AULAS 08 E 09
Distribuição de probabilidade normal

Ernesto F. L. Amaral

10 e 12 de setembro de 2013
Metodologia de Pesquisa (DCP 854B)

Fonte:
Triola, Mario F. 2008. "Introdução à estatística". 10 ª ed. Rio de Janeiro: LTC. Capítulo 6 (pp.192-249).

ESTRUTURA DA AULA

– A distribuição normal padrão

– Aplicações da distribuição normal

– Distribuições amostrais e estimadores

– O Teorema Central do Limite

– A normal como aproximação da binomial

– Determinação de normalidade

A DISTRIBUIÇÃO NORMAL PADRÃO

LEMBREMOS

– **Variável aleatória** é uma variável que tem um único valor numérico, determinado pelo acaso, para cada resultado de um experimento.

– **Distribuição de probabilidade** descreve a probabilidade de cada valor da variável aleatória.

– **Variável aleatória discreta** tem uma quantidade finita de valores ou uma quantidade enumerável de valores.

– **Variável aleatória contínua** tem infinitos valores, sem saltos ou interrupções.

GRÁFICOS DAS DISTRIBUIÇÕES

– O **histograma de probabilidade** é um gráfico de uma distribuição de probabilidade discreta.

– A **curva de densidade** é um gráfico de uma distribuição de probabilidade contínua, em que:

  – A área total sob a curva tem que ser igual a 1.

  – Cada ponto na curva tem que ter uma altura vertical que é 0 ou maior, não estando abaixo do eixo *x*.

DISTRIBUIÇÕES DE PROBABILIDADE

– Como a **área total** sob o gráfico de uma distribuição de probabilidade é igual a 1, há correspondência entre área e probabilidade (ou frequência relativa).

– Isto possibilita **calcular probabilidades** com utilização das áreas.

– **É importante:**
  – Desenvolver a habilidade para determinar áreas correspondentes a várias regiões sob o gráfico da distribuição.
  – Encontrar valores da variável *z* que correspondem a áreas sob o gráfico.

DISTRIBUIÇÕES UNIFORMES

– Na **distribuição uniforme**, uma variável aleatória contínua apresenta valores de probabilidade que se espalham uniformemente sobre a faixa de valores possíveis.

– Em geral, a área de um retângulo se torna 1 quando fazemos sua altura igual ao valor de 1/amplitude.

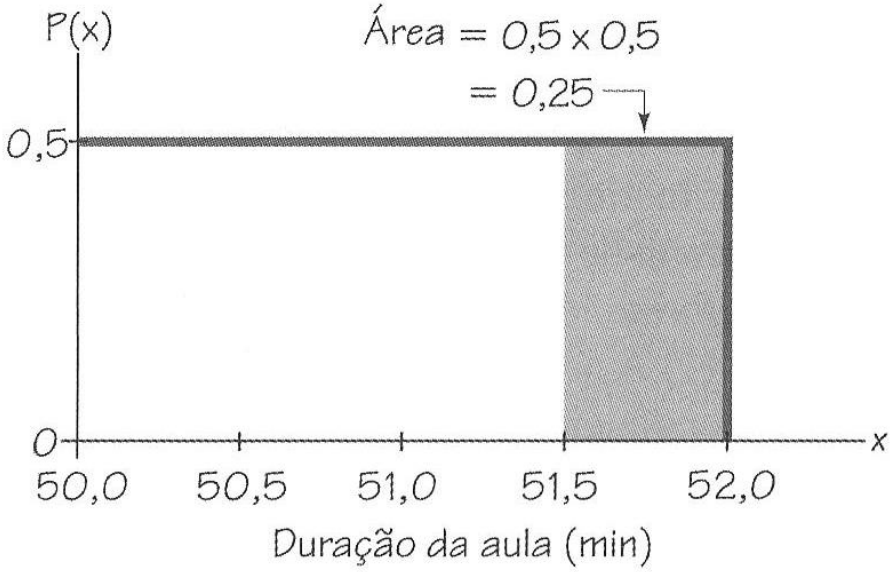


FIGURA 6-3 Usando a Área para Achar a Probabilidade

DISTRIBUIÇÃO NORMAL

– As **distribuições normais** são importantes, porque elas ocorrem frequentemente em situações reais e desempenham papel importante nos métodos de inferência estatística.

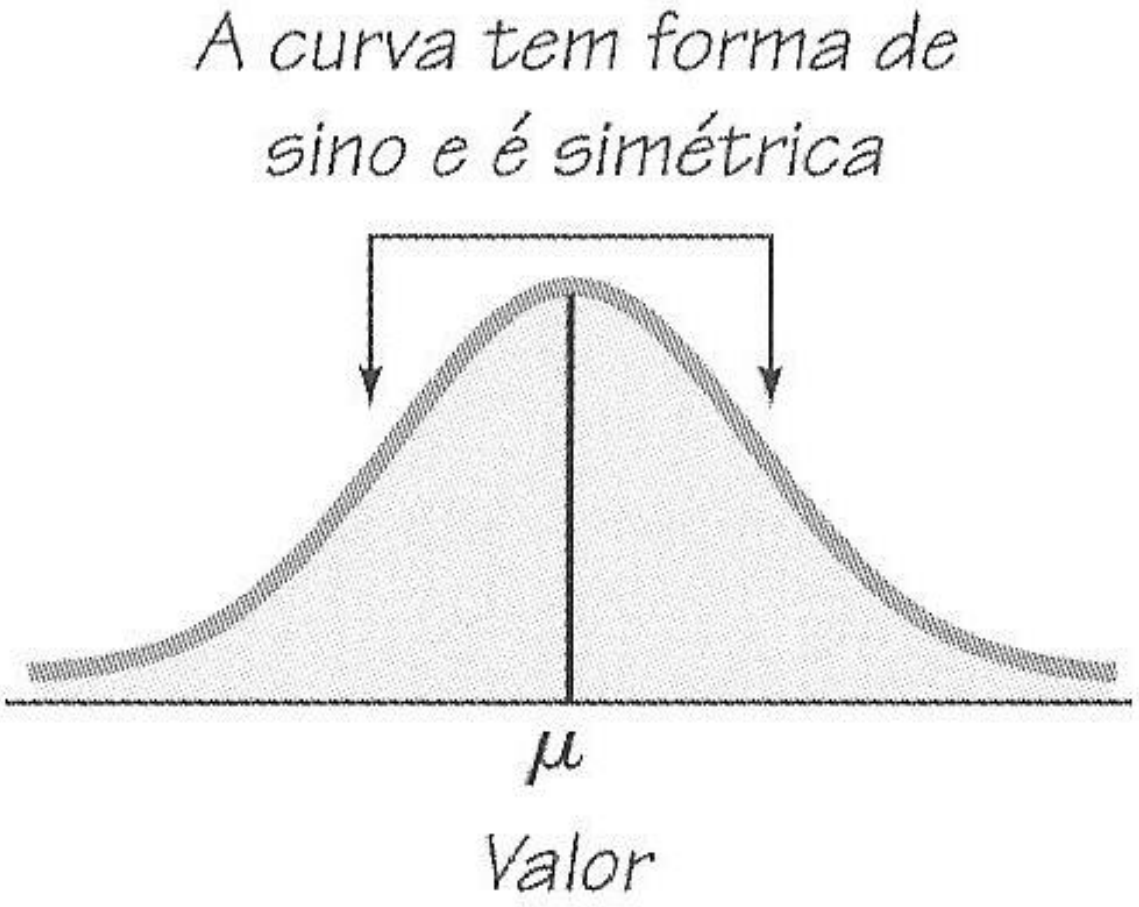
– A distribuição é normal se uma variável aleatória contínua tem uma distribuição com um gráfico simétrico em forma de sino.

– Qualquer distribuição normal é determinada pela média (μ) e desvio padrão (σ):

$$y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

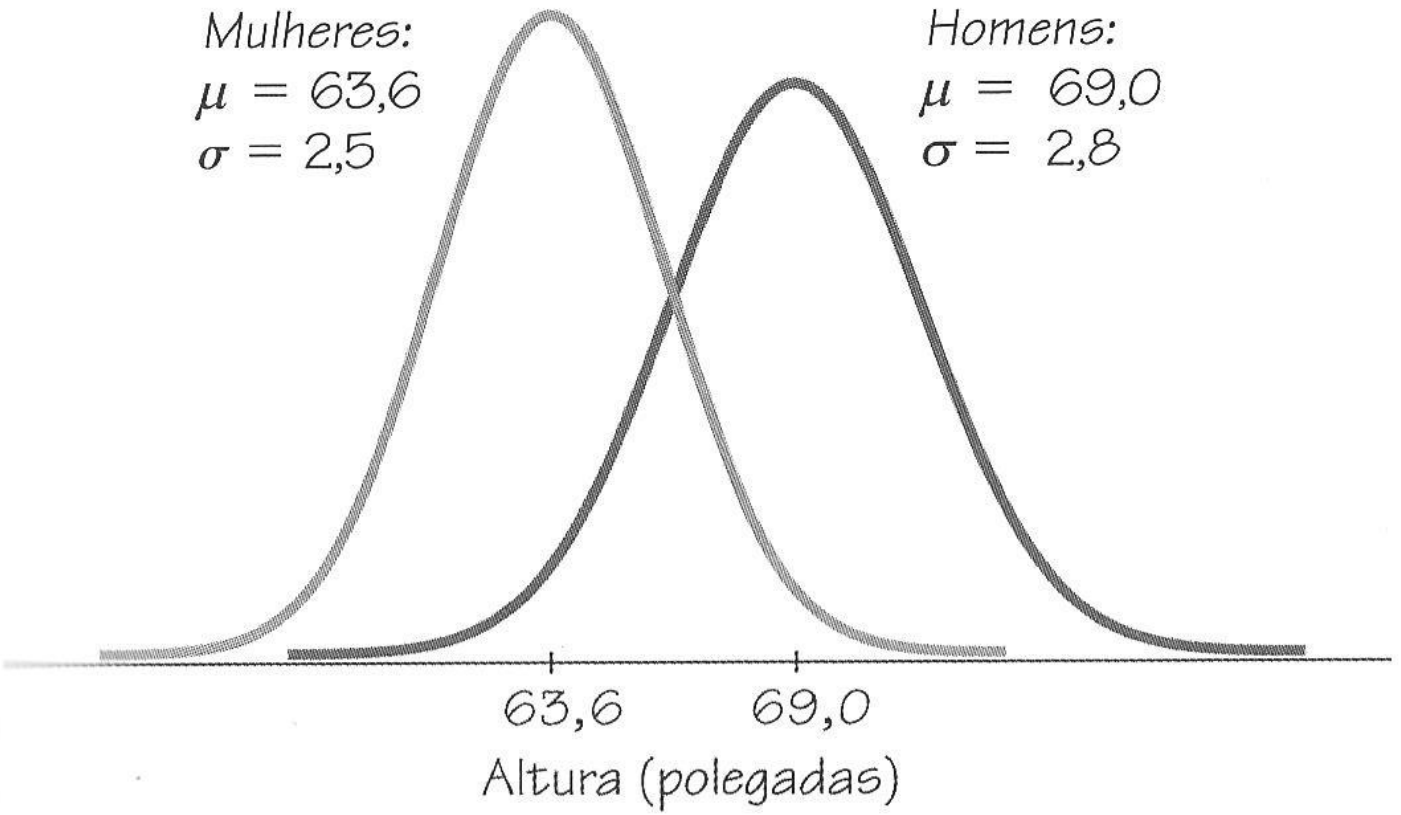
GRÁFICO DA DISTRIBUIÇÃO NORMAL

– De posse de valores específicos para μ e σ, podemos fazer o seguinte gráfico da distribuição normal.



