<u>Miscelânea</u>

1 Aproximação Binomial-Normal e Intervalos de Confiança para proporções

1 Seja Y uma variável aleatória com distribuição binomial de parâmetros $n = 10$ e $p = 0.4$.
a) Determine o valor exato e a aproximação Normal para $P(7 \leq Y)$. Calcule pela aproximação Normal o valor $P(6.5 \leq Y)$ e conclua qual dos dois valores dados pela aproximação representa melhor o verdadeiro valor de $P(7 \leq Y)$, onde Y é Binomial.
b) Determine o valor exato e a aproximação Normal para $P(Y < 5)$. Calcule pela aproximação Normal o valor $P(Y \leq 4.5)$ e conclua qual dos dois valores dados pela aproximação representa melhor o verdadeiro valor de $P(Y < 5)$, onde Y é Binomial.

2 De um lote de produtos manufaturados, extraímos 100 itens ao acaso. Se 10% dos itens são defeituosos, calcular a probabilidade de 12 itens dentre os selecionados serem defeituosos. Use a aproximação Normal.

3 Seja p a proporção de indivíduos com glaucoma na cidade de Campinas. Se o Ministério da Saúde informa que a proporção atual é igual a 0.1 e temos uma amostra de 19 indivíduos selecionados ao acaso desta população, responda:
a) Qual é a probabilidade de que a proporção amostral

$$p^* = \frac{n^o \text{ de portadores na amostra}}{19}$$

seja maior ou igual a 5/19? Calcule a probabilidade exata.
b) Resolva o item a) utilizando a aproximação Normal dada pelo Teorema Central do Limite.
c) Compare o valor exato de $P(5/19 \leq p^*)$ calculado em a) com os valores calculados por aproximação Normal: $P_N(5/19 \leq p^*)$ (do item b)), $P_N(4.5/19 \leq p^*)$, $P_N(5.5/19 \leq p^*)$, estes últimos calculados utilizando a aproximação Normal.
Determine qual deles representa melhor o verdadeiro valor dado no item a).

d) Calcule a probabilidade exata $P(p^* \leq 2/19)$ e compare o resultado com as aproximações pela Normal:
$P_N(p^* \leq 1.5/19), P_N(p^* \leq 2/19), P_N(p^* \leq 2.5/19)$. Qual destes valores representa melhor o verdadeiro valor $P(p^* \leq 2/19)$?

4 Uma amostra aleatória de 625 donas-de-casa revela que 70% delas preferem a marca X de detergente. Construir o intervalo de confiança de 90% (intervalo por estimativa pontual e intervalo conservador) para $p =$ proporção das donas-de-casa que preferem X.

5 Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis a seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato em questão.

(a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja no máximo 0.01 com probabilidade de 80%.

(b) Se na amostra fina (com tamanho dado por (a)) observou-se que 55% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de confiança (95%) para a proporção p.

6 Suponha que estejamos interessados em estimar a porcentagem de consumidores de um certo produto. Se uma amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

(a) O intervalo de confiança para p, com coeficiente de confiança 95%.

(b) O tamanho da amostra para que o erro da estimativa não exceda 0.02 unidades com probabilidade de 95%.

7 Um pedido de auxílio, feito pelo correio, teve 412 respostas a 5000 cartas enviadas, e outro pedido, mais dispendioso, teve 312 respostas a 3000 cartas enviadas. Obtenha o intervalo de confiança de 90% para a diferença de proporções entre os dois pedidos.

2 Teste de hipóteses para proporções

1 Os dados correspondem a uma distribuição $Bin(n, p)$. Conduzir o seguinte teste
$H_0 : p = 0.75$ vs $H_1 : p < 0.75$. Assuma $n = 150$.
a) Se $p^* = 0.72$ for uma estimativa pontual de p. Determine a força da evidência contida nos dados (p-valor).
b) Verifique se a entimativa p^* apresenta evidência suficiente para rejeitar H_0 ao nível $\alpha = 0.01$.

