Apostila de Estatística Básica

Curso de Psicologia

2º Semestre - Volume II

Probabilidades, Distribuição Normal, Teste de Hipóteses, Distribuição Qui-Quadrado e Correlação.

Prof. Dr. Celso Eduardo Tuna

Capítulo 8 - Probabilidade

8.1 Conceito

Intuitivamente pode-se definir probabilidade como:

$$p(A) = \frac{\text{número de casos favoráveis a A}}{\text{número total de casos possíveis}}$$

Ao conjunto desses casos possíveis dá-se o nome de espaço amostral (S). E ao conjunto de casos favoráveis a A dá-se o nome de evento A.

Ex 1) Probabilidade de se obter um número par como resultado de um lançamento de um dado:

S = {1,2,3,4,5,6} e A = {2,4,6}, então p = 3/6 = 1/2 =0,5 ou 50 %

Ex 2) Probabilidade de se obter o número 4 como resultado de um lançamento de um dado:

S = {1,2,3,4,5,6} e A = {4}, então p = 1/6 0,167 ou 16,7 %

Ex 3) Probabilidade de se obter um número diferente de 4 no lançamento de um dado:

S = {1,2,3,4,5,6} e A = {1,2,3,5,6}, então p = 5/6 0,833 ou 83,3 %

8.2 Eventos Complementares

O evento do exemplo 3 é denominado de **complementar** do evento do exemplo 2. Ou seja, se p é a probabilidade de um evento ocorrer e q é a probabilidade de que ele não ocorra, então:

$$p + q = 1 \Rightarrow q = 1 - p$$

8.3 Exercícios

1) Uma pesquisa do PC World foi realizada com 4000 proprietários de computadores pessoais, e verificou que 992 dos computadores apresentaram falhas num intervalo de 02 anos após a compra. Tomando por base esses resultados, qual a probabilidade de você comprar um computador pessoal e ele apresentar problema nos próximos dois anos ?

Resp: a) p = 0,248 ou 24,8 %

2) Um grupo de alunos é composto de 15 homens e 35 mulheres. O professor sorteia aleatoriamente alguém do grupo. Qual a probabilidade de não ser mulher ?

Resp: a) p = 0,3 ou 30 %

3) A tabela abaixo descreve os alunos registrados pelo período de uma semana num curso. A distribuição segue de acordo com o sexo e com a idade.

Idade	Sexo		Total
	Feminino	Masculino	
Abaixo de 20 anos	20	15	35
Entre 20 e 40 anos	65	150	215
Acima de 40 anos	50	95	145
Total	135	260	395

Se um aluno é escolhido ao acaso, qual a probabilidade:

a) de ser mulher ?
b) de ser mulher e ter acima de 40 anos ?
c) de ser homem e ter menos de 20 anos ?
d) de ser mulher entre 20 e 40 anos ?
e) de ser homem e ter menos de 40 anos ?
f) ter entre 20 e 40 anos ?

Resp: a) p = 0,342 b) p = 0,127 c) p = 0,038 d) p = 0,165 e) p = 0,418 f) p = 0,544

4) Um casal planeja ter 3 filhos. Determine a probabilidade de nascerem:
a) três homens;
b) dois homens e uma mulher.

Resp: a) p = 1/8 b) p = 3/8

EXERCÍCIOS COMPLEMENTARES

Probabilidades:

1) Uma sala contém 10 homens e 20 mulheres, sendo que a metade dos homens e três quartos das mulheres têm olhos castanhos. Uma pessoa é escolhida ao acaso. Determine:

a) a probabilidade de ser homem;
b) a probabilidade de ser homem e ter olhos castanhos;
c) a probabilidade de ser mulher com olhos castanhos;
d) a probabilidade de ter olhos castanhos;
e) a probabilidade de ser mulher ou ter olhos castanhos;
f) a probabilidade de ser mulher, dado que a pessoa escolhida tem olhos castanhos.

Resposta: a) 1/3 b) 1/6 c) 1/2 d) 2/3 e) 5/6 f) 3/4

2) Seja uma família sorteada de uma população de 120 famílias, as quais se distribuem conforme a tabela.

Matriculada no programa bolsa escola	Grau de instrução do chefe da casa			Total
	Nenhum	1º grau	2º grau	
Sim	30	20	25	75
Não	10	15	20	45
Total	40	35	45	120

Calcule a probabilidade de a família sorteada :

a) Ser matriculada no programa bolsa escola:
b) ter o chefe da casa com o 2º grau:
c) Ser matriculada no programa bolsa escola e o chefe da casa ter o 2º grau:
d) Ser matriculada no programa bolsa escola, considerando que o sorteio tenha sido restrito às famílias cujo chefe da casa tem o 2º grau:

a) 62,5% b) 37,5% c) 20,83% d) 55,55%

3) Dos 100 alunos de uma turma, 40 gostam de álgebra, 30 gostam de geometria, 10 gostam de álgebra e geometria, e há os que não gostam de álgebra nem de geometria. Um aluno é escolhido ao acaso. Qual a probabilidade de ele gostar de:

a) Álgebra? b) geometria? c) álgebra e geometria? d) álgebra ou geometria?

Resp. a) 0,4 b) 0,3 c) 0,1 d) 0,6

Capítulo 9 - Distribuição Normal

9.1 Distribuição de Probabilidade

Seja a seguinte distribuição de freqüência:

Número de Professores faltosos por dia, em 1 mês	Freqüências
0	22
1	5
2	2
3	1
Total	30

Através dos dados apresentados pode-se calcular a probabilidade de em um dia:

não faltar nenhum professor: P = 22/30 = 0,73
faltar 1 professor: P = 5/30 = 0,17
faltarem 2 professores: P = 2/30 = 0,07
faltarem 3 professores: P = 1/30 = 0,03

Podemos então elaborar uma tabela denominada **distribuição de probabilidade**:

Número de Professores faltosos por dia, em 1 mês	Probabilidade
0	0,73
1	0,17
2	0,07
3	0,03
Total	1,00

Pode-se então determinar uma função que associe a variável faltas com a sua probabilidade, denominada função probabilidade denominada por:

$$F(x) = P\ (X = x_i)$$

9.2 Distribuição Normal

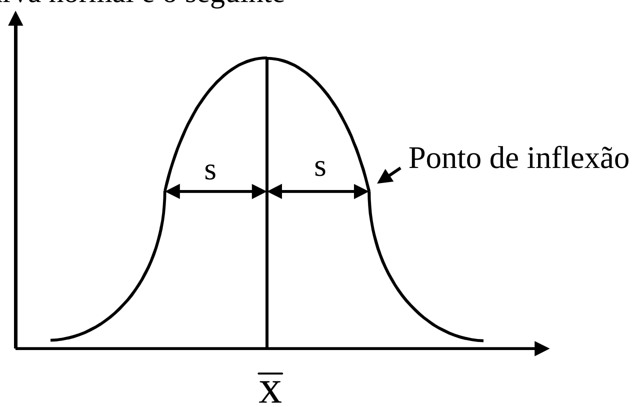
Relembrando:

Variável é o conjunto de resultados possíveis de um fenômeno. A variável pode ser qualitativa, quando seus valores são expressos por atributos (ex: sexo, cor), ou pode ser quantitativa, quando seus valores são expressos em números.

A variável quantitativa pode ser contínua, quando assume qualquer valor entre dois limites (ex: peso, altura, medições), ou pode ser discreta, quando só pode assumir valores pertencentes a um conjunto enumerável (ex: número de filhos, contagens em geral).

Entre as distribuições teóricas de variável contínua, a mais empregada é a distribuição normal.

O aspecto gráfico da curva normal é o seguinte



Onde $\overline{x}$ é a média e s é o desvio padrão.

Quando nos referimos a uma distribuição normal, cita-se a média e o seu desvio padrão. $N(\overline{x}, s)$

A equação da curva é a seguinte: $Y = \dfrac{1}{s\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{X-\overline{x}}{s}\right)^2}$

Quando temos em mão uma variável aleatória com distribuição normal, nosso principal interesse é obter a probabilidade de essa variável aleatória assumir um valor em um determinado intervalo. Essa probabilidade é representada pela área sob a curva dentro desse intervalo. A área total sob a curva é 1. O cálculo desse valor é difícil, sendo então esse já tabelado.

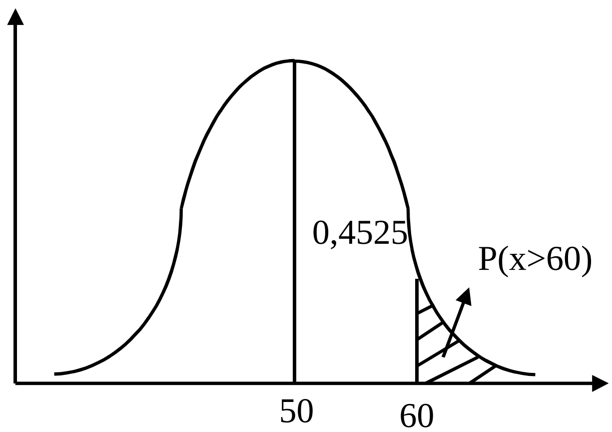
Exemplo:

1) Seja um teste de inteligência aplicado a um grupo de 50 adolescentes do 3º ano colegial. Obteve-se uma distribuição normal com média 50 e desvio padrão 6. Pergunta-se qual a proporção de alunos com notas superiores a 60 ?

