Intervalos Estatísticos para uma Única Amostra

Roteiro

1. Introdução
2. Intervalo de Confiança para Média
   i. População normal com variância conhecida
   ii. População normal com variância desconhecida
3. Intervalo de Confiança para Variância
4. Intervalo de Confiança para Proporção
5. Roteiro para Construção de Intervalos
6. Intervalos de Tolerância e Previsão
7. Referências

Introdução

Estimação Intervalar

- Exemplo:
  - √ Estimativa da viscosidade média de produto químico
  - √ $\bar{x} = 1000$
  - √ Dificilmente $\mu = \bar{x}$
- Quão próximo x está de μ?
  - √ Entre 900 e 1100?
  - √ Entre 990 e 1010?
  - √ Qual intervalo é mais informativo?

- Estimação intervalar:
  - √ Limites que representam um intervalo de valores plausíveis para um parâmetro
- Intervalo de confiança:
  - √ Estimativa de intervalo para um parâmetro de uma população
  - √ Não podemos estar certos de que o intervalo contém o parâmetro verdadeiro (desconhecido) da população
  - √ Usamos somente uma amostra para estimar intervalo
  - √ Intervalo de confiança é construído de modo que tenhamos alta confiança de que ele contém parâmetro

Intervalo de Tolerância

- Exemplo:
  - √ Dados de viscosidade de produto químico
  - √ Limites de 95% dos valores de viscosidade
    - $\mu - 1{,}96\ \sigma$, $\mu + 1{,}96\ \sigma$ (distribuição normal)
    - μ e σ desconhecidos
      (não é um intervalo útil)
  - √ Podem-se usar estimativas pontuais ($\bar{x}$ e s)
    - $\bar{x} - ks$, $\bar{x} + ks$
    - k > 1,96 para considerar erro de estimação

Importante

- Intervalos de confiança e de tolerância:
  - √ Delimitam elementos desconhecidos de uma população

Intervalo de Previsão

- Fornece limites em uma (ou mais) observação(ões) futura(s) de população
  - √ Exemplo: delimitar uma nova medida de viscosidade
  - √ Com amostra grande o intervalo de previsão tende ao intervalo de tolerância
  - √ Em tamanhos mais modestos, os intervalos de previsão e de tolerância são diferentes

Objetivos

- Intervalo de Confiança:
  - √ delimita os parâmetros da população ou da distribuição
- Intervalo de tolerância
  - √ delimita uma proporção selecionada de uma população
- Intervalo de Previsão
  - √ delimita observações futuras provenientes da população ou da distribuição

Intervalo de Confiança para Média – Distribuição Normal com Variância Conhecida

Situação Especial

- Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória proveniente de
  - √ População **normal**, com
  - √ Média μ desconhecida
  - √ Variância σ^2 **conhecida**
- Cenário não-realista!
  - √ Em geral ambos os parâmetros são desconhecidos

Média Amostral

- Distribuição amostral: $\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$
- Média amostral padronizada:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{N}(0,1)$$

- Intervalo de confiança para μ: $l \leq \mu \leq u$
  - √ Extremos l e u calculados a partir dos dados amostrais
  - √ Diferentes valores de l e u para diferentes amostras
  - √ Extremos são variáveis aleatórias L e U

- Suponha que possamos determinar valores de L e U tal que essa afirmação de probabilidade seja verdadeira

$$\mathrm{P}\{L \le \mu \le U\} = 1 - \alpha, \text{ com } 0 \le \alpha \le 1$$

  √ Há uma probabilidade 1–α de selecionar uma amostra para a qual o IC conterá o valor verdadeiro de μ.

- Particular amostra com $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$.
  √ Calcula-se o intervalo resultante de confiança para m
  √ $l \le \mu \le u$
  √ l e u: limites inferior e superior de confiança
  √ $\gamma = 1 - a$: coeficiente de confiança

- Determinação intervalo de confiança para μ:
  √ População normal e variância conhecida

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$$

  – Não depende do parâmetro desconhecido μ!

Então
$$\mathrm{P}\left\{-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le z_{\alpha/2}\right\} = 1 - \alpha$$

sendo $z_{\alpha/2}$ o percentil superior com α/2(100)% da normal padrão

logo
$$\mathrm{P}\left\{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

Intervalo de Confiança para a Média, Variância Conhecida

- Seja $\bar{x}$ a média de amostra aleatória, de tamanho n, oriunda de população normal com variância σ^2 conhecida
  √ Intervalo com 100(1 – α)% de confiança para μ:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Exemplo 8-1

- Teste de impacto em barras entalhadas
  √ Técnica Charpy V-notch (ASTM E23)
  √ Mede energia de impacto
  √ Usada para determinar se material experimenta ou não transição do dúctil-frágil com decréscimo temperatura
  √ População de medidas de energia (J) de impacto em A238, com desvio padrão σ = 1 J

√ Amostra de medidas de energia (J) de impacto em corpos de prova de aço A238

- n = 10
- Média amostral: $\bar{x}$ = 64,46 J
- Coeficiente de confiança: 1 – α = 95%
- $z_{\alpha/2} = z_{0,025} = 1,96$

√ Intervalo com 95% de confiança para μ:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64,46 - 1,96\frac{1}{\sqrt{10}} \leq \mu \leq 64,46 + 1,96\frac{1}{\sqrt{10}}$$

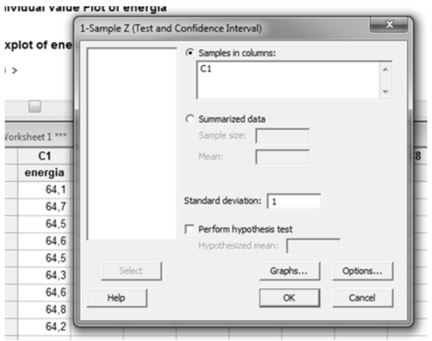
$$63,84 \leq \mu \leq 65,08$$

– Baseado nos dados amostrais, essa é uma faixa de valores altamente plausíveis para a energia média de impacto para o aço A238 a 60° C

- Saída Minitab:

√ Dados: *BD_estatistica.xlsx/impacto*

Stat > Basic Statistics > 1–Sample Z ➔



- Saída

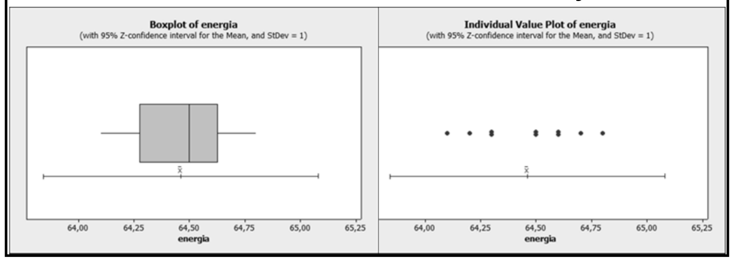
```
MTB > OneZ C1;
SUBC>   Sigma 1.

One-Sample Z: C1

The assumed standard deviation = 1


Variable   N    Mean  StDev  SE Mean       95% CI
C1        10  64,460  0,227    0,316  (63,840; 65,080)
```

- Gráficos dados e o intervalo de confiança:



Intervalo de Confiança – Interpretação

- O valor de μ é desconhecido:

√ A afirmação 63,84 ≤ μ ≤ 65,08 é tanto correta quanto falsa

- Interpretação correta:

√ Um IC é um intervalo aleatório

(os extremos são variáveis aleatórias)

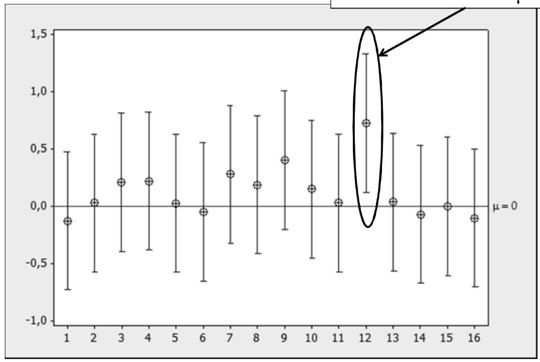
√ São construídos os intervalos com (1 – α) 100% de confiança de um número infinito de amostras

√ (1 – α) 100% desses intervalos conterão o valor verdadeiro de μ.

