UNIDADE 5

Distribuição de Probabilidades Discretas e Contínuas

Objetivos específicos de aprendizagem

Ao finalizar esta Unidade, você deverá ser capaz de:

- Identificar e aplicar modelos probabilísticos discretos;
- Identificar e aplicar modelos probabilísticos contínuos (distribuição normal);
- Saber quando e como utilizar as distribuições amostrais;
- Calcular e interpretar intervalos de confiança; e
- Dimensionar amostras para serem utilizadas em pesquisas e projetos.

INTRODUÇÃO

Caro estudante,

Como você progrediu nos conhecimentos básicos de probabilidade, agora iremos trabalhar com as chamadas distribuições de probabilidades. Essas distribuições auxiliam no cálculo de probabilidades e, ainda, nos processos de estimação e de decisão, conforme veremos na próxima Unidade. Estudaremos as distribuições de amostragem e dimensionamento de amostras que, também, serão vistas nesta Unidade.

Bons estudos e conte conosco para auxiliá-lo sempre que necessário.

Vamos começar com alguns conceitos preliminares.

Para que você tenha condições de entender as distribuições, é necessário conhecer bem o que é uma **variável aleatória***, que pode ser discreta ou contínua.

*** Variável aleatória** – função que associa valores reais aos eventos de um espaço amostral. Fonte: Elaborado pelo autor.

Um exemplo de uma variável aleatória discreta (v.a.) é a quantidade de ações que tiveram queda em um determinado dia, em uma carteira composta por cinco ações diferentes. A função será dada por:

> X = "quantidade de ações que tiveram queda em um determinado dia" define uma variável aleatória discreta, que pode assumir os valores 0, 1, 2, 3, 4, 5.

Vamos considerar agora uma situação na qual se verificou o tempo gasto por um funcionário público para atender um contribuinte. A função será:

> Y= "tempo gasto por um funcionário público para atender um contribuinte" define uma variável aleatória contínua, que pode assumir infinitos valores.

Vamos trabalhar aqui principalmente com as variáveis aleatórias discretas. Se uma variável aleatória X pode assumir os valores x_1, x_2,..., x_n com probabilidades respectivamente iguais a p_1, p_2,..., p_n, e $\sum_{i=1}^{n} p_i = 1$, temos então definida uma **distribuição de probabilidade***.

*** Distribuição de probabilidade** – é um tipo de distribuição que descreve a chance que uma variável pode assumir ao longo de um espaço de valores. Fonte: Elaborado pelo autor.

É importante ressaltarmos que a variável aleatória tem notação de letra maiúscula e seus possíveis valores minúsculos, como utilizamos anteriormente.

Se a variável X em questão for discreta, sua distribuição é caracterizada por uma **função de probabilidade (P(X=x))**, que associa probabilidades não nulas aos possíveis valores da variável aleatória.

Distribuições Discretas

Imagine uma situação na qual somente podem ocorrer dois possíveis resultados, "sucesso" e "fracasso". Veja alguns exemplos:

- uma venda é efetuada ou não em uma ligação de *call center*;
- um contribuinte pode ser adimplente ou inadimplente;
- uma guia recolhida pode ter seu preenchimento ocorrido de forma correta ou incorreta; e
- um consumidor que entra em uma loja pode comprar ou não comprar um produto.

Essas situações correspondem à Distribuição de Bernoulli. Ou seja, se associarmos uma variável aleatória x aos possíveis resultados do experimento de forma que X=1 se o resultado for "sucesso" e X=0 se o resultado for "fracasso", então, a variável aleatória X, assim definida, tem Distribuição de Bernoulli, com p sendo a probabilidade de ocorrer "**sucesso**" e q = (1-p) a probabilidade de ocorrer "**fracasso**".

Neste momento, você deve saber que quando estamos falando de sucesso, devemos relacioná-lo com o objetivo do exercício ou do problema a ser resolvido, que, muitas vezes, pode não ser algo bom.

Ampliando nossa discussão, é importante mencionarmos ainda que a função de probabilidade da Distribuição de Bernoulli é dada por:

$$P(X = x) = \begin{cases} p \text{ para } x = 1, \\ q = 1 - p \text{ para } x = 0 \\ 0 \text{ para } x \text{ diferente de } 0 \text{ ou } 1 \end{cases}$$

Sendo assim, a média e a variância serão obtidas por:

- Média = p (onde p corresponde à probabilidade de sucesso).
- Variância = p · q (onde q corresponde à probabilidade de fracasso).

Essa obtenção da estimativa de média e desvio padrão é importante, pois tais medidas podem ser usadas para caracterizar a situação e também para a definição da média e do desvio padrão da distribuição binomial que iremos ver posteriormente.

Contextualizando a Distribuição de Bernoulli, temos a seguinte situação: a experiência tem mostrado que até fevereiro o motorista que é parado em uma *blitz* tem 60% de chance de estar adimplente em relação ao Imposto sobre a Propriedade de Veículos Automotores (IPVA). Temos, portanto, uma probabilidade de sucesso **(o motorista não estar devendo o IPVA)** de 0,6 e uma probabilidade de estar devendo de 0,4 **(vem da diferença q = 1 – 0,6)**.

Distribuição Binomial

Para que uma situação possa se enquadrar em uma distribuição binomial, deve atender as seguintes condições:

- são realizadas n repetições (tentativas) independentes;
- cada tentativa é uma prova de Bernoulli (somente podem ocorrer dois possíveis resultados); e

- a probabilidade p de sucesso em cada prova é constante.

Se uma situação atende a todas as condições anteriores, então a variável aleatória X = número de sucessos obtidos nas n tentativas terá uma distribuição binomial com n tentativas e p probabilidades de sucesso.

Agora você deve parar a sua leitura e lançar uma moeda 30 vezes para cima. Após fazer isso e anotar os resultados, veja se o experimento que acabou de fazer se encaixa em uma distribuição binomial (condições apresentadas anteriormente).

Simbolicamente, temos: $X \sim B(n,p)$ com a interpretação:

A variável aleatória X tem distribuição binomial (B) com n ensaios e uma probabilidade p de sucesso (em cada ensaio).

A função de probabilidade utilizada para cálculo de probabilidades, quando a situação se enquadra na distribuição binomial, será dada por meio da seguinte expressão:

$P(X=x)=C_n^x p^x q^{n-x}$ onde:

p é probabilidade de "sucesso" em cada ensaio;

q = 1-p é a probabilidade de "fracasso" em cada ensaio;

$C_n^x = \dfrac{n!}{x!(n-x)!}$, onde n! é o fatorial de n, é combinação de n valores tomados x a x

Lembre-se dos conceitos de análise combinatória vistos no segundo grau!

Exemplo

Vamos considerar que algumas pessoas entram em uma loja no período próximo ao dia das mães. Sabemos que a probabilidade de uma pessoa do gênero masculino comprar um presente é de 1/3. Se entrarem quatro pessoas do gênero masculino na tal loja, qual a probabilidade de que duas venham a comprar presentes?

Se essas quatro pessoas entram na loja e duas delas compram, podemos colocar as possibilidades da seguinte forma (C ➔ compra e não-C ➔ não compra). O espaço amostral associado ao experimento é:

C, C, não-C, não-C **ou** C, não-C, não-C, C **ou** C, não-C, C, não-C **ou** não-C, não-C, C, C **ou** não-C, C, não-C, C **ou** não-C, C, C, não-C

Logo, calculando as probabilidades usando as regras do **"e"** (multiplicação, pois são independentes) e do **"ou"** (soma), a **probabilidade de 2 clientes do gênero masculino comprarem presentes** é:

$$p = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3} + \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3} + \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}$$

$$p = 6\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}$$

$$p = 6\cdot\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^2$$

$$p = \frac{24}{81} \cong 29{,}63\%$$

Agora, vamos calcular utilizando a função de probabilidade apresentada anteriormente e verificar que o resultado será o mesmo.

$$P(X=2) = C_4^2.\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^2 = \frac{4!}{2!.(4-2)!}.\frac{1}{9}.\frac{4}{9} = \frac{4.3.2.1}{2.1.2.1}.\frac{4}{81} = \frac{24}{81} \cong 0{,}2963 \quad ou \quad 29{,}63\%$$

Os valores da média e da variância da distribuição binomial são:

Média = n.p

Variância = n.p.q

Exemplo

Em uma determinada repartição pública, 10% das guias preenchidas estão incorretas. Essas guias correspondem a uma liberação na qual cinco guias devem estar preenchidas conjuntamente. Considere que cada guia tem a mesma probabilidade de ser preenchida incorretamente (como se houvesse repetição no experimento de retirar guias).