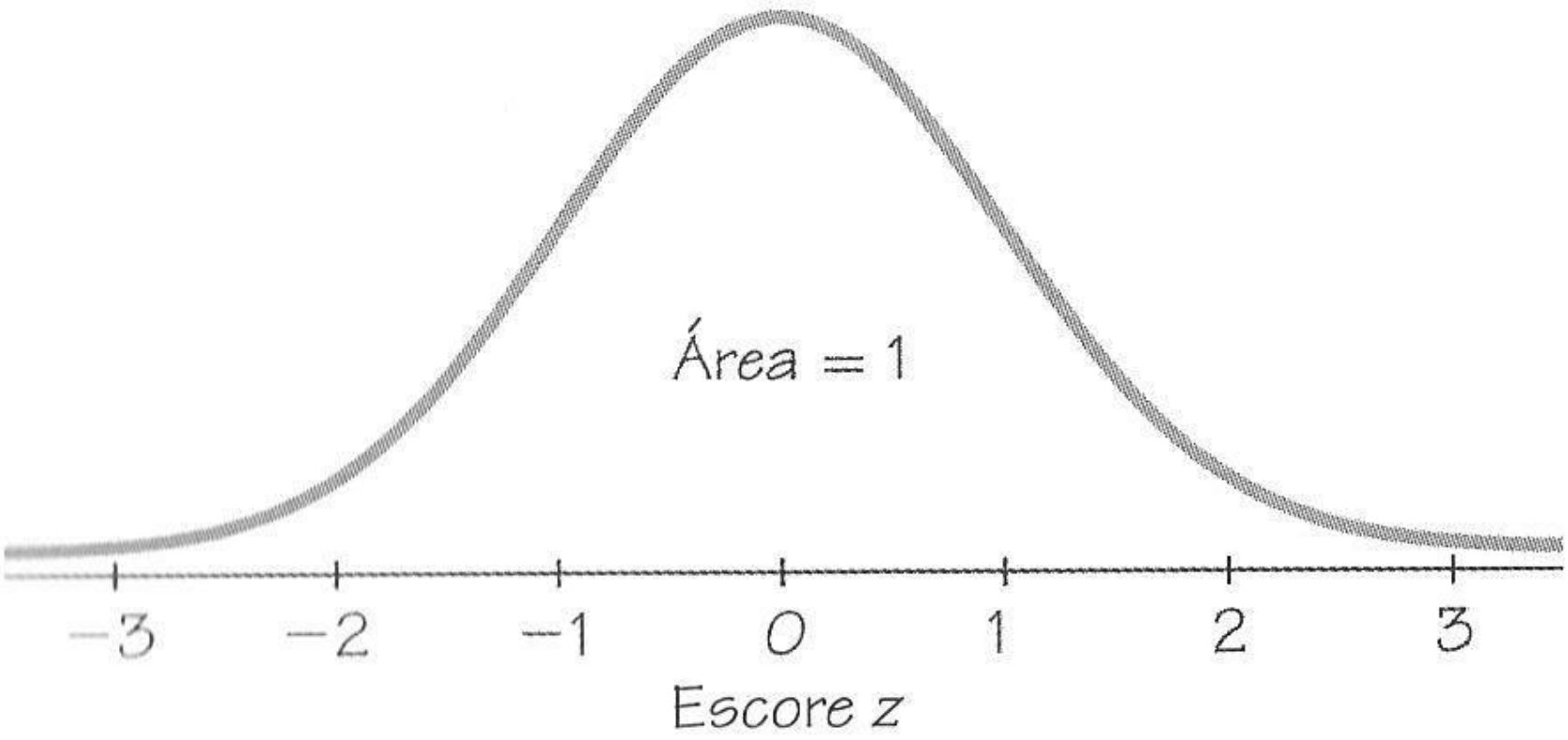
VARIAÇÃO NAS DISTRIBUIÇÕES NORMAIS

– Há muitas distribuições normais diferentes, dependendo de dois parâmetros: a média populacional (μ) e o desvio padrão populacional (σ).



DISTRIBUIÇÃO NORMAL PADRÃO

– A distribuição normal padrão é uma distribuição de probabilidade normal com média (μ) igual a 0 e desvio padrão (σ) igual a 1.



ENCONTRE PROBABILIDADES A PARTIR DE ESCORES *z*

– Usando a tabela das páginas 618-619, é possível achar áreas (ou probabilidades) para muitas regiões diferentes.
  – Se refere à distribuição normal padrão (μ=0 e σ=1).
  – Possui resultados para escores *z* negativos e positivos.

– **Escore *z*:** distância na escala horizontal da distribuição normal padrão:
  – Parte inteira e decimal: coluna à esquerda da tabela.
  – Parte do centésimo: linha no topo da tabela.

– **Área:** região sob a curva (valores no corpo da tabela).

Escores *z* POSITIVOS

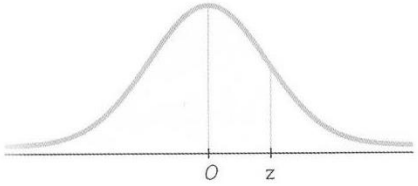


TABELA A.2 Área Acumulada à ESQUERDA *(continuação)*

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495 *	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949 *	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

– **Área acumulada à esquerda** de z=1,13 é igual a 0,8708.

– Há uma probabilidade de 0,8708 de **selecionarmos aleatoriamente** um escore *z* menor que 1,13.

– **No Stata:**

di normal(1.13)

DIFERENTES COMANDOS NO STATA (*help functions*)

normal(z)

```
Domain:       -8e+307 to 8e+307
Range:        0 to 1
Description:  returns the cumulative standard normal distribution.
```

normalden(z)

```
Domain:       -8e+307 to 8e+307
Range:        0 to .39894 ...
Description:  returns the standard normal density.
```

normalden(z,s)

```
Domain z:     -8e+307 to 8e+307
Domain s:     1e-308 to 8e+307
Range:        0 to 8e+307
Description:  returns the rescaled standard normal density.
              normalden(z,1) = normalden(x) and normalden(z,s) = normalden(z)/s.
```

normalden(x,m,s)

```
Domain x:     -8e+307 to 8e_307
Domain m:     -8e+307 to 8e_307
Domain s:     1e-308 to 8e+307
Range:        0 to 8e+307
Description:  returns the normal density with mean m and standard deviation s.
                  normalden(x,0,1) = normalden(x)
                  normalden(x,m,s) = normalden((x-m)/s)/s
```

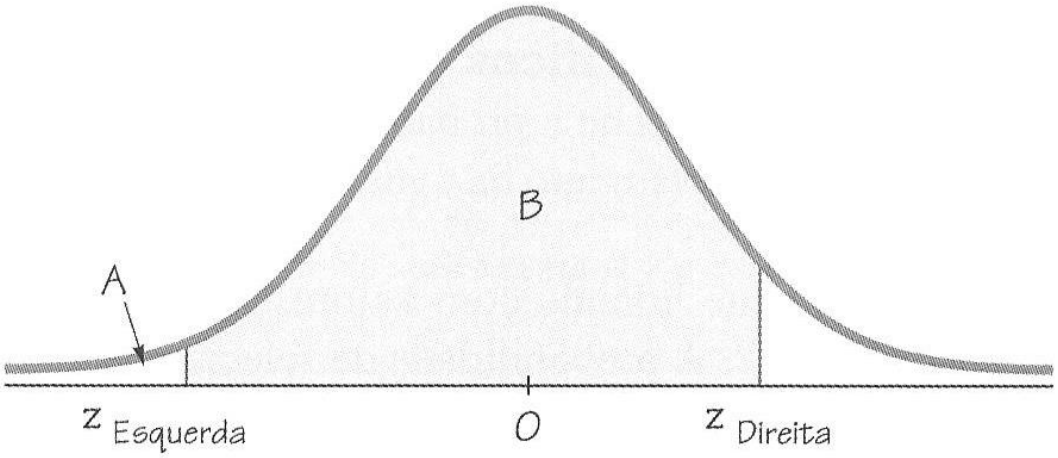
ENCONTRANDO PROBABILIDADES

– Para encontrar o valor da probabilidade, primeiro desenhe um gráfico, sombreie a região desejada e pense em uma maneira de achar a área correspondente.

– *P(a<z<b)*: probabilidade do escore *z* estar entre *a* e *b*.

– *P(z>a)*: probabilidade do escore *z* ser maior que *a*.

– *P(z<a)*: probabilidade do escore *z* ser menor que *a*.



Área sombreada B = (áreas A e B combinadas) − (área A)
= (área da Tabela A-2 usando $z_{Direita}$) − área da Tabela A-2 usando $z_{Esquerda}$)

PROBABILIDADE DE VALOR EXATO É IGUAL A ZERO

– Com uma distribuição de probabilidade contínua, a probabilidade de se obter qualquer valor único exato é zero:

$$P(z = a) = 0$$

– Por exemplo, há uma probabilidade 0 de selecionarmos aleatoriamente uma pessoa com altura exatamente igual a 1,763947 metros.

– Um ponto isolado na escala horizontal é representado por uma linha vertical, e não uma área sob a curva:

$$P(a \leq z \leq b) = P(a < z < b).$$

– A probabilidade de se obter um valor **no máximo igual *b*** é igual à probabilidade de se obter um valor **menor que *b***.

– É importante saber interpretar frases-chave: no máximo, pelo menos, mais do que, não mais do que...

ENCONTRE ESCORES *z* A PARTIR DE ÁREAS

– Encontramos áreas (probabilidades), a partir de escores *z*.

– Agora vamos encontrar o escore *z* (distâncias ao longo da escala horizontal) a partir da área (regiões sob a curva).

– **Desenhe** uma curva em forma de sino e identifique a região sob a curva que corresponde à probabilidade dada.

– **Na tabela** (pp.618-619), use a área acumulada à esquerda, localize a probabilidade mais próxima e identifique escore *z*.

– **No Stata:**

  – *z* (1,645) que separa 95% inferiores dos demais:

  di invnormal(.95)

  – *z* (1,96) que separa 97,5% inferiores dos demais:

  di invnormal(.975)

APLICAÇÕES DA DISTRIBUIÇÃO NORMAL

APLICAÇÕES DA DISTRIBUIÇÃO NORMAL

– Vamos tratar de métodos para trabalhar com distribuições normais que não são padrões (ou $\mu \neq 0$, ou $\sigma \neq 1$, ou ambos).

– Podemos fazer conversão para transformar qualquer distribuição normal em distribuição normal padrão.