√ Intervalo com 95% de confiança para população normal

– Simulação com n =10, μ =0

Intervalo falha em conter o verdadeiro valor de μ



- Na prática:
  - √ O IC é construído a partir de uma única amostra aleatória
  - √ Esse intervalo poderá conter ou não o verdadeiro valor de μ.
  - √ Não é razoável vincular um nível de probabilidade a esse evento específico

- Afirmação apropriadas:
  - √ O intervalo observado [l, u] envolve o valor verdadeiro de μ, com confiança de $(1 - \alpha)$ 100%
  - √ Essa afirmação tem uma interpretação de frequência
  - √ Não sabemos se a afirmação é verdadeira para essa amostra específica
  - √ O **método** usado para obter [l, u] resulta em afirmações corretas $(1 - \alpha)$ 100% das vezes

Nível de Confiança e Precisão de Estimação

- Comprimento do intervalo de confiança: $2\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$
  - √ Com 95% de confiança $2\left(1,96\frac{\sigma}{\sqrt{n}}\right) = 3,92\frac{\sigma}{\sqrt{n}}$
  - √ Com 99% de confiança $2\left(2,58\frac{\sigma}{\sqrt{n}}\right) = 5,16\frac{\sigma}{\sqrt{n}}$
  - √ O IC com 99% é maior que o IC com 99%
    (o nível de confiança é maior)

√ Comprimento do IC é medida de **precisão** da estimação
– Precisão é inversamente proporcional ao comprimento

- Desejável:
  √ IC curto o suficiente (preciso) para tomada de decisão e com confiança adequada
- Solução:
  √ Escolher o tamanho da amostra n grande o suficiente para construir IC com **precisão** (comprimento) especificada, com a **confiança** prescrita.

Escolha do Tamanho da Amostra

- Precisão do IC: $2\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$
- Erro ao usar $\bar{x}$ para estimar μ: $\mathrm{E} = |\bar{x} - \mu|$
- Tamanho da amostra:
  √ Escolher *n* tal que $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \mathrm{E}$
  √ Comprimento do intervalo resultante: 2E

- Tamanho da amostra com erro especificado:
  √ Se $\bar{x}$ for usada como estimativa de μ, podemos estar $(1 - \alpha)100\%$ confiantes de que o erro $|x - \mu|$ não excederá o valor E especificado quando o tamanho da amostra for

$$n = \left(\frac{z_{\alpha/2}\sigma}{\mathrm{E}}\right)^2$$

– Deve ser arredondado para número inteiro

Exemplo 8-2

- Continuação exemplo 8.1
  √ Teste Charpy V-notch
  √ Quantos corpos de prova para assegura que o IC de 95% para μ para o aço A238, cortado a 60° C, tivesse um comprimento máximo de 1,0 J.
  - Erro de estimação: E = 0,5
  - σ = 1 J
  - $z_{\alpha/2}$ = 1,96

$$n = \left(\frac{z_{\alpha/2}\sigma}{\mathrm{E}}\right)^2 = \left[\frac{(1,96)(1)}{0,5}\right]^2 = 15,37$$

  √ Tamanho da amostra: n = 16

- Os intervalos de confiança obtidos são bilaterais:
  - √ Fornecem limites de confiança inferior (*l*) e superior (*u*)

Limites Unilaterais

- Limites unilaterais de confiança para μ.
  - √ Estabelecer $l = -\infty$ ou $u = \infty$
  - √ Trocar $z_{\alpha/2}$ por z_α.
- Limites unilaterais de confiança para a média com variância conhecida
  - √ Limite superior com (1 – α)100% de confiança para μ.

$$\mu \le u = \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

  - √ Limite inferior com (1 – α)100% de confiança para μ.

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = l \le \mu$$

Exemplo 8-3

- Continuação exemplo 8.1
  - √ Teste CharpyV-notch
  - √ Intervalo unilateral com 95% de confiança para a energia média de impacto.
    - Média amostral: 64,46 J
    - σ = 1 J
    - $z_\alpha = 1{,}64$
    - *n* = 10

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu$$

$$64,46 - 1,64\frac{1}{10} \le \mu$$

$$63,94 \le \mu$$

- √ O limite inferior de intervalo unilateral é sempre maior que o limite equivalente em IC bilateral
  - – $z_\alpha < z_{\alpha/2}$.
  - – Se o interesse é apenas o limite inferior para μ, então prefere-se o IC unilateral por fornecer a mesma confiança com limite inferior maior
- √ Similarmente, um limite unilateral superior é sempre menor do que um limite bilateral de igual confiança

Método Geral para Deduzir IC

- IC para um parâmetro desconhecido θ.
  - √ Amostra aleatória $X_1, ..., X_n$ de tamanho *n*.
  - √ Suponha existir estatística $g(X_1, ..., X_n; \theta)$ com as seguintes propriedades:
    - $g(X_1, ..., X_n; \theta)$ depende da amostra e de θ.
    - A distribuição de probabilidades de $g(X_1, ..., X_n; \theta)$ não depende de θ ou de qualquer outro parâmetro desconhecido
  - √ Exemplo: $\theta = \mu$
    - A estatística $g(X_1, X_2, \ldots, X_n) = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ satisfaz as duas condições

√ Encontrar as constantes C_L e C_U tal que

$$\mathrm{P}\{C_L \le g(X_1, X_2, \ldots, X_n) \le C_U\} = 1 - \alpha$$

- C_L e C_U não dependem de θ.

√ No exemplo:

- $C_L = -z_{\alpha/2}$ e $C_U = z_{\alpha/2}$.

√ Manipular as desigualdades de modo que:

$$\mathrm{P}\{L(X_1, X_2, \ldots, X_n) \le \theta \le U(X_1, X_2, \ldots, X_n)\} = 1 - \alpha$$

- Limite inferior de confiança: $L(X_1, X_2, \ldots, X_n)$
- Limite superior de confiança: $U(X_1, X_2, \ldots, X_n)$

- No exemplo:

$$L(X_1, X_2, \ldots, X_n) = \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$U(X_1, X_2, \ldots, X_n) = \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- A grandeza $g(X_1, ..., X_n; \theta)$ é denominada grandeza pivotal
  - √ Em nosso exemplo
    - Grandeza pivotal: $g(X_1, X_2, \ldots, X_n) = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Intervalo de Confiança para a Média – Amostra Grande

- Não requer suposições de população normal e variância conhecida
- Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de população qualquer com média μ e variância σ^2 desconhecidas
  - √ O tamanho amostral n é grande o suficiente para permitir a aplicação do TCL

- Pelo TCL: $$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$$
  - √ σ é desconhecido!
  - √ Como *n* é grande, a troca de σ pelo desvio padrão amostral S tem pouco efeito na distribuição de Z.
  - √ Assim

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \overset{as.}{\sim} \mathrm{N}(0,1), \text{ para } n \text{ suficientemente grande}$$

- Intervalo com (1 – α)100% de confiança para μ para amostras grandes

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

  - √ O resultado se mantém independente da forma da distribuição da população
  - √ Em geral, *n* deveria ser no mínimo 40 para usar esse resultado de forma confiável

    (O TCL geralmente se mantém com $n \geq 30$)
  - √ Aqui recomenda-se tamanho amostral maior pois a troca de σ por S implica maior variabilidade

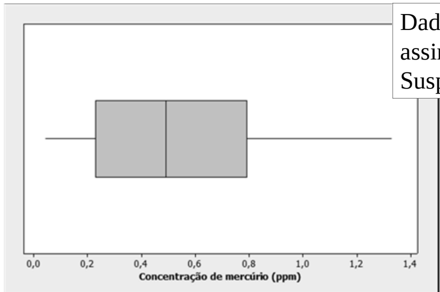
Exemplo 8.4

- Contaminação por mercúrio
  - √ Investigação por mercúrio em peixe de boa grande
  - √ Amostra de peixes de 53 lagos da Flórida
  - √ Medidas de concentração (em ppm) de mercúrio no tecido muscular

- Saídas Minitab:
  - √ Estatística Descritiva

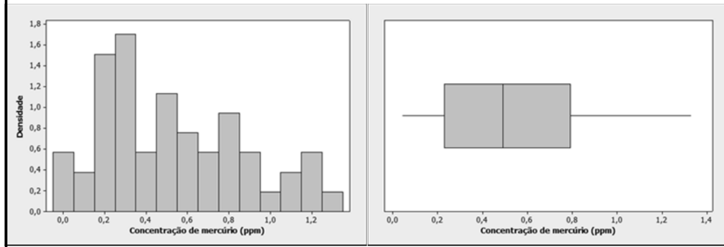
Descriptive Statistics: concentração

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
concentração	53	0,5250	0,0479	0,3486	0,0400	0,2300	0,4900	0,7900	1,3300



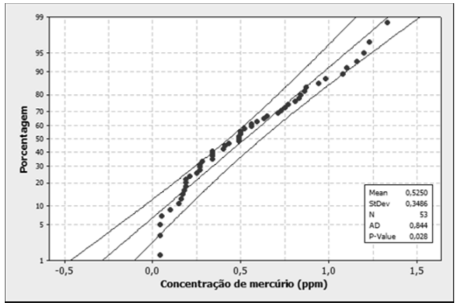
Dados amostrais indicam assimetria
Suspeita de não-normalidade