Como na binomial são n ensaios de Bernoulli e a distribuição tem média p, a média da binomial será n.p. Raciocínio semelhante é feito para a variância.

a) Qual a probabilidade de haver exatamente três guias incorretas nas cinco guias para liberação?

O **sucesso** é a ocorrência de guias preenchidas incorretamente.

p = 0,1 n = 5

$$P(X=3) = C_5^3 \cdot 0{,}1^3 \cdot 0{,}9^2 = 0{,}0081$$

b) Qual a probabilidade de haver duas ou mais guias incorretas nas cinco guias para liberação?

$$P(X \geq 2) = P(X{=}2) + P(X{=}3) + P(X{=}4) + P(X{=}5)$$
$$= 1 - [P(X{=}0) + P(X{=}1)] = 0{,}0815$$

c) Qual a probabilidade de um conjunto de cinco guias não apresentar nenhuma guia incorreta?

$$P(X=0) = C_5^0\, 0{,}1^0 \cdot 0{,}9^5 = 0{,}5905$$

Antes de prosseguir, desta vez com o estudo da Distribuição de Poisson, você deve realizar as Atividades 1 e 2, ao final desta Unidade, para aplicar os conhecimentos já adquiridos sobre a distribuição binomial. Lembre-se de que as respostas se encontram no final do livro.

Distribuição de Poisson

Você pode empregar a Distribuição de Poisson em situações nas quais não se está interessado no número de sucessos obtidos em n tentativas, como ocorre no caso da distribuição binomial, entretanto, esse número de sucessos deve estar dentro de um intervalo contínuo, ou seja, **o número de sucessos ocorridos durante um intervalo contínuo**, que pode ser um intervalo de tempo, espaço etc.

Imagine que você queira estudar o número de suicídios ocorridos em uma cidade durante um ano ou o número de acidentes automobilísticos ocorridos em uma rodovia em um mês ou o número de defeitos encontrados em um rolo de arame ovalado de 500m. Essas situações são exemplos daquelas que se enquadram na Distribuição de Poisson.

Note que nos exemplos anteiores não há como você determinar a probabilidade de ocorrência de um sucesso, mas sim a frequência média de sua ocorrência, como dois suicídios por ano, que denominaremos $\boldsymbol{\lambda}$.

Em uma situação com essas características, a variável aleatória X = número de sucessos em um intervalo contínuo, terá uma Distribuição Poisson, com λ (frequência média de sucesso). Simbolicamente, podemos utilizar a notação $\mathbf{X \sim P(\lambda)}$.

A variável aleatória X tem uma Distribuição de Poisson (P) com uma frequência média de sucesso λ.

A função de probabilidade da Distribuição de Poisson será dada por meio da seguinte expressão:

$$P(X = x) = e^{-\lambda}.\frac{\lambda^x}{x!}$$

Onde:

e =2,7182 (base dos logarítmos neperianos); e

λ corresponde a frequência média de sucesso no intervalo contínuo que se deseja calcular a probabilidade.

Exemplo

A análise dos dados dos últimos anos de uma empresa de energia elétrica forneceu o valor médio de um blecaute por ano. Pense na probabilidade de isso ocorrer no próximo ano:

a) Nenhum blecaute.

b) De 2 a 4 blecautes.

c) No máximo 2 blecautes.

Note que o exemplo afirma que a cada ano acontece em média um blecaute, ou seja, o **número de sucesso ocorrido em um intervalo contínuo**. Verificamos que a variável tem Distribuição Poisson:

$$P(X = x) = e^{-\lambda}.\frac{\lambda^x}{x!}$$

Veja que aqui não é necessário fazer regra de três, pois as perguntas são no intervalo de um ano. Então: $\lambda = 1$:

a) $P(x=0) = \dfrac{e^{-1}.1^0}{0!} = \dfrac{0{,}3679.1}{1} = 0{,}3679 \text{ } ou \text{ } 36{,}79\%$

b) $P(x=2) + P(x=3) + P(x=4) = \dfrac{e^{-1}.1^2}{2!} + \dfrac{e^{-1}.1^3}{3!} + \dfrac{e^{-1}.1^4}{4!} =$

$= 0{,}1839 + 0{,}061 + 0{,}015$

$= 0{,}2599 \text{ } ou \text{ } 25{,}99\%$

c) Como já temos os valores de x = 0 e x = 2 basta calcularmos para x = 1 e somarmos os resultados.

$$P(x=1) = \frac{e^{-1}.1^1}{1!} = \frac{0{,}3679.1}{1} = 0{,}3679 \text{ } ou \text{ } 36{,}79\%$$

$P(x \le 2) = P(x=0) + P(x=1) + P(x=2) =$
$0{,}3679 + 0{,}3679 + 0{,}1839 = 0{,}9197 \text{ } ou \text{ } 91{,}97\%$

Vejamos uma aplicação da Distribuição de Poisson considerando que o Corpo de Bombeiros de uma determinada cidade recebe, em média, três chamadas por dia. Queremos saber, então, qual a probabilidade do Corpo de Bombeiros receber:

Como o intervalo em que se deseja calcular a probabilidade é um dia, o λ será igual a 3.

a) **4 chamadas em um dia**: verificamos que a variável tem Distribuição Poisson, pois temos número de chamadas (variável discreta) por dia (intervalo contínuo). A probabilidade será calculada por meio da expressão:

$$P(X = x) = e^{-\lambda}.\frac{\lambda^x}{x!}$$

Como não é necessário fazer regra de três, pois as perguntas são no intervalo de um dia, então: λ = 3. Substituindo na expressão, teremos:

$$P(X = 4) = e^{-3}\frac{3^4}{4!} = 0{,}1680$$

b) **Nenhuma chamada em um dia**: nesse caso, o intervalo continua sendo um dia. Logo, o lambda (λ) continua sendo o mesmo, ou seja, λ = 3. Substituindo então na expressão, teremos:

$$P(X = 0) = e^{-3}\frac{3^0}{0!} = 0{,}0498$$

c) **20 chamadas em uma semana**: nesse caso o intervalo em que se deseja calcular a probabilidade é de uma semana, ou seja, sete dias. Então, em uma semana, a frequência média de chamadas será de 7 dias vezes 3 chamadas/dia:

λ = 21 chamadas por semana

Substituindo os valores, teremos a seguinte probabilidade:

$$P(X = 20) = e^{-21}\frac{21^{20}}{20!} = 0{,}0867$$

Uma característica da Distribuição de Poisson é que as estatísticas da distribuição (média e variância) apresentam o mesmo valor, ou seja, são iguais a λ . Então, teremos:

$$\text{Média} = \text{Variância} = \lambda$$

Antes de discutir as distribuições contínuas, vamos aplicar os conhecimentos relacionados à Distribuição de Poisson realizando a Atividade 3 ao final desta Unidade. É importante salientarmos que nesta Unidade a resolução das atividades de aprendizagem serão solicitadas ao longo do texto para facilitar a sua compreensão dos conceitos e de como utilizá-los.

Distribuições Contínuas

Dentre as várias distribuições de probabilidade contínuas, abordaremos aqui apenas a distribuição normal, pois ela apresenta grande aplicação em pesquisas científicas e tecnológicas. Grande parte das variáveis contínuas de interesse prático segue essa distribuição, aliada ao Teorema do Limite Central (TLC), que é a base das estimativas e dos testes de hipóteses realizados sobre a média de uma população qualquer, e garante que a distribuição amostral das médias segue uma distribuição normal, independentemente da distribuição da variável em estudo, como será visto mais adiante.