2 Suspeita-se da honestidade de um dado de 6 faces. Procurando suporte para tal afirmação considera-se o número de vezes que a face 2 é obtida numa sequência de n lançamentos independentes.
a) Determine a hipótese nula H_0 e a alternativa H_1.
b) Em $n = 20$ lançamentos independentes obtem-se 2 vezes a face 2. Calcule a força da evidência contida nos dados e responda: Para que níveis de significância α, a hipótese H_0 é rejeitada?. Interprete.
Calcule o π-value utilizando a aproximação normal e responda: Para que níveis de significância α, a hipótese H_0 é rejeitada?. Compare.
c) Em $n = 20$ lançamentos independentes obtem-se 6 vezes a face 2. Calcule a força da evidência contida nos dados e determine se os dados resultam significantes ao nível $\alpha = 0.10$.

3 Membros de uma associação profissional desejam provar que menos da metade dos eleitores apoiam as medidas tomadas pela equipe econômica do governo para enfrentar a crise financeira internacional. Seja p a proporção de eleitores que apoiam as medidas.
a) Determine a hipótese nula e a alternativa de um teste que permita avaliar a situação.
b) Se uma pesquisa com 500 eleitores selecionados ao acaso revela que 228 apoiam as medidas econômicas, podemos dizer que os dados são significantes ao nível $\alpha = 0.05$?.

4 Dois grupos, A e B, são formados por pessoas distintas que possuem a mesma enfermidade. É ministrado um soro ao grupo A mas não ao grupo B. Das 100 pessoas que formaram o grupo A, 75 se curaram e, das 100 pessoas que formaram o grupo B, 65 obtiveram a cura. Verifique se o soro é eficiente na cura da enfermidade.

5 Numa amostra aleatória de visitantes de um museu, 22 de 100 famílias provenientes da região Sul e 33 de 120 famílias provenientes de São Paulo compraram alguma coisa nas lojas do museu. Podemos considerar que a proporção de pessoas, provenientes da região Sul e de São Paulo, que compraram alguma coisa nas lojas do museu, são iguais?

6 Um método de borrifar nuvens (para provocar chuva) obteve sucesso em 54 dentre 150 tentativas, enquanto que o outro método obteve sucesso em 33 dentre 100 tentativas. Pode-se concluir que o primeiro método é superior ao segundo?

7 Sendo X o número de sucessos em $n = 10$ provas de Bernoulli, queremos testar $H_0 : p = 0,6$. Se o teste for unilateral e rejeitarmos H_0 para valores pequenos de X, detemine o p-valor se o valor observado de X for 3. Conclua sobre a rejeição ou não de H_o.

8 Membros de uma associação patronal desejam demonstrar que mais de 60% dos seus associados apoiam a política de privatização do governo. Determine a região crítica do teste de hipótese para essa situação, para um nível de significância $\alpha = 0.05$, supondo que os dados são colhidos de uma amostra com 80 associados selecionados ao acaso.

3 Estimação por Intervalo para populações Normais

1 Considere a seguinte amostra aleatória de tamanho 20 proveniente de uma distribuição Normal de média desconhecida μ e variância σ^2: 13.736, 14.579, 14.025, 13.542, 14.294, 13.815, 13.615, 13.633, 13.893, 14.105, 14.129, 15.029, 13.814, 14.516, 13.982, 14.174, 13.900, 14.319, 13.822, 13.728

a) Calcule o $100\gamma\%$ I.C. para μ sabendo que $\sigma^2 = 0.36$, para os níveis de confiança $\gamma = 0.9, 0.95, 0.99$. Calcule o comprimento de cada intervalo de confiança. Que evidência há na relação entre o comprimento do intervalo e o nível de confiança?

b) Calcule o $100\gamma\%$ I.C. para μ supondo σ^2 desconhecido, para os níveis de confiança $\gamma = 0.9, 0.95, 0.99$.