- Histograma dos dados:



√ Dados amostrais indicam que distribuição da concentração de mercúrio não é normal

- Gráfico de probabilidade normal dos dados



√ Há evidências amostrais que indicam a não-normalidade da população

- Intervalo aproximado de 95% de confiança para μ:
  - √ Suposição de normalidade não é necessária (*n* > 40)

$$\bar{x} - z_{0,025}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{0,025}\frac{s}{\sqrt{n}}$$

$$0,5250 - 1,96\,\frac{0,3486}{\sqrt{53}} \leq \mu \leq 0,5250 + 1,96\,\frac{0,3486}{\sqrt{53}}$$

$$0,4311 \leq \mu \leq 0,6189$$

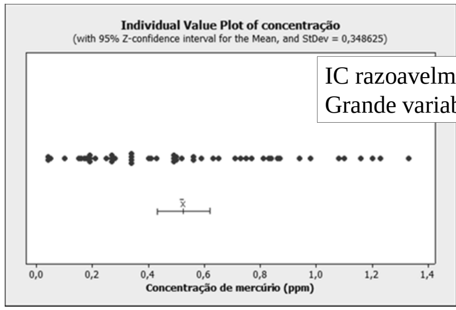
  - √ Intervalo é razoavelmente largo
    - Há grande variabilidade nas medidas de concentração de mercúrio

- Saída Minitab

One-Sample Z: concentração

```
The assumed standard deviation = 0,348625

Variable       N    Mean   StDev  SE Mean         95% CI
concentração  53  0,5250  0,3486   0,0479  (0,4311; 0,6188)
```



IC razoavelmente largo
Grande variabilidade nas medidas

Intervalo de Confiança para um Parâmetro, Amostra Grande

- Suponha θ um parâmetro de uma população e $\hat{\theta}$ um estimador de θ.
  - √ Se $\hat{\theta}$ apresentar:
    - i. Uma distribuição normal aproximada
    - ii. Aproximadamente não viciado para θ.
    - iii. Desvio padrão $\sigma_{\hat{\theta}}$ que pode ser estimado dos dados

  Então $$\frac{(\hat{\theta} - \theta)}{\sigma_{\hat{\theta}}}$$

  tem aproximadamente distribuição normal padrão

- Atendidas essas condições, um IC aproximado para θ no caso de amostras grandes é:

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$$

  - √ Em geral, os EMV satisfazem as 3 condições listadas
  - √ Pode-se usar a expressão quando $\sigma_{\hat{\theta}}$ for função de parâmetros desconhecidos
    - Usar dados amostrais para calcular estimativas de parâmetros desconhecidos e substituí-las na expressão

Construção de Intervalos de Confiança para a Média

- Procedimentos estudos até aqui:
  - √ Intervalos de confiança exatos para média:
    - População normal, com variância conhecida
    - Amostras de qualquer tamanho
  - √ Intervalos de confiança aproximados para a média:
    - População com qualquer distribuição de probabilidade
    - Variância desconhecida

      (Usa-se s como estimativa pontual de σ)
    - Amostras grandes ($n \geq 40$)

- Como construir intervalos de confiança para a média para população com:
  - √ variância desconhecida e
  - √ amostra pequena?

Intervalo de Confiança para Média – Distribuição Normal com Variância Desconhecida

- Suposição razoável:
  - √ População normal
    - – Muitas populações encontradas na prática são bem aproximadas pela normal
    - – Assim, procedimentos de construção de IC baseados na normalidade têm larga aplicabilidade
    - – Desvios moderados de normalidade têm pequeno efeito nas conclusões do IC
- Se normalidade não é suposição razoável:
  - √ Alternativa
    - – Usar procedimentos não-paramétricos para construir IC

Hipóteses para Desenvolvimento do Procedimento

- Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória proveniente de
  - √ População **normal**, com
  - √ Média μ desconhecida
  - √ Variância σ^2 **desconhecida**
- Estatísticas amostrais:
  - √ Média amostral: $\overline{X}$
  - √ Variância amostral: S^2.

- A variância σ^2 é desconhecida
  - √ Não é possível calcular a quantidade Z

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$$

  - √ Um procedimento natural é trocar σ pelo desvio-padrão da amostra S
  - √ Calcula-se então a quantidade T:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

    - – Qual a distribuição da estatística T?

- Qual o efeito na distribuição ao trocar σ por S?
  - √ Se a amostra for grande muda-se "muito pouco"
    - Usa-se a normal para construir o IC

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \overset{as.}{\sim} \mathrm{N}(0,1), \text{ para } n \text{ suficientemente grande}$$

  - √ Em geral, no entanto, *n* é pequeno
    - Emprega-se uma outra distribuição na construção do IC

Distribuição *t*

- Seja X_1, X_2, ..., X_n uma amostra aleatória proveniente de uma população normal, com média μ e variância σ^2 desconhecidas.
  - √ A variável aleatória

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim \mathrm{t}_{(n-1)}$$

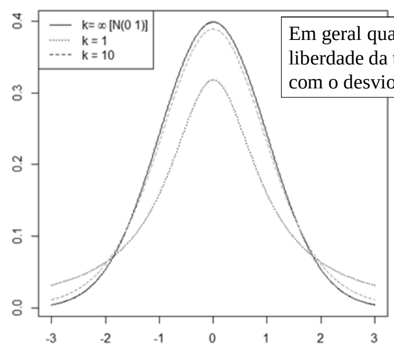
tem uma distribuição *t* com n–1 graus de liberdade

Distribuição *t* – Função de Densidade

- Seja *X* uma variável aleatória com distribuição *t* com k graus de liberdade:

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left[\left(1+\frac{x^2}{k}\right)\right]^{\frac{k+1}{2}}}, \quad -\infty < x < \infty; k = 1, 2 \ldots$$

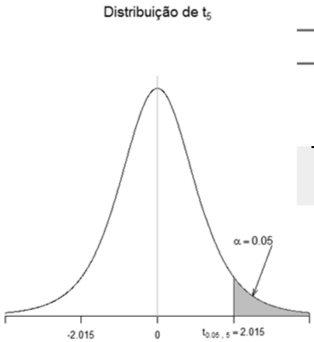
  - √ *k* é o número de graus de liberdade
  - √ Média da distribuição *t*: $\mathrm{E}[X] = 0, \forall k > 1$
  - √ Variância da distribuição *t*: $\mathrm{Var}[X] = \dfrac{k}{k-2}, \forall k > 2$



Em geral quantidade de graus de liberdade da t está associada com o desvio-padrão estimado

- √ Distribuições simétricas em torno de zero e unimodais
- √ Distribuição *t*: tem caudas mais pesadas que a normal
- √ Forma limite da t = normal padrão quando k → ∞

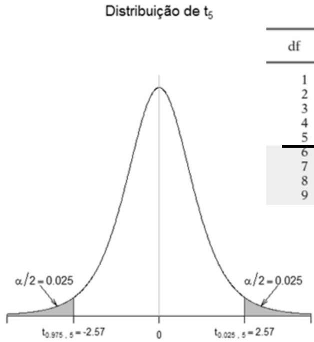
- $t_{\alpha,\,k}$: Percentis da t
  - √ Seja a variável aleatória T ~ t_k, então $P\{T_k > t_{\alpha,k}\} = \alpha$
- Exemplo – Unilateral
  - √ α = 0,05 e k = 5



df	.25	.20	.15	.10	.05	.025
						Upper-tai
1	1.000	1.376	1.963	3.078	6.314	12.71
2	0.816	1.061	1.386	1.886	2.920	4.303
3	0.765	0.978	1.250	1.638	2.353	3.182
4	0.741	0.941	1.190	1.533	2.132	2.776
5	0.727	0.920	1.156	1.476	$\boxed{2.015}$	2.571
6	0.718	0.906	1.134	1.440	1.943	2.447
7	0.711	0.896	1.119	1.415	1.895	2.365
8	0.706	0.889	1.108	1.397	1.860	2.306
9	0.703	0.883	1.100	1.383	1.833	2.262

$$P\{T_5 > 2,015\} = 0,05$$

- Exemplo – Bilateral:
  - √ α = 0,05 e k = 5

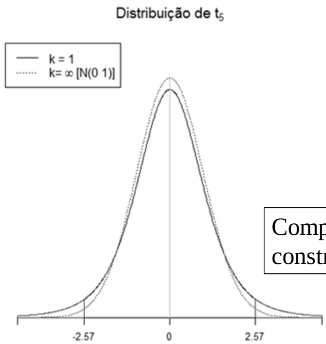


df	.25	.20	.15	.10	.05	.025
						Upper-tail
1	1.000	1.376	1.963	3.078	6.314	12.71
2	0.816	1.061	1.386	1.886	2.920	4.303
3	0.765	0.978	1.250	1.638	2.353	3.182
4	0.741	0.941	1.190	1.533	2.132	2.776
5	0.727	0.920	1.156	1.476	2.015	$\boxed{2.571}$
6	0.718	0.906	1.134	1.440	1.943	2.447
7	0.711	0.896	1.119	1.415	1.895	2.365
8	0.706	0.889	1.108	1.397	1.860	2.306
9	0.703	0.883	1.100	1.383	1.833	2.262