Distribuição Normal

A função densidade de probabilidade da distribuição normal é dada por:

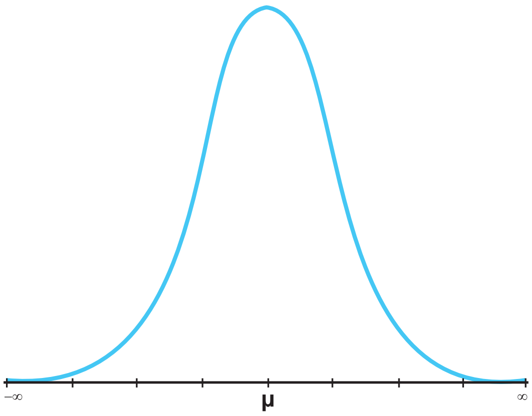
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in R.$$

Onde:

μ e σ são a média e o desvio padrão, respectivamente, da distribuição de probabilidade.

π corresponde a 3,1415 e *exp* a uma função exponencial.

O gráfico da distribuição normal, utilizando a função mostrada anteriormente e os conceitos vistos nas disciplinas *Matemática Básica e Matemática para Administradores*, é dado por:

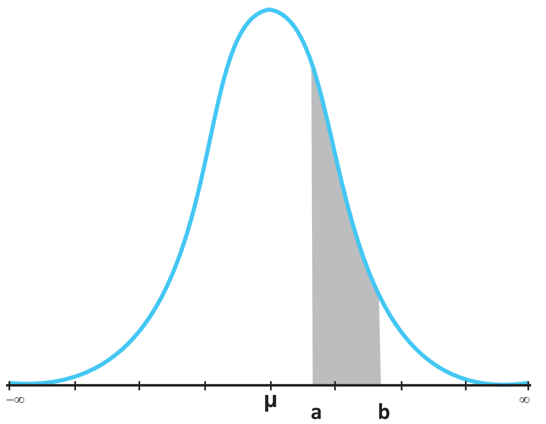


Você encontrará a seguir as principais propriedades da distribuição normal:

- é simétrica em relação ao ponto $x = \mu$ (50% abaixo e 50% acima da média);
- tem forma **campanular***;
- as três medidas de posição – média, mediana e moda – se confundem no ponto máximo da curva ($x = \mu$);
- fica perfeitamente definida conhecendo-se a média e o desvio padrão, pois outros termos da função são constantes; e
- toda a área compreendida entre a curva e o eixo x é igual a 1 (conceito de probabilidades).

*** Campanular** – relativo à campânula; objeto em forma de sino. Fonte: Houaiss (2009).

Portanto, a área sob a curva entre os pontos a e b, em que $a < b$, representa a probabilidade da variável X assumir um valor entre a e b (área escura), como observamos a seguir.



Desse modo, você pode associar que, no caso das distribuições contínuas, a área do gráfico corresponde a probabilidades.

Então, veja a notação utilizada para a distribuição normal:

$$X \sim N(\mu \, , \, \sigma^2)$$

A variável x tem distribuição normal com média μ e variância σ^2.

Para calcularmos as probabilidades via distribuição normal, é necessário o conhecimento de cálculo integral. Assim, procuramos tabelar os valores de probabilidade que seriam obtidos por meio da integração da função densidade de probabilidade normal em um determinado intervalo.

A dificuldade para se processar esse tabelamento se prendeu na infinidade de valores que μ(média) e σ(desvio padrão) poderiam assumir. Nessas condições, teríamos que dispor de uma tabela para cada uma das infinitas combinações de μ e σ, ou seja, em cada situação que se quisesse calcular uma probabilidade.

Para resolver esse problema, podemos obter uma nova forma para a distribuição normal, que não seja influenciada por μ e σ. O problema foi solucionado mediante emprego de uma nova variável definida por:

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

Essa variável transforma todas as distribuições normais em uma distribuição normal reduzida ou padronizada, de média zero e desvio padrão um. Então, temos: $Z \sim N(0,1)$.

Assim, utilizamos apenas uma tabela para o cálculo de probabilidades para qualquer que seja a curva correspondente a uma distribuição normal.

Portanto, para um valor de $x = \mu$ em uma distribuição normal qualquer, corresponde o valor:

$$z = \frac{x-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$ na distribuição normal reduzida.

Para $x = \mu + \sigma$, temos:

$$z = \frac{x-\mu}{\sigma} = \frac{\mu+\sigma-\mu}{\sigma} = \frac{\sigma}{\sigma} = 1$$ e, assim por diante.

Podemos definir a distribuição normal reduzida ou padronizada como sendo uma distribuição da variável Z que apresenta distribuição normal com média zero e variância 1 ($Z \sim N(0;1)$).

Na Tabela 15, que apresenta a distribuição normal padronizada, as áreas ou probabilidades fornecidas estão entre zero e o valor de Z, como vemos a seguir.

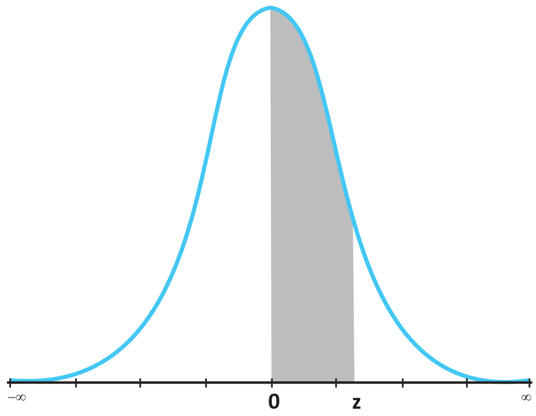


Tabela 15: Área sob a curva normal padronizada compreendida entre os valores 0 e Z

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Fonte: Elaborada pelo autor

Veja que na Tabela 15 os valores apresentados na primeira coluna correspondem a parte inteira e decimal do valor de Z (por exemplo 1,5), enquanto os valores da primeira linha correspondem a parte centesimal (por exemplo 8). Assim, teremos o valor de Z = 1,58. Já os valores encontrados no meio da tabela correspondem às probabilidades dos respectivos valores compreendidos entre zero e Z.

Para que você possa entender a utilização da distribuição normal, vamos considerar a arrecadação como um tributo de uma pequena cidade. Verificamos que essa arrecadação seguia uma distribuição normal com duração média de R$ 60.000,00 e desvio padrão de R$ 10.000,00. Procuramos, então, responder os seguintes questionamentos:

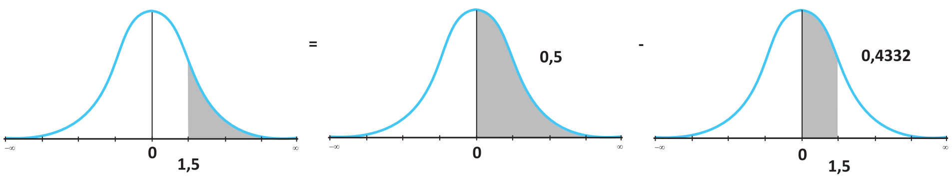
a) Qual a probabilidade de uma arrecadação ser maior do que R$ 75.000,00?

Como a variável arrecadação apresenta distribuição aproximadamente normal com média 60000 e variância de 10000^2 [$X \sim N(60000;10000^2)$] e procura-se calcular a $P(X > 75000) = ?$

Primeiramente, precisamos transformar a variável X em Z e, depois, substituindo na expressão, teremos:

$$z = \frac{x-\mu}{\sigma} = \frac{75000-60000}{10000} = 1{,}50$$

Olhando esse valor na Tabela 15, z = 1,50 (1,5 na primeira coluna e o zero na primeira linha), encontraremos no meio da tabela o valor de 0,4332 que corresponde à probabilidade de z estar entre zero e 1,5, como você pode observar a seguir.



A área escura da figura corresponde a $P(X>75000)$, que é a mesma coisa que: $P(z > 1{,}50)$. Então:

$P(z > 1{,}50)$ [Figura 1] = $P(0<z<+\infty)$[Figura 2] –

$P(0 < z < 1{,}50)$ [Figura 3] = 0,5 – 0,4332 = 0,0668.

