Estatística aplicada

Tópicos especiais em Estatística Aplicada

Prof. Celso J. Munaro (cjmunaro@gmail.com)

VI - Intervalo de confiança para uma amostra

Caps 7 e 8 de [1]

Estatística aplicada

Tópicos especiais em Estatística Aplicada

Prof. Celso J. Munaro (cjmunaro@gmail.com)

VI - Intervalo de confiança para uma amostra

Caps 7 e 8 de [1]

Introdução

A inferência estatística pode ser dividida em duas áreas principais: estimativa de parâmetros e teste de hipóteses.

Exemplo de estimativa de parâmetros: medição da resistência à tração de um componente usado em um chassi de automóvel. Como a variabilidade na resistência à tração está naturalmente presente entre os componentes individuais devido a diferenças nos lotes de matérias-primas, processos de fabricação e procedimentos de medição, interessa estimar a resistência à tração média dos componentes.

Na prática, serão usados dados de amostra para calcular um número que é, em certo sentido, um valor razoável (ou suposição) da média verdadeira. Esse número é chamado de estimativa pontual. Veremos que é possível estabelecer a precisão da estimativa.

Introdução

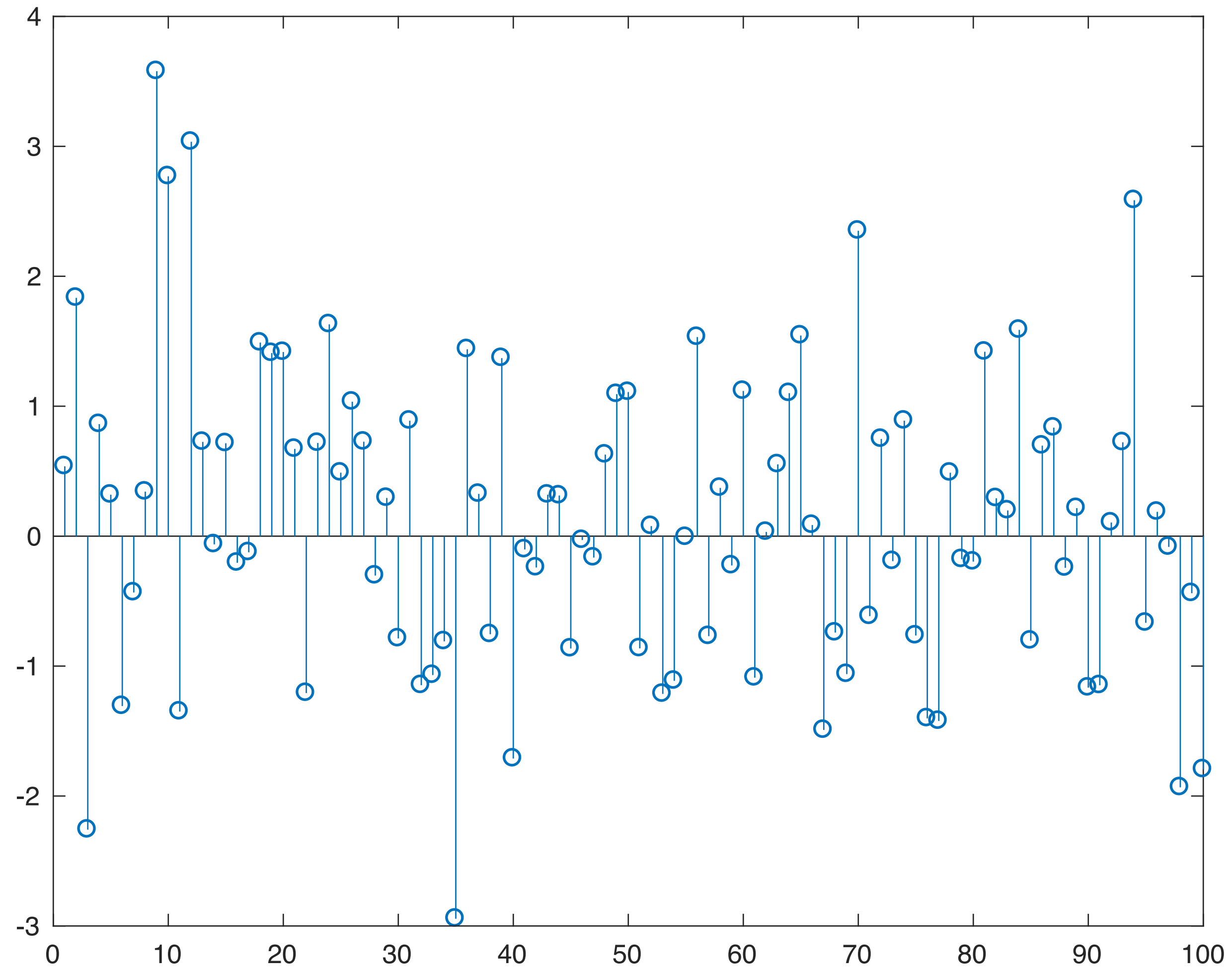
Se X é uma variável aleatória com distribuição $f(x)$, caracterizada por um parâmetro desconhecido θ, e se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de tamanho n de X, então $\hat{\Theta} = h(X_1, X_2, \ldots X_n)$ é chamado um estimador pontual de θ.

Observe que $\hat{\Theta}$ é uma variável aleatória, pois é função de uma variável aleatória.

Assim que as amostras de X são selecionadas, $\hat{\Theta}$ assume um valor particular $\hat{\theta}$, chamado de estimativa de θ.

Introdução

Exemplo: 100 amostras de uma variável aleatória com distribuição $N(\mu, \sigma)$



$$\mu(X) = 0.1231$$

$$\mu(X(1:50)) = 0.2840$$

$$\mu(X(51:100)) = -0.0379$$

Introdução

Problemas usuais de interesse em estimação pontual:

- A média μ de uma população
- O desvio padrão σ de uma população
- A diferença da média entre duas populações $\mu_1 - \mu_2$

Introdução

Definição

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \tag{7-1}$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \tag{7-2}$$

is called the **bias** of the estimator $\hat{\Theta}$.

Introdução

Exemplo 7.1

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\overline{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

If $\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\overline{X}) = \mu \qquad (5\text{-}40a)$$

Mean is an unbiased estimator

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\overline{X}) = \frac{\sigma^2}{p} \qquad (5\text{-}40b)$$

Introdução

Exemplo 7.1. Sample variance

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1}E\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1}E\sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2) - nE(\overline{X}^2)\right]$$

Introdução

Exemplo 7.1. Sample variance

The last equality follows from Equation 5-37 in Chapter 5. However, since $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

$$\begin{aligned} E(S^2) &= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2) - n(\mu^2+\sigma^2/n)\right] \\ &= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) \\ &= \sigma^2 \end{aligned}$$

A variância amostral S^2 é uma estimativa não polarizada da variância σ^2

Introdução

A média amostral é uma estatística; isto é, é uma variável aleatória que depende dos resultados obtidos em cada amostra em particular. Como uma estatística é uma variável aleatória, ela possui uma distribuição de probabilidade.

Definição:

The probability distribution of a statistic is called a **sampling distribution.**

Introdução

Distribuição amostral da média

Vamos determinar a distribuição amostral da média amostral $\bar{X}$

Suponha que uma amostra aleatória de tamanho n seja retirada de uma população normal com média μ e variância σ^2.

Cada observação desta amostra, $X_1, X_2, \ldots, X_3$ é uma variável aleatória independente com distribuição normal e média μ e variância σ^2.

Introdução

Distribuição amostral da média

A média amostral $$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

tem uma distribuição normal de média $$\mu_{\overline{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

e variância $$\sigma^2_{\overline{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Introdução

Se estivermos amostrando uma população que tem uma distribuição de probabilidade desconhecida, a distribuição amostral da média da amostra ainda será aproximadamente normal com a média μ e variância σ^2/n, se n for grande.

Introdução

Teorema do limite central:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

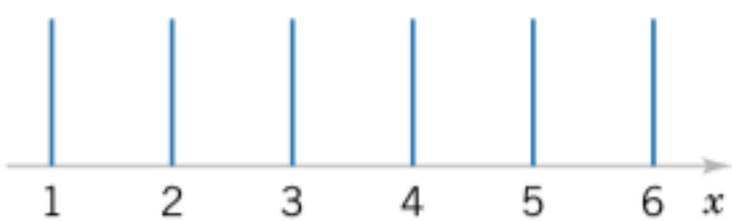
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7-6}$$

