de Santana.

Co-orientadora: Marli Ranal.

Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Agronomia.

Inclui bibliografia.

1. Estatística agrícola - Teses. 2. Sementes - Teses. 3. Germinação - Teses. I. Santana, Denise Garcia de. II. Ranal, Marli. III. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Agronomia. III. Título.

CDU: 311.1:631

CELSO CURSINO

**RESTRICÇÕES DA CORRELAÇÃO NOS TESTES DE GERMINAÇÃO DE
SEMENTES E EMERGÊNCIA DE PLÂNTULAS**

Dissertação apresentada à Universidade Federal de Uberlândia, como parte das exigências do Programa de Pós-graduação em Agronomia – Mestrado, área de concentração em Fitotecnia, para obtenção do título de “Mestre”.

APROVADA em 27 de dezembro de 2006

Prof. Dr. Ednaldo Guimarães

UFU

Prof^a Dra. Maria Eugênia Lisei de Sá

EMATER

Prof^a. Dra. Marli Ranal

UFU

Prof^a. Dra. Denise Garcia de Santana
ICIAG-UFU
(Orientadora)

UBERLÂNDIA
MINAS GERAIS – BRASIL
2007

À minha esposa, às minhas filhas e aos meus netos, pelo
farto suprimento de carinho;
Ao meu filho Bruno Leonardo, *in memoriam*, por ter me
sustentado com entusiasmo,
enquanto eu *só estudava*;
Ao meu pai, *in memoriam*, por eu não ter conseguido
herdar ou cultivar suas virtudes de humildade e
temperança, mas pela riqueza que me coube
como legado da ânsia de perscrutar,
pela geometria, o significado
mais intrínseco das leis
e fenômenos exatos.

AGRADECIMENTOS

À professora Ms. Maria Tereza de Queiroz Piacentini, pelas pérolas da língua portuguesa de seus livros de ortografia e pela revisão deste trabalho, sou devedor.

Às minhas orientadoras professoras Dra. Denise Garcia de Santana e Dra. Marli Ranal, ilhas de dedicação e denodo nesta carreira do magistério, tão cheia de circunstâncias adversas, como eu poderia ser suficientemente grato? A elas coube receber este Engenheiro que ci-vil “estranho no ninho” da Agronomia, dando-me o aconchego da paciência e do desapego, iniciando-me na inebriante arte da criação científica. Estendo a todo o Instituto de Ciências Agrárias o meu reconhecimento do esforço extra para tornar factíveis as possibilidades do ambiente multidisciplinar.

À minha irmã, professora Dra. Sídnei Cursino Guimarães, meu carinho por me fazer acreditar que ainda há dignidade em ser mestre, em uma docência apaixonada, livre de toda mágoa.

SUMÁRIO

RESUMO	<i>i</i>
ABSTRACT	<i>ii</i>
1 INTRODUÇÃO	<i>1</i>
2 REVISÃO BIBLIOGRÁFICA	<i>4</i>
3 MATERIAL E MÉTODOS	<i>10</i>
3.1 Testes de germinação de sementes por envelhecimento acelerado.....	<i>10</i>
3.2 Teste de velocidade de emergência em campo	<i>11</i>
3.3 Efeito dos tratamentos na medida de correlação	<i>12</i>
3.4 Simulação de conjunto de dados para situações específicas	<i>13</i>
3.5 Variável auxiliar Z.....	<i>13</i>
4 RESULTADOS E DISCUSSÃO	<i>16</i>
4.1 Testes de germinação por envelhecimento acelerado.....	<i>16</i>
4.2 Testes de velocidade de emergência em campo.....	<i>19</i>
4.3 Entre variáveis de testes diferentes entre si.....	<i>24</i>
4.4 Efeitos dos tratamento.....	<i>28</i>
4.5 Dados simulados	<i>34</i>
4.6 Dispersões curvilíneas	<i>41</i>
4.7 Conclusão	<i>44</i>
5 REFERÊNCIAS	<i>45</i>
6 ANEXO	<i>49</i>

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

DIC	Delineamento inteiramente casualizado
DBC	Delineamento em blocos casualizados
ICIAG	Instituto de Ciências Agrárias (UFU)
$H_o : \rho = 0$	Teste de Hipóteses – Hipótese nula definida com $\rho = 0$
$\mu\text{mol m}^{-2} \text{s}^{-1}$	micromol por metro quadrado por segundo
r	Coefficiente de Correlação de Pearson – Coeficiente do produto-momento de Pearson
R^2	Coefficiente de determinação
V.Emer 5ºdia	Médias do teste de velocidade de emergência medidas em plântulas emersas por dia, leitura no quinto dia. Por analogia, V.Emer 6ºdia, a leitura do sexto dia, e assim sucessivamente.
36°C24h	Variável de germinação em teste de germinabilidade por Envelhecimento acelerado realizado em estufa a com temperatura de 36°C e tempo de permanência 24 horas (por analogia, seguem-se 38°C48h para 38°C e tempo de permanência 48 horas, e assim por diante, seqüencialmente)

RESUMO

CURSINO, CELSO. **Restrições da correlação nos testes de germinação de sementes e emergência de plântulas.** 2006. 61p. Dissertação (Mestrado em Agronomia/Fitotecnia) – Universidade Federal de Uberlândia. Uberlândia.¹

Coeficiente de Pearson r é usado para comparar experimentos científicos. Em tecnologias de sementes serve para comparar resultados de procedimentos que medem vigor. Quando se prognosticam resultados de correlações baseados em condições similares e eles não acontecem, a correlação de Pearson enfrenta críticas, atribuídas principalmente a duas causas. Primeiramente pela estatística, para cuja utilização da correlação de Pearson existem prescrições nem sempre observadas, talvez por não serem entendidas como pressuposições. Exigem-se variáveis métricas naturalmente associadas, com distribuição normal bivariada, pareamento, homoscedasticidade, nuvem de dispersão retilínea; detecção de *outliers*. Somam-se observações práticas quanto à validade restrita a um trecho da série de dados, da criação de faixas de valores para considerá-la de “baixa” a “alta”, da necessidade da análise gráfica, da interpretação de significância, entre outras. A segunda causa seria justamente a existência de variação biológica devido a fatores diversos externos e interno às sementes, servindo às vezes de sustentáculo para conclusões de interesse do pesquisador. No objetivo de identificar aplicabilidade das correlações e as causas de resultados estranhos, foram comparados dados existentes no Laboratório de Análises de Sementes do ICIAG da Universidade Federal de Uberlândia-MG, testes germinação de envelhecimento acelerado em condições ideais de repetibilidade em laboratório, e teste de emergência de plântulas em campo, e outras variáveis simuladas, havendo incidência de resultados estranhos. A representação gráfica normal da dispersão entre X e Y mostra satisfatoriamente o correlacionamento de variáveis naturalmente associadas com n grande. Entretanto, se a covariância não é tão óbvia, a dispersão $Y=f(X)$ não é suficiente para mostrar crescimento ou decréscimo simultâneo entre as variáveis. Usando metodologia alternativa de plotagem das variáveis em relação a uma variável auxiliar Z , de mesmos n elementos que X e Y , pôde-se estudar individualmente o comportamento das variáveis. O método gráfico permitiu taxar correlações em válidas ou não pela similaridade das variáveis, comparável à homoscedasticidade; verificar *outliers* em n pequeno ou grande; agrupamento de *outliers* em “trecho dissidente” e mostrar efeito de tratamentos. Nos casos analisados, concluiu-se que, comparando vigor de sementes com resultados só de laboratório, tão bem como no seu relacionamento com os de campo; e entre dados simulados, as inconsistências de resultados de correlações são preponderantes por não seguirem as prescrições da literatura, entre outras. A magnitude das distorções por causas estatísticas não deixou espaço para mensurar efeitos da variação de condições biológicas de sementes, alterações temporais relativas a manuseio ou edafoclimáticas.

Palavras-chave: 1. Falhas em correlações – 2. Confiabilidade de correlações – 3. Prescrições da correlação – 4. Homoscedasticidade nas correlações – 5. Detectar *outliers* na correlação

¹ Comitê Orientador: Denise Garcia Santana– UFU (Orientadora) e Marli Ranal - UFU (Co-orientadora).

ABSTRACT

CURSINO, CELSO. **Restrictions of the correlation in the tests of seed germination and seedling emergence.** Uberlândia: UFU, 2007. 61p. Dissertation (Master Program Agronomy/Crop Science) Federal University of Uberlandia, Uberlândia.²

Coefficient of Pearson r is used to compare scientific tests. In seeds technology it is used to compare results of procedures that measure vigour. When the correspondingly similar results are not found in very similar conditions, Person's correlation faces criticism mainly due to two factors. The first one comes from statistics for whose usage of Person's correlation there are prescriptions that are not always observed, when they are not understood as assumption. Variables naturally associated are required with bivariated normal distribution, pairing; homoscedasticity, rectilinear dispersion; detection of outliers. Added to them, there are practical observations in what refers the correlation to be valid only in a restrict range of the data series, the necessity to create value ranges to consider this correlation as 'good' or 'bad', the need of the graphical analysis, the use and interpretation of the significance, among others. The second cause of odd results would be the existence of several biological factors, which are sometimes support for the reseracher conclusions. With the objective of identifying applicability of correlations and the causes for odd results of r , there have been compared data existent in the Seeds Analysis Laboratory of ICIAG of the Universidade Federal de Uberlândia-MG, as well as tests of germination of acelerated aging in optimal conditions of repetibility done in laboratory, and tests of field seedling emergency, as well as other simulated variables. The results showed odd results. The normal scattergram between X and Y is enough clear to elucidate only correlated variables of large samples. Although, if the covariance is not as obvious the dispersion $Y=f(X)$ is not enough to show simultaneous increasing or decreasing between variables. With an alternative methodology of plotting the variables related to another auxiliar variable Z of the same n elements of X and Y , we could study the variable behavior in an individual way. It was possible to create graphic criteria to assess non-valid correlations, such as similarity of variables comparable to homoscedastity; influence of outliers on small or big n ; grouping of outliers in a 'dissident range', influence of treatments effect. In the analysed cases, we concluded that, comparing seeds vigour with only laboratory results, as well as its relation with the field results and among simulated data, the results inconsistency of correlations are prevalent as they do not follow the literature prescriptions, among others. The magnitude of the distortions due to statistical causes did not leave space for measuring effects of the variation of the biological seeds conditions, temporal alterations related to management or the edafoclimatic ones.

Keywords: 1. Failure in correlations – 2. Correlation reliability – 3. Prescriptions of correlation use – 4. Homoscedasticity in correlations – 5. Outliers detection in the correlation.

² Advising Committee: Denise Garcia Santana - UFU (Advisor) – and Marli Ranal – UFU (Co-Advisor)

1 INTRODUÇÃO

Correlação de Pearson mede a associação linear entre duas variáveis pareadas, tomadas do mesmo indivíduo, amostra ou repetição. A correlação é uma medida estatística bastante utilizada por tecnologistas de sementes para comparações entre metodologias para testes de vigor, especialmente em trabalhos incluindo emergência em campo (EGLI; TEKRONI, 1977; JOHNSON; WAX, 1978; BISHNOI; DELOUCHE, 1980; DELOUCHE; BASKIN, 1973; GREGORY; HOPPER; JIVIDEN, 1986; ORDOVINI et al. 1988; FREITAS, 2000). Em algumas publicações, a medida é interpretada como coeficiente de determinação (R^2) do modelo de regressão e em outras, é aplicada incorretamente. As distorções são, em parte, justificadas pela similaridade algébrica entre os coeficientes de correlação e de determinação (SOKAL; ROHLF, 1995) e pela literatura, que não é clara ao distinguir as duas técnicas estatísticas, repassando a idéia de que o coeficiente de correlação é uma extensão do modelo de regressão. Para acrescentar, alguns livros discutem essas duas técnicas no mesmo capítulo.

A principal diferença entre as duas técnicas é a natureza das variáveis envolvidas. No modelo de regressão, a variável dependente (Y) e a independente (X) apresentam uma relação funcional ou de causa e efeito. Para a medida de correlação, não há necessidade de definição da natureza das variáveis, e, portanto, esta medida pode ser aplicada independentemente das variáveis terem ou não relação funcional ou de causa e efeito, ainda que sejam afetadas por causas comuns (SOKAL; ROHLF, 1995). Esta busca de relações de causa e efeito entre as variáveis com o uso da correlação, é também um dos equívocos de interpretação da medida e tema de discussão na literatura (SOKAL; ROHLF, 1995; SCHIELD, 1995). Por outro lado, não se exclui que a correlação entre duas variáveis pode ser um sinal de causalidade, mas não o suficiente para inferi-la (SCHELD, 1995).

Existem condições específicas para o uso de correlação, mas sua utilização como medida de correlacionamento entre duas variáveis sem verificação dessas condições é observada em trabalhos científicos. Em tecnologia de sementes, a primeira verificação é quanto ao pareamento das variáveis, que em muitas situações entre experimentos de campo e laboratório não é bem definido. As condições mais explícitas na literatura estatística em geral, apontam para a validade do coeficiente de correlação quando a

nuvem de dispersão ser aderente ao modelo de regressão linear e; assumindo que exista um correlacionamento natural entre as variáveis, prescrevem a verificação da influência de *outliers* (NETER; WASSERMAN; KUTNER, 1985). Outras condições menos explícitas relacionam a influência do tamanho da amostra no teste de significância da correlação (MEKO, 1981); a interpretação prática do próprio teste de significância e a necessidade de testar se as variáveis são homogêneas (MILLER, 1994). Davis (1971) sugere que sejam definidas faixas de valores de correlação para cada área de conhecimento. A validade da correlação é confinada a “trechos” da série de dados (MEKO, 1981).

Um outro problema de aplicação e interpretação da correlação ocorre nos modelos de delineamentos experimentais, onde as variáveis são afetadas pelos efeitos dos tratamentos, como nos modelos de delineamento inteiramente casualizados e dos efeitos de tratamentos e blocos, como nos delineamentos em blocos casualizados. Estes efeitos podem superestimar ou subestimar o valor da correlação, podendo-se atribuir associação a variáveis não correlacionadas, pelo menos não linearmente, ou ainda, não se atribuir associação a variáveis correlacionadas. Essas particularidades fazem com que a correlação, apesar de ser uma das medidas estatísticas mais utilizadas e descritas na literatura, deva ser reavaliada.

Se entendida a verificação do conjunto de condições e ressalvas do uso da correlação como uma prescrição, e tendo ele sido atendido, a correlação se aplica quando se supõe haver variação proporcional e linear entre as variáveis. Este é o conceito de covariância, ou seja, uma variável cresce ou decresce simultaneamente com o incremento ou decréscimo da outra (STEEL; TORRIE, 1980; NETER; WASSERMAN; KUTNER, 1985 e SOKAL; ROHLF, 1995). Neste caso o r mede a covariância sem efeito de escala. Em algumas circunstâncias, porém, mesmo verificando o atendimento às condições para aplicação, e observando que dois conjuntos de variáveis parecem manter covariância, persistem resultados de experimentos agrônômicos de germinação de sementes e emergência de plântulas com valores que se distanciam dos esperados obtidos em situações idênticas. Com isto, aumenta a lista de prescrições para o uso de correlação.

Uma vez que as explicações para os problemas em foco não são abordadas de forma a definir uma prescrição, e as condições encontradas não são suficientes para elucidar a aplicabilidade da medida de correlação para condições particulares, o presente estudo teve como objetivos, analisar situações em que o coeficiente de

correlação pode ou não ser indicado para associar o relacionamento entre variáveis, sobretudo na comparação de germinação de sementes em laboratório com emergência de plântulas em campo, a influência dos *outliers*, significância estatística e prática, além dos fatores de estabilidade.

2 REVISÃO BIBLIOGRÁFICA

A correlação é uma medida estatística bastante utilizada por tecnologistas de sementes para comparações entre metodologias dos testes de vigor, especialmente em trabalhos incluindo emergência em campo (EGLI; TEKRONI, 1977; JOHNSON; WAX, 1978; BISHNOI; DELOUCHE, 1980; GREGORY; HOPPER; JIVIDEN, 1986; ORDOVINI et al., 1988; Freitas et al., 2000). Uma vez que variáveis entre testes de campo e laboratório são por natureza independentes, a inexistência de relação funcional entre elas não é questionada e, portanto, os pesquisadores não aventam a possibilidade de usar modelos de regressão. Nesses experimentos, a escolha por correlação é imediata, mas a natureza independente das variáveis, apesar de necessária, não é condição suficiente para sua aplicação.

Toledo e Marcos Filho; Toledo (1977) sustentam a hipótese de alta correlação entre os resultados de germinação em laboratório e a emergência das plântulas em campo somente verificada quando as condições de campo são plenamente favoráveis. Essa afirmativa é verdadeira pela própria hipótese de que um lote com alta qualidade no teste de germinação, também apresentará esta alta qualidade nas condições de campo se este estiver com alto vigor, caso contrário, a correlação é baixa. Contudo, para a medida de correlação, não é restritivo que as variáveis a serem correlacionadas sejam afetadas por causas totalmente distintas, como as condições climáticas, por exemplo. Da mesma forma que não se exclui a possibilidade de ambas serem influenciadas por causas comuns, como citado por Miller (1994) e Davis; Samaniego; Watnik (1997).