Transformando a nota 60 em desvios reduzidos tem-se:

$$z = \frac{60-50}{6} = 1{,}67$$

Consultando a tabela verifica-se:



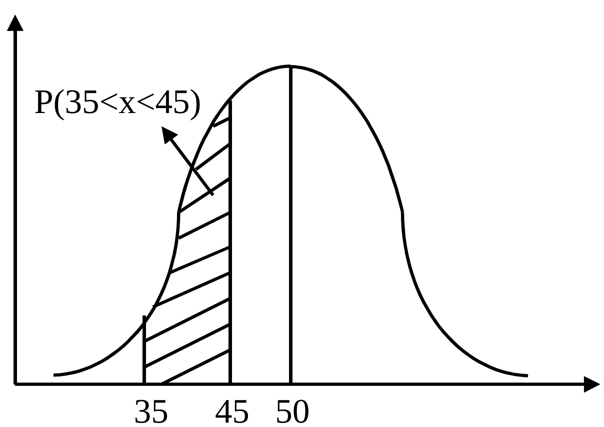
Probabilidade da nota ser superior a 60 é 0,5 - 0,4525 = 0,0475 ou 4,75 %

2) Com os dados do problema anterior, averiguar o número de alunos com notas entre 35 e 45.

Calculando os desvios reduzidos tem-se:

$$z_1 = \frac{45-50}{6} = -0{,}83$$
$$z_2 = \frac{35-50}{6} = -2{,}5$$

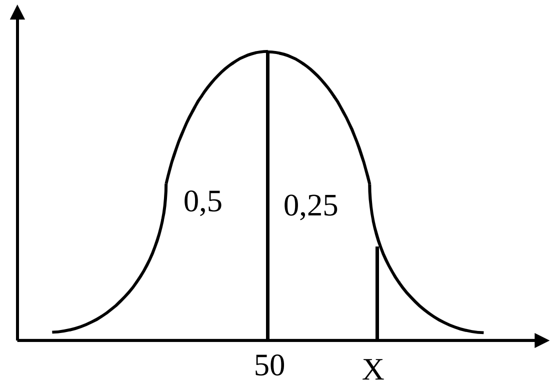
Consultando a tabela verifica-se:



Probabilidade (área) entre 0 e 2,5 = 0,4938
Probabilidade (área) entre 0 e 0,83 = 0,2967
Então Probabilidade (área) entre 2,5 e 0,83 = 0,4938 - 0,2967 = 0,1971
O número de alunos é 0,1971 x 50 = 9,855= 10 pessoas

3) Com os dados do problema anterior, qual é a nota abaixo da qual estão 75% dos alunos ?

Consultando a tabela, a área é de 0,5 + 0,25 = 0,75
O valor de z correspondente a área de 0,2486 é 0,67
O valor de z correspondente a área de 0,2518 é 0,68
Pode-se adotar um valor médio z = 0,675



$$0{,}675 = \frac{x - 50}{6} \Rightarrow x = 50 + 6 \cdot 0{,}675 = 54{,}05$$

4) Um teste padronizado de escolaridade tem distribuição normal com média 100 e desvio padrão 25. Determine a probabilidade de um indivíduo submetido ao teste ter nota:

a) maior que 120
b) entre 75 e 125
c) entre 115 e 125
d) qual é a nota abaixo da qual estão 70% dos alunos

Resp: a) p = 21,19 % b) p = 68,26% c) p = 11,55% d) 113

5) Os salários dos funcionários de uma escola têm distribuição normal com média de R$ 1500,00, e desvio padrão de R$ 200,00. Qual a proporção de funcionários que ganham:

a) entre R$ 1400 e R$ 1600 ?
b) acima de R$ 1500 ?
c) acima de R$ 1400 ?
d) abaixo de R$ 1400 ?
e) acima de R$ 1650 ?

Resp: a) p = 38,3 % b) p = 50% c) p = 69,15% d) p = 30,85% e) p = 22,66%

6) Determinar os valores de z simétricos em relação a origem, que entre si abrangem 95 % da área total.

Resp: z = 1,96

7) Determinar os valores de z simétricos em relação a origem, que entre si abrangem 99 % da área total.

Resp: z = 2,575

EXERCÍCIOS COMPLEMENTARES

Distribuição Normal

4) Seja um teste de inteligência aplicado a um grupo de 1000 alunos de uma escola superior. Obteve-se uma distribuição normal, com média de 32 e desvio padrão de 4. Pergunta-se.

a) Qual o número de alunos com notas superiores a 38 ?
b) Qual o número de alunos com notas inferiores a 35 ?
c) Qual o número de alunos com notas compreendidas entre 27 e 31 ?

Resposta: a) 67 b) 773 c) 296

5) A renda anual média de uma grande comunidade pode ser aproximada por uma distribuição normal com média de R$ 7.000,00 e desvio padrão de R$ 3.000,00. a) Que porcentagem da população terá renda superior a R$ 13.000,00 ? b) Abaixo de qual renda temos 15% da população?

Resposta: a) 2,28 % b) R$ 3880

6) Os resultados de um concurso de habilitação tiveram distribuição normal com média 50 e desvio padrão 10. Os candidatos serão classificados conforme o seguinte critério decrescente:
A - 10 % das notas B - 15 % das notas C - 50 % das notas D - 15 % das notas E - 10 % das notas. Determine as notas limites para a classificação dos candidatos.

Resposta: A-acima de 62,8 B-entre 56,7 e 62,8 C-entre 43,3 e 56,7 D-entre 37,2 e 43,3 F-abaixo de 37,2

Capítulo 10 - Testes de Hipóteses

Hipótese estatística é uma afirmação a respeito da distribuição de uma ou mais variáveis.

A prova ou o teste de uma hipótese estatística é uma regra que, obtidos os valores amostrais, conduz a uma decisão de aceitar ou rejeitar a hipótese considerada.

Erro do tipo 1 - considerar falsa uma hipótese verdadeira
Erro do tipo 2 - considerar verdadeira uma hipótese falsa

10.1 Teste da Média

Exemplo: Aplicou-se um teste de QI a um grupo de 2970 crianças de mesma idade. Obteve-se os seguintes resultados:

Média = 104 Desvio padrão = 17,03

Deseja-se, a partir desses dados, comprovar a hipótese de que a média da população de onde foi extraída a amostra acima seja igual a 100, ou seja, admitir que essas 2970 crianças não são mais inteligentes que a média; já que no teste em estudo o QI médio é igual a 100.

Hipótese - média igual a 100

Nível de significância de 5% - é o risco de rejeitar uma hipótese , que na realidade, é verdadeira.

Qual será o limite para aceitar a hipótese, admitindo um erro de 5%?

Aplicando a fórmula do erro padrão da média $s_{\overline{x}} = \dfrac{\sigma}{\sqrt{N}}$

No caso: $s_{\overline{x}} = \dfrac{17,03}{\sqrt{2970}} = 0,31$

Prova Unicaudal (a região de rejeição está situada em umas das caudas apenas da curva normal).

Consultando a tabela da curva normal, vemos que uma área de 5 % corresponde a um desvio reduzido de 1,65, então:

$z = \dfrac{104 - 100}{0,31} = 12,90$ que é maior que 1,65, caindo na zona de rejeição da igualdade.

Prova Bicaudal (a região de rejeição está situada em ambas as caudas da curva normal).

O desvio reduzido que corresponde a área de 0,5 - 0,025 = 0,475 é o 1,96

$z = \dfrac{104 - 100}{0,31} = 12,90$ que é maior que 1,96, caindo na zona de rejeição da igualdade.

Exercícios:

1) A média obtida através dos anos em um teste vocacional foi de 100 pontos. Com o objetivo de saber se a nova classe (calouros) é típica com respeito a vocação, tomou-se uma amostra de 50 alunos. O resultado foi uma média 95 com desvio padrão de 10. Pode-se afirmar, a um nível de significância de 5 %, que essa nova turma é igual às anteriores ?

Resposta: Não

2) - A média em dias de internação de crianças que sofreram acidente de trânsito e que não estavam usando o cinto de segurança é de 1,39 dias. Em um levantamento de 123 crianças que estavam usando o cinto, a média foi de 0,83 dias e desvio padrão de 0,16 dias. Podemos concluir que o uso do cinto diminui o tempo médio de internação?

Resposta: Sim

3) A fim de acelerar o tempo que um analgésico leva para surtir efeito, um químico analista acrescentou certo ingrediente à fórmula original, que acusava um tempo médio de 43 minutos para fazer efeito. Em 49 observações com a nova fórmula, obteve-se um tempo médio de 41 minutos, com desvio padrão de 10 minutos. A nova fórmula é melhor, pior ou igual a anterior ?
Adote α = 5 %.

Resposta: Igual

10.2 - Teste de Hipóteses acerca de proporções

Considerando que uma proporção é caso especial da média, a hipótese pode ser testada com o emprego de:

$$z = \frac{\hat{p} - P}{\sqrt{\frac{P \cdot Q}{n}}}$$

onde n é tamanho da amostra
$\hat{p}$ é a proporção da amostra (acerto)
P é proporção da população (acerto)
Q é proporção da população (erro)
P + Q = 1
Q = 1 - P

Pode-se ter teste unicaudal (unilateral) ou teste bicaudal (bilateral).

Exemplo 1 : O fabricante de uma droga medicinal afirma que ela é 90 % eficaz na cura de uma alergia, em um período de 8 horas. Em uma amostra de 200 pessoas que tinham alergia a droga curou 160 pessoas. Determine se a afirmação do fabricante é verdadeira ? Usar α = 5 %.

$\hat{p}$ = 160/200 = 0,8   P = 0,9   Q = 1 - 0,9 = 0,1   n = 200

Calculando o valor de z tem-se:

$$z = \frac{0{,}8 - 0{,}9}{\sqrt{\frac{0{,}9 \cdot 0{,}1}{200}}} = -4{,}71$$

para α = 5 %, teste unicaudal => z_0 = -1,645, rejeição

Resposta: ao nível de 5 %, unicaudal, rejeita-se a hipótese. A afirmação do fabricante não é verdadeira.

Exemplo 2: A experiência tem demonstrado que 60 % dos estudantes são aprovados num exame de inglês para admissão a uma universidade. Se 60 dos 110 estudantes de uma certa cidade fossem aprovados, pode-se concluir que estes estudantes são inferiores em inglês ? Usar α = 5 %.