– Se convertermos valores para escores padronizados, os procedimentos para trabalhar com distribuições normais serão os mesmos daqueles usados para distribuição normal padrão:

$$z = (x - \mu) / \sigma$$

EQUIVALÊNCIA ENTRE NORMAL E NORMAL PADRÃO

– A área em qualquer distribuição normal limitada por um escore *x* é igual à área limitada pelo escore *z* equivalente na distribuição normal padrão.

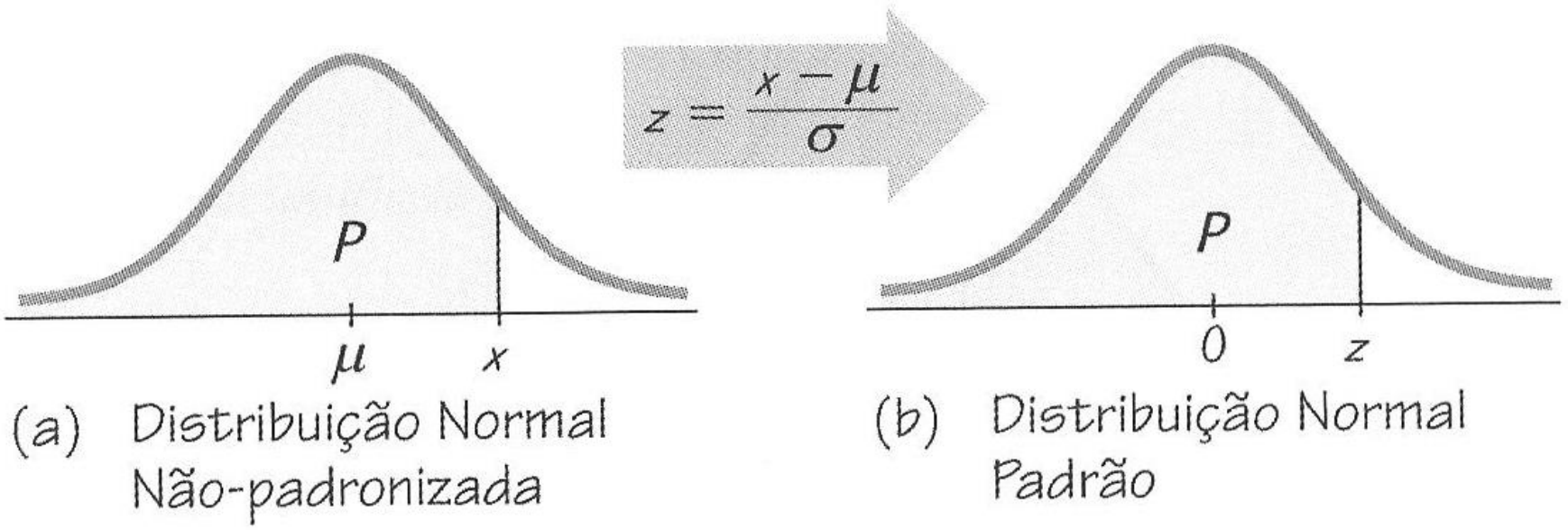


FIGURA 6-12 Conversão de uma Distribuição Normal Não-padronizada para a Distribuição Normal Padrão

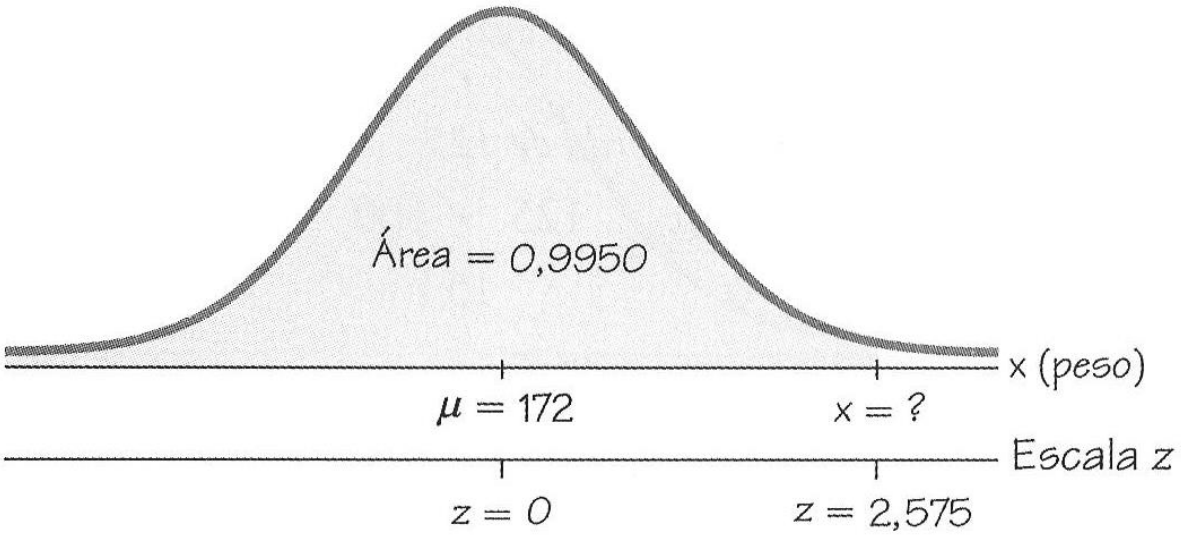
ENCONTRE VALORES *x* A PARTIR DE ÁREAS

– **Desenhe** uma curva em forma de sino e identifique a região sob a curva que corresponde à probabilidade dada.

– Encontre o escore *z* correspondente à área acumulada à esquerda de *x*.

– Sendo $\mu = 172$ libras e $\sigma = 29$ libras, use μ, σ e *z* para calcular *x*, com esta fórmula $x = \mu + (z*\sigma)$.

– Com a área (0,9950), encontramos o escore *z* (2,575) e depois aplicamos a fórmula acima: x = 246,675 libras.



DISTRIBUIÇÕES AMOSTRAIS E ESTIMADORES

DISTRIBUIÇÕES AMOSTRAIS E ESTIMADORES

– **Distribuição amostral de uma estatística** (média amostral) é a distribuição de todos valores da estatística, quando todas amostras possíveis de mesmo tamanho *n* tiverem sido extraídas da mesma população.

– Algumas estatísticas (proporção e média) são boas para estimação de valores de parâmetros populacionais.

– A distribuição amostral de uma estatística é geralmente representada por uma tabela, histograma de probabilidade ou fórmula.

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

– **Distribuição amostral da proporção** é a distribuição de probabilidade das proporções amostrais, com todas amostras tendo o mesmo tamanho amostral *n* tiradas de uma mesma população.

– Proporções amostrais tendem a atingir o alvo da proporção populacional:

  – Todas proporções amostrais possíveis têm uma média igual à proporção populacional.

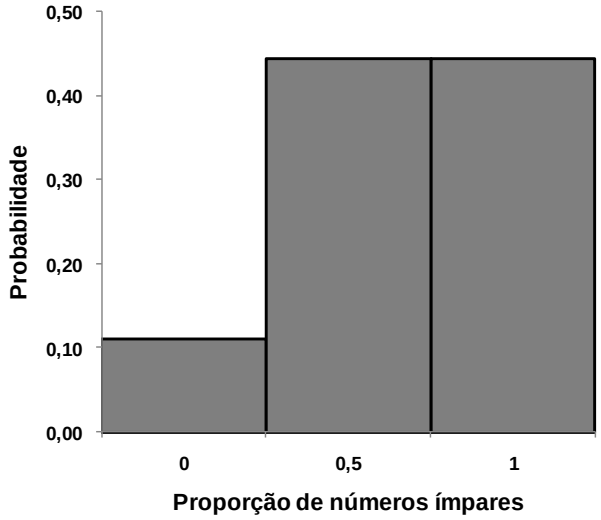
– Sob certas condições, a distribuição das proporções amostrais pode ser aproximada por uma distribuição normal.

DISTRIBUIÇÃO AMOSTRAL DE PROPORÇÕES DE NÚMEROS ÍMPARES, CONSIDERANDO AMOSTRA DE n=2

- População: 1; 2; 5
- Proporção populacional: 2/3 = 0,67

Amostra tamanho n=2	Proporção números ímpares	Probabilidade
1; 1	1	1/9
1; 2	0,5	1/9
1; 5	1	1/9
2; 1	0,5	1/9
2; 2	0	1/9
2; 5	0,5	1/9
5; 1	1	1/9
5; 2	0,5	1/9
5; 5	1	1/9

Proporção de números ímpares	Probabilidade
0	1/9 = 0,11
0,5	4/9 = 0,44
1	4/9 = 0,44



- Em geral, a distribuição amostral de proporções tem média igual à proporção populacional (proporções amostrais apontam para proporção populacional). Neste caso, média das proporções é de 6/9 = 2/3 = 0,67.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

– **Distribuição amostral da média** é a distribuição de probabilidade das médias amostrais, com todas amostras tendo o mesmo tamanho amostral *n* tiradas de uma mesma população.

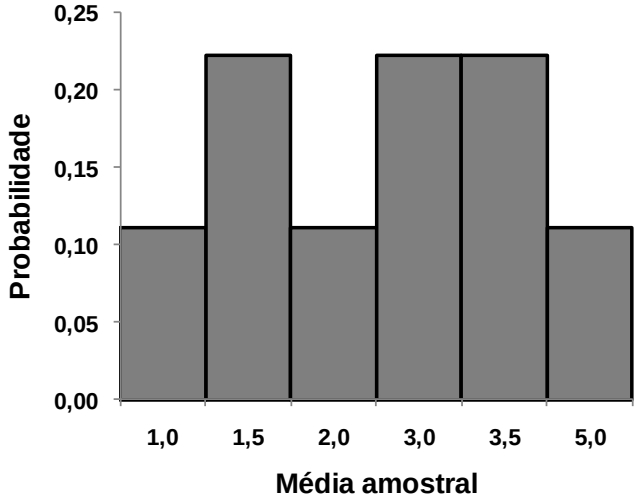
– A distribuição amostral da média é tipicamente representada como uma distribuição de probabilidade no formato de tabela, histograma de probabilidade ou fórmula.

DISTRIBUIÇÃO AMOSTRAL DE MÉDIAS, CONSIDERANDO AMOSTRA DE n=2

- População: 1; 2; 5
- Média populacional: 8/3 = 2,7

Amostra x1	Amostra x2	Média de x1 e x2	Probabilidade
1	1	1,0	1/9
1	2	1,5	1/9
1	5	3,0	1/9
2	1	1,5	1/9
2	2	2,0	1/9
2	5	3,5	1/9
5	1	3,0	1/9
5	2	3,5	1/9
5	5	5,0	1/9

Média de x1 e x2	Probabilidade
1,0	1/9 = 0,11
1,5	2/9 = 0,22
2,0	1/9 = 0,11
3,0	2/9 = 0,22
3,5	2/9 = 0,22
5,0	1/9 = 0,11



- Em geral, a distribuição das médias amostrais tem média igual à média populacional (médias amostrais apontam para média populacional). Neste caso, a média das médias é de 24/9 = 2,7.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA (cont.)