O pareamento, ou associação entre pares de dados, é observado quando se comparam lotes diferentes mas as leituras se referem a condições idênticas, como são em uma mesma temperatura e dia de leitura em testes de envelhecimento acelerado. Mesmo com a possibilidade de falsos pareamentos, a medida é amplamente descrita na literatura de germinação de sementes, mas também é alvo de crítica de muitos pesquisadores, dentre eles, Marcos Filho et al. (1984). Segundo os autores, alguns resultados de correlação fogem do esperado e se tornam de difícil interpretação. Estas críticas podem ser devidas à falta de verificação das condições de uso da medida, à forma como são conduzidos os experimentos e à própria interpretação dos resultados.

Os preceitos básicos relativos ao cálculo do coeficiente de correlação de Pearson (1930) são que os dados não sejam colhidos a partir de classes ou intervalos e não estejam sujeitos a erros de medidas, aferição de instrumentos e padrão de comportamento na medição, sendo recomendável o mesmo avaliador ou medidor para todas as amostras. Para outros tipos de variáveis como as nominais dicotômicas (como macho/fêmea) ou ordinais existem outros coeficientes que não o de Pearson (1930), utilizados para medir essas correlações do tipo “não-paramétricas”, como o coeficiente de correlação de Kendal (GRASLAND, 2000; DEL SIEGLE, 2006).

Devido à necessidade de as variáveis serem aleatórias, Bewick; Cheek; Ball (2003) citam que a obtenção de dados distinguindo grupos, como mulheres e homens, resulta na formação de *clusters* de pontos na dispersão que aumentam o valor do coeficiente de correlação encontrado.

No coeficiente de correlação linear de Pearson (1930), o termo linear é proveniente da constatação de que a medida está restrita à relação linear de covariância, relacionada à reta. A literatura especifica a necessidade de as variáveis a serem correlacionadas serem aleatórias e independentes, ou seja, de que não deve existir uma associação funcional e de dependência entre elas. No entanto, autores como Neter; Wassermann; Kutner (1985) mostram a possibilidade de a reta de maior aderência ser uma indicação de que a nuvem de dispersão se distribua ao longo de uma direção reta, como se fosse o eixo da elipsóide que envolve esta nuvem. Meko (1981) soma a estas observações o fato de que correlações não podem ser aplicadas quando a nuvem de dispersão tende a modelos não-lineares.

Grasland (2000) complementa, citando a classificação de monotônica, quando uma distribuição linear tem tendência estritamente crescente ou decrescente, ou seja, não comporta pontos de inflexão de mínima ou de máxima. Para que o coeficiente de Pearson (1930) seja válido, uma vez que correlação só é válida no trecho linear, este autor afirma que toda relação suscetível ao coeficiente de Pearson (1930) precisa estar em um trecho de característica monotônica.

Sokal; Rohlf (1995) exemplificam com dados amostrais que a plotagem gráfica das variáveis nem sempre oferece a dispersão de contorno bem caracterizado. Estes autores mostram que o contorno vai assumindo claramente formato elíptico à medida que a quantidade de elementos (n) cresce. Sustentam que o formato tende a um círculo quando o correlacionamento entre as variáveis é nulo e se torna uma elipsóide à medida

que se aproxima dos valores -1 ou $+1$, inclinando-se para baixo ou para cima, respectivamente. Rummel (1976) já havia mostrado graficamente que a mera observação do formato da nuvem de dispersão tendendo a uma envoltória elíptica mais ou menos alongada em torno do maior eixo, permite avaliar, com relativa proximidade, os valores de r sem necessidade de efetuar cálculos.

Meko (1981) estudou séries temporais observando que a correlação se aplica e o resultado, para ser válido, precisa estar confinado a um trecho específico da série. Neste dado trecho, a nuvem de dispersão deveria ter o aspecto elíptico alongado em torno de uma reta, aspecto não curvilíneo.

Ainda se referindo à nuvem de dispersão, Miller (1994) e Del Siegle (2006) rescrevem a necessidade da variância ser homoscedástica para a validade do uso de correlações. Entretanto, não se conseguiu encontrar alguma indicação de procedimentos deles para esta verificação, a menos da observação gráfica sem também indicar o modo de constatar-la. Garson (1999) indica a necessidade de comparar os coeficientes de variação para comprovar homogeneidade das variáveis cidade, exigindo que sejam parecidos, sem especificar limites ou padrões de variação.

Em capítulo específico para o tema, Sokal; Rohlf (1995) mostram que a correlação poderia ser aplicada a qualquer conjunto de pares sem se importar com o significado implícito de sua distribuição, que poderia ser normal ou não, servindo meramente de um índice matemático de pouco interesse, sem função de estimador amostral estatístico. No entanto, sugerem que para a correlação ter significado prático, as amostras devem ser formadas por variáveis com distribuição normal bivariada. Assim, o coeficiente de correlação amostral r pode bem ser o estimador para o coeficiente de correlação populacional ρ . Constata-se atualmente que o estimador r é ligeiramente enviesado e subestima o coeficiente de correlação populacional quando a amostra (n) é pequena. Os autores sugerem uma correção proposta por Kendall (STUART, 1961; *in* SIEGEL, 2006).

A representação gráfica desta distribuição é uma plotagem tridimensional. Se a distribuição é normal bivariada, o aspecto da curva deverá ter o mesmo da distribuição normal monovariada quando cada uma das curvas for observada em vista de topo, ora assumindo o traçado $Z=f(x)$, ora $Z=f(y)$. Os autores mostram em desenho plotado em perspectiva isométrica, que se duas variáveis são assumidas como independentes (portanto, não indicadas para análise de regressão), mas sendo correlacionadas

positivamente em certo grau, então resultará que altos valores de Y serão acompanhados de altos valores de X e vice-versa. Isto somente pode ser visualizado graficamente quando a quantidade de pontos for maior que 100. A constatação de normalidade de cada amostra pode ser testada mas não significa que a normalidade bivariada seja atendida.

Não havendo distribuição gaussiana bivariada, caso muito freqüente, o coeficiente de correlação de Pearson (1930) deve redundar em conclusões errôneas (GRASLAND, 2000). O autor acrescenta que a ausência de relação linear que possa ser medida por correlação de Pearson (1930) não significa que nenhuma relação haja entre as características das duas variáveis estudadas, apenas não há correlação linear.

Pela prescrição de distribuição normal, não podem existir *outliers*. Sokal; Rohlf (1995), NeterWasserman; Kurtner (1985) afirmam que os valores de correlação são fortemente influenciados por eles, sem definir uma sistemática de identificá-los. Weldon (2000) acrescenta que uma das características do gráfico de dispersão dos resíduos padronizados da correlação é que ele permite identificar *outliers* em pares que fogem muito do contorno elíptico da nuvem de dispersão. No entanto, o contorno elíptico bem definido de variáveis correlacionadas não é fácil de ser obtido e claramente detectado a menos de amostras muito grandes ou quando a correlação paramétrica é alta (SOKAL; ROHLF, 1995).

Os problemas atribuídos à forma como são conduzidos os experimentos envolvem diversos fatores. Em tecnologia de sementes as atividades estão concentradas no desenvolvimento de novas variedades e híbridos de alta produtividade com sementes de vigor, no conceito de que sejam aptas a resistir a condições adversas, em ambientes os mais diversos, mantendo o máximo poder de germinação pelo maior período possível de armazenagem, capazes de estabelecer, após a sementeira, um estande de mínima perda e maior lucro (EGLI; TEKRONY; WIRALAGA, 1990; CARVALHO; VIEIRA, 1994). Sementes de variedades, híbridos e outros são arranjadas em lotes, de acordo principalmente com seu local de cultivo e época de colheita e devem apresentar potencialmente maior vigor. Labouriau (1983), Caldwell; Delouche, (1960) e Marcos Filho; Cícero; Da Silva, (1987) citam que é de se esperar que sementes de lotes com melhor desempenho de germinação em laboratório forneçam subsídios que permitam inferir um melhor desempenho de emergência de suas plântulas em campo. A definição de vigor referida tem caráter econômico de muita relevância, por ser base para

argumentação comercial entre o que é semente e o que será descartado para grão (AOSA, 1983; BRASIL, 1992). Andrade et al. (2001) concluíram que, quando o cultivo não sofre estresse e as condições de campo são favoráveis, deve haver equivalência da emergência no campo com a germinação em laboratório, também preconizados pelas RAS - Regras para a Análise de Sementes (BRASIL, 1992; MARCOS FILHO; TOLEDO, 1977; DELOUCHE; BASKIN, 1973).

O efeito de tratamento foi observado por Santana; Ranal, (2006) para experimentos inteiramente casualizados. Elas constatam que as correlações entre as amostras existem, mas parecem estar majoradas pelo efeito de tratamento e estes efeitos podem ser padronizados. Na mesma ótica das referidas autoras, pode-se pressupor que efeitos de tratamentos serão também encontrados quando os delineamentos experimentais mantêm a casualização sob controle do pesquisador. As amostras de sementes são normalmente distribuídas em arranjos de lotes visando análise de variância relacionada aos delineamentos experimentais. Nos experimentos que comparam desempenho de lotes de sementes com vigor, as variáveis estão agrupadas em função de lotes, que podem ser considerados um agrupamento tal como tratamento. Isto poderá ser observado, tanto em DIC, quanto em DBC.

A interpretação dos resultados costuma ser problemática, principalmente porque os pesquisadores tendem a exigir mais da correlação do que ela é capaz de medir. Começa com a apresentação de resultados de correlação em conjunto com medidas de determinação (R^2) afetas à análise de regressão. Estas deveriam ser utilizadas apenas quando existe dependência entre as variáveis. Seria, então, proposto um modelo em que o pesquisador deveria sugerir a partir da definição da natureza das variáveis envolvidas. Uma tem natureza de variável dependente (Y) e a outra independente (X). Elas apresentam uma relação funcional ou de causa e efeito. No caso de sementes, não se pode estabelecer que um lote que tenha maior ou menor vigor seja “explicado” pelo maior ou menor vigor de outro lote. Essa confusão é, em parte, justificada pela similaridade algébrica entre os coeficientes de correlação e de determinação (SOKAL; ROHLF, 1995) e pela literatura, que não é clara ao distinguir as duas técnicas estatísticas, repassando a idéia de que o coeficiente de correlação é uma extensão do modelo de regressão. Para agravar, alguns livros discutem essas duas técnicas no mesmo capítulo (SANTANA; RANAL, 2006).

Em revisão realizada por Santana; Ranal (2006) foi constatado que, depois de determinar se o valor da correlação é suficientemente alto para atribuir alguma associação entre as variáveis, o pesquisador depara-se com o problema da significância. A correlação amostral r é uma medida descritiva, tornando-se inferencial quando ao valor r é associado o valor da significância. Na maioria das vezes, o teste t de “Student” para a significância do coeficiente de correlação proposto por Pearson (1930) verifica apenas se a correlação entre as variáveis é nula ($H_o : \rho = 0$) e não define a associação no caso contrário. Acrescenta-se também, a tendência da correlação em apresentar valores significativos com o aumento do tamanho da amostra (COCHRAN; SNEDECOR, 1989; ZAR, 1999), mesmo para variáveis com valores baixos de correlação. Desta forma, o pesquisador deve atentar para a diferença entre significância estatística e prática (MILLER, 1994). Este mesmo autor relata resultado de sua pesquisa analisando trabalhos científicos publicados, que em sua maioria o autor informa significância, probabilidade associada à correlação, mas sequer informa qual é a hipótese de nulidade. Questiona se a intenção do autor é fazer inferências com a correlação medida com dados amostrais, verificando se teria ou não probabilidade de ser zero enquanto a populacional não for. Por outro lado, se não há teste de hipóteses, estes pesquisadores devem estar imaginando a correlação descritivamente. Davis (1971) ressalta que deveriam ser adotados predicados bem definidos para magnitudes de correlação definindo como “perfeita” a correlação $r = 1,0$, que valores entre $r = 0,70$ e $r = 0,99$ correlações “muito altas”, de $r = 0,50$ a $r = 0,60$ “substanciais”, de $r = 0,30$ a $r = 0,49$ “moderados”, de $r = 0,10$ a $r = 0,29$ “baixos” e, abaixo disso, “desprezíveis”. Por suas observações a definição de uma correlação “boa” é afeta à experiência de cada pesquisador e é específica de cada área de conhecimento. Para o autor, portanto, em cada uma destas áreas de conhecimento, para cada tipo de comparações entre amostras, deveria haver um padrão ou tabela padronizada de valores. Acrescenta que as faixas de valores serviriam, então, para atender ao propósito de balizar o pesquisador no propósito de assumir que determinados resultados encontrados possam ser considerados “bons”.

3 MATERIAL E MÉTODOS

As sementes de sorgo (*Sorghum bicolor*) utilizadas nos testes de germinabilidade por envelhecimento acelerado e de velocidade de emergência de plântulas, foram provenientes de diferentes safras agrícolas com qualidade fisiológica distintas. As amostras dos diferentes lotes foram fornecidas por firmas produtoras de sementes, contendo especificações quanto à cultivar ou híbrido comercialmente utilizado. As metodologias e os dados utilizados foram gentilmente cedidos por Santos (1985).

3.1 Testes de germinação de sementes por envelhecimento acelerado

No teste de envelhecimento acelerado, as sementes foram submetidas a condições de alta umidade e temperatura para que fosse manifestado nos lotes o vigor após o estresse. Com o estresse, a taxa de deterioração é aumentada e os lotes menos vigorosos tendem a diminuir ainda mais sua germinação quando comparados aos mais vigorosos.

O delineamento utilizado foi o inteiramente casualizado em esquema fatorial 8 x 3, sendo 8 lotes e 3 temperaturas (36°C, 38°C e 40°C), com quatro repetições de 50 sementes, totalizando 200 sementes por tratamento. Análises individuais foram executadas para cada tempo de permanência na câmara (24 h, 48 h, 72 h, 96 h e 120 h).

As sementes foram colocadas em caixas plásticas tipo gerbox contendo uma bandeja de tela de metal, onde foram distribuídas as sementes de maneira uniforme, preenchendo totalmente os espaços vazios da tela. No interior do gerbox foram adicionados 50 mL de água, formando uma câmara úmida. As caixas de germinação foram levadas à B.O.D com umidade relativa de 90% a 100% com temperatura e tempo de permanência estabelecidos de acordo com o tratamento.

Depois de serem determinados os tempos de permanência pelos tratamentos para o envelhecimento, quantidades suficientes de sementes para quatro sub-amostras em número suficiente de 200 sementes foram colocadas para germinar em germinadores. A leitura da germinação ocorreu no décimo dia, e o critério de classificação das plântulas e sementes foram os mesmos mencionados na caracterização do perfil da amostra para a germinação, sendo os dados expressos em porcentagem de plântulas normais.

As comparações de resultados foram feitas com quase todas as combinações possíveis, porém, devido ao padrão constante de comportamento, somente algumas correlações foram apresentadas neste trabalho. A seleção de alguns resultados, tanto dos testes de germinação de sementes quanto de velocidade de emergência foi relativamente aleatória, mantendo-se sempre que possível o lote 8 de menores resultados como referência. Foram mantidas as quantidades de valores de médias das repetições, as quantidades de lotes e de resultados, somente onde houve resposta (germinação) consideravelmente altas. Foram eliminados os testes cujas condições adversas extremas resultaram em baixa germinação, bem como foram excluídas sementes com defeitos e visualmente indicadoras de baixo vigor.

3.2 Teste de velocidade de emergência em campo

No teste de emergência de plântulas utilizou-se o delineamento de blocos casualizados com oito tratamentos correspondentes aos oito lotes, com quatro repetições de 100 sementes cada um, totalizando 400 sementes por lote. Sementes dos 8 lotes foram semeadas em parcelas constituídas por uma linha de 5 m de comprimento, onde foram distribuídas 100 sementes espaçadas de 0,05 m, em sulco de 0,03 m de profundidade, sendo a distância entre as linhas de 0,50 m. As leituras de emergência foram feitas em intervalos de 7 dias num total de 3 leituras, considerado a partir do início da semeadura. Para efeito de avaliação, quando a plúmula rompe o coleóptilo são contadas como plântulas emergidas. .

Além do percentual de emergência de plântulas, das contagens diárias calculou-se o tempo médio de emergência, baseado em Labouriau (1983) pela expressão:

$$\bar{t} = \frac{\sum_{i=1}^k n_i t_i}{\sum_{i=1}^k n_i} \quad (0 < \bar{t} \leq k),$$

onde t_i : tempo entre o início do experimento e a i -ésima observação (dia ou hora); n_i : número de plântulas emergidas no tempo t_i (não o número acumulado, mas o número referido para a i -ésima observação) e k : último tempo de emergência de plântulas.