$\hat{p}$ = 60/110 = 0,55   P = 0,6   Q = 1 - 0,6 = 0,4   n = 110

Calculando o valor de z tem-se:

$$z = \frac{0{,}55 - 0{,}6}{\sqrt{\frac{0{,}6 \cdot 0{,}4}{110}}} = -1{,}07$$

para α = 5 %, teste unicaudal => z_0 = -1,645, aceitação

Resposta: ao nível de 5 %, unicaudal, aceita-se a hipótese. Os estudantes dessa cidade são iguais aos demais.

Exercício 1: A proporção de aprovação ao final do ano nas escolas da periferia de São Paulo é de 86 %. Dos 200 alunos de uma escola localizada nessa região a proporção de aprovados foi de 92 %. Pode-se afirmar que os alunos dessa escola são melhores que os outros ? Usar α = 5 %.

Resposta: ao nível de 5 %, unicaudal, rejeita-se a hipótese. Os estudantes dessa escola são mesmo melhores que os demais.

Exercício 2: Uma amostra de 200 proprietários de carro de uma cidade mostrou que 48 deles tinham sido multados naquele ano. A média anual nacional é de 30 % dos motoristas são multados por ano. Pode-se afirmar que os motoristas dessa cidade são menos infratores que os demais ? Usar α = 5 %.

Resposta: ao nível de 5 %, unicaudal, rejeita-se a hipótese. Os motoristas dessa cidade são menos infratores que a maioria.

Exercício 3: Se você lançar um dado 240 vezes e obtiver 52 seis, concluirá que o dado favorece o número seis ? Usar α = 5 %.

Resposta: ao nível de 5 %, unicaudal, rejeita-se a hipótese. O dado favorece o seis.

10.3 - Teste de Hipóteses acerca de diferenças entre médias aritméticas

Aqui, trabalha-se com distribuições amostrais da diferença, obtendo z pela fórmula:

$$z = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Exemplo: Suponhamos um teste de inteligência aplicado a 318 meninos e 197 meninas de 13 anos de idade, obtendo-se os seguintes resultados:

$\overline{x}_1 = 38$

$\sigma_1 = 12$

$\overline{x}_2 = 36$ $\quad$ $n_1 = 318$ e $n_2 = 197$

$\sigma_2 = 13$

Hipótese: as médias são iguais. Teste bicaudal com nível de significância de 5 % (z = +- 1,96)

Calculando z:

$$z = \frac{(38 - 36)}{\sqrt{\frac{12^2}{318} + \frac{13^2}{197}}} = \frac{2}{\sqrt{1{,}3107}} = 1{,}75$$

Atinge-se a zona de aceitação, então, as médias são iguais, ou seja, os 318 meninos são iguais as 197 meninas.

Exercício 1: Examinaram duas classes constituídas de 40 e 50 alunos, respectivamente. Na primeira, a média foi 74 com desvio padrão 8. Enquanto que na Segunda a média foi 78 com desvio padrão 7. Há uma diferença significativa entre os aproveitamentos das duas classes no nível de significância de 5 %?

Resp: Atinge-se a zona de rejeição, então, as médias são mesmo diferentes, ou seja, a sala com 50 alunos obteve uma média maior que a de 40 alunos.

Exercício 2 : A altura média de 50 estudantes do sexo masculino que tiveram participação superior à média nas atividades atléticas colegiais era de 178,23 cm, com desvio padrão de 6,35 cm. Enquanto que os 50 que não mostraram nenhum interesse nessas atividades apresentaram a altura média de 175,45 cm, com desvio de 7,11 cm. Testar a hipótese dos estudantes do sexo masculino que participam de atividades atléticas serem mais altos que os demais. Adote α = 5 %

Resp: Atinge-se a zona de rejeição, então, as médias são mesmo diferentes, ou seja, os estudantes do sexo masculino que participam de atividades atléticas são mesmo mais altos que os demais.

Exercício 3 : No estudo de efeito de doses diárias de vitamina C sobre os resfriados registrou-se o número de resfriados contraídos por cada participante durante um certo período de tempo experimental, resultando no quadro abaixo:

	Nº de pessoas	Nº médio de resfriados	Desvio padrão do no de resfriados
Tomou vitamina C	407	1,38	1,23
Tomou vitamina C falsa	411	1,48	1,14

Pergunta-se: as doses diárias de vitamina C têm efeito sobre o no de resfriados contraídos? Adote α = 5 %.

Resp: Atinge-se a zona de aceitação, então, as médias são iguais, ou seja, as doses diárias de vitamina C não têm efeito sobre o nº de resfriados contraídos.

EXERCÍCIOS COMPLEMENTARES

Teste da Média

7) No exame de inglês TOEFL (Test of English as a Foreign Language) utilizado por alunos estrangeiros para ingressar nas faculdades americanas, verificou-se através dos anos anteriores que a média obtida foi de 450 pontos. Em 2002, 136 brasileiros realizaram o teste e obtiveram uma média de 465 pontos com desvio padrão de 80 pontos. Pode-se concluir, no nível de significância de 5% que os brasileiros possuem um melhor desempenho nesse teste ?

Resposta: Sim

8) Em uma pesquisa verificou-se que em média os estudantes universitários estudam em casa 8 horas por semana. Em uma faculdade, 36 alunos foram entrevistados e resultou em uma média de 7,5 horas de estudo semanal com um desvio padrão de 2 horas. Pode-se dizer, com um nível de significância de 5%, que esses alunos estudam menos que os demais?

Resposta: Não

Teste de Proporção

9) Em 1990, 5,8 % dos candidatos a emprego submetidos a um teste de drogas foram reprovados. Em 2000, 58 dos 1520 candidatos foram também reprovados no teste. Pode-se concluir que houve diminuição, aumento, ou ficou inalterado do uso de drogas ? $\alpha = 5\%$

Resposta: Diminuição

10) Em 2000, o IBGE observou que 9 % dos estudantes no Brasil que terminavam o segundo grau, ingressavam na faculdade. Em Lorena, 60 dos 500 secundaristas entrevistados se matricularam na faculdade. Pode-se concluir que Lorena está acima da média nacional em relação a porcentagem de estudantes ingressantes na faculdade ? $\alpha = 5\%$

Resposta: Sim, está.

Teste da diferença entre médias

11) Considere o quadro abaixo e responda: Pode-se concluir que as crianças nascidas em hospital particular são mais pesadas do que as crianças nascidas em hospital público? $\alpha = 5\%$

Pesos, em kg, de recém-nascidos, em dois hospitais diferentes			
Hospital	Tamanho da amostra	Média (kg)	Desvio padrão (kg)
Particular	50	3,1	1,6
Público	50	2,7	1,4

Resposta: Não

12) Considere o quadro abaixo e responda: Pode-se concluir que os estudantes de escolas públicas tiram notas inferiores aos da escola particular num exame de vestibular ? adote $\alpha = 5\%$

Nota no exame vestibular			
Escola	Tamanho da amostra	Média (kg)	Desvio padrão (kg)
Particular	100	80	16
Pública	100	70	20

Resposta: Sim.

Capítulo 11 - Distribuição Qui-Quadrado (χ^2)

11.1 Cálculo do Qui-Quadrado

Assim como a distribuição normal, a distribuição qui-quadrado pode também ser representada por uma equação.

Utiliza-se a χ^2 para medir a discrepância entre valores observados e os resultados teóricos de uma distribuição hipotética. Para tanto, foi demonstrado que:

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t}, \text{ onde}$$

f_e - freqüências efetivamente obtidas ou freqüências empíricas
f_t - freqüências teóricas

11.2 Aplicação: Provas de Independência.

Uma das aplicações mais usuais do χ^2 refere-se a "provas de independência" em que desejamos saber se duas variáveis estão relacionadas ou não. A hipótese que se testa é a da independência, ou seja, não possuem relação entre si. Neste caso deve-se determinar um nível de significância (risco de se rejeitar uma hipótese verdadeira). Usualmente utiliza-se o de 5 %.

Exemplo 1:
No estudo de um teste de aptidão artística, um dos itens consistia na escolha entre 3 desenhos geométricos de aspecto variado. Cada uma das pessoas deveria indicar sua preferência por um dos desenhos designados por A, B, e C. Foi apresentado para 60 pessoas, sendo que 30 escolheram A, 18 escolheram B e 12 escolheram o C. Deseja-se saber se essa escolha foi ditada pelo bom gosto, ou se foram feitas ao acaso.

Hipótese: a proporção de pessoas que escolhe cada desenho é a mesma (20, 20, 20).

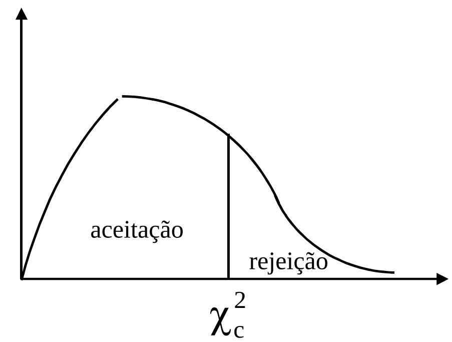
Desenho	f_e	f_t
A	30	20
B	18	20
C	12	20
Total	60	60

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t} = \frac{(30-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(12-20)^2}{20} = 5 + 0{,}2 + 3{,}2 = 8{,}4$$

υ = 3-1 = 2

Consulta-se a tabela para χ^2 0,95 (5% de significância, 1 - 0,05 = 0,95) e υ = 2, determinando um qui-quadrado crítico $\chi^2_C = 5{,}99$

Conclusão: Como o χ^2 calculado é maior que o crítico 8,4 > 5,99, rejeita-se a hipótese de independência, ou seja, a escolha dos desenhos foi pelo bom gosto.