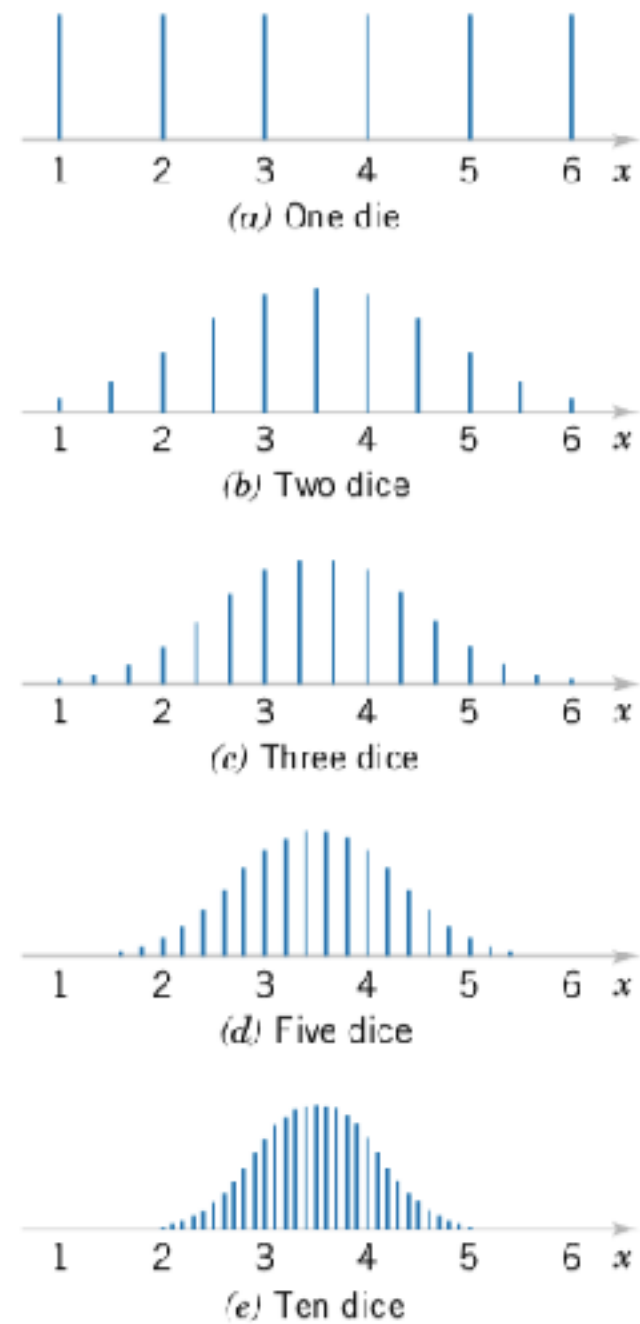
as $n \to \infty$, is the standard normal distribution.

Introdução

Teorema do limite central:
exemplo jogando de 1 a 6 dados simultaneamente e calculando a média do resultado

Para um dado, a distribuição é uniforme





Introdução

Exemplo 7.13

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Introdução

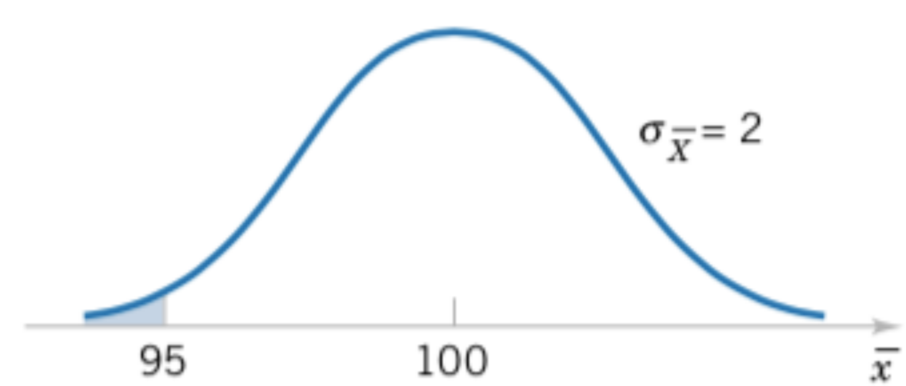
Exemplo 7.13

Note that the sampling distribution of X is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-7. Standardizing the point $\overline{X} = 95$ in Fig. 7-7, we find that

$$z = \frac{95 - 100}{2} = -2.5$$



and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$

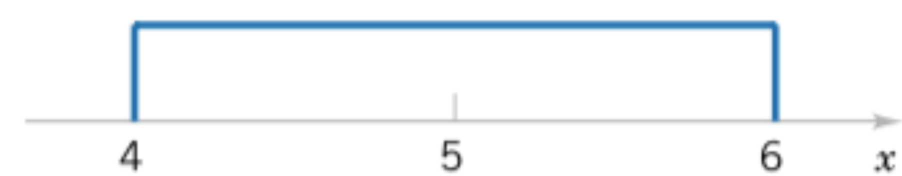
Introdução

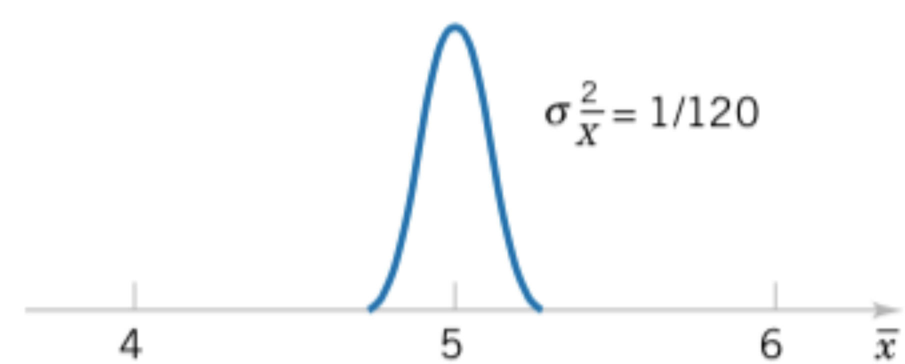
Exemplo 7.14

Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.





The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6 - 4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of $\overline{X}$ is approximately normal with mean $\mu_{\overline{X}} = 5$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n = 1/[3(40)] = 1/120$. The distributions of X and $\overline{X}$ are shown in Fig. 7-8.

Introdução

Definição

If we have two independent populations with means μ_1 and μ_2 and variances σ_2^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}9)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

Introdução

Exemplo 7.15

The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. What is the probability that the difference in the two sample means $\overline{X}_2 - \overline{X}_1$ is at least 25 hours? Assume that the old and improved processes can be regarded as independent populations.

Introdução

Exemplo 7.15

X_1 : distribuição normal com média $\mu = 5000$h e desvio padrão $\sigma_1/\sqrt(n_1) = 40/\sqrt(16) = 10$h

X_2 : distribuição normal com média $\mu = 5050$h e desvio padrão $\sigma_2/\sqrt(n_2) = 30/\sqrt(25) = 6$h

$X_1 - X_2$: distribuição normal

Média $\mu_1 - \mu_2 = 5000 - 5050 = 50$h

Variância $\sigma_2^2/n_2 + \sigma_1^2/n_1 = 6^2 + 10^2 = 136$h

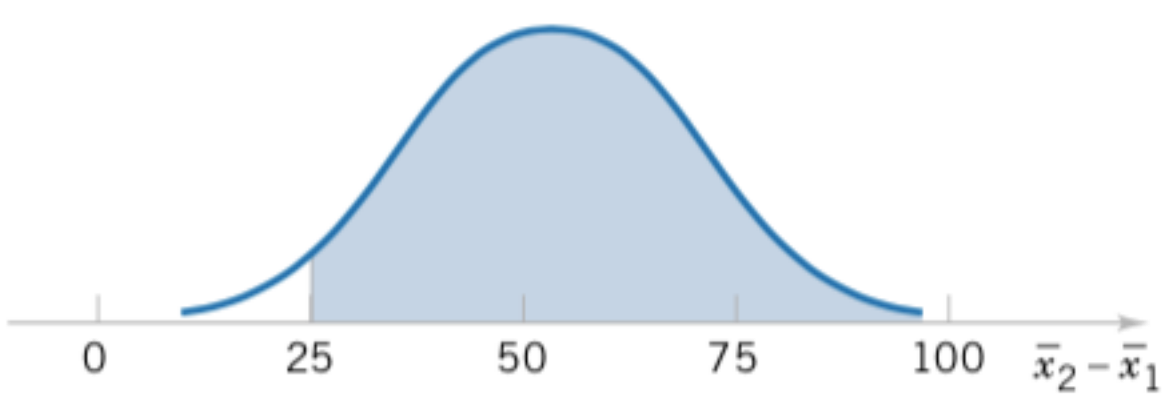
Introdução

Exemplo 7.15

$X_1 - X_2$: distribuição normal

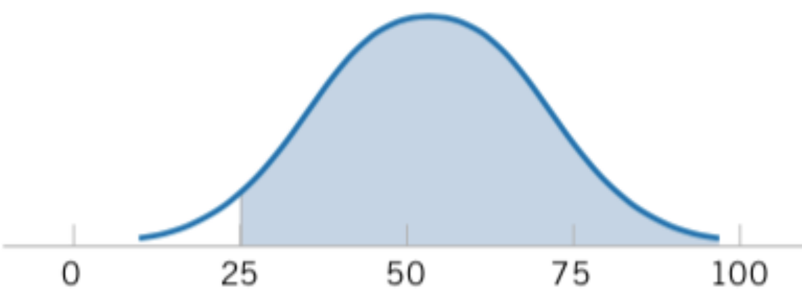
Média $\mu_1 - \mu_2 = 5000 - 5050 = 50$h

Variância $\sigma_2^2/n_2 + \sigma_1^2/n_1 = 6^2 + 10^2 = 136$h



Introdução

Exemplo 7.15



Corresponding to the value $\bar{x}_2 - \bar{x}_1 = 25$ in Fig. 7-9, we find that

$$z = \frac{25 - 50}{\sqrt{136}} = -2.14$$

and we find that

$$\begin{aligned} P(\bar{X}_2 - \bar{X}_1 \geq 25) &= P(Z \geq -2.14) \\ &= 0.9838 \end{aligned}$$

Intervalo de confiança

Suponha que a média de uma variável aleatória seja μ. Ao amostrar esta variável, obtemos uma estimativa $\hat{\mu}$ desta média.

