Variável aleatória

Quando uma variável tem resultados ou valores que tendem a variar de uma observação para outra em razão de fatores relacionados com a chance, nós chamamos de variável aleatória. Definimos uma variável aleatória associada a uma amostra ou experimento, de tal modo que seus resultados possíveis sejam numéricos. Por exemplo, a jogada de uma moeda tem dois resultados, K ou C, que não são numéricos. Podemos então considerar nossa variável aleatória sendo o número de caras em uma jogada, que tem os valores numéricos possíveis 0 e 1. Uma variável aleatória (v.a) é uma função com valores numéricos, cujo valor é determinado por fatores de chance. As variáveis aleatórias podem ser discretas ou contínuas. Uma variável aleatória é dita discreta se toma valores que podem ser contados, e é dita contínua quando pode tomar qualquer valor de um determinado intervalo.

Distribuição binomial

É utilizada para calcular a probabilidade de experimentos que apresentam duas possibilidades, sucesso ou fracasso.

A expressão para o calculo da probabilidade de uma distribuição binomial é dada por: $P(x)=C_{n,x}.p^x.q^{n-x}$, onde $C_{n,x}$ é a combinação de n tomados x a x, p é a probabilidade de sucesso, q é a probabilidade de fracasso (q=1-p), n é o número de observações ou provas idênticas e x é o número de sucessos esperados.

Exemplo: Uma moeda é lançada quatro vezes, qual a probabilidade de sair cara. a) uma vez, b) três vezes, c) pelo menos uma vez.

a) Vamos considerar a probabilidade de sucesso sendo a probabilidade de sair cara, logo p=1/2 e a probabilidade de fracasso sendo a probabilidade de coroa, isto é q=1-p=1-1/2=1/2, logo a probabilidade ocorrer um sucesso será $P(1)=C_{4,1}(1/2)^1.(1/2)^3=0,25$; b) $P(3)=C_{4,3}(1/2)^3(1/2)^1=0,25$, c) $P(1)+P(2)+P(3)+P(4)=1-P(0)=1-C_{4,0}.(1/2)^0.(1/2)^4=1-0,0625=0,9375$.

- Média e variância da distribuição binomial

A média da distribuição binomial é o número de observações vezes a probabilidade de ocorrência do evento, isto é $\mu=np$. A variância é igual ao número de observações vezes a probabilidade de sucesso vezes a probabilidade de fracasso, isto é $\sigma^2=n.p.q$, logo o desvio padrão é $\sigma = \sqrt{n.p.q}$.

Exercício: Um exame de múltipla escolha consiste em 10 questões, cada uma com 4 opções. A aprovação no exame exige do aluno pelo menos nota seis, ou seja o acerto de pelo menos seis questões. Qual a chance de aprovação: a) se o aluno nada estudou? b) se o aluno estudou suficiente para poder eliminar duas escolhas, devendo escolher apenas entre duas opções.
Resp: a) p=1/4 e q=3/4 P(6)+P(7)+P(8)+P(9)+P(10)=0,0197 , b) p=1/2 e q=1/2 P(6)+P(7)+P(8)+P(9)+P(10)=0,3770.

Exercício uma equipe de basquete tem probabilidade 0,88 de vitórias sempre que joga. Se o time atuar 4 vezes, determine a probabilidade de que vença: a) todas as partidas; b)

exatamente 2 partidas; c) pelo menos uma partida; d) no máximo 3 partidas; e) mais da metade das partidas.
Resp: a) 59,9695%; b) 6,6908%; c) 99,9793%; d) 40,0305%; e) 92,6802%.

Exercício: Supondo que a probabilidade de um casal ter filhos com olhos verdes é de 17%, em 400 famílias com 4 crianças cada uma, quantas se esperaria que tivesse: a) dois filhos com olhos verdes; b) nenhum dos filhos com olhos verdes.

a) 48 famílias; b) 190 famílias.

Distribuição de Poisson

É uma distribuição discreta de probabilidade, dita uma regra matemática, que serve para descrever a probabilidade do número de ocorrências num campo ou intervalo contínuo (tempo ou espaço) de eventos que não ocorrem com muita freqüência. Somente um valor é necessário para determinar a probabilidade de um determinado número de sucessos, que é o número médio de sucessos para a específica dimensão de tempo ou espaço. A expressão da distribuição de Poisson é: P(x)= $\frac{e^{-\lambda t}(\lambda t)^x}{x!}$; onde λ é o número médio de sucessos em um determinado intervalo de tempo ou espaço, t é o intervalo de tempo ou espaço contínuo de observações que se está analisando e x é o número de sucessos no intervalo desejado.

Exemplo: Uma pizzaria recebe em média 8 chamadas por hora. Qual a probabilidade de que, em uma hora selecionada aleatoriamente, sejam recebidas exatamente 5 chamadas?

Temos que a média é λ=8, o intervalo de tempo é t=1 e o número de sucessos no intervalo dado é x=5, logo P(5)= $\frac{e^{-8.1}(8.1)^5}{5!} = 0{,}091603662$.

Exemplo: Em uma certa rodovia aparecem em média na proporção de um buraco a cada 500m. Qual a probabilidade de aparecer 4 buracos em um intervalo de 1500m?

Temos que a média é de 1 buraco a cada 500m, logoλ=1/500, o intervalo de espaço é t=1500 e o número de sucessos no intervalo dado é x=4, logo P(4)= $\frac{e^{-3}(3)^4}{4!} = 0{,}168031356$.

Exercício: Suponha que haja 265 erros de pontuação distribuídos aleatoriamente em um contrato comercial de 458 linhas. Encontre a probabilidade de conter em uma dada linha 2 erros apenas.
Resp: 0,093852722.

Exercício: Um a cada cem carros cai num buraco de uma avenida. Se cem carros passarem, qual a probabilidade de dois carros caírem no buraco?
Resp: 0,183940.

Exercício: Numa central telefônica chegam 300 telefonemas por hora. Qual a probabilidade de que em um minuto não haja nenhum chamado.
Resp: 0,006738.

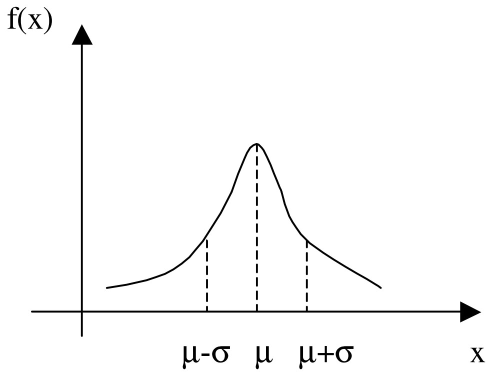
Distribuição normal

É utilizada na análise de variáveis aleatórias contínuas (intervalos reais).

Formula: $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, onde:

x = Variável aleatória normal.
σ = Desvio padrão.
μ = Média.

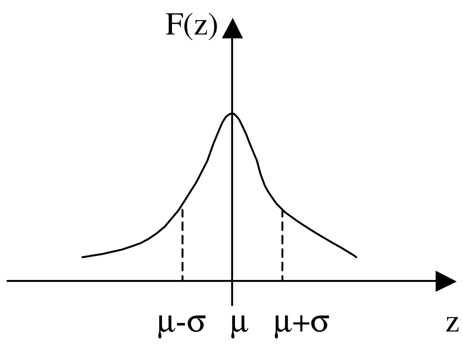
Gráfico da distribuição f(x).



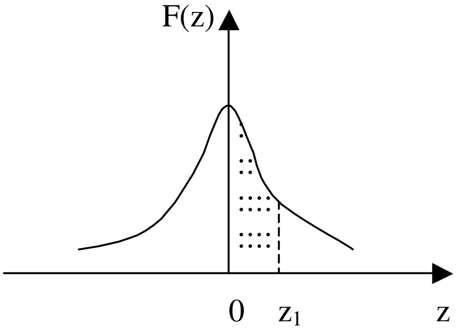
Características do gráfico de f(x).

1-Simétrico em torno da média.
2- Não toca o eixo 0x (0x é uma assíntota).
3-A distribuição normal fica delimitada entre a média e o desvio padrão.
4-A área sob o gráfico de f é igual a 1 (100%).
5-A área entre dois pontos, correspondente a probabilidade do valor de uma variável aleatória entre estes dois pontos.
6-Apresenta um único pico que é a média, logo a média, mediana e moda apresentam mesmo valor.

Fazendo $z = \frac{x-\mu}{\sigma}$, podemos simplificar nossa função f(x), desta forma o gráfico da distribuição normal fica como no gráfico abaixo.



Temos que a probabilidade de uma variável aleatória z estar entre 0 e z_1 sendo a área sob a curva da função f(z) que vai de 0 a z_1, logo é a integral $\int_0^{z_1} f(z)dz$.



Temos uma tabela da distribuição normal que nos dá a área sob a curva f(z) para valores que vão de 0 a z.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	,01179	0,1217	,01255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	,03389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,42650	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,10	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,20	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,30	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4995	0,4996	0,4997
3,40	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,50	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,60	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4999	0,4999	0,4999
3,70	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,80	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,90	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
4,00	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Use a tabela acima para cálculo de probabilidades da distribuição normal para resolver os seguintes exercícios:

1) Com o auxílio da tabela da distribuição normal, encontre a área para os seguintes valores de z:
   a) $0<z<1{,}23$; b)$-2{,}15< z< 0$; c) $-1{,}56 < z < 1{,}48$
   Resp: a) 0,3907 ; b) 0,4842 ; c) 0,4406+0,4306=0,8712

2) Estima-se que a vida útil de um aparelho de TV, segue uma distribuição normal, com Média μ = 4000 horas e desvio padrão de 500 horas. Qual a probabilidade de um aparelho escolhido aleatoriamente durar entre 4000 e 4500 horas?

   Para x=4000, temos, $z=\dfrac{x-\mu}{\sigma}=\dfrac{4000-4000}{500}=0$.
   Para x=4500, temos, $z=\dfrac{x-\mu}{\sigma}=\dfrac{4500-4000}{500}=\dfrac{500}{500}=1$.
   Logo P(4000< x < 4500) = P(0<z<1)=0,3413=34,13%.

3) Uma máquina empacotadora de milho, foi calibrada de modo que sua média é μ=15Kg de milho colocados no saco, com desvio padrão de 0,3Kg. Sabendo-se que a distribuição dos pesos segue uma distribuição normal, determine a probabilidade de um saco escolhido aleatoriamente, ter peso entre 14,1Kg e 14,7Kg.
   Para x=14,1, temos, $z=\dfrac{x-\mu}{\sigma}=\dfrac{14{,}1-15}{0{,}3}=\dfrac{-0{,}9}{0{,}3}=-3$.
   Para x= 14,7, temos, $z=\dfrac{x-\mu}{\sigma}=\dfrac{14{,}7-15}{0{,}3}=\dfrac{-0{,}3}{0{,}3}=-1$.
   Logo P(14,1< x < 14,7) = P(-3<z<-1)=0,4986 – 0,3413 = 0,8399%.

4) As vendas de uma loja seguem aproximadamente uma distribuição normal, com média \$400,00 e desvio padrão de \$100,00. Qual a probabilidade de que em um determinado dia a loja venda: a) Entre \$450,00 e \$650,00 ; b) Entre \$350,00 e \$500,00
   Resp: a)30,23% ; b) 53,28%

5) A média de um concurso foi de μ = 75, com desvio padrão σ = 8.Determine: a) Os escores reduzidos de dois candidatos, cuja pontuação foram 95 e 60, respectivamente. b) As pontuações de dois candidatos cujo escores reduzidos foram, respectivamente, -0,5 e 1,5
   Resp: a) z = -1,88 e z = 2,5 ; b) x=71 e x=87

6) Os pesos de 500 estudantes são normalmente distribuidos com média μ=64,8Kg e desvio padrão σ=4,8Kg. Encontre o número de alunos que pesam: a) entre 60 e 75Kg ; b) Mais de 62,7Kg.
   Resp: a) 412 alunos ;b) 333 alunos.

7) As alturas de 20.000 alunos de um colégio têm distribuição normal, com média 1,64m e desvio padrão 0,16m. a) Qual o número esperado de alunos com altura superior a 1,52m.? b) Qual o intervalo simétrico em torno da média, que conterá 78% das alturas dos alunos?
   Resp; a) $0{,}7734\times 20.000$ = 15468 alunos b) entre 1,4432m e 1,8368m.

8) A duração de certo componente eletrônico pode ser considerada normalmente distribuída com média 850 dias e desvio padrão de 45 dias. Calcular a probabilidade de um componente durar: a) Entre 700 e 1000 dias; b) Mais de 800 dias; c) Menos de 750 dias; c) Exatamente 1000 dias. *Resposta: a)1; b) 0,8665; c) 0.*

9) Os pesos de 600 estudantes são normalmente distribuídos com média de 65,3 Kg e desvio padrão de 5,5 Kg. Encontre o número de alunos que pesam: a) Entre 60 e 70 Kg; b) Mais de 63,2 Kg. *Resposta: a) 380; b) 389.*

10) Em uma distribuição normal, 28% dos elementos são superiores a 34 e 12% inferiores a 19. Encontrar a média e a variância da distribuição. *Resposta: a) 73,47; b) 29,03.*

Fundamentos de Amostragem

- População: É o conjunto formado por indivíduos ou objetos que têm pelo menos uma variável comum e observável. Podendo ser finita ou infinita.

Exemplos: Temos os seguintes exemplos de população finita: Número de alunos matriculados em uma determinada faculdade, número de livros de uma biblioteca, número de funcionários de uma empresa, etc. Temos os seguintes exemplos de população infinita: Número de ligações recebidas por uma empresa, número de acidentes de trânsito que ocorrerá no próximo ano, número de pessoas que entram e saem em um determinado shopping, etc. Dizemos que estes exemplos fazem parte de uma população infinita por não termos como contar ou controlar estes números.

- Amostra: É um subconjunto de uma população.
- Amostragem: É o processo de seleção de uma amostra, que possibilita o estudo das características da população.
- Censo: Examina todos os elementos da população.

Vantagens e desvantagens da amostragem e do censo:

Em caso de populações infinitas utilizamos a amostragem. No caso de testes destrutivos nós também utilizamos a amostragem, por exemplo, para testarmos a vida útil de determinado lote de lâmpadas, nós tomamos uma amostra deste lote e a testamos. Quando queremos obter informações mais rápidas nós utilizamos a amostragem.