- Para um **tamanho amostral fixo**, a média de todas as possíveis médias amostrais é igual à média da população.
- O **valor de uma estatística** (por exemplo, média amostral) depende dos valores incluídos na amostra e, em geral, varia de uma amostra para outra.
  - Essa variabilidade de uma estatística é chamada de variabilidade amostral.
- À medida que o **tamanho da amostra** aumenta, a distribuição amostral das médias amostrais tende a se tornar uma distribuição normal.
  - Por isso, em muitos casos usaremos a média amostral com o propósito de fazer alguma inferência sobre a média populacional.

ESTATÍSTICAS COM ESTIMADORES DE PARÂMETROS

Tabela 6-7 Distribuições Amostrais de Estatísticas (para Amostras de Tamanho 2 Tiradas com Reposição da População 1, 2, 5)

Amostra	Média $\bar{x}$	Mediana	Amplitude	Variância s^2	Desvio Padrão s	Proporção de Números Ímpares	Probabilidade
1, 1	1,0	1,0	0	0,0	0,000	1	1/9
1, 2	1,5	1,5	1	0,5	0,707	0,5	1/9
1, 5	3,0	3,0	4	8,0	2,828	1	1/9
2, 1	1,5	1,5	1	0,5	0,707	0,5	1/9
2, 2	2,0	2,0	0	0,0	0,000	0	1/9
2, 5	3,5	3,5	3	4,5	2,121	0,5	1/9
5, 1	3,0	3,0	4	8,0	2,828	1	1/9
5, 2	3,5	3,5	3	4,5	2,121	0,5	1/9
5, 5	5,0	5,0	0	0,0	0,000	1	1/9
Média dos Valores da Estatística	8/3	8/3	16/9	26/9	1,3	2,3	
Parâmetro Populacional	8/3	2	4	26/9	1,7	2,3	
A estatística amostral atinge o alvo do parâmetro populacional?	Sim	Não	Não	Sim	Não	Sim	

- Estatísticas que atingem parâmetro (**estimadores não-viesados**): média, variância, proporção.
- Estatísticas que não atingem parâmetro (**estimadores viesados**): mediana, amplitude, desvio padrão.

POR QUE AMOSTRAR COM REPOSIÇÃO?

– Quando selecionamos amostras pequenas de grandes populações, não há diferença significativa se amostramos com ou sem reposição.

– Amostragem com reposição resulta em eventos independentes que não são afetados pelos resultados anteriores.

– Eventos independentes são mais fáceis de serem analisados e resultam em fórmulas mais simples.

– Como não é prático obter todas amostras possíveis, podemos tirar conclusões importantes e significativas sobre toda população, usando apenas uma amostra.

O TEOREMA CENTRAL DO LIMITE

ALGUNS PRINCÍPIOS

– Ao selecionar uma **amostra aleatória** de uma população com média (μ) e desvio padrão (σ):

  – **Se $n>30$**, então as médias amostrais têm uma distribuição que pode ser aproximada por uma distribuição normal com média (μ) e desvio padrão ($\sigma/\sqrt{n}$), independente da distribuição da população original.

  – **Se $n\leq30$** e a população original tem uma **distribuição normal**, então as médias amostrais têm uma distribuição normal com média (μ) e desvio padrão ($\sigma/\sqrt{n}$).

  – **Se $n\leq30$**, mas a população original **não tem uma distribuição normal**, então os métodos a seguir não se aplicam.

TEOREMA CENTRAL DO LIMITE (TCL)

– O **teorema central do limite** diz que...

  – se tamanho amostral é grande o suficiente...

  – a distribuição das médias amostrais pode ser aproximada por uma distribuição normal...

  – mesmo que a população original não seja normalmente distribuída.

PRESSUPOSTOS DO TCL

– A **variável aleatória *x*** tem uma distribuição (que pode ou não ser normal) com média μ e desvio padrão σ.

– **Amostras aleatórias simples (AAS)**, com mesmo tamanho amostral *n*, são selecionadas da população.

  – AAS são amostras selecionadas de uma população de modo que todas possíveis amostras de tamanho *n* têm a mesma chance de ser escolhidas.

CONCLUSÕES DO TCL

– **Distribuição** das médias amostrais irá se aproximar de uma distribuição normal à medida que *n* aumentar.

– A **média** de todas médias amostrais é a média μ da população.

– O **desvio padrão** de todas médias amostrais é $\sigma/\sqrt{n}$.

REGRAS PRÁTICAS UTILIZADAS

– Para amostras aleatórias de tamanho ***n* maior que 30**, a distribuição das médias amostrais pode ser bem aproximada pela distribuição normal, mesmo se a população original não for normalmente distribuída.

  – Populações com distribuições muito diferentes da normal requerem tamanhos amostrais maiores que 30, mas estas são exceções.

  – A aproximação é melhor, quando *n* aumenta.

– Se a população original for **normalmente distribuída**, médias amostrais serão normalmente distribuídas para qualquer tamanho amostral *n*.

NOTAÇÕES DAS ESTATÍSTICAS

– Teorema central do limite envolve a distribuição da população original e a distribuição das médias amostrais.

– Se todas possíveis amostras de tamanho *n* são selecionadas de uma população com média μ e desvio padrão σ, a média das médias amostrais é designada por:

$$\mu_{\bar{x}} = \mu$$

– O desvio padrão das médias amostrais é chamado de erro padrão da média e é designado por:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

APLICAÇÃO DO TEOREMA CENTRAL DO LIMITE

– Ao lidar com um valor individual de uma população normalmente distribuída, use:

$$z = \frac{x - \mu}{\sigma}$$

– Ao lidar com uma média de alguma amostra, certifique-se de usar $\sigma/\sqrt{n}$ como desvio padrão das médias amostrais:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

APLICAÇÃO DO TEOREMA CENTRAL DO LIMITE

– Com dados amostrais, supomos que homens têm pesos normalmente distribuídos, μ de 172 libras e σ de 29 libras.

– Probabilidade de um homem selecionado aleatoriamente ter peso maior que 175 libras:

$$z = (x - \mu) / \sigma = (175 - 172) / 29 = 0,10$$

P(z>0,1) = 1 – (área à esquerda de z) = 1 – 0,5398 = 0,4602

– Probabilidade de 20 homens selecionados aleatoriamente terem peso acima de 175 libras:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{175 - 172}{\frac{29}{\sqrt{20}}} = \frac{3}{6,4845971} = 0,46$$

P(z>0,46) = 1 – (área à esquerda de z) = 1 – 0,6772 = 0,3228

CORREÇÃO PARA POPULAÇÃO FINITA

– Na aplicação do teorema central do limite, supomos que a população seja infinitamente grande para utilizar:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

– Quando amostramos com reposição, a população é infinita.

– Porém, aplicações reais envolvem amostras sem reposição.

– Quando tirar amostras sem reposição e tamanho amostral *n* for maior que 5% do tamanho finito *N* da população, ajuste o desvio padrão das médias amostrais (erro padrão da média), multiplicando-o pelo fator de correção para população finita:

$$\sqrt{\frac{N-n}{N-1}}$$

A NORMAL COMO APROXIMAÇÃO DA BINOMIAL

USO DA NORMAL COMO APROXIMAÇÃO DA BINOMIAL

– Se uma distribuição binomial satisfaz as condições de $np>=5$ e $nq>=5$, então as probabilidades de uma distribuição binomial podem ser aproximadas por uma distribuição normal:

  – Com média ($\mu=np$).

  – Com desvio padrão ($\sigma=\sqrt{npq}$).

  – Com números inteiros discretos x ajustados pela **correção de continuidade**:

    – O valor discreto x (número de sucessos) é substituído pelo intervalo de x–0,5 a x+0,5.

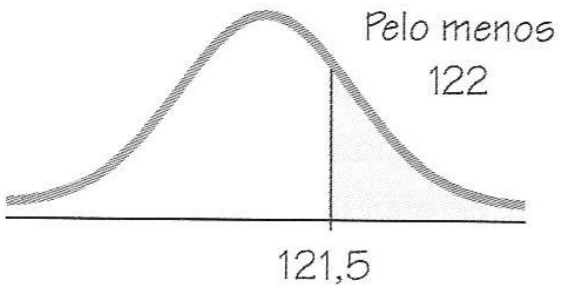
  – Use o escore $z=(x–\mu)/\sigma$ para encontrar a área correspondente à probabilidade desejada.

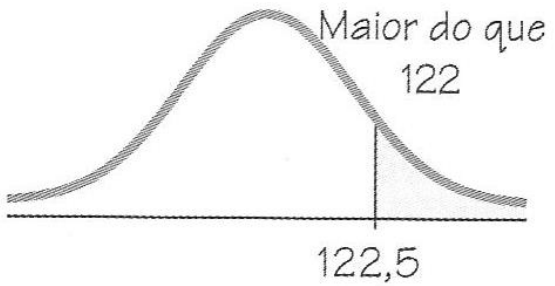
CORREÇÕES DE CONTINUIDADE

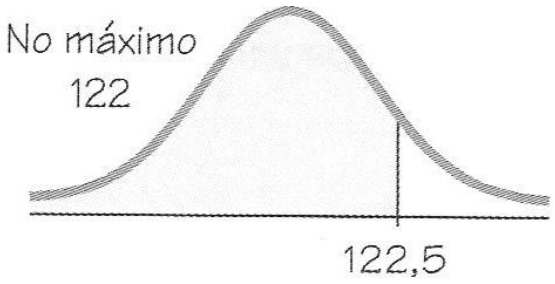
– Ao usar distribuição normal como aproximação da binomial, representamos o único valor *x* pelo intervalo *x*–0,5 a *x*+0,5.

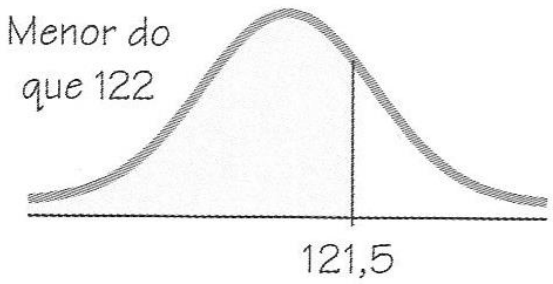
– Se você está tentando achar a probabilidade de obter pelo menos 122 homens entre 213 pessoas, o número discreto é *x*=122, limite inferior é 121,5 e limite superior é 122,5.

