**MAE-5755- Métodos Estatísticos Aplicados às Ciências Biológicas**

**Lista de exercícios**

1. Um criminologista desejava estudar a relação entre: X (densidade populacional = número de pessoas por unidade de área) e Y (índice de assaltos = número de assaltos por 100000 pessoas) em grandes cidades .Para isto sorteou 10 cidades observando em cada uma delas os valores de X e Y . Os resultados obtidos foram:

Cidade 1 2 3 4 5 6 7 8 9 10

X 59 49 75 65 89 70 54 78 56 60

Y 190 180 195 186 200 204 192 215 197 208

1. Classifique as variáveis envolvidas.
2. Calcule a média, mediana, desvio-padrão e a distância interquartis para cada variável.
3. Construa o diagrama de dispersão entre Y e X. Comente.
4. As notas finais de um curso de Estatística foram: 7, 5, 4, 5, 6, 3, 8, 4, 5, 4, 6, 4, 5, 6, 4, 6, 6, 3, 8, 4, 5, 4, 5, 5 e 6.
5. Determine a mediana, os quartis e a média.
6. Separe o conjunto de dados em dois grupos denominados *aprovados*, com nota pelo menos igual a 5 e *reprovados*. Compare a variância desses dois grupos.
7. O gráfico abaixo considera as variáveis pressão sistólica e idade de imigrantes com menos de dez anos (Migra1) e com mais de dez anos (Migra2) desde a migração.



A dispersão dos pontos indica que:

1. existem muitos pontos aberrantes
2. existe correlação linear positiva entre as variáveis para o grupo Migra2
3. independentemente do tempo desde a migração as variáveis são altamente correlacionadas
4. d) existe correlação linear positiva entre as variáveis para o grupo Migra1
5. Considerando a reta de regressão ajustada:

## Diâmetro = 7,68 + 0,185 Volume.

Podemos dizer que:

1. o diâmetro não pode ser estimado a partir do volume
2. o coeficiente de correlação linear entre as variáveis é nulo
3. há um aumento médio de 0,185 unidades no diâmetro com o aumento de uma unidade de volume
4. o valor estimado do diâmetro é 7,68 unidades para volumes iguais a 1 unidade

5) Considere o seguinte resumo descritivo da pulsação de estudantes com atividade física intensa e fraca:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Atividade | N | Média | Mediana | Desvio padrão | Mínimo | Máximo | 1ºquartil | 3ºquartil |
| Intensa | 30 | 79,6 | 82 | 10,5 | 62 | 90 | 70 | 79 |
| Fraca | 30 | 73,1 | 70 | 9,6 | 58 | 92 | 63 | 77 |

Estes resultados indicam que:

1. 25% e 50% dos estudantes com atividade física intensa e fraca, respectivamente, tiveram pulsação inferior a 70
2. a pulsação de um estudante com fraca atividade física é provavelmente inferior a 63
3. a atividade física não tem efeito na média da pulsação dos estudantes
4. quaisquer 15 estudantes com fraca atividade física têm pulsação inferior a 70

6) Os gráficos *boxplot* dos dados de pulsação dos estudantes submetidos a atividade física intensa e fraca são, respectivamente:



1. A e B
2. B e D
3. A e C
4. B e C

7) Os histogramas a seguir mostram a distribuição das temperaturas (°C) ao longo de vários dias de investigação para duas regiões (R1 e R2). Podemos dizer que:



1. as temperaturas das regiões R1 e R2 têm mesma média e mesma variância
2. nada podemos afirmar sobre as médias e as variâncias
3. a temperatura média da região R2 é maior que a de R1
4. as temperaturas das regiões R1 e R2 têm mesma média e variância diferentes
5. Na companhia A, a média dos salários é 10000 unidades e o 3o quartil é 5000.
6. Se você se apresentasse como candidato a funcionário nessa firma e se o seu salário fosse escolhido ao acaso entre todos os possíveis salários, o que seria mais provável: ganhar mais ou menos que 5000 unidades?
7. Suponha que na companhia B a média dos salários seja 7000 unidades, a variância praticamente zero e o salário também seja escolhido ao acaso. Em qual companhia você se apresentaria para procurar emprego, com base somente no salário?
8. Um experimento foi realizado em dois laboratórios de modo independente com o objetivo de verificar o efeito de três tratamentos (A1, A2 e A3) na concentração de uma substância no sangue de animais. As concentrações observadas nos dois laboratórios são apresentas abaixo