Temos como desvantagem do censo o seu alto custo. Porém na análise de populações pequenas é recomendado o censo, por exemplo, se formos analisar os moradores de uma determinada casa, nós tomaremos toda população. Uma vantagem do censo é que temos uma precisão completa da análise.

Tipos de Amostragem

- Amostragem aleatória

É a retirada de uma amostra da população de forma aleatória

- Amostragem com reposição e sem reposição

Vantagens e desvantagens

Alto custo: com reposição pelo fato de podermos analisar o mesmo elemento mais de uma vez.

Caráter destrutivo: com reposição não é possível.

Classificação das Amostragens

- Amostragem probabilística: É a probabilidade de seleção de uma amostra na qual cada unidade amostral da população tem probabilidade diferente de zero e conhecida de pertencer à amostra
- Amostragem não probabilística: Neste plano de amostra a probabilidade de seleção é desconhecida para alguns ou todos elementos da população, podendo alguns elementos ter probabilidade nula de pertencer à amostra. Ex: Amostra por voluntários.

As amostragens probabilísticas subdividem-se em:

- Aleatória ou Sistemática: É a amostragem em que todos os elementos da população têm a mesma probabilidade de serem selecionados
- Estratificação: Propõe a divisão da população em subgrupos homogêneos. Ex: Pesquisa entre pessoas de classes diferentes.
- Por conglomerado: Propõe a divisão em subgrupos heterogêneos.

Inferências estatísticas

Estimação: Estimar é a ação de fazer uma suposição generalizada a respeito de um todo baseado em informações lógicas. Exemplo: Uma pesquisa mostrou que em 1999, 90% dos acidentes de trânsito foram com homens e 1% com mulheres. A mesma pesquisa foi feita em 2000 e mostrou que 95% eram homens e 5% mulheres. Não podemos concluir que o número de acidentes com homens aumentaram, pois não foi aplicado à pesquisa nenhum fator que minimizasse as margens de erros. Usamos então o fator de correção que minimizam estes tipos de erros e falaremos sobre eles mais tarde.

- Parâmetro: É uma função do conjunto de valores da população. Ex: Média Aritmética, variância, etc.
- Estimativa: É o valor assumido pelo parâmetro em determinada amostra.

Estimar parâmetros é basear-se nos resultados da amostra para estimá-los à população. No exemplo acima temos que o parâmetro usado para análise foi a proporção. A estimativa para caracterizar a população foi de 90% para os homens em 1999.

Estimadores mais usados

Parâmetro Populacional	*Estimadores*
Média	$\bar{x}$
Diferença entre as médias de duas populações	$\bar{x}_1 - \bar{x}_2$
Proporção	p
Diferença entre proporções de duas populações	$p_1 - p_2$

Desvio padrão σ

Formas de estimativas

- Estimativa pontual: Determina um valor específico de um parâmetro.

- Estimativa intervalar: Dá um intervalo de valores possíveis onde está o valor parâmetro populacional.

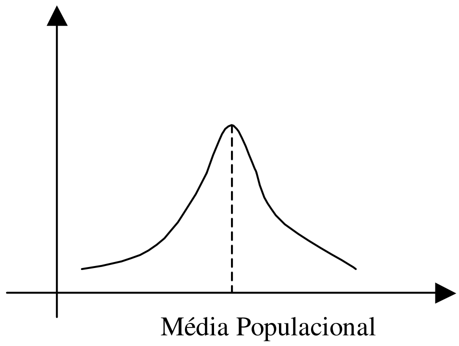
Exemplo: Temos a porcentagem de acidentes registrado envolvendo homens e mulheres no ano de 1999:

Estimativa pontual: 90% de homens e 10% de mulheres.
Estimativa Intervalar: entre 88% e 92% de homens e entre 8% e 12% de mulheres.

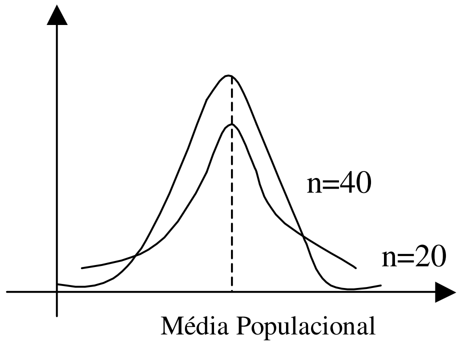
Distribuição de médias amostrais

É uma distribuição de probabilidade que indica a probabilidade das médias amostrais. Esta distribuição é uma função da média e do desvio padrão populacional e do tamanho da amostra.

- As médias amostrais tendem a agrupar-se em torno da média populacional.



- As distribuições amostrais de grandes amostras têm menor variabilidade que as de pequeno tamanho amostral. (n=tamanho da amostra)



Definimos por $\mu_{\bar{x}}$ sendo a média da população amostral e por μ_x a média da população.
A média de uma distribuição amostral é sempre igual a média populacional, isto é, $\mu_{\bar{x}} = \mu_x$. Quando o tamanho da população é muito grande ou infinita, ou quando n≥30, o desvio padrão da

distribuição de média amostral será $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$, onde $\sigma_{\bar{x}}$ é o desvio (ou dispersão) padrão amostral, σ_x é o desvio (ou dispersão) padrão da população e n é o tamanho da amostra.

Notamos que a dispersão (desvio padrão) da distribuição amostral ($\sigma_{\bar{x}}$) depende da dispersão (desvio padrão) da população (σ_x) e do tamanho da amostra (n). Uma amostra maior resulta em uma menor variabilidade entre possíveis médias amostras. Populações muito dispersas geram maior variabilidade entre as médias amostrais.

Teorema do limite central

Há uma tendência para as distribuições de média e de proporções aproximarem-se da distribuição normal. Se uma população tem distribuição normal, a distribuição das médias amostrais também será normal, para qualquer tamanho de amostra. Se uma população não tem distribuição normal, para amostras grandes a distribuição das médias amostrais pode ser aproximada para normal.

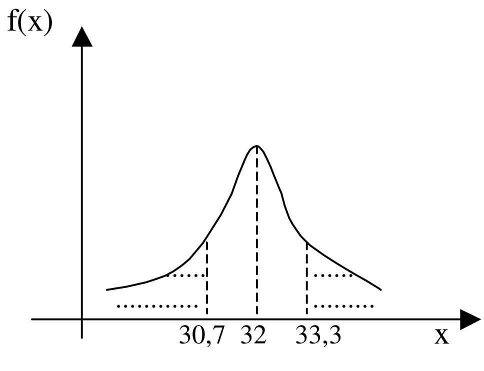
Como regra adotamos que para n ≥30, aproximamos a distribuição das médias amostrais para normal. Este resultado é conhecido como teorema do limite central.

Exemplos:

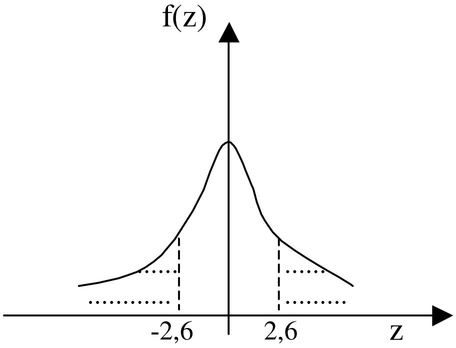
1) De uma grande população de média 32 e desvio padrão 3,7 retira-se uma amostra de 56 elementos. Determine:

a) A média da distribuição amostral.
b) O desvio padrão da distribuição amostral
c) A percentagem das possíveis médias amostrais que deferirão por mais de 1,3 da média da população

a) A média da população é igual à média da distribuição amostral, logo $\mu_{\bar{x}} = \mu_x$ =32.
b) $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{3,7}{\sqrt{56}} = 0,5$
c) Como n=56≥30 logo podemos supor a distribuição amostral normal. Como queremos a proporção das médias que diferirão por mais de 1,3 da média populacional, então estamos interessados nas proporções das médias amostrais que estarão abaixo de 32-1,3=30,7 e nas que estarão acima de 32+1,3=33,3



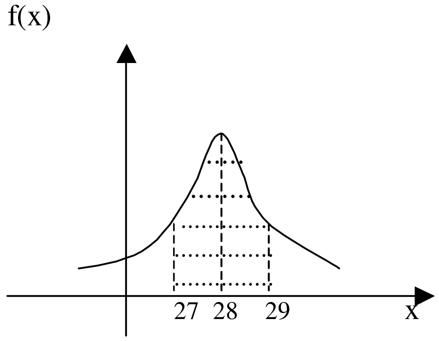
Queremos determinar P(x<30,7)+P(x>33,7), fazendo a mudança de variável de x para z temos que para x=30,7 temos $z = \frac{x-\mu}{\sigma} = \frac{30,7-32}{0,5} = -2,6$ e para x=33,3 temos $z = \frac{x-\mu}{\sigma} = \frac{33,3-32}{0,5} = 2,6$, logo queremos determinar P(z<-2,6)+P(z>2,6).



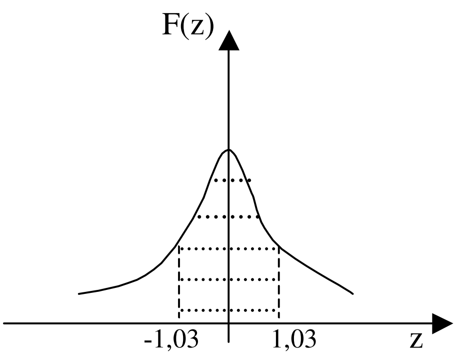
Logo utilizando a tabela normal temos que P(z<-2,6)+P(z>2,6)=0,0047+0,0047=0,0094= =0,94%.

2) Um fabricante de sapatos diz que seus sapatos têm média de vida de 28 meses e desvio padrão de 6 meses. Qual a percentagem de amostras de tamanho 38 terão média de vida em um intervalo de 1 mês em torno da média? Qual será sua percentagem se n=62?

Temos sempre que a média da população e igual a média da população amostral, isto é $\mu_x = \mu_{\bar{x}} = 28$, como n=38>30 temos que $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{6}{\sqrt{38}} = 0{,}9733$. A porcentagem ou probabilidade que queremos corresponde à área hachurada no gráfico abaixo.

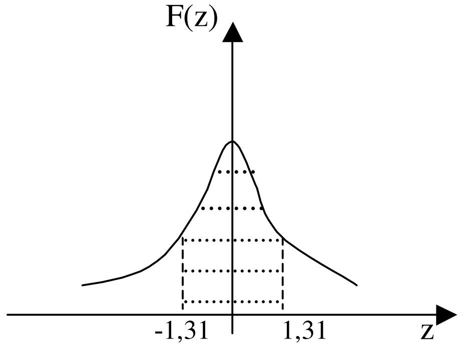


Fazendo a mudança de variável de x para z temos que para x=27$\Rightarrow$ $z = \frac{x-\mu}{\sigma} = \frac{27-28}{0{,}9733} = -1{,}03$, para x=29$\Rightarrow$ $z = \frac{x-\mu}{\sigma} = \frac{29-28}{0{,}9733} = 1{,}03$. Desta forma a porcentagem que queremos corresponde à área do gráfico abaixo.



Logo a porcentagem que queremos é P(27<x<29)= P(-1,03<z<1,03)=0,3485+0,3485=0,6970=69,70%.

Se n=62, então $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{6}{\sqrt{62}} = 0,7620$. Fazendo a mudança de variável de x para z temos que para x=27⇒ $z = \frac{x-\mu}{\sigma} = \frac{27-28}{0,7620} = -1,31$, para x=29⇒ $z = \frac{x-\mu}{\sigma} = \frac{29-28}{0,7620} = 1,31$. Desta forma a porcentagem que queremos corresponde à área do gráfico abaixo.



Logo a porcentagem que queremos é P(27<x<29)= P(-1,31<z<1,31)=0,4049+0,4049=0,8098=80,98%. Notamos que se o tamanho da amostra for maior, teremos uma maior concentração do gráfico da função f(x) em torno da média é por esta razão que temos a proporção de n=62 maior do que para n=38.

Exercício: Qual a probabilidade de se obter uma média amostral superior a 28,7 meses em uma amostra de 66 observações.
Resposta: 17,11%

Distribuição de proporções amostrais.

Esta distribuição indica a probabilidade de determinado conjunto de proporções amostrais, dados o tamanho da amostra e a proporção populacional.

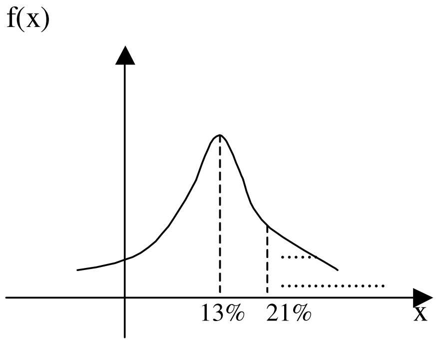
A média da distribuição de proporção amostral ($\bar{p}$) é igual a proporção populacional (p), isto é $\bar{p} = p$.

Quando temos o tamanho da amostra maior ou igual a trinta (n≥30), a distribuição de proporções amostrais se aproxima da distribuição normal, então quando n≥30, nós vamos aproximar a distribuição de proporções amostrais para a distribuição normal.

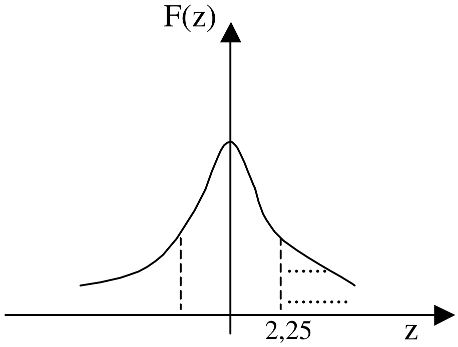
Quando a população é muito grande ou infinita nós calculamos o desvio padrão da distribuição de proporções amostrais pela expressão $\sigma_p = \sqrt{\frac{p.(1-p)}{n}}$.

Exemplo: Um comerciante compra jarros em grandes lotes. Periodicamente ele verifica os lotes para determinar a proporção de quebrados. Se um lote contém 13% de jarros quebrados, qual a probabilidade do comerciante em uma amostra de 89 jarros encontrar 21% ou mais de jarros quebrados?