Retirou-se a probabilidade encontrada de 0,5, pois esse valor corresponde à probabilidade de zero até o infinito.

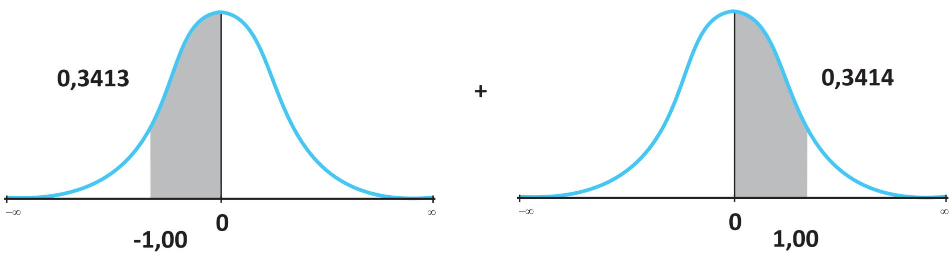
b) Qual a probabilidade da arrecadação estar entre R$ 50.000,00 e R$ 70.000,00?

P(50000 < X < 70000) = ?

Primeiramente, precisamos transformar a variável X em Z e, depois, substituindo na expressão de Z, teremos valores de Z_1 e Z_2, relacionados aos valores de X_1=50000 e X_2=70000:

$$z_1 = \frac{x-\mu}{\sigma} = \frac{50000-60000}{10000} = -1{,}00$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{70000-60000}{10000} = 1{,}00$$



Podemos verificar que:

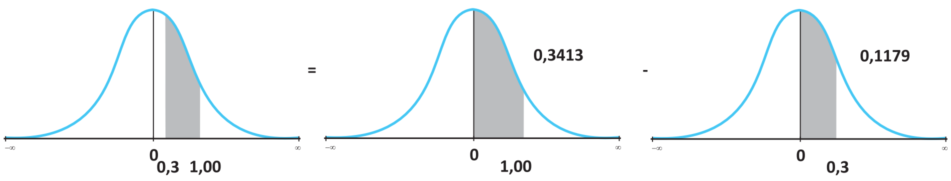
P(50000 < X < 70000) = P(– 1,00 < z < 1,00) = 0,3413 + 0,3413 = 0,6826

c) Qual a probabilidade da arrecadação estar entre R$ 63.000,00 e R$ 70.000,00?

P(63000 < X < 70000) = ?

$$z_1 = \frac{x-\mu}{\sigma} = \frac{63000-60000}{10000} = 0{,}30$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{70000-60000}{10000} = 1{,}00$$



$$P(63000 < X < 70000) = P(\ 0{,}30 < z < 1{,}00) =$$
$$0{,}3413 - 0{,}1179 = 0{,}2234$$

Destacamos que existem outras distribuições tanto discretas quanto contínuas que não foram abordadas neste livro. Portanto, recomendamos que você procure outras fontes de conhecimento, a começar por fazer uma pesquisa na internet sobre essas distribuições.

Antes de prosseguir, você deve realizar as Atividades 4 e 5 ao final desta Unidade, na qual você terá a oportunidade de verificar o seu grau de compreensão sobre a distribuição normal.

Distribuições amostrais

Com as distribuições amostrais, você pode inferir propriedades de um agregado maior (a população) a partir de um conjunto menor (a amostra), ou seja, inferir sobre parâmetros populacionais, dispondo apenas de estatísticas amostrais. Portanto, torna-se necessário um estudo detalhado das distribuições amostrais, que são base para intervalos de confiança e testes de hipóteses.

Para que você tenha condições de fazer afirmações sobre um determinado parâmetro populacional (ex: μ), baseados na estimativa $\bar{x}$, obtida a partir dos dados amostrais, é necessário conhecer a relação existente entre $\bar{x}$ e μ, isto é, o comportamento de $\bar{x}$, quando se extraem todas as amostras possíveis da população, ou seja, sua distribuição amostral.

Para obtermos a distribuição amostral de um estimador, é necessário conhecer o processo pelo qual as amostras foram retiradas, isto é, se amostras foram retiradas **com reposição** ou **sem reposição**. Neste material, iremos considerar apenas as situações de amostragens com reposição.

Dessa forma, a partir do comportamento da estatística amostral, podemos aplicar um teorema muito conhecido na estatística como Teorema do Limite Central (TLC). Esse teorema propõe que, se retirarmos todas as possíveis amostras de tamanho n de uma população, independente de sua distribuição, e verificarmos como as estatísticas amostrais obtidas se distribuem, teremos uma distribuição **aproximadamente normal**, com $\mu_{x} = \mu$ (**média das medias amostrais igual à média populacional**) e variância das médias $\sigma^2_{\bar{x}} = \dfrac{\sigma^2}{n}$ (**variância das médias**

mostrais igual à variância da população dividida pelo tamanho da amostra), independentemente da distribuição da variável em questão.

Portanto, considerando a distribuição amostral de médias, quando se conhece a variância populacional ou a amostra é grande ($n > 30$), utilizamos a estatística z da distribuição normal vista anteriormente, independentemente da distribuição da população.

Então, por meio do TLC, a estatística será dada por: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$.

Confira a indicação de um programa para cálculo amostral na seção Complementando ao final desta Unidade.

Distribuição t de *Student*

Na prática, muitas vezes não se conhece σ^2 e trabalha-se com amostras pequenas, ou seja, menor ou igual a 30. Assim, você conhece apenas sua estimativa s (desvio padrão amostral). Substituindo σ por seu estimador s, na expressão da variável padronizada, obtemos a seguinte variável:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$ **(expressão semelhante a Z)**

Essa variável segue uma distribuição t de *Student* com $(n - 1)$ **graus de liberdade***.

O **n – 1** corresponde ao divisor do cálculo da variância amostral, ou seja, o número de variáveis na amostra que variam livremente na definição da estatística.

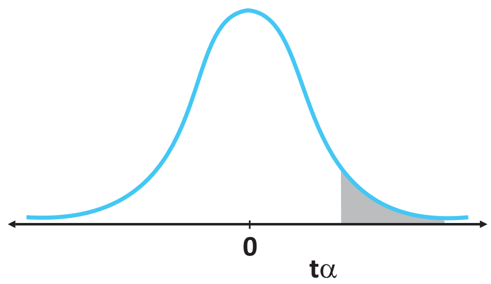
A distribuição t de *Student* apresenta as seguintes características:

- é simétrica em relação à média, que é zero;
- tem forma campanular (semelhante a normal);

*** Graus de liberdade (GL)** – é o número de determinações independentes (dimensão da amostra) menos o número de parâmetros estatísticos a serem avaliados na população. Fonte: Elaborado pelo autor.

- quando n tende para infinito, a distribuição t tende para a distribuição normal. Na prática, a aproximação é considerada boa quando n >30; e
- possui n-1 graus de liberdade.

Vamos aprender a utilizar a tabela da distribuição de t de *Student*. Na tabela t de *Student*, na primeira linha, temos o valor de **α que corresponde à probabilidade (área) acima de um determinado valor da tabela**. Veja a seguir o conceito de **α (área mais escura)**.



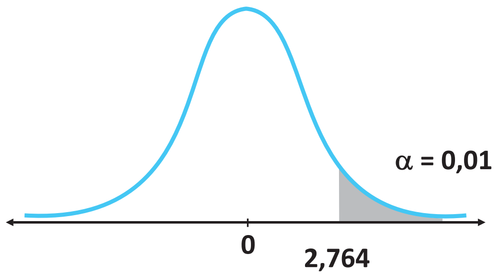
Observe que na Tabela 16, a seguir, temos, na primeira coluna, os graus de liberdade (GL) e, no centro da tabela, os valores da **estatística t de *Student***. Na primeira linha, temos os valores de α.