Quão próxima esta estimativa é de μ ?

Em que intervalo de valores se espera que a média estimada esteja?

Intervalo de confiança

Uma estimativa do intervalo onde o parâmetro de uma população possa estar é denominado **intervalo de confiança.**

Não podemos ter certeza de que o intervalo contém o parâmetro verdadeiro e desconhecido da população - usamos apenas uma amostra da população completa para calcular a estimativa pontual e o intervalo.

Entretanto, o intervalo de confiança é construído para que tenhamos alta confiança de que ele contém o parâmetro de população desconhecido.

Intervalo de confiança para média de uma distribuição normal, variância conhecida

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. From the results of Chapter 5 we know that the sample mean $\overline{X}$ is normally distributed with mean μ and variance σ^2/n. We may **standardize** $\overline{X}$ by subtracting the mean and dividing by the standard deviation, which results in the variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{8-3}$$

Now Z has a standard normal distribution.

Intervalo de confiança para média de uma distribuição normal, variância conhecida

A **confidence interval** estimate for μ is an interval of the form $l \leq \mu \leq u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

$$P\{L \leq \mu \leq U\} = 1 - \alpha \qquad (8\text{-}4)$$

where $0 \leq \alpha \leq 1$. There is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting **confidence interval** for μ is

$$l \leq \mu \leq u \qquad (8\text{-}5)$$

The end-points or bounds l and u are called the **lower-** and **upper-confidence limits,** respectively, and $1 - \alpha$ is called the **confidence coefficient.**

Intervalo de confiança para média de uma distribuição normal, variância conhecida

In our problem situation, because $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution, we may write

$$P\left\{-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - \alpha$$

Now manipulate the quantities inside the brackets by (1) multiplying through by $\sigma/\sqrt{n}$, (2) subtracting $\overline{X}$ from each term, and (3) multiplying through by -1. This results in

$$P\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \qquad (8\text{-}6)$$

Intervalo de confiança para média de uma distribuição normal, variância conhecida

Definição

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \tag{8-7}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Intervalo de confiança para média de uma distribuição normal, variância conhecida

Exemplo 8.1

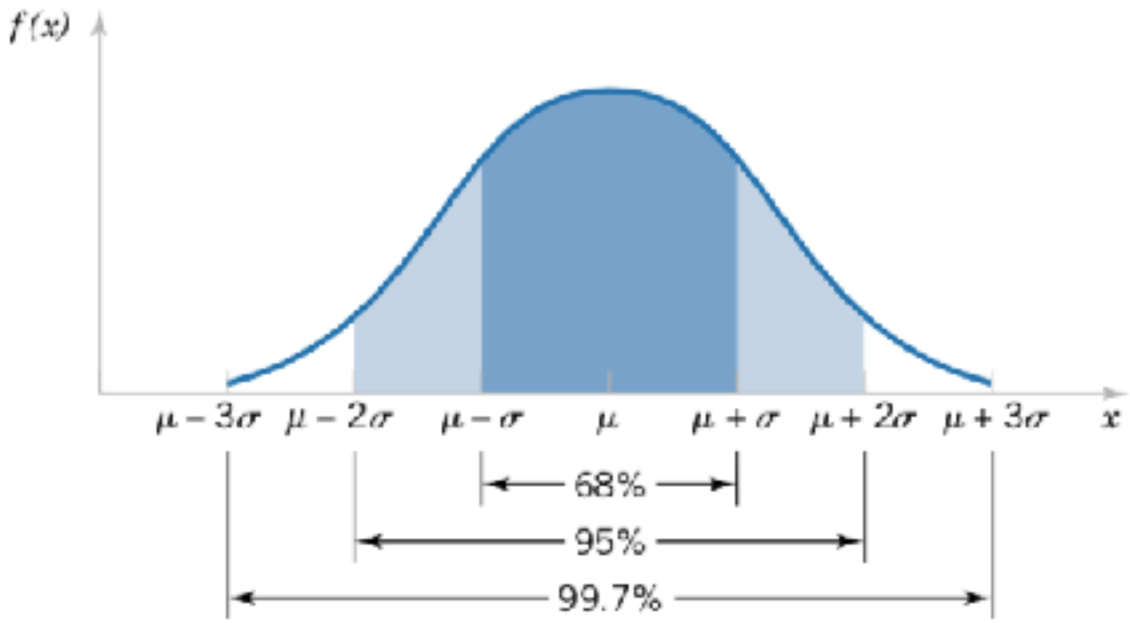
ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60ºC are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1J$.

Queremos obter um intervalo para a média μ do impacto da energia, com nível de confiança de 95%

Intervalo de confiança para média de uma distribuição normal, variância conhecida

Exemplo 8.1

Queremos obter um intervalo para a média μ do impacto da energia com nível de confiança de 95%.



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

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$\bar{x} = 64.46 \qquad \sigma = 1$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \leq \mu \leq 65.08$$

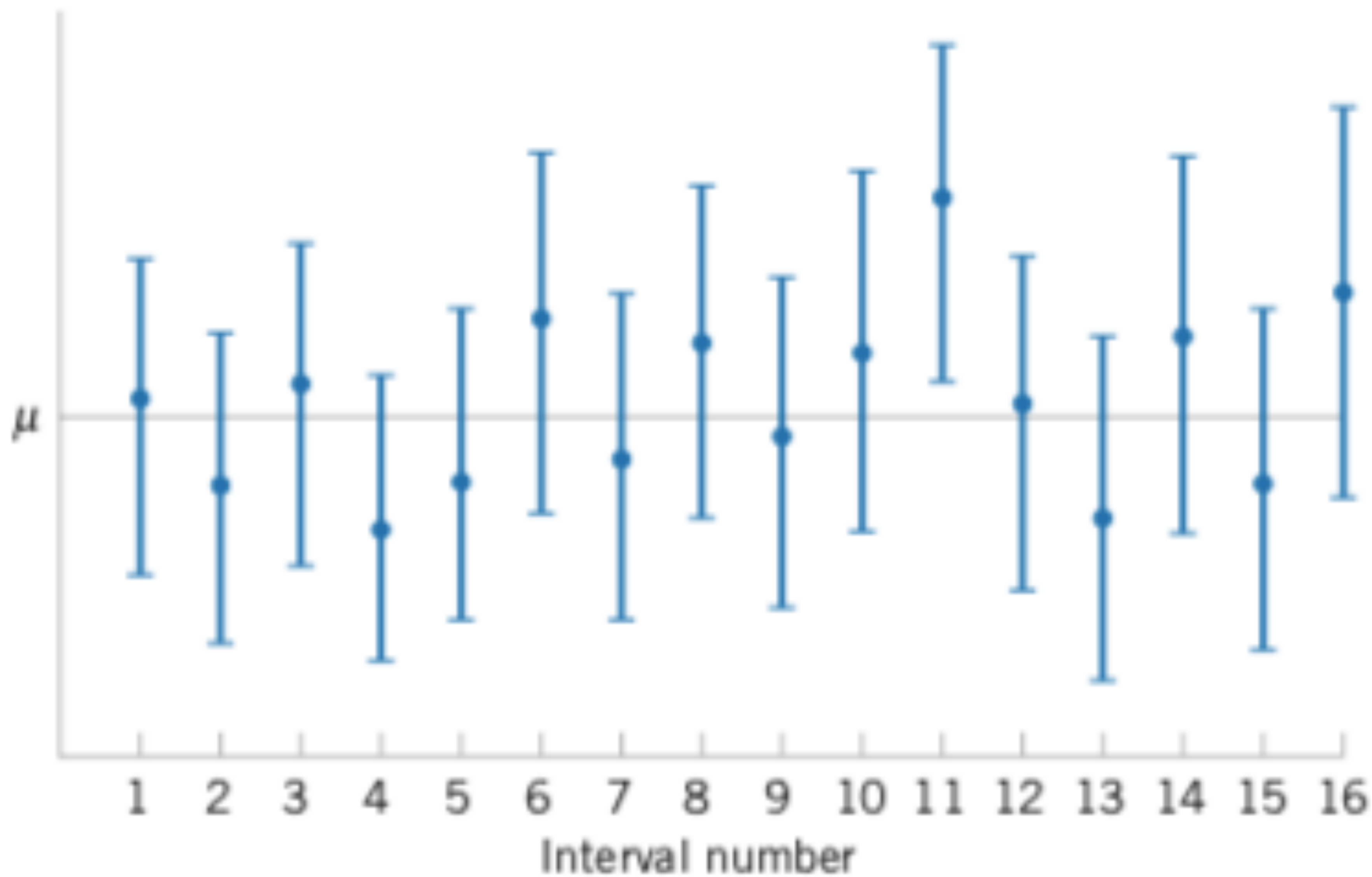
Interpretação do intervalo de confiança

O intervalo de confiança é obtido a partir de amostragem e seus limites são variáveis aleatórias. É incorreto afirmar que a probabilidade da média estar no intervalo é de 95% .

A interpretação correta é de um intervalo de confiança da média μ com nível de confiança $100(1-\alpha)\,\%$ é: se um número infinito de amostras for coletado, $100(1-\alpha)\,\%$ dos intervalos vai conter a média real da população μ.