Queremos a probabilidade P(x>21%), o que corresponde à área hachurada na distribuição de proporções amostrais abaixo.



Como n=89>30 podemos aproximar a distribuição de proporções amostrais para a distribuição normal. Temos que o desvio padrão da distribuição de proporções é $\sigma_p = \sqrt{\frac{p.(1-p)}{n}} = \sqrt{\frac{0{,}13.(1-0{,}13)}{89}} = 0{,}0356 = 3{,}56\%$ Logo para x=0,21=21%, temos que z=$\frac{0{,}21-0{,}13}{0{,}0356} = \frac{21\%-13\%}{3{,}65\%} = 2{,}25$. Logo P(x>21%)≅P(z>2,25)=0,5-0,4878=0,0122=1,22%.



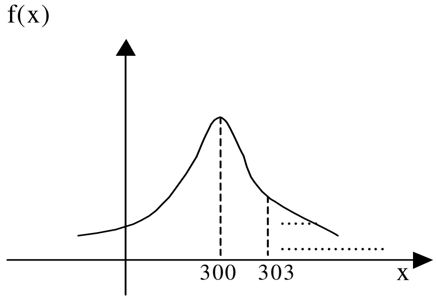
Amostragem de uma população finita.

A maior parte das amostras são feitas sem reposição por motivos de custos e conveniência. Se o tamanho da amostra for pequeno em relação à população a amostragem com reposição e sem reposição não apresentam grandes variações. Se a amostragem for maior que 5% da população as amostragens começam a diferir, neste caso as fórmulas do desvio padrão das médias amostrais e das proporções amostrais devem sofrer uma correção. Utilizamos então o fator de correção finita dado por $\sqrt{\frac{N-n}{N-1}}$, (N é o tamanho da população e n é o tamanho da amostra) que deve ser multiplicado pelo desvio padrão das médias e das proporções amostrais para fazermos as devidas correções. Desta forma ficamos com o desvio das médias amostrais corrigido sendo $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$ e o desvio padrão das proporções amostrais corrigido sendo $\sigma_p = \sqrt{\frac{p.(1-p)}{n}}.\sqrt{\frac{N-n}{N-1}}$.

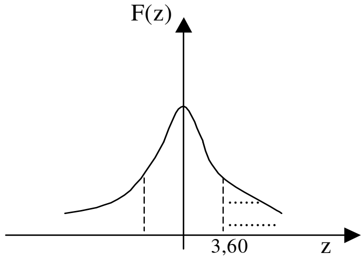
Exemplo: Uma máquina de encher recipientes de refrigerantes está regulada para encher o recipiente com 300ml e desvio padrão de 5 ml. Se o processo tem distribuição normal, qual a probabilidade de em uma amostra de 30 refrigerantes de um lote de 172 refrigerantes termos uma média superior a 303 ml?

Neste caso temos uma população finita que é o lote de 172 refrigerantes, logo N=172 e temos o tamanho da amostra sendo n=30≥30, logo podemos aproximar a distribuição das médias amostrais para a distribuição normal. Como o tamanho da amostra corresponde a mais que 5% da população. De fato n/N=30/172=0,1744=17,44%. Temos então que corrigir o desvio padrão da distribuição. Logo $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} = \frac{5}{\sqrt{30}}.\sqrt{\frac{172-30}{172-1}} = 0{,}8319.$

Na distribuição das médias amostrais queremos a probabilidade P(x>303) que corresponde à área abaixo.



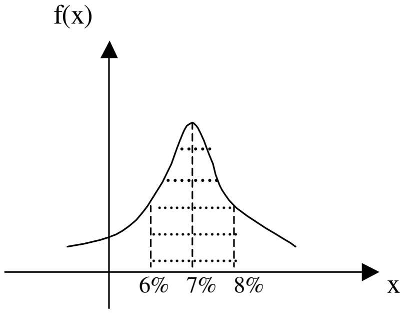
Transformando a variável x para z temos: se x=303, então z=(303-300) / 0,8319=3,60. Logo P(x>303)≅P(z>3,60)=0,5-0,4998=0,0002=0,02%



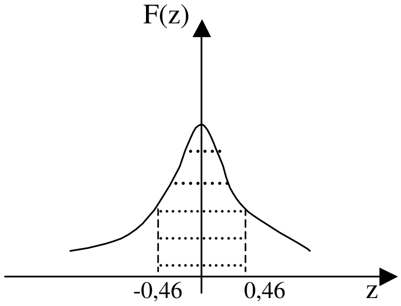
Exemplo: Uma fábrica de CD's tem uma média de 7% de CD's fabricados com defeito. Uma amostra de 123 CD's é retirada de um lote de 1021 CD's. Qual a probabilidade da proporção amostral de CD's com defeito esteja entre 6% e 8%?

Neste caso temos uma população finita que é o lote de 1021 CD's, logo N=1021 e temos o tamanho da amostra sendo n=123>30, logo podemos aproximar a distribuição das proporções amostrais para a distribuição normal. Como o tamanho da amostra corresponde a mais que 5% da população. De fato n/N=123/1021=0,1205=12,05%. Temos então que corrigir o desvio padrão da distribuição. Logo $\sigma_p = \sqrt{\frac{p.(1-p)}{n}}.\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0{,}07.(1-0{,}07)}{123}}.\sqrt{\frac{1021-123}{1021-1}} = 0{,}0216$

Na distribuição das porporções amostrais queremos a probabilidade P(6%<x<8%) que corresponde à área abaixo.



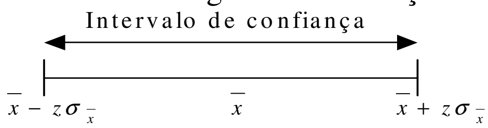
Transformando a variável x para z temos: se x=6%=0,06, então z=(0,06-0,07)/0,0216=-0,46 e se x=8%=0,08, então z=(0,08-0,07)/0,0216=0,46. Logo P(6%<x<8%)≅P(-0,46<z<0,46)=0,1772+0,1772=0,3544=35,44%



Estimação da média de uma população

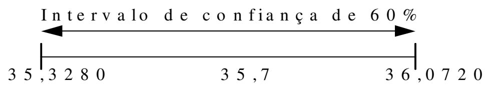
Grandes amostras produzem médias amostrais mais próximas da média da distribuição amostral (média populacional), pois temos que o desvio padrão da distribuição amostral diminui quando n cresce ($\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$). Para estimarmos da média de uma população nós precisamos do desvio padrão, que pode ser o desvio padrão da população se este for conhecido, senão utilizaremos o desvio padrão amostral que calcularemos utilizando os dados amostrais.

Quando o desvio padrão populacional é conhecido e a população tem distribuição normal ou n≥30 nós estimamos a média populacional sendo: $\mu_x = \bar{x} \pm z\sigma_{\bar{x}}$ (estimativa intervalar) ou $\mu_x = \bar{x}$ (estimativa pontual), onde $\bar{x}$ é a média da amostra (média amostral), $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ e z é a variável aleatória normal que determinará o grau de confiança do intervalo.

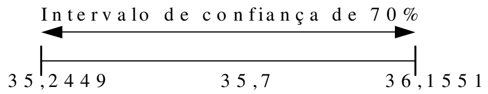


Exemplo: Foi extraída uma amostra de 47 funcionários de uma empresa e foi calculada a média amostral de 35,7 anos de idade, supondo o desvio padrão desta população igual a 3 anos. Determine o intervalo de confiança para a média de idade dos funcionários com 60%, 70% e 80% de confiança.

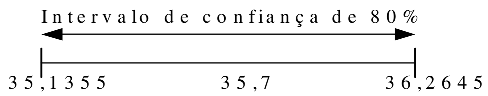
Como n=47>30 temos que a distribuição da população se aproxima da distribuição normal. Para termos um intervalo de confiança de 60%, nós tomamos um valor de z que nos dá uma porcentagem de 60% em torno da média, isto é, procuramos na tabela da distribuição normal o valor de z que nos dá o valor mais próximo e superior a 30% de área, logo o valor –z nos dá os outros 30%. Este valor corresponde a z=0,85, que nos dá uma área de 30,23%. Calculando o desvio padrão temos que $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{3}{\sqrt{47}} = 0{,}4376$. Logo estimamos a média dos funcionários em um intervalo de confiança de 60% da seguinte maneira: $\bar{x} \pm z\sigma_{\bar{x}} \Leftrightarrow 35{,}7 \pm 0{,}85.0{,}4376 \Leftrightarrow 35{,}7 \pm 0{,}3720$. Isto é, nós temos 60% de probabilidade de que a média da população esteja neste intervalo.



Para 70% temos z=1,04 que corresponde a um área de 0,3508, logo o intervalo de confiança será $\bar{x} \pm z\sigma_{\bar{x}} \Leftrightarrow 35{,}7 \pm 1{,}04.0{,}4376 \Leftrightarrow 35{,}7 \pm 0{,}4551$

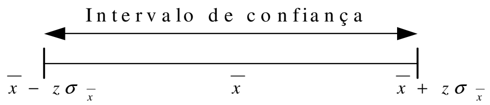


Para 80% temos z=1,29 que corresponde a um área de 0,4015, logo o intervalo de confiança será $\bar{x} \pm z\sigma_{\bar{x}} \Leftrightarrow 35{,}7 \pm 1{,}29.0{,}4376 \Leftrightarrow 35{,}7 \pm 0{,}5645$



Erro de estimação

O erro em um intervalo de estimação da média de uma população é a diferença entre a média amostral e a média populacional. Como o intervalo tem centro na média amostral e a média da população supostamente está no intervalo com determinada confiança, então temos que o erro máximo é dado por $e = z\sigma_{\bar{x}}$, onde *e* é o erro de estimação, logo $e = z\dfrac{\sigma_x}{\sqrt{n}}$.



Determinação do tamanho da amostra

Para determinarmos o tamanho de uma amostra em função do erro desejado, do desvio padrão da população e da confiança também desejada na estimação, nós utilizamos a expressão de n abaixo decorrente da expressão do erro.

$$e = z\frac{\sigma_x}{\sqrt{n}} \Rightarrow n = \frac{z^2\sigma_x{}^2}{e^2} \Rightarrow n = (\frac{z\sigma_x}{e})^2$$

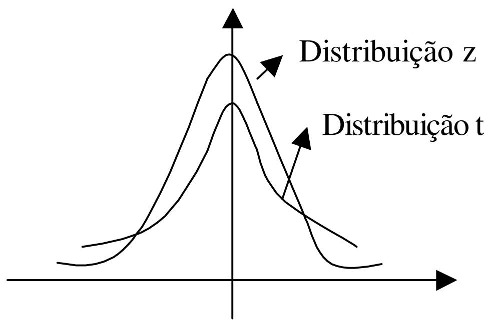
Exemplo: Qual o tamanho da amostra que devemos tomar, para termos um intervalo de 80% de confiança, com um erro de 0,83 em torno da média, onde o desvio padrão da população é de 3,7?

Para termos um intervalo de 80% de confiança nós tomamos z=1,29, que corresponde a uma área de 0,4015, logo o tamanho da amostra será $n = (\frac{z\sigma_x}{e})^2 = (\frac{1{,}29.3{,}7}{0{,}83})^2 = 33{,}07$, vamos tomar então n=34, por uma medida de segurança, isto é, para garantirmos no mínimo nossas exigências de erro e confiança.

\# Estimação da média de uma população quando o desvio padrão da população (σ_x) é desconhecido

Se σ_x é desconhecido utilizamos o desvio padrão amostral (s_x) para estimar o desvio padrão da população σ_x. Isto é, substituímos nas equações $\sigma_{\bar{x}}$ por $s_{\bar{x}}$, onde $s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$. Temos na prática que se o desvio padrão amostral ($s_{\bar{x}}$) é uma boa aproximação do desvio padrão da população da população (σ_x).

Quando o tamanho da amostra for menor que 30, devemos usar a distribuição de student (t), pois ela se aproximará melhor da distribuição das médias amostrais e por ser a mais indicada quando não conhecemos o desvio padrão da população (σ_x). Para usarmos a distribuição t, temos que ter uma população com distribuição normal ou próxima da normal. Porém se o tamanho da amostra for maior ou igual a 30, esta informação é desnecessária e poderemos usar a distribuição normal

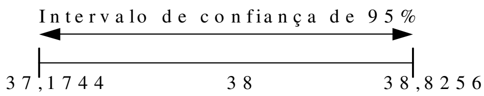


Temos que o desvio padrão amostral é calculado pela expressão $s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$, onde n-1 é o número de graus de liberdade que utilizaremos para determinarmos os valores de t na tabela da distribuição de student logo abaixo. Desta forma quando não tivermos o desvio padrão da população nós estimaremos a média populacional pelo intervalo $\bar{x} \pm t\frac{s_x}{\sqrt{n}}$. Porém se o tamanho da amostra for maior ou igual a 30, nós podemos substituir o t pelo z, isto é, a distribuição de student.