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Laboratório 1 | | |  |  | Laboratório 2 | | |
|  | A1 | A2 | A3 |  |  | A1 | A2 | A3 |
|  | 8 | 4 | 3 |  |  | 4 | 6 | 5 |
|  | 3 | 8 | 2 |  |  | 5 | 7 | 4 |
|  | 1 | 10 | 8 |  |  | 3 | 7 | 6 |
|  | 4 | 6 | 7 |  |  | 4 | 8 | 5 |
| Total | 16 | 28 | 20 |  | Total | 16 | 28 | 20 |

1. O que você pode comentar sobre as médias dos três tratamentos nos dois laboratórios?
2. Sem nenhum cálculo, apenas olhando os dados, em qual dos dois laboratórios será observado o maior valor da estatística F na análise de variância?
3. Num conjunto de dados, o primeiro quartil é 10, a mediana é 15 e o terceiro quartil é 20. Indique quais das seguintes afirmativas são verdadeiras, justificando sua resposta:
4. o intervalo interquartis é 5
5. o valor 32 seria considerado *outlier* segundo o critério utilizado na construção do *boxplot*
6. a mediana ficaria alterada de 2 unidades se um ponto com valor acima do terceiro quartil fosse substituído por outro 2 vezes maior
7. o valor mínimo é maior do que zero.
8. A bula de um medicamento A para dor de cabeça afirma que o tempo médio para que a droga faça efeito é de 60 seg com desvio padrão de 10 seg. A bula de um segundo medicamento B afirma que a média correspondente é de 60 seg com desvio padrão de 30 seg Sabe-se que as distribuições são simétricas. Indique quais das seguintes afirmativas são verdadeiras, justificando sua resposta:
9. os medicamentos são totalmente equivalentes com relação ao tempo para efeito pois as médias são iguais
10. com o medicamento A, a probabilidade de cura de sua dor de cabeça antes de 40 seg é maior do que com o medicamento B
11. com o medicamento B a probabilidade de você ter sua dor de cabeça curada antes de 60 seg é maior que com o medicamento A
12. Na tabela abaixo estão indicadas as durações de 335 lâmpadas.

|  |  |
| --- | --- |
| Duração (horas) | Número de lâmpadas |
| 0 - 100 | 82 |
| 100 - 200 | 71 |
| 200 - 300 | 68 |
| 300 - 400 | 56 |
| 400 - 500 | 43 |
| 500 - 800 | 15 |

1. Esboce o histograma correspondente.
2. Calcule os quantis de ordem p=0.1, 0.3, 0.5, 0.7 e 0.9.
3. A tabela abaixo representa a distribuição do número de dependentes por empregados da Empresa Mirante.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

n° de Dependentes n° de Empregados

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. 40
2. 50
3. 30
4. 20
5. 10

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Total 100

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A mediana, média e moda são, respectivamente:

1. 50; 15; 50
2. 1; 2,1; 1
3. 50,5; 50; 50
4. 1; 1; 1
5. Qual a porcentagem de empregados da Empresa Mirante com 2 ou mais dependentes?
6. 40%
7. 50%
8. 60%
9. 75%
10. No artigo intitulado “Estimativas do Valor Energético a partir de Características Químicas e Bromatológicas dos Alimentos” (Rev. Bras. Zootec., 30(6):1837-1856, 2001) estudou-se a disponibilidade de energia de alimentos considerando os nutrientes digestíveis totais (NDT) e também as análises químicas e metabólicas das dietas. Nele se apresentam os seguintes gráficos:

#### 

1. Para qual variável observou-se uma correlação linear maior entre a variável explicativa e o NDT? Justifique.
2. Calcule o valor esperado de NDT se o valor de porcentagem da digestibilidade da matéria seca for 47.
3. Os dados da seguinte tabela sugerem que tipo de relação entre as variáveis X e Y?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 3 | 4 | 6 | 8 | 9 | 11 | 14 |
| Y | 1 | 2 | 4 | 4 | 5 | 7 | 8 | 9 |