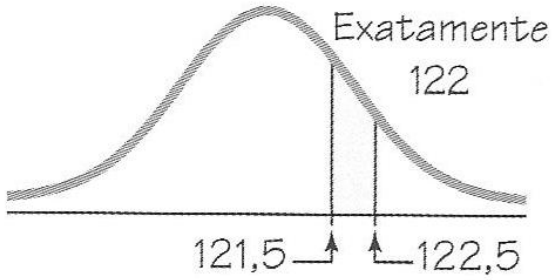
– Determine qual probabilidade você deseja estimar:





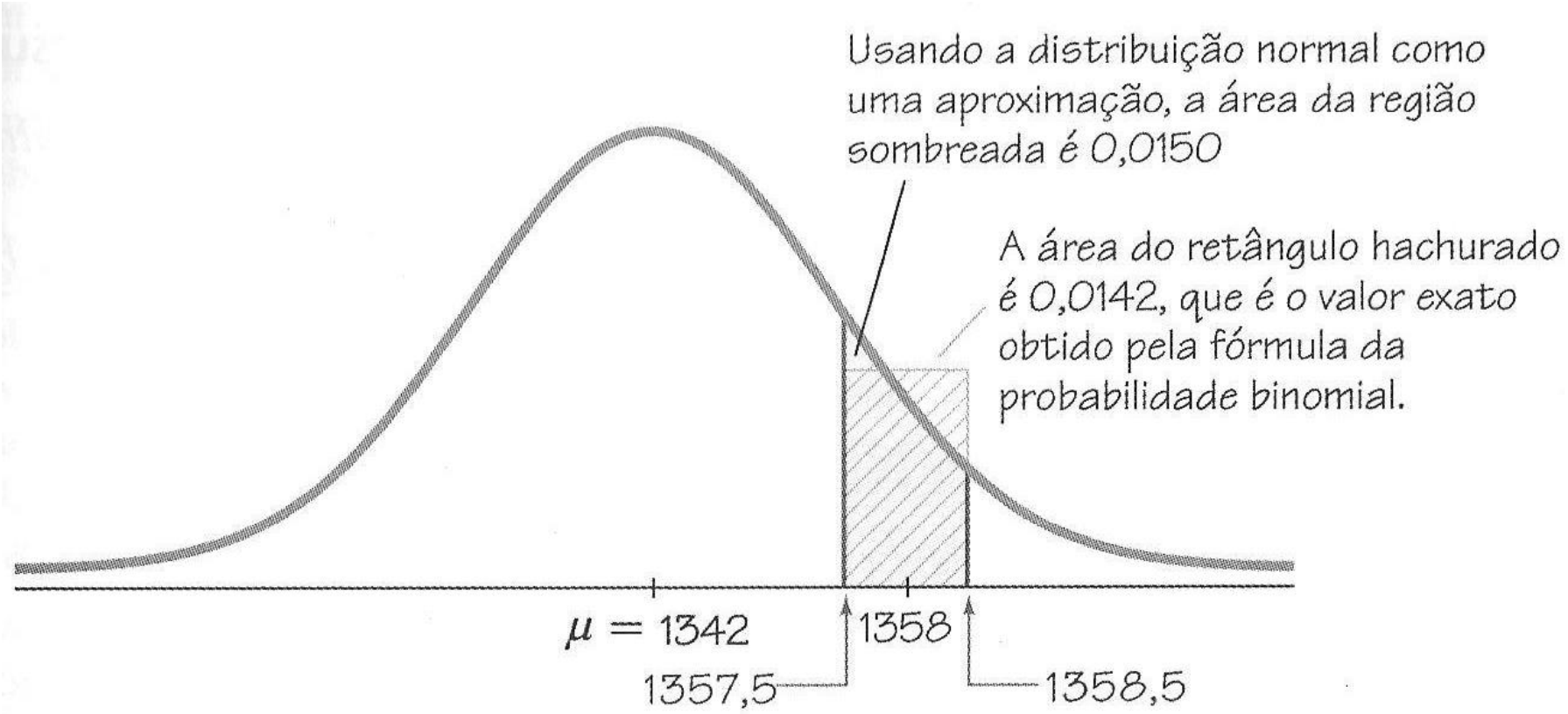






LIMITAÇÃO DO USO DA NORMAL COMO BINOMIAL

– Uso da distribuição normal resulta em valor aproximado, já que área correta exata é a de um retângulo.



DETERMINAÇÃO DE NORMALIDADE

DETERMINAÇÃO DE NORMALIDADE

– Alguns métodos estatísticos exigem que os dados amostrais tenham sido selecionados aleatoriamente de uma população que tenha distribuição normal.

– Podemos analisar histogramas, valores extremos (*outliers*) e gráficos de quantis normais para determinar se as exigências para uma distribuição normal são satisfeitas.

GRÁFICOS QUANTIL-NORMAL

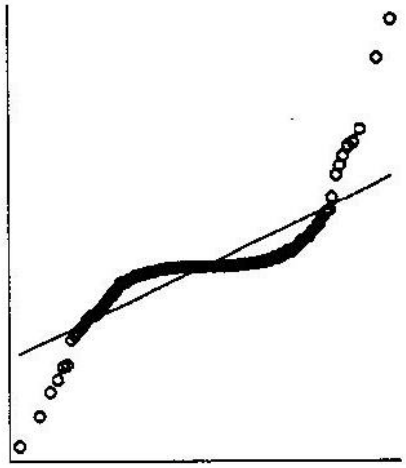
– Um gráfico dos quantis normais (ou gráfico de probabilidades normais) é um gráfico de pontos (x, y) em que um eixo possui o conjunto original de dados amostrais e o outro eixo apresenta o escore *z*, correspondente ao valor esperado do quantil da distribuição normal padrão.

– Se os pontos não se aproximam de uma reta ou se os pontos exibem um padrão simétrico que não seja um padrão linear, então os dados parecem provir de uma população que não tem distribuição normal.

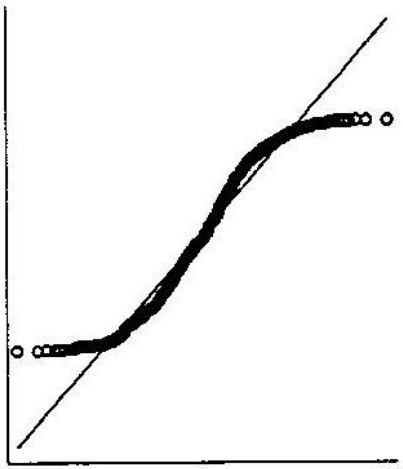
– Se o padrão dos pontos é razoavelmente próximo de uma reta, então os dados parecem provir de uma população com distribuição normal.

– Se a variável seguisse uma distribuição normal, os pontos se encontrariam exatamente sobre a linha diagonal.

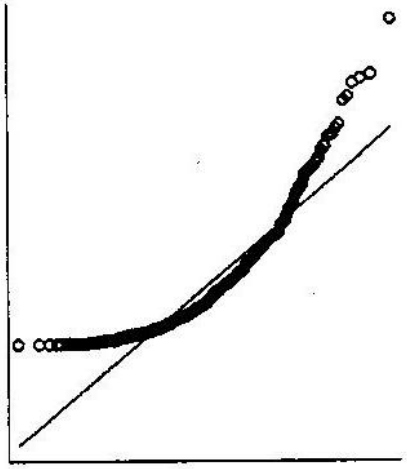
EXEMPLOS DE GRÁFICOS QUANTIL-NORMAL



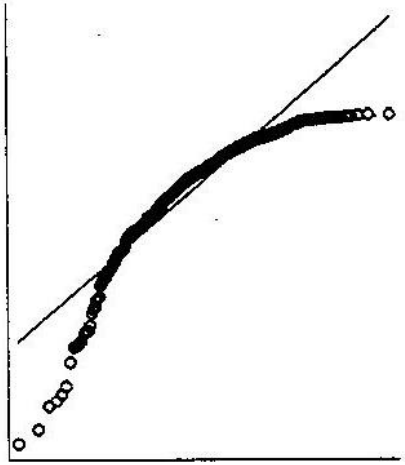
Heavy Tails, High and Low Outliers



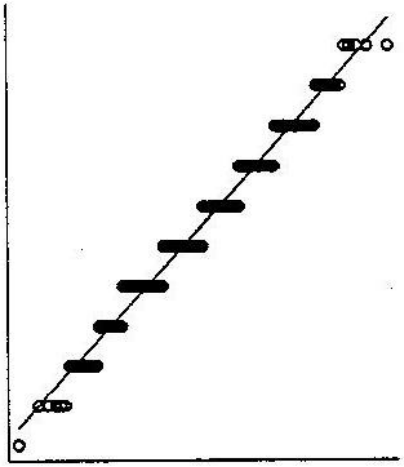
Light Tails, No Outliers



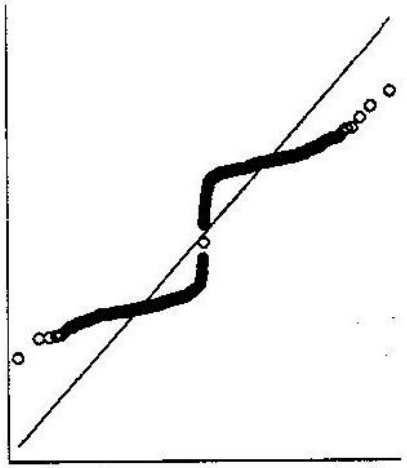
Positive Skew, High Outliers



Negative Skew, Low Outliers



Granularity
(discrete values)



Two Peaks, Central Gap
(bimodal)

Figure 1.10 Quantile-normal plots reflect distribution shape.

Hamilton (1992: 16).

TRANSFORMAÇÃO DE DADOS

– Lawrence Hamilton (“Regression with graphics”) pág.18-19:

Y^3 >>> q=3

Y^2 >>> q=2

Y^1 >>> q=1

$Y^{0,5}$ >>> q=0,5

log(Y) >>> q=0

$-(Y^{-0,5})$ >>> q=–0,5

$-(Y^{-1})$ >>> q=–1

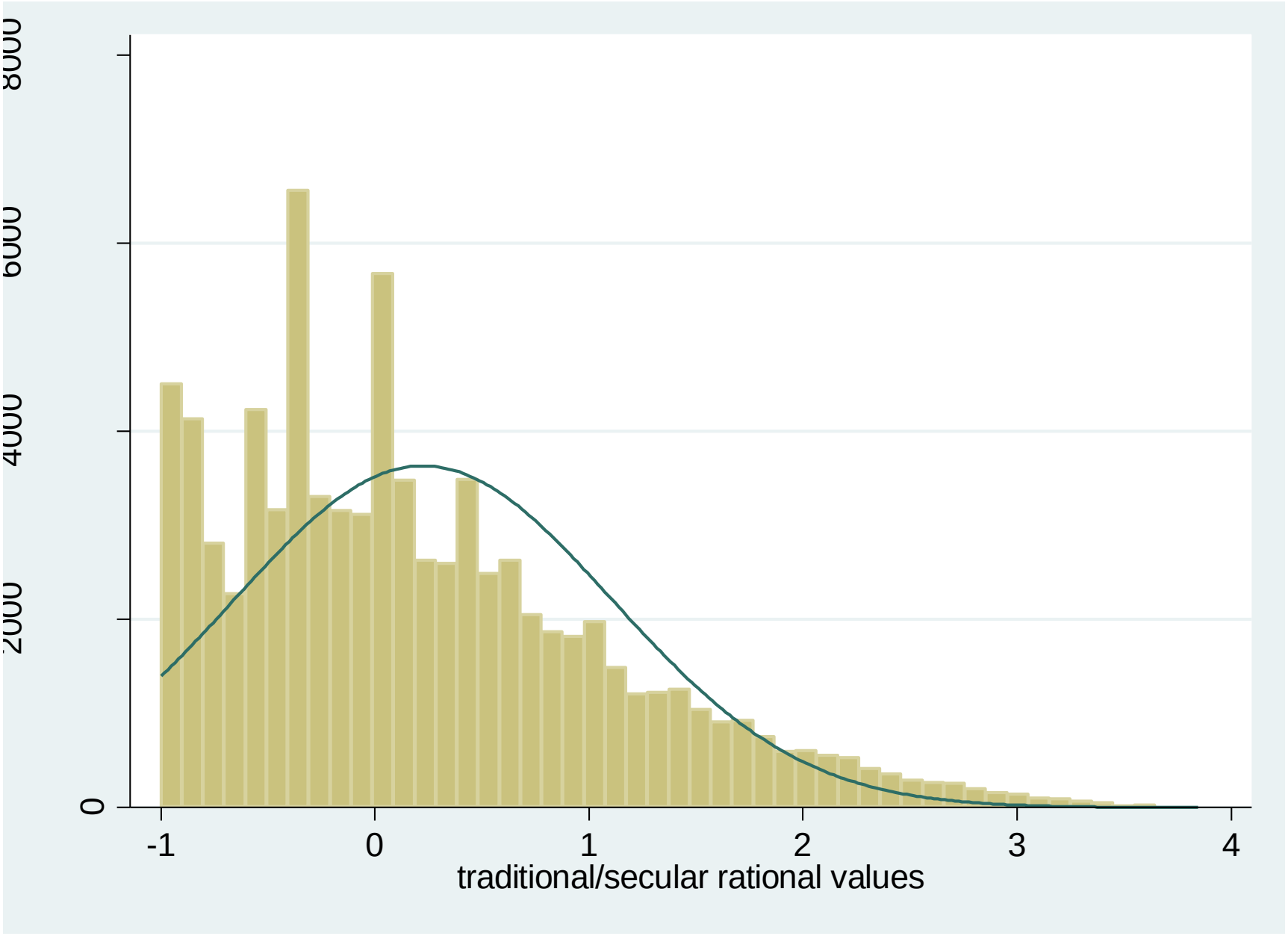
– q>1: reduz concentração à direita.