Tabela 16: Limites unilaterais da distribuição t de *Student* ao nível α de probabilidade

	α								
GL	0.250	0.200	0.150	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656	318.289
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.328
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.894
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
240	0.676	0.843	1.039	1.285	1.651	1.970	2.342	2.596	3.125
480	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107
700	0.675	0.842	1.037	1.283	1.647	1.963	2.332	2.583	3.102
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098

Fonte: Elaborada pelo autor

Para exemplificar o uso da tabela, considere que desejamos encontrar a probabilidade ser maior do que um valor de t igual a 2,764 trabalhando com uma amostra de tamanho n = 11. Portanto, teremos 10 graus de liberdade e, nessa linha, procuremos o valor que desejamos encontrar: 2,764. Subindo na tabela em direção ao α, encontraremos um valor de 0,01 na primeira linha, ou seja, essa é a probabilidade de ser maior do que 2,764 com 10 graus de liberdade.

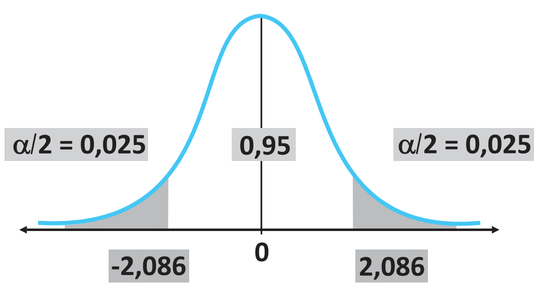


Vamos resolver outro exemplo:

Encontre o valor de t tal que a probabilidade de t (distribuição) esteja entre -t e t e seja igual a 0,95 com 20 graus de liberdade. Isso pode ser representado da forma a seguir:

$$t \,/\, P\,(-t < t < t\,) = 0{,}95 \text{ com 20 gl}$$

A área do meio corresponde a uma probabilidade de 0,95. Então, como a probabilidade total é igual a 1, sobrou 0,05 de probabilidade para ser dividida pelas áreas do lado direito e esquerdo. Observando o valor de $\alpha/2=0{,}025$ (área à direita do valor tabelado) na tabela de t de *Student* e com 20 graus de liberdade, encontraremos o valor de 2,086. Do outro lado, teremos um valor negativo, pois ele está à esquerda da média igual a zero, como você pode ver a seguir.

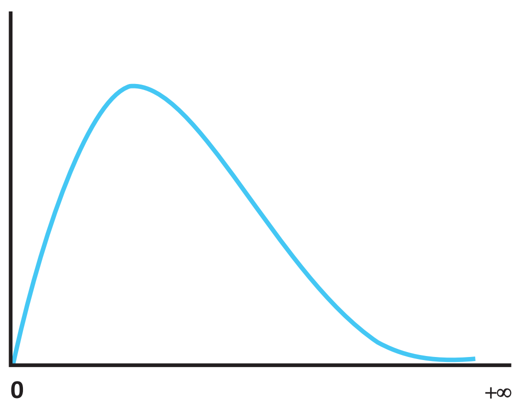


Distribuição de Qui-Quadrado

Retirando uma amostra de n elementos de uma população normal com média μ e variância σ^2, podemos demonstrar que a distribuição amostral da variância amostral segue uma **distribuição de χ^2 (qui-quadrado)** com n-1 graus de liberdade. A variável da estatística de qui-quadrado será dada por:

$\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$ tem distribuição χ^2 com n-1 graus de liberdade.

Essa distribuição é sempre positiva, o que pode ser comprovado pela própria definição da variável. É, ainda, assimétrica à direita, como você pode ver no gráfico da distribuição, a seguir.



Por meio da Tabela 17, você pode ver como é feita a utilização da distribuição de qui-quadrado com graus de liberdade (GL).

Tabela 17: Limites unilaterais da distribuição de χ^2 ao nível α de probabilidade

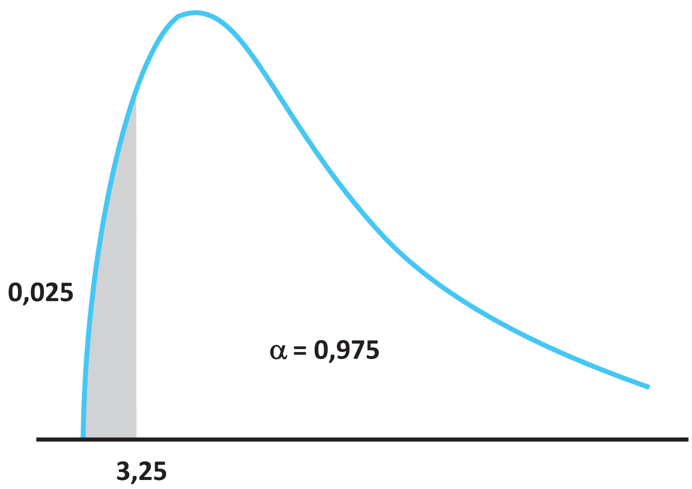
	α												
GL	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005
1	0.0000	0.0002	0.0010	0.0039	0.0158	0.1015	0.4549	1.3233	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	1.3863	2.7726	4.6052	5.9915	7.3778	9.2104	10.5965
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.2125	2.3660	4.1083	6.2514	7.8147	9.3484	11.3449	12.8381
4	0.2070	0.2971	0.4844	0.7107	1.0636	1.9226	3.3567	5.3853	7.7794	9.4877	11.1433	13.2767	14.8602
5	0.4118	0.5543	0.8312	1.1455	1.6103	2.6746	4.3515	6.6257	9.2363	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	3.4546	5.3481	7.8408	10.6446	12.5916	14.4494	16.8119	18.5475
7	0.9893	1.2390	1.6899	2.1673	2.8331	4.2549	6.3458	9.0371	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.3444	1.6465	2.1797	2.7326	3.4895	5.0706	7.3441	10.2189	13.3616	15.5073	17.5345	20.0902	21.9549
9	1.7349	2.0879	2.7004	3.3251	4.1682	5.8988	8.3428	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.1558	2.5582	3.2470	3.9403	4.8652	6.7372	9.3418	12.5489	15.9872	18.3070	20.4832	23.2093	25.1881
11	2.6032	3.0535	3.8157	4.5748	5.5778	7.5841	10.3410	13.7007	17.2750	19.6752	21.9200	24.7250	26.7569
12	3.0738	3.5706	4.4038	5.2260	6.3038	8.4384	11.3403	14.8454	18.5493	21.0261	23.3367	26.2170	28.2997
13	3.5650	4.1069	5.0087	5.8919	7.0415	9.2991	12.3398	15.9839	19.8119	22.3620	24.7356	27.6882	29.8193
14	4.0747	4.6604	5.6287	6.5706	7.7895	10.1653	13.3393	17.1169	21.0641	23.6848	26.1189	29.1412	31.3194
15	4.6009	5.2294	6.2621	7.2609	8.5468	11.0365	14.3389	18.2451	22.3071	24.9958	27.4884	30.5780	32.8015
16	5.1422	5.8122	6.9077	7.9616	9.3122	11.9122	15.3385	19.3689	23.5418	26.2962	28.8453	31.9999	34.2671
17	5.6973	6.4077	7.5642	8.6718	10.0852	12.7919	16.3382	20.4887	24.7690	27.5871	30.1910	33.4087	35.7184
18	6.2648	7.0149	8.2307	9.3904	10.8649	13.6753	17.3379	21.6049	25.9894	28.8693	31.5264	34.8052	37.1564
19	6.8439	7.6327	8.9065	10.1170	11.6509	14.5620	18.3376	22.7178	27.2036	30.1435	32.8523	36.1908	38.5821
20	7.4338	8.2604	9.5908	10.8508	12.4426	15.4518	19.3374	23.8277	28.4120	31.4104	34.1696	37.5663	39.9969
21	8.0336	8.8972	10.2829	11.5913	13.2396	16.3444	20.3372	24.9348	29.6151	32.6706	35.4789	38.9322	41.4009
22	8.6427	9.5425	10.9823	12.3380	14.0415	17.2396	21.3370	26.0393	30.8133	33.9245	36.7807	40.2894	42.7957
23	9.2604	10.1957	11.6885	13.0905	14.8480	18.1373	22.3369	27.1413	32.0069	35.1725	38.0756	41.6383	44.1814
24	9.8862	10.8563	12.4011	13.8484	15.6587	19.0373	23.3367	28.2412	33.1962	36.4150	39.3641	42.9798	45.5584
25	10.5196	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366	29.3388	34.3816	37.6525	40.6465	44.3140	46.9280
26	11.1602	12.1982	13.8439	15.3792	17.2919	20.8434	25.3365	30.4346	35.5632	38.8851	41.9231	45.6416	48.2898
27	11.8077	12.8785	14.5734	16.1514	18.1139	21.7494	26.3363	31.5284	36.7412	40.1133	43.1945	46.9628	49.6450
28	12.4613	13.5647	15.3079	16.9279	18.9392	22.6572	27.3362	32.6205	37.9159	41.3372	44.4608	48.2782	50.9936
29	13.1211	14.2564	16.0471	17.7084	19.7677	23.5666	28.3361	33.7109	39.0875	42.5569	45.7223	49.5878	52.3355
30	13.7867	14.9535	16.7908	18.4927	20.5992	24.4776	29.3360	34.7997	40.2560	43.7730	46.9792	50.8922	53.6719
40	20.7066	22.1642	24.4331	26.5093	29.0505	33.6603	39.3353	45.6160	51.8050	55.7585	59.3417	63.6908	66.7660
50	27.9908	29.7067	32.3574	34.7642	37.6886	42.9421	49.3349	56.3336	63.1671	67.5048	71.4202	76.1538	79.4898
60	35.5344	37.4848	40.4817	43.1880	46.4589	52.2938	59.3347	66.9815	74.3970	79.0820	83.2977	88.3794	91.9518
100	67.3275	70.0650	74.2219	77.9294	82.3581	90.1332	99.3341	109.1412	118.4980	124.3421	129.5613	135.8069	140.1697
120	83.8517	86.9233	91.5726	95.7046	100.6236	109.2197	119.3340	130.0546	140.2326	146.5673	152.2113	158.9500	163.6485