1. não há associação entre X e Y
2. há relação linear positiva
3. há relação linear negativa
4. há relação quadrática
5. Os dados abaixo foram retirados de Altman (1991) e representam a relação entre os resultados do mapeamento do fígado e os da anatomia patológica (padrão áureo). O interesse é verificar quão bom é o mapeamento do fígado na identificação de anormalidade patológica.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Situação A   |  |  |  |  | | --- | --- | --- | --- | |  | Patologia | |  | | Mapeamento | Positivo | Negativo | Total | | Positivo | 231 | 32 | 263 | | Negativo | 27 | 54 | 81 | | Total | 258 | 86 | 344 |   Situação B   |  |  |  |  | | --- | --- | --- | --- | |  | Patologia | |  | | Mapeamento | Positivo | Negativo | Total | | Positivo | 77 | 96 | 173 | | Negativo | 9 | 162 | 171 | | Total | 86 | 258 | 344 |   Para as situações A e B, calcule a prevalência de anormalidade, sensibilidade, especificidade, valores preditivo positivo e negativo. Quais as conclusões? |

1. Um estudo foi realizado com o objetivo de avaliar a influência da exposição ao material particulado fino (MP2,5) na capacidade vital forçada (% do predito) em indivíduos que trabalham em ambiente externo. Deseja-se verificar se o efeito da exposição depende da ocorrência de hipertensão ou diabetes. Os 101 trabalhadores na amostra foram classificados quanto à exposição e presença de diabetes ou hipertensão. As médias da capacidade vital forçada em cada combinação das categorias de diabetes ou hipertensão e exposição estão representadas na figura abaixo.
2. Comente os resultados obtidos
3. As médias amostrais sugerem a existência de efeito de interação entre exposição e diabetes ou hipertensão?



1. Em um teste de esforço cardiopulmonar aplicado a 55 mulheres e 104 homens, foram medidas entre outras, as seguintes variáveis:

* Grupo: (Normais, Cardiopatas ou DPOC (portadores de doença pulmonar obstrutiva crônica)
* VO2MAX: consumo máximo de O2 (ml/min)
* VCO2MAX:consumo máximo de CO2 (ml/min)

Algumas medidas descritivas e gráficos são apresentados abaixo.

## VO2MAX

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grupo** | **n** | **Média** | **Mediana** | **Desvio padrão** |
| Normais | 56 | 1845 | 1707 | 795 |
| Cardiopatas | 57 | 1065 | 984 | 434 |
| DPOC | 46 | 889 | 820 | 381 |

## VCO2MAX

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grupo** | **n** | **Média** | **Mediana** | **Desvio padrão** |
| Normais | 56 | 2020 | 1847 | 918 |
| Cardiopatas | 57 | 1206 | 1081 | 479 |
| DPOC | 46 | 934 | 860 | 430 |

Coeficiente de correlação entre VO2MAX e VCO2MAX = 0,92

1. Que grupo tem a maior variabilidade?



1. Compare as médias e as medianas dos 3 grupos.
2. Compare os intervalos interquartílicos dos 3 grupos para cada variável. Você acha razoável usar a distribuição normal para esse conjunto de dados?
3. O que representam os asteriscos nos boxplots?
4. Que tipo de função você ajustaria para modelar a relação entre o consumo máximo de CO2 e o consumo máximo de O2? Por quê?
5. Há informações que necessitam verificação quanto a possíveis erros? Quais?
6. Para avaliar a associação entre a persistência do canal arterial em recém nascidos pré-termo (RNPT) e óbito ou hemorragia intracraniana, um pesquisador obteve os dados dispostos nas Tabelas 1 e 2.