A distribuição t é simétrica!

$$P\{T_5 > 2,57\} = 0,025$$
$$P\{T_5 < -2,57\} = 0,025$$
$$P\{-2,57 \le T_5 \le 2,57\} = 0,95$$

- Exemplo – Normal padrão e t_5:
  - √ α = 0,05 – Bilateral
    - Z ~ N (0, 1) e T ~ t_5.



$$P\{T_5 > 2,57\} = 0,025$$
$$P\{Z > 2,57\} = 0,005$$

$$P\{-2,57 \le T_5 \le 2,57\} = 0,95$$
$$P\{-1,96 \le Z \le 1,96\} = 0,95$$

Compare os intervalos de 95% de confiança construídos com a t e com a Normal

Intervalo de Confiança *t* para μ

√ População normal e variância desconhecida

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

– Não depende dos parâmetros desconhecidos μ e σ!

Então

$$P\left\{-t_{\alpha/2,n-1} \le \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \le t_{\alpha/2,n-1}\right\} = 1 - \alpha$$

sendo $t_{\alpha/2,\,(n-1)}$ o percentil superior com α/2(100)% da t com n–1 graus de liberdade

logo

$$P\left\{\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

Intervalo de Confiança para a Média, Variância Desconhecida

- Seja $\bar{x}$ a média de amostra aleatória, de tamanho n, oriunda de população normal com variância σ^2 desconhecida
  - √ Intervalo com $100(1 - \alpha)\%$ de confiança para μ:

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Limites Unilaterais de Confiança

- São facéis de usar, como no caso dos intervalos de confiança construídos a partir da distribuição normal.
  - √ Limite superior com $(1 - \alpha)100\%$ de confiança para μ.

$$\mu \leq u = \bar{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}$$

  - √ Limite inferior com $(1 - \alpha)100\%$ de confiança para μ.

$$\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} = l \leq \mu$$

Exemplo 8-5

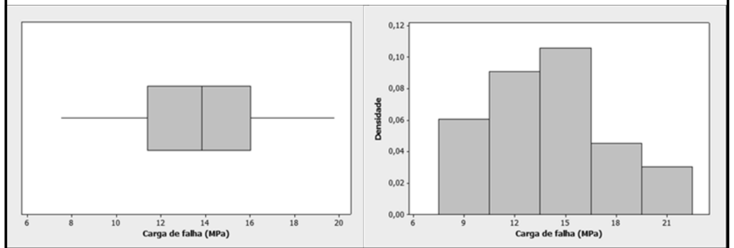
- Testes de adesão de liga U-700
  - √ Medidas de carga no ponto de falha de corpos de prova
  - √ Amostra de tamanho $n = 22$
  - √ Variância populacional desconhecida
  - √ Busca-se intervalo de 95% de confiança para a média populacional
  - √ Dados: *BD_producao.xlsx/adesao*

- Estatísticas descritivas da amostra:

Descriptive Statistics: carga

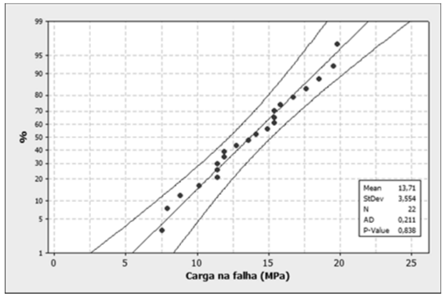
Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
carga	22	13,714	0,758	3,554	7,500	11,400	13,850	16,025	19,800

- Gráficos univariados:



  - √ Os dados aparentam ser normais
  - √ Suposição de normalidade é crucial para estimar μ.

- Gráfico de probabilidade:



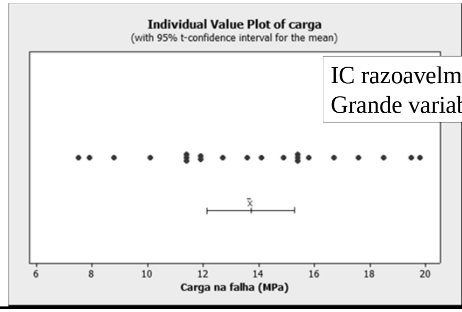
√ O gráfico de probabilidade normal reforça as evidências amostrais sobre a suposição de que a população é normalmente distribuída

- Saída Minitab

Stat > Basic Statistics > 1–Sample t ➔

One-Sample T: carga

Variable	N	Mean	StDev	SE Mean	95% CI
carga	22	13,714	3,554	0,758	(12,138; 15,289)



IC razoavelmente largo
Grande variabilidade nas medidas

- Intervalo t de 95% de confiança para μ:
  - √ Parâmetro da t: 22 – 1 = 21 graus de liberdade

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$13,71 - 2,080\,\frac{3,55}{\sqrt{22}} \leq \mu \leq 13,71 + 2,080\,\frac{3,55}{\sqrt{22}}$$

$$12,14 \leq \mu \leq 15,28$$

  - √ Intervalo é razoavelmente amplo
    - Há grande variabilidade nas medidas do teste tratativo de adesão

Tamanho de Amostra para Construção de Intervalos de Confiança *t*

- Não é fácil selecionar o tamanho da amostra n para obter precisão de estimação especificada
  - √ O comprimento do intervalo envolve s que é desconhecido antes da coleta dos dados
  - √ O percentil t depende do tamanho da amostra n.
- Um n apropriado pode ser obtido apenas através de tentativa e erro
  - √ Os resultados dependerão da confiabilidade da "tentativa" para σ.

Intervalo de Confiança para Variância – População Normal

Distribuição χ^2

√ Seja X_1, X_2, ..., X_n uma amostra aleatória proveniente de população normal, com média Média μ e variância σ^2. Seja S^2 a variância da amostra. Então a variável aleatória

$$X^2 = \frac{(n-1)}{\sigma^2} S^2 \sim \chi^2_{(n-1)}$$

√ tem distribuição qui-quadrado (χ^2) com n – 1 graus de liberdade

Distribuição χ^2 – Função de Densidade

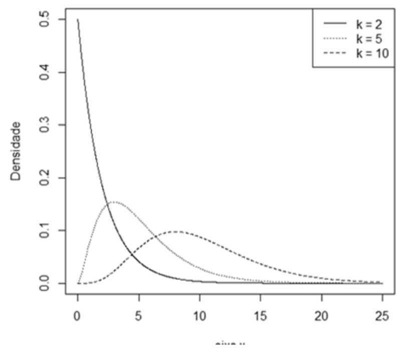
- Seja *X* uma variável aleatória com distribuição χ^2 com k graus de liberdade:

$$f_X(x) = \frac{1}{2^{\frac{k}{2}}\,\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1}\,\mathrm{e}^{-\frac{x}{2}}\,,\; x > 0$$

√ *k* é o número de graus de liberdade

√ Média da distribuição χ^2: $\mathrm{E}[X] = k$

√ Variância da distribuição χ^2: $\mathrm{Var}[X] = 2k$



√ Distribuições unimodais e assimétricas à direita

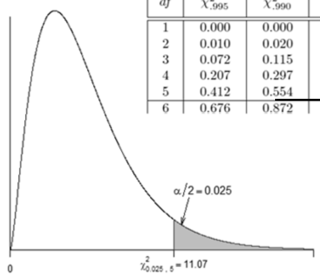
√ À medida que k aumenta, distribuição torna-se mais simétrica

√ Forma limite da χ^2 = normal padrão quando $k \to \infty$

- $\chi^2_{\alpha,\,k}$: Percentis da χ^2:
  - √ Seja a variável aleatória $X^2 \sim \chi^2_k$, então $P\{X^2 > \chi^2_{\alpha,k}\} = \alpha$
- Exemplo – Unilateral
  - √ α = 0,05 e k = 5

Distribuição de χ^2_5

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143
5	0.412	0.554	0.831	1.145	1.610	9.236	$\boxed{11.070}$	12.833
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449

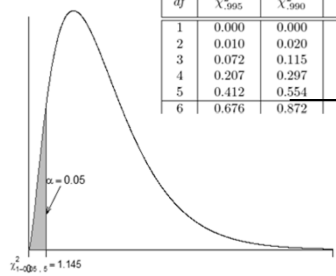


$$P\{\chi^2_5 > 11,070\} = 0,05$$

- $\chi^2_{\alpha,\,k}$: Percentis da χ^2:
  - √ Seja a variável aleatória $X^2 \sim \chi^2_k$, então $P\{X^2 < \chi^2_{1-\alpha,k}\} = \alpha$
- Exemplo – Unilateral
  - √ α = 0,05 e k = 5

Distribuição de χ^2_5

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143
5	0.412	0.554	0.831	$\boxed{1.145}$	1.610	9.236	11.070	12.833
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449



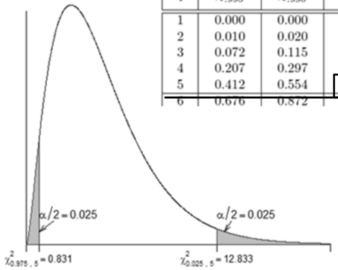
$$P\{\chi^2_5 < 1,145\} = 0,05$$

A distribuição χ^2 não é simétrica!