2 Considere a seguinte amostra aleatória de tamanho 15 proveniente de uma distribuição Normal de média μ e variância desconhecida σ^2: 5.055, 6.916, 5.812, 5.044, 4.914, 5.665, 4.772, 5.502, 3.841, 5.782, 4.579, 5.477, 7.158, 5.254, 5.276

a) Calcule o $100\gamma\%$ I.C. para σ sabendo que $\mu = 5$, para os níveis de confiança $\gamma = 0.9, 0.95, 0.99$. Calcule o comprimento de cada intervalo de confiança.

b) Calcule o $100\gamma\%$ I.C. para σ supondo μ desconhecido, para os níveis de confiança $\gamma = 0.9, 0.95, 0.99$. Calcule o comprimento de cada intervalo de confiança.

3 Considere a média amostral de uma a.a. de uma população com média μ e variância igual 10. Encontre o valor de n para que o intervalo aleatório$(\bar{X} - 0,5; \bar{X} + 0,5)$ tenha probabilidade aproximada de 0,954 de conter μ.

4 Sejam $X_1, ..., X_n$ a.a. tendo distribuição $N(\mu, \sigma^2)$. A margem de erro de um IC de nível $100(1-\alpha)\%$ para μ utilizando-se $\bar{X}$, se σ conhecido, é dado por $e = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}$. Considerando $\sigma = 4$, qual é o tamanho amostral de modo a ter 90% de certeza de que o erro de estimação não exceda 0,8?

4 Teste de hipóteses para populações Normais

1 Denotemos por μ a verdadeira média de nível de radioatividade (picocuries por litro). O valor 5 pCi/L é considerado como linha divisoria entre água segura e não segura. Qual dos seguintes testes recomenda conduzir?
$H_0 : \mu = 5$ vs $H_1 : \mu > 5$
$H_0 : \mu = 5$ vs $H_1 : \mu < 5$
Explique seu raciocínio em termos dos erros tipo I e II.

2 Para cada situação apresentada a seguir, verifique se os dados apresentam evidência suficiente para rejeitar a hipótese nula sendo que s denota o desvio padrão amostral $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

a) População normal, $n = 15$, $\overline{X} = 83.9$, $s = 18.2$, $\alpha = 10\%$, para o teste $H_0 : \mu = 85$ vs $H_1 : \mu < 85$.

b) População normal, $n = 15$, $\overline{X} = 79.1$, $s = 11.8$, $\alpha = 10\%$, para o teste $H_0 : \mu = 76$ vs $H_1 : \mu \neq 76$.

3 Sabendo que a resistência à tensão, de uma peça de algodão possui distribuição normal.

a) A resistência é medida em 15 peças selecionadas ao acaso, observando-se uma média amostral igual a 39.3 e um desvio padrão amostral igual a 2.6. Verifique se os dados são significantes ao nível $\alpha = 10\%$, para o teste $H_0 : \mu = 40$ vs $H_1 : \mu \neq 40$.

b) Determine a região crítica dos teste enunciado em a) para $\alpha = 10\%$.

c) A resistência é medida em 54 peças selecionadas ao acaso, observando-se uma média amostral igual a 42.4 e um desvio padrão amostral igual a 3.1. Calcule a força da evidência contida nos dados e determine para quais níveis de significância H_0 é rejeitada.

d) Melhorias implementadas no tratamento da fibra de algodão permitem suspeitar que a resistência tem aumentado. Perante esta afirmação reformule o teste. Se essa resistência foi medida em 15 peças observando-se uma média amostral de 41.3 com um desvio padrão amostral igual a 2.6. Verifique se os dados são significantes ao nível $\alpha = 0.05$. Determine a região crítica do teste para $\alpha = 0.05$.

4 A demanda biológica de oxigênio (DBO) é um índice de poluição controlado nas indústrias de papel e celulose (para preservar o equilibrio ambiental toda indústria de papel deve consumir uma quantidade de oxigênio que não supere um "valor limite"). Em 43 medidas coletadas numa indústria, no peíodo: Setembro 1999-Fevereiro 1999, a média e o desvio padrão dos dados observados foram 3242 ppd e 757 ppd, respectivamente. Aquela empresa tinha estabelecido como valor limite 3000 ppd para o DBO médio. Julgaria que os dados amostrais suportam que a meta foi atingida ao nível $\alpha = 5\%$?