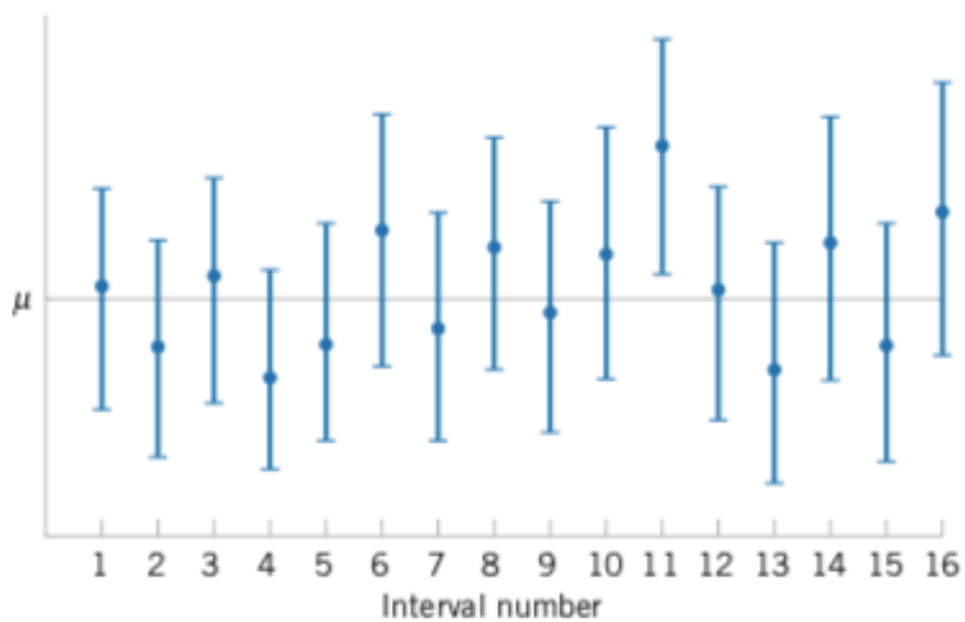
Interpretação do intervalo de confiança

$100(1-\alpha)\,\%$ dos intervalos vai conter a média real da população μ.



Interpretação do intervalo de confiança

$100(1-\alpha)\,\%$ dos intervalos vai conter a média real da população μ.



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

Aumentar o nível de confiança aumenta o tamanho do intervalo.

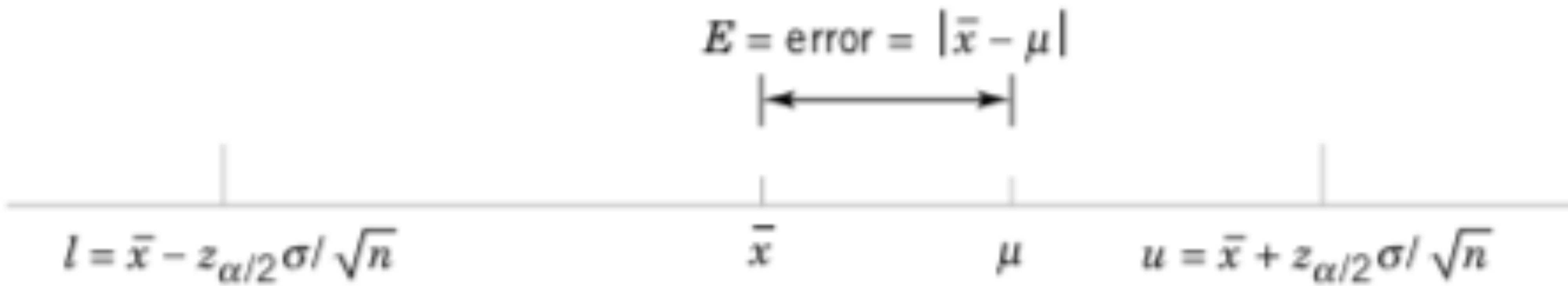
Para $100(1-\alpha) = 99\,\%$, $z_{\alpha/2} = z_{0.025} = 2.58$

Interpretação do intervalo de confiança

O comprimento de um intervalo de confiança é uma medida da precisão da estimativa. A precisão é inversamente proporcional ao nível de confiança.

É desejável obter um intervalo de confiança que seja curto o suficiente para fins de tomada de decisão e que também tenha confiança adequada.

Uma maneira de conseguir isso é escolher o tamanho da amostra n grande o suficiente para fornecer um IC de comprimento ou precisão especificados com a confiança prescrita.

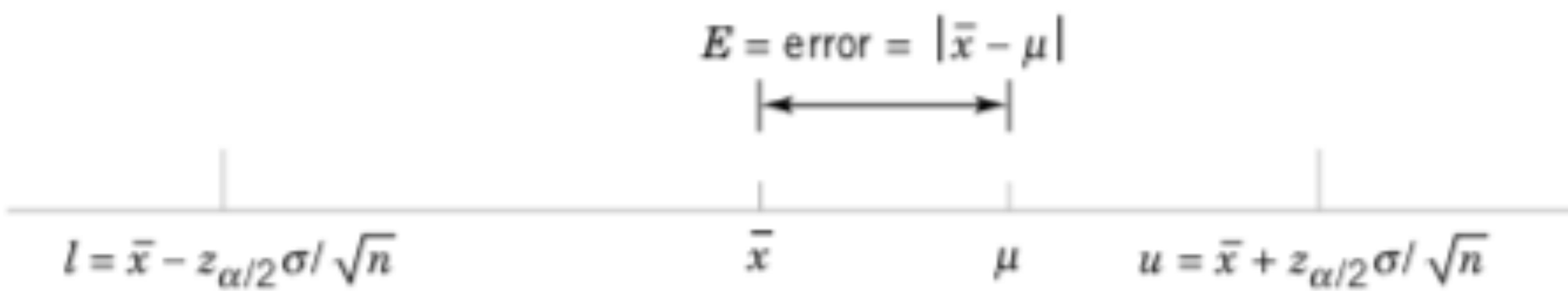


Interpretação do intervalo de confiança

Definição:

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \qquad (8\text{-}8)$$



Interpretação do intervalo de confiança

Exemplo 8.2:

To illustrate the use of this procedure, consider the CVN test described in Example 8-1, and suppose that we wanted to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at 60°C has a length of at most 1.0*J*. Since the bound on error in estimation E is one-half of the length of the CI, to determine n we use Equation 8-8 with $E = 0.5$, $\sigma = 1$, and $z_{\alpha/2} = 0.025$. The required sample size is 16

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left[\frac{(1.96)1}{0.5}\right]^2 = 15.37$$

and because n must be an integer, the required sample size is $n = 16$.

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

O IC pode ser construído mesmo que a população não tenha distribuição normal, desde que n seja grande.

Para n grande, a estimativa da variância também é boa e não há problema.

O problema surge quando n é pequeno.

Neste caso, a estimativa S^2 deve ser usada para substituir σ^2.

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

Para σ conhecida,

$Z = (\bar{X} - \mu)/(\sigma/\sqrt(n))$ t

em distribuição normal padronizada.

Para σ desconhecida, Z se torna

$T = (\bar{X} - \mu)/(S/\sqrt(n))$

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

Definição: distribuição T

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{8-15}$$

has a t distribution with $n - 1$ degrees of freedom.

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

A distribuição t tem pdf,

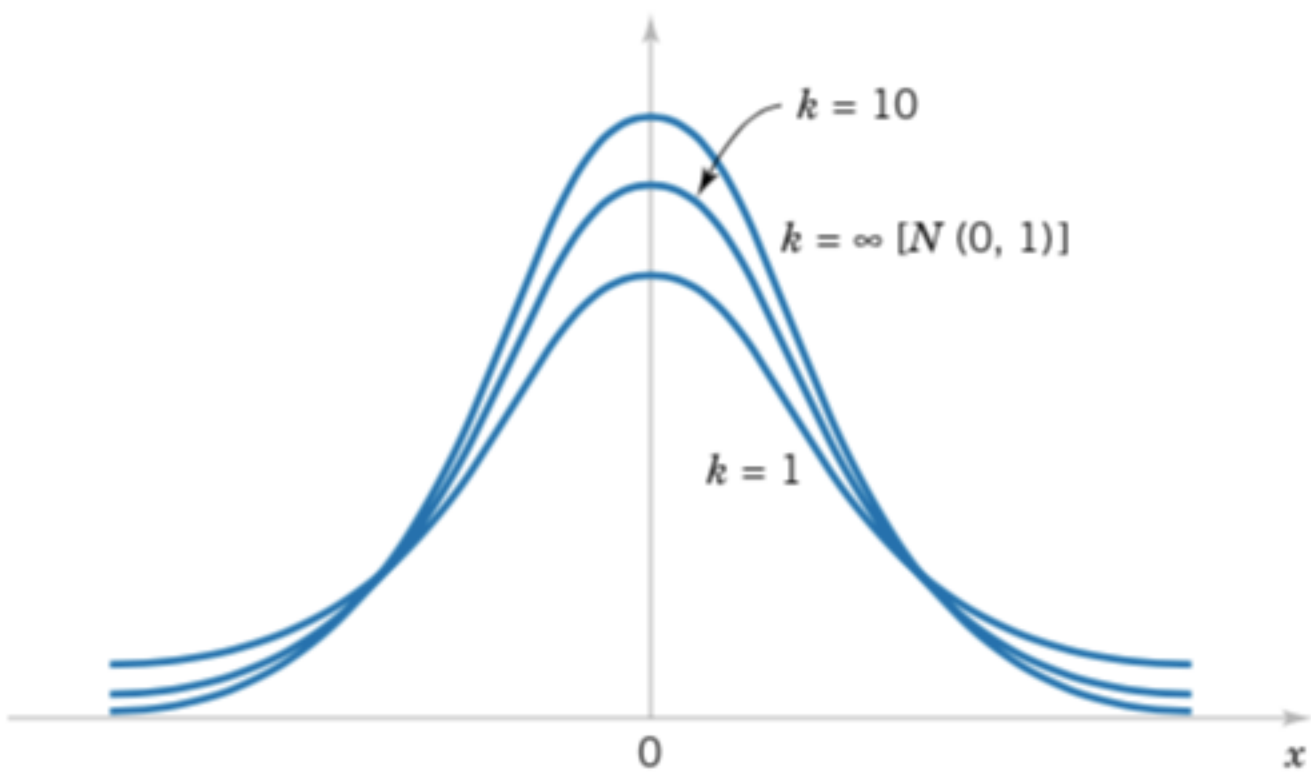
$$f(x) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{[(x^2/k)+1]^{(k+1)/2}} \quad -\infty < x < \infty$$