- Exemplo – Bilateral:
  - √ α = 0,05 e k = 5

Distribuição de χ^2_5

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143
5	0.412	0.554	$\boxed{0.831}$	1.145	1.610	9.236	11.070	$\boxed{12.833}$
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449



A distribuição χ^2 não é simétrica!

$$P\{\chi^2_5 > 12,833\} = 0,025$$
$$P\{\chi^2_5 < 0,831\} = 0,025$$
$$P\{0,831 \le \chi^2_5 \le 12,833\} = 0,95$$

Intervalo de Confiança χ^2 para σ^2

√ População normal e variância desconhecida

$$X^2 = \frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$$

– Não depende dos parâmetros desconhecidos μ e σ!

Então

$$P\left\{\chi^2_{(1-\alpha/2),(n-1)} \le \frac{n-1}{\sigma^2}S^2 \le \chi^2_{\alpha/2,(n-1)}\right\} = 1-\alpha$$

sendo $\chi^2_{\alpha/2,(n-1)}$ o percentil superior com α/2(100)% da χ^2, com n–1 graus de liberdade

logo

$$P\left\{\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,(n-1)}} \le \sigma^2 \le \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,(n-1)}}\right\} = 1-\alpha$$

Intervalo de Confiança para a Variância

- Seja s^2 a variância de amostra aleatória de n observações provenientes de população normal com variância σ^2 desconhecida
  - √ Intervalo com $100(1-\alpha)\%$ de confiança para σ^2:

$$\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,(n-1)}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{(1-\alpha/2),(n-1)}}$$

  - √ Intervalo com $100(1-\alpha)\%$ de confiança para σ:

$$\sqrt{\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,(n-1)}}} \leq \sigma \leq \sqrt{\frac{(n-1)\,s^2}{\chi^2_{(1-\alpha/2),(n-1)}}}$$

Limites Unilaterais de Confiança

- São fáceis de usar, como no caso dos intervalos de confiança construídos anteriormente para μ.
  - √ Limite superior com $(1-\alpha)100\%$ de confiança para σ^2.

$$\sigma^2 \leq u = \frac{(n-1)\,s^2}{\chi^2_{(1-\alpha),(n-1)}}$$

  - √ Limite inferior com $(1-\alpha)100\%$ de confiança para σ^2.

$$\frac{(n-1)\,s^2}{\chi^2_{\alpha,(n-1)}} = l \leq \sigma^2$$

Exemplo 8-6

- Enchimento garrafas de detergente:
  - √ Máquina automática para enchimento de garrafas de detergente
  - √ Amostra aleatória de 20 garrafas
  - √ Variância amostral: $s^2 = 0{,}0153$ (onça fluida)^2.
  - √ Se variância for muito grande, existirá proporção inaceitável de garrafas cujo enchimento não foi completo e cujo enchimento foi em demasia
  - √ Volume de enchimento distribuído de forma aproximadamente normal.

- Intervalo superior de confiança de 95% para σ^2:

$$\sigma^2 \leq \frac{(20-1)\,s^2}{\chi^2_{0{,}95,(20-1)}}$$

$$\sigma^2 \leq \frac{(19)\,0{,}0153}{10{,}117} = 0{,}0287 \text{ (onça fluida)}^2$$

- Intervalo superior de confiança de 95% para σ:

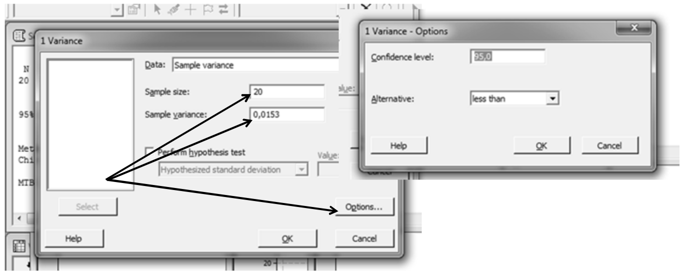
$$\sigma \leq \sqrt{\frac{(19)\,0{,}0153}{10{,}117}}$$

$$\sigma \leq 0{,}17 \text{ onça fluida}$$

  - √ Com um nível de confiança de 95%, os dados indicam que o desvio-padrão do processo poderia ser tão grande quanto 0,17 onça fluida

- Minitab:
  - √ Comandos

Stat > Basic Statistics > 1–Sample t ➔



√ Saída:

```
Test and CI for One Variance

Method

The chi-square method is only for the normal distribution.
The Bonett method cannot be calculated with summarized data.


Statistics

 N  StDev  Variance
20  0,124    0,0153


95% One-Sided Confidence Intervals

            Upper Bound   Upper Bound
Method        for StDev  for Variance
Chi-Square        0,170        0,0287
```

Intervalo de Confiança para Proporção – Amostra Grande

Proporção de População

- Amostra aleatória de tamanho *n* retirada de uma grande (possivelmente infinita) população.
  - √ X: quantidade de observações amostrais que pertencem a uma categoria de interesse ($X \leq n$)
  - √ Estimador pontual da proporção *p* da população:

$$\hat{P} = \frac{X}{n}$$

  - √ *n* e *p* são parâmetros de uma binomial

- Distribuição amostral do estimador $\hat{P}$:

$$\hat{P} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

√ Se *p* não estiver muito perto de 0 ou 1 e se *n* for suficientemente grande

– Na prática, np e n(1 – p) ≥ 5

Intervalo Aproximado de Confiança para *p*

√ Aproximação normal para uma proporção binomial

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{as.}{\approx} \mathrm{N}(0,1)$$

– Se *n* for grande o suficiente

– Depende do parâmetro desconhecido *p*.

Então
$$\mathrm{P}\left\{-z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right\} \approx 1 - \alpha$$

sendo $z_{\alpha/2}$ o percentil superior com α/2(100)% da normal padrão

logo
$$\mathrm{P}\left\{\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\} \approx 1 - \alpha$$

- Infelizmente os limites superior e inferior do intervalo contêm o parâmetro desconhecido *p*!
- Uma solução satisfatória é trocar p por $\hat{P}$

$$\mathrm{P}\left\{\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right\} \approx 1 - \alpha$$

Intervalo de Confiança para uma Proporção Binomial

- Seja $\hat{p}$ a proporção de observações em uma amostra aleatória, de tamanho *n* que pertença a uma classe de interesse

√ Intervalo aproximado com 100(1 – α)% de confiança para *p*:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

√ Requer que *np* e *n*(1 – *p*) ≥ 5

√ Nos casos em que *n* for pequeno , deve-se usar outros métodos (numéricos ou baseados na binomial)

Exemplo 8-7

- Mancais de eixos de manivelas de motores
  - √ Amostra de tamanho 85
  - √ 10 motores da amostra têm acabamento mais rugoso que o especificado
  - √ Estimativa pontual da proporção populacional de mancais não-conformes

$$\hat{p} = \frac{10}{85} = 0,12$$

- Intervalo bilateral de 95% de confiança para *p*:

$$\hat{p} - z_{0,025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0,025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0,12 - 1,96\sqrt{\frac{(0,12)(0,88)}{85}} \leq p \leq 0,12 + 1,96\sqrt{\frac{(0,12)(0,88)}{85}}$$

$$0,05 \leq p \leq 0,19$$

Escolha do Tamanho da Amostra

- Precisão do IC: $2\left(z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$
- Erro ao usar $\hat{P}$ para estimar *p*: $E = |\hat{P} - p|$
- Tamanho da amostra:
  - √ Escolher *n* tal que $z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = E$
  - √ Comprimento do intervalo resultante: 2E

- Tamanho da amostra com erro especificado:
  - √ Se $\hat{p}$ for usada como estimativa de *p*, podemos estar $(1 - \alpha)100\%$ confiantes de que o erro $|\hat{p} - p|$ não excederá o valor E especificado quando o tamanho da amostra for

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

    - Deve ser arredondado para número inteiro
  - √ É necessária uma estimativa para cálculo de *n*:
    - Pode-se usar estimativa $\hat{p}$ de amostra anterior
    - Pode-se utilizar uma amostragem preliminar (piloto)

- O máximo de $p(1 - p)$ dá-se para $p = 0,5$!
  - √ Pode-se usar este fato no cálculo de n

$$n = \left(\frac{z_{\alpha/2}}{\mathrm{E}}\right)^2 (0,25)$$

  - √ É um cálculo conservativo, ou seja, estamos no mínimo 100(1 – a)% confiantes de que o erro em estimar $\hat{p}$ através de p é menor do que E, se o tamanho da amostra for n.