3.3 Efeito dos tratamentos na medida de correlação

As expressões para cálculo da correlação, os dados originais e os resíduos aplicados para medidas de germinação (germinabilidade, tempo médio, medidas de velocidade, de uniformidade e sincronia) obtidas em delineamentos experimentais foram extraídos de Santana; Ranal (2006), gentilmente cedidos pelas autoras. No primeiro experimento, sementes de *Brassica chinensis* L. var. *parachimensis* (couve-da-Malásia) foram semeadas em gerbox contendo 150 g de solo ou substrato comercial (Plantimax®, sendo umedecidas com solução da fração alcalóides totais extraída da casca de *Copaifera langsdorffii* Desf. a 0; 62,5; 125,0 e 250,0 mg mL⁻¹, com extrato aquoso de folhas fotossinteticamente ativas da mesma espécie (extrato bruto) e com água como tratamento controle. O extrato aquoso foi obtido fotosinteticamente de folhas ativas, secas a 40-45°C, e trituradas em água na proporção de 1:10 em peso por volume. A quantidade de água ou solução adicionada em cada gerbox foi calculada usando a capacidade do substrato no campo, sendo 130 mL para o solo e 115 mL para o substrato comercial. O pH da solução da *Copaifera langsdorffii* foi de 5,7. Quando necessário, água destilada foi adicionada aos substratos. O experimento foi feito sob iluminação contínua de luz branca fluorescente (32,86 μmol m⁻² s⁻¹) entre 24°C (média das mínimas temperaturas) e 26°C (média das máximas temperaturas). O delineamento experimental foi o inteiramente casualizado, com 10 tratamentos e seis repetições, totalizando 60 parcelas para cada medida usada para o cálculo da correlação.

No segundo experimento, sementes de *Matayba guianensis* Aubl. coletadas de indivíduos estabelecidos na borda da mata de galeria da Estação Ecológica do Panga, no Instituto de Biologia da UFU, Uberlândia-MG, foram semeadas em três tipos de solo (solo de floresta de galeria, solo de cerrado e solo de campo sujo), localizados na mesma Estação Ecológica. Solo e sementes sem resíduos foram colocadas em caixas gerbox e submetidas a iluminação contínua com luz branca (26,73 μmol m⁻² s⁻¹) entre 23,6°C e 24,2 °C (médias de mínimas e de máximas temperaturas). O delineamento experimental foi inteiramente casualizado, com três tratamentos e quatro repetições (mata, cerrado e campo sujo), gerando 12 parcelas para cada medida de germinação usada para o cálculo da correlação.

Os testes de germinação para *Brassica ghinensis* var. *parachimensis* foram avaliados diariamente contando novas plântulas emersas após o sexto dia depois de semeada e para a *Matayba guianensis* após o 31º dia. Três variáveis foram calculadas:

G: germinabilidade (%); $t = \sum n_i t_i / \sum n_i$ tempo médio de germinação (LABOURIAU, 1983); $CV_T = s_t / \bar{t} \cdot 100$; coeficiente de variação do tempo de germinação (SANTANA, RANAL, 2006), em que s_t é o desvio padrão do tempo de germinação (LABOURIAU 1983).

3.4 Simulação de conjunto de dados para situações específicas

Para evidenciar situações não caracterizadas com dados originais, foram criadas variáveis com comportamento gráfico sob controle, onde se variou o tamanho da amostra, a tendência da reta em relação à variável auxiliar Z , de cada variável e a dispersão. Partiu-se de resíduos gerados aleatoriamente, mediante escolha controlada de uma faixa de dispersão em relação à média, em cada conjunto, somados algebricamente com os valores obtidos dentro de uma equação de reta com inclinação escolhida. A partir daí foi possível criar nuvens de dispersões com aderência controlada a esta reta regressora que funcionou como geratriz. Garantiram-se as condições de estudo com tamanho de amostra e dispersões variadas sujeitas às condições ideais de modelos lineares.

3.5 Variável auxiliar Z

Normalmente nos gráficos de dispersão que representam o relacionamento entre duas variáveis X e Y os dados são plotados a partir da tendência da regressora de equação $Y=f(X)$ que explicita a variação conjunta delas, ou covariância. Quando existe covariância natural entre as variáveis, esta representação gráfica é suficiente e muito elucidativa. No entanto, para explicitar a covariância nos casos críticos, uma nova metodologia alternativa foi proposta. A partir dos conceitos de regressão linear, foram plotadas variáveis X e Y como dependentes (as mesmas do relacionamento $Y=f(X)$), desta feita para observar separadamente o comportamento individual de cada uma das duas.

Sendo $X(x_1, x_2, \dots, x_n)$ e $Y(y_1, y_2, \dots, y_n)$ as duas variáveis dependentes a serem plotadas no mesmo gráfico, cada uma foi plotada em relação a uma variável auxiliar criada como independente, definida por Z ($Z \in N^+ \quad z=1,2,3..n$), com incremento unitário, onde n é o número de elementos das variáveis X e Y . Além da visualização

gráfica da dispersão individual de cada uma das variáveis, definem-se suas equações genéricas como:

$$Y = \alpha_Y + \beta_{ZY} \cdot Y_i + \varepsilon_i \quad (E 1)$$

$$X = \alpha_X + \beta_{ZX} X_i + \varepsilon_i \quad (E 2)$$

onde: X e Y são as expressões usuais de retas em função de seus parâmetros alfa e dos coeficientes angulares, respectivamente β_{ZY} e β_{ZX} acrescidos de cada resíduo.

Para que ambas as variáveis possam ser traçadas no mesmo diagrama sem perda de resolução gráfica, equalizaram-se as escalas multiplicando-se todos os dados de uma delas por potências de 10. Este procedimento é típico de mudança de escala e não interfere no resultado do cálculo do coeficiente de Pearson (1930) entre variáveis.

Também é sabido que no cálculo da correlação nenhuma ordenação entre os pares é exigida, ou seja, qualquer que seja a ordenação feita entre as variáveis, mantendo-se o pareamento original, o resultado do r calculado é o mesmo.

O objetivo do método alternativo proposto é a visualização gráfica do comportamento das variáveis constatando se há similaridade no comportamento, a partir das curvas de evolução ou perfis de cada variável, e da covariância, ou seja, crescimento ou diminuição. Esta só é possível quando se re-ordenam as duas variáveis a partir de uma delas, mantendo-se o pareamento.

Num mesmo gráfico podem ser traçadas diversas variáveis que estão sendo comparadas entre si mas sempre uma delas deve ser escolhida como referência para ordenação. Se as demais se mostrarem com perfis similares entre si, mas muito diferente da escolhida, pode-se, alternativamente, mudar a escolha da variável ordenada e observar o comportamento das demais. No caso extremo de dissimilaridade entre todas, ou entre a maioria das que se pretendem comparar, a plotagem terá de ser de duas a duas, mostrando comportamentos distintos entre os pares.

Variáveis com perfis idênticos foram denominadas “similares”, ainda que se ficasse fortemente inclinado a caracterizá-las como homogêneas, ou homoscedásticas. Isto porque não foram encontrados parâmetros na literatura (MILLER, 1994; GARSON, 1999 E DEL SIEGLE, 2006) para homogeneidade de variâncias devidas à resíduos por tendência.

O artifício do uso da variável auxiliar Z com incrementos unitários permitiu gerar modelos de regressão linear com inclinações diversas sem que a inclinação interferisse nos resultados. Constatou-se, pois, que a magnitude dos coeficientes lineares das retas regressoras $Y=f(Z)$ e $X=f(Z)$ obtidos não interferem no resultado da correlação, a menos de inclinações próximas de nula. A inclinação deste eixo da elipse está associada aos valores máximos e mínimos das variáveis: amplitude. O traçado genérico pode ser visto no exemplo da Figura 1. Desde que exista uma inclinação bem definida verificou-se que eventual achatamento pela mudança de escala, mudando substancialmente a amplitude, também não altera o valor do r calculado. Se houvesse interferência, a utilização do recurso de mudança de escala de uma delas seria inviável.

Definiram-se os coeficiente angulares b_x ou b_y como sendo respectivamente cada uma das inclinações das variáveis a serem correlacionadas, quando traçadas pelo modelo linear de regressão em função da variável auxiliar Z . As inclinações também são denominadas de tendência da variável X e tendência da variável Y

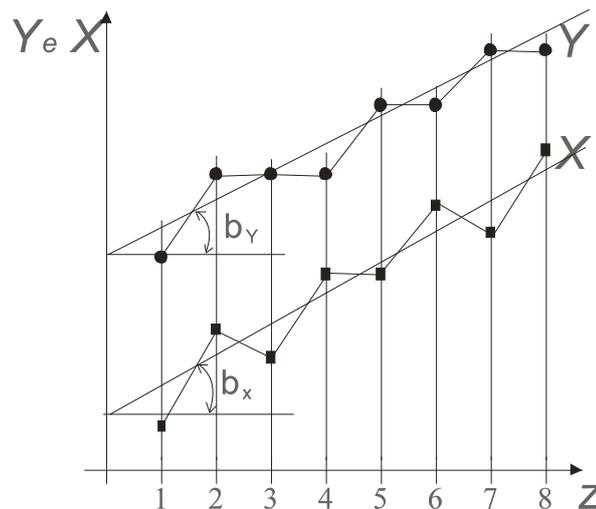


FIGURA 1– Variáveis individuais X e Y em função da variável auxiliar Z , $Y=f(Z)$ e $X=f(Z)$

4 RESULTADOS E DISCUSSÃO

4.1 Testes de germinação por envelhecimento acelerado

Em testes de germinação por envelhecimento acelerado, os dados aproveitados de 8 lotes ($n = 8$) foram organizados nas linhas da Tabela 1A do Anexo A. Cada lote foi submetido a 15 condições ($n = 15$) correspondente a 3 temperaturas (36°C, 38°C e 40°C) vezes 5 dias (24 h, 48 h, 72 h, 96 h e 120 h), ofereceram uma matriz de 8 X 15. A leitura dos vetores linhas e colunas oferece 15 variáveis “condições”, $n = 8$ dados de lotes; e 8 variáveis *lotes*, $n = 15$ dados.

Variáveis *condições* e variáveis *lotes* foram comparadas entre si, sempre precedidos pela análise gráfica da forma alternativa proposta neste trabalho, conforme as orientações de Meko (1981), Miller (1994), e Schield (1995) quanto à necessidade de se observar o comportamento gráfico.

A Figura 2a mostra as variáveis condições traçadas em função da variável Z, todas com perfis semelhantes.

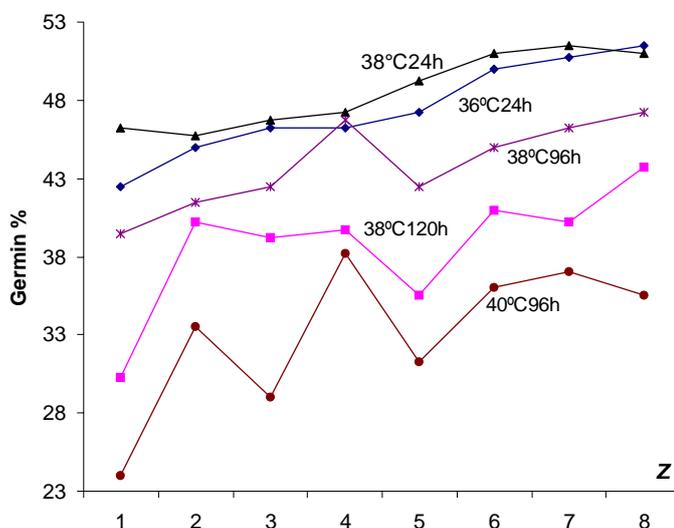


FIGURA 2a. Germinabilidade (%) de sementes de sorgo (*Sorghum bicolor*) oriundas de 8 lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU Santos, (1985). As condições de germinação (temperatura e tempo de permanência na condição térmica) estão apresentadas ao lado das curvas, previamente ordenadas pela de 36°C24h, em função da variável auxiliar Z.

Na Figura 2a, por questão de clareza, foram plotadas somente cinco das 15 condições. Ao serem as variáveis ordenadas na plotagem pela $36^{\circ}\text{C}24\text{h}$, verificam-se perfis similares, mas com uma relativa diferença de ordens de grandeza das variações. Mesmo que todas elas tenham mostrado tendências de crescimento, as correlações encontradas, conforme Tabela 1, variaram bastante. Esta pôde ser uma das constatações facultadas pelo gráfico. Como a expressão da correlação de Pearson (1930) submete os dados a uma padronização para eliminar os efeitos de escala, se há poucos dados, n pequeno, o fator de padronização não chega a homogeneizar os resíduos e qualquer pequena alteração em qualquer dado altera substancialmente o resultado.

TABELA 1. Valores de correlação r encontrados na comparação uma a uma, de variáveis contendo oito dados de lotes, correspondentes às 15 combinações de condições de germinabilidade de sementes de sorgo (*Sorghum bicolor*) em 3 condições de temperatura (indicadas no cabeçalho e na lateral da tabela) e 5 tempos de permanência (indicados na segunda linha de cabeçalho e encabeçando linhas da tabela).

		36°C				38°C					40°C				
		48h	72h	96h	120h	24h	48h	96h	72h	120h	24h	48h	72h	96h	120h
36C	24h	0,91	0,51	0,79	0,76	0,93	0,81	0,83	0,83	0,31	0,92	0,94	0,77	0,71	0,64
	48h		0,33	0,58	0,60	0,88	0,56	0,64	0,60	0,06	0,85	0,83	0,63	0,42	0,43
	72h			0,46	0,64	0,30	0,59	0,63	0,56	0,27	0,40	0,53	0,11	0,45	0,22
	96h				0,73	0,71	0,86	0,84	0,87	0,32	0,79	0,75	0,62	0,73	0,42
	120h					0,48	0,82	0,82	0,78	0,34	0,57	0,62	0,34	0,81	0,32
38C	24h					0,70	0,67	0,70	0,27	0,95	0,95	0,89	0,55	0,71	
	48h						0,77	0,79	0,51	0,81	0,83	0,64	0,85	0,54	
	72h							0,98	0,55	0,66	0,69	0,48	0,83	0,49	
	96h								0,59	0,67	0,71	0,60	0,88	0,61	
	120h									0,22	0,28	0,27	0,75	0,46	
40C	24h										0,97	0,81	0,55	0,55	
	48h											0,84	0,61	0,68	
	96h												0,57	0,88	
	120h													0,65	

Comparando-se dados de $36^{\circ}\text{C}24\text{h}$ e $36^{\circ}\text{C}48\text{h}$ obteve-se $r = 0,91$, ao passo que entre $36^{\circ}\text{C}48\text{h}$ e $36^{\circ}\text{C}72\text{h}$, obteve-se $r = 0,33$. Observa-se que na comparação entre ambas as condições de temperatura e tempo de permanência, pela similaridade das curvas, há condições que são teoricamente favoráveis a correlações altas entre os pares de dados. São exemplos de variações bruscas entre tantas outras mostradas pela Tabela

1 que não puderam ser explicadas por nenhuma causa, como falha de medida ou de variação biológica. Sendo assim, todas as correlações da Tabela 1, têm validade duvidosa pelo **baixo n** , ainda que em alguns casos ofereçam altos valores de r .

Analisando-se de outra forma, comparam-se variáveis *lotes* com 15 dados mostrados na Figura 2b.

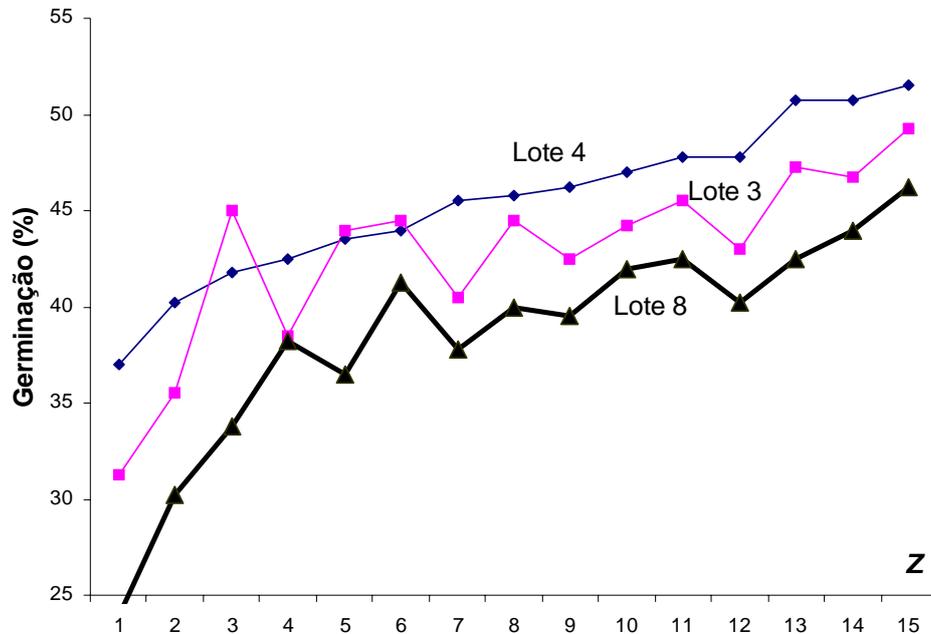


FIGURA 2b. Germinabilidade (%) de sementes de sorgo (*Sorghum bicolor*) oriundas de 8 lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU Santos, (1985). Na condição de germinação à temperatura de 36°C e tempo de permanência térmica por 24 horas, as 3 variáveis *lotes* identificadas ao lado das curvas, são traçadas em função da variável auxiliar Z de incremento unitário, ordenadas pela *lot* 4.

