

Teste de Normalidade das Cargas Fatoriais

Até o presente momento os intervalos de confiança e os valores-p para o TIAF são sugeridos com base em métodos não paramétricos, que não exigem a normalidade das distribuições por amostragem das estimativas obtidas: utilizou-se o intervalo de confiança percentílico e o valor-p foi calculado diretamente das distribuições por amostragem empíricas das estimativas. Contudo, sabe-se que as alternativas não paramétricas por não exigirem o modelo gaussiano como pressuposto básico e sim distribuição livre, constituem em testes de significância com poder bem menor que os dos correspondentes paramétricos.

O poder de um teste é a probabilidade de rejeitar a hipótese nula, quando ela é realmente falsa e, por isso mesmo, deve ser rejeitada. Em consequência, ao admitir uma hipótese nula como falsa é mais provável que o pesquisador o rejeite pelo uso adequado de testes paramétricos dos que os não paramétricos.

Para minimizar a probabilidade de erros de decisão nos TIAF'S realizados, seria mais prudente “conectar” os intervalos de confiança e os valores-p construídos ao modelo de distribuição de probabilidades das estimativas investigadas. A curva normal por ser a mais freqüente distribuição ajustável à descrição de variáveis aleatórias e por otimizar o poder de um teste de hipótese será utilizada como prova de aderência aos dados.

O pesquisador sempre deve lembrar que os testes de significância são menos úteis em amostras pequenas (menos que 30) e muitos sensíveis em amostras grandes (que excedem a 1000 observações). Logo, o pesquisador deve usar também testes gráficos para avaliar o real grau de desvio da normalidade.

Para realizar o teste de aderência supra citado, isto é, o teste de normalidade aos dados, se aplicará, primeiro, o **Gráfico de Probabilidades** (*P-P-Probability Plots*), que indica graficamente o ajustamento das cargas fatoriais estimadas ao modelo gaussiano. Este tipo de gráfico indica no eixo das abscissas as freqüências relativas acumuladas observadas nas reamostragens e no eixo das ordenadas a função distribuição de probabilidades esperadas. A diagonal principal do gráfico representa um ajustamento perfeito das reamostragens a

função de distribuição de probabilidades. Quanto mais os pontos se afastam da diagonal, ou se distribuem segundo um determinado padrão, menor é o ajustamento das reamostragens à distribuição teórica em análise. Além do gráfico de ajustamento à normal, um gráfico do ajustamento dos resíduos ($y_{obs} - y_{est}$) será indicado conjuntamente. Se as reamostragens são perfeitamente normais, os resíduos distribuir-se-ão segundo uma faixa horizontal em torno do zero, sem denotar qualquer padrão de distribuição. Por simplificação, somente serão mencionados os resultados e interpretações verificadas, pois estes gráficos são bastante simples e fáceis de serem produzidos.

Para confirmar os resultados encontrados pelo teste do Gráfico de Probabilidades se recorrerá ao **Teste de Kolmogorov-Smirnov** (Siegel, Sidney; 1975). Kolmogorov-Smirnov desenvolveram um método, em geral mais poderoso que o do Qui-Quadrado (χ^2), para testar a aderência, em que a variável de teste é a maior diferença observada entre a função de distribuição acumulada do modelo e a da amostra.

A função de distribuição acumulada do modelo testado, ou função de repartição, dá as probabilidades acumuladas em cada ponto, ou seja, $F(x) = P(X \leq x)$. A função de distribuição acumulada da amostra corresponderá ao gráfico das freqüências relativas acumuladas (ogiva). Designa-se essa segunda função por $G(x)$. O teste consta simplesmente da verificação do valor:

$$d = \max | F(x) - G(x) |$$

e da comparação com um valor crítico tabelado em função do nível de significância (α) e do tamanho da amostra (n). Se d for maior que o valor crítico, rejeita-se H_0 . Os valores críticos para os níveis de significância usuais estão tabelados. Um método mais prático e rigoroso de se testar a hipótese nula é calcular o nível de credibilidade da hipótese nula, isto é, o seu valor-p. O valor-p pode ser então definido como a probabilidade condicional do valor crítico “ d ” ser igual ou maior que o valor efetivamente observado, dado a hipótese nula do

modelo como verdadeira, isto é , procura-se a probabilidade na cauda , além deste valor observado de “d”.

Para realizar o teste de aderência das cargas fatoriais em função das variáveis de cada problema se utilizará uma amostra de três bases de dados do elenco de bases aplicadas neste estudo (seis bases ao todo, ver capítulo 8): a base busca de fatores, a base creme dental e a base modo de vida.

A Base Busca de Fatores é fruto de uma pesquisa realizada junto a 212 servidores de reformatórios de menores infratores, que derem a sua opinião sobre o grau de influência de 23 atributos na prática de ato infracional praticada por adolescente.