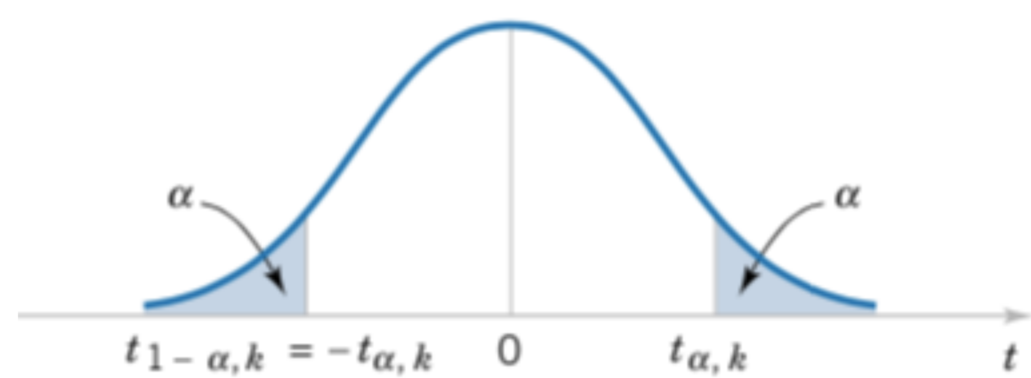
sendo k o número de graus de liberdade. A média é zero e a variância é $k/(k-2)$, para $k > 2$.

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

$$f(x) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{[(x^2/k)+1]^{(k+1)/2}} \quad -\infty < x < \infty$$

sendo k o número de graus de liberdade. A média é zero e a variância é $k/(k-2)$, para $k > 2$.





Intervalo de confiança para média de uma distribuição normal, variância desconhecida

IC para a média, distribuição t

It is easy to find a $100(1-\alpha)$ percent confidence interval on the mean of a normal distribution with unknown variance by proceeding essentially as we did in Section 8-2.1. We know that the distribution of $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is t with $n-1$ degrees of freedom. Letting $t_{\alpha/2,n-1}$ be the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom, we may write:

$$P(-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}) = 1 - \alpha$$

or

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1 - \alpha$$

Rearranging this last equation yields

$$P(\overline{X} - t_{\alpha/2,n-1}S/\sqrt{n} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1}S/\sqrt{n}) = 1 - \alpha \tag{8-17}$$

This leads to the following definition of the $100(1-\alpha)$ percent two-sided confidence interval on μ.

Intervalo de confiança para média de uma distribuição normal, variância desconhecida

Definição:

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a **100(1 − α) percent confidence interval on μ** is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \tag{8-18}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

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

σ **conhecido**

Intervalo de confiança para variância de uma distribuição normal

Definicão:

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

$$X^2 = \frac{(n-1)\,S^2}{\sigma^2} \qquad (8\text{-}19)$$

has a chi-square (χ^2) distribution with $n - 1$ degrees of freedom.

pdf da distribuição chi-quadrado:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \qquad x > 0$$

k = graus de liberdade

$$\mu = k, \sigma^2 = 2k$$

Intervalo de confiança para variância de uma distribuição normal

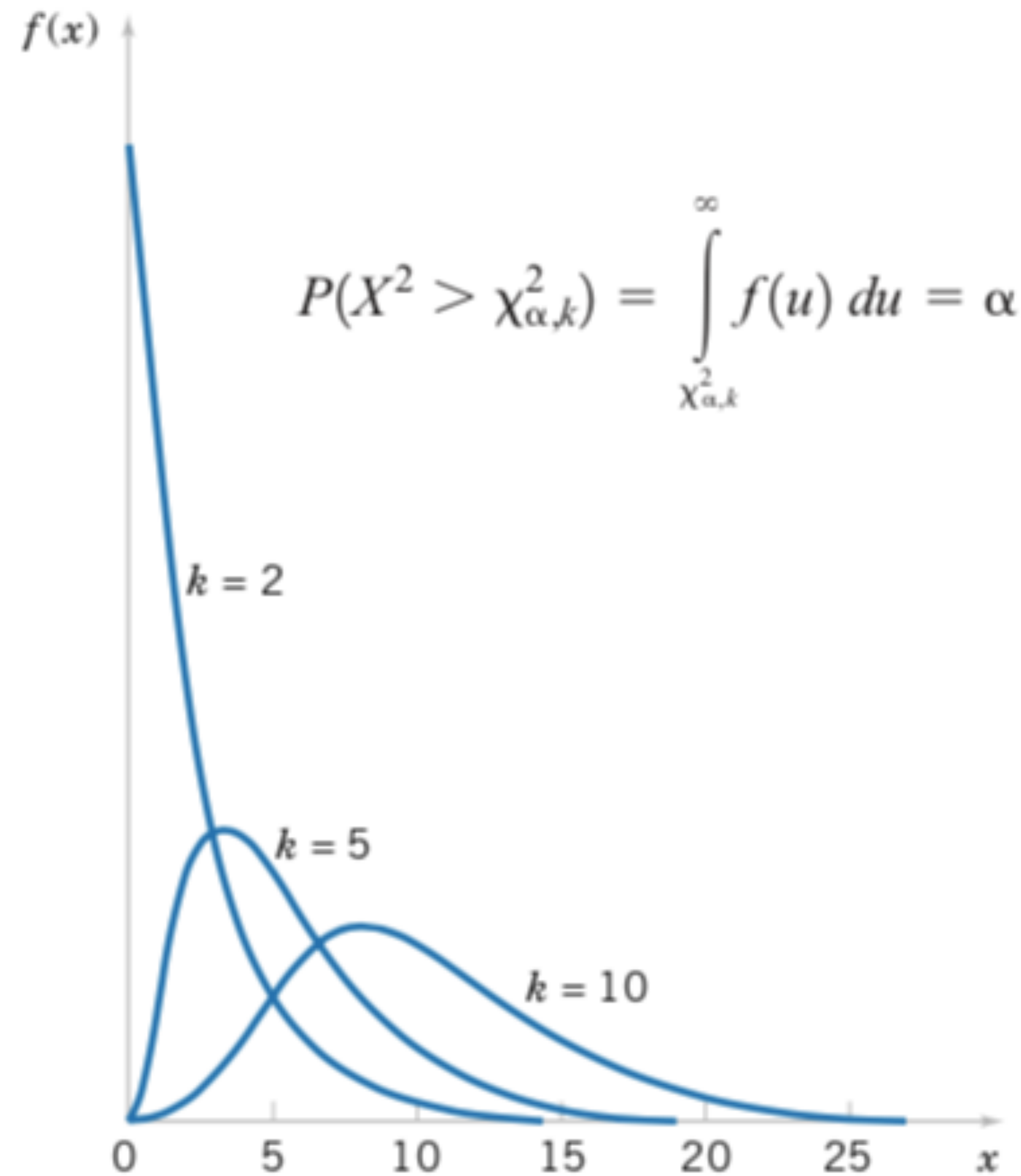
Definição:

pdf da distribuição chi-quadrado:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \qquad x > 0$$