5 Uma empresa mineira acredita que a exploração de urânio é possível numa certa região, isto é, na região a concentração média de urânio é superior a 10. Admitindo-se que a distribuição desta concentração é normal e que as medições em 13 pontos selecionados ao acaso na região são 7.92, 10.29, 19.89, 17.73, 10.36, 13.50, 8.81, 6.18, 7.02, 11.71, 8.33, 9.32, 14.61

a) Verifique se há evidência suficiente contra a hipótese de abandono da área .

b) Qual seria a região crítica do teste ao nível de significância $\alpha = 2\%$?

6 Em 18 condenações por "posse de drogas" num tribunal norte-americano as condenações atribuídas tiveram média de 38 meses e desvio padrão amostral de 4 meses. Considerando que as condenações são normalmente distribuídas, verifique se os dados suportam ao nível de significância $\alpha = 5\%$ a suspeita de que nesse tribunal as condenações por "posse de drogas" é em média maior do que 36 meses. Faça o mesmo teste considerando $\alpha = 1\%$. Interprete.

7 Um fabricante de aparelhos de TV afirma que são necessários no máximo 250 microamperes (μA) para atingir um certo grau de brilho num tipo de TV. De uma amostra de 20 aparelhos obtivemos uma média amostral de $\overline{X} = 257.3$ μA. Denotemos por m a verdadeira média de

μA necessário para atingir o grau de brilho desejado e suponhamos que m é a média de uma população normal com σ conhecido e igual a 15.

a) Calcule a força da evidência contida nos dados para o nível $\alpha = 0.05$ conduzindo o teste cuja hipótese nula especifica que m é no máximo 250 μA.

b) Calcule a região crítica do teste para o nível $\alpha = 0.05$.

c) Se $m = 260$, Qual é a probabilidade de cometer um erro tipo II?

d) Para qual valor de n (tamanho amostral) a probabilidade de cometer o erro tipo II resulta igual a 0.01.

8 O ponto de desvanecimento de cada uma de 16 amostras de uma certa marca de vegetais hidrogenados foi determinado, resultando numa média amostral $\overline{X} = 94.32$. Considerando que o ponto de desvanecimento possui distribuição normal de desvio conhecido $\sigma = 1.20$.

a) Verifique se a amostra apresenta evidência suficiente para rejeitar H_0 ao nível $\alpha = 0.01$, calculando o π-value onde $H_0 : \mu = 95$ vs $H_1 : \mu \neq 95$.

b) Se $\alpha = 0.01$ e $\mu' = 94$. Qual é a probabilidade de cometer erro tipo II?

c) Se $\alpha = 0.01$. Que valor de n (tamanho amostral) é necessário para obter uma probabilidade de cometer erro tipo II, com $\mu' = 94$ igual a 0.1?

9 O ponto médio desejado de SiO_2 em certo tipo de cimento aluminoso é de 5.5. Para provar se o verdadeiro ponto médio da porcentagem numa planta de produção em particular é 5.5, foram coletadas 16 amostras. Supondo que a porcentagem de SiO_2 numa amostra está normalmente distribuída com desvio conhecido e igual a $\sigma = 0.3$ e sabendo que na amostra selecionada obteve-se $\overline{X} = 5.25$, responda:

a) Os dados indicam de forma conclusiva que o verdadeiro ponto médio de porcentagem não é $\mu = 5.5$?.

b) Se $\alpha = 0.01$, qual o valor de n (tamanho amostral) é necessário para obter uma probabilidade de cometer erro tipo II, com $\mu' = 5.6$ igual a 0.01?