Graus de liberdade (n -1)	α bicaudal														
	0,1000	0,0900	0,0800	0,0700	0,0600	0,0500	0,0400	0,0300	0,0200	0,0100	0,0050	0,0025	0,0010	0,0005	0,0001
	α unicaudal														
	0,0500	0,0450	0,0400	0,0350	0,0300	0,0250	0,0200	0,0150	0,0100	0,0050	0,0025	0,0013	0,0005	0,0003	0,0001
1	6,3137	7,0264	7,9158	9,0579	10,5789	12,7062	15,8945	21,2051	31,8210	63,6559	127,3211	254,6608	636,5776	1.273,155	6.370,544
2	2,9200	3,1040	3,3198	3,5782	3,8964	4,3027	4,8487	5,6428	6,9645	9,9250	14,0892	19,9629	31,5998	44,7035	100,1358
3	2,3534	2,4708	2,6054	2,7626	2,9505	3,1824	3,4819	3,8961	4,5407	5,8408	7,4532	9,4646	12,9244	16,3261	28,0142
4	2,1318	2,2261	2,3329	2,4559	2,6008	2,7765	2,9985	3,2976	3,7469	4,6041	5,5975	6,7582	8,6101	10,3051	15,5345
5	2,0150	2,0978	2,1910	2,2974	2,4216	2,5706	2,7565	3,0029	3,3649	4,0321	4,7733	5,6042	6,8685	7,9756	11,1759
6	1,9432	2,0192	2,1043	2,2011	2,3133	2,4469	2,6122	2,8289	3,1427	3,7074	4,3168	4,9807	5,9587	6,7882	9,0804
7	1,8946	1,9662	2,0460	2,1365	2,2409	2,3646	2,5168	2,7146	2,9979	3,4995	4,0294	4,5946	5,4081	6,0815	7,8883
8	1,8595	1,9280	2,0042	2,0902	2,1892	2,3060	2,4490	2,6338	2,8965	3,3554	3,8325	4,3336	5,0414	5,6170	7,1200
9	1,8331	1,8992	1,9727	2,0554	2,1504	2,2622	2,3984	2,5738	2,8214	3,2498	3,6896	4,1458	4,7809	5,2911	6,5938
10	1,8125	1,8768	1,9481	2,0283	2,1202	2,2281	2,3593	2,5275	2,7638	3,1693	3,5814	4,0045	4,5868	5,0489	6,2119
11	1,7959	1,8588	1,9284	2,0067	2,0961	2,2010	2,3281	2,4907	2,7181	3,1058	3,4966	3,8945	4,4369	4,8633	5,9232
12	1,7823	1,8440	1,9123	1,9889	2,0764	2,1788	2,3027	2,4607	2,6810	3,0545	3,4284	3,8065	4,3178	4,7166	5,6950
13	1,7709	1,8317	1,8989	1,9742	2,0600	2,1604	2,2816	2,4358	2,6503	3,0123	3,3725	3,7345	4,2209	4,5972	5,5134
14	1,7613	1,8213	1,8875	1,9617	2,0462	2,1448	2,2638	2,4149	2,6245	2,9768	3,3257	3,6746	4,1403	4,4995	5,3644
15	1,7531	1,8123	1,8777	1,9509	2,0343	2,1315	2,2485	2,3970	2,6025	2,9467	3,2860	3,6239	4,0728	4,4168	5,2387
16	1,7459	1,8046	1,8693	1,9417	2,0240	2,1199	2,2354	2,3815	2,5835	2,9208	3,2520	3,5805	4,0149	4,3464	5,1339
17	1,7396	1,7978	1,8619	1,9335	2,0150	2,1098	2,2238	2,3681	2,5669	2,8982	3,2224	3,5430	3,9651	4,2858	5,0431
18	1,7341	1,7918	1,8553	1,9264	2,0071	2,1009	2,2137	2,3562	2,5524	2,8784	3,1966	3,5101	3,9217	4,2332	4,9663
19	1,7291	1,7864	1,8495	1,9200	2,0000	2,0930	2,2047	2,3457	2,5395	2,8609	3,1737	3,4811	3,8833	4,1869	4,8988
20	1,7247	1,7816	1,8443	1,9143	1,9937	2,0860	2,1967	2,3362	2,5280	2,8453	3,1534	3,4554	3,8496	4,1461	4,8382
21	1,7207	1,7773	1,8397	1,9092	1,9880	2,0796	2,1894	2,3278	2,5176	2,8314	3,1352	3,4324	3,8193	4,1095	4,7847
22	1,7171	1,7734	1,8354	1,9045	1,9829	2,0739	2,1829	2,3202	2,5083	2,8188	3,1188	3,4118	3,7922	4,0769	4,7358
23	1,7139	1,7699	1,8316	1,9003	1,9783	2,0687	2,1770	2,3132	2,4999	2,8073	3,1040	3,3931	3,7676	4,0475	4,6939
24	1,7109	1,7667	1,8281	1,8965	1,9740	2,0639	2,1715	2,3069	2,4922	2,7970	3,0905	3,3761	3,7454	4,0207	4,6543
25	1,7081	1,7637	1,8248	1,8929	1,9701	2,0595	2,1666	2,3011	2,4851	2,7874	3,0782	3,3606	3,7251	3,9965	4,6194
26	1,7056	1,7610	1,8219	1,8897	1,9665	2,0555	2,1620	2,2958	2,4786	2,7787	3,0669	3,3464	3,7067	3,9744	4,5868
27	1,7033	1,7585	1,8191	1,8867	1,9632	2,0518	2,1578	2,2909	2,4727	2,7707	3,0565	3,3334	3,6895	3,9540	4,5565
28	1,7011	1,7561	1,8166	1,8839	1,9601	2,0484	2,1539	2,2864	2,4671	2,7633	3,0470	3,3213	3,6739	3,9348	4,5309
29	1,6991	1,7540	1,8142	1,8813	1,9573	2,0452	2,1503	2,2822	2,4620	2,7564	3,0380	3,3103	3,6595	3,9177	4,5053
30	1,6973	1,7520	1,8120	1,8789	1,9546	2,0423	2,1470	2,2783	2,4573	2,7500	3,0298	3,2999	3,6460	3,9017	4,4820
40	1,6839	1,7375	1,7963	1,8617	1,9357	2,0211	2,1229	2,2503	2,4233	2,7045	2,9712	3,2266	3,5510	3,7884	4,3213
50	1,6759	1,7289	1,7870	1,8516	1,9244	2,0086	2,1087	2,2338	2,4033	2,6778	2,9370	3,1840	3,4960	3,7230	4,2282
60	1,6706	1,7232	1,7808	1,8448	1,9170	2,0003	2,0994	2,2229	2,3901	2,6603	2,9146	3,1562	3,4602	3,6808	4,1688
70	1,6669	1,7192	1,7765	1,8401	1,9118	1,9944	2,0927	2,2152	2,3808	2,6479	2,8987	3,1366	3,4350	3,6508	4,1269
80	1,6641	1,7162	1,7732	1,8365	1,9078	1,9901	2,0878	2,2095	2,3739	2,6387	2,8870	3,1220	3,4164	3,6287	4,0955
90	1,6620	1,7138	1,7707	1,8337	1,9048	1,9867	2,0839	2,2050	2,3685	2,6316	2,8779	3,1108	3,4019	3,6118	4,0722
100	1,6602	1,7120	1,7687	1,8315	1,9024	1,9840	2,0809	2,2015	2,3642	2,6259	2,8707	3,1018	3,3905	3,5984	4,0536
200	1,6525	1,7036	1,7596	1,8217	1,8915	1,9719	2,0672	2,1857	2,3451	2,6006	2,8385	3,0621	3,3398	3,5387	3,9698
250	1,6510	1,7020	1,7578	1,8197	1,8894	1,9695	2,0645	2,1826	2,3414	2,5956	2,8322	3,0543	3,3299	3,5268	3,9546
500	1,6479	1,6987	1,7543	1,8158	1,8851	1,9647	2,0591	2,1763	2,3338	2,5857	2,8195	3,0387	3,3101	3,5038	3,9220
750	1,6469	1,6976	1,7531	1,8145	1,8836	1,9631	2,0573	2,1742	2,3313	2,5824	2,8154	3,0336	3,3036	3,4960	3,9116
1000	1,6464	1,6970	1,7525	1,8139	1,8829	1,9623	2,0564	2,1732	2,3301	2,5807	2,8133	3,0310	3,3002	3,4922	3,9069
10000	1,6450	1,6956	1,7509	1,8121	1,8810	1,9602	2,0540	2,1704	2,3267	2,5763	2,8076	3,0241	3,2915	3,4820	3,8918

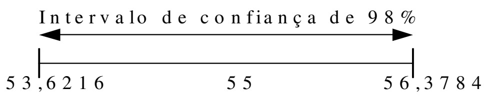
Exemplo: Uma amostra de uma população de distribuição normal de tamanho 25 tem média 38 e desvio padrão 2. Determine um intervalo de confiança para a média desta população com uma confiança de 95%.

Como n=25<30 nós vamos usar a distribuição de student. O grau de liberdade é n-1=25-1=24, logo devemos olhar na tabela da distribuição de student na linha do grau de liberdade 24 à área de uma cauda (unicaudal) da distribuição de student que corresponde a 0,025 ou a área das duas caudas (bicaudal) que corresponde a 0,05, pois a tabela da distribuição de student ou contrario da tabela da distribuição normal nos dá a área das caudas da distribuição. Então o lhando na tabela de student, temos que o valor de t que corresponde à área de uma cauda igual a 0,025 é t=2,0639. Logo determinamos o intervalo de confiança sendo $\bar{x} \pm t\frac{s_x}{\sqrt{n}} \Leftrightarrow 38 \pm 2{,}06339.\frac{2}{\sqrt{25}} \Leftrightarrow 38 \pm 0{,}8256$



Exemplo: Suponha que uma amostra de tamanho 35, média 55 e desvio padrão 3,5 foi retirada de uma população para análise. Determine um intervalo de 98% de confiança para a média desta população.

Como n=35>30 nós usaremos a distribuição normal. Olhando na tabela da distribuição normal temos que o valor de z que corresponde a uma área de 0,4901 é z=2,33, logo o nosso intervalo de confiança é o seguinte: $\bar{x} \pm z\frac{s_x}{\sqrt{n}} \Leftrightarrow 55 \pm 2{,}33.\frac{3{,}5}{\sqrt{35}} \Leftrightarrow 55 \pm 1{,}3784$.



Estimação da média populacional para populações finitas

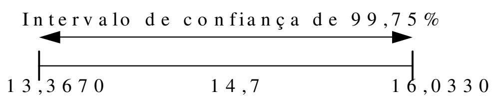
Se tivermos que o tamanho da amostra (n) corresponde a mais que 5% da população (N), então temos que corrigir o desvio padrão, onde o fator de correção é $\sqrt{\frac{N-n}{N-1}}$.

Se o desvio padrão da população σ_x for conhecido nós estimaremos a média da população com o desvio padrão corrigido da seguinte maneira: $\bar{x} \pm z\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, onde o erro de estimação agora será dado por $e = z\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

Se o desvio padrão da população σ_x for desconhecido nós usamos o desvio padrão amostral s_x e estimaremos a média da população com desvio padrão corrigido da seguinte maneira: $\bar{x} \pm t\frac{s_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, onde o erro de estimação agora será dado por $e = t\frac{s_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

Exemplo: Uma amostra de 17 elementos é retirada de uma população com distribuição próxima da normal composta de 202 elementos. A média amostral é 14,7 e o desvio padrão de 1,6. Estime a média populacional em um intervalo com 99,75% de confiança.

Como n=17<30, devemos usar a distribuição de student. Temos que a amostra de 17 elementos corresponde a mais que 5% da população de 202 elementos. De fato n/N=17/202=0,0842=8,42%. Logo devemos utilizar o fator de correção finita $\sqrt{\frac{N-n}{N-1}}$. Como o grau de liberdade é n-1=17-1=16, olhando na tabela de student à área de duas caudas de 0,0025 obtemos t=3,5805. Logo estimamos a média da população com 99,75% de confiança no intervalo seguinte: $\bar{x} \pm t\frac{s_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 14,7 \pm 3,5805.\frac{1,6}{\sqrt{17}}.\sqrt{\frac{202-17}{202-1}} = 14,7 \pm 1,3330$.



Se o tamanho da amostra n fosse maior ou igual a 30, nós utilizaríamos a distribuição normal.

Determinação do tamanho da amostra para desvios padrões corrigidos

Se o desvio padrão populacional σ_x for conhecido então o erro é dado por $e = z\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, logo $n = \frac{z^2.\sigma_x{}^2.N}{e^2.(N-1)+z^2.\sigma_x{}^2}$.

Se o desvio padrão populacional σ_x for desconhecido então o erro é dado por $e = t\frac{s_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, logo $n = \frac{t^2.s_x{}^2.N}{e^2.(N-1)+t^2.s_x{}^2}$.

Intervalos de confiança unilaterais para médias de uma população

Utilizamos os intervalos de confiança unilaterais para estimarmos a média de uma população, quando estamos apenas interessados nos limites inferiores ou superiores dos intervalos de confiança.

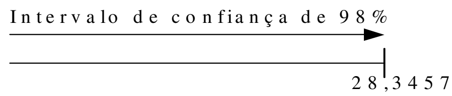
Exemplo: Uma empresa de distribuição de energia elétrica de uma cidade quer saber o mínimo de energia consumida por residência, para que a empresa tenha lucro.

Exemplo: Uma empresa que produz cabos de energia elétrica deseja saber o máximo de amperagem que os seus cabos suportam.

Considere o seguinte exemplo: Uma amostra de 113 observações é retirada de uma população. A média amostral é 27,9 e o desvio padrão de 2,3.

a) Determine um limite superior para a média com 98% de confiança.
b) Qual a probabilidade da média ser maior que 28,1?

a) Como n=113>30, usaremos a distribuição normal. Logo olhando na tabela da distribuição normal o valor limite superior de z que nos dá uma área próxima de 98% é z=2,06, isto é a área à esquerda de z=2,06 é de 0,9803=98,03%. Desta forma estimamos o limite máximo da média populacional sendo: $\bar{x} + z\frac{s_x}{\sqrt{n}} \Leftrightarrow 27{,}9 + 2{,}06.\frac{2{,}3}{\sqrt{113}} \Leftrightarrow 27{,}9 + 0{,}4457 \Leftrightarrow 28{,}3457$, isto é, a probabilidade de que a média populacional seja menor que 28,3457 é de 98,03%.

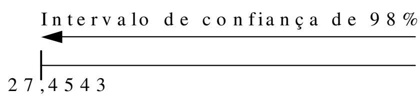


b) Queremos a probabilidade da média populacional ser maior que 28,1, isto é P(μ_x>28,1), logo transformando μ_x para variável z usando a média amostral de 27,9 e o desvio padrão da distribuição sendo $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{2{,}3}{\sqrt{113}} = 0{,}2164$ temos: z=(28,1-27,9)/0,2164=0,92. Logo P(μ_x>28,1) ≅P(z>0,92)=0,5-0,3212=0,1788=17,88%.