Fonte: Elaborada pelo autor

Para obter probabilidades ou o valor da estatística de qui-quadrado, você irá proceder do mesmo modo que na tabela da distribuição t de *Student*. Na primeira linha, temos os valores de α, na primeira coluna temos os graus de liberdade e no meio da tabela temos os valores da estatística de qui-quadrado.

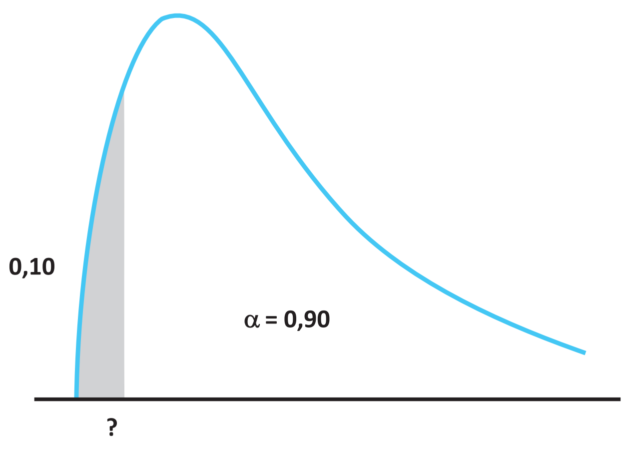
Vamos, então, aprender a olhar a tabela de qui-quadrado.

Encontre a probabilidade de o valor de qui-quadrado ser maior do que 3,25 com 10 graus de liberdade, ou seja, $P(x^2 > 3{,}25) = ?$



Para 10 graus de liberdade e um valor de 3,25 (valor aproximado) na tabela, encontraremos na parte superior um valor de $\alpha = 0{,}975$, que corresponde à probabilidade procurada.

Agora, sabemos que a probabilidade de ser maior que um determinado valor de qui-quadrado é igual a 0,90 ($P(x^2 > ?) = 0{,}9$ com 15 graus de liberdade. Então, o valor da interrogação (?) será obtido na tabela de qui-quadrado.



Observando a tabela de qui-quadrado com 15 graus de liberdade e um valor de $\alpha = 0{,}90$, encontraremos no meio da tabela um valor de 8,55, que será o valor de qui-quadrado, cuja probabilidade de ser maior do que ele é de 0,90 (α).

Distribuição de F

A **distribuição de F de Fischer-Snedecor** corresponde à distribuição da razão de duas variâncias. Temos, então, duas populações que apresentam variâncias populacionais e delas são retiradas amostras nas quais são calculadas variâncias amostrais. A relação entre essas variâncias é que nos dá a distribuição de F. A estatística da distribuição é apresentada a seguir:

$$F = \frac{s_A^2 / \sigma_A^2}{s_B^2 / \sigma_B^2}$$

Segue uma distribuição F com $v_1 = n_1 - 1$ e $v_2 = n_2 - 1$ graus de liberdade para o numerador e o denominador, respectivamente.

Uma das tabelas de F de Snedecor é apresentada a seguir:

Tabela 18: Limites unilaterais da distribuição F de Fischer–Snedecor ao nível de 10% de probabilidade

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857	60.195	60.473	60.705	60.902	61.073	61.220	61.740	62.529	62.794	63.061	63.194
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401	9.408	9.415	9.420	9.425	9.441	9.466	9.475	9.483	9.487
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222	5.216	5.210	5.205	5.200	5.184	5.160	5.151	5.143	5.138
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907	3.896	3.886	3.878	3.870	3.844	3.804	3.790	3.775	3.768
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282	3.268	3.257	3.247	3.238	3.207	3.157	3.140	3.123	3.114
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920	2.905	2.892	2.881	2.871	2.836	2.781	2.762	2.742	2.732
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684	2.668	2.654	2.643	2.632	2.595	2.535	2.514	2.493	2.482
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519	2.502	2.488	2.475	2.464	2.425	2.361	2.339	2.316	2.304
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396	2.379	2.364	2.351	2.340	2.298	2.232	2.208	2.184	2.172
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302	2.284	2.269	2.255	2.244	2.201	2.132	2.107	2.082	2.069
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227	2.209	2.193	2.179	2.167	2.123	2.052	2.026	2.000	1.986
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.166	2.147	2.131	2.117	2.105	2.060	1.986	1.960	1.932	1.918
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.116	2.097	2.080	2.066	2.053	2.007	1.931	1.904	1.876	1.861
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.073	2.054	2.037	2.022	2.010	1.962	1.885	1.857	1.828	1.813
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.037	2.017	2.000	1.985	1.972	1.924	1.845	1.817	1.787	1.771
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	2.005	1.985	1.968	1.953	1.940	1.891	1.811	1.782	1.751	1.735
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.978	1.958	1.940	1.925	1.912	1.862	1.781	1.751	1.719	1.703
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.954	1.933	1.916	1.900	1.887	1.837	1.754	1.723	1.691	1.674
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.932	1.912	1.894	1.878	1.865	1.814	1.730	1.699	1.666	1.649
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.913	1.892	1.875	1.859	1.845	1.794	1.708	1.677	1.643	1.626
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.896	1.875	1.857	1.841	1.827	1.776	1.689	1.657	1.623	1.605
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.880	1.859	1.841	1.825	1.811	1.759	1.671	1.639	1.604	1.586
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.866	1.845	1.827	1.811	1.796	1.744	1.655	1.622	1.587	1.568
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.853	1.832	1.814	1.797	1.783	1.730	1.641	1.607	1.571	1.552
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841	1.820	1.802	1.785	1.771	1.718	1.627	1.593	1.557	1.538
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.830	1.809	1.790	1.774	1.760	1.706	1.615	1.581	1.544	1.524
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.820	1.799	1.780	1.764	1.749	1.695	1.603	1.569	1.531	1.511
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.811	1.790	1.771	1.754	1.740	1.685	1.592	1.558	1.520	1.500
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.802	1.781	1.762	1.745	1.731	1.676	1.583	1.547	1.509	1.489
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794	1.773	1.754	1.737	1.722	1.667	1.573	1.538	1.499	1.478
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.737	1.715	1.695	1.678	1.662	1.605	1.506	1.467	1.425	1.402
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.703	1.680	1.660	1.643	1.627	1.568	1.465	1.424	1.379	1.354
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.680	1.657	1.637	1.619	1.603	1.543	1.437	1.395	1.348	1.321
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680	1.653	1.629	1.609	1.590	1.574	1.513	1.403	1.358	1.307	1.278
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.636	1.612	1.592	1.573	1.557	1.494	1.382	1.336	1.282	1.250
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.625	1.601	1.580	1.562	1.545	1.482	1.368	1.320	1.265	1.232
240	2.727	2.325	2.107	1.968	1.871	1.799	1.742	1.696	1.658	1.625	1.598	1.573	1.552	1.533	1.516	1.451	1.332	1.281	1.219	1.180