O lote 8 tem valores de percentual de germinação bem inferiores aos demais lotes. Uma vez que é objetivo da tecnologia de semente isolar por comparação lotes com maior vigor, sendo o lote 8 de menores percentuais de germinação, seria intuitivo pressupor que os valores de correlações obtidos entre o lote 8 e os demais fossem diferentes, e menores, daquelas correlações entre os lotes de alto percentual de germinação. As séries de dados são plotadas em função da variável auxiliar Z e todas variáveis, mesmo as não plotadas, têm os perfis bastante similares, inclusive do lote 8. As correlações obtidas foram altas e estáveis porque a quantidade de elementos, obtidos entre as colunas da Tabela 1.A, relativas a cada lote, cresceu para $n = 15$ (3 temperaturas e 5 tempos de permanência).

É usual em tecnologia de sementes que se faça a comparação apenas de

percentual de germinação entre variáveis *condições*, correlacionando-se entre si todas as 15 temperatura X tempo de permanência, contendo normalmente 8 dados (médias de repetições). No entanto, ao se compararem variáveis *lotes* por correlações, utilizando-se os 15 dados de condições encontrados em cada lote (3 diferentes temperaturas e 5 tempo de exposição), oferecem-se $n = 15$, maiores, portanto, que os 8 usuais. Isto facultaria resultados de correlações mais confiáveis. Comparariam justamente lotes, e não testes.

Os resultados de r encontrados atenderiam às prescrições da correlação, mostrando graficamente similaridade das variáveis, aproximando-se da prescrição de Garson (1999) quanto à homogeneidade, mas parecem falhar no quesito independência dos dados, prescrito Del Siegle (2006) e Grasland (2000). Não são independentes porque foram submetidos a efeito de tratamentos e, conseqüentemente, mantiveram comportamentos parecidos. Da mesma forma como os efeitos de tratamentos foram identificados nos resultados de correlações em amostras inteiramente casualizadas, pode-se verificar que a covariância foi forçada. Assim, as sementes se submetem a efeitos idênticos, afetando para mais ou para menos sua germinação. Os resultados de correlação daí obtidos, mostrados na Tabela 2, são bastante altos e poderiam ser considerados estáveis e válidos, mas são de utilidade prática duvidosa porque mesmo o lote 8, com dados bastante diferentes, tem resultados de r que se equiparam aos demais.

TABELA 2. Valores de correlação r encontrados na comparação uma a uma, de variáveis lotes indicados no cabeçalho e indentando cada linha, de $n = 15$ combinações de condições de germinabilidade de sementes de sorgo (*Sorghum bicolor*) em 3 condições de temperatura (indicadas no cabeçalho e na lateral da tabela) e 5 tempos de permanência.

	Lote 2	Lote 3	Lote 4	Lote 5	Lote 6	Lote 7	Lote 8
Lote1	0,962	0,702	0,878	0,814	0,838	0,826	0,863
Lote 2		0,664	0,832	0,811	0,814	0,832	0,876
Lote 3			0,848	0,584	0,834	0,743	0,882
Lote 4				0,755	0,883	0,853	0,924
Lote 5					0,658	0,620	0,777
Lote 6						0,901	0,933
Lote 7							0,8439

4.2 Testes de velocidade de emergência em campo

Os dados de velocidade de emergência obtidos em campo ofereceram gráficos de

cada seqüência de dados em relação a Z . Também puderam ser analisados sob duas óticas relativas às linhas e colunas da matriz conforme Tabela 2A do Anexo A. Primeiramente, os vetores colunas oferecem variáveis *lotes*, $n = 10$ dados (dias de leitura de 5 a 14), uma para cada dia, plotados na Figura 3a. em função da variável auxiliar Z .

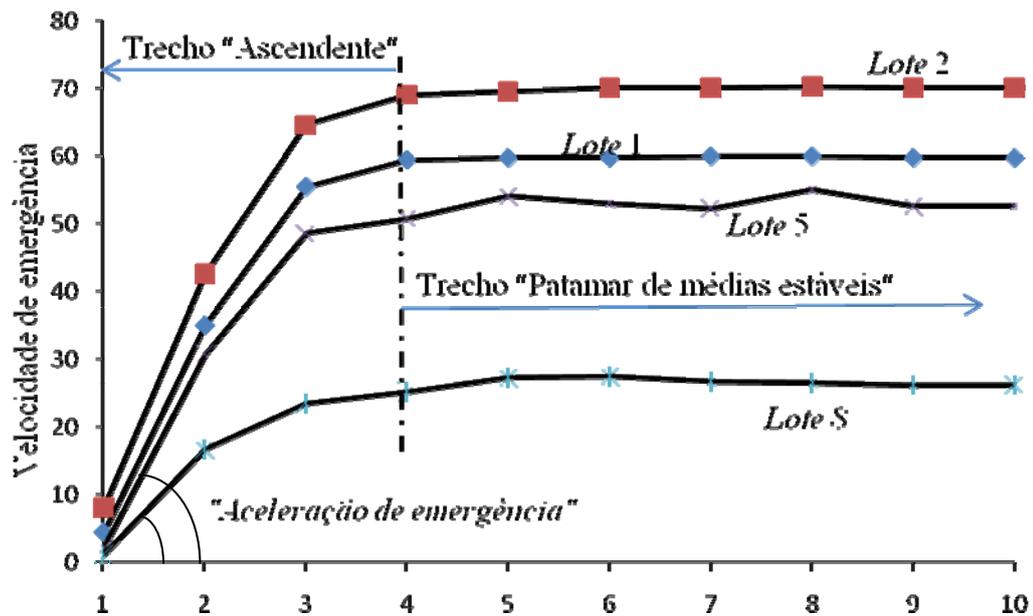


FIGURA 3a. Emergência de plântulas, lotes (plântulas emersas.dia⁻¹) de sementes de sorgo (*Sorghum bicolor*) oriundas de 8 lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU Santos, (1985). As variáveis lotes são identificadas ao lado das curvas traçadas em função da variável auxiliar Z todas naturalmente ordenadas. Coeficiente angular do trecho ascendente entendido como “aceleração de emergência”. Trecho no patamar estável após 3ª leitura, sem covariância.

Era de se esperar que a plotagem destes dados da Figura 3a, com leituras de percentuais acumulados de emergência do quinto ao 14º dia de leituras diárias, com pequeno $n = 10$, trouxesse o mesmo tipo de problema do estudo de caso em germinação de sementes por envelhecimento acelerado. Isto não aconteceu devido a dois motivos constatados:

1. a geometria idêntica de todos os perfis das variáveis ordenadas demonstra graficamente a similaridade gráfica das variáveis, parecendo se aproximar da homogeneidade preconizada por Garson (1999).
2. a existência em todas elas de valores velocidade de emergência muito pequenos, como são próprios do teste, nos inícios de cada série de dados de todas as variáveis. Eles “forçam” a inclinação da nuvem de dispersão

causando acentuadas tendências a todas. Isto pode ser atribuído à amplitude forçada, funcionando como se fossem *outliers* (sozinhos ou em grupo), sobretudo no início das séries.

Ainda que na nuvem de dispersão tais valores possam permanecer dentro da elipse, estes valores atuam como *outliers*, mesmo que seu valor não ofereça resíduos com valor acima de 3 a 4 vezes o desvio padrão. Pode-se dizer que tais valores fazem parte do padrão ou interpretá-los como *outliers*, depende da visão do pesquisador. Como todos eles têm a mesma característica, um “defeito inato”, independentemente de cálculo, seria possível prever que todos os valores de correlações calculadas nestas condições, ofereceriam resultados próximos de $r = 1,00$, como se pôde constatar na Tabela 3, a seguir:

TABELA 3. Valores de correlação r encontrados na comparação de variáveis *lotes* indicados no cabeçalho e indentando cada linha, de $n = 10$ dias de leitura de sementes de sorgo (*Sorghum bicolor*) emersas por dia do quinto ao 14º dia.

	<i>Lote 2</i>	<i>Lote 4</i>	<i>Lote 5</i>	<i>Lote 8</i>
<i>Lote 1</i>	1,00	0,991	0,988	0,973
<i>Lote 2</i>		0,999	0,990	0,978
<i>Lote 4</i>			0,989	0,9701
<i>Lote 5</i>				0,9814

Observa-se que mesmo o lote 8, com valores de velocidade de emergência bem inferiores aos outros lotes, mantém resultados altos de correlações com os demais.

Conforme a prescrição proposta por Sokal; Rohlf (1995) e Neter; Wasserman; Kurtner (1985), a necessária distribuição normal (bivariada) entre as variáveis a serem comparadas, seria suficiente para justificar inexistência de *outliers*.

No entanto, foi observado que estas variáveis em estudo seguem distribuições do tipo normal, com ligeira assimetria, média de 2,5. Portanto, os valores iniciais das séries também não se qualificariam como *outliers*, segundo os citados autores. Não sendo *outliers*, os valores de correlação entre estas variáveis não devem ser fortemente influenciados por eles.

Verifica-se que a proposição de Weldon (2000) de identificar *outliers* quando fogem muito do contorno elíptico da nuvem de dispersão, pode ser melhor aproveitada quando se analisam as variáveis individualmente, com a utilização da variável auxiliar Z .

Os mesmos lotes 1, 2, 4, 5 e 8 plotados na Figura 3a, se forem retirados os dados correspondentes ao primeiro dia de leitura, continuarão com $n = 8$, pequeno, mas os valores de correlação já cairão de próximos de 1,00 para uma média de 0,80, conforme Tabela 4.

TABELA 4. Valores de correlação r encontrados na comparação de variáveis *lotes* indicados no cabeçalho e indentando cada linha, de $n = 9$ dias de leitura (sem a leitura do 5º dia) de sementes de sorgo (*Sorghum bicolor*) emersas por dia.

	<i>Lote 2</i>	<i>Lote 4</i>	<i>Lote 5</i>	<i>Lote 8</i>
<i>Lote 1</i>	0,99	0,71	0,81	0,86
<i>Lote 2</i>		0,70	0,83	0,87
<i>Lote 4</i>			0,82	0,79
<i>Lote 5</i>				0,86

Se for também retirado o segundo dia de leitura, $n = 8$, as correlações caem para média de 0,60 com valores bem mais baixos, mas estáveis. Esta é a característica do índice velocidade de emergência, manter o perfil da variável em velocidade crescente até se estabilizar. Este é o princípio básico da covariância, base da correlação (crescimento ou decréscimo em conjunto). Ele é obedecido em todas as variáveis, nas séries como um todo. Os valores extremos iniciais “puxam” a inclinação da nuvem de dispersão.

Em segundo lugar, se a análise for direcionada para comparar os vetores linhas, eram 10 variáveis condições *dias*, $n = 8$ médias de repetição, e algumas das curvas são apresentadas na Figura 3b. Ainda que este $n = 8$ seja relativamente pequeno, as características de similaridade de perfis de variáveis, descritas anteriormente, são mantidas. Há um crescente de velocidade de emergência de plântulas em função do desempenho de cada lote, sem *outliers*, pelo menos como usualmente caracterizados pela literatura.

Também, por motivo de clareza, apenas 3 das 9 condições *dias* são apresentados na Figura 3b

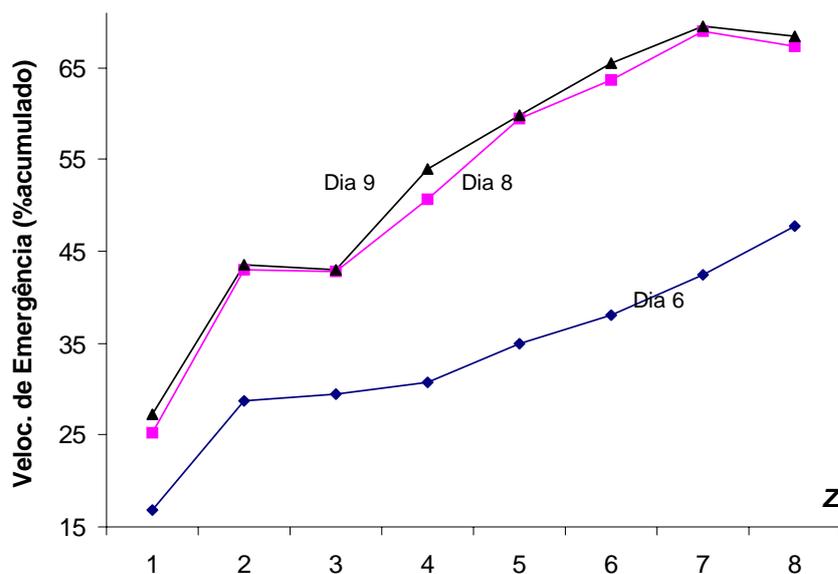


FIGURA 3b. Emergência de plântulas, velocidade (plântulas emersas.dia⁻¹) de sementes de sorgo (*Sorghum bicolor*) oriundas de dados de 8 lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU Santos, (1985). As variáveis *dias* de leitura são identificadas ao lado das curvas traçadas em função da variável auxiliar *Z* de incremento unitário, ordenadas pela curva *Dia 6*.

Os perfis são bem semelhantes, apresentados pelas curvas de todos os dias de leitura. Mesmo no primeiro dia, ainda com baixos valores e velocidade, e mesmo com $n = 8$, pequeno, a similaridade dos resultados em altos valores das correlações podem ser previsto máximos, próximos de $r = 1,00$.

A Tabela 5, a seguir, parece confirmar esta previsão. Nem mesmo reduzindo a quantidade de dados para um n ainda menor a estabilidade destas correlações pôde ser afetada em simulações.

TABELA 5. Valores de correlação r encontrados na comparação de variáveis dia de leitura, indicados no cabeçalho e indentando cada linha, de $n = 8$ médias de lotes, de dados de velocidade de emergência (plântulas emersas. dia⁻¹) de sementes de sorgo (*Sorghum bicolor*).

	<i>Dia 08</i>	<i>Dia 09</i>	<i>Dia 11</i>	<i>Dia 13</i>
<i>Dia 06</i>	0,96	0,95	0,96	0,95
<i>Dia 08</i>		0,99	0,99	0,99
<i>Dia 09</i>			0,99	0,99
<i>Dia 11</i>				0,99

Mesmo com baixo n , a correlação se manteve estável e alta, mas também

sujeitas às mesmas críticas da falta de padronização do efeito de tratamento conforme Santana; Ranal, (2006), tanto em relação ao crescente observado pela seleção de lotes, quanto pelos valores crescentes de velocidade. Os resultados foram os previsivelmente esperados.

Pode-se observar neste trabalho, que estas variáveis têm o comportamento bastante próximo da classificação de “naturalmente correlacionadas” dada por Miller (1994) e Schield (1995) passíveis, segundo os autores, a ter a possibilidade de um caráter inferencial.

Nesta ótica, os perfis dos gráficos de velocidade de emergência levados a efeito em laboratório são idênticos aos dos perfis em testes realizados no campo. Assim, as variáveis se mostram mais suscetíveis à tarefa comercial de inferência de produtividade na projeção desde dados de laboratório até o real em campo.

4.3 Entre variáveis de testes diferentes entre si.

Os resultados de testes de envelhecimento acelerado têm curvas com perfis bastante dissemelhantes das curvas de velocidade de emergência. Por isto, é intuitivo que serão encontrados resultados de correlações discrepantes, ainda que conduzidos com os mesmos lotes. É natural que existam as diferenças de condições biológicas que culminam em reações diferentes entre eles. Fatores diversos devem influir para isto. No entanto, como se observou nas variações geométricas das curvas, já demonstrado, as variações nos resultados atribuídas a fatores biológicos parecem ser desprezíveis face a ordem de grandeza de variação de resultados por causas estatísticas. Os resultados estranhos parecem vir da não observação efetiva das restrições elencadas na literatura para a correta utilização do coeficiente de Pearson (1930).

Verifica-se uma influência marcante de as curvas serem tão diferentes. Na comparação entre variáveis *lotes*, os dados têm diferentes dimensionais, até o pareamento em quantidade de dados é complicado de se escolher. Como as variáveis originais têm n condições diferentes entre si (15 para germinação por envelhecimento acelerado e 10 de velocidade de emergência), há necessidade de adequações escolhendo-se, a critério do pesquisador, um trecho de 10 dados da maior série ($n = 15$).

Neste exemplo, foram escolhidos para comparação, as curvas mostradas na Figura 4a, ambas são as variáveis *lote 2*, de cada conjunto de dados de germinação por envelhecimento acelerado e de velocidade de emergência.