Exemplo: Considere o exemplo anterior:

a) Determine um limite inferior para a médias com 98% de confiança.
b) Qual a probabilidade da média populacional ser menor que 27,3?

a) Como n=113>30, usaremos a distribuição normal. Logo olhando na tabela da distribuição normal o valor limite inferior de z que nos dá uma área próxima de 98% é z=-2,06, isto é a área à direita de z=-2,06 é de 0,9803=98,03%. Desta forma estimamos o limite mínimo da média populacional sendo: $\bar{x} - z\frac{s_x}{\sqrt{n}} \Leftrightarrow 27{,}9 - 2{,}06.\frac{2{,}3}{\sqrt{113}} \Leftrightarrow 27{,}9 - 0{,}4457 \Leftrightarrow 27{,}4543$, isto é, a probabilidade de que a média populacional seja maior que 27,4543 é de 98,03%.



b) Queremos a probabilidade da média populacional ser menor que 27,3, isto é P(μ_x<27,3), logo transformando μ_x para variável z usando a média amostral de 27,9 e o desvio padrão da distribuição sendo $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{2,3}{\sqrt{113}} = 0,2164$ temos: z=(27,3-27,9)/0,2164=-2,77. Logo P(μ_x<27,3) ≅P(z<-2,77)=0,5-0,4972=0,0028=0,28%.

Estimação da proporção de uma população

Usamos a proporção amostral para estimar a proporção populacional. A estimativa pontual é quando estimamos a proporção populacional pela amostral, isto é, p=x/n, onde x é o número de itens na amostra e n é o tamanho da amostra.

Exemplo: Em um lote de CD's é retirada uma amostra de 123 CD's e observamos que 47 estão com defeito. Estime a proporção de CD's com defeito deste lote pontualmente.

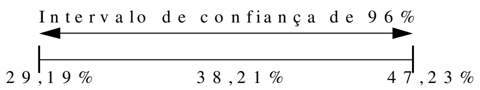
p=x/n=47/123=0,3821=38,21%

A estimativa intervalar de proproções populacionais é feita da seguinte maneira: $\frac{x}{n} \pm z\sqrt{\frac{\frac{x}{n}.(1-\frac{x}{n})}{n}}$, onde x é o número de itens da amostra, z é a variável aleatória normal e n o tamanho da amostra. O desvio padrão da proporção é calculado da seguinte maneira: $\sigma_p = \sqrt{\frac{\frac{x}{n}.(1-\frac{x}{n})}{n}} = \sqrt{\frac{p.(1-p)}{n}}$, pois p=x/n. Na estimativa intervalar de proporções populacionais não utilizamos a distribuição de student.

Exemplo: Considere o exemplo anterior. Estime a proporção populacional com uma confiança de 96%.

Temos que p=x/n=47/123=0,3821=38,21% e $\sigma_p = \sqrt{\frac{p.(1-p)}{n}} = \sqrt{\frac{0,3821(1-0,3821)}{123}} = 0,0438$. O valor de z que nos dá uma confiança mais próxima de 96% é z=2,06, que nos dá uma área de 0,4803%. Logo estimamos a proporção populacional da seguinte maneira:

$$\frac{x}{n} \pm z\sqrt{\frac{\frac{x}{n}.(1-\frac{x}{n})}{n}} \Leftrightarrow p \pm z.\sqrt{\frac{p.(1-p)}{n}} \Leftrightarrow 0,3821 \pm 2,06.\sqrt{\frac{0,3821.(1-0,3821)}{123}} \Leftrightarrow 0,3821 \pm 0,0902$$
$$\Leftrightarrow 38,21\% \pm 9,02\%$$



Erro de estimação da proporção populacional

O erro de estimação da proporção populacional é a diferença entre a proporção amostral e a proporção populacional, como supomos que a proporção amostral no intervalo de estimação da proporção populacional, logo o erro máximo é dado por $e = z.\sqrt{\frac{p.(1-p)}{n}}$, notamos que o erro aumenta quando $p.(1-p)$ aumenta, isto é quando $-p^2+p$ aumenta, considerando a função f(p)=$-p^2+p$, temos que seu ponto de mínimo é *p*=1/2, logo o erro é máximo quando p=1/2=0,5=50%. Desta forma quando tivermos situações onde a proporção *p* não for conhecida, nós vamos supor que p=1/2, pois para este valor de *p* teremos um erro máximo, desta forma estaremos aumentando o desvio padrão e aumentando a confiança de nossa estimação.

Exemplo: Qual o tamanho da amostra que devemos tomar para estimarmos a porporção de uma população com um nível de significância de 6% e um erro de 0,03?

Nível de significância é o complementar do nível de confiança e é denotado pela letra grega alfa (α), isto é, α=1-NC, logo NC=1-α=1-0,06=0,94=94%, onde NC é o nível de confiança.

Como a proporção *p* não é conhecida tomamos *p*=0,5, para um nível de confiança de 94% tomamos z=1,89 que nos dá uma área de 0,4706. Como o erro é dado por $e = z.\sqrt{\frac{p.(1-p)}{n}}$, temos que $n = \frac{z^2.p.(1-p)}{e^2} = \frac{1{,}89^2.0{,}5.(1-0{,}5)}{0{,}03^2} = 992{,}25$, vamos tomar então n=993 para garantirmos no mínimo nossas exigências de confiança e erro.

Exercício: De uma amostra de 153 observações de um lote de parafusos, foi detectado 29 parafusos com defeito. Determine o erro de estimação da proporção populacional se α=4%.
Resp: 6,53%.

Exercício: De um lote de 1791 canetas, foi retirada uma amostra de 93 canetas, desta amostra foram encontradas 13 canetas com defeito. Estime a proporção de canetas defeituosas deste lote com um nível de confiança de 97,5%. *Resp:13,98%* $\pm$ *7,84%.*

\# Determinação do tamanho da amostra em populações finitas

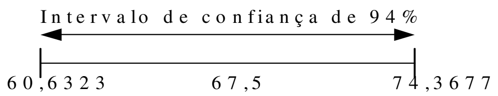
O erro de estimação da proporção populacional de populações finitas é dado por $e = z.\sqrt{\frac{p.(1-p)}{n}}.\sqrt{\frac{N-n}{N-1}}$, logo temos que $n = \frac{z^2.p.(1-p).N}{e^2.(N-1)+z^2.p.(1-p)}$, onde $p = \frac{x}{n}$.

Exercício: Qual o tamanho da amostra que devemos retirar de um lote de 2795 relógios, para estimarmos a proporção de defeituosos com um nível de confiança de 95,8% e com um erro de 0,023? *Resp: n=1155.*

Exemplo: Foi retirada uma amostra dos pesos dos estudantes de uma faculdade conforme os seguintes dados: 68 ; 70 ; 56 ; 78 ; 49 ; 62 ; 67 ; 91 ; 58 ; 62 ; 71 ; 78. Estime a média populacional com um nível de significância de 6%.

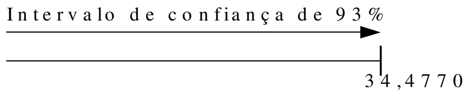
Temos que a média dos dados é $\bar{x} = 67{,}5$ e o desvio padrão amostral s_x é dado pela expressão $s_x = \sqrt{\frac{\sum(\bar{x}-x_i)^2}{n-1}} = \sqrt{\frac{1417}{11}} = 11{,}3498$. Podemos calcular o desvio padrão amostral na HP-12C, limpando os registradores estatísticos teclando (f) (reg) e entrando com os dados amostrais teclando (Σ+), depois teclamos (g) (s) para obtermos o desvio padrão amostral.

Com n=12<30 vamos utilizar a distribuição de student, o valor de t que nos dá um nível de significância de 6%, para um grau de liberdade de 11 é t=2,0961, logo estimamos a média populacional sendo: $\bar{x} \pm t\frac{s_x}{\sqrt{n}} \Rightarrow 67{,}5 \pm 2{,}0961.\frac{11{,}3498}{\sqrt{12}} \Rightarrow 67{,}5 \pm 6{,}8677$.



Exercício: Uma amostra de 41 observações de uma população apresentou uma média de 33,7. Se a variância populacional é de 11,3, determine um limite máximo para a média populacional com α=7%.

Resp: $\mu_{\bar{x}} = 34{,}4770$



Exercício: Em um estudo sobre o grau de impurezas em lotes de 3,2Kg de determinado composto medicinal, o erro máximo tolerável é de 1,03Kg. Se o grau de impurezas apresenta um desvio padrão populacional de 5,03g. Determine o tamanho da amostra que devemos tomar neste estudo, se desejamos um nível de significância de 4%. *Resp: n=102.*

Testes de hipóteses (ou significância)

A finalidade do teste de hipóteses é avaliar afirmações sobre valores de parâmetros populacionais.

$1^{\underline{0}}$) Passo: Formular a hipótese nula H_0 e a alternativa H_1.
$2^{\underline{0}}$) Passo: Escolher a distribuição adequada (t ou z)
$3^{\underline{0}}$) Passo: Escolher o nível de significância (ou seja, os valores críticos).
$4^{\underline{0}}$) Passo: Calcular a estatítica teste e compara-la com os valores críticos.
$5^{\underline{0}}$) Passo: Rejeitar a hipótese nula se a estatística teste execeder os valores críticos, caso contrário aceitar a hipótese nula H_0.

A Hipótese nula H_0 é uma afirmação que diz que o parâmetro populacional é como o especificado, isto é, a afirmação é verdadeira.

A hipótese alternativa H_1 é uma afirmação que oferece uma alternativa ao especificado, isto é, o parâmetro é maior ou menor que o valor especificado.

a. Testes de hipóteses bilaterais e unilaterais

Exemplo: Um fabricante de parafusos para automóveis afirma que o diâmetro de seus parafusos é de 6mm. Uma amostra de 43 parafusos apresentou uma média de 5,8mm e desvio padrão de 0,2mm. Assumindo α=4% o que podemos dizer sobre a afirmação do fabricante.

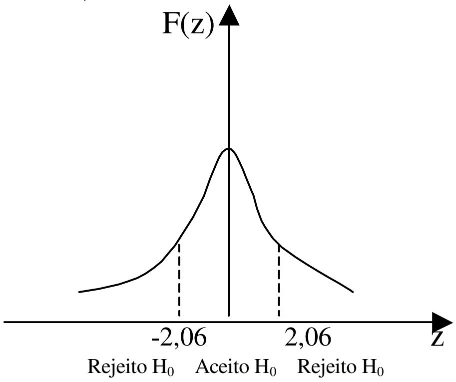
1º) Passo: Vamos formular a hipótese nula H_0 e a hipótese alternativa H_1. Seja H_0 a afirmação do fabricante, isto é μ=6mm e a alternativa H_1 que $\mu \neq$6mm.

H_0:μ= 6mm.

H_1: $\mu \neq$6mm.

2º) Passo: Como n=43>30 usaremos a distribuição normal. Como o parafuso tem um limite máximo e um limite mínimo de diâmetro para ser útil, temos então um teste bilateral.

3º) Passo: Como α=4%, temos então que o nível de confiança NC=96%, logo os valores de z que nos dá uma área mais próxima de 0,96 em torno o eixo vertical F(z) são z=-2,06 e z=2,06. Estes valores de z são ditos valores críticos.

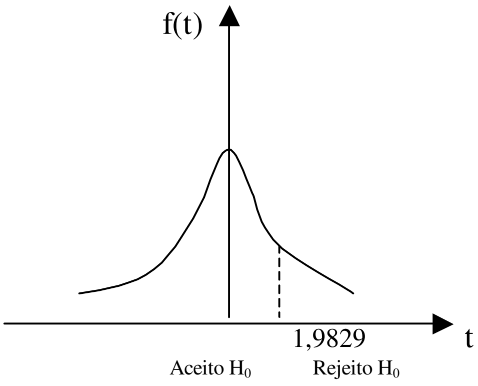


4º) Passo: Transformando a média amostral obtida $\bar{x} = 5{,}8$, para a variável z_{teste}, usando a média μ da população informada pelo fabricante temos: $z_{teste} = \dfrac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}} = \dfrac{5{,}8 - 6}{\frac{0{,}2}{\sqrt{43}}} = -6{,}56$.

5º) Passo: Como z_{teste}=-6,56<-2,06, devemos rejeitar a hipótese H_0, isto é, a média da população não é de 6mm, logo aceitamos a hipótese alternativa, ou seja, a média da população é diferente de 6mm.

Exemplo: Um gerente de um determinado banco disse que em média seus clientes esperam 30 minutos para serem atendidos. Um teste foi feito com 23 clientes e observamos uma média de 34 minutos, com um desvio padrão de 5 minutos. Considerando α=3% o que podemos dizer da afirmação do gerente?

Vamos tomar a hipótese nula sendo H_0: μ=30 minutos, e a alternativa sendo H_1: μ >30 minutos. Como n=23<30 vamos usar a distribuição de student. O grau de liberdade é n-1=23-1=22, logo o valor de t que nos dá uma área à direita de 0,03 é t=1,9829, isto é, para valores de t_{teste} maiores que 1,9829 rejeitamos H_0, para valores de t_{teste} menores que 1,9829 aceitamos H_0.



Calculando t_{teste} temos $t_{teste} = \dfrac{\overline{x}-\mu}{\dfrac{s_x}{\sqrt{n}}} = \dfrac{34-30}{\dfrac{5}{\sqrt{23}}} = 3{,}8367$. Como o $t_{teste}=3{,}8367>1{,}9829$ nós rejeitamos H_0, isto é não aceitamos a afirmação do gerente.

Tipos de erros associados aos testes de hipóteses

Temos o erro chamado de erro tipo I, que seria considerar H_0 falsa, isto é, rejeitar H_0, sendo H_0 verdadeira. A probabilidade de cometermos este erro é igual ao nível de significância α.

Temos o erro chamado de erro tipo II, que seria aceitar H_0 sendo H_0 falsa, indicamos este tipo de erro por β.

Exercícios:

1) Um fabricante de determinada lâmpada afirma que a vida útil de suas lâmpadas é de no mínimo 8000 horas. Uma amostra de 43 observações apresentou média de 7800 horas de vida útil. Sabendo-se que o desvio padrão populacional é de 400 horas, o que podemos dizer da afirmação do fabricante supondo $\alpha=5\%$?
Resp: H_0: $\mu=8000$ horas e H_1: $\mu<8000$ horas. Rejeitamos H_0.