10 Um experimento para comparar a resistência de coesão à tensão do morteiro modificado de látex de polímeros, com a resistência do morteiro não modificado, supondo que os dados tem distribuição Normal; resultou em $\overline{X} = 18.12\ kfg/cm^2$ para o morteiro modificado e em $\overline{Y} = 16.87\ kfg/cm^2$ para o morteiro não modificado. Sejam μ_1 e μ_2 as verdadeiras resitências de coesão à tensão para os morteiros modificado e não modificado respectivamente. Verifique se os dados suportam a rejeição de H_0. Onde $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 > 0$ com nível de significância $\alpha = 0.01$, nas seguintes situações:

a) Se para o morteiro modificado foi utilizada uma amostra de tamanho $m = 40$ e para o não modificado foi utilizada uma amostra de tamanho $n = 32$. Os valores dos desvios são conhecidos σ_1 e σ_2 (associados respectivamente ao morteiro modificado e ao não modificado), $\sigma_1 = 1.6$ e $\sigma_2 = 1.4$. Proponha uma estatística para conduzir o teste e verifique se os dados indicam a rejeição de H_0.

b) Se para o morteiro modificado foi utilizada uma amostra de tamanho $m = 40$ e para o não modificado foi utilizada uma amostra de tamanho $n = 32$. Os valores dos desvios são conhecidos σ_1 e σ_2 (associados respectivamente ao morteiro modificado e ao não modificado), $\sigma_1 = \sigma_2 = 1.6$. Proponha uma estatística para conduzir o teste e verifique se os dados indicam a rejeição de H_0.

c) Sabendo que

$$m = 30, \sum_{i-1}^{m}(X_i - \overline{X})^2 = 40.1,\ n = 22, \sum_{j=1}^{n}(Y_j - \overline{Y})^2 = 53.22.$$

Assumindo que os desvios σ_1 e σ_2 são desconhecidos e iguais, proponha uma estatística para o teste e determine se os dados indicam a rejeição de H_0.

11 Os estudantes universitários homens entediam-se mais facilmente que as estudantes mulheres?. Esta pergunta foi examinada pelo artigo "Boredom in Young Adults Gender and Cultural Comparisons"(*J. of Cross Cultural Psych.* pp. 209-223). Os autores aplicaram uma escala denominada *Escala Proneness de tedio* a 97 estudantes homens e a 148 estudantes mulheres, todos eles de universidades norteamericanas. Assumindo que a classificação fornecida pela escala Proneness possui distribuição normal verifique se a seguinte informação apoia a hipótese da

investigação.
Faça o teste adequado utilizando um nível de significância $\alpha = 0.05$ e os dados da seguinte tabela

Gênero	Tamanho amostral	Média am.	Desvio verdadeiro (σ)
Homens	97	10.40	4.83
Mulheres	148	9.26	4.86

12 Denotemos por μ_1 e μ_2 aos verdadeiros pontos médios de durações de superfícies de rodagem para duas marcas competidoras de medida FR78-15 de pneus radiais. Faça o seguinte teste de hipótese assumindo que a duração das superfícies de rodagem possui distribuição normal $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$ com nível de significância $\alpha = 0.05$, usando a seguinte informação: $m = 40$, $\overline{X} = 36500$, $\sigma_1 = 2200$ (valor verdadeiro do desvio) e $n = 40$, $\overline{Y} = 33400$, $\sigma_2 = 1900$ (valor verdadeiro do desvio).

13 (Teste sobre a variância com a média desconhecida) Seja $X_1, ..., X_n$ uma amostra aleatória normal $N(\mu, \sigma^2)$, μ desconhecido. Considere $H_o : \sigma^2 = 100$ versus $H_1 : \sigma^2 \neq 100$. Sabendo que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (distribuição Qui-Quadrado com $n-1$ graus de liberdade) onde $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ e utilizando $\alpha = 0,05$, $n = 16$, obtenha a região crítica do teste. Sabendo-se que $S^2 = 169$, qual é a conclusão do teste? Lembre-se que a distribuição Qui-Quadrado não é simétrica.