Exercício 1: Influem as cores no sabor ? Apresentou-se a 100 pessoas quatro garrafas de suco de laranja, de diferentes cores, pedindo que indicassem a de suco mais ácido. Os resultados obtidos foram: (adote 5% de significância)

Cor do suco	f_e	f_t
Amarelo Claro	32	
Amarelo Vivo	22	
Laranja Claro	13	
Laranja Forte	13	
Indiferente	20	
Total	100	

11.3 - Tabelas de Dupla Entrada ou Maiores.

No caso de tabelas maiores que as anteriormente vistas, pode-se aplicar uma regra prática para determinar o grau de liberdade do problema.

O grau de liberdade υ é determinado por: $\upsilon = (L - 1) \times (C - 1)$, onde:
L é o número de linhas da tabela e C é o número de colunas.

No cálculo das freqüências teóricas deve-se observar que essas devem ser proporcionais aos seus totais. Portanto:

$$f_T = \frac{(\text{Total da coluna}) \times (\text{Total da linha})}{(\text{Total geral})}$$

Exemplo 1: Freqüências observadas num estudo de permissividade relacionada com orientação política gerou os seguintes resultados:

Método de educação das crianças	Orientação Política		Total
	Liberal	Conservador	
Permissivo	5 (7,5)	10 (7,5)	15
Não Permissivo	15 (12,5)	10 (12,5)	25
Total	20	20	40

Existe relação entre a orientação política e a permissividade na educação das crianças ?

Lembre-se que a hipótese a ser testada é a da independência.

Cálculo das freqüências teóricas: $f_{T1} = 15 \times 20 / 40 = 7{,}5$ $f_{T2} = 25 \times 20 / 40 = 12{,}5$

$\upsilon = (2 - 1) \times (2 - 1) = 1$

Consulta-se a tabela para $\chi^2\, 0{,}95$ (5% de significância, 1 - 0,05 = 0,95) e $\upsilon = 1$, determinando um qui-quadrado crítico $\chi^2_C = 3{,}84$

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t} = \frac{(5-7{,}5)^2}{7{,}5} + \frac{(10-7{,}5)^2}{7{,}5} + \frac{(15-12{,}5)^2}{12{,}5} + \frac{(10-12{,}5)^2}{12{,}5} =$$
$$0{,}83 + 0{,}83 + 0{,}5 + 0{,}5 = 2{,}66$$

Conclusão: Como o χ^2 calculado é menor que o crítico 2,66 < 3,84, aceita-se a hipótese de independência, ou seja, de acordo com os resultados obtidos, a orientação política não está relacionada com a permissividade. Ou seja, não podemos afirmar que os liberais são menos permissivos na educação que os conservadores, apesar de os dados nos persuadirem a chegar a essa conclusão.

Exercício 1: De uma amostra de 36 alunos do 2º grau, perguntou-se sobre o objetivo de prosseguir os estudos (cursar faculdade) ou não. O resultado foi de que 21 prosseguem e 15 outros não. Em seguida, foi perguntado se fumavam, resultando nos dados da tabela abaixo:

Fuma	Vai Cursar a Faculdade ?		Total
	Sim	Não	
Sim	15 ( )	5 ( )	
Não	6 ( )	10 ( )	
Total			

Existe relação entre a fumar e o desejo de prosseguir nos estudos ?

Lembre-se que a hipótese a ser testada é a da independência.

Exercício 2 – Entre 270 empregados de uma indústria, foi feita uma pesquisa para saber se o ajustamento à função apresentava alguma relação com o nível educacional. O resultado obtido aparece na tabela abaixo. Verifique, ao nível de significância de 5% se existe relação entre as duas variáveis, ou seja, se as variáveis são independentes ou não.
Ho - as variáveis são independentes

	Ajustamento	Desajustamento	Total
Nível Primário	45 ( )	85 ( )	
Nível Secundário	59 ( )	61 ( )	
Nível Superior	16 ( )	4 ( )	
Total			

Lembre-se que a hipótese a ser testada é a da independência.

11.4 - Uso do χ^2 em amostras muito pequenas. A correção de Yates

Aplica-se a problemas de grau de liberdade 1 e quando as freqüências teóricas forem menores que 10 unidades.

A correção de Yates tem como objetivo diminuir o tamanho do χ^2 esperado. Reduz-se então meia unidade (0,5) todas as diferenças entre as frequencias observadas (empíricas) e as teóricas. A fórmula "corrigida " é a seguinte:

$$\chi^2 = \sum \frac{(|f_e - f_t| - 0,5)^2}{f_t}$$

Exemplo 1: Aplique a correção de Yates: De uma amostra de 36 alunos do 2º grau, perguntou-se sobre o objetivo de prosseguir os estudos (cursar faculdade) ou não. O resultado foi de que 21 prosseguem e 15 outros não. Em seguida, foi perguntado se fumavam, resultando nos dados da tabela abaixo:

Fuma	Vai Cursar a Faculdade ?		Total
	Sim	Não	
Sim	15 (11,67)	5 (8,33)	20
Não	6 (9,33)	10 (6,67)	16
Total	21	15	36

Existe relação entre a fumar e o desejo de prosseguir nos estudos ?

Lembre-se que a hipótese a ser testada é a da independência.

Cálculo das freqüências teóricas: f_{T1} = 20 x 21 / 36 = 11,67 f_{T2} = 20 x 15 / 36 = 8,33
f_{T3} = 16 x 21 / 36 = 9,33 f_{T4} = 16 x 15 / 36 = 6,67

υ = (2 - 1) x (2 - 1) = 1

Consulta-se a tabela para χ^2 0,95 (5% de significância, 1 - 0,05 = 0,95) e υ = 1, determinando um qui-quadrado crítico $\chi^2_C = 3,84$

$$\chi^2 = \sum \frac{\left(|f_e - f_t| - 0{,}5\right)^2}{f_t} = \frac{\left(|15 - 11{,}67| - 0{,}5\right)^2}{11{,}67} + \frac{\left(|5 - 8{,}33| - 0{,}5\right)^2}{8{,}33} +$$

$$+ \frac{\left(|6 - 9{,}33| - 0{,}5\right)^2}{9{,}33} + \frac{\left(|10 - 6{,}67| - 0{,}5\right)^2}{6{,}67} = 0{,}69 + 0{,}96 + 0{,}86 + 1{,}20 = 3{,}71$$

Conclusão: Como o χ^2 calculado é menor que o crítico 3,71 < 3,84, aceita-se a hipótese de independência, rejeitada anteriormente sem a correção de Yates, ou seja, de acordo com os resultados obtidos, não existe relação entre fumar e o desejo de prosseguir nos estudos. Portanto, não podemos afirmar que há uma maior incidência de fumantes no grupo que deseja continuar os estudos do que no outro grupo.

Exercício 1 – Aplicando a correção de Yates, realize uma prova de qui-quadrado para o seguinte problema 2 x 2, ao nível de significância de 5%, ou seja, verifique se as variáveis são independentes ou não :

	Cabelos Escuros	Cabelos Claros	Totais
Olhos Escuros	20 ( )	14 ( )	
Olhos Claros	5 ( )	10 ( )	
Totais			

Ho - as variáveis são independentes

Capítulo 12 - Correlação

12.1 Conceito

A correlação expressa a relação entre duas ou mais variáveis. Se duas ou mais variáveis variam concumitantemente, diz-se que estão correlacionadas.

Exemplo: A estatura de uma pessoa e o seu peso. Para uma estatura maior corresponde, em geral, a um peso maior. Dizemos, por isso, que entre as variáveis peso e estatura existe correlação.

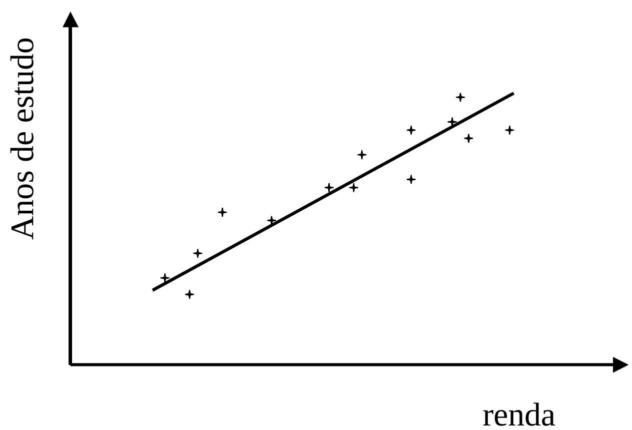
12.2 Correlação Positiva, Negativa e Curvilínea

a) **Correlação positiva**: valores elevados de uma variável corresponde a valores elevados da outra. Exemplo peso e altura
b) **Correlação negativa**: valores elevados de uma variável corresponde a valores baixos da outra e vice-versa. Exemplo: reprovações e nível de escolaridade.
c) **Correlação curvilínea**: começa negativa e termina positiva ou vice-versa. Exemplo: tamanho da família e situação sócio econômica.

12.3 Representação Gráfica

As correlações variam com respeito a sua força. Podemos visualizar essa força num diagrama de dispersão que é um gráfico capaz de mostrar a maneira pela qual os valores de duas variáveis, X e Y, distribuem-se ao longo da faixa dos possíveis resultados.

Exemplo: Renda x Anos de estudo



A força da correlação entre X e Y aumenta a medida que os pontos se agrupam em torno de uma linha reta imaginária.

12.4 Coeficiente de Correlação

Expressa numericamente a força e o sentido da correlação. Os coeficientes oscilam entre -1 e 1
C = -1 -> correlação negativa perfeita
-1 < C < - 0,6 -> correlação negativa forte
-0,6 < C < - 0,3 -> correlação negativa moderada
-0,3 < C < 0,0 -> correlação negativa fraca
0,0 < C < 0,3 -> correlação positiva fraca
0,3 < C < 0,6 -> correlação positiva moderada
0,6 < C < 1 -> correlação positiva forte
C = 1 -> correlação positiva perfeita

12.5 Coeficiente de Correlação para dados nominais dispostos numa tabela 2 x 2.

Coeficiente ϕ (fi)

$\phi = \sqrt{\frac{\chi^2}{N}}$, onde χ^2 é o Qui-quadrado calculado e N é o tamanho da amostra

Vamos verificar o exemplo anterior onde comparou-se o objetivo de prosseguir nos estudos e o hábito de fumar.