k = graus de liberdade

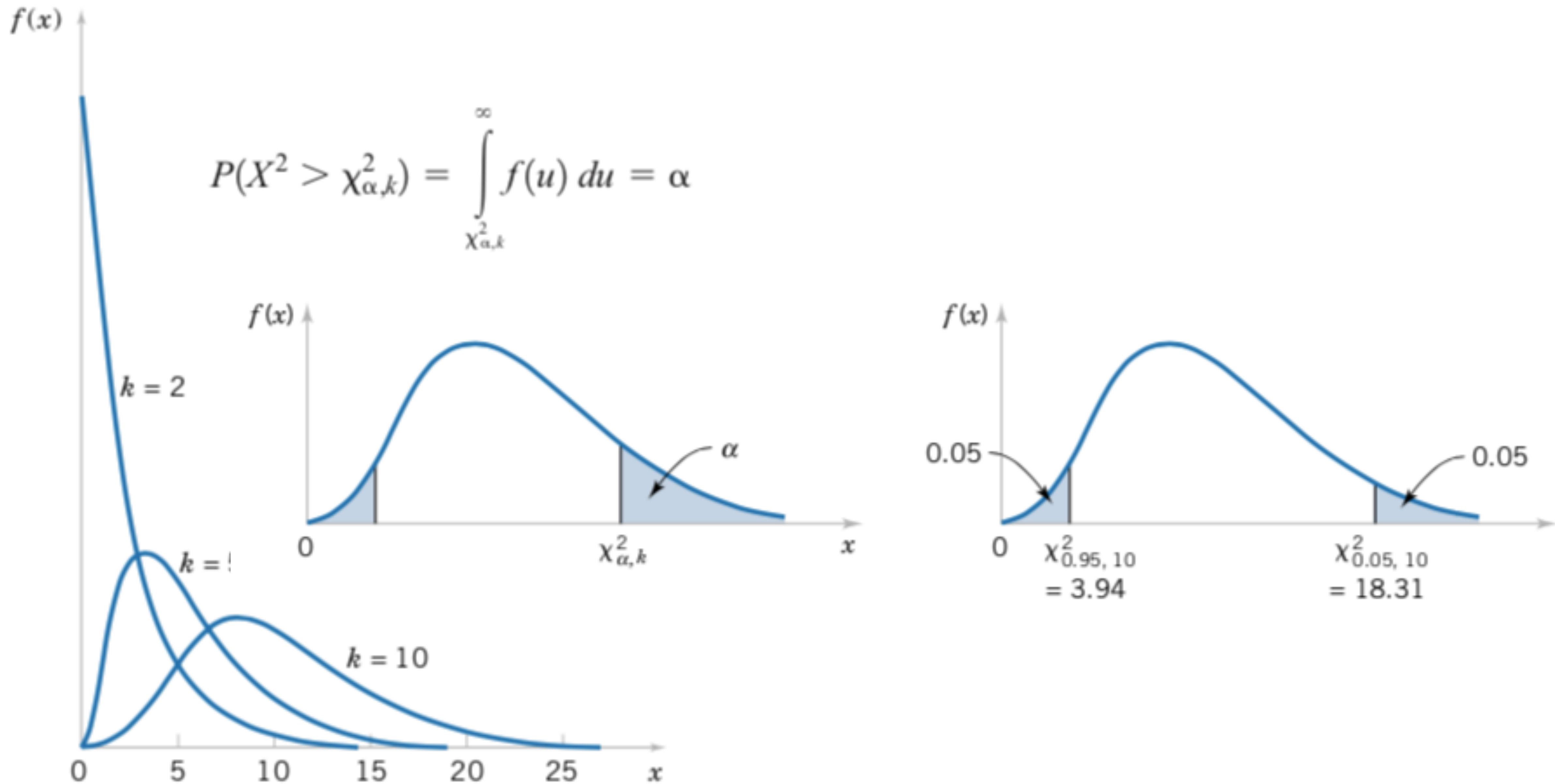
$$\mu = k, \sigma^2 = 2k$$



$$P(X^2 > \chi^2_{\alpha,k}) = \int_{\chi^2_{\alpha,k}}^{\infty} f(u)\, du = \alpha$$

Intervalo de confiança para variância de uma distribuição normal

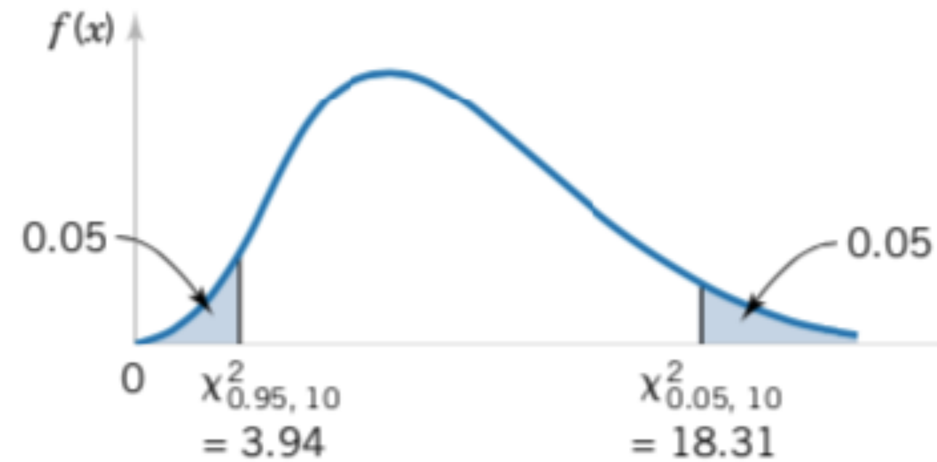
$$P(X^2 > \chi^2_{\alpha,k}) = \int_{\chi^2_{\alpha,k}}^{\infty} f(u)\,du = \alpha$$



Intervalo de confiança para variância de uma distribuição normal

Construção do IC $P(X^2 > \chi^2_{0.05,10}) = P(X^2 > 18.31) = 0.05$

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



is chi-square with $n - 1$ degrees of freedom, we may write

$$P(\chi^2_{1-\alpha/2,n-1} \le X^2 \le \chi^2_{\alpha/2,n-1}) = 1 - \alpha$$

$= 0.05$

so that

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

This last equation can be rearranged as

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

Intervalo de confiança para variância de uma distribuição normal

Definição

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then **a 100(1 − α)% confidence interval on σ^2 is**

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{8-21}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A **confidence interval for σ** has lower and upper limits that are the square roots of the corresponding limits in Equation 8-21.

Intervalo de confiança para variância de uma distribuição normal

Limites inferiores e superiores para σ

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are

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

respectively.

Intervalo de confiança para variância de uma distribuição normal

Exemplo:

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed. A 95% upper-confidence interval is found from Equation 8-22 as follows:

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.95,19}}$$

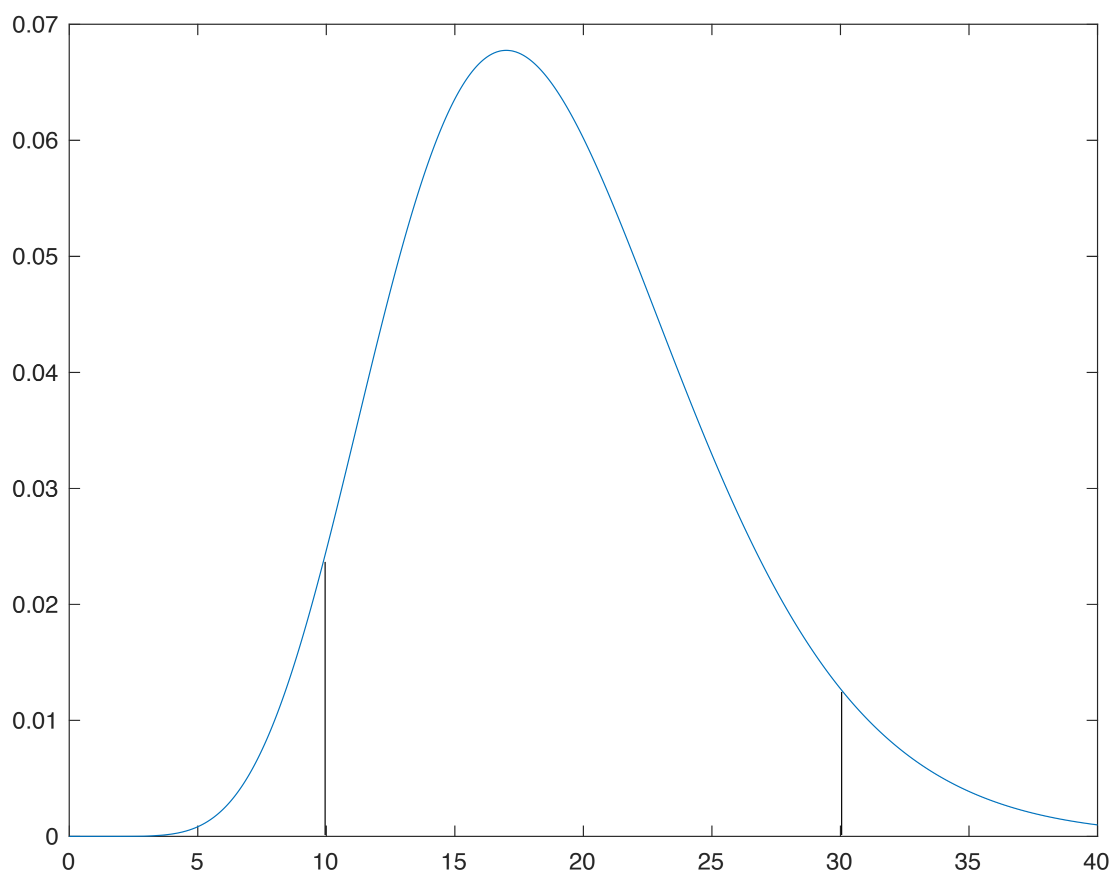
0.05

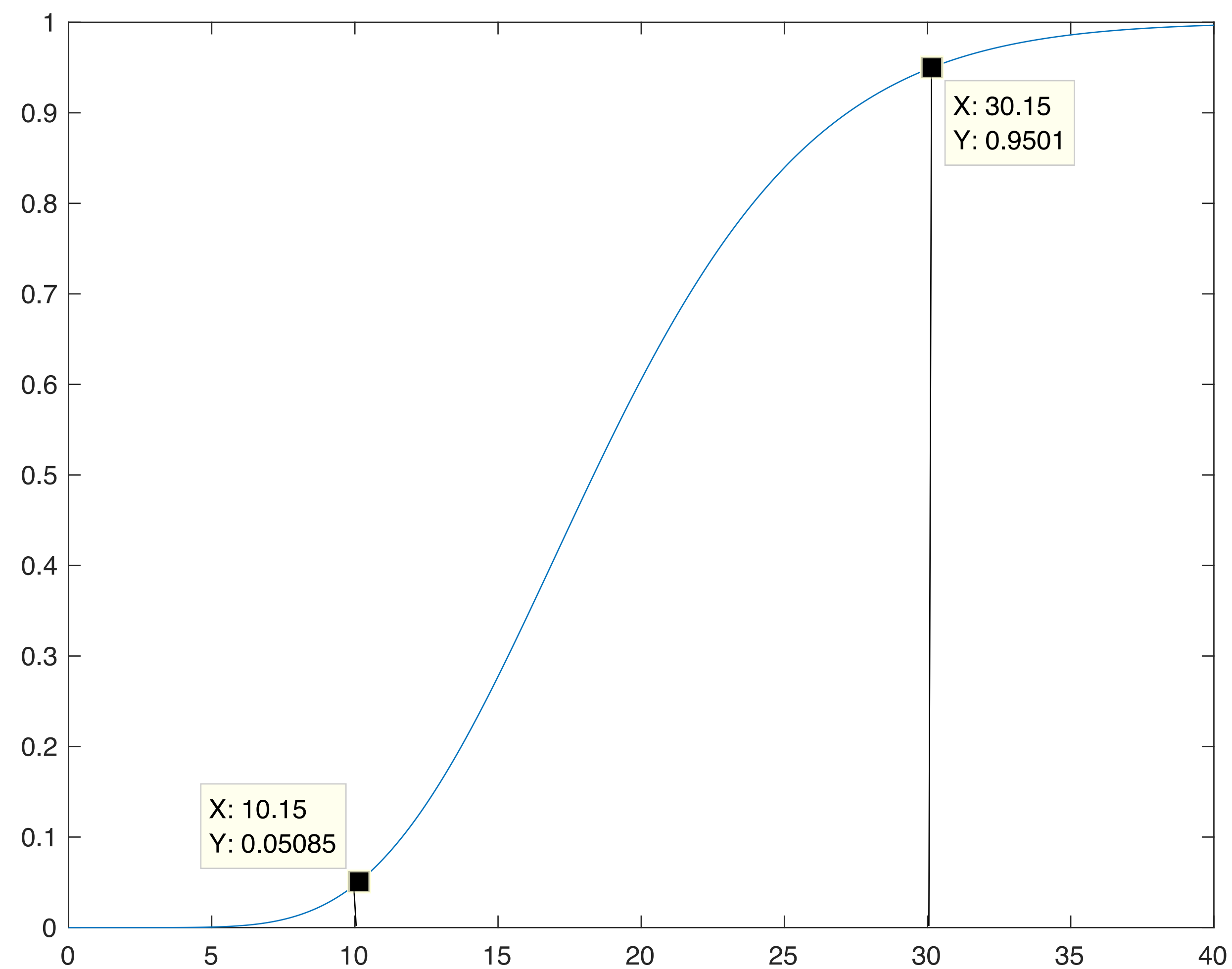
$$\sigma^2 \leq \frac{(19)0.0153}{10.117} = 0.0287 \text{ (fluid ounce)}^2$$

$$\sigma < 0.17$$

Intervalo de confiança para variância de uma distribuição normal

Exemplo: pdf e cdf para k=19





chi2inv([0.95 0.05],19)=[30.14 10.11]

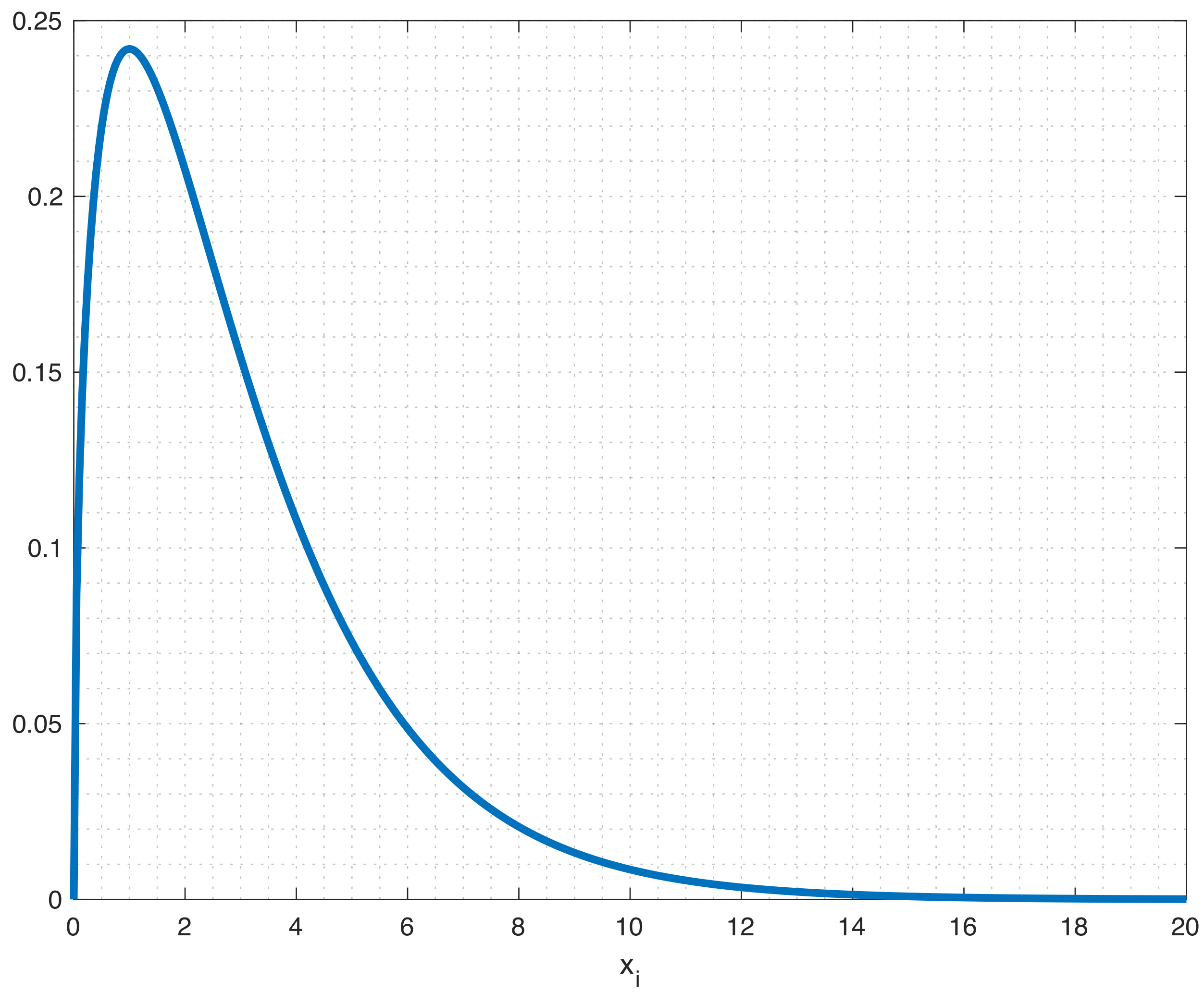
0.27/[30.15 10.11] = [0.0096 0.0287]

Distribuição chi-quadrado

Se $Z_1, Z_2, \ldots, Z_n$ são variáveis aleatórias independentes com distribuição normal padronizada, então a soma de seus quadrados $Q = \sum_{i=1}^{k} = Z_i^2$ tem distribuição chi-quadrado com k graus de liberdade.