14 Uma fábrica de embalagens para produtos químicos está estudando dois processos, A e B, para combater a corrosão de suas latas especiais. Considerando $X_1, ..., X_n$ e $Y_1, ..., Y_m$ a.a. dos processos A e B, respectivamente. As amostras X_i, $i = 1, ..., n$ e Y_i, $i = 1, ..., m$ são independentes e possuem distribuição, respectivamente, $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$, σ_1^2 e σ_2^2 conhecidos. Queremos testar o efeito dos tratamentos A e B, isto é, queremos testar as hipóteses $H_o : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. A região crítica associada às hipóteses é $RC = \{\hat{X} - \hat{Y} | \hat{X} - \hat{Y} \leq c_1 ou \hat{X} - \hat{Y} \geq c_2\}$. Utilizando $\alpha = 0,05$ e os dados apresentados na seguinte tabela, conclua sobre os dois tratamentos.

Método	Amostra	Média	Desvio Padrão
A	15	48	10
B	12	52	15

15 (Teste sobre a variância com a média conhecida) Seja $X_1, ..., X_n$ uma amostra aleatória normal $N(\mu, \sigma^2)$, μ conhecido igual a 45. Considere $H_o : \sigma^2 = 7$ versus $H_1 : \sigma^2 < 7$. Sabendo que $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$ (distribuição Qui-Quadrado com n graus de liberdade) e utilizando $\alpha = 0,05$, $n = 15$, obtenha a região crítica do teste. Sabendo-se que $\sum_{i=1}^{n}(X_i - \mu)^2 = 42$, qual é a conclusão do teste? Utilizando $\alpha = 0,01$, qual é a conclusão do teste?

5 Testes Qui-Quadrado

1 Suponha que as proporções de quatro tipos de sangue de certa raça sejam iguais a 0.16, 0.48, 0.2, 0.16. Dadas as frequências observadas de 180, 360, 130 e 100 para uma outra raça, verifique se ela possui a mesma distribuição quanto aos tipos de sangue.

2 Os números de acidentes automobilísticos semanais de determinada comunidade foram os seguintes: 11, 8, 17, 5, 13, 9, 17, 6, 10, 4. Tais frequências estão de acordo com a crença de que as condições dos acidentes são as mesmas, neste período de 10 semanas ?

3 Uma amostra de cinquenta peças produzidas por uma máquina forneceu a distribuição de comprimentos das peças dada a seguir, com valores em mm. A especificação de produção indica que o comprimento das peças tem distribuição normal de média 500 mm e desvio padrão 10 mm. Através dos valores observados, concordamos ou discordamos dessa especificação (use o nível descritivo)? Se a especificação não for a postulada, você consegue explicar o que está errado?

Tabela 1: Especificação

Comprimento (mm)	Frequências
$[480, 485)$	1
$[485, 490)$	5
$[490, 495)$	11
$[495, 500)$	14
$[500, 505)$	9
$[505, 510)$	5
$[510, 515)$	4
$[515, 520]$	1
Total	50

4 Os dados seguintes representam os resultados de uma investigação da distribuição do sexo das crianças de 32 famílias possuindo cada uma delas 4 crianças. Use a distribuição Binomial com $n = 4$ e $p = 0.5$ para calcular as frequências esperadas. Aplique depois o teste Qui Quadrado para verificar se este modelo de distribuição Binomial é satisfatório neste caso.

Tabela 2: Dados

Número de filhos	0	1	2	3	4
Número de famílias	4	10	8	7	3

5 Uma pesquisa eleitoral foi realizada com o objetivo de estudar a influência da idade na preferência por dois candidatos presidenciais A e B. A população de eleitores foi dividida em três faixas de iddade sendo que de cada uma delas foi obtida uma amostra de 200 indivíduos. Em seguida, os eleitores selecionados em cada amostra tiveram suas opiniões registradas. Os resultados obtidos foram os seguintes:

Tabela 3: Dados

Intervalo	Preferem A	Preferem B	Indecisos
$[16, 30)$	67	117	16
$[30, 50)$	109	74	17
$[50, \infty)$	118	64	18

Estabeleça uma hipótese apropriada e efetue o teste correspondente.

6 Numa hipotética epidemia 900 crianças contraíram a doença. Das 450 que não receberam tratamento, 104 sofreram efeitos posteriores. Das 450 restantes que efetivamente receberam tratamento, 166 sofreram efeitos posteriores. Teste a hipótese de que a chance de sofrer efeitos posteriores é a mesma para indivíduos que receberam e não receberam tratamento.