– q=1: dados originais.

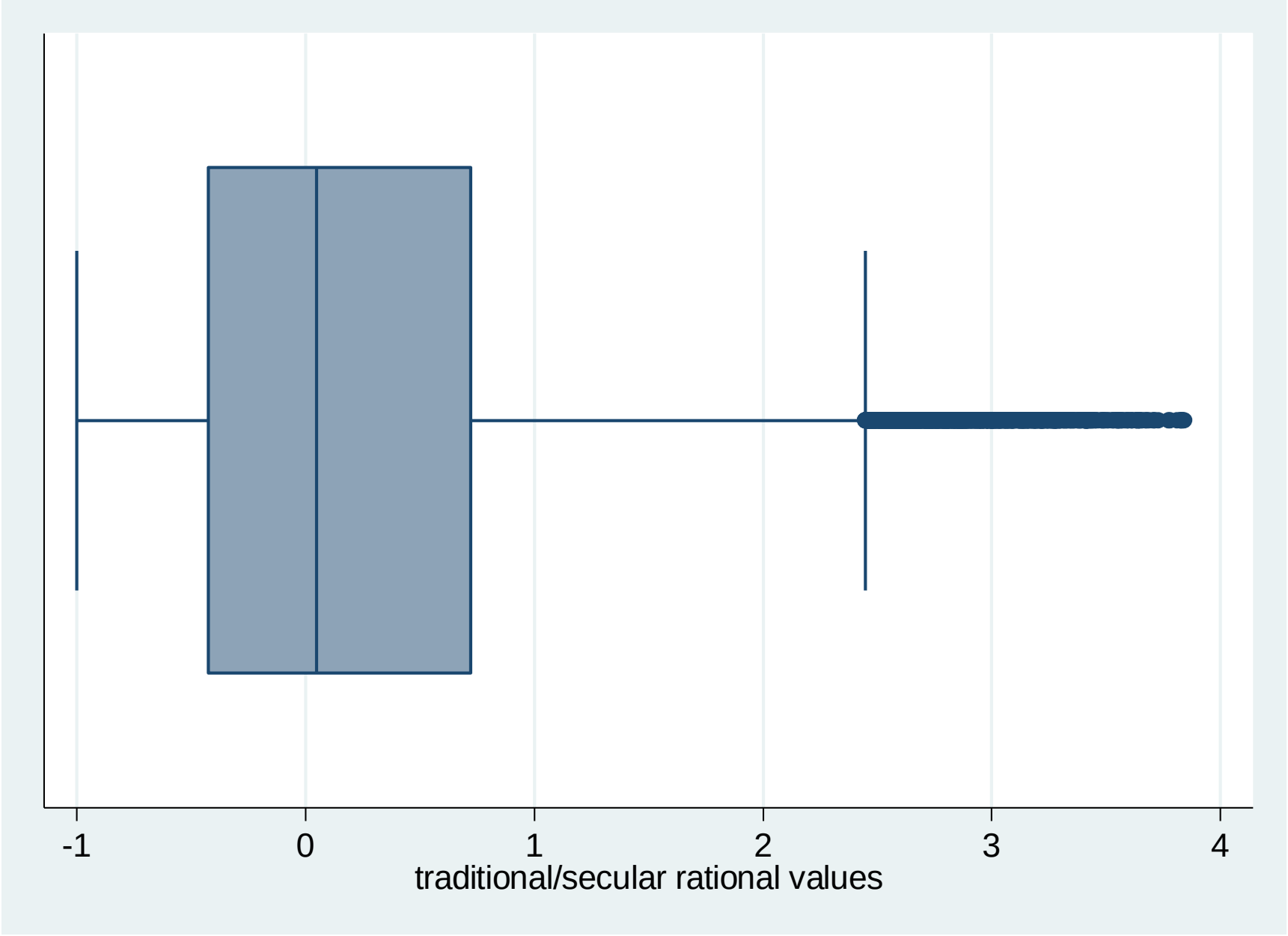
– q<1: reduz concentração à esquerda.

– *log*(*x*+1) viabiliza transformação quando *x*=0. Se distribuição de *log*(*x*+1) for normal, é chamada de distribuição lognormal.

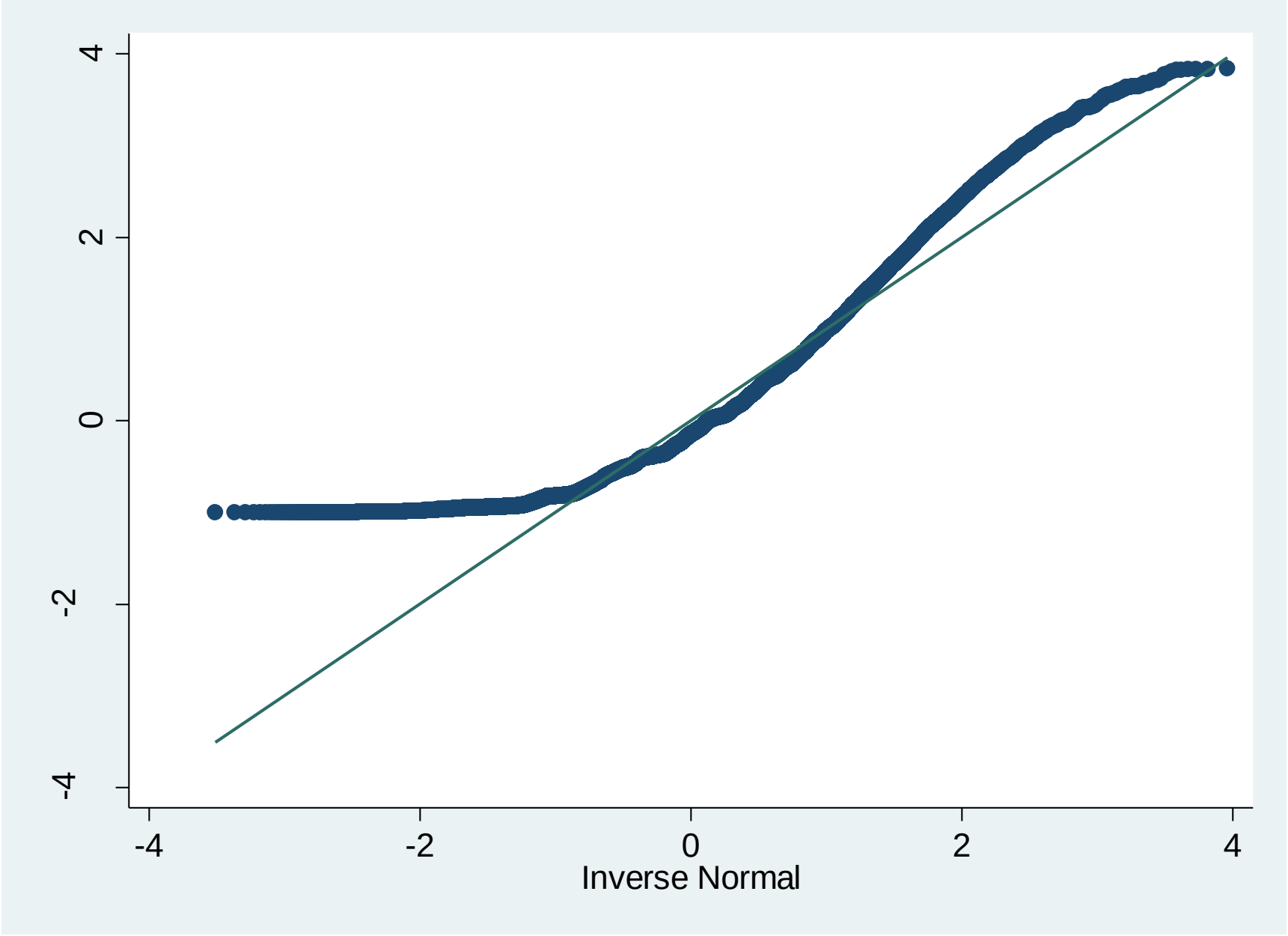
ÍNDICE VALORES RACIONAIS (TRADICIONAL/SECULAR)