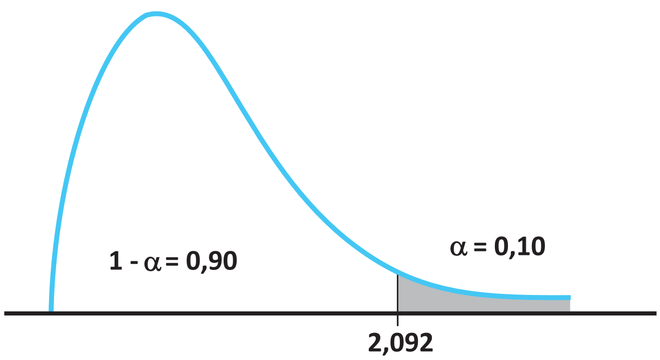
Fonte: Elaborada pelo autor

Note que, no caso da tabela de F, o valor de α que corresponde à área extrema à direita da curva é apresentado no título da tabela, pois, para cada valor de α, temos uma tabela diferente.

Encontramos uma aplicação prática da distribuição de F na verificação da homogeneidade das variâncias provenientes de duas populações normais e independentes. Então, encontre o valor de F1 cuja probabilidade de ser maior do que ele é 0,10 com 5 e 25 graus de liberdade, ou seja, $P(F > F1) = 0{,}10$ com $v_1 = 5$ e $v_2 = 25$ gl.

Como temos a probabilidade do resultado ser maior do que um valor de F, esse valor corresponde ao valor de α. Precisaremos, então, trabalhar com a tabela que apresenta 10% de probabilidade no título, como a Tabela 18.

Observando $v_1 = 5$ e $v_2 = 25$, encontraremos um valor de F igual a 2,092.



Noções de Estimação

Um dos principais objetivos da estatística inferencial consiste em estimar os valores de parâmetros populacionais desconhecidos (estimação de parâmetros) utilizando dados amostrais. Então, qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória, desde que esta amostra represente bem a população.

Os parâmetros populacionais mais comuns a serem estimados são: a média, o desvio padrão e a proporção.

A estatística inferencial apresenta uma relevância alta, já que a utilização de dados amostrais está associada à maioria das decisões que um gestor ou um pesquisador deve tomar. Consiste em tirar conclusões de uma população a partir de amostra representativa dessa população, tendo isso grande importância em muitas áreas do conhecimento.

A partir de uma amostra de 800 clientes (escolhidos aleatoriamente entre todos os clientes que abasteceram na primeira quinzena de um determinado mês) de um posto de gasolina que possuem carros populares, verificou-se que o consumo médio de gasolina foi de R$ 200,00 por quinzena.

Reflita sobre a afirmação a seguir:

Podemos inferir que o consumo médio da população de clientes da primeira quinzena do mês em estudo, proprietários de carros populares que abastecem nesse posto de gasolina, é de R$ 200,00.

Esta é uma estimativa que chamamos de **pontual**, ou seja, inferimos sobre a população considerando apenas o valor da estimativa. Essas estimativas por ponto não nos dão uma ideia sobre confiança e sobre as margens de erro que deveriam ser aplicadas ao resultado. Tudo que nós sabemos, por exemplo, é que o consumo médio de gasolina foi estimado em R$ 200,00 por quinzena, independentemente do tamanho da amostra e da variabilidade inerente aos dados. Se fosse usado um tamanho grande de amostra e houvesse pouca variabilidade, teríamos grandes razões para acreditar no resultado; mas não sabemos nada quando temos apenas uma estimativa por ponto.

Entretanto, podemos estimar ou fazer inferências sobre os valores da população usando uma segunda abordagem chamada de **estimativas por intervalos** ou **intervalos de confiança**, que dão o intervalo dentro do qual se espera que esteja o valor da população, com uma dada probabilidade ou um nível de confiança. Nesse caso, poderíamos inferir, por exemplo, que o consumo de carros populares que abastecem no posto de gasolina está no intervalo de R$180,00 a R$ 220,00 e, ainda, afirmaríamos isso com, por exemplo, 95% de certeza.

Como a estimativa por intervalos nos fornece uma informação mais precisa em relação ao parâmetro, esta é a melhor forma de se estimar o parâmetro populacional. Então, para você estimar parâmetros populacionais por meio de dados amostrais, é necessário o conhecimento da distribuição amostral da estatística que está sendo usada como estimador.

Na seção Distribuições Amostrais, abordamos esse assunto. Se julgar necessário, retome o conteúdo.

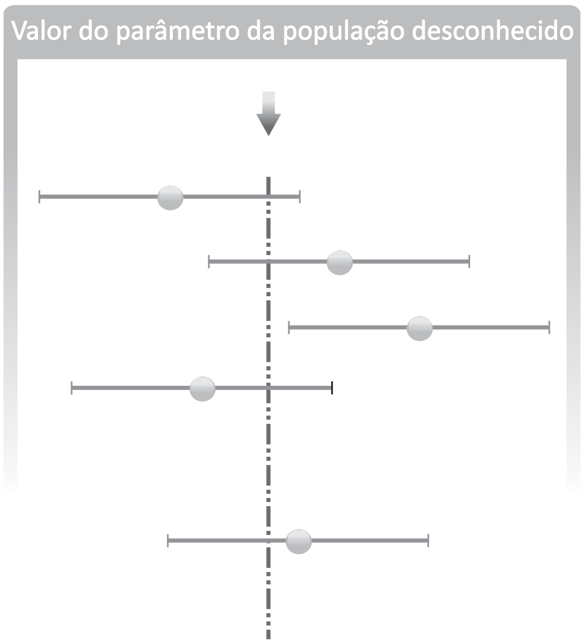
Em resumo, podemos dizer que a estimativa pontual fornece uma estimativa única de um parâmetro e que a estimativa intervalar nos dá um intervalo de valores possíveis, no qual se admite que esteja o parâmetro populacional com uma probabilidade conhecida.

Estimação por Intervalos

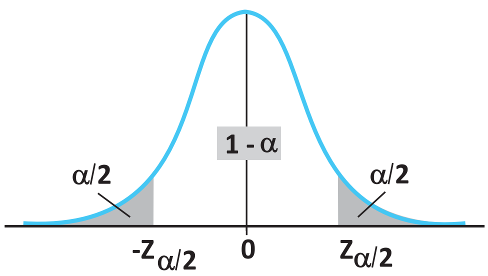
Você irá ver agora que um **intervalo de confiança** dá um intervalo de valores, centrado na estatística amostral, no qual julgamos, com um risco conhecido de erro, estar o parâmetro da população.

É o nível de significância que nos dá a medida da **incerteza** dessa inferência. O α geralmente assume valores entre 1 e 10%.

Então, a partir de informações de amostras, devemos calcular os limites de um intervalo, valores críticos, que em (1-α)% dos casos inclua o valor do parâmetro a estimar e em α% dos casos não inclua o valor do parâmetro, como podemos ver no desenho abaixo.



O nível de confiança $1 - \alpha$ é a probabilidade de o intervalo de confiança conter o parâmetro estimado. Em termos de variável normal padrão Z, isso representa a área central sob a curva normal entre os pontos $-Z$ e Z.