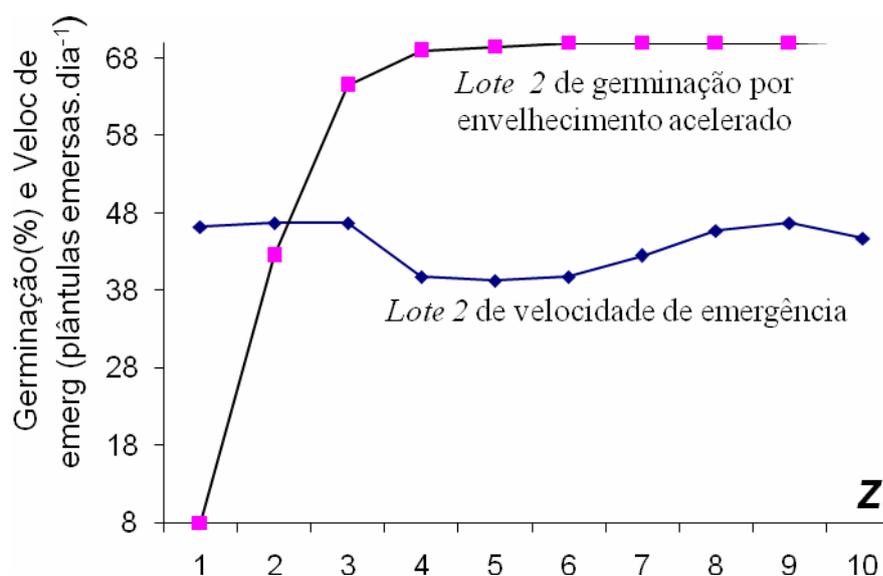


FIGURA 4a. Ambas variáveis *lote 2*, identificadas ao lado das curvas pelas procedências de testes distintos de velocidade de emergência (plântulas emersas.dia⁻¹); e de germinabilidade (%) em envelhecimento acelerado, de sementes de sorgo (*Sorghum bicolor*) oriundas de safras e qualidade fisiológica distintas existentes no LASEM-UFU Santos, (1985) traçadas em função de Z, ordenadas pelos dados do *lote 2* de germinação.

Puderam ser observadas discrepâncias na correlação entre resultados de envelhecimento acelerado de sementes contra os de velocidade de emergência de plântulas. Os resultados são aleatórios. Conforme Figuras 4a quando se comparam seqüência de dados diferentes dentro de um mesmo *lote 2*, os perfis diferentes permitem identificar falta de homogeneidade no senso estrito deste trabalho.

Da tabela 1A do Anexo A, relativa à variável de envelhecimento acelerado, escolheram-se apenas 10 (dos 15) valores variando da temperatura 36°C, todos os dias até 40°C e permanência 24 h para assegurar pareamento com os 10 valores de velocidade de emergência. E da tabela 2A do Anexo A, velocidade de emergência, foram escolhidos todos os 10 valores. Feito isto para todos os lotes. As correlações encontradas nas comparações de todos as variáveis *lotes* de velocidade de emergência, contra e apresentados apenas os resultados contra o *lote 2* de germinação por envelhecimento acelerado, resultaram na Tabela 5.

TABELA 5. Valores de correlação r encontrados na comparação de variáveis todos os *lotes* (de 1 a 8) de velocidades de emergência indicados no cabeçalho, (plântulas emersas.dia⁻¹), contra lote 1 de germinação (%) e identando a linha, de $n = 10$ médias escolhidas, de sementes de sorgo (*Sorghum bicolor*).

	<i>Lote1</i>	<i>Lote2</i>	<i>Lote3</i>	<i>Lote4</i>	<i>Lote 5</i>	<i>Lote6</i>	<i>Lote7</i>	<i>Lote8</i>
<i>Lote 1</i>	-0,35	-0,506	-0,266	-0,396	0,201	-0,18	-0,34	-0,13

Observam-se, como esperado, valores de correlações totalmente discrepantes entre si, variando de valores negativos, como $r = -0,506$ a positivos $r = 0,201$. Não têm sentido prático, independentemente de faltarem ao atendimento de outras prescrições como é o caso de $n = 10$, pequeno.

Quando comparadas variáveis *condições* ($n = 8$) de germinação por envelhecimento acelerado com variáveis *condições dias* ($n = 8$) de velocidade de emergência, as condições comparadas são algumas mostradas na Figura 4b

A validade prática dos resultados deste procedimento, comparando testes de natureza diferente deixa dúvidas porque não são variáveis “naturalmente correlacionadas”, conforme Miller (1994) e Schield (1995), e esta falta de identidade de naturezas podem ser observadas na Figura 4b.

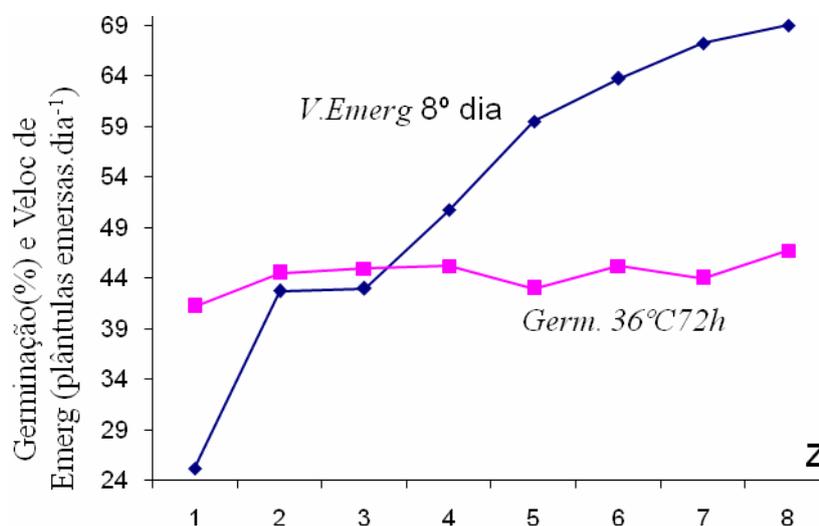


FIGURA 4b. Variáveis *dias* de velocidade de emergência (plântulas emersas.dia⁻¹); e *condições* de germinabilidade (%) em envelhecimento acelerado, de sementes de sorgo (*Sorghum bicolor*) oriundas de lotes de safras e qualidade fisiológica distintas existentes no LASEM-UFU Santos, (1985). A velocidade de emergência no 8º dia de leitura e a germinação (%) na temperatura 36°C e tempo de exposição 72h, são identificadas ao lado das curvas traçadas em função de Z, ordenadas pela curva *Germ36°C72h*.

A plotagem mostra seqüências com perfis não homogêneos, com tendências bastante diferentes, permitindo-se intuir que todas as comparações irão resultar r bastante aleatórios como mostra a Tabela 8.

TABELA 8. Valores de correlação r encontrados na comparação entre variáveis condições *dias* para velocidade de emergência (plântulas emersas.d⁻¹) identicadas no cabeçalho, e *condições* de germinação por envelhecimento acelerado identando cada linha na primeira coluna, para sementes de sorgo (*Sorghum bicolor*)

	Condições <i>dias</i> em Velocidade de Emergência				
	5º dia	6º dia	7º dia	8º dia	9º dia
Germ36°C24h	0,01	0,54	0,46	0,43	0,42
Germ 36°C 48h	0,04	0,51	0,37	0,35	0,32
Germ 36°C 72h	0,64	0,58	0,62	0,63	0,63
Germ 36°C 96h	-0,24	0,25	0,27	0,25	0,28
Germ 36°C 120h	0,20	0,61	0,65	0,64	0,64
Germ 38°C 24h	-0,14	0,38	0,26	0,23	0,22
Germ 38°C 48h	0,11	0,57	0,61	0,61	0,62
Germ 38°C 72h	-0,09	0,52	0,48	0,45	0,47
Germ 38°C 96h	-0,20	0,42	0,38	0,35	0,37
Germ 38°C 120h	-0,10	0,57	0,56	0,52	0,56
Germ 40°C 24h	0,00	0,45	0,39	0,37	0,37

Os resultados de correlações parecem mostrar absoluta falta de relacionamento comparando variáveis condições, que são muito diferentes entre si, tanto provenientes de medidas emergência de plântulas como de germinação de semente. Mesmo sabendo que existe uma “correlação” natural entre os eventos de germinabilidade de sementes em laboratório e emergência de plântulas em campo, de sementes de mesmo lote, os números comparados são de natureza diferentes. Esta falta absoluta de covariância justifica-se pelos perfis diferentes das curvas, aproximando-se do conceito de heteroscedasticidade entre variáveis. O n pequeno cujo efeito nocivo se mostra menos danoso em variáveis muito homogêneas, como foi visto nos exemplos anteriores, neste caso tem influência desprezível face ao comportamento totalmente distinto entre as variáveis. Os valores transitam entre correlações positivas de até $r = 0,64$ e negativos de $r = -0,24$, muitos deles próximos de zero.

A análise prévia do comportamento gráfico poderia e deveria ter sido utilizada para descartar o uso de correlações na finalidade precípua de medir estes

relacionamentos. Não significa, no entendimento deste trabalho, que não haja correlacionamento entre os fenômenos biológicos, mas apenas que os resultados de correlações não têm significado prático.

Em tecnologia de sementes surgem sempre novas propostas de testes destinados a medir vigor de sementes, resultando em novos índices cada vez considerando novos fatores antes não levados em conta. No entanto, fica patente que há uma preocupação acentuada em “validar” pelo uso da correlação os resultados de um teste com outros. Algumas vezes foram encontrados testes cujos índices resultantes são combinação linear entre o anterior e eventual correlação entre as variáveis só podem resultar em $r = 1,00$. Outras vezes, as variáveis são tão matematicamente diferentes da outra, que a comparação entre um conjunto de resultados de uma, se houver algum correlacionamento, é fruto de tamanha aleatoriedade que não terá nenhum sentido prático. E o pior é que não invalida a importância da medida em si.

Além disso, a utilização da correlação, como base precípua para inferir boa produtividade de sementes testadas, deixa de existir porque o objetivo seria, no mesmo tipo de teste, saber se os resultados promissores de laboratório resultariam em estantes produtivos e lucrativos em campo. E desta forma, comparar um teste com outro deixa de ter este objetivo e o objetivo de compará-los perde a essência.

4.4 Efeitos dos tratamento

Em alguns dos testes de vigor de sementes conduzidos em laboratório e levados a cabo por Santana; Ranal, (2006), conforme Tabela 3A do Anexo A deste trabalho, também se verificou a ocorrência de variáveis naturalmente associadas. Alguns dos índices comparados são uns derivados dos outros, em combinação linear, predispondo-os a correlações de valor máximo. São os casos de velocidade de emergência e velocidade média.

O efeito constatado por Santana; Ranal, (2006) de que a correlação entre os dados está majorada pelo efeito dos tratamentos fica patente quando comparada com a correlação calculada entre os resíduos. No entanto, tal efeito só pôde ser observado com precisão nas variáveis naturalmente associadas, de perfis semelhantes, já que a tendência do comportamento gráfico dos resíduos permanece idêntica às dos dados originais. Isto acarreta correlações também consistentes, mas em valor menor. As tendências com dados originais após ordenamento dos pela seqüência de Sincronismo, e

com resíduos, são mostradas nas Figuras de 5a e 5b. As escalas são corrigidas com multiplicadores em potência de 10 na medida do necessário.

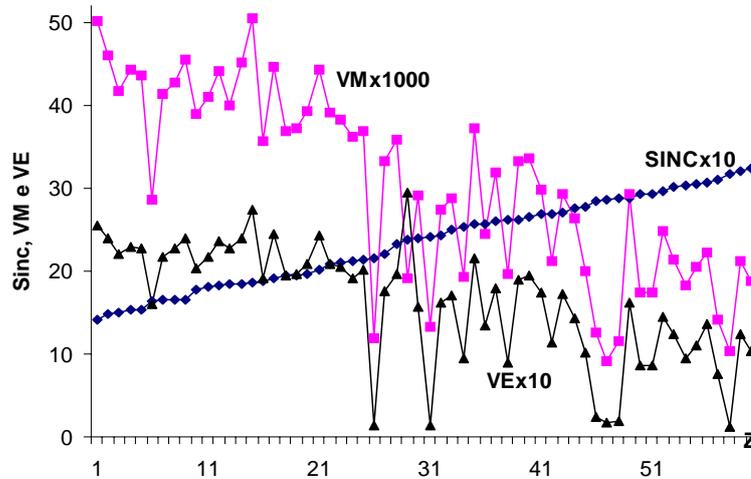


FIGURA 5a. Variáveis velocidade de emergência VE (plântulas emersas.dia⁻¹); de sincronismo $SINC$, e de velocidade média VM , de sementes de couve da malásia (*Brassica chinensis* L. var. *parachimensis*) conforme dados cedidos por Santana, Ranal, (2006). A VM com correção de escala 10^3 , $SINC$ com correção de escala 10, e VE com correção 10, são identificadas ao lado das curvas traçadas em função da variável auxiliar Z , ordenadas pela curva $SINC \times 10$.

Observa-se que o artifício da utilização da variável auxiliar permite verificar que realmente existe uma covariância (crescimento acompanhado de crescimento ou decréscimo) entre as três variáveis, ordenadas pelo índice de sincronismo. No entanto, também é possível intuir que o ordenamento pela velocidade de emergência ou velocidade média deixaria estas duas seqüências VE e VM com perfis mais homogêneos entre si, uma vez que o comportamento dos dados é idêntico por serem uma, combinação linear da outra. As correlações obtidas na comparação entre elas ofereceu os resultados da Tabela 8.

TABELA 8. Correlações entre testes de velocidade de emergência VE (plântulas emersas.dia⁻¹); de sincronismo $SINC$, e de velocidade média VM , de sementes de couve da malásia (*Brassica chinensis* L. var. *parachimensis*) conforme dados cedidos por Santana, Ranal, (2006), com $n = 60$.

	VE	VM
$SINC$	-0,82	-0,70
VE		0,91

A Tabela 8 confirma as previsões de altas correlações, demonstrando numericamente o que a Figura 5a apontava graficamente. As variáveis VE e VM por crescerem no mesmo sentido, são positivas e os perfis de ambos preconizavam alto valor de r . Por outro lado, comparando entre si variáveis Sincronismo, os coeficientes são negativos mas também altos. O $n = 60$ é suficientemente grande para atestar sua validade.

Observa-se na Figura 5a, anterior, que as variáveis VE e VM são semelhantes e o traçado das curvas poderia ter sido conduzido com a ordenação a partir de uma delas. Foi feito isto utilizando-se as variáveis de resíduos, resultantes do procedimento sugerido para isolar a influência numérica dos efeitos de tratamentos, mostrados na Figura 5b.

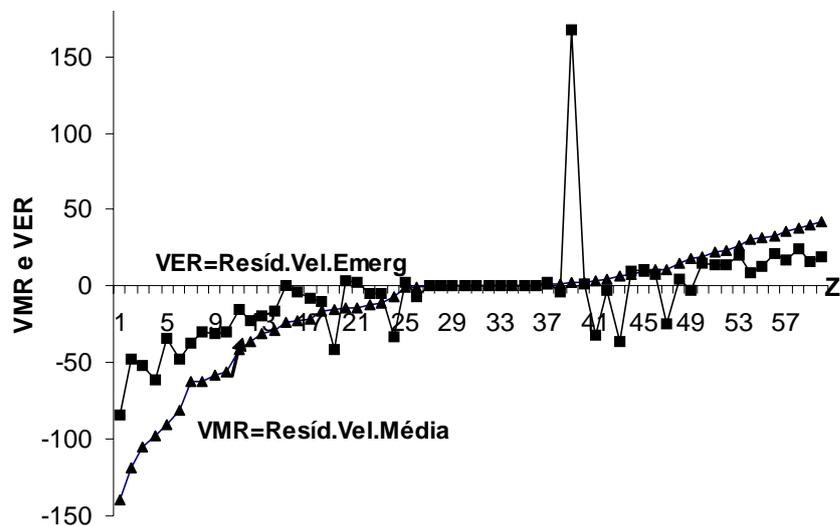


FIGURA 5b. Variáveis de resíduos de testes de velocidade de emergência VER (plântulas emersas. dia^{-1}) e de velocidade média VMR , de sementes de couve da malásia (*Brassica chinensis* L. var. *parachimensis*) conforme dados cedidos por Ranal, Santana (2006). VMR e VER são identificadas ao lado das curvas traçadas em função da variável auxiliar Z de incremento unitário, ordenadas pela curva VMR .

A comparação entre os resíduos dos dados continua com similaridade gráfica entre VM e VE , covariando positivamente e com muita similaridade de formato com os dados originais. Entre estes VMR/VER e os resíduos de Sincronismo ficam mantidos também os mesmos perfis encontrados no traçado gráfico dos dados originais e mostram tendências bem definidas. O fenômeno analisado por Santana; Ranal, (2006) parece ficar claro com a Tabela 9.

TABELA 9. Correlações entre resíduos de dados de velocidade de emergência *VER* (plântulas emersas.dia⁻¹); de sincronismo *SINCR*, e de velocidade média *VMR*, de sementes de couve da malásia (*Brassica chinensis* L. var. *parachimensis*) conforme dados cedidos por Santana, Ranal, (2006), com $n = 60$.