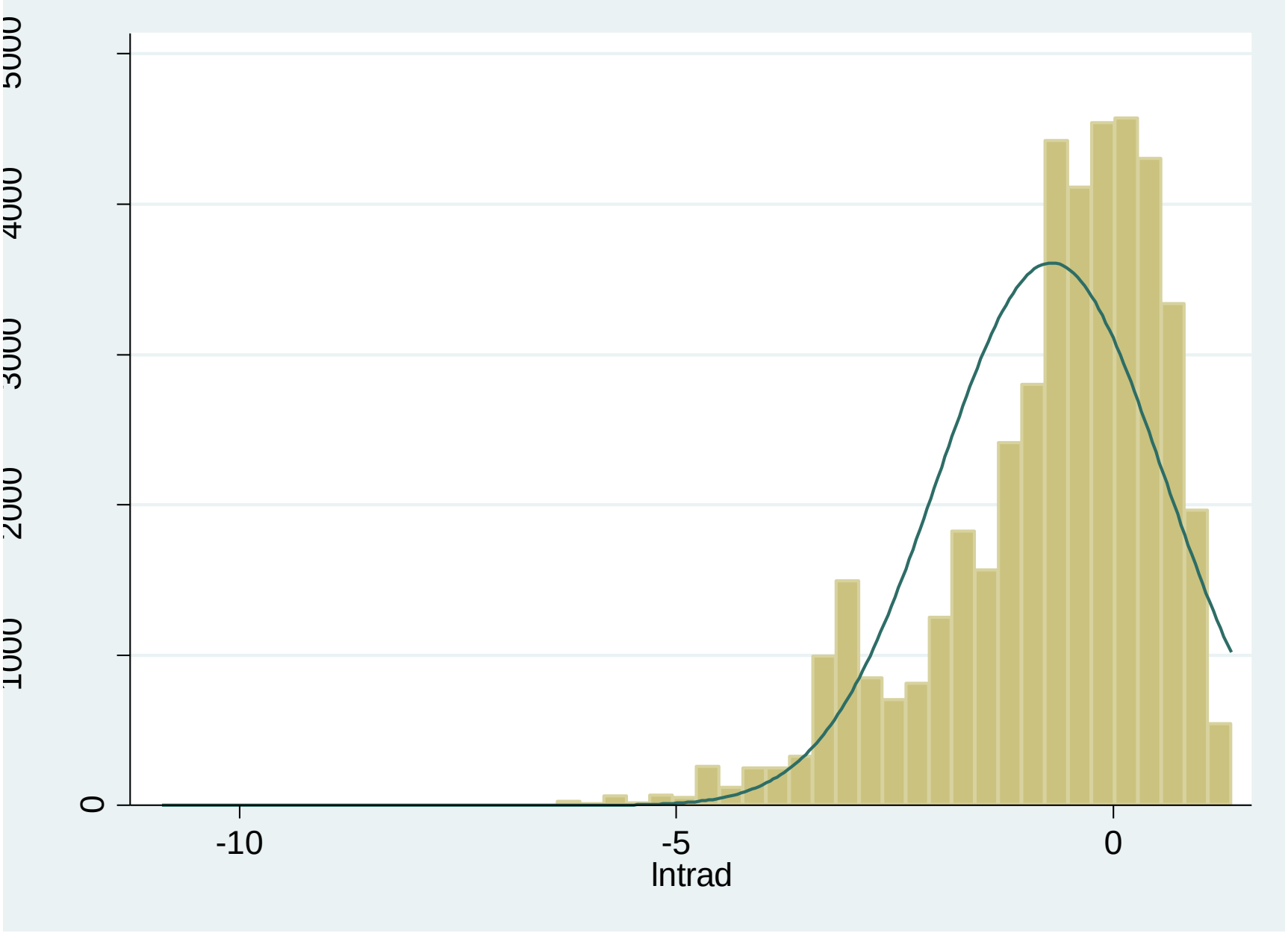
ÍNDICE VALORES RACIONAIS (TRADICIONAL/SECULAR)



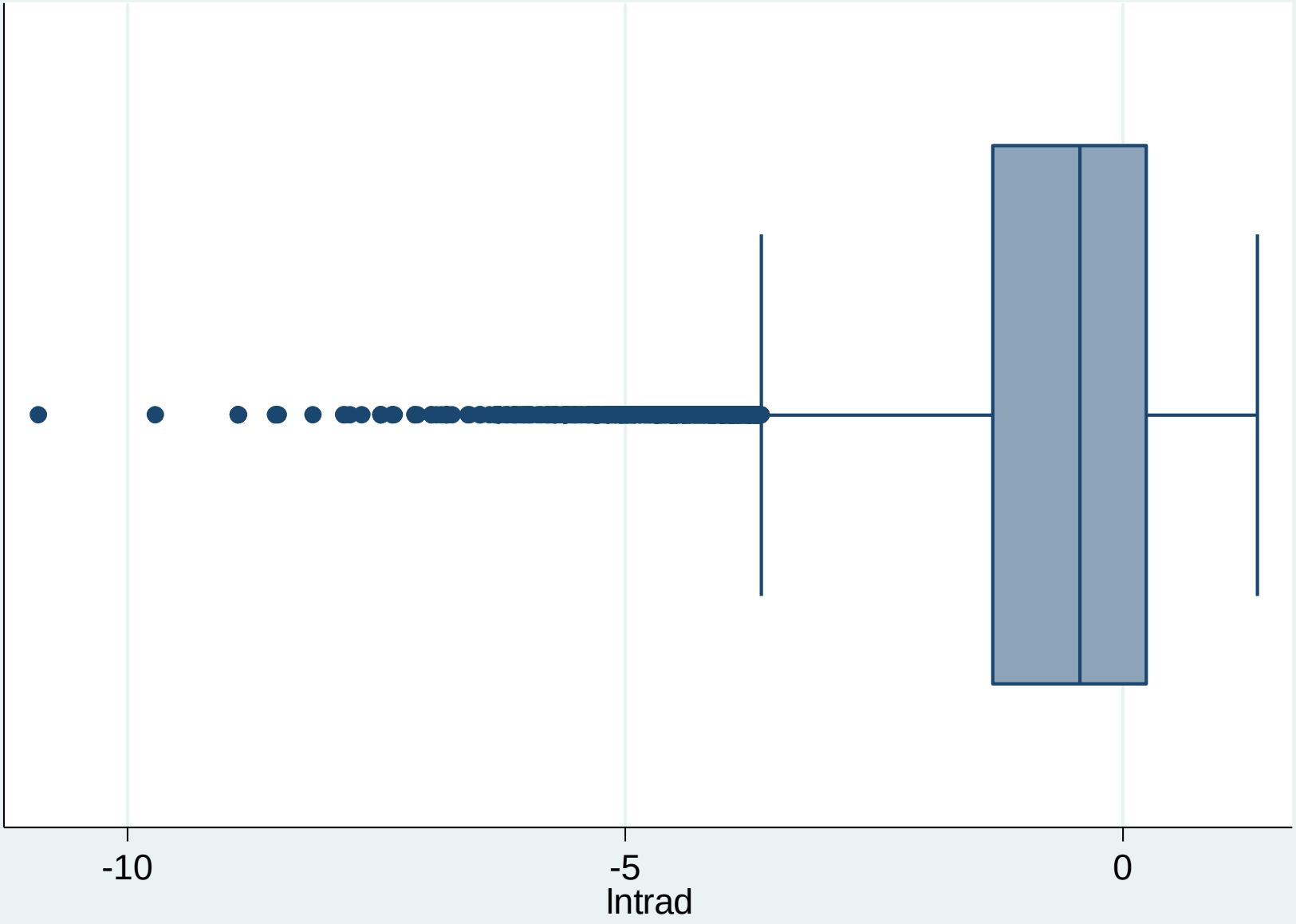
ÍNDICE VALORES RACIONAIS (TRADICIONAL/SECULAR)