Você pode observar que a área total sob a curva normal é unitária. Se a área central é $1 - \alpha$, o ponto $-z$ representa o valor de Z, que deixa à sua esquerda a área $\alpha/2$, e o ponto z representa o valor de Z, que deixa à sua direita a área $\alpha/2$.

Vamos aprender agora a construir o intervalo de confiança para uma média quando o desvio padrão populacional é conhecido ou a amostra é grande.

Vamos imaginar a seguinte situação: o Departamento de Recursos Humanos de uma prefeitura informa que o tempo de execução de tarefas que envolvem participação manual varia de tarefa para tarefa, mas que o desvio padrão permanece aproximadamente constante, em 3 minutos. Novas tarefas estão sendo implantadas na prefeitura. Uma amostra aleatória do tempo de execução de 50 das novas tarefas forneceu o valor médio de 15 minutos. Determine um intervalo de confiança de 95% para o tempo médio de execução de uma dessas novas tarefas.

Primeiramente, você precisa identificar que o desvio padrão populacional é conhecido e também a amostra é considerada grande ($n > 30$). Então, a construção do intervalo de confiança será feita utilizando a média amostral. Utilizaremos, para a obtenção dos limites de confiança, a curva normal padrão Z.

Como os limites são dados por meio da estatística calculada a partir dos dados amostrais e da margem de erro (fornecido pela estatística da distribuição multiplicada pelo desvio padrão da distribuição amostral), teremos, nessa situação, os limites calculados por meio da seguinte expressão:

$$\boxed{\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}}, \text{ onde } \boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Erro padrão da média, obtido a partir do cálculo do desvio padrão da média de uma amostra aleatória simples retirada de uma população.

Margem de erro $(e = z_{\alpha/2}\,\sigma_{\bar{x}})$

Logo, o intervalo de confiança tem centro na média amostral:

Calculando, teremos:

$$1\text{-}\alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow \alpha/2 = 0{,}025$$

Olhando na tabela de Z, você encontrará $Z_{\alpha/2} = 1{,}96$

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1{,}96.\frac{3}{\sqrt{50}} = 0{,}8315$$

$$P(\bar{x} - e < \mu < \bar{x} + e) = (1 - \alpha)$$

$$P(15 - 0{,}8315 < \mu < 15 + 0{,}8315) = 0{,}95$$

$$P(14{,}168 < \mu < 15{,}831) = 0{,}95$$

Interpretação do resultado: em cada grupo de 100 amostras retiradas de 50 pessoas, espera-se que, em 95 delas, a média esteja dentro do intervalo de 14,168 a 15,831.

Antes de continuar a leitura, você deve realizar, ao final desta Unidade, a Atividade 6, na qual irá aplicar os conhecimentos relacionados à amostra e ao intervalo de confiança. Em caso de dúvida, faça contato com seu tutor.

Dimensionamento de Amostras

Desenvolvendo a expressão de erro mostrada anteriormente, obteremos o tamanho de amostra para estimação da média populacional quando o desvio padrão populacional for conhecido, como mostramos a seguir:

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = z_{\alpha/2} \cdot \frac{\sigma}{e} \quad \Rightarrow \quad n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e} \right)^2 \quad \Rightarrow \quad n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{e^2}$$

Imagine a seguinte situação: que tamanho de amostra será necessário para produzir um intervalo de 95% de confiança para a verdadeira média populacional, com erro de 1,0, se o desvio padrão da população é 10,0?

Substituindo esses valores na expressão, teremos:

$$n_o = \frac{Z^2_{\alpha/2} . \sigma^2}{e^2} = \frac{1{,}96^2 . 10^2}{1^2} = 384{,}16 \cong 385$$

Você pode alterar a confiança que teremos um diferente valor de Z e também o erro. Isso irá depender da precisão que você irá desejar nas suas estimativas.

Quando trabalhamos com proporção de sucesso, podemos substituir a variância por p.q (proporção de sucesso vezes a proporção de fracasso) da Distribuição de Bernoulli.

$$n = \frac{Z^2_{\alpha/2} . \hat{p}\hat{q}}{e^2}$$

Onde $\hat{p}$ e $\hat{q}$ correspondem às estimativas de sucesso e de fracasso, respectivamente, obtidos a partir de resultados amostrais.

Vamos ver uma aplicação?

Um setor da prefeitura que cuida da documentação de imóveis está interessado em **estimar** a proporção de pessoas que compram novos imóveis na cidade para melhor dimensionar o setor de atendimento. Para isso, amostrou 80 pessoas do seu cadastro, verificando que 30 delas teriam comprado imóvel no último ano. Determine o tamanho da amostra necessário para estimar com 95% de confiança essa proporção e com erro máximo de 4%.

Substituindo os valores, teremos:

$$\hat{p} = \frac{30}{80} = 0{,}375 \qquad e \qquad \hat{q} = 1 - \hat{p} = 1 - 0{,}375 = 0{,}625$$

$$n = \frac{Z^2_{\alpha/2} \cdot \hat{p}\hat{q}}{e^2} = \frac{1{,}96^2 \cdot 0{,}375 \cdot 0{,}625}{0{,}04^2} = 562{,}73 \cong 563$$

Complementando...

Através do *link* que apresentamos a seguir, você poderá fazer cálculos das distribuições de probabilidade discretas ou contínuas, de dimensionamento de amostras e de intervalos de confiança.

Programa estatístico Bioestat. Disponível em: <http://www.mamiraua.org.br/download/Default.aspx?dirpath=e:\home\mamiraua\Web\download\BioEstat 5 Portugues&tipo=diretorio>. Acesso em: 19 nov. 2010.

Resumindo

Nesta Unidade, você aprendeu sobre as principais distribuições de probabilidade, sejam elas discretas ou contínuas, e como utilizá-las. Também conheceu as distribuições de amostragem e, quando utilizá-las, e noções básicas de estimação (intervalos de confiança) e dimensionamento de amostras. Essas informações serão muito importantes para a compreensão da próxima Unidade.

Atividades de aprendizagem

Para verificar se você está acompanhando o que apresentamos nesta Unidade, procure responder às atividades propostas, a seguir. Se tiver dificuldades para resolvê-las, consulte seu tutor.

1. No Brasil, a proporção de microempresas que fecham em até um ano de atividade é de 10%. Em uma amostra aleatória de 20 microempresas, qual a probabilidade de 5 terem fechado em até um ano de sua criação?
2. Entre 2.000 famílias de baixa renda e com quatro crianças, considerando-se que a chance de nascer uma criança do sexo masculino é igual a do sexo feminino, em quantas família se esperaria que tivessem:

    a) Dois filhos do sexo masculino.

    b) Um ou dois filhos do sexo masculino.

    c) Nenhum filho do sexo feminino.

3. A ouvidoria de uma prefeitura recebe em média 2,8 reclamações/hora, segundo uma Distribuição de Poisson. Determine a probabilidade de chegarem duas ou mais reclamações em um período de:

    a) 30 minutos.

    b) 1 hora.

    c) 2 horas.

4. As rendas mensais de funcionários do setor de arrecadação de uma prefeitura são normalmente distribuídas com uma média de R$ 2.000,00 e um desvio padrão de R$ 200,00. Qual é o valor de Z para uma renda X de R$ 2.200,00 e de R$ 1.700,00?
5. O uso diário de água por pessoa em uma determinada cidade é normalmente distribuído com média μ igual a 20 litros e desvio padrão σ igual a 5 litros.

   a) Que percentagem da população usa entre 20 e 24 litros por dia?
   b) Que percentagem usa entre 16 e 20 litros?
   c) Qual é a probabilidade de que uma pessoa selecionada ao acaso use mais do que 28 litros?

6. Considere que as despesas mensais com alimentação em restaurantes de comida a quilo para um casal são normalmente distribuídas com desvio padrão de R$ 3,00. Uma amostra de 100 casais revelou uma despesa média de R$ 27,00. Determine o intervalo de confiança de 95% para a despesa com alimentação de casais.