	<i>VE</i>	<i>VM</i>
<i>SINC</i>	-0,70	-0,55
<i>VE</i>		0,74

Comparando a Tabela 9 com a Tabela 8, verificam-se as reduções numéricas, em valores absolutos, de $r = 0,82$ para $r = 0,70$, de $r = 0,70$ para $r = 0,55$ e de $r = 0,91$ para $r = 0,74$. Houve uma redução sistemática e substancial variando de 15% a 22% no coeficiente de correlação. Esta verificação parece comprovar numericamente o que o comportamento gráfico havia identificado previamente, mostrando a similaridade entre os perfis das curvas originais das variáveis com o das curvas traçadas com resíduos. Houve ligeira mudança visualmente explícita nas nuvens de dispersão.

A Figura 5b mostra que, na variável *VMR*, apenas um dos dados no meio da série com valor de 167,90 deveria ser descartado como *outlier*. Ele é do tipo indicado por Sokal; Rohlf (1995) e, Neter; Wasserman; Kurtner (1985), que dizem que as correlações são fortemente influenciados por *outliers*. Observou-se que o descarte deste valor e do seu par na variável comparada, não interferiu nos valores das correlações encontradas, nem nos dados originais, nem com os resíduos.

Atribuiu-se esta pouca influência ao número n suficientemente grande e por ele estar no meio da série, confirmando as afirmações de Cochran; Snedecor (1989) e Zar (1999), dizendo que o aumento do tamanho da amostra, mesmo para variáveis com valores baixos de correlação tem significância estatística maior.

Seguindo as orientações de Weldon, (2000), para a identificação no gráfico de dispersão da relação pronta $Y = f(X)$ dos resíduos padronizados da correlação, verifica-se que ele foge da nuvem de dispersão e se confirmaria realmente como um *outlier*, sem, no entanto, ter interferido no resultado de r . Quando estes *outliers* são encontrados nos finais das séries, sua influência é mais acentuada.

Com esta constatação, procurou-se neste estudo analisar situações onde *outliers* atuam incrementando os efeitos identificados pelos referidos autores como forte influência nos resultados da correlação.

Por outro lado, as observações de Meko (1981), de as correlações serem confinadas a trechos bem definidos da série, parecem se confirmar graficamente. Um conjunto de valores *outliers* distorce mais fortemente o valor da correlação. A este conjunto que forma um trecho de vários elementos da série que fogem sistematicamente do perfil preponderante da variável, foi dado o nome neste trabalho de “trecho dissidente”.

A Figura 6, obtida da plotagem entre dados ordenados pela seqüência de velocidade de emergência (*VEx100*) e de germinação (*GERM*), mostra a clara influência do trecho dissidente quando as variáveis têm *n* suficientemente grande e são homogêneas, ou seja, são naturalmente correlacionadas.

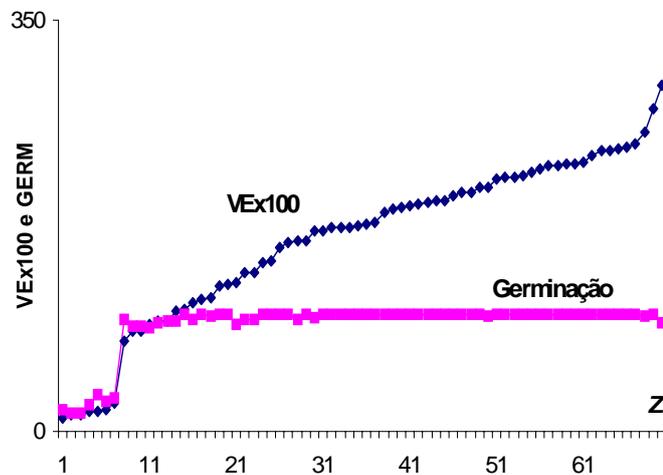


FIGURA 6. Variáveis de velocidade de emergência *VE* (plântulas emersas.dia⁻¹) e de germinação, de sementes de mataíba (*Matayba guianensis* Aubl.) coletadas de indivíduos estabelecidos na mata de galeria na E.E. Panga – UFU, dados cedidos por Ranal, Santana (2006). *VM* e *GERM* são identificadas ao lado das curvas traçadas em função da variável auxiliar *Z* de incremento unitário, ordenadas pela curva *VEX100*.

As duas séries de dados mostram em suas curvas, que só têm em comum a ocorrência de um trecho com brusca queda de valores de ambas as variáveis. Como se trata de valores que incluem as 8 repetições, pode-se intuir que se trata de ocorrência de um trecho dissidente marcado por um conjunto de vários *outliers*. Depreende-se que apenas uma das médias fugiu (composta por oito repetições aí plotadas), se desviou dos valores “normais”, e que provavelmente houve alguma falha de medida por motivos biológicos, temporal ou até de ocorrência previsível, mas que afetou ambos as séries de dados.

Em termos de pesquisa científica, seria altamente sugerível que se verificassem as condições de tal ocorrência ou para eliminar o conjunto de dados, tratando-os como efetivos *outliers*, ou para incorporar a ocorrência como parte de um comportamento padrão, neste tipo de ensaio.

A correlação calculada para toda a série de 60 dados mostrados na Figura 6, resulta em $r = 0,75$. No entanto, observando o trecho inicial apresentado na Figura 6, constata-se um trecho dissidente.

Para efeito de simulação de efeito de trecho dissidente, à medida que se eliminam os valores iniciais o valor obtido da correlação muda substancialmente. Eliminando-se seqüencialmente os 10 primeiros, a correlação cai para $r = 0,43$; os quatorze primeiros, cai para $r = 0,10$.

Finalmente, eliminando-se os 29 primeiros a correlação passa a valer $r = -0,49$, ou seja, deixa de ser positivamente correlacionada em $r = 0,75$ em toda a série para ser negativamente correlacionada em $r = 0,49$, ambas de valores absolutos expressivos (correlações de moderada a alta, conforme tabela para os casos de Davis (1971)).

É muito interessante observar esta seqüência na Figura 6, que permitiria antecipar este efeito. O trecho inicial obrigaria a tendência de ambas as séries a ter inclinação preponderantemente positiva, o que resultaria em covariância de crescimento conjunto. Se for eliminado o trecho inicial, a série de dados de germinação se apresenta com tendência, coeficiente angular, na, enquanto que a de *VE* tem inclinação positiva.

O coeficiente de Pearson calculado na série ordenada a partir de sincronismo entre o 31º até o final, entre os dados originais é $r = -0,49$ e, na mesma série entre os resíduos ela resulta em $r = 0,06$, ou seja, quase nula porque foram eliminados os efeitos de tratamento conforme Santana, Ranal, (2006).

Ainda assim o valor original de $r = 0,75$ não poderia ser enquadrado na definição de “boas correlações”, conforme faixa proposta por Davis (1971). Este assunto será melhor abordado na seção de dados simulados.

O fato de a correlação entre as variáveis ser zero não significa que não haja relacionamento entre os fenômenos, apenas que as variáveis selecionadas não são adequadas para serem comparadas. Somente havia tendência a um decréscimo nos valores de germinação pelo efeito de tratamento, desqualificando as variáveis que não sejam independentes.

Observa-se em todo o contexto deste item que, mesmo tendo as tendências naturalmente associadas, variáveis de testes agrônômicos quando comparadas, têm variações bruscas de r , constataadamente matemáticas, que desestimulam supor que sejam causadas por mudança de vigor ou qualidade, bem como pelos fatores biológicos ou edafoclimáticos, que seriam os grandes objetivos das pesquisas.

Conforme Del Siegle (2006) e Grasland (2000), se os dados são colhidos a partir de classes ou intervalos, como o são os lotes, talvez algum coeficiente de correlação não-paramétrico seja mais indicado para compará-los, se é que haja algum interesse em correlacioná-los e que o resultado tenha alguma utilidade prática e construtiva.

4.5 Dados simulados

Softwares estatísticos como Statistica® e Prophet® apresentaram resultados idênticos ao do MSEXCEL® até na segunda casa decimal. O MsExcel®, não obstante sujeito a críticas diversas conhecidas na literatura sobre a precisão de funções estatísticas, apresentou uma grande vantagem em relação aos demais. Possibilitou acompanhar as variações de resultados numéricos dinâmica e simultaneamente aos efeitos gráficos. Quando se simulam alterações forçadas nos dados individuais, automaticamente são atualizados os resultados numéricos e gráficos. Este artifício foi utilizado manualmente nas simulações feitas.

Nos dados simulados, apresentados na Tabela 4A, no Anexo A, para $n = 20$, a Figura 7a, abaixo, representa as plotagens contra Z de um conjunto de dados V , a princípio gerada com tendência bem definida, com pouca dispersão, e outro conjunto $VZero$, a princípio gerada com baixa tendência, próxima de inclinação zero. Após serem ordenadas pela $VZero$, as inclinações se invertem.

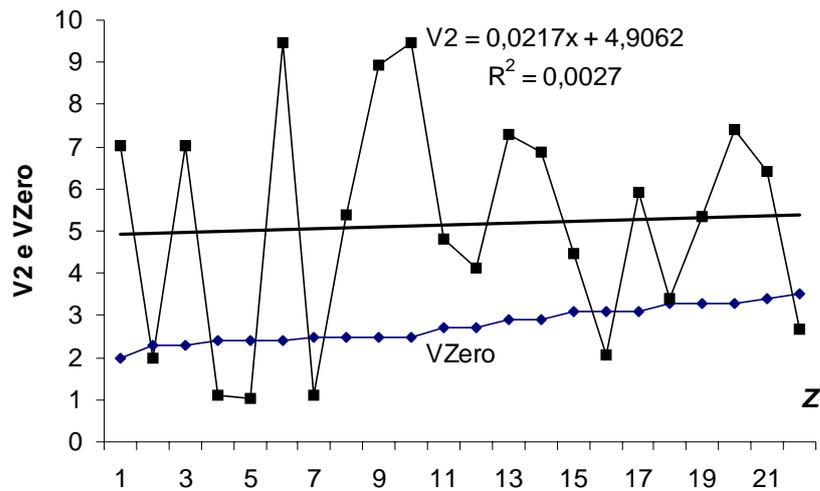


FIGURA 7a. Curvas $VZero$ originalmente de inclinação nula, passa a ter coeficiente angular diferente de zero e $V2$ passa a ter inclinação nula após ser sido ordenado pela $Vzero$. A identificação de $Vzero$ e $V2$ (com equações de regressão), estão indicadas ao lado das curvas.

$VZero$ originalmente gerada a partir de resíduos aleatórios criando uma nuvem de dispersão sobre uma geratriz de inclinação zero foi submetida à correlação com $V2$, de tendência bem definida. Ao ser ordenada pela variável $VZero$, a nuvem de dispersão dos dados de $VZero$, com coeficiente angular originalmente zero, passa a ter uma inclinação. Em contrapartida, a seqüência $V2$ que antes tinha inclinação bem definida, passa a ter inclinação zero.

Esta situação define bem o caso em que não existe covariância, quando uma das variáveis tem inclinação, ou seja, tendência, a outra se comporta com resíduos variando apenas em relação à média, sem tendência. O resultado da correlação neste caso é o esperado de $r = -0,007$, próxima de zero. Esta é, portanto, a situação definida por Rummel (1976), entre tantos, que dizem que a correlação linear é zero porque o crescimento de uma das variáveis não é acompanhado do crescimento ou decréscimo da outra. Neste caso, os textos didáticos mostram normalmente a relação $Y=f(X)$ com a nuvem de dispersão em formato elíptico, tendo o maior eixo se posicionado na horizontal, ou seja, com tendência zero, na plotagem da nuvem de dispersão entre ambas as variáveis.

Uma outra definição para esta situação é de que sendo $r = 0$, não significa falta de correlação entre duas variáveis, apenas reflete a ausência de relação do tipo linear entre essas variáveis (CRUZ, REGAZZI, 1997; DEL SIEGLE, 2006).

Pôde-se observar, na plotagem da Figura 7b, que a modificação de apenas três valores nos extremos da série eleva o resultado dos anteriormente quase zero, para $r = 0,29$. Para visualização, as curvas foram ordenadas graficamente a partir de *V2-Alterada*,.

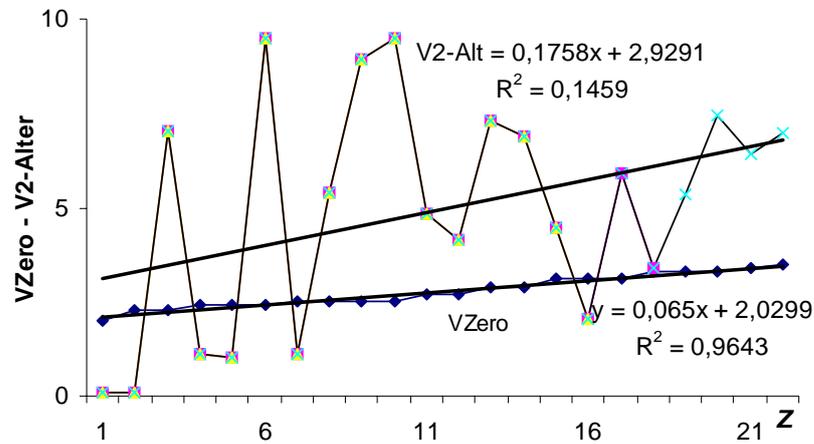


FIGURA 7b. *V2-Alterada* e *VZero* indicadas ao lado da curva, com mudança de apenas um elemento, em função da variável auxiliar *Z*.

Nota-se, portanto, que se alguma das seqüências de dados não apresentar tendência, mesmo com $n = 20$, não tão pequeno -, alterando-se os elementos extremos, o novo coeficiente angular da série já proporciona um relacionamento de correlação entre as variáveis.

Quando as seqüências de dados têm ambas dispersões grandes, como as *V9* e *V10*, ao serem ordenadas por *V9*, o gráfico mostra que *V10* passa a ter um perfil bastante aleatório, característico de falta de tendência, como mostra a Figura 7c. Este é outro caso em que se identifica uma correlação zero por ser realmente nula, ou seja, não existe covariância entre os dados comparados.

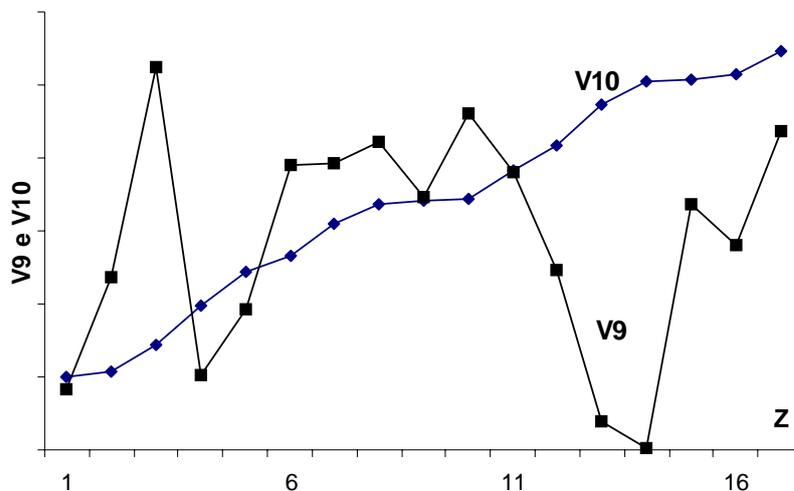


FIGURA 7c. Variáveis $V9$ e $V10$, indicadas ao lado das curvas de grandes dispersões após ordenadas pela $V9$ em função da variável auxiliar Z .

Apesar de o valor deste coeficiente de correlação de Pearson (1930) ter resultado próximo de zero, ele parece poder ser classificado de “válido” e “estável”. No entanto, quanto menor o n , mais instável o r se torna, podendo oscilar bastante com a alteração de poucos dos valores.

Nos exemplos anteriores se configuram correlações instáveis em que três fatores parecem ser condicionantes para esta instabilidade, nas situações:

1. quanto menor o n mais a correlação fica suscetível ao efeito das tendências próximas de zero oferecerem resultados instáveis.
2. quanto maior é a dispersão de uma ou das duas variáveis, mais fácil de ocorrer instabilidade.
3. Quanto mais um *outlier*, (no senso deste trabalho) ou preferencialmente um grupo deles (aqui denominado de “trecho dissidente”), se mostrar graficamente ocupando o extremo de ambas as séries, é previsível um valor alto de correlação com possibilidade de ter significado falso.

Não se buscou equacionar limites de influência associando as influências nocivas de n pequeno, dispersão elevada e trecho dissidente neste trabalho, ficando como desafio para novas pesquisas.