2) Um fabricante de monitores afirma que a vida útil de seus monitores é de 15000 horas. Uma amostra de 7 monitores apresentou uma média de 15990 horas e desvio padrão de 1552 horas. O que podemos dizer da afirmação do fabricante supondo $\alpha=0{,}05$? *Resp: Aceitamos a afirmação do fabricante.*

3) Determine o intervalo de confiança 95% para a percentagem populacional de defeituosos para os seguintes dados: N=1000, n=100 e x=10. *Resp: 10%$\pm$5,58%..*

4) Estime a média populacional com 90% de confiança para as seguintes situações: a) $\overline{x}=30$, $\sigma_x=4$, n=200 e N=2000; b) $\overline{x}=45$, $\sigma_x=6$, n=48 e N=600; c) $\overline{x}=10$, $\sigma_x=2$, n=120 e N=840. *Resp: a)30$\pm$0,4426; b) 45$\pm$1,3717; c) 10$\pm$0,2790*

5) Numa amostra de 300 observações acusou 30 pneus defeituosos numa remessa. Usando uma confiança de 99%, determine o erro de estimação. *Resp: e=0,04468.*

6) Numa tentativa de melhorar o atendimento no SAC, os funcionários procuraram estimar o tempo médio que gastam com cada cliente. Numa amostra aleatória de 60 clientes

colhida no período de três semanas acusou uma média de 30 minutos com desvio padrão de 6 minutos. Construa o intervalo de confiança de 98% para o verdadeiro tempo médio gasto em cada atendimento. *Resp: 30±1,8048.*

7) Numa estação de trem de grande circulação foi colhida uma amostra aleatória de 100 observações com média 30 e desvio padrão de 7. Determine com 99% de confiança uma cota superior para a média. *Resp: 30+1,8060.*

8) Um maratonista afirma ser capaz de dar 200 voltas em uma pista de corrida em 3 horas. Um estudo realizado através de uma associação atlética utilizando-se de 14 maratonistas, revelou que eles são capazes de executar 180 voltas no mesmo período de tempo, havendo um desvio padrão de 25 voltas. Assumindo α=5%, o que podemos dizer acerca do preparo físico do maratonista? *Resp:* $H_0{:}\mu=200$; $H_1{:}\mu\neq200$; $Vc=\pm1,96$; $t_{teste}=-2,99$. *Rejeitamos a afirmação do maratonista.*

9) Um pescador afirma pescar 18 quilos de peixe em uma hora. Uma pesquisa realizada em um grupo de 10 pescadores revelou que em média eles pescam 12 quilos de peixe em uma hora, com um desvio padrão de 4 quilos. Pode-se aceitar a afirmação do pescador? $H_0=\mu=18$; $H_1{:}\mu<18$; $t_{teste}=-4,7434$. *Rejeitamos.*

10) Uma embalagem de repelente para insetos indica um conteúdo de 100 ml. De uma amostra de 16 embalagens verificou-se uma média de 95,7ml com desvio de 3,4 ml. O que podemos dizer sobre a afirmação do fabricante? $H_0{:}\ \mu=100$; $H_1{:}\mu\neq100$. *Rejeitamos.*

11) Uma empresa que produz cordas afirma que suas cordas suportam 720 quilos. Uma amostra de 35 cordas foi testada e apresentou uma média de 698 quilos com desvio de 35 quilos. O que podemos dizer sobre a afirmação do fabricante? *Resp:* $H_0{:}\mu=720$; $H_1{:}\mu<720$; $t_{teste}=-3,326$. *Rejeitamos a afirmação do fabricante.*

12) O juizado de pequenas causas de uma cidade diz que em média seus processos são concluídos em 233 dias. Uma pesquisa foi feita com 50 usuários deste juizado e obtemos uma média de 231,5 dias para a conclusão de seus processos com um desvio padrão de 7,6 dias. Que podemos dizer sobre a afirmação deste juizado de pequenas causas? *Resp:*$H_0{:}\mu=233$; $H_1{:}\ \mu>233$; $z_{teste}=-1,3956$. *Aceitamos a afirmação do juizado.*

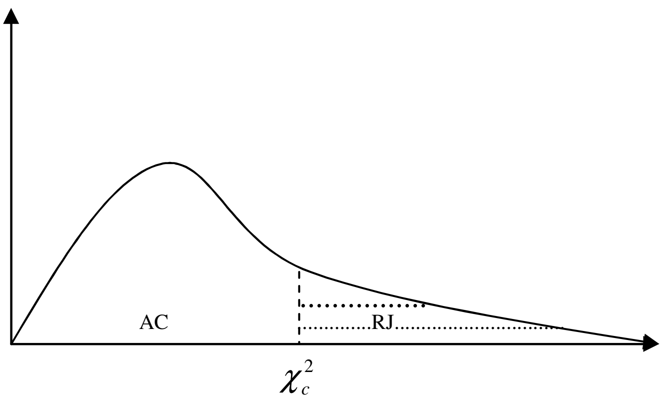
\# Testes ou provas não-paramétricos

Os testes não-paramétricos são bastante utilizados para tomada de decisões em relação a certos dados obtidos de pesquisas nas áreas das ciências humanas. Não é preciso admitir hipóteses sobre a distribuição de probabilidade da população, de onde foi retirada a amostra para análise, por essa razão os testes não-paramétricos são também conhecidos por testes ou provas livres de distribuição.

Teste Qui-Quadrado (ou de adequação do ajustamento)

É o mais conhecido dos testes não-paramétricos. Considere um espaço amostral Ω de um experimento aleatório. Sejam os eventos A_1 , A_2 , ... , A_k associados a esse experimento aleatório. Suponha que esse experimento aleatório foi repetido *n* vezes em idênticas condições. Sejam F_{o1} , F_{o2} , ... , F_{ok} as freqüências observadas dos eventos A_1 , A_2 , ... , A_k e F_{e1} , F_{e2} , ... , F_{ek} as freqüências esperadas (ou teóricas) desse K eventos. Faremos o teste do Qui-Quadrado para verificarmos se há adequação de ajustamento entre as freqüências observadas e as esperadas. Ou seja, fazemos o teste para sabermos se as discrepâncias (F_{oi} - F_{ei}), para i=1 , 2 , ... , k são devidas ao acaso, ou se existe de fato uma diferença significativa entre as freqüências observadas e as esperadas. Os passos para se fazer o teste do Qui-Quadrado são os seguintes:

1) Formular as hipóteses H_0 e H_1. (H_0 afirma não haver discrepâncias entre as freqüências, H_1 afirma que há discrepância entre as frequências)
2) Fixa-se o nível de significância α do teste e acha-se o valor crítico χ_c^2 na tabela da distribuição Qui-Quadrado usando que o grau de liberdade é gl=k-1, onde k é o número de eventos.
3) Determina-se no gráfico da distribuição Qui-Quadrado a região de aceitação de H_0 (AC) e de rejeição de H_0 (RJ).



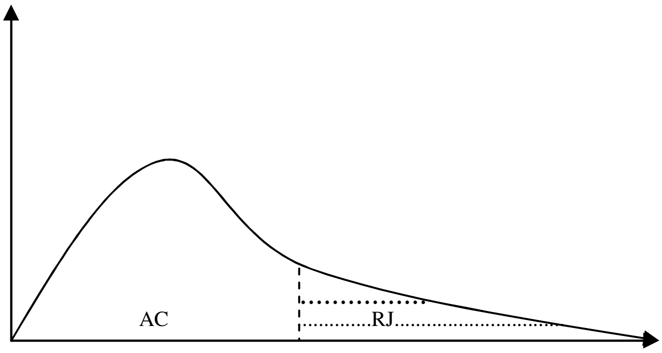
4) Calcula-se o valor teste utilizando a formúla:

$$\chi_t^2 = \sum_{i=1}^{K} \frac{(F_{oi} - F_{ei})^2}{F_{ei}} = \frac{(F_{o1} - F_{e1})^2}{F_{e1}} + \frac{(F_{o2} - F_{e2})^2}{F_{e2}} + ... + \frac{(F_{ok} - F_{ek})^2}{F_{ek}}$$

5) Se $\chi_t^2 < \chi_c^2$, então aceitamos H_0. Se $\chi_t^2 > \chi_c^2$ rejeitamos H_0.

Exemplo: Suponha que lançamos uma moeda 124 vezes e observamos cara (K) 81 vezes e coroa (C) 43 vezes. Vamos fazer o teste do Qui-Quadrado para sabermos se a moeda é honesta, usando um nível se significância de 5%, ou seja, α =5%.

1) H_0: A moeda é honesta
   H_1: A moeda não é honesta
2) Para α =5% temos na tabela da distribuição Qui-Quadrado χ_c^2 =3,841, pois o grau de liberdade é *gl*=k-1=2-1=1, lembrando que neste exemplo temos k=2, que é o número de eventos, ou seja, cara (K) e coroa (C).
3) Determina-se no gráfico da distribuição Qui-Quadrado as regiões de aceitação de H_0 (AC) e de rejeição de H_0 (RJ).



4) Calcula-se o valor teste χ_t^2:

As freqüências esperadas (ou teóricas) para os eventos A_1=K cara e A_2=C coroa, são exatamente a metade para cada evento, ou seja, F_{e1}=62 e F_{e2}=62, pois lançamos a moeda 124 vezes, logo esperamos que ocorra cara 62 vezes e coroa 62 vezes, isto é, metade de 124 para cada evento. A freqüência observada para o evento cara A_1=K foi F_{o1}=81 e para o coroa A_2=C foi F_{o2}=43.

$$\chi_t^2 = \sum_{i=1}^{2} \frac{(F_{oi} - F_{ei})^2}{F_{ei}} = \frac{(81-62)^2}{62} + \frac{(43-62)^2}{62} = 11,645$$

5) Como $\chi_t^2 = 11,645 > \chi_c^2 = 3,841$ rejeitamos H_0, ou seja, consideramos que a moeda não é honesta com um nível de significância α =5%, isto é, com um risco de 5%

Regressão e correlação

São técnicas estreitamente relacionadas que envolvem uma forma de estimação. Essa técnicas são utilizadas para estimar uma relação, que possa existir na população.

A correlação mede a força, ou grau, de relacionamento entre duas variáveis.
A regrassão nos dá uma equação matemática, que descreve o relacionamento entre as variáveis.

Regressão linear.

É a técnica de estabelecer uma equação matemática linear, que descreve o relacionamento entre duas variáveis.

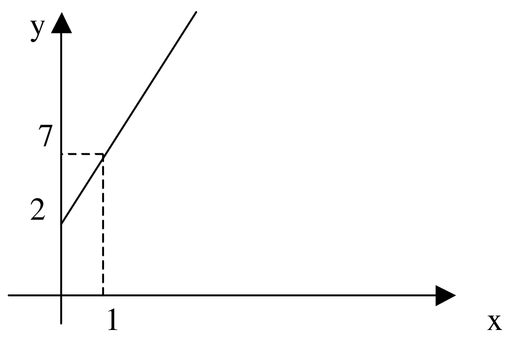
Ex: Dureza e resistência de um metal.
Ex: Procura de automóveis usados × aumento de carros novos (Causa e efeito).
Ex: Estimação de lucros(predizer valores futuros).

Função linear.

Y = a + bx ou f(x) = a + bx, pois y = f(x).
b:Coeficiente angular.
a: Coeficiente linear.

Os coeficientes a e b, são determinados com base nos dados amostrais.

Ex: y = 2 + 5x

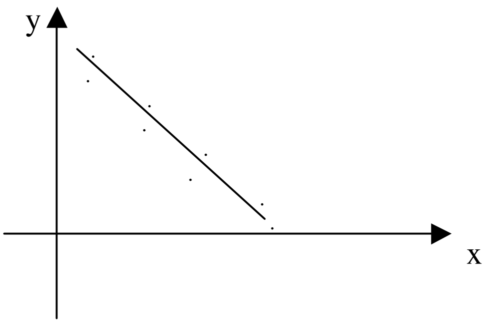


A reta intercepta o eixo y no ponto (0;a), isto é, y=a, neste caso no ponto (0;2), isto é, y = 2. Este ponto é chamado intercepto-y.

O coeficiente angular indica a variação de y, por unidade de variação de x. No exemplo para cada unidade de variação de x, correspondem a 5 unidades de variação de y.

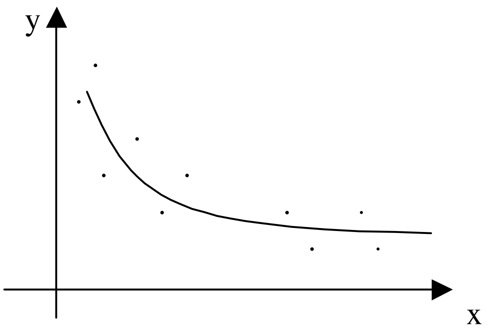
Decisão por um tipo de relação.

Nem todas as situações são bem aproximadas por uma função linear. Uma forma simples de verificar isto é colocar os dados no gráfico.
Ex:



Linear

Exemplo:

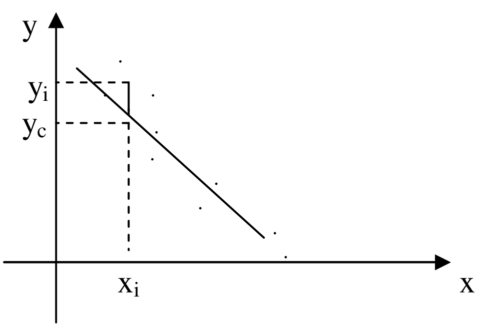


Determinação da função linear (método dos mínimos quadrados)

É o método mais usado para ajustar uma linha reta a um conjunto de pontos $(x_i;y_i)$. Esta reta tem as seguintes características:

1-A soma dos desvios padrões verticais dos pontos em relação à reta é zero.

$\sum(y_i - y_c) = 0$; onde y_i são as ordenadas dos pontos e y_c são as ordenadas dos pontos que estão sobre a reta que tem como abscissa x_i.



2-A soma dos quadrados desses desvios é mínimo.

$\sum (y_i - y_c)^2$ é mínimo

Os valores dos coeficientes de a e b, da equação da reta de ajuste $y_c = a + bx$, são calculados resolvendo as seguintes equações ditas normais:

1) $\sum y = na + b\sum x$

2) $\sum xy = a\sum x + b\sum x^2$; onde n é o número de pares de dados (x;y) observados.