Exemplo 8-8

- Mancais de eixos de manivela (cont. Ex. 8-7)
  - √ Determinar tamanho amostra que o erro de estimação seja menor que 0,05, com uma confiança de 95%

$$n = \left(\frac{z_{0,025}}{\mathrm{E}}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1,96}{0,05}\right)^2 0,12(0,88) \approx 163$$

  - √ Se quisermos estar no **mínimo** 95% confiantes:

$$n = \left(\frac{z_{0,025}}{\mathrm{E}}\right)^2 (0,25) = \left(\frac{1,96}{0,05}\right)^2 (0,25) \approx 385$$

Limites Unilaterais de Confiança

- Pode-se encontrar limites unilaterais aproximados de confiança para p.
  - √ Limite superior com $(1 - \alpha)100\%$ de confiança para p.

$$p \leq u = \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  - √ Limite inferior com $(1 - \alpha)100\%$ de confiança para p.

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = l \leq p$$

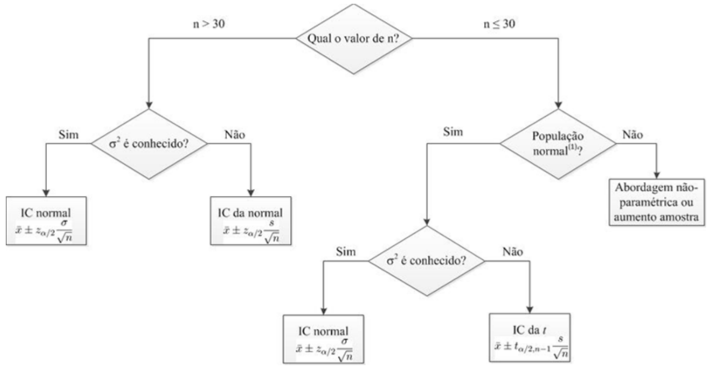
Roteiro para Construção de Intervalos de Confiança

Roteiro

- Etapa primordial:
  - √ Coincidir objetivo do estudo com cálculo apropriado
- Comentários:
  - √ Determine o parâmetro (e a distribuição dos dados) que estará limitado pelo intervalo de confiança ou testado pela hipótese
  - √ Verifique se outros parâmetros são conhecido ou necessários de serem estimados

---

- Intervalo de confiança para a média populacional μ:



√ (1) Normal ou aproximadamente normal (pelo menos unimodal e simétrica)

- Intervalo de confiança para a variância populacional σ^2
  - √ População normal
  - √ Qualquer tamanho amostral

$$\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,(n-1)}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{(1-\alpha/2),(n-1)}}$$

- Intervalo de confiança para a desvio-padrão populacional σ:
  - √ População normal
  - √ Qualquer tamanho amostral

$$\sqrt{\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,(n-1)}}} \leq \sigma \leq \sqrt{\frac{(n-1)\,s^2}{\chi^2_{(1-\alpha/2),(n-1)}}}$$

- Intervalo de confiança para proporção binomial de população *p*:
  - √ Amostra grande
  - √ *p* não muito próximo de 0 ou 1
  - √ np e $n(1 - p) \geq 5$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Intervalos de Tolerância e de Previsão

Intervalo de Previsão para Observação Futura

- Problema:
  - √ Prever uma observação futura de uma variável
  - √ Diferente de estimar a média da variável
  - √ Intervalo de confiança não é apropriado

√ Suponha uma amostra aleatória X_1, X_2, ..., X_n, proveniente de uma população.

√ Deseja-se prever o valor X_{n+1}.

– Uma única observação futura

√ Estimação pontual de X_{n+1}: $\bar{X}$

√ Erro de previsão: $X_{n+1} - \bar{X}$

√ Valor esperado do erro de previsão

$$\mathrm{E}(X_{n+1} - \bar{X}) = \mu - \mu = 0$$

√ Variância do erro de previsão:

$$\mathrm{Var}[X_{n+1} - \bar{X}] = \sigma^2 + \frac{\sigma^2}{n} = \sigma^2\left(1 + \frac{1}{n}\right)$$

- A observação futura é independente da média atual da amostra!

- O erro de previsão é normalmente distribuído:

$$Z = \frac{X_{n+1} - \bar{X}}{\sigma\sqrt{1 + \frac{1}{n}}} \sim \mathrm{N}(0,1)$$

- Se σ^2 é desconhecido, troca-se σ por S, obtendo-se a estatística T:

$$T = \frac{X_{n+1} - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} \sim t_{n-1}$$

Intervalo de Previsão

- Um intervalo de previsão de $(1 - \alpha)(100)\%$ para uma observação futura a partir de uma distribuição normal:

$$\bar{x} - t_{\alpha/2,(n-1)}\, s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{\alpha/2,(n-1)}\, s\sqrt{1+\frac{1}{n}}$$

- Comentários:
  - √ O IP para X_{n+1} será sempre maior que o IC para μ:
    - Há mais variabilidade associada com o erro de previsão
  - √ Erros de estimação:
    - Erro de previsão: $\bar{X} - X_{n+1}$
      (diferença dentre duas variáveis aleatórias)
    - Erro estimação média: $\bar{X} - \mu$
      (diferença entre variável aleatória e constante)
  - √ Quando *n* torna-se grande ($n \to \infty$)
    - Comprimento IC → 0
      (torna-se o valor único de μ)
    - Comprimento IP → $2z_{\alpha/2}\, \sigma$.

√ Sempre haverá incerteza sobre o valor futuro Xn+1.
- Mesmo quando não for necessário estimar qualquer dos parâmetros da distribuição

Exemplo 8-9

- Adesão em uma liga (continuação Ex. 8-5):
  - √ Amostra com n = 22
  - √ Estimativas pontuais: $\bar{x} = 13,71$ $s = 3,55$
  - √ Intervalo com 95% de confiança: $12,4 \le \mu \le 15,28$
  - √ Planeja-se testar um 23º corpo-de-prova
  - √ Intervalo de previsão de 95% para esse corpo-de-prova

$$\bar{x} - t_{0,025,(n-1)}\, s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{0,025,(n-1)}\, s\sqrt{1+\frac{1}{n}}$$

$$13,71 - (2,080)(3,55)\sqrt{1+\frac{1}{22}} \le X_{23} \le 13,71 + (2,080)(3,55)\sqrt{1+\frac{1}{22}}$$

$$6,16 \le X_{23} \le 21,26$$

- Intervalo de previsão é consideravelmente maior que o IC

Intervalo de Tolerância

- Seja uma população normal com média μ e variância σ^2, desconhecidas.
- Deseja-se estabelecer um intervalo que englobe β% dessa população:
  - √ Ex.: o intervalo [μ – 1,96 σ, μ + 1,96 σ] engloba β=95% da população
  - √ Ele é denominado **intervalo de tolerância**.
  - √ Mas, e se μ e σ^2 forem desconhecidos?

- Se se μ e σ^2 forem desconhecidos,
  - √ Calcular média e desvio-padrão amostral da amostra de tamanho n
  - √ O intervalo $[\bar{x} - 1,96s, \bar{x} + 1,96s]$ conterá 95% dos valores da população?
  - √ Devido à variabilidade amostral é provável que esse intervalo contenha menos de 95% dos valores da população.
  - √ Solução:
    - – Trocar 1,96 por algum valor k de maneira que a proporção da distribuição contida no intervalo de de 95% ter algum nível de confiança

Desenvolvimento Teórico

- Seja uma amostra aleatória X_1, X_2, ..., X_n, proveniente de uma população normal com média μ e variância σ^2, desconhecidas. Sua função de distribuição acumulada é F.
- Sejam as estatísticas:

$$L = L(X_1, X_2, \ldots, X_n)$$
$$U = U(X_1, X_2, \ldots, X_n)$$
$$L < U$$

- Assuma que L e U são tais que

$$\mathrm{P}\{F(U) - F(L) \geq \beta\} \geq \gamma \text{ , com } \beta, \gamma \in (0,1)$$

- [L, U] é um intervalo de tolerância que engloba 100xβ% da população com uma confiança γ.