**Tabela 1:** Frequência absoluta e relativa de óbitos.

|  |  |  |  |
| --- | --- | --- | --- |
| **PCA** | **Óbito** | |  |
| **Sim** | **Não** | **Total** |
| Presente | 8 (38,1%) | 13 (61,9%) | 21 (100,0%) |
| **Ausente** | 1 (2,5%) | 39 (97,5%) | 40 (100,0%) |
| **Total** | 9 (14,8%) | 52 (85,2%) | 61 (100,0%) |

**Tabela 2**: Frequência absoluta e relativa de hemorragia intracraniana.

|  |  |  |  |
| --- | --- | --- | --- |
| **PCA** | **Hemorragia intracraniana** | |  |
| **Sim** | **Não** | **Total** |
| Presente | 7 (33,3%) | 14 (66,7%) | 21 (100,0%) |
| **Ausente** | 7 (17,5%) | 33 (82,5%) | 40 (100,0%) |
| **Total** | 14 (23,0%) | 44 (77,0%) | 61 (100,0%) |

Os resultados da análise estão resumidos na Tabela 3

**Tabela 3:** Valores p dos testes de homogeneidade das distribuições correspondentes a RNPT com e sem PCA e razões de chances (odds ratios).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variável** | | **Nível descritivo** | | **Razões de chances (odds ratios) e**  **Intervalo de Confiança (95%)** | | |
|  | | **(p-value)** | | **Estimativa** | **Limite inferior** | **Limite superior** |
| Óbito | 0,001 | | 24,0 | | 2,7 | 210,5 |
| **Hemorragia intracraniana** | 0,162 | | 2,4 | | 0,7 | 8,0 |

1. Interprete as estimativas das razões de chances, indicando claramente a que pacientes elas se referem.
2. Analogamente, interprete os intervalos de confiança correspondentes, indicando claramente a que pacientes eles se referem.
3. Com base nos resultados anteriores, o que você pode concluir sobre a associação entre persistência do canal arterial e óbito para RNPT em geral? E sobre a associação entre a persistência do canal arterial e a ocorrência de hemorragia interna? Justifique suas respostas.
4. Qual a hipótese nula testada em cada caso?
5. Qual a interpretação dos níveis descritivos (p-value) em cada caso?

**Fonte**: Afiune, J.Y. (2000). Avaliação ecocardiográfica evolutiva de recém-nascidos pré-termo, do nascimento até o termo. Tese de doutorado, FMUSP.

1. Num estudo realizado por Humberto F.G. Freitas e A.J. Mansur no Instituto do Coração da FMUSP, candidatos a transplante foram acompanhados durante o período de espera por um coração. O tempo até o evento de interesse (aqui chamado de tempo de sobrevivência) foi definido como o número de dias decorridos entre a primeira consulta de avaliação e o procedimento cirúrgico. Entre possíveis fatores que poderiam influenciar o tempo até o transplante está a *presença de insuficiência tricúspide*.



Para avaliar a importância desse fator, foram construídas curvas de sobrevivência pelo método de Kaplan-Meier e realizada uma análise baseada no modelo de riscos proporcionais de Cox, com ajuste por sexo, idade e etiologia. Os resultados estão indicados na tabela abaixo.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fator** | **Número de casos** | **Nível descritivo**  **(p-value)** | **Risco relativo** | **Intervalo de confiança (95%)** | |
| **limite. inferior** | **limite superior** |
| **Insuficiência tricúspide** | 868 | 0,039 | 1,25 | 1,01 | 1,54 |

1. Estime descritivamente a proporção de pacientes com e sem insuficiência tricúspide cujo tempo até a ocorrência do transplante é de 1500 dias.
2. Existem evidências de que a presença de insuficiência tricúspide contribui para um pior prognóstico? Justifique sua resposta.
3. Interprete o risco relativo apresentado na tabela acima
4. Qual a razão para se incluir um intervalo de confiança na análise.

**Fonte**: Lima, A.C.P., Singer, J.M. e Fusaro, E.R. (2000). Relatório de análise estatística sobre o projeto “Prognóstico de pacientes com insuficiência cardíaca encaminhados para tratamento cirúrgico.” São Paulo: Centro de Estatística Aplicada do IME/USP.