Fuma	Vai Cursar a Faculdade ?		Total
	Sim	Não	
Sim	15 (11,67)	5 (8,33)	20
Não	6 (9,33)	10 (6,67)	16
Total	21	15	36

$\upsilon = (2 - 1) \times (2 - 1) = 1$ e $\chi^2_C = 3{,}84$ $\chi^2 = 5{,}13$

Pode-se então calcular o **coeficiente ϕ (fi)** que é uma medida capaz de calcular o grau de associação em tabelas 2 x 2.

No exemplo: $\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{5{,}13}{36}} = 0{,}38$ indicando uma **correlação moderada** entre prosseguir os estudos e o hábito de fumar**.**

12.6 Coeficiente de Correlação para dados nominais dispostos numa tabela de ordem superior a 2 x 2.

Coeficiente de Contingência C.

$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$, onde χ^2 é o Qui-quadrado calculado e N é o tamanho da amostra

Vamos verificar o exemplo anterior de uma tabela 3 x 3 utilizada na comparação de vários grupos em que se testou a independência das variáveis nível educacional e o ajustamento à função. Nesse caso determinou-se:

	Ajustamento	Desajustamento	Total
Nível Primário	45 (57,78)	85 (72,22)	130
Nível Secundário	59 (53,33)	61 (66,67)	120
Nível Superior	16 (8,89)	4 (11,11)	20
Total	120	150	270

$\upsilon = (3 - 1) \times (2 - 1) = 2$ $\chi^2_C = 5{,}99$ $\chi^2 = 16{,}41$

Com isso rejeitou-se a hipótese de independência, e portanto, o coeficiente de de contingência C pode ser determinado.

$C = \sqrt{\dfrac{\chi^2}{\chi^2 + N}} = \sqrt{\dfrac{16{,}41}{16{,}41 + 270}} = 0{,}24$ indicando uma **correlação fraca** entre nível educacional e o ajustamento à função**.**

12.7 V de Cramér. Uma alternativa para o Coeficiente de Contingência C.

Alguns estatísticos utilizam o valor V de Cramér ao invés do C. O V de Cramér é definido por:

$$V = \sqrt{\frac{\chi^2}{N \cdot (k - 1)}}$$

onde: χ^2 é o Qui-quadrado calculado; N é o tamanho da amostra e k é o número de linhas ou colunas (usar o menor).

Para o exemplo anterior: $V = \sqrt{\dfrac{\chi^2}{N \cdot (k-1)}} = \sqrt{\dfrac{16{,}41}{270 \cdot (2-1)}} = 0{,}246$ indicando também uma correlação fraca entre nível educacional e o ajustamento à função**.**

Exercícios:

1) Dado o quadro determine o **coeficiente** ϕ **(fi)**

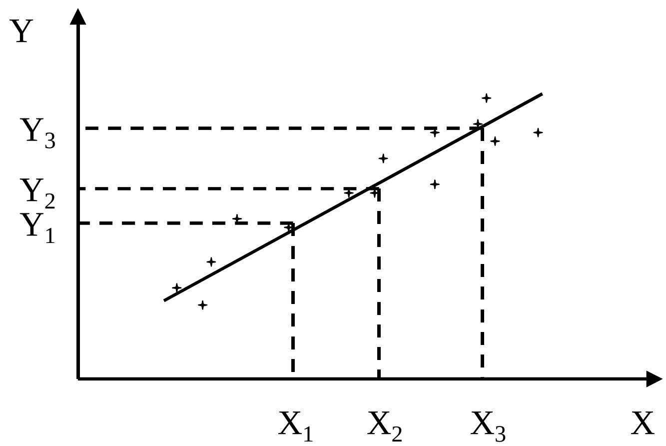
Assistiu às aulas	Passaram no Exame		Total
	Sim	Não	
Sim	22 ( )	8 ( )	
Não	10 ( )	18 ( )	
Total			

2) Dado o problema calcule C e V

Região	Candidato			Total
	A	B	C	
Sul	20 ( )	17 ( )	5 ( )	
Centro	15 ( )	16 ( )	16 ( )	
Norte	4 ( )	14 ( )	18 ( )	
Total				

12.8 Relação entre duas variáveis quantitativas.

Se retirarmos de uma população, uma amostra casual de tamanho N, teremos para cada elemento da amostra um par de observações: um valor de X e um valor de Y. Esses pares determinam N pontos no plano que podem ser representados graficamente num sistema de eixos cartesianos.



Ao gráfico acima dá-se o nome de diagrama de dispersão, esses nos fornece uma idéia intuitiva da eventual relação entre as duas variáveis.

Pode-se medir essa correlação através do Coeficiente de Correlação Linear de Pearson (r)

$$r = \frac{\left[n \cdot \sum (x_i \cdot y_i)\right] - \left[\left(\sum x_i\right) \cdot \left(\sum y_i\right)\right]}{\sqrt{[n \cdot \sum x_i^2 - \left(\sum x_i\right)^2] \cdot [n \cdot \sum y_i^2 - \left(\sum y_i\right)^2]}}$$

onde $-1 \leq r \leq 1$

Exemplo: Vamos comparar a correlação das notas de matemática com as de estatística de uma amostra aleatória de 10 alunos de uma classe:

Nº	Notas		$X_i \cdot Y_i$	X_i^2	Y_i^2
	Matemática (X_i)	Estatística (Y_i)			
1	5	6	30	25	36
2	8	9	72	64	81
3	7	8	56	49	64
4	10	10	100	100	100
5	6	5	30	36	25
6	7	7	49	49	49
7	9	8	72	81	64
8	3	4	12	9	16
9	8	6	48	64	36
10	2	2	4	4	4
Total	65	65	473	481	475

Logo:

$$r = \frac{\left[n \cdot \sum (x_i \cdot y_i)\right] - \left[\left(\sum x_i\right) \cdot \left(\sum y_i\right)\right]}{\sqrt{[n \cdot \sum x_i^2 - \left(\sum x_i\right)^2] \cdot [n \cdot \sum y_i^2 - \left(\sum y_i\right)^2]}} =$$

$$= \frac{\left[10 \cdot 473\right] - \left[65 \cdot 65\right]}{\sqrt{[10 \cdot 481 - 65^2] \cdot [10 \cdot 475 - 65^2]}} = \frac{4730 - 4225}{\sqrt{[4810 - 4225] \cdot [4750 - 4225]}} =$$

$$= \frac{505}{\sqrt{585 \cdot 525}} = \frac{505}{554{,}189} = 0{,}911$$

Correlação Forte

Exercício 1: Relação entre nível educacional do respondente e do respectivo pai, medidos em anos de freqüência à escola.

Criança	Anos de Escola		$X_i \cdot Y_i$	X_i^2	Y_i^2
	Pais (X_i)	Filhos (Y_i)			
A	12	12			
B	10	8			
C	6	6			
D	16	11			
E	8	10			
F	9	8			
G	12	11			
Total =					

Exercício 2: Uma agência estudou a demanda de matrículas em relação ao desconto promocional dado e obteve os seguintes valores:

	Demanda de Matrículas (X_i)	Desconto Promocional % (Y_i)	$X_i \cdot Y_i$	X_i^2	Y_i^2
1	6	3			
2	17	5			
3	27	8			
4	20	13			
5	45	16			
6	28	17			
7	34	20			
8	53	22			
Total =					

Determine o grau de correlação.

Resp: r = 0,846 correlação positiva forte

EXERCÍCIOS COMPLEMENTARES

Distribuição Qui-Quadrado

1) – Dado o problema 3 x 3 abaixo tabulado, verifique, ao nível de significância de 5% se existe relação entre as duas variáveis, ou seja, se as variáveis são independentes ou não. Caso exista relação, calcule o coeficiente de correlação.

	Consumo de Álcool			
Consumo de Drogas	Alto	Moderado	Baixo	Totais
Alto	5 ( )	7 ( )	20 ( )	
Médio	10 ( )	8 ( )	15 ( )	
Baixo	15 ( )	6 ( )	10 ( )	
Total				

2) – Será que o trabalho interfere realmente na produção escolar? Foi feito um levantamento entre 1100 alunos, dos quais 640 trabalham e 460 não trabalham, e perguntados se já haviam ou não repetido alguma matéria. Os resultados estão tabulados abaixo. Verifique, ao nível de significância de 5%, se existe relação entre as duas variáveis, ou seja, se as variáveis são independentes ou não. Caso exista relação, calcule o coeficiente de correlação.

	Já repetiram	Nunca repetiram	Total
Trabalham	430 ( )	210 ( )	
Não Trabalham	120 ( )	340 ( )	
Total			

3) Apresentou-se a seguinte distribuição de preferência de cia aérea nos vôos entre São Paulo Rio na ponte aérea. Deseja-se saber se essa escolha foi ditada pela preferência, ou se foram feitas ao acaso. Adote um nível de significância de 5 %.

Cia Aérea	f_e	f_t
Gol	21	
Tam	17	
Varig	14	
BRA	8	
Indiferente	20	
Total		

Correlação

4) Os dados abaixo relacionam os anos de estudo (x_i) e a renda mensal em mil reais (y_i) de 6 pessoas sorteadas ao acaso.

No	Anos de escola (x_i)	Renda mensal (y_i) x R$ 1000	($x_i \cdot y_i$)	$(x_i)^2$	$(y_i)^2$
1	20	3,0			
2	5	1,0			
3	10	2,0			
4	15	2,5			
5	7	0,7			
6	3	0,5			
Totais					

Determine o Coeficiente de Correlação de Pearson e o grau e correlação das duas variáveis.

5) Os dados abaixo relacionam horas de estudo (x_i) e a nota no exame (y_i) de 5 pessoas sorteadas ao acaso.