- Tomam-se duas estatísticas tais que:

$$L = L(X_1, X_2, \ldots, X_n) = \bar{x} - ks$$
$$U = U(X_1, X_2, \ldots, X_n) = \bar{x} + ks$$

- Não é trivial encontrar o valor de k dados $\bar{x}$, s, β e γ!
- Howe, 1969 propôs a determinação de k (bilateral) através da expressão:

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z^2_{(1-\beta)/2}}{\chi^2_{(\gamma),(n-1)}}}$$

  - √ em que $z_{1-\beta/2}$ é o percentil superior (1 – β/2)% da distribuição normal e $\chi^2_{g,\, n-1}$ é o percentil superior γ% da distribuição $\chi^2_{\gamma,\, n-1}$.

- Os valores de *k* podem ser encontrados na Tabela XII do Apêndice, para γ = 90%, 95% e 99% de confiança:

Table XI Factors for Tolerance Intervals

Values of *k* for Two-Sided Intervals

Confidence Level	0.90			0.95			0.99		
Percent Coverage	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
2	15.978	18.800	24.167	32.019	37.674	48.430	160.193	188.491	242.300
3	5.847	6.919	8.974	8.380	9.916	12.861	18.930	22.401	29.055
4	4.166	4.943	6.440	5.369	6.370	8.299	9.398	11.150	14.527
5	3.949	4.152	5.423	4.275	5.079	6.634	6.612	7.855	10.260
6	3.131	3.723	4.870	3.712	4.414	5.775	5.337	6.345	8.301
7	2.902	3.452	4.521	3.369	4.007	5.248	4.613	5.488	7.187
8	2.743	3.264	4.278	3.136	3.732	4.891	4.147	4.936	6.468
9	2.626	3.125	4.098	2.967	3.532	4.631	3.822	4.550	5.966
10	2.535	3.018	3.959	2.839	3.379	4.433	3.582	4.265	5.594
11	2.463	2.933	3.849	2.737	3.259	4.277	3.397	4.045	5.308
12	2.404	2.863	3.758	2.655	3.162	4.150	3.250	3.870	5.079
13	2.355	2.805	3.682	2.587	3.081	4.044	3.130	3.727	4.893
14	2.314	2.756	3.618	2.529	3.012	3.955	3.029	3.608	4.737
15	2.278	2.713	3.562	2.480	2.954	3.878	2.945	3.507	4.605

Exemplo 8-10

- Adesão em uma liga (continuação Ex. 8-9):
  - √ Amostra com n = 22
  - √ Estimativas pontuais: $\bar{x} = 13,71$ $s = 3,55$
  - √ Intervalo de tolerância para a carga na falha que inclua 90% dos valores da população, com confiança de 95%:

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z^2_{(1-0,90)/2}}{\chi^2_{(0,95),(n-1)}}}$$

$$= \sqrt{\frac{(22-1)\left(1+\frac{1}{22}\right)(1,645)^2}{11,591}}$$

$$k = 2,2639$$

- Valor de k pode ser obtido na Tabela XII do Apêndice para n = 22, β = 90% e γ = 95%

Table XI Factors for Tolerance Intervals

Values of *k* for Two-Sided Intervals

Confidence Level	0.90			0.95			
Percent Coverage	0.90	0.95	0.99	0.90	0.95	0.99	0.90
2	15.978	18.800	24.167	32.019	37.674	48.430	160.193
3	5.847	6.919	8.974	8.380	9.916	12.861	18.930
4	4.166	4.943	6.440	5.369	6.370	8.299	9.398
5	3.949	4.152	5.423	4.275	5.079	6.634	6.612
19	2.172	2.588	3.399	2.337	2.784	3.656	2.703
20	2.152	2.564	3.368	2.310	2.752	3.615	2.659
21	2.135	2.543	3.340	2.286	2.723	3.577	2.620
22	2.118	2.524	3.315	$\boxed{2.264}$	2.697	3.543	2.584
23	2.103	2.506	3.292	2.244	2.673	3.512	2.551
24	2.089	2.489	3.270	2.225	2.651	3.483	2.522
25	2.077	2.474	3.251	2.208	2.631	3.457	2.494

- Intervalo desejado de tolerância:

$$[\bar{x} - ks\,;\bar{x} + ks]$$

$$[13,71 - (2,264)(3,55)\,;13,71 + (2,264)(3,55)]$$

$$[5,67\,;21,74]$$

  - √ Podemos estar 95% confiantes de que no mínimo 90% dos valores de carga na falha para essa liga estão entre 5,67 e 21,74 MPa

- Quando $n \to \infty$:
  - √ $k \to z_{(1-\gamma)/2}$.
  - √ IP se aproxima do intervalo de tolerância que contém $100x(1 - \alpha)\%$ da distribuição
- O pacote tolerance do R fornece um conjunto de funções para estimar e plotar os limites de tolerância de uma ampla gama de distribuições contínuas e discretas

Gráficos de Probabilidade

- Seção 6.6 – pág. 134

- Como saber se uma distribuição de probabilidades é um modelo razoável para os dados?
  - √ Pode-se fazer uma verificação de suposições:
    - Forma da distribuição, frequência esperada das observações
- Verificação gráfica:
  - √ Histogramas
    - Dão uma ideia da forma da distribuição,
    - Em geral não são indicadores confiáveis (a menos que o tamanho amostral seja grande)

- Gráfico de probabilidades:
  - √ Procedimento geral é simples
  - √ Mais confiável que histograma para tamanhos amostrais pequenos ou moderados
  - √ Usa eixos especiais, projetados para a distribuição hipotética

- Procedimento:
  - √ Ordenação das observações amostrais:
    - $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.
  - √ Plotam-se os pontos $(x_{(j)}; (j - 0{,}5)/n)$
    (observação, frequência acumulada observação)
  - √ Usa-se uma escala de probabilidade
  - √ Distribuição descreve adequadamente os dados:
    - pontos cairão, aproximadamente, ao longo de uma linha reta
  - √ Modelo hipotético não é apropriado
    - – os pontos desviam-se significativamente de uma linha reta
  - √ É subjetivo determinar se os pontos seguem ou não uma linha reta!

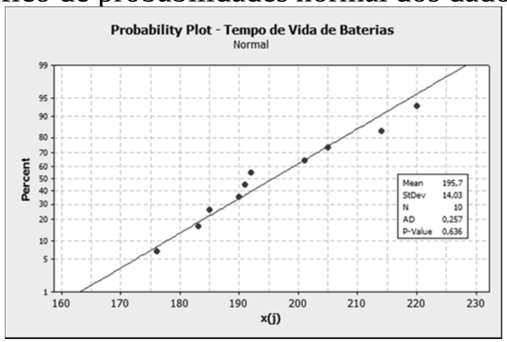
Exemplo 6-7

- Observações sobre tempo de vida de bateria (min.)
  - √ O modelo normal é adequado aos dados?

j	$x_{(j)}$	(j – 0,5)/10	z_j
1	176	0,05	–1,64
2	183	0,15	–1,04
3	185	0,25	–0,67
4	190	0,35	–0,39
5	191	0,45	–0,13
6	192	0,55	0,13
7	201	0,65	0,39
8	205	0,75	0,67
9	214	0,85	1,04
10	220	0,95	1,67

$$\frac{j - 0,5}{n} = \mathrm{P}\{Z \leq z_j\} = \Phi(z_j)$$

- Gráfico de probabilidades normal dos dados:



  - √ Ser mais influenciado pelos pontos do meio que pelos dos extremos
  - √ Eixo y com escala de probabilidades (escala z)

- Pode ser também construído como:
  - √ $x_{(j)}$ vs. escores padronizados (z_j):

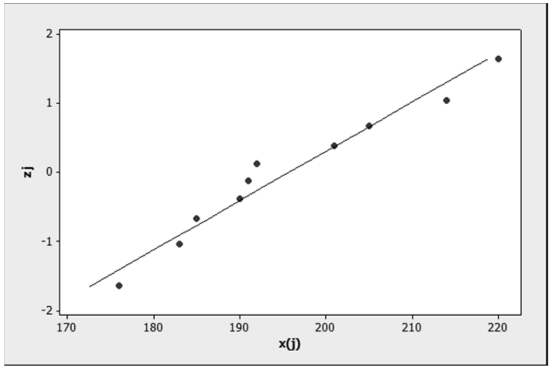
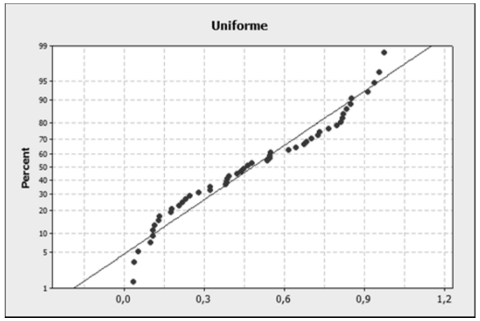