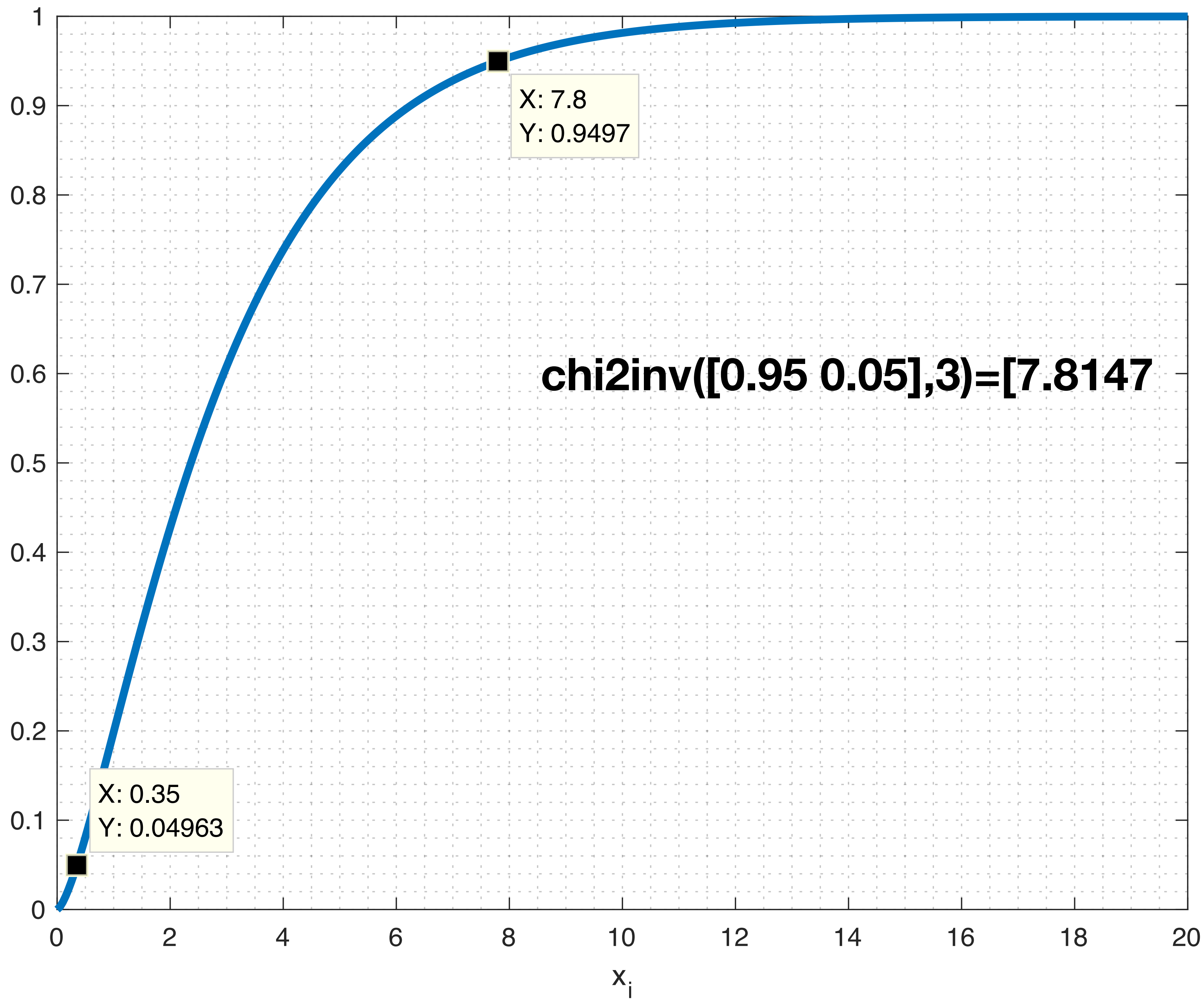
Notação: $Q = X^2(k)$

Distribuição chi-quadrado k=3



pdf , k=3

Distribuição chi-quadrado k=3

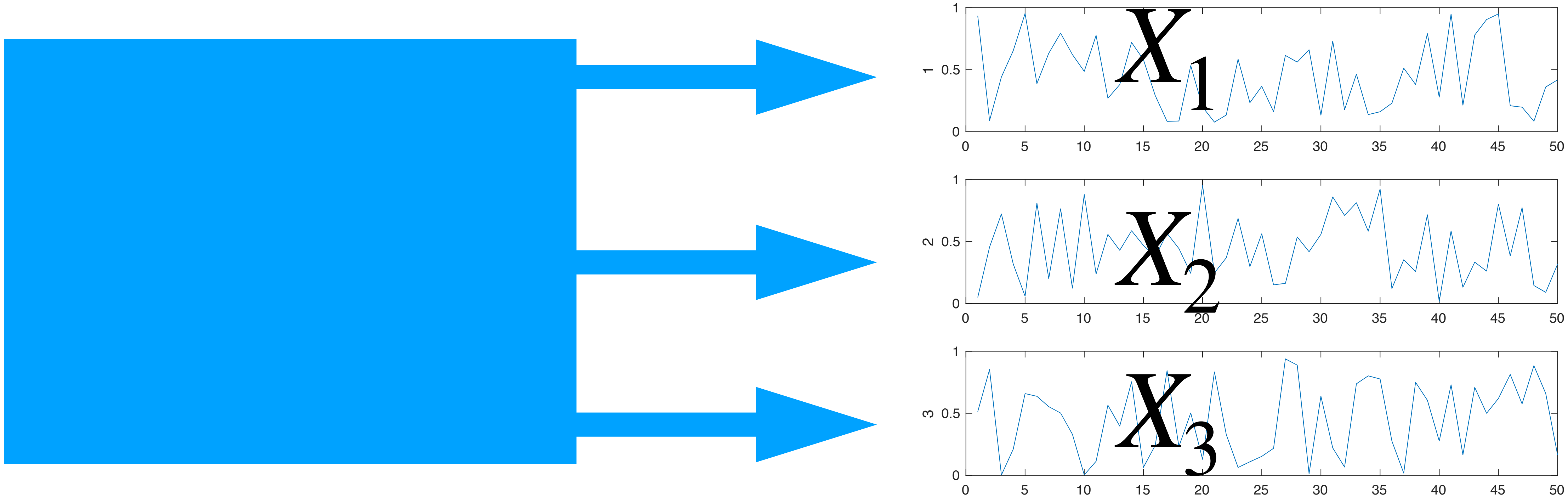


cdf, k=3

chi2inv([0.95 0.05],3)=[7.8147 0.3518]

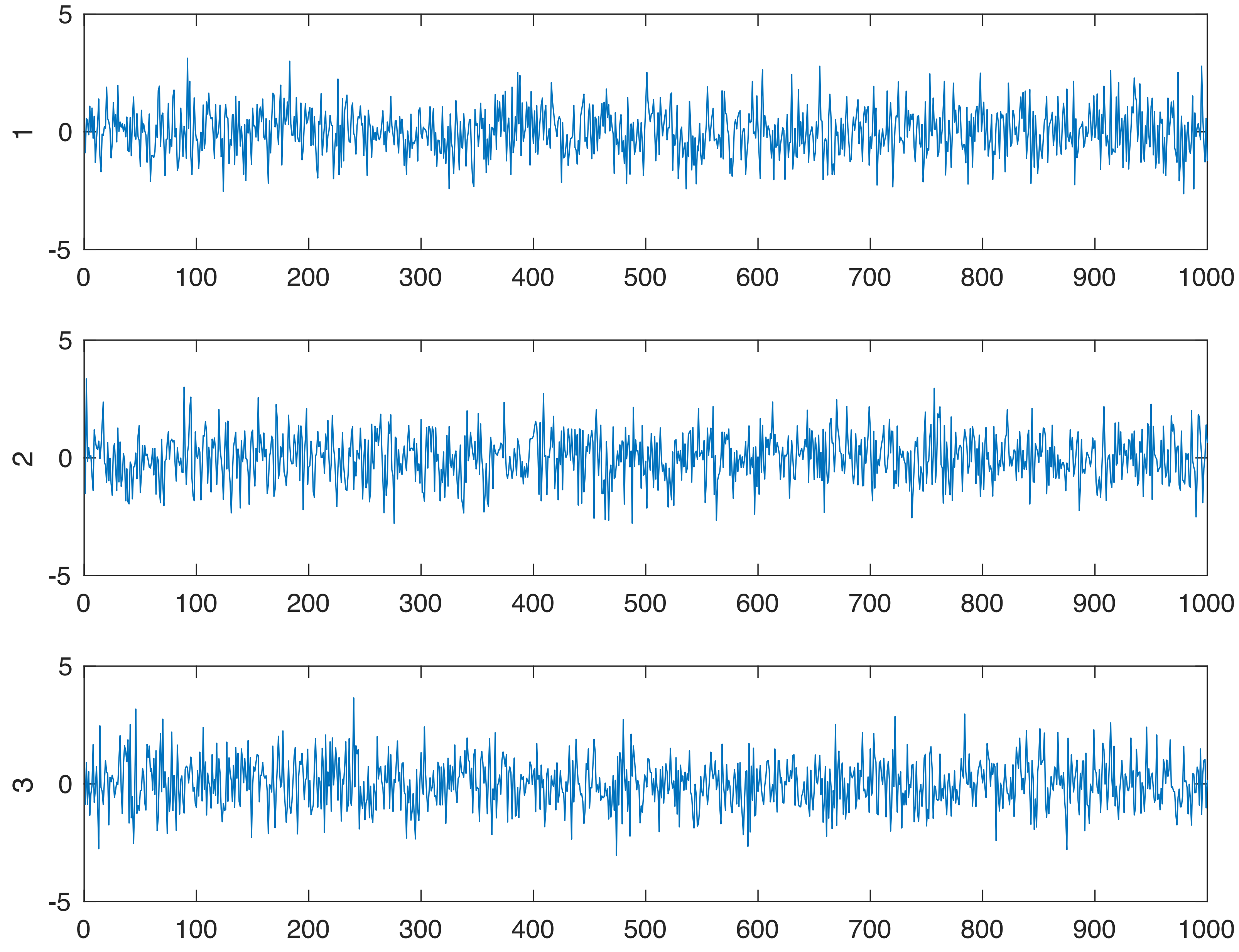
Aplicação

Três variáveis aleatórias com distribuição normal padronizadas



Aplicação

Três variáveis aleatórias com distribuição normal **padronizadas**



Aplicação

Soma das 3 variáveis ao quadrado: $Q = x_1^2 + x_2^2 + x_3^2$

chi2inv([0.95 0.05],3)=
[7.8147   0.3518]

$$P(0.35 < Q < 7.81) = 0.9$$