No	Horas de estudo (x_i)	Nota no exame (y_i)	($x_i \cdot y_i$)	$(x_i)^2$	$(y_i)^2$
1	12	10			
2	10	9			
3	9	9			
4	8	7			
5	5	6			
Totais					

Determine o Coeficiente de Correlação de Pearson (r), e classifique o grau e correlação das duas variáveis.

Respostas:1) $X^2 = 8{,}68$. Não tem relação. 2) $X^2 = 180{,}82$ $\phi=0{,}41$ moderada 3) $X^2 = 6{,}87$ não tem relação 4) $r = 0{,}96$ $y = 0{,}1534\,x + 0{,}083$ 5) $r = 0{,}95$ $y = 0{,}6045\,x + 2{,}8806$

Apêndice:

<u>Tamanho da Amostra para populações finitas</u>

$$n = \frac{z^2 \cdot (x/n) \cdot [1-(x/n)] \cdot N}{(N-1) \cdot e^2 + z^2 \cdot (x/n) \cdot [1-(x/n)]}$$

n = tamanho da amostra
N = tamanho da população
e = % de erro na forma unitária
z = intervalo de confiança, 1,96 para 95% de confiança (valor usual)
2,58 para 99% de confiança.
x/n = proporção esperada. O valor de n é máximo para x/n = 0,50

Resultando em:

$$n = \frac{1{,}96^2 \cdot 0{,}50 \cdot [1-0{,}50] \cdot N}{(N-1) \cdot e^2 + 1{,}96^2 \cdot 0{,}50 \cdot [1-0{,}50]} =$$

$$n = \frac{0{,}9604 \cdot N}{(N-1) \cdot e^2 + 0{,}9604}$$

Exemplo:

erro 2%	0,02
z=	1,96
x/n =	0,5

População	Amostra
100	96
200	185
300	267
400	343
500	414
600	480
700	542
800	600
900	655
1000	706
1100	755
1200	800
1300	844
1400	885
1500	923
1600	960
1700	996
1800	1029
1900	1061
2000	1091

População	Amostra
10000	1936
20000	2144
30000	2223
40000	2265
50000	2291
60000	2309
70000	2321
80000	2331
90000	2339
100000	2345

População	Amostra
100000	2345
200000	2373
300000	2382
400000	2387
500000	2390
600000	2391
700000	2393
800000	2394
900000	2395
1000000	2395

População	Amostra
1000000	2395
2000000	2398
3000000	2399
4000000	2400
5000000	2400
6000000	2400
7000000	2400
8000000	2400
9000000	2400
10000000	2400
115000000	2401

Cálculo do erro

$$e = z \cdot \sqrt{\frac{(x/n) \cdot [1 - (x/n)]}{n}}$$ para população desconhecida

$$e = z \cdot \sqrt{\frac{(x/n) \cdot [1 - (x/n)]}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$ para população conhecida

para z = 1,96 e x/n = 0,50 tem-se:

$$e = 0{,}98 \cdot \sqrt{\frac{1}{n}}$$ para população desconhecida

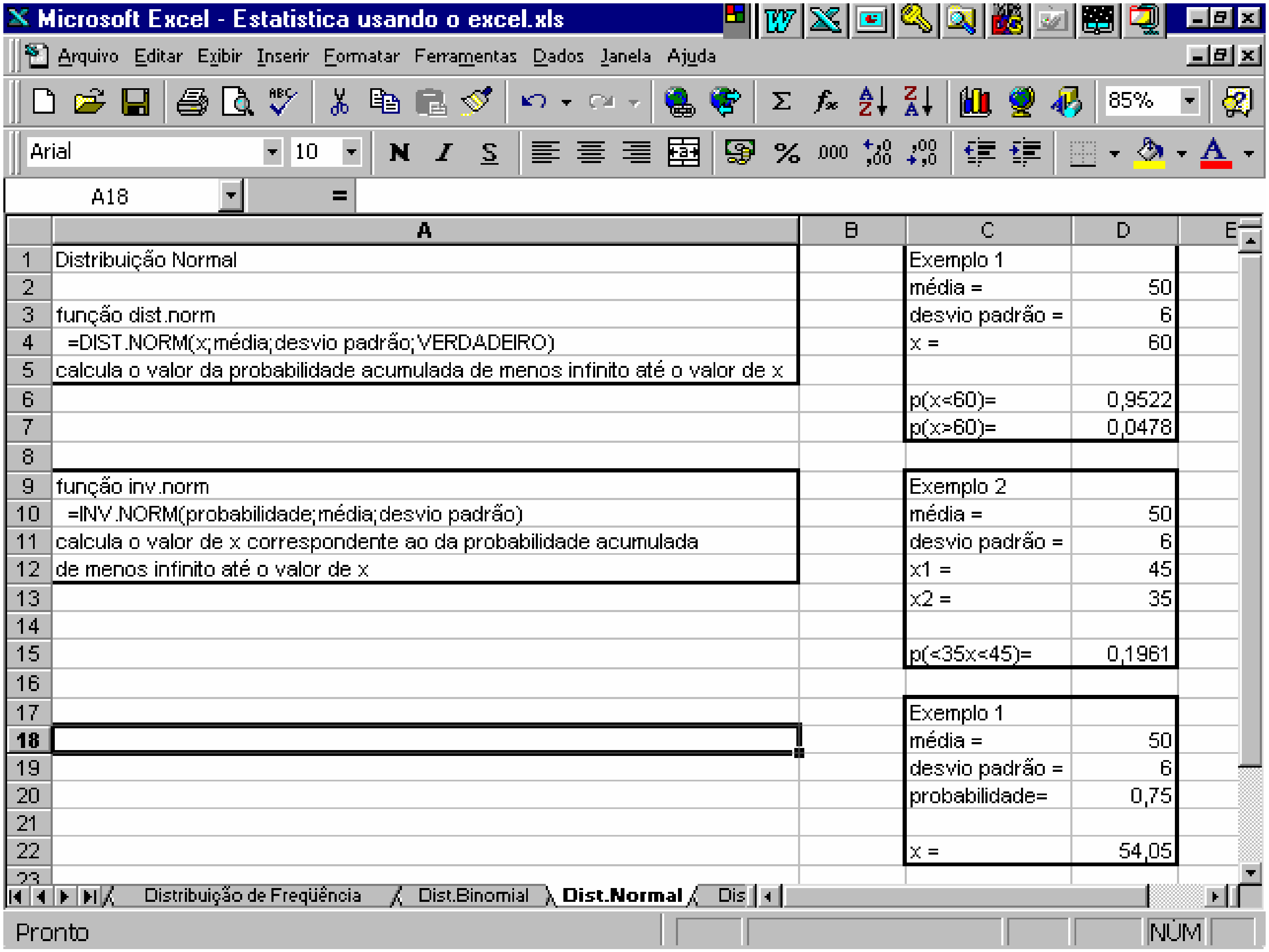
$$e = 0{,}98 \cdot \sqrt{\frac{N - n}{n \cdot (N - 1)}}$$ para população conhecida

População =	100
Amostra	Erro
10	0,30
20	0,20
30	0,15
40	0,12
50	0,10
60	0,08
70	0,06
80	0,05
90	0,03
100	0,00

Bibliografia

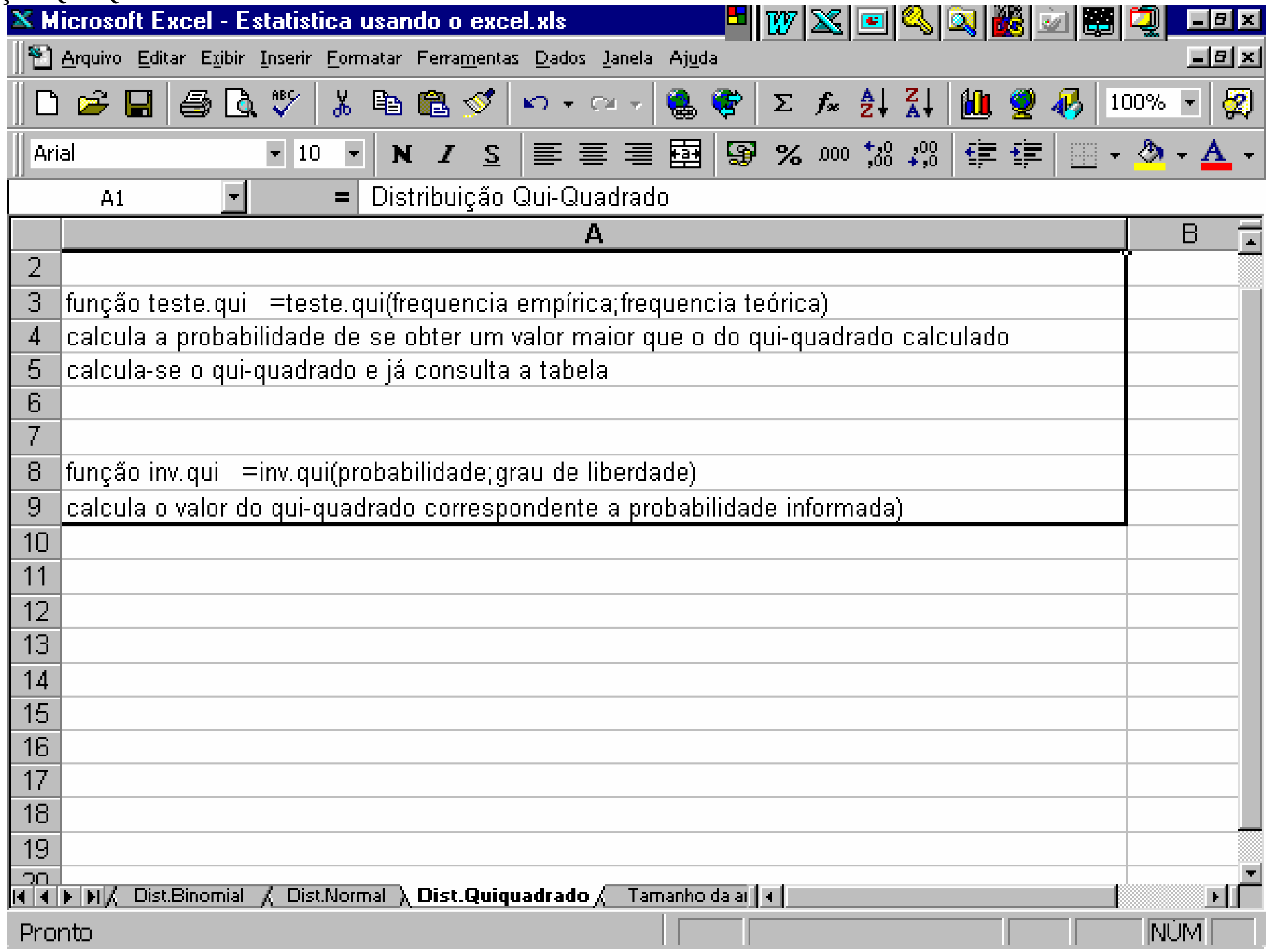
STEVENSON, W. J. Estatística Aplicada à Administração. São Paulo: Editora HARBRA Ltda, 1981.