Gráfico de Probabilidades Normal

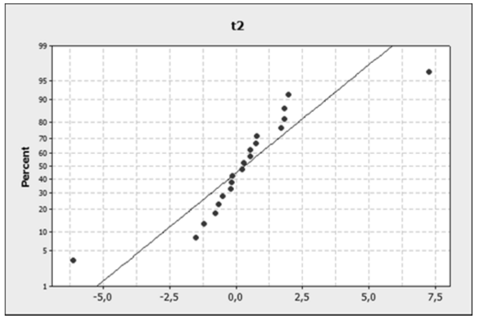
- Pode ser útil na identificação de distribuições que sejam simétricas mas que tenham caudas mais pesadas (ou mais leves) que a normal

- Distribuição de cauda leve



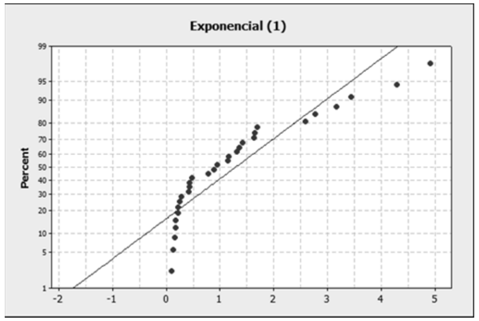
  - √ Pontos à esquerda tendem a ficar abaixo da linha e à direita tendem a ficar acima
    - As menores e maiores observações não serão tão extremas como se esperaria de uma normal

- Distribuição de cauda pesada



  - √ Pontos à esquerda tendem a ficar acima da linha e à direita tendem a ficar abaixo
  - √ Gráfico em forma de S

- Distribuição assimétrica



√ Pontos de ambas as extremidades tendem a estar abaixo da linha

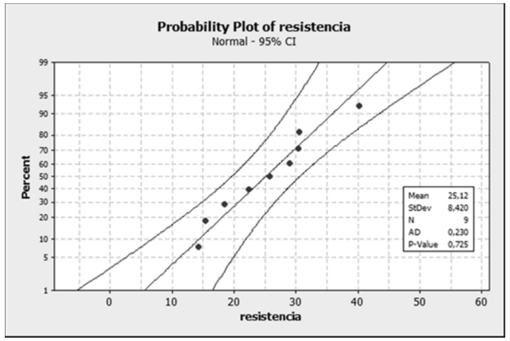
√ Gráfico tem forma curvada

Exemplos de Aplicação

Exemplos 8-81 e 8.84

- Resistência do concreto à compressão
  - √ Mistura com cinza
  - √ Tamanho amostral: 9

a) Gráfico de probabilidade dos dados: normal



b) Intervalo unilateral inferior com 99% de confiança

One-Sample T: resistencia

Variable	N	Mean	StDev	SE Mean	99% Lower Bound
resistencia	9	25,12	8,42	2,81	16,99

√ $t_{0,01;\,8}$ = 2,896

$$\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} = l \leq \mu$$
$$25,12 - 2,89646\frac{8,42}{\sqrt{9}} \leq \mu$$
$$16,99 \leq \mu$$

c) Intervalo bilateral com 98% de confiança

One-Sample T: resistencia

Variable	N	Mean	StDev	SE Mean	98% CI
resistencia	9	25,12	8,42	2,81	(16,99; 33,25)

√ $t_{0,01;\,8}$ = 2,896

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$
$$13,71 - 2,896\,\frac{8,42}{\sqrt{9}} \leq \mu \leq 13,71 + 2,896\,\frac{8,42}{\sqrt{9}}$$
$$16,99 \leq \mu \leq 33,25$$

d) Intervalo unilateral superior com 99% de confiança – variância

Test and CI for One Variance: resistencia

Method

The chi-square method is only for the normal distribution.

Statistics

Variable	N	StDev	Variance
resistencia	9	8,42	70,9

99% One-Sided Confidence Intervals

Variable	Method	Upper Bound for StDev	Upper Bound for Variance
resistencia	Chi-Square	18,56	344,5

√ $\chi^2_{0,99;\,8}$ = 1,647

$$\sigma^2 \leq \frac{(9-1)\,s^2}{\chi^2_{0,99,(9-1)}}$$
$$\sigma^2 \leq \frac{(8)\,70,90194}{1,64650} = 344,498\ \text{MPa}^2$$

e) Intervalo bilateral com 98% de confiança – variância

Test and CI for One Variance: resistencia

Method

The chi-square method is only for the normal distribution.

Statistics

Variable	N	StDev	Variance
resistencia	9	8,42	70,9

98% Confidence Intervals

Variable	Method	CI for StDev	CI for Variance
resistencia	Chi-Square	(5,31; 18,56)	(28,2; 344,5)

√ $\chi^2_{0,99;\,8}$ = 1,647

√ $\chi^2_{0,01;\,8}$ = 20,090

$$\frac{(9-1)\,s^2}{\chi^2_{0,01,(9-1)}} \leq \sigma^2 \leq \frac{(9-1)\,s^2}{\chi^2_{(0,99),(9-1)}}$$
$$\frac{(8)\,70,90194}{20,0902} \leq \sigma^2 \leq \frac{(8)\,70,90194}{1,64650}$$
$$28,233 \leq \sigma^2 \leq 344,498\ \text{MPa}^2$$

f) Intervalos bilaterais com 98% de confiança

√ Média

$\bar{x} = 22,922$

$s^2 = 39,827$

$s = 6,311$

```
One-Sample T

N   Mean  StDev  SE Mean         98% CI
9  22,92   6,31     2,10  (16,83; 29,02)
```

√ Variância

```
Test and CI for One Variance

Method

The chi-square method is only for the normal distribution.


Statistics

N  StDev  Variance
9   6,31      39,8

98% Confidence Intervals

                CI for          CI for
Method          StDev         Variance
Chi-Square  (3,98; 13,91)  (15,9; 193,5)
```

g) Intervalos bilaterais com 98% de confiança

√ Média

$\bar{x} = 25,011$

$s^2 = 70,844$

$s = 8,417$

```
One-Sample T: C8

Variable  N   Mean  StDev  SE Mean         98% CI
C8        9  25,01   8,42     2,81  (16,88; 33,14)
```

√ Variância

```
Test and CI for One Variance: C8

Method

The chi-square method is only for the normal distribution.


Statistics

Variable  N  StDev  Variance
C8        9   8,42      70,8

98% Confidence Intervals

                             CI for          CI for
Variable  Method             StDev         Variance
C8        Chi-Square  (5,31; 18,55)  (28,2; 344,2)
```

Exercício 8.91

- Investigação sobre erros em cabeamentos
  - √ Amostra: 1600 aviões
  - √ Aviões com problemas cabeamento: 8 (sucessos)

a) Intervalo de confiança para proporção $\hat{p} = \frac{8}{1600} = 0,005$

```
MTB > POne 1600 8;
SUBC>   Confidence 99;
SUBC>   Use2.
```

```
Test and CI for One Proportion

Sample  X     N  Sample p              99% CI
1       8  1600  0,005000  (0,000458; 0,009542)

Using the normal approximation.
```

$$\hat{p} - z_{0,005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{0,005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0,005 - 2,576\sqrt{\frac{(0,005)(0,995)}{1600}} \le p \le 0,12 + 2,576\sqrt{\frac{(0,005)(0,995)}{1600}}$$

$$0,000459 \le p \le 0,009542$$

b) Tamanho amostral – estimativa preliminar

$$n = \left(\frac{z_{0,005}}{\mathrm{E}}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{2,576}{0,008}\right)^2 0,005(0,995) \approx 516$$

√ Os tamanhos amostrais serão diferentes se calculados por meio da distribuição exata (binomial)

- Probabilidade de sucesso é muito baixa

- Calcular tamanho amostral pela distribuição exata
- Checar saída do minitab

b) Tamanho amostral – estimativa conservativa

√ No mínimo 99% confiantes

$$n = \left(\frac{z_{0,005}}{\mathrm{E}}\right)^2 (0,25) = \left(\frac{2,576}{0,008}\right)^2 (0,25) \approx 25921$$

√ Saída Minitab

Stat > Power and Sample Size > Sample Size for Estimation

Sample Size for Estimation

```
Method

Parameter            Proportion
Distribution         Binomial
Proportion           0,5
Confidence level     99%
Confidence interval  Two-sided


Results

  Margin  Sample
of Error    Size
   0,008   26039
```

Referências

Bibliografia Recomendada

- Montgomery, D. C. (LTC)
  Estatística Aplicada e Probabilidade para Engenheiros
- Pinheiro, J. I. D et al. (Campus)
  Probabilidade e Estatística: Quantificando a Incerteza