A Base Creme Dental foi gerada pela aplicação de um questionário a 40 respondentes que deram a sua opinião sobre os benefícios que esperam da compra de um creme dental. Foram listados 6 benefícios que constituíram em variáveis do modelo.

A Base Modo de Vida surgiu de um *survey* envolvendo somente 25 respondentes sobre o comportamento de donas de casa na hora da compra de bens e serviços. Foram listados 7 comportamentos que constituíram em variáveis do modelo.

Vale lembrar que no capítulo 3, colocou-se que segundo Hair (1998) a análise fatorial não deve ser utilizada em amostras inferiores a 50 observações. Contudo, para ilustrar justamente os problemas que incorrem na quebra deste importante pressuposto utilizou base de dados inferiores a aquele limite.

Os gráficos de probabilidades não serão ilustrados no texto do trabalho, mas seus os resultados serão comunicados nas análises e interpretações realizadas em cada base de dados.

6.1

Teste de Normalidade da “Base Busca de Fatores”

6.1.1

Teste de Normalidade das Variáveis no Método *Bootstrap*

Tabela 4: Caso 1: Busca de Fatores
Resultado do Teste de Normalidade para o Caso 1 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,50	Normal	0,10	Normal
V2	0,36	Normal	0,58	Normal
V3	0,11	Normal	0,98	Normal
V4	0,95	Normal	0,98	Normal
V5	0,33	Normal	0,89	Normal
V6	0,39	Normal	0,69	Normal
V7	0,99	Normal	0,54	Normal
V8	0,39	Normal	0,97	Normal
V9	0,51	Normal	0,77	Normal
V10	0,04	Nnormal	0,96	Normal
V11	0,05	Nnormal	0,99	Nnormal
V12	0,60	Normal	0,15	Normal
V13	0,51	Normal	0,79	Normal
V14	0,33	Normal	0,79	Normal
V15	0,09	Normal	0,75	Normal
V16	0,25	Normal	0,79	Normal
V17	0,29	Normal	0,85	Normal
V18	0,85	Normal	0,60	Normal

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V19	0,98	Normal	0,96	Normal
V20	0,76	Normal	0,64	Normal
V21	0,98	Normal	0,33	Normal
V22	0,49	Normal	0,68	Normal
V23	0,48	Normal	0,95	Normal

Nota: Teste ao nível de 5% de significância e Normal = aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *bootstrap*, fator 1, o gráfico de ajustamento à normal indica que a maioria das variáveis se ajusta ao modelo normal, pois se observa em muitos atributos um afastamento cíclico fraco em relação à diagonal. Exceções são constatadas nas variáveis $V_{10}, V_{11}, V_{14}, V_{16}, V_{17}, V_{18}, V_{19}$ e V_{23} . O gráfico do ajustamento dos resíduos revela um tímido comportamento cíclico em torno de zero para a maioria das variáveis do Caso 1, denotando fraco afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

No mesmo método, fator 2, o gráfico de ajustamento à normal indica também que a maioria das variáveis se ajusta ao modelo normal, pois se observa, igualmente ao fator 1, que em muitos atributos existe um certo afastamento cíclico fraco em relação à diagonal. Exceções são constatadas nas variáveis $V_1, V_2, V_{10}, V_{11}, V_{12}, V_{14}, V_{16}, V_{17}, V_{21}$ e V_{23} . O gráfico do ajustamento dos resíduos revela um tímido comportamento cíclico em torno de zero para a maioria das variáveis do Caso 1, denotando fraco afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

O teste de normalidade da Tabela 4 confirma o que o gráfico de probabilidades havia indicado: as estimativas das cargas fatoriais obtidas pelo método *bootstrap* são em geral normalmente distribuídas, o que justificaria a

construção alternativa de um intervalo de confiança *bootstrap* normal, isto tanto para o fator 1 quanto para o fator 2. Contudo, como também existem variáveis que não se ajustam à normalidade, adotando-se um critério de uniformidade e comparabilidade, os intervalos de confiança serão construídos seguindo o procedimento percentílico.