Na seção anterior parece ter ficado patente que a definição de “altas correlações” conforme descrição de Davis (1971), precisaria de complementação. Um valor de $r = 0,75$ conforme o encontrado na correlação entre as variáveis citadas na Figura 6, não

corresponde a uma correlação consistente. Parece que a solução prática não se trata apenas da definição de faixas adequadas para caracterizar correlações. Da mesma forma não se restringe a que elas sejam “boas” numa faixa conforme experiência em cada área de conhecimento, ou da experiência do pesquisador. Com a falta de similaridade gráfica entre as variáveis, somada com o n pequeno, o resultado deveria ser considerado ao acaso, sem associação com o tipo de teste, mesmo com a “experiência” do pesquisador. Desta forma, a sugestão de se apresentar uma “tabela” orientativa de correlações válidas para uma determinada área de conhecimento científico deveria incluir também especificações de variáveis com comportamento homogêneos na comparação entre lotes de um mesmo tipo de teste. Se um pesquisador seguir comparando variáveis dissociadas vai ter sempre uma “tabela” variando entre a estranheza e o bizarro. A “tabela” só seria esclarecedora se relacionasse variáveis sabidamente correlacionáveis, obedecendo as condições de similaridade gráfica das variáveis, de n suficientemente grande, com ou sem a padronização do efeito de tratamento.

As variáveis $V3$ e $V4$ apresentadas no apêndice deste trabalho, com $n = 8$ foram transcritas da Tabela 4A do Anexo A, para este item (Tabelas 10 e 11).

TABELA 10. Variáveis simuladas $V3$ e $V4$ a partir de uma geratriz de reta com coeficiente angular e resíduos gerados randomicamente, obtidos da Tabela 4A do Anexo A, depois de ordenadas a partir de $V3$.

V3	4.00	4.10	5.10	5.40	6.00	6.10	6.40	7.20
V4	2.00	2.90	4.30	3.10	4.50	4.30	5.20	5.40

A correlação obtida entre as variáveis acima resultou em $r = 0,91$. É considerada “alta” com cada, nas faixas propostas por Davis (1971). O elemento de valor 6,00 na Tabela 10 está em evidência para posteriormente ser substituído e verificados eventuais efeitos sobre o resultado da correlação.

Desta Tabela 10 foi obtido o gráfico da Figura 8 que se segue.

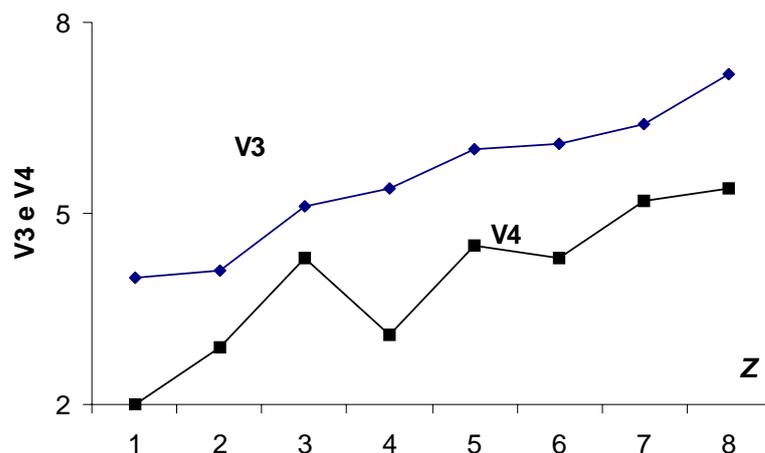


FIGURA 8. Variáveis $V3$ e $V4$, indicadas ao lado das curvas, obtidas com dados simulados com pequenas dispersões, ordenadas a por $V3$, com $n = 8$.

A aparência dos perfis de $V3$ e $V4$ dá indicativos de similaridade gráfica das variáveis e se poderia intuir que a correlação seria alta e estável, ainda que com n pequeno. Acontece que com mera mudança de um dos elementos da $V3$, passando de 6,0 para 4,1 conforme mostra a Tabela 12, o coeficiente sofre alteração abrupta:

TABELA 11. Variáveis simuladas $V3$ e $V4$ alterando-se, a partir da Tabela 10, apenas um dado de $V3$ que era 6,10, para 4,10, hachurado em cinza.

V3	4.00	4.10	5.10	5.40	4.10	6.10	6.40	7.20
V4	2.00	2.90	4.30	3.10	4.50	4.30	5.20	5.40

Com a mudança, houve uma alteração do valor de $r = 0,91$ para $r = 0,74$. Não se pode nem imaginar atribuir a este valor original, nem seu substituto, o nome de de *outlier*. Ele continua absolutamente dentro da elipsóide da nuvem de dispersão. No entanto a correlação teve uma queda de quase 25%. Parece uma sensibilidade excessiva para uma pequena mudança de apenas um dos 16 elementos que compõem ambas as séries. Esta ordem de grandeza parece inviabilizar a ferramenta no limite de sensibilidade para medir variações de germinabilidade de sementes ou emergência de plântulas, que sofrerão, entre lotes, pequenas variações não tão drásticas.

Quando n é pequeno, a variação brusca de r deve ser atribuída às influências do caráter geométrico da correlação. Observou-se que a simples alteração de um dos

elementos mudou a direção da tendência de uma das seqüências de dados. Se considerada numa possível reta regressora, neste caso ela teria a nuvem de dispersão relativamente bem aderente (R^2 acima de 0,65). A mudança brusca de direção da nuvem de dispersão - coeficiente angular de uma possível reta regressora entre as dispersões - foi alterada pela mudança da posição relativa dos resíduos dos pares (x,y) . O novo valor de 4,1 leva o resíduo para o início da série. A nova dispersão passa a ter três valores iniciais da $V3$ “puxando” fortemente uma eventual reta de regressão que posiciona a nuvem de dispersão dela. Daí, o motivo da alteração ser identificada como meramente geométrica, e os resíduos padronizados na equação de Pearson (1930) “pulam” de quadrantes (o valor do coeficiente de Pearson é um somatório de produtos de resíduos que ocupam quadrantes com sinais iguais ou opostos). A alteração brusca do valor se dá exclusivamente porque a reta regressora muda bruscamente de direção. São mudanças que configuram comportamento quase que aleatório.

Observou-se que à medida que aumenta o número n de elementos, sem ocorrência de trechos dissidentes, a correlação fica menos sensível à mudança de valores individuais, ou seja, eventual ocorrência de *outliers*, conforme os descrevem Neter; Wasserman; Kurtner (1985), ainda que com valores grandes, não altera muito o resultado do coeficiente de Pearson (1930).

A formulação do critério para se definir n como suficientemente grande para que um *outlier*, cuja valoração também não foi definida numericamente, ficam como desafios para novos estudos.

As seqüências de dados $V5$ e $V6$ foram geradas, respectivamente, pelo prolongamento das séries numéricas de $V3$ e $V4$, mantendo-se a mesma lei de formação e aleatorização de resíduos. Com este novo e maior valor de n , o coeficiente r tem comportamento estável e, neste caso, alto, com $r = 0,97$. A mudança do quarto elemento do antigo 6,0 para o novo valor de teste atribuído de 4,1 só ocasiona uma leve mudança para 0,96, como era de se esperar.

Conforme as observações deste trabalho, estes valores de $r = 0,60$ e $r = -0,41$ não demonstram covariâncias, no sentido prático. Talvez seja esta a consideração de Meko (1981) dizer que nem sempre a correlação com significância teórica tem significância prática. Neste caso, nenhuma inferência é presumível, ainda que com eventuais altos valores de coeficiente de Pearson (1930), que pareceram estar incorporados de uma tácita aleatoriedade.

Se a dispersão de cada variável for verificada muito diferente uma da outra, os resultados de correlação são imprevisíveis. Se ambas as dispersões são grandes, mesmo com n suficientemente grande, a correlação linear tende a ser pequena mas parece ser suficientemente confiável em oferecer r próximos de zero.

Quando a dispersão é alta, na maioria dos casos encontrados neste trabalho, ficou-se muito tentado a afirmar que, mesmo encontrando um $r = 0,38$, esta seria uma correlação inválida. Esta situação, graficamente, não pareceu representar uma cotendência, mas uma aleatoriedade.

4.6 Dispersões curvilíneas

O caso normal relatado na literatura (MEKO, 1981) é o da tendência curvilínea verificada no resultado da regressão entre os resíduos padronizados do relacionamento entre as duas variáveis. Observou-se, porém, que, tanto pode haver tendência curvilínea da nuvem dos resíduos da regressão de cada variável em relação a Z , quanto tendência curvilínea na dispersão dos resíduos padronizados na regressão de $Y=f(X)$.

Tendências curvilíneas das variáveis

Nas Figuras 9a e 9b a seguir, foram construídas a partir das seqüências de dados $V5$ e $V6$ da Tabela 4A do Anexo A. Observou-se que as curvas plotadas antes de serem ordenadas por uma delas, apresentavam tendências curvilíneas simétricas. No entanto, elas obedeceram a leis de formação que permitiram identificá-las com perfis bastante similares, ainda que simétricos.

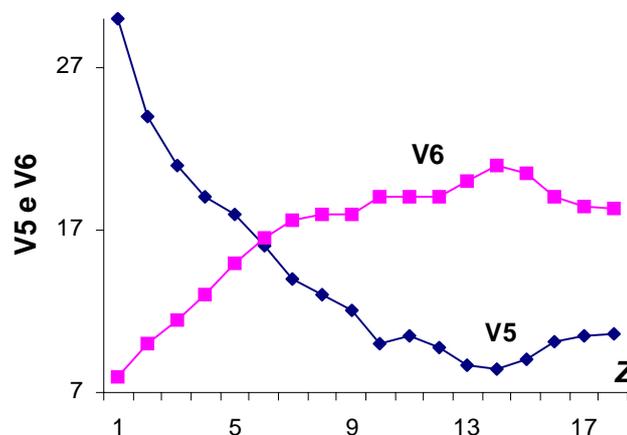


FIGURA 9a. Curvas identificadas na figura $V5$ e $V6$ antes de serem ordenadas traçadas em função da variável auxiliar Z .

correlacionáveis Após análise mais minuciosa, os perfis muito similares por si só explicam se tratar de variáveis naturalmente correlacionáveis. Com o ordenamento por uma delas, os resíduos ainda que pareados tomam as formas retilíneas e a plotagem usual de $Y=f(X)$, como mostrada na Figura 9b mostra uma nuvem de dispersão muito aderente, própria de altas correlações. Neste caso, este tipo de traçado gráfico entre duas variáveis, parece ser suficientemente esclarecedor, mais apropriado neste caso que o da plotagem individual de cada uma contra a variável auxiliar Z .

Assim, verifica-se que, mesmo com tendências individuais de dispersões curvilíneas de cada perfil da série de dados plotada contra Z , a dispersão plotada da relação $Y=f(X)$ é retilínea e oferece uma correlação estável.

O gráfico da regressão entre $V6=f(V5)$ apresentado na Figura 9b, mostrou uma alta aderência dos pontos à reta e, portanto, pôde-se prever, antes do cálculo, que a correlação entre elas seria alta e estável, desde que o n seja suficientemente grande. O valor de coeficiente de correlação de Pearson (1930) $r = 0,989$ foi, como esperado, naturalmente alto.

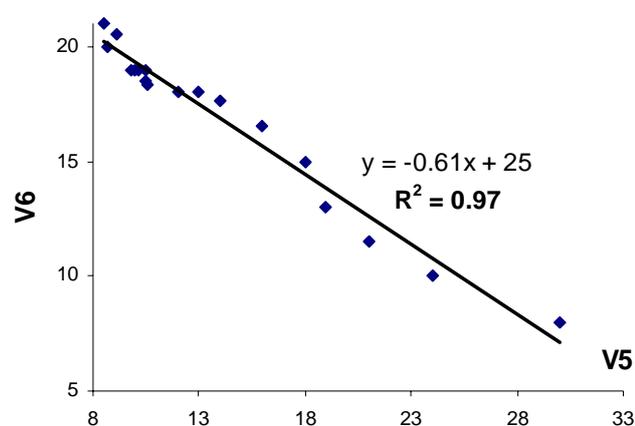


FIGURA 9b. Variável indicada na curva $y = V6$ plotados em função de $V5$

Tendências curvilíneas na regressão $Y = f(X)$

A Figura 9c mostra um exemplo de tendência curvilínea na relação entre as duas variáveis, com uma nuvem de dispersão que ocuparia alternadamente posições acima e abaixo dela.

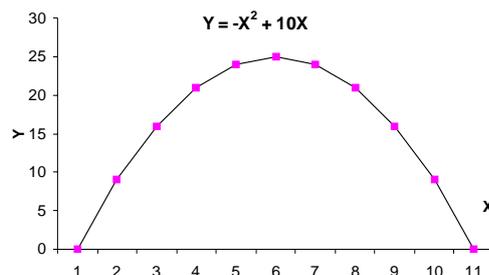


FIGURA 9c – Curva de aspecto parabólico $Y=f(X)$ genérica

Conforme Neter; Wasserman; Kutner (1985); Miller (1994) e Meko (1981), é fácil intuir que não pode haver correlação, já que não atende à condição de nuvem de dispersão linear. As observações práticas confirmam esta condição como *sine-qua-non* para a validade do coeficiente de correlação linear de Pearson (1930).

Verificou-se nos casos analisados que, se as variáveis são naturalmente associadas, sem “trechos dissidentes”, o número de elementos pode crescer indefinidamente que o coeficiente r não perde sua capacidade de responder a uma correlação estável, sem “saltos”. A equação proposta por Fisher, conforme Cochran; Snedecor, (1989) que leva em conta o grau de liberdade ($n - 2$), mostra que realmente n não interfere no coeficiente r , mas interfere na significância, caso as variáveis não sejam naturalmente correlacionáveis.

Outras situações e testes foram avaliados mas não foram inseridos neste trabalho por questão de foco. Técnicas de correlações multivariadas, por exemplo, mostraram-se de uma eficácia muito interessante ao permitirem agrupar, tanto lotes, quanto condições de testes, os mais semelhantes; e visualizá-los graficamente em dendrogramas. O pareamento por tipo de dado permite evidenciar associação natural e consistente entre as variáveis lotes e condições de testes de uma e de outra forma.

No entanto, a similaridade entre duas variáveis ainda não confere ao lote, ou à condição de teste, o atestado de ser melhor, mas simplesmente de serem semelhantes em maior ou menor grau.

4.7 Conclusão

Diante de tão diversificados exemplos práticos e simulados, conseguidos a partir da utilização da ferramenta alternativa do traçado gráfico individual das variáveis em relação à variável auxiliar Z , pela consistência repetitiva dos resultados auferidos, as conclusões encontradas parecem convergir para a generalidade. Por força da prudência foram particularizadas.

Conclui-se que, nos casos analisados, comparando vigor de sementes com resultados entre testes de laboratório, tão bem como no relacionamento destes com os de campo, será difícil distinguir eventuais influências que possam ser atribuídas a variações biológicas das sementes, de alterações temporais, de manuseio ou edafoclimáticas nos resultados das correlações pela, uma vez que preponderam os problemas estatísticos. Resultados estranhos devem ser atribuídos à inobservância das prescrições de uso da correlação, incluindo efeitos de *outliers* (*strictu sensu* do presente trabalho), sobretudo quando agrupados em “trechos dissidentes”, de variáveis não serem similares, definidas neste trabalho como não-homoscedáticas, ou não-naturalmente associadas, em muitos casos os problemas parecem ser agravados pelas amostras pequenas.

Correlação se mostra uma ferramenta indicada na comparação de variáveis *lotes*, sobretudo levando-se em conta $n = 15$ de condições de testes estáveis, na análise de vigor de sementes. Resultados parecem ser mais elucidativos quando os correlacionamentos são levados a cabo entre a testes semelhantes, tanto entre aqueles feitos só em laboratório, quanto entre estes e os de campo, mostrado pela semelhança dos perfis das curvas quando traçadas em relação à variável auxiliar Z .

Correlações conduzidas com os resíduos dos dados originais mostraram que testes com dados com correlações altas, continuam correlacionados mas o valor de r diminui quando se utilizam resíduos.

5 REFERÊNCIAS

- ANDRADE, R. V. et al. Qualidade fisiológica das sementes de milho híbrido simples HS 200 em relação ao tamanho. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 3, p. 576-582, maio/jun. 2001.
- ASSOCIATION OF OFFICIAL SEED ANALYSTS. **Seed vigor testing handbook**. East Lansing, 1983. 88 p. (Handbook of seed testing. Contribution, 32).
- BEWICK, V.; CHEEK, L.; BALL, J. Statistics review 7: correlation and regression. **Critical Care**, Fullerton, v. 7, n. 6, p. 451-459, 2003. Disponível em: <<http://www.pubmedcentral.gov/articlerender.fcgi>>. Acesso em: maio 2006.
- BISHNOI, U. R.; DELOUCHE, J. C. Relationship of vigour tests and seed lots to cotton seedling establishment. **Seed Science and Technology**, Zurich, v. 8, p. 341-346, 1980.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Departamento Nacional de Produção Vegetal. **RAS: regras para análise de sementes**. Brasília, DF, 1992. 365 p.
- CALDWELL, W. P.; DELOUCHE, J. C. Seed vigor and vigor tests. **Proceedings of the Association of Official Seed Analysts**, Genebra, v. 50, p. 124-129, 1960.
- CARVALHO, N. M.; VIEIRA, R. D. **Testes de vigor em sementes**. Jaboticabal: FUNEP, 1994. 164 p.
- COCHRAN, W. G.; SNEDECOR, G. W. **Statistical methods**. 8. ed. Ames: Iowa State University Press, 1989. 503 p.
- CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa, MG: UFV, 1997. 442 p.
- DAVIS, J.A. **Elementary survey analysis** Englewood, NJ; Prentice-Hall, 1971
- DAVIS, U. C.; SAMANIEGO, F. J.; WATNIK, M. R. The separation principle in linear regression. **Journal of Statistics Education**, California, v. 5, n. 31, 1997. Disponível em: <<http://www.amstat.org/publications/jse/v5n3/samaniego.html>>. Acesso em: nov. 2006.
- DEL SIEGLE, D. **EPSY 341: principles and methods in educational research**. Connecticut: Neag School of Education, University of Connecticut, 2006. Disponível

em: <<http://www.gifted.uconn.edu/siegle/EPsy341/EPsy341FALL.htm>>. Acesso em: fev. 2006.