Resolvendo as equações acima temos:

$$a = \frac{\sum y - b\sum x}{n} \qquad \text{e} \qquad b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

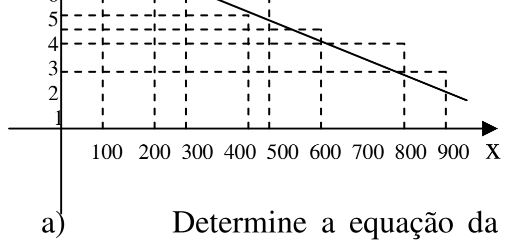
A reta de regressão passa pelo ponto $(\bar{x};\bar{y})$, isto é, determinada a equação de ajuste $y_c = a + bx$, tem-se que $\bar{y} = a + b\bar{x}$, onde $\bar{x}$ e $\bar{y}$ são respectivamente a média das abscissas x e das ordenadas y, dos pares de dados (x;y).

Exemplo: Os dados abaixo representam a freqüência de acidentes e o nível de esforço preventivo educacional. (Exemplo do Stevenson, p. 350)

Acidentes por milhão de homens/horas (y)	Homens/horas por mês para educação(x)
7	200
6,4	500
5,2	450
4,0	800
3,1	900
8,0	150
6,5	300
4,4	600

Suponha que queiramos saber, se há uma relação entre o nível de esforço preventivo, e a freqüência de acidentes, isto é, se a freqüência de acidentes depende do nível de esforço preventivo. O nível de esforço preventivo seria a variável independente (x) e a freqüência de acidentes a variável dependente (y).

Colocando os dados no gráfico, vemos que eles podem ser aproximados por uma equação linear, pois estão próximos a uma reta. Veja o gráfico abaixo.



a) Determine a equação da reta, que melhor descreva a relação entre acidentes e horas de esforço preventivo.

b) Estime a freqüência de acidentes se o esforço preventivo for de 700 horas.

a) $$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{8.18720 - 3900.44,6}{8.2415000 - 1521000} = -0,0059$$

$$a = \frac{\sum y - b\sum x}{n} = \frac{44,6 + 0,0059.3900}{8} = 8,4513$$

Como a equação da reta de ajuste é y = a + bx, temos y = 8,45 – 0,0059x.

b) y = 8,4513 – 0,0059.700 = 4,3213

Exemplo: (Stevenson, p.351 ex.9): Determine uma equação preditora do montante de seguro, em função da renda anual, com base nos seguintes dados:

Renda anual(em $1000) (x)	Seguro (y)
20	10
25	12
26	15
18	10
16	15
17	20
32	30
13	5
38	40
40	50
42	40

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 1,3210$$

$$\sum x = 287$$

$$\sum x^2 = 8571$$

$$a = \frac{\sum y - b\sum x}{n} = -12,0121$$

$$\sum y = 247$$

$$\sum y^2 = 7819$$

$$\sum xy = 7875$$

y = -12,0121 + 1,3210x

a) Qual o montante do seguro se a renda for 29?

Y = -12,0121 + 1,3210.29 = 26,2969

c) Qual a renda anual se o montante do seguro for 25?
25 = -12,0121 + 1,3110x ⇒ x = 28,0182 (HP 12C x = 28,0178)

O coeficiente de determinação, r^2

$$r^2 = \frac{\left(\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}\right)^2}{\left[\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right] \cdot \left[\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2\right]}$$

Temos que $0 < r^2 < 1$. Se r^2 estiver próximo de 1, significa que os pontos estão próximos da reta de ajuste, logo temos uma boa estimação. Se estiver próximo de zero, segnifica que temos dados distantes da reta de ajuste, logo temos uma péssima estimação.

Exemplo: No exemplo da freqüência de acidentes × nível de esforço preventivo educacional temos:

$r^2 = 0{,}9084$. Temos um valor próximo de 1, significando que 90,84% da variação dos acidentes estão relacionados com a variação do nível de esforço preventivo, e apenas 9,16% da variação não são explicados pela variação do nível de esforço.

Correlação (o coeficiente r de Pearson)

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$

Temos que $-1 < r < 1$. Se r estiver próximo de –1, então os dados estão próximos da reta de ajuste, que é decrescente. Se r estiver próximo de 1, então os dados estão próximos da reta de ajuste, que é crescente. Se r estiver próximo de zero, temos que os dados estão distantes da reta de ajuste, logo temos uma péssima estimação.

Exercícios:

1) Se deseja avaliar a relação existente entre o número de horas (x_i) e a nota obtida (y_i). Os dados estão apresentados na tabela seguinte. Qual a equação da reta ajustada entre x e y? Qual a qualidade do ajuste?

i	1	2	3	4	5	6
x_i	1	3	4	5	6	6,5
y_i	6	7	7.5	8	8,5	8,7

Resp: y = 55146 + 0,4946x ; r^2 = 0,9997 = 99,97%

2) Os dados da tabela abaixo representam o consumo e a renda disponível. Com base nos dados apresentados, responda ‘as questões apresentadas a seguir.

Anos	Consumo(y), em R$ milhões	Renda(x), em R$ milhões

1960	158	189
1961	160	209
1962	163	220
1963	165	235
1964	170	250

a) Verificar a qualidade do ajuste dos dados à reta de regressão.
b) Se a qualidade do ajuste for boa, encontre a reta de regressão.
c) Qual o consumo esperado para uma renda de 400 milhões de reais?
Resp: a) $r^2 = 0{,}9543$; b) $y = 120{,}4445 + 0{,}1938x$; c) R$ 197,9703 milhões

3) A seguir estão apresentadas as vendas e os custos da industria água fria Ltda. Com base nos valores apresentados determine: a) a equação da de ajuste; b) Os coeficientes de correlação e determinação; c) o volume dos custos fixos projetados para vendas de \$300 e \$650; d) o volume de vendas previsto, para custos iguais a \$220.

Resp: a) $y = 20{,}5716 + 0{,}4582x$; b) $r = 0{,}9826$ e $r^2 = 0{,}9656$; c) 158,0175 e 318,3710; d) 435,2878

Número índices

São usados para indicar variações relativas em quantidades, preços, ou valores de um artigo, durante dado período de tempo.

- Número índices simples.

Avalia a variação relativa de um único item ou variável econômica, entre dois períodos de tempo.

Relativo de preço = $\frac{p_n}{p_0}.100.$

Relativo da quantidade = $\frac{q_n}{q_0}.100.$

Relativo do valor = $\frac{p_n q_n}{p_0 q_0}.100.$

Onde, p_0 = preço de um item no ano – base
q_0 = quantidade de um item no ano – base
p_n = preço de um item em determinado ano
q_n = quantidade de um item em determinado ano

Exemplo: Considere a tabela abaixo, que indica o preço e o volume médio de um determinado modelo de moto:

Ano	Preço médio de venda(\$)	N$^{\underline{0}}$ Vendido	Receita(em 1000)(\$)
1996	1500	30	45
1997	1800	33	59,4
1998	2100	31	65,1

1999	2300	35	80,5
2000	2350	42	98,7

Tomando 1996 como ano – base, temos o preço de \$1500 relativo a 100% e os outros preços medidos em relação a este. Analogamente o volume de 30 será relacionado a 100% e a receita de 45.000 relacionada também a 100%, os outros volumes e receitas medidos em relação a estes.

Os números – índices (relativos) para preço, quantidade e valor de 1998, tomando 1996 como ano – base são os seguintes:

Preço: $\frac{p_{1998}}{p_{1996}}.100\% = p_{\frac{1996}{1998}} = \frac{2100}{1500}.100 = 140$ ou 140%. Como 140%-100% = 40%, este índice indica um aumento de 40% no preço em relação ao preço de 1996.

Quantidade: $\frac{q_{1998}}{q_{1996}}.100\% = q_{\frac{1996}{1998}} = \frac{31}{30}.100 = 103{,}3$ ou 103,3%. Como 103,3%-100% = 3,3%, este índice indica um aumento de 3,3% da quantidade vendida em 1998, em relação a quantidade de 1996.

Valor: $\frac{p_{1998}.q_{1998}}{p_{1996}.q_{1996}}.100\% = p.q_{\frac{1996}{1998}} = \frac{65{,}1}{45}.100 = 144{,}7$ ou 144,7%. Como 144,7%-100% = 44,7%, este índice indica um aumento de 44,7% na receita de 1998, em relação à receita de 1996.

Elos e cadeias relativos.

Elos de relativos: $p_{\frac{1}{2}}, p_{\frac{2}{3}}, p_{\frac{3}{4}},....$, representam os preços relativos de cada intervalo de tempo, referidos ao anterior.

Relativos em cadeia: $p_{\frac{1}{n}} = p_{\frac{1}{2}}.p_{\frac{2}{3}}....p_{\frac{n-1}{n}}$

$$p_{\frac{1}{n}} = \frac{p_n}{p_1} = \frac{p_2}{p_1}.\frac{p_3}{p_2}.\frac{p_4}{p_3}....\frac{p_{n-1}}{p_{n-2}}.\frac{p_n}{p_{n-1}}$$

Exemplo: Os elos relativos dos preços no período de 1997 a 2000 são: 111; 109; 132; 121; respectivamente. Determine o preço relativo de 2000, tomando 1996 como ano-base.

$\frac{p_{1996}}{p_{1997}} = p_{\frac{1997}{1996}} = 1{,}11$; $\frac{p_{1997}}{p_{1998}} = p_{\frac{1998}{1997}} = 1{,}09$; $\frac{p_{1998}}{p_{1999}} = p_{\frac{1999}{1998}} = 1{,}32$; $\frac{p_{1999}}{p_{2000}} = p_{\frac{2000}{1999}} = 1{,}21$. Logo

$\frac{p_{1996}}{p_{2000}} = p_{\frac{2000}{1996}} = \frac{p_{1996}}{p_{1997}}.\frac{p_{1997}}{p_{1998}}.\frac{p_{1998}}{p_{1999}}.\frac{p_{1999}}{p_{2000}}$ = 1,11.1,09.1,32.1,21 = 1,93 = 193 ou 193%. Logo um aumento de 93%

Método agregativo simples.

Índice de preço agregativo simples: $\frac{\sum p_n}{\sum p_0}$, onde p_n são os preços em um determinado ano e p_0 os preços do ano-base.

Exemplo: A tabela abaixo apresenta o preço médio da saca de 50Kg do feijão, açúcar efarinha nos anos de 1996, 1998, e 2000. Calcule o índice agregativo simples dos preços para o ano de 2000, tomando como ano-base 1996.

	1996	1998	2000
Feijão	50	55	53
Açúcar	30	36	40
Farinha	40	38	40

$\frac{\sum p_n}{\sum p_0} = \frac{53+40+40}{50+30+40} = 1{,}11 \Rightarrow$111 ou 111%. Logo um aumento de 11%.

\# Método das médias simples dos relativos: $\frac{\sum \frac{p_n}{p_0}}{N}$, onde N é o número de relativos.

Exemplo: Considere o exemplo acima. Calcule a média simples dos relativos.

$\frac{\frac{53}{50}+\frac{40}{30}+\frac{40}{40}}{3} = 1{,}13 \Rightarrow$ 113 ou 113%. Logo um aumento de 13%.

\# Método agregativo ponderado.

São mais precisos que o agregativo simples, pois considera-se o preço, e a quantidade.

Índice de Laspeyres ou método do ano-base

Preço: $\frac{\sum p_n q_0}{\sum p_0 q_0}.100$

Quantidade: $\frac{\sum q_n p_0}{\sum p_0 q_0}.100$

Valor: $\frac{\sum p_n q_n}{\sum p_0 q_0}.100$

Índice de Paasche ou método do determinado ou época atual.

Preço: $\frac{\sum p_n q_n}{\sum p_0 q_n}.100$

Quantidade: $\frac{\sum p_n q_n}{\sum q_0 p_n}.100$

Índice de Fischer ou método ideal.

Utiliza os dois índices anteriores.

$IF_p = \sqrt{IL_P . IP_P}$

IL_P é o índice de Laspeyres para preços.
IP_P é o índice de Paasche para preços.

$IF_Q = \sqrt{IL_Q . IP_Q}$

IL_Q é o índice de Laspeyres para quantidade.
IP_Q é o índice de Paasche para quantidade.

Ex: calcule os índices de Laspeyres, Paasche e Fischer, dos produtos dados na tabela abaixo:

ANO	1999	1999	1999	2000	2000	2000
	Preço	Quantidade	Valor	Preço	quantidade	Valor
Feijão	50	20	1000	60	30	1800
Açucar	30	50	1500	40	60	2400
			$\sum p_0 q_0 = 25000$			$\sum p_n q_n = 42000$

	Laspeyres	Laspeyres	Paasche	Paasche	Fischer	Fischer
ANO	1999	2000	1999	2000	1999	2000
Preços	100	128	100	127	100	128
Quantidade	100	132	100	131	100	131,5

Laspeyres.

Preço: $\frac{\sum p_n q_0}{\sum p_0 q_0}.100 = \frac{60.20 + 40.50}{2500}.100 = 1{,}28.100 = 128 \text{ ou } 128\%$

Quantidade: $\frac{\sum q_n p_0}{\sum p_0 q_0}.100 = \frac{30.50 + 60.30}{2500}.100 = 1{,}32.100 = 132 \text{ ou } 132\%$

Paasche.

Preço: $\frac{\sum p_n q_n}{\sum p_0 q_n}.100 = \frac{4200}{50.30 + 60.30}.100 = 1{,}27.100 = 178 \text{ ou } 127\%$

Quantidade: $\frac{\sum p_n q_n}{\sum q_0 p_n}.100 = \frac{42000}{20.60 + 50.40}.100 = 1{,}31.100 = 131 \text{ ou } 131\%$

- Fischer.

Preço: $IF_p = \sqrt{IL_P . IP_P} = \sqrt{128.127} = 128$

Quantidade: $IF_Q = \sqrt{IL_Q . IP_Q} = \sqrt{132.131} = 131{,}5$

Exercícios:

1) Calcule o índice de preço de Laspeyres, admitindo como ano-base 1991, e utilizando os dados da tabela abaixo.

	1991	1991	1992	1992	1993	1993	1994	1994
	Preço	Quantidade	Preço	Quantidade	Preço	Quantidade	Preço	Quantidade
Lápis	0,5	125	0,5	140	0,3	234	0,4	179
Caneta	3,2	103	3,3	105	3,4	96	3,5	69
Grafite	1,4	56	1,3	62	1,4	81	1,5	85

Resp: 100%, 101%, 99%, 105%.