6.1.2

Teste de Normalidade das Variáveis no Método *Jackknife*

Tabela 5: Caso 1: Busca de Fatores
Resultado do Teste de Normalidade para o Caso 1 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,00	Nnormal	0,00	Nnormal
V2	0,00	Nnormal	0,00	Nnormal
V3	0,00	Nnormal	0,00	Nnormal
V4	0,00	Nnormal	0,02	Nnormal
V5	0,00	Nnormal	0,00	Nnormal
V6	0,00	Nnormal	0,00	Nnormal
V7	0,00	Nnormal	0,00	Nnormal
V8	0,00	Nnormal	0,00	Nnormal
V9	0,00	Nnormal	0,00	Nnormal
V10	0,00	Nnormal	0,00	Nnormal
V11	0,00	Nnormal	0,03	Nnormal
V12	0,00	Nnormal	0,00	Nnormal

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V13	0,00	Nnormal	0,00	Nnormal
V14	0,00	Nnormal	0,00	Nnormal
V15	0,00	Nnormal	0,00	Nnormal
V16	0,00	Nnormal	0,00	Nnormal
V17	0,00	Nnormal	0,00	Nnormal
V18	0,00	Nnormal	0,00	Nnormal
V19	0,00	Nnormal	0,00	Nnormal
V20	0,01	Nnormal	0,00	Nnormal
V21	0,00	Nnormal	0,00	Nnormal
V22	0,00	Nnormal	0,00	Nnormal
V23	0,00	Nnormal	0,00	Nnormal

Nota: Teste ao nível de 5% de significância e Normal = aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *jackknife*, fator 1, o gráfico de ajustamento à normal indica que a grande maioria das variáveis se não se ajusta ao modelo normal, pois se observa em muitos atributos um afastamento cíclico acentuado em relação à diagonal. Exceções são constatadas nas variáveis V_8 e V_{20} . O gráfico do ajustamento dos resíduos revela um fraco comportamento aleatório para a maioria das variáveis do Caso 1, isto é, um acentuado padrão de distribuição, denotando algum afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

No mesmo método, fator 2, o gráfico de ajustamento à normal indica também que a maioria das variáveis não se ajusta ao modelo normal, pois se observa, igualmente ao fator 1, que em muitos atributos existe um certo afastamento cíclico em relação à diagonal. O gráfico do ajustamento dos resíduos, como no fator 1, revela um certo padrão de comportamento dos afastamentos para a maioria das variáveis do Caso 1, denotando algum afastamento em relação à

normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

Pelo que indica o Teste de Kolmogorov-Smirnov as estimativas das cargas fatoriais obtidas pelo método *jackknife* não se aderem ao modelo normal, conclusão análoga à observada pela investigação da normalidade pelos Gráficos de Probabilidades, o que justificaria a adoção de intervalos de confiança *jackknife* percentílico, isto tanto para o fator 1, quanto para o fator 2.

6.2

Teste de Normalidade da “Base Creme Dental”

6.2.1

Teste de Normalidade das Variáveis no Método *Bootstrap*

Tabela 6: Caso 2 : Creme Dental – Resultado do Teste de Normalidade para o Caso 2 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,00	Nnormal	0,00	Nnormal
V2	0,00	Nnormal	0,00	Nnormal
V3	0,00	Nnormal	0,00	Nnormal
V4	0,00	Nnormal	0,00	Nnormal
V5	0,00	Nnormal	0,00	Nnormal
V6	0,00	Nnormal	0,00	Nnormal

Nota: Teste ao nível de 5% de significância e Normal = aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

O Teste de Kolmogorov-Smirnov revela (ver Tabela 6), ao nível de 5% de significância, que as estimativas geradas pelo método *bootstrap* não seguem o modelo normal de probabilidade. Isto tanto para o fator 1 quanto para o fator 2. Os Gráficos de Probabilidades revelam para a maioria das variáveis um afastamento

em relação à diagonal, o que confirma os resultados do teste de aderência de utilizado (*K-S*). Os gráficos de afastamento de resíduos não apresentam um comportamento horizontal em torno de zero, indicando também um afastamento das estimativas da normalidade.

6.2.2

Teste de Normalidade das Variáveis no Método *Jackknife*

Tabela 7: Resultado do Teste de Normalidade para o Caso 2 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,26	Normal	0,46	Normal
V2	0,85	Normal	0,08	Normal
V3	0,05	Nnormal	0,41	Normal
V4	0,41	Normal	0,36	Normal
V5	0,05	Nnormal	0,53	Normal
V6	0,50	Normal	0,03	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

Pela observação do Gráfico de Probabilidades se observa que para a maioria das variáveis existe um ajustamento razoável dos valores estimados à curva normal, estando os pontos mais próximos da diagonal. Já os gráficos do ajustamento dos resíduos, denotam em algumas variáveis um comportamento cíclico em torno de zero, denotando algum afastamento em relação à normal, isto é menos evidente nos gráficos de ajustamento em que os pontos estão mais próximos da diagonal. Estes resultados coincidem com os testes de hipóteses de aderência obtidos através do Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

O *TIAF* baseado na normalidade das distribuições por amostragem das estimativas somente teria sentido no método *jackknife*, para o Caso Creme Dental naquelas variáveis onde a normalidade é garantida.