DELOUCHE, J. C.; BASKIN, C. C. Accelerated aging techniques for predicting the relative storability of seed lots. **Seed Science and Technology**, Zurich, v. 1, n. 2, p. 427-452, 1973.

EGLI, D. B.; TEKRONY, D. M. Relationship between laboratory indices of soybean seed vigor and field emergence. **Crop Science**, Madison, v. 17, n. 4, p. 573-77, 1977.

EGLI, D. B.; TEKRONY, D. M.; WIRALAGA, R. A. Effect of soybean seed vigor and size on seedling growth. **Journal of Seed Technology**, Lincoln, v. 13, p. 1-12, 1990.

FREITAS, R. A. et al. Correlação entre testes para avaliação da qualidade de sementes de algodão e a emergência das plântulas em campo. **Revista Brasileira de Sementes**, Brasília, DF, v. 22, p. 97-103, 2000.

GARSON, G. D. **Assumptions of correlation per Garson**. New York: NC State University, 1999. PA765. Disponível em: <<http://www.nyu.edu/its/socsci/Docs/correlate.html>>. Acesso em: mar. 2006.

GRASLAND, C. **Corrélation statistique**. 2000. Disponível em: <<http://www.grasland.cicrp.jussieu.fr/grasland/STAT98>>. Acesso em: mar. 2006.

GREGORY, J.; HOPPER, N.; JIVIDEN, G. Evaluation of several laboratory tests for determining cottonseed quality. In: THE BELTWIDE COTTON CONFERENCE, 1986, Memphis. **Proceedings...** Memphis: Nacional Cotton Council, 1986. p. 62-63.

JOHNSON, R. R.; WAX, L. M. Relationship of soybean germination and tests to field performance. **Agronomy Journal**, Madison, v. 70, p. 273-278, 1978.

KENDALL, M. G.; STUART, A. **The advanced theory of statistics**. London: Charles Griffin, 1961. v. 2.

LABOURIAU, L. G. **A germinação das sementes**. Washington: Organização dos Estados Americanos, Programa Regional de Desenvolvimento Científico e Tecnológico, 1983. 174 p. (Série de Biologia. Monografia, 24).

MARCOS FILHO, J.; TOLEDO, F. F. **Manual das sementes: tecnologia da produção**. São Paulo: Ceres, 1977.

MARCOS FILHO, J.; CÍCERO, S. M.; DA SILVA, W. R. **Avaliação da qualidade das sementes**. Piracicaba: FEALQ, 1987.

MARCOS FILHO, J. et al. Testes para avaliação do vigor de sementes de soja e suas relações com a emergência das plântulas em campo. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 19, n. 5, p. 605-613, 1984.

MEKO, D. M. **Applications of Box-Jenkins methods of time-series analysis to reconstruction of drought from tree rings**. 1981. cap. 9. Thesis (Ph. D.) - University of Arizona, Tucson, 1981.

MILLER, L. E. Correlations: description or inference? **Journal of Agricultural Education**, Carbondale, v. 35, n. 5-7, 1994.

NETER, J.; WASSERMAN, W.; KUTNER M. H. **Applied linear statistical models: regression, analysis of variance and experimental designs**. 2. ed. Homewood: R. D. Irwin, 1985. p. 109-141.

ORDOVINI, A. F. et al. Seeds vigour tests in onion (*Allium cepa* L.) and their correlation with field emergence. In: INTERNATIONAL CONGRESS OF ISTA, 25., 1988, Pretoria. **Proceedings...** Zürich: ISTA, 1988. p.104.

PEARSON, E.S. The test of significance for the correlation coefficient. **Journal of the American Statistical Association**, New York, v. 26, p. 128-134, 1930.

RUMMEL, R. J. **Understanding correlation**. Honolulu: University of Hawaii, 1976. Disponível em: <<http://www.mega.nu:8080/ampp/rummel/uc>>. Acesso em: jul. 2006.

SANTANA, D. G.; RANAL, M. A. Linear correlation in experimental design models applied to seed germination. **Seed Science and Technology**, Zurich, v. 34, p. 241-247, 2006.

SANTOS, L. E. **Base de dados com testes em sementes de sorgo**. 1985. Monografia (Graduação) – Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, 1985.

SCHIELD, M. Correlation, determination and causality in introductory statistics. In: ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, 1995, Orlando. **Proceedings of the Section on Statistical Education**. Orlando: American Statistical Society, 1995. p. 1-6.

SOKAL, R. R.; ROHLF, F. J. **Biometry**: the principles of statistics in biological research. 3. ed. New York: Freedman and Company, 1995. 897 p.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**: a biometrical approach. 2. ed. New York: McGrawHill Book, 1980.

WELDON, K. L. A simplified introduction to correlation and regression. **Journal of Statistics Education**, Alexandria, v. 8, n. 3, 2000. Disponível em: <<http://www.amstat.org/publications/jse/secure/v8n3/weldon.cfm>>. Acesso em: nov. 2006.

ZAR, J. H. **Biostatistical analysis**. 4. ed. New York: Prentice Hall, 1999. 929 p.

6 ANEXO

Anexo A – Tabelas de dados

TABELA 1A. Germinabilidade (%) de sementes de Sorgo (*Sorghum bicolor*) oriundas de oito lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU (1985). Os lotes são identificados no cabeçalho e as condições de temperatura e tempo de permanência, em colunas iniciais, indentando as linhas.

Temperatura	Tempo Sob estresse	Lote 1	Lote 2	Lote 3	Lote 4	Lote 5	Lote 6	Lote 7	Lote 8
36°C	24h	45,0	46,3	47,3	50,8	46,3	50,0	51,5	42,5
	48h	45,0	46,8	46,8	50,8	43,8	48,3	51,0	44,0
	72h	43,0	46,8	44,5	44,0	45,3	45,3	45,0	41,3
	96h	38,8	39,8	38,5	42,5	43,5	45,8	46,0	38,3
	120h	40,3	39,3	35,5	40,3	39,8	41,0	43,8	30,3
38°C	24h	45,8	46,8	49,3	51,5	47,3	51,0	51,0	46,3
	48h	46,0	45,8	45,5	47,8	47,5	50,5	48,0	42,5
	72h	42,5	44,8	43,0	47,8	47,5	45,8	48,5	40,3
	96h	41,5	42,5	42,5	46,3	46,8	45,0	47,3	39,5
	120h	40,8	39,8	40,5	45,5	45,8	42,0	38,5	37,8
40°C	24h	42,0	43,5	44,3	47,0	43,0	48,5	47,0	42,0
	48h	41,0	42,8	44,5	45,8	42,5	47,3	46,0	40,0
	72h	36,5	33,3	44,0	43,5	39,3	45,5	44,0	36,5
	96h	33,5	29,0	31,3	37,0	38,3	36,0	35,5	24,0
	120h	36,5	33,3	45,0	41,8	41,0	41,0	41,0	33,8

TABELA 2A. Médias de velocidade de emergência (plântulas emersas.dia⁻¹) de sementes de Sorgo (*Sorghum bicolor*) oriundas de oito lotes de safras e qualidade fisiológica distintas, existentes no LASEM-UFU (1985). Os lotes são identificados no cabeçalho e os dias de leitura na colunas inicial, indentando as linhas do quinto ao 14° dia

Médias de 8 repetições de Velocidade de Emergência do 5° ao 14° dia								
	Lote1	Lote2	Lote3	Lote4	Lote5	Lote6	Lote7	Lote8
Dia 5	4,3	8,0	4,8	2,8	2,0	4,5	2,0	1,0
Dia 6	35,0	42,5	29,5	47,8	30,8	38,0	28,8	16,8
Dia 7	55,5	64,5	40,3	65,8	48,5	59,8	40,8	23,5
Dia 8	59,5	69,0	42,8	67,3	50,8	63,8	43,0	25,3
Dia 9	59,8	69,5	43,0	68,5	54,0	65,5	43,5	27,3
Dia 10	59,8	70,0	43,3	68,8	53,0	65,8	44,0	27,5
Dia 11	60,0	70,0	42,5	69,0	52,3	65,3	43,5	26,8
Dia 12	60,0	70,3	42,5	69,5	55,0	65,5	44,0	26,5
Dia 13	59,8	70,0	42,3	67,0	52,5	64,8	46,5	26,3
Dia 14	59,8	70,0	42,3	67,0	52,5	64,8	46,5	26,3

TABELA 3A. Variáveis de testes de velocidade média VM (plântulas emersas.dia⁻²) e respectivos resíduos VMR; de velocidade de emergência VE (plântulas emersas.dia⁻¹) e respectivos resíduos VER, germinação (%) Germ e respectivos resíduos (GermR) compostas cada uma de 60 dados incluindo repetições. A linha de cabeçalho identifica as variáveis, a coluna Z, é variável auxiliar numeradas de 1 a 60 relativa às repetições das médias.

Z	VM	VMR	VE	VER	Germ	GermR
1	0,03	0,00	1,74	0,04	100,00	3,67
2	0,02	-0,01	1,36	-0,34	96,00	-0,33
3	0,03	0,00	1,62	-0,08	96,00	-0,33
4	0,02	-0,01	0,86	-0,84	90,00	-6,33
5	0,04	0,01	2,08	0,38	98,00	1,67
6	0,05	0,02	2,55	0,85	98,00	1,67
7	0,03	0,00	1,78	0,01	100,00	0,00
8	0,03	0,00	1,73	-0,04	100,00	0,00
9	0,04	0,00	1,92	0,15	100,00	0,00
10	0,03	0,00	1,60	-0,17	100,00	0,00
11	0,03	0,00	1,62	-0,15	100,00	0,00
12	0,04	0,00	1,97	0,20	100,00	0,00
13	0,02	-0,01	1,45	-0,30	100,00	0,00
14	0,04	0,00	1,91	0,17	100,00	0,00
15	0,03	0,00	1,94	0,20	100,00	0,00
17	0,03	0,00	1,89	0,14	100,00	0,00
18	0,02	-0,01	1,13	-0,61	100,00	0,00
19	0,04	0,01	2,15	0,40	100,00	0,00
20	0,02	0,00	1,04	-0,10	100,00	1,33
21	0,02	0,00	1,23	0,10	100,00	1,33
22	0,03	0,01	1,43	0,29	100,00	1,33
23	0,01	-0,01	0,77	-0,37	96,00	-2,67
24	0,02	0,00	1,25	0,11	100,00	1,33
25	0,02	0,00	1,10	-0,03	96,00	-2,67
26	0,01	0,00	0,19	0,02	26,00	3,33
27	0,01	0,00	0,14	-0,03	16,00	-6,67
28	0,01	0,00	0,12	-0,05	18,00	-4,67
29	0,01	0,00	0,14	-0,03	16,00	-6,67
30	0,01	0,00	0,24	0,07	28,00	5,33
31	0,01	0,00	0,17	0,01	32,00	9,33
32	0,04	0,00	2,04	-0,19	100,00	0,00
33	0,05	0,01	2,74	0,51	100,00	0,00
34	0,03	-0,01	1,76	-0,48	100,00	0,00
35	0,04	0,00	2,17	-0,07	100,00	0,00
36	0,04	0,00	2,44	0,21	100,00	0,00
37	0,04	0,00	2,26	0,03	100,00	0,00
38	0,04	0,00	2,27	0,09	100,00	0,00
39	0,04	0,00	2,42	0,24	100,00	0,00
40	0,04	0,00	2,35	0,17	100,00	0,00
41	0,04	0,00	1,95	-0,23	100,00	0,00
42	0,05	0,01	2,39	0,21	100,00	0,00
43	0,03	-0,01	1,71	-0,48	100,00	0,00
44	0,04	0,00	2,30	0,04	100,00	0,00
45	0,04	0,00	2,21	-0,05	100,00	0,00
46	0,04	-0,01	1,96	-0,29	100,00	0,00
47	0,05	0,00	2,39	0,13	100,00	0,00
48	0,04	0,00	2,28	0,02	100,00	0,00
49	0,05	0,00	2,40	0,14	100,00	0,00
50	0,04	0,01	2,16	0,30	100,00	0,67
51	0,04	0,00	2,08	0,21	100,00	0,67

TABELA 3A - Cont

Z	VM	VMR	VE	VER	Germ	GermR
52	0,02	-0,01	1,35	-0,52	96,00	-3,33
53	0,04	0,00	2,03	0,16	100,00	0,67
54	0,03	-0,01	1,56	-0,30	100,00	0,67
55	0,04	0,00	2,01	0,15	100,00	0,67
56	0,02	0,00	0,86	-0,41	90,00	-1,67
57	0,02	0,00	2,95	1,68	92,00	0,33
58	0,02	0,00	1,02	-0,25	94,00	2,33
59	0,02	0,00	0,95	-0,32	94,00	2,33
60	0,02	0,00	0,90	-0,37	88,00	-3,67

TABELA 4A. Variáveis simuladas de $V1$ a $V10$ com inclinações de reta regressora geratriz maior que zero, e $Vzero$ com geratriz de inclinação nula, indicadas na linha de cabeçalho, Z é a seqüência incremental da variável auxiliar Z , de acordo com a quantidade n de elementos de cada uma.

Z	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	VZero
1	5,00	2,00	4,00	2,00	4,00	2,00	30	8	252,88	1,06	2,30
2	5,10	1,11	4,10	2,90	4,10	2,90	24	10	235,98	20,12	2,40
3	6,40	1,11	5,40	3,10	5,40	3,10	21	11,5	53,61	118,00	2,50
4	7,00	4,48	6,00	4,50	6,00	4,50	19	13	253,68	168,12	3,10
5	6,10	3,41	5,10	4,30	5,10	4,30	18	15	71,40	261,97	3,30
6	8,20	4,82	7,20	5,40	7,20	5,40	16	16,5	172,32	230,37	2,70
7	7,10	1,02	6,10	4,30	6,10	4,30	14	17,6	50,23	41,47	2,40
8	8,30	2,06	6,40	5,20	6,40	5,20	13	18	272,65	218,80	3,10
9	8,90	5,34			7,90	6,40	12	18	208,23	122,71	3,30
10	8,20	2,66			7,20	5,10	10	19	170,16	172,88	3,50
11	9,80	7,28			8,80	6,80	10,5	19	191,79	190,76	2,90
12	9,40	5,92			8,40	7,10	9,8	19	133,45	195,16	3,10
13	10,50	7,42			9,50	7,90	8,7	20	257,22	139,81	3,30
14	11,10	4,14			10,10	6,90	8,5	21	168,61	211,17	2,70
15	11,30	5,40			10,30	8,10	9,1	20,5	98,75	51,06	2,50
16	11,20	8,92			10,20	9,20	10,1	19	155,00	195,81	2,50
17	11,90	6,88			10,90	8,60	10,5	18,5	121,89	95,79	2,90
18	12,60	6,42			11,60	8,90	10,6	18,3	252,88	1,06	3,40
19	13,70	7,04			12,70	9,50	30	8	235,98	20,12	2,00
20	12,70	9,48			11,70	10,30	24	10	53,61	118,00	2,50
19	13,70	7,04			12,70	9,50	30	8	235,98	20,12	2,30
20	12,70	9,48			11,70	10,30	24	10	53,61	118,00	2,40

ÍNDICE

RESUMO	<i>i</i>
ABSTRACT	<i>ii</i>
1 INTRODUÇÃO	1
2 REVISÃO BIBLIOGRÁFICA	4
3 MATERIAL E MÉTODOS	10
3.1 Testes de germinação de sementes por envelhecimento acelerado	10
3.2 Teste de velocidade de emergência em campo	11
3.3 Efeito dos tratamentos na medida de correlação	12
3.4 Simulação de conjunto de dados para situações específicas	13
3.5 Variável auxiliar Z	13
4 RESULTADOS E DISCUSSÃO	16
4.1 Testes de germinação por envelhecimento acelerado	16
4.2 Testes de velocidade de emergência em campo	19
4.3 Entre variáveis de testes diferentes entre si	24
4.4 Efeitos dos tratamento	28
4.5 Dados simulados	34
4.6 Dispersões curvilíneas	41
Tendências curvilíneas das variáveis.....	41
Tendências curvilíneas na regressão $Y = f(X)$	42
4.7 Conclusão	44
5 REFERÊNCIAS	45
6 ANEXO	49