2) Obtenha os índices de quantidade de Paasche, utilizando os dados da tabela da questão 1).
Resp: 100%; 105%; 109%; 89%.

3) Um apartamento foi comprado por R$10.000 em 1998 e por R$ 10.5000 no ano seguinte. Calcule o relativo de preço em 1999, considerando como ano-base o ano de 1998.
Resp: 1,05 ou 105 ou 105%

4) Um corretor vendeu 800 apartamentos em 1997 e em 1995 vendeu 6000. Calcule o relativo de quantidade em 1995, considerando 1995 como ano-base.
Resp: 133%

5) Uma concessionária vendeu em 1996, 16.000 carros com preço unitário de R$10.000. Em 1998 vendeu 19.000 carros com preço unitário de R$12.000. Calcule o relativo de valor em 1998. Resp: 142,5

6) Os preços de uma saco de arroz eram de R$1,00 em 1995 e R$2,00 em 2000. Calcule o preço relativo tomando 1995 com ano-base e 2000 como o ano dado.
Resp:2,0 ou 200%

Exércícios

1º) Dado o rol abaixo construir a distribuição em classes de freqüência.
ROL: 23 ; 25,1 ; 25,2 ; 32,4 ; 34,3 ; 34,3 ; 34,3 ; 37,4 ; 48; 49 ; 49; 49; 51,2; 51,5 ; 53 ; 61 ; 65,5 ; 65,6 ; 67,7 ; 67,8 ; 70,3 ; 79,4 ; 79,6 ; 79,6 ; 79,7 ; 81 ; 81 ; 81 ; 81,2 ; 89 ; 89 ; 93 ; 93,9 ; 95 ; 101 ; 103 ; 107 ; 107 ; 107 ; 107,3 ; 108.

2º) Determine a média e a mediana do ROL e da distribuição em classes de freqüência que você construiu no item anterior. A média e a mediana da distribuição em classes são valores próximos da média e da mediana do ROL?

3º) Em cada uma das situações abaixo utilize o cálculo da média correta. Escolha entre o cálculo da média aritmética, da média geométrica ou da média harmônica. Justifique a utilização de cada média em cada situação.

a) Um motorista foi de Lauro de Freitas para Camaçarí à velocidade média de 65 km/h. Depois foi de Camaçarí para Feira de Santana à velocidade média de 73 km/h. Depois voltou de Feira de Santana para Salvador à velocidade média de 81 km/h. Calcule a velocidade média de todo este percurso.

b) Uma empresa alimentícia teve uma receita de \$3.000.000,00 no ano de 1980 e uma receita de \$15.000.000,00 no ano 2000. Estime a receita desta empresa no ano de 1990.

4º) Dado o rol abaixo construir a distribuição em classes de freqüência.ROL: 23,3 ; 24,2 ; 25 ; 26 ; 26,7 ; 34,6 ; 34,7 ; 45 ; 45,8 ; 46,1 ; 46,1 ; 47 ; 47 ; 49 ; 49,2 ; 50; 57,2 ; 58; 59 ; 63 ; 68,4 ; 68,5 ; 76 ; 77 ; 78 ; 78,1 ; 78,1 ; 79,8

5º) Determine a média e a mediana do ROL e da distribuição em classes de freqüência que você construiu no item anterior. A média e a mediana da distribuição em classes são valores próximos da média e da mediana do ROL?

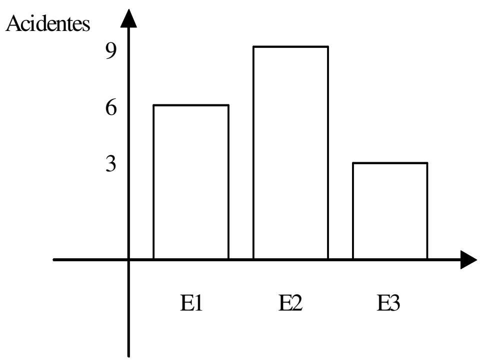
6º) As notas de 9 candidatos a uma vaga de eletrotécnico em uma industria foram as seguintes: 5,0 ; 7,2 ; 6,8 ; 5,3 ; 8,3 ; 7,2 ; 8,0 ; 8,2 ; 6,8. Sabendo-se que a nota de corte seria determinada pelo 3º quartil, determine as notas dos candidatos selecionados.

7º) Dado o rol abaixo construir a distribuição em classes de freqüência.

ROL: 2,3 ; 2,3 ; 2,5 ; 2,5 ; 2,5 ; 3 ; 4,5 ; 6 ; 6 ; 7 ; 8 ; 11,9 ; 11,9 ; 12,3 ; 14 ; 14 ; 15 ; 16 ; 17
22,1 ; 22,4 ; 22,4 ; 22,5 ; 22,5 ; 23 ; 23,7 ; 24,1 ; 24,9 ; 25,3 ; 26,2 ; 27 ; 27 ; 27

8º) Determine a média e a mediana do ROL e da distribuição em classes de freqüência que você construiu no item anterior. A média e a mediana da distribuição em classes são valores próximos da média e da mediana do ROL?

9º) Em três empresas do pólo petroquímico de Camaçarí obtivemos os seguintes dados referentes a números de acidentes de trabalho no ano de 2002: Na empresa E1 tivemos 6 acidentes. Na empresa E2 ocorreram 9 acidentes e na empresa E3 ocorreu 3 acidentes. Sabendo-se que nas empresas E1, E2 e E3 existem respectivamente 300, 450 e 150 funcionários, podemos representar estes dados no gráfico abaixo e concluir que a empresa E2 é menos eficiente na prevenção de acidentes comparada com as empresas E1 e E3? Por quê? Em caso negativo represente estes dados em um gráfico correto.



10º) A tabela abaixo apresenta o IPC (Índice de Preços ao Consumidor) nas regiões do Brasil, e seus respectivos pesos proporcionais à população de cada região. Calcule o INPC (Índice Nacional de Preços ao Consumidor) com base nos IPC's ponderados de cada região.

REGIÃO	IPC	PESOS
Sudeste	23,40	3,74
Centro-Oeste	25,90	1,73
Nordeste	31,80	1,24
Norte	27,70	0,78
Sul	21,20	1,07

11º) Um grupo de pessoas foi classificado segundo o peso e o nível de colesterol no sangue, onde a proporção encontra-se na tabela abaixo:

	Peso			
Colesterol	Excesso	Normal	Baixo	Total
Alto	0,20	0,18	0,17	0,55
Normal	0,12	0,10	0,23	0,45
Total	0,32	0,28	0,40	1,00

a) Qual a probabilidade de uma pessoa escolhida ao acaso ter colesterol normal?
b) Se escolhermos uma pessoa desse grupo e ela tiver colesterol alto, qual a probabilidade que ela tenha excesso de peso?

12º) Em uma fábrica temos duas máquinas, máquina A e máquina B. A máquina A é responsável por 53,7% da produção, e a máquina B por 46,3% da produção. A máquina A produz 1,3% de peças com defeito e a máquina B 2,5% de peças com defeito.

a) Se uma peça é escolhida ao acaso e verificamos que ela é defeituosa, qual a probabilidade dela ter sido produzida pela máquina B?
b) Se escolhemos uma peça ao acaso, qual a probabilidade dela ser boa?

13º) Os diâmetros dos parafusos de determinado lote estão dados em uma distribuição em classes de freqüência abaixo:

Classes (Diâmetros)	fi
2,894 ⊢— 2,924	356
2,924 ⊢— 2,954	476
2,954 ⊢— 2,984	212
2,984 ⊢— 3,014	132
3,014 ⊢— 3,044	76
$\sum$	1252

a) Qual a média desses parafusos?
b) Determine o diâmetro do parafuso que representa a mediana.

14º) Um grupo de pessoas foi classificado segundo o peso e pressão arterial, onde a proporção encontra-se na tabela abaixo:

	Peso			
Pressão	Excesso	Normal	Baixo	Total
Alta	0,32	0,11	0,10	0,53
Normal	0,18	0,16	0,13	0,47
Total	0,50	0,27	0,23	1,00

a) Qual a probabilidade de uma pessoa escolhida ao acaso ter peso baixo e pressão alta?
b) Se escolhermos uma pessoa desse grupo e ela tiver excesso de peso, qual a probabilidade que ela tenha também pressão alta?

15º) Em uma fábrica temos três máquinas, máquina A, máquina B e máquina C. A máquina A é responsável por 36,3% da produção, a máquina B por 27,9% da produção e a máquina C por 35,8% da produção. A máquina A produz 1,5% de peças com defeito a máquina B 1,9% de peças com defeito e a máquina C 2,1% de peças com defeito.

a) Se uma peça é escolhida ao acaso e verificamos que ela é perfeita, qual a probabilidade dela ter sido produzida pela máquina C?
b) Se escolhermos uma peça ao acaso, qual a probabilidade dela ser defeituosa?

16º) Cinco empresas petroquímicas do pólo de Camací-Ba informaram ao sindicato a média dos salários pagas por elas, conforme tabela abaixo. Qual a média geral dos salários pagos por essas empresas?

EMPRESA	MÉDIA SALARIAL	FUNCIONÁRIO
Buteno	1245,89	456
Cherênio	1763,92	412
Fogênio	1097,93	567
Norteno	978,13	674
Soluteno	1117,43	540

17º) Uma moeda é viciada de modo que a probabilidade de sair K (cara) é 7/3 da probabilidade de sair C (coroa). Se lançarmos essa moeda qual a probabilidade de sair:

c) Cara (K)
d) Coroa (C))

18º) (2,0 pontos) Em uma urna temos 20 bolas enumeradas de 1 a 20.

c) Se retirarmos uma bola par dessa urna, qual a probabilidade dessa bola ser maior ou igual a 13?
d) Qual a probabilidade de uma bola menor que 17 ser retirada dessa urna?
e) Qual a probabilidade de uma bola que seja múltipla de 3 e múltipla de 2 ser retirada dessa urna?

19º) Em uma urna temos quatro moedas. A moeda M1 é uma moeda normal, a moeda M2 é viciada de tal modo que sair cara (K) é 1219 vezes mais provável que sair coroa (C)), a moeda M3 tem duas caras (K) e a moeda M4 tem duas coroas (C). Uma moeda é escolhida ao acaso e lançada.

a) Se o resultado obtido foi coroa (C), qual a probabilidade da moeda lançada ter sido a moeda M2?
b) Qual a probabilidade de observarmos moeda M3 e cara (K)?
c) Qual a probabilidade de observarmos coroa (C)?

20º) Os registros de uma pequena empresa indicam que 43,98% das faturas por elas emitidas são pagas com atraso. De 5 faturas expedidas, determine a probabilidade de ao menos 2 serem pagas com atraso.

21º) Os defeitos em fios de tear de uma determinada industria são bem aproximados por um processo de Poisson com média de 0,7 defeitos a cada 7 metros de fio. Determine a probabilidade de em 5 metros de fio termos no máximo 1 defeitos.

22º) Os pesos de 2346 funcionários de uma empresa são normalmente distribuídos com média de 78,8Kg e desvio padrão de 6,3Kg. Encontre o número de funcionários que pesam mais de 78 Kg.

23º) A média amostral das idades de 31 estudantes de uma determinada faculdade é de 28,3 anos com desvio padrão de 4,1 anos. Estime a média populacional dos estudantes desta faculdade com uma confiança de 75,93%.

24º) Um fabricante de sapatos diz que seus sapatos têm média de vida de 27,9 meses e desvio padrão de 7,3 meses. Qual a percentagem de amostras de tamanho 32 terão média de vida em um intervalo de 2 meses em torno da média?

25º) Uma industria de cd's tem uma proporção de 0,0987 dos cd's que são fabricados com algum tipo de defeito. Determine a probabilidade de em uma amostra de 9 cd's termos mais que 7 cd's com defeito.

26º) Uma industria de tecidos tem sua produção apresentando defeitos nos tecidos em uma média de 0,63 defeitos a cada 2,17m^2 de tecido produzido, sabendo-se que estes defeitos são bem aproximados pela distribuição de Poisson, determine a probabilidade de termos no mínimo 1 defeito em 12,89m^2 de tecido produzido.

27º) Os pesos dos funcionários de uma empresa são normalmente distribuídos com média de 76,7Kg e desvio padrão de 6,4Kg. Qual a probabilidade de um funcionário escolhido aleatoriamente ter peso maior que 77,5Kg?

28º) As alturas de 8 estudantes de uma certa faculdade de administração são as seguintes: 1,71 ; 1,67 ; 1,78 ; 1,81 ; 1,77 ; 1,81 ; 1,65 ; 1,69. Estime a média populacional das alturas dos estudantes desta faculdade com um nível de significância de 0,0005.

29º) Utilizando 9 sons diferentes de um aparelho eletrônico, um rapaz criou sinais escolhendo 3 seqüências de sons desses 9 sons para se comunicar com sua namorada.

a) Quantos sinais este rapaz pôde produzir?
b) Se escolhermos uma dessas seqüências aleatoriamente, qual a probabilidade de que seja uma seqüência de três sons iguais?

30º) Dos pedidos de financiamento de estudos ao governo federal 61,73% são atendidos. Determine a probabilidade de em 9 pedidos termos 3 _não_ atendidos.

31º) Em uma determinada estrada passam 1,7 carros em cada 5 minutos. Determine a probabilidade de passar 7 carros em 1 hora.

32º) Os pesos de 1235 funcionários de uma empresa são normalmente distribuídos com média de 78,3Kg e desvio padrão de 3,3Kg. Encontre o número de funcionários que pesam entre 77Kg e 79Kg.

33º) Foi retirada uma amostra de tamanho 39 de um determinado lote de pneus novos e verificou-se que em média estes pneus rodam 47698Km. Estime a média de quilômetros que estes pneus podem rodar com uma confiança de 77,98% sabendo-se que o desvio padrão da amostra é de 546Km.

34º) Um comerciante compra jarros em grandes lotes. Periodicamente ele verifica os lotes para determinar a proporção de quebrados. Se um lote contém 9,83% de jarros quebrados, qual a probabilidade do comerciante em uma amostra de 37 jarros encontrar 8% ou menos de jarros quebrados?