Os intervalos de confiança construídos para as estimativas geradas, tanto no *bootstrap* quanto no *jackknife*, continuarão sendo os percentílicos, para manter a uniformidade de procedimento em todas as variáveis da base e facilitar a análise de comparativa.

6.3

Teste de Normalidade da Base “Modo de Vida”

6.3.1

Teste de Normalidade das Variáveis no Método *Bootstrap*

Tabela 8: Caso 3 : Modo de Vida
Resultado do Teste de Normalidade para o Caso 3 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,000	Nnormal	0,000	Nnormal
V2	0,000	Nnormal	0,000	Nnormal
V3	0,000	Nnormal	0,000	Nnormal
V4	0,000	Nnormal	0,000	Nnormal
V5	0,000	Nnormal	0,000	Nnormal
V6	0,000	Nnormal	0,000	Nnormal
V7	0,000	Nnormal	0,000	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

Pela observação do Gráfico de Probabilidades se observa que para todas as variáveis no fator 1 o ajuste à curva normal não pode ser considerado, uma vez

evidenciado pontos das variáveis distanciados da diagonal. Os gráficos do ajustamento dos resíduos, no fator 1, confirmam os resultados indicados pelos gráficos de ajustamento à normal: para a grande maioria das variáveis existe um certo comportamento tendência e cíclico para o comportamento dos resíduos, se distanciando do que seria esperado: uma nuvem aleatória e espaçada de pontos. Estes resultados coincidem com os testes de hipóteses de aderência obtidos através do Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

No fator 2, constatam-se resultados análogos ao fator 1, com a exceção para a variável 3, que diverge do resultado de teste de aderência de Kolmogorov-Smirnov (distribuição não normal, ao nível 5 % de significância), parecendo tanto pelo gráfico de ajustamento à normal quanto no gráfico do Ajustamento dos resíduos haver ajustamento à distribuição gaussiana.

6.3.2

Teste de Normalidade das Variáveis no Método *Jackknife*

Tabela 9: Caso 3: Modo de Vida
Resultado do Teste de Normalidade Para o Caso 3 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,97	Normal	0,15	Normal
V2	0,99	Normal	0,00	Nnormal
V3	0,74	Normal	0,00	NNormal
V4	0,59	Normal	0,01	Nnormal
V5	0,76	Normal	0,03	Nnormal
V6	1,00	Normal	0,00	Nnormal
V7	0,76	Normal	0,04	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal = rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *jackknife*, no fator 1, observa-se do Gráfico de Probabilidades que a maioria das variáveis se ajusta à curva normal, uma vez evidenciado pontos das variáveis próximos à diagonal (com exceção para a variável V_1). Os gráficos do ajustamento dos resíduos, no fator 1, confirmam os resultados indicados pelos gráficos de ajustamento à normal: para a grande maioria das variáveis existe um certo comportamento aleatório e espaçado dos pontos. Estes resultados praticamente coincidem com os testes de hipóteses de aderência de Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

No fator 2, constatam-se resultados opostos ao fator 1: a maioria das variáveis não parece ser ajustar à curva normal, com exceção para a variável V_1 , que apresenta no gráfico de ajustamento à normal pontos toleravelmente próximos à diagonal e comportamento razoavelmente não tendencial e aleatório dos pontos dos resíduos.

O TIAF baseado na normalidade das distribuições por amostragem das estimativas somente teria sentido no método *jackknife* para o Caso 3, no fator 1, mas vale lembrar que a qualidade da distribuição por amostragem das cargas fatoriais nas componentes principais 1 e 2 na técnica *jackknife*, neste estudo de caso, é inferior ao do método *bootstrap*, sendo então recomendável trabalhar com os intervalos de confiança do *bootstrap* ao utilizar o TIAF, ao invés do intervalo de confiança *jackknife*, isto é, utilizar-se do intervalo de confiança percentílico, uma vez que o modelo gaussiano não é garantido à distribuição das estimativas *bootstrap*. Um outro motivo que se leva a adotar o intervalo de confiança percentílico é que não existe uma uniformidade no ajuste à normal entre as variáveis, entre os fatores e entre os estudos de caso, o que dificultaria a análise comparativa inferencial dos resultados e a homogeneidade do uso de modelos.

6.4

Procedimentos Metodológicos a serem seguidos com base nos Testes de Normalidades

Verificou-se pelos testes de normalidades realizados que não existe uma regularidade e uniformidade na aderência das distribuições amostrais das cargas fatoriais ao modelo normal: o resultado foi específico para certo estudo de caso, para certo fator e para certa variável. Portanto, para se manter uma uniformidade na construção dos intervalos de confiança para as diversas variáveis e para os diversos problemas, será utilizada a estimação percentílica.