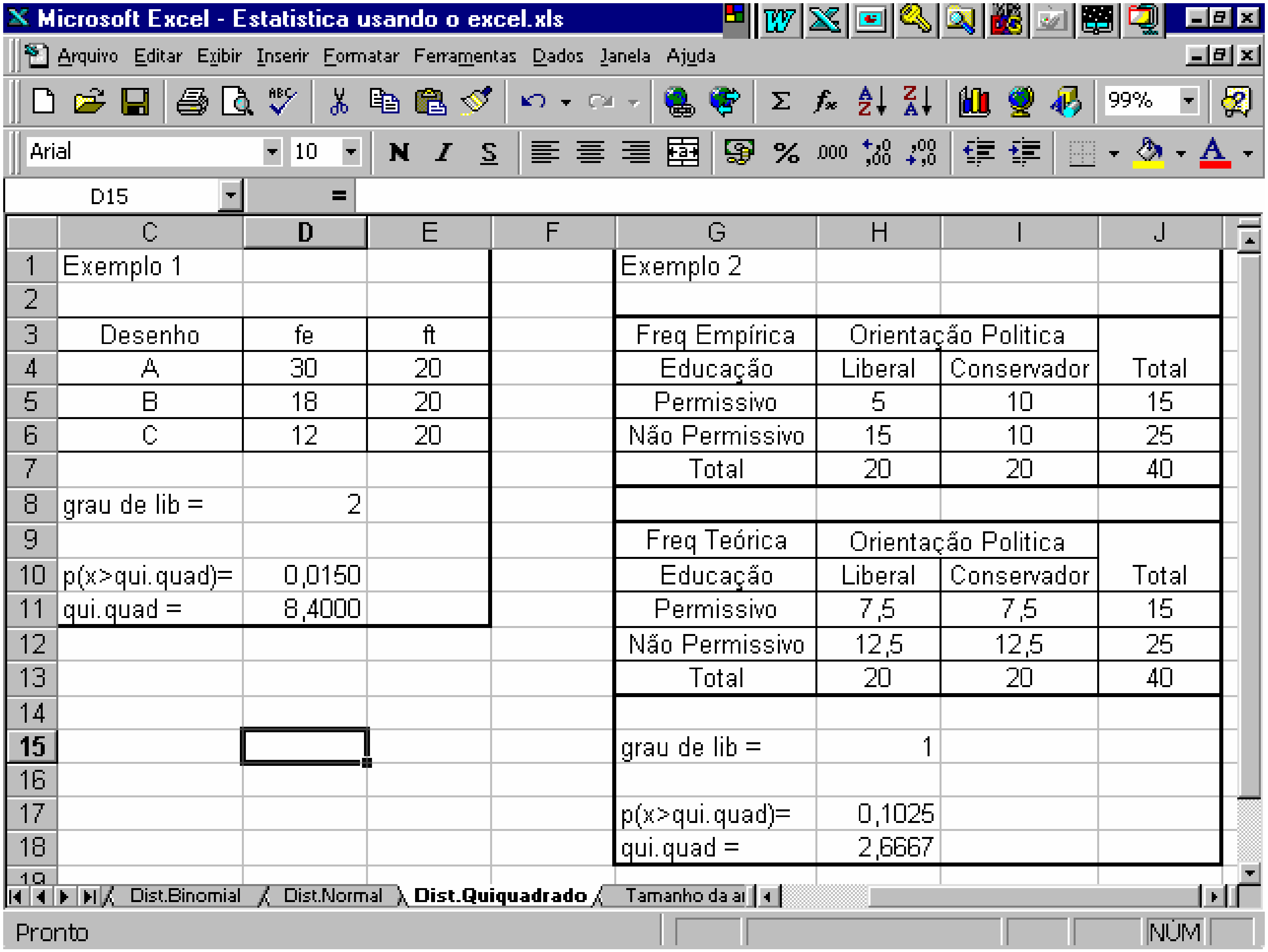
Distribuição Normal



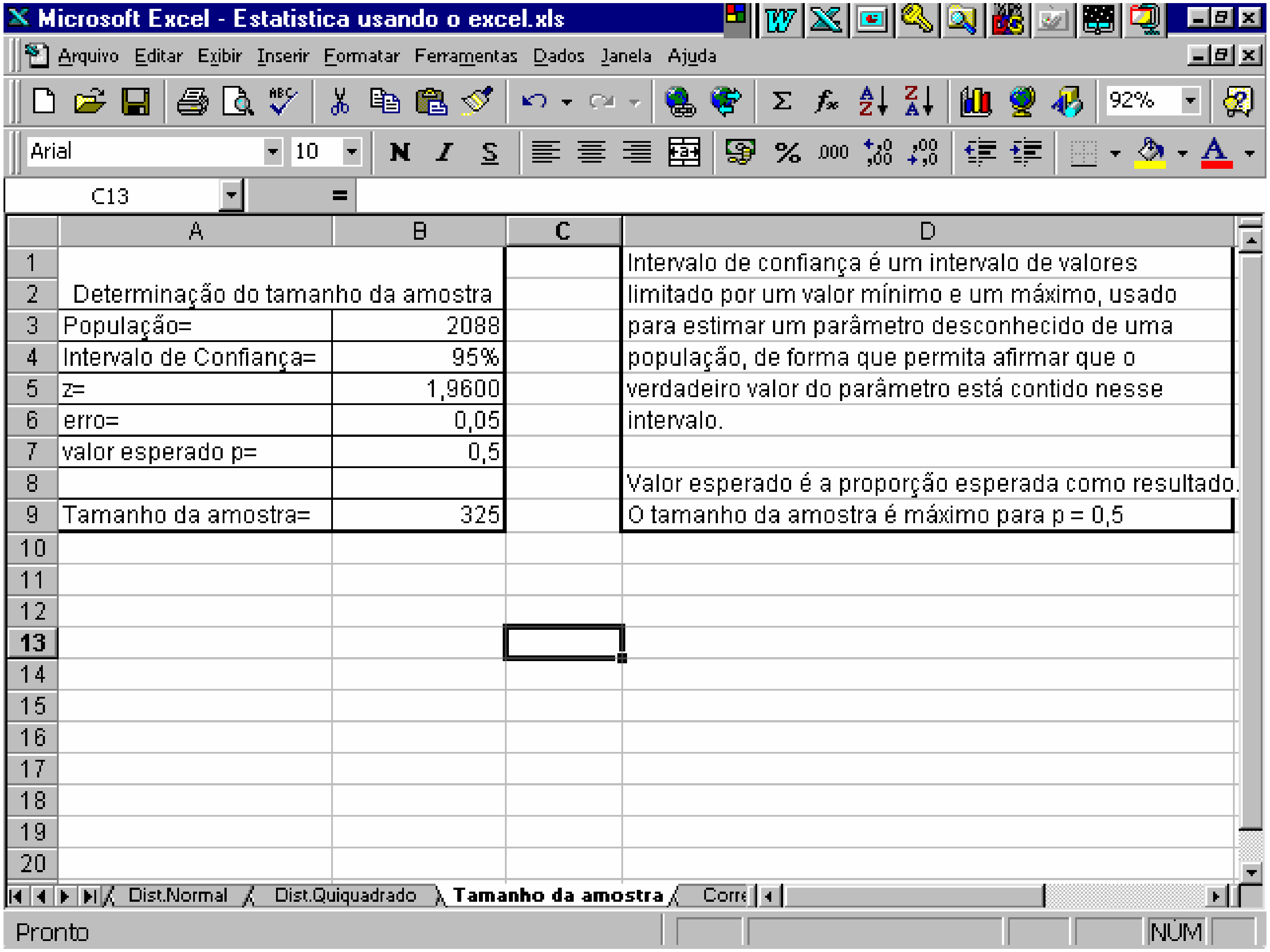
	A	B	C	D
1	Distribuição Normal		Exemplo 1	
2			média =	50
3	função dist.norm		desvio padrão =	6
4	=DIST.NORM(x;média;desvio padrão;VERDADEIRO)		x =	60
5	calcula o valor da probabilidade acumulada de menos infinito até o valor de x			
6			p(x<60)=	0,9522
7			p(x>60)=	0,0478
8				
9	função inv.norm		Exemplo 2	
10	=INV.NORM(probabilidade;média;desvio padrão)		média =	50
11	calcula o valor de x correspondente ao da probabilidade acumulada		desvio padrão =	6
12	de menos infinito até o valor de x		x1 =	45
13			x2 =	35
14				
15			p(<35x<45)=	0,1961
16				
17			Exemplo 1	
18			média =	50
19			desvio padrão =	6
20			probabilidade=	0,75
21				
22			x =	54,05