7 Os resultados da classificação de 100 pessoas segundo a cor dos olhos e do cabelo foram os seguintes: Você diria que a cor dos olhos independe da cor do cabelo?

Tabela 4: Dados

	Castanhos	Azuis	Cinza
Claro	13	18	9
Escuro	37	12	11

8 No exame final de um curso compareceram 105 estudantes do sexo masculino e 40 do sexo feminino. O professor anotou se o estudante escolhia uma carteira nas filas da frente ou nas de trás. Os resultados dessa observação estão apresentados na tabela abaixo.

Tabela 5: Dados

	Filas da Frente	Filas de trás
Homens	35	70
Mulheres	20	20

A escolha da fila é a mesma para homens e mulheres?

9 Uma amostra de 200 adultos foi entrevistada a respeito de certo projeto de lei. Os resultados são os que seguem,

Tabela 6: Dados

	Favoráveis	Contrários
Homens casados	56	24
Homens solteiros	15	25
Mulheres casadas	24	16
Mulheres solteiras	13	27

Verifique se a opinião independe do sexo; além disso, verifique se a opinião independe do estado civil.

10 Uma agência de turismo classifica uma amostra de 200 clientes com menos de 40 anos segundo o sexo e o passeio preferido (praias ou serras). Após a aplicação de um teste Qui Quadrado apropriado, obteve-se um $p - valor = 0.451$. A mesma pesquisa é repetida para uma amostra de 200 clientes com 40 anos ou mais e o valor do $p - valor = 0.001$. Os dados são apresentados a seguir.

Tabela 7: Menos de 40 anos

	Praias	Serras
Masculino	70	30
Feminino	65	35

Tabela 8: 40 anos ou mais

	Praias	Serras
Masculino	50	50
Feminino	70	30

a. Formule as hipóteses para cada caso

b. Que tipo de teste Qui Quadrado foi aplicado?. Justifique.

c. Interprete os resultados.

6 Regressão Linear e Método de Mínimos Quadrados

1 Uma companhia local de fornecimento de energia selecionou uma residência típica para desenvolver um modelo para o consumo de energia (em kilowatts por dia) como uma função da temperatura (graus centígrados) diária nos meses de inverno. A seguinte informação foi coletada em 15 dias.

Medição	1	2	3	4	5	6	7
Temperatura	0	8	7.5	13.5	14	8.5	4.5
Energia	70	57	60	63	57	66	67

Medição	8	9	10	11	12	13	14	15
Temperatura	-11	-7.5	-8.5	1.5	0.5	2	-6	-4
Energia	107	96	88	80	64	79	82	97

a) Graficar os dados.

b) Calcular o coeficiente de correlação amostral.

c) Calcular a reta de ajuste pelo método dos Mínimos Quadrados. Graficar a reta junto aos dados.

d) Calcular o Erro Quadrático Médio associado ao modelo linear.

e) Qual o consumo aproximado se a temperatura média do dia é 6 graus Centígrados?.

2 A distância de detenção (pés) de um carro numa determinada estrada foi estudada como função da velocidade (mi/h). Os dados coletados são

Medição	1	2	3	4	5	6
Velocidade	20.5	20.5	30.5	40.5	48.8	57.8
distância	15.4	13.3	33.9	73.1	113.0	142.6

Seja $Y =$ distância. Ajustar Y e $\sqrt{Y}$ como funções lineares da velocidade. Qual dos ajustes acha mais adequado? Justifique.

3 Graficar os seguintes pares de pontos (x,y):

x	2.38	0.46	-1.14	0.38	0.67	-0.94	0.27	0.3
y	13.16	7.38	2.57	7.15	8	3.22	6.82	6.7

x	-0.16	-0.17	2.83	-1.28	1.44	1.67	0.27
y	5.52	5.5	14.48	2.17	10.1	11	6.8

(a) Calcular a reta de regressão pelo método de mínimos quadrados. Graficar esta reta junto ao gráfico anterior.
(b) Repetir o item anterior, trocando o primeiro ponto (2.38, 13.16) por (2.38, 22).
(c) Calcular o erro quadrático médio em ambos os casos e compará-los.