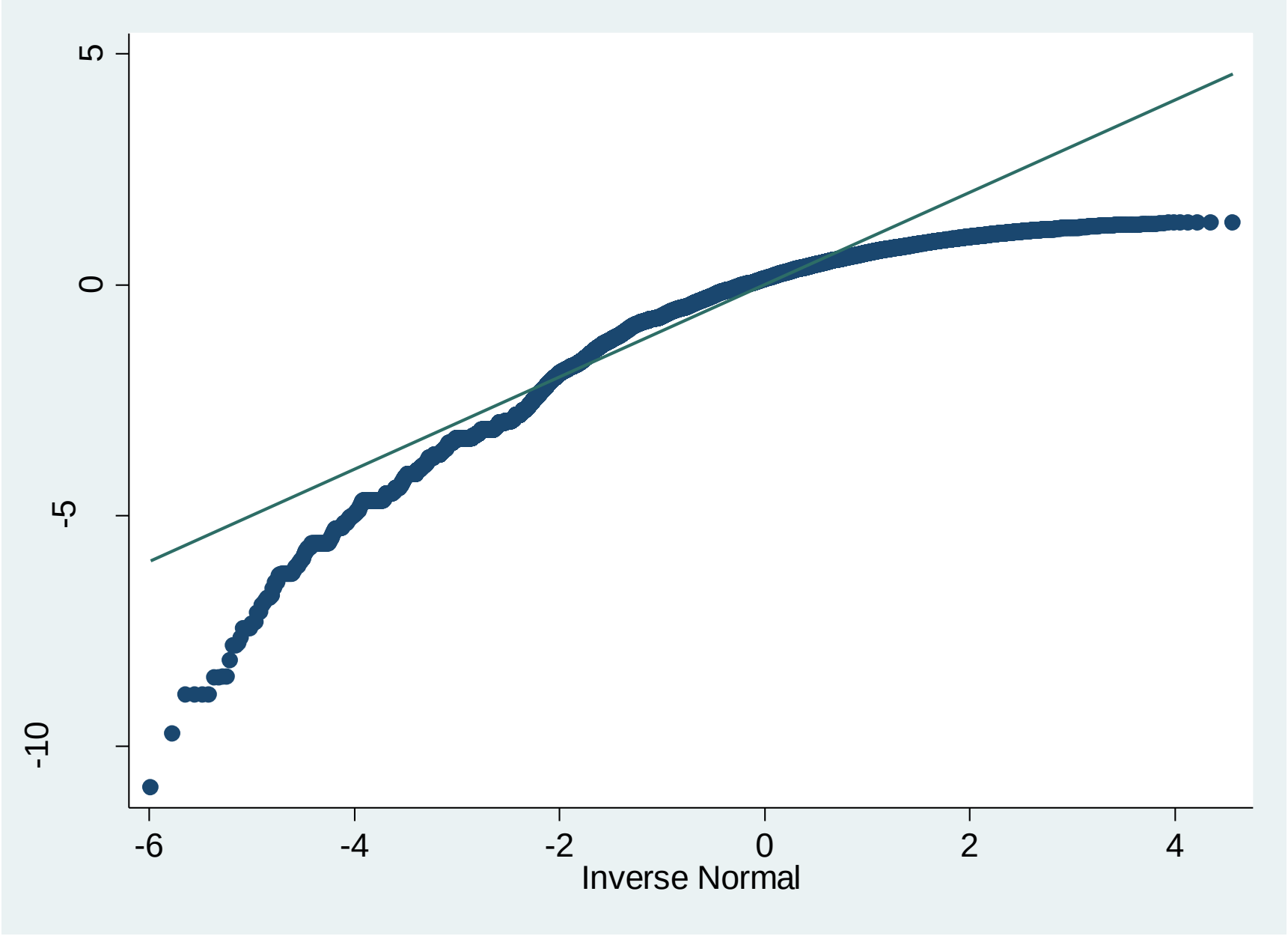
LOGARITMO DO ÍNDICE VALORES RACIONAIS



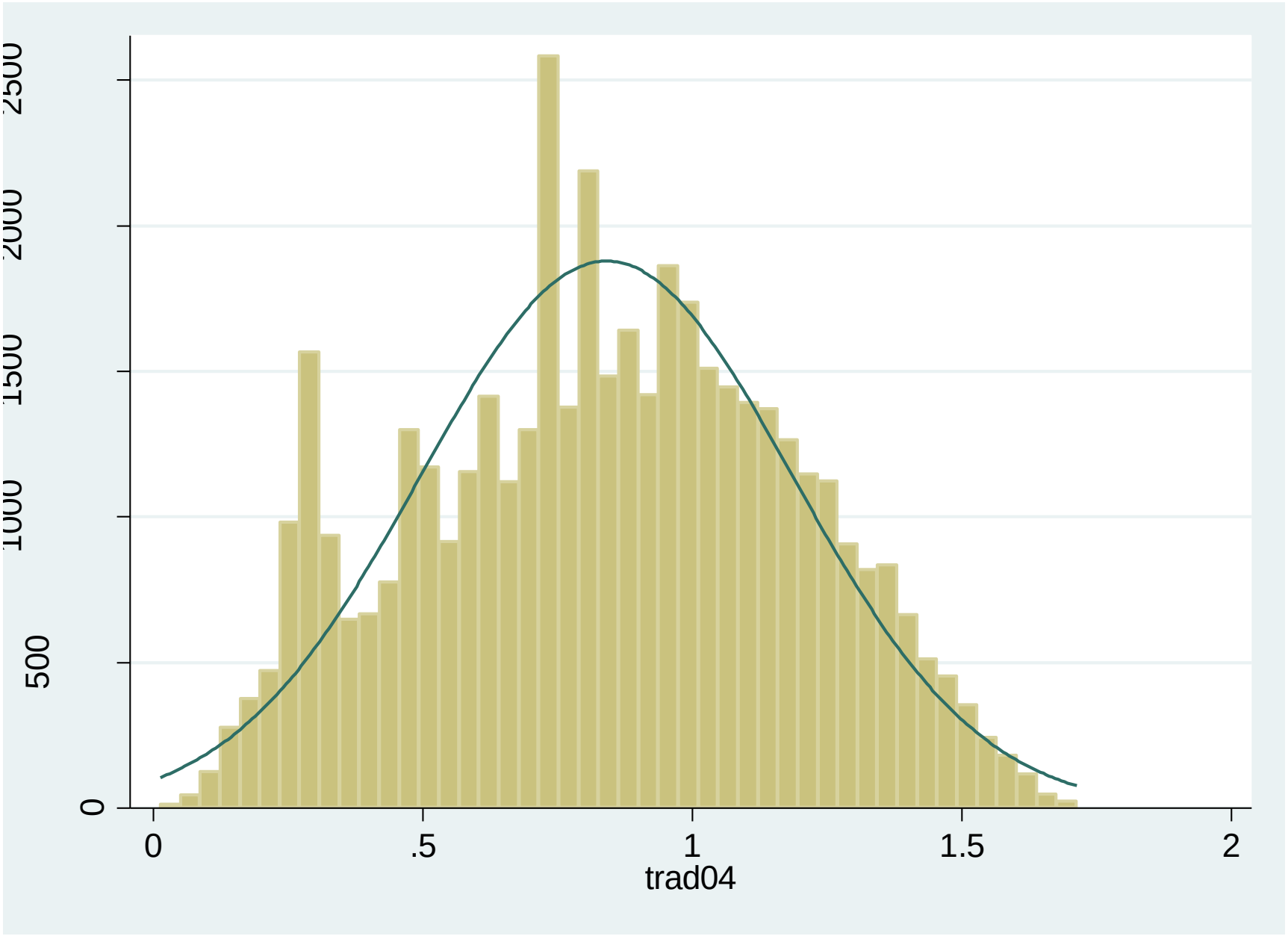
LOGARITMO DO ÍNDICE VALORES RACIONAIS



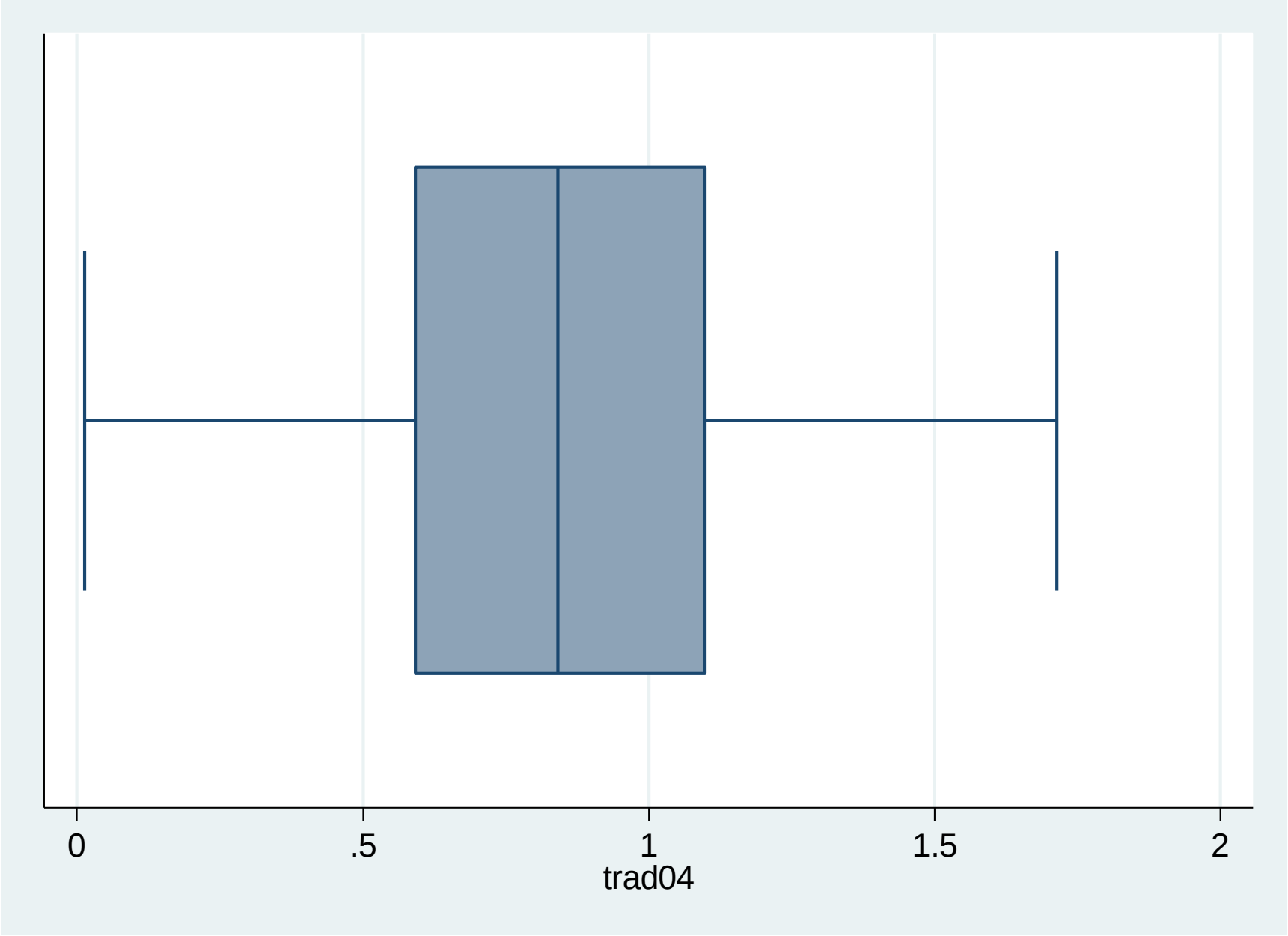
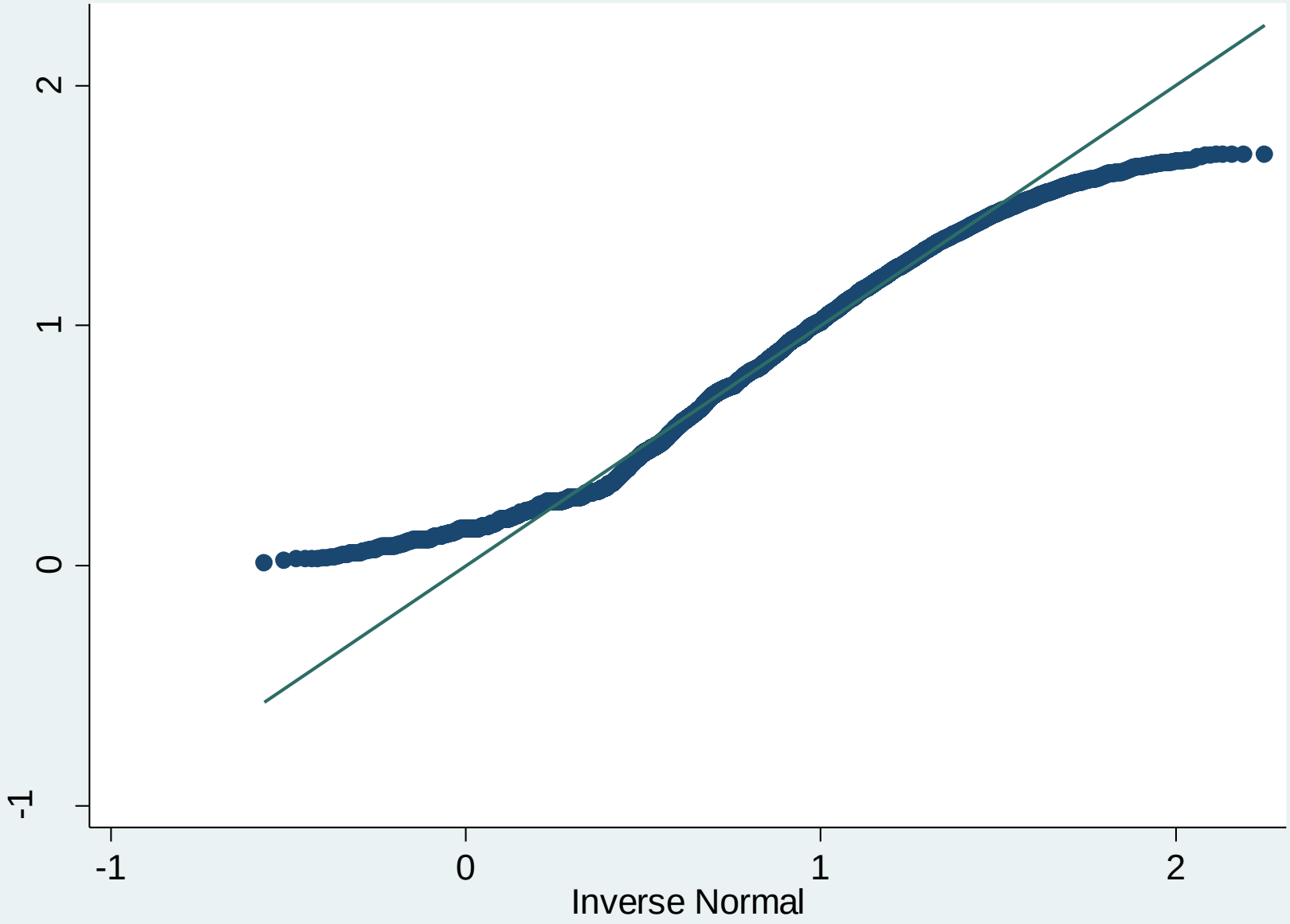
LOGARITMO DO ÍNDICE VALORES RACIONAIS



ÍNDICE VALORES RACIONAIS ELEVADO A 0,4



ÍNDICE VALORES RACIONAIS ELEVADO A 0,4



ÍNDICE VALORES RACIONAIS ELEVADO A 0,4