Distribuição Qui-Quadrado

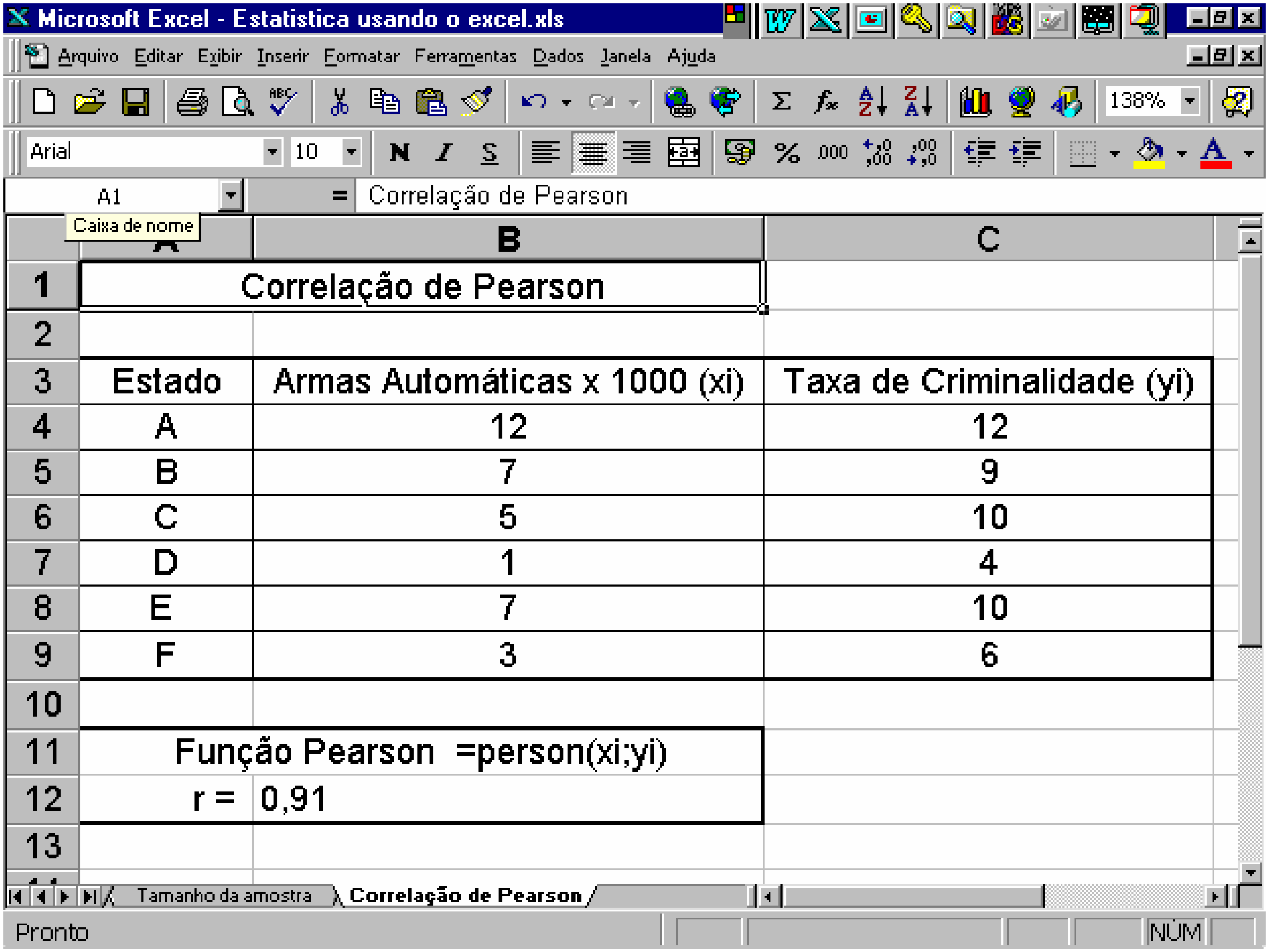



Microsoft Excel - Estatistica usando o excel.xls
Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda
D15
Exemplo 1
Desenho fe ft
A 30 20
B 18 20
C 12 20
grau de lib = 2
p(x>qui.quad)= 0,0150
qui.quad = 8,4000
Exemplo 2
Freq Empírica Orientação Politica
Educação Liberal Conservador Total
Permissivo 5 10 15
Não Permissivo 15 10 25
Total 20 20 40
Freq Teórica Orientação Politica
Educação Liberal Conservador Total
Permissivo 7,5 7,5 15
Não Permissivo 12,5 12,5 25
Total 20 20 40
grau de lib = 1
p(x>qui.quad)= 0,1025
qui.quad = 2,6667
Dist.Binomial Dist.Normal Dist.Quiquadrado Tamanho da a
Pronto
NUM

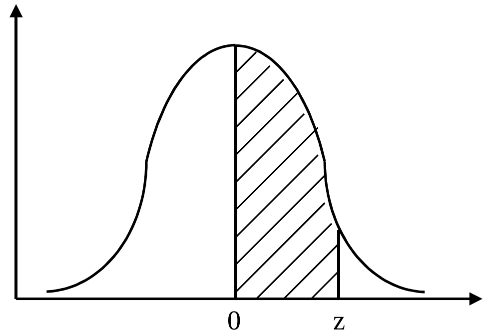
Tamanho da Amostra



Correlação

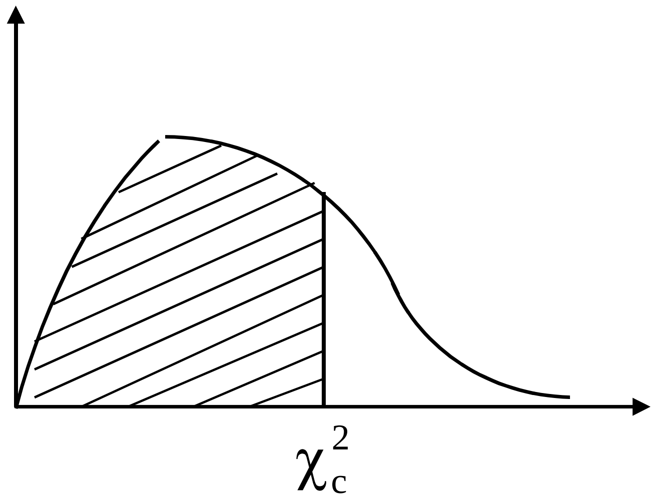

Microsoft Excel - Estatistica usando o excel.xls
Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda
138%
Arial 10
A1 = Correlação de Pearson
Caixa de nome
Correlação de Pearson
Estado
Armas Automáticas x 1000 (xi)
Taxa de Criminalidade (yi)
A 12 12
B 7 9
C 5 10
D 1 4
E 7 10
F 3 6
Função Pearson =person(xi;yi)
r = 0,91
Tamanho da amostra
Correlação de Pearson
Pronto
NÚM

ÁREA SUBTENDIDA PELA CURVA NORMAL REDUZIDA DE 0 A Z



z	0	1	2	3	4	5	6	7	8	9
0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

DISTRIBUIÇÃO QUI-QUADRADO



ν	0,5%	1,0%	2,5%	5,0%	10,0%	25,0%	50,0%
1	7,88	6,63	5,02	3,84	2,71	1,323	0,455
2	10,60	9,21	7,38	5,99	4,61	2,773	1,386
3	12,84	11,34	9,35	7,81	6,25	4,108	2,366
4	14,86	13,28	11,14	9,49	7,78	5,385	3,357
5	16,75	15,09	12,83	11,07	9,24	6,626	4,351
6	18,55	16,81	14,45	12,59	10,64	7,841	5,348
7	20,28	18,48	16,01	14,07	12,02	9,037	6,346
8	21,95	20,09	17,53	15,51	13,36	10,219	7,344
9	23,59	21,67	19,02	16,92	14,68	11,389	8,343
10	25,19	23,21	20,48	18,31	15,99	12,549	9,342
11	26,76	24,73	21,92	19,68	17,28	13,701	10,341
12	28,30	26,22	23,34	21,03	18,55	14,845	11,340
13	29,82	27,69	24,74	22,36	19,81	15,984	12,340
14	31,32	29,14	26,12	23,68	21,06	17,117	13,339
15	32,80	30,58	27,49	25,00	22,31	18,245	14,339
16	34,27	32,00	28,85	26,30	23,54	19,369	15,338
17	35,72	33,41	30,19	27,59	24,77	20,489	16,338
18	37,16	34,81	31,53	28,87	25,99	21,605	17,338
19	38,58	36,19	32,85	30,14	27,20	22,718	18,338
20	40,00	37,57	34,17	31,41	28,41	23,828	19,337

BIBLIOGRAFIA:

COSTA NETO, P. L. de O. Estatística. São Paulo: Editora Edgard Blucher Ltda, 17º ed. 1999.

CRESPO, A. A. Estatística Fácil. São Paulo: Editora Saraiva, 17º ed. 1999.

DOWNING, D. , CLARK, J. Estatística Aplicada. São Paulo: Editora Saraiva, 2000.

LAPPONI, J. C. Estatística Usando Excel. São Paulo: Editora Lapponi, 2000.

LEVIN, J. Estatística Aplicada a Ciências Humanas, 2ª edição. São Paulo: Editora Harper & Row do Brasil Ltda, 1978.

NICK, E. , KELLNER, S. R. O. Fundamentos de Estatística para as Ciências do Comportamento. Rio de Janeiro: Editora Renes, 1971.

SIEGEL, S. Estatística Não Paramétrica. São Paulo: Editora McGraw-Hill do Brasil Ltda, 1975.

STEVENSON, W. J. Estatística Aplicada à Administração. São Paulo: Editora Harper & Row do Brasil Ltda, 1981.

TRIOLA, M. F. Introdução à Estatística. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 7ª ed. 1999.