4 Graficar os seguintes pares de pontos (x,y):

x	0.34	1.38	-0.65	0.68	1.40	-0.88	-0.30	-1.18	0.50	-1.75
y	0.27	1.34	-0.53	0.35	1.28	-0.98	-0.72	-0.81	0.64	-1.59

(a) Ajustar uma reta $y = ax + b$ pelo método de mínimos quadrados e graficá-la junto ao gráfico anterior.
(b) Ajustar uma reta $x = cy + d$ pelo método de mínimos quadrados e graficá-la junto ao gráfico anterior.
(c) As retas dos itens (a) e (b) são iguais? Se não forem, qual é a razão?

5 Os seguintes dados são resultantes de um experimento para investigar a relação entre a força aplicada x e o tempo y para a ruptura de placas de metal laminado.

x	22.5	25	28	30.5	38	40.5	42.5	48	54.5	55	70
y	44	42	33.5	28	18	13.6	15	10.3	9	6.3	4

(a) Graficar os pontos (x,y). é razoável pensar em um ajuste linear?
(b) Graficar os pontos (ln(x), ln(y)). Calcular a reta de regressão pelo método de mínimos quadrados e graficá-la junto ao gráfico anterior.
(c) Quanto será aproximadamente o tempo necessário para a ruptura quando a força aplicada x for igual a 62?

6 Consideremos o modelo linear

$$Y_i = ax_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

onde $\epsilon_1, \ldots \epsilon_n$ são variáveis aleatórias independentes com

$$E(\epsilon_i) = 0 \quad e \quad Var(\epsilon_i) = \sigma^2 < \infty.$$

(a) Obter o estimador $\hat{a}$ de a por mínimos quadrados.
(b) Calcular $E(\hat{a})$ e $Var(\hat{a})$.

7 Estimação por Máxima Verossimilhança

1 Considere uma amostra aleatória de tamanho 2 proveniente da distribuição exponencial, digamos (X_1, X_2) amostra de

$$f(x) = \lambda \exp(-\lambda x).$$

Obtenha o estimador de Máxima Verossimilhança de λ desde a amostra. Sabendo que a amostra assume o valor $(2.5, 3)$, i.e. X_1 toma o valor 2.5 e X_2 o valor 3, calcule o valor do estimador de Máxima Verossimilhança de λ.

2 Considere uma amostra aleatória de tamanho 4 (X_1, X_2, X_3, X_4) proveniente da distribuição Binomial Negativa $BN(\theta, r)$ com θ desconhecido $0 < \theta < 1$,

$$p(x) = \begin{pmatrix} r - x - 1 \\ r - 1 \end{pmatrix} \theta^r (1-\theta)^x, \;\; x = 0, 1, 2, \cdots$$

Assuma $r = 2$ e obtenha o estimador de Máxima Verossimilhança de θ desde a amostra. Sabendo que a amostra assume os valores $(2, 4, 1, 5)$, calcule o estimador nesse caso.

3 Suponha que X é uma v.a. $Poisson(\lambda)$, com λ desconhecido e positivo.

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \; x = 0, 1, 2, \cdots$$

Considere uma amostra aleatória de tamanho n e calcule o estimador de Máxima Verossimilhança para λ desde a amostra. Calcule o estimador sabendo que a amostra tem tamanho 3 e assume os seguintes valores $(1, 1, 3)$.

4 Suponha que X é uma v.a. $Geom(p)$, com $p > 0$ desconhecido.

$$p(x) = (1-p)^x p, \; x = 0, 1, 2, \cdots$$

Considere uma amostra aleatória de tamanho n e calcule o estimador de Máxima Verossimilhança para p desde a amostra. Calcule o estimador sabendo que a amostra tem tamanho 6 e assume os seguintes valores $(0, 5, 1, 7, 1, 3)$.