1. Em um estudo comparativo de duas drogas para hipertensão os resultados abaixo foram usados para descrever a eficácia e a tolerabilidade das drogas ao longo de 5 meses de tratamento

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tabela 1. **Efeitos colaterais ao longo do estudo.** | | | | | |
| Efeito | Droga 1 | |  | Droga 2 | |
| Colateral | n | % |  | n | % |
| **não** | **131** | **61,22** |  | **144** | **65,45** |
| **sim** | **83** | **38,79** |  | **76** | **34,54** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| Tabela 2. **Variação da pressão arterial entre o fim e o início do estudo.** | | | | | |
| Variação | Droga 1 | |  | Droga 2 | |
| Pressão | n | % |  | n | % |
| 0 |-- 5 | 9 | 4.20561 |  | 5 | 2.27273 |
| 5 |-- 10 | 35 | 16.3551 |  | 29 | 13.1818 |
| 10 |-- 20 | 115 | 53.7383 |  | 125 | 56.8181 |
| 20 |-- 30 | 54 | 25.2336 |  | 56 | 25.4545 |
| 30 |-- 40 | 1 | 0.46729 |  | 5 | 2.27273 |

|  |
| --- |
|  |
| **Figura 1.** Histogramas da variação da pressão arterial entre o fim e o início do estudo para cada uma das drogas empregadas. |

1. Com a finalidade de melhorar a apresentação dos resultados, faça as alterações que você julgar necessárias em cada uma das tabelas e figura.
2. Calcule a média, o desvio padrão e a mediana da variação de pressão arterial para cada uma das duas drogas.
3. Compare os resultados apresentados abaixo, obtidos diretamente dos dados da amostra, com os que você obteve no item (b).

|  |  |  |  |
| --- | --- | --- | --- |
| **Tabela 2.** Medidas descritivas da variação da pressão arterial entre o fim e o início do estudo. | | | |
| **Droga** | **Média** | **DP** | **Mediana** |
| 1 | 15,58 | 6,09 | 15,49 |
| 2 | 16,82 | 6,37 | 17,43 |

1. Os dados encontrados no arquivo Braga (1998) são provenientes de um estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca. As variáveis medidas durante a realização do teste foram observadas em quatro momentos distintos: repouso (REP), limiar anaeróbio (LAN), ponto de compensação respiratório (PCR) e pico (PICO). As demais variáveis são referentes às características demográficas e clínicas dos pacientes e foram registradas uma única vez.
2. Descreva a distribuição da variável consumo de oxigênio (VO2) em cada um dos quatro momentos de avaliação utilizando medidas resumo (mínimo, máximo, quartis, mediana, média, desvio padrão, etc), *boxplots* e histogramas. Você identifica algum paciente com valores de consumo de oxigênio discrepantes? Interprete os resultados.
3. Descreva a distribuição da classe funcional NYHA por meio de uma tabela de freqüências. Utilize um método gráfico para representar essa tabela.
4. Os dados a seguir são provenientes de um estudo sobre AIDS. Deseja-se comparar os dois grupos de doentes com relação à distribuição das variáveis contínuas **Tempo de diagnóstico** e **Ganho de peso por semana** e com relação à distribuição das varáveis qualitativas nominais doença sexualmente transmissível (**DST)** e método anticoncepcional (**MAC**). Critique a forma de apresentação dos dados e construa uma planilha adequada para a análise estatística dos dados.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grupo I** | | **Tempo de** |  |  | Ganho de Peso |
|  | **registro** | **Diagnóstico** | **DST** | **MAC** | **por Semana** |
|  | 2847111D | pré natal | não | Pílula | 11Kg em 37 semanas |
|  | 3034048F | 6 meses | não | pílula | ? |
|  | 3244701J | 1 ano | não | Condon | ? |
|  | 2943791B | pré natal | não | não | 8 Kg em 39 semanas |
|  | 3000327F | 4 anos | condiloma/ sífilis | não | 9Kg em 39 semanas |
|  | 3232893D | 1 ano | não | DIU | 3Kg em 39 semanas |
|  | 3028772E | 3 anos | não | não | 3 kg em 38 semanas |
|  | 3240047G | pré natal | não | pílula | 9 Kg em 38 semanas |
|  | 3017222G |  | HPV | CONDON | falta exame clínico |
|  | 3015834J | 2 anos | não | condon | 14 Kg em 40 semanas |
| **Grupo II** | | **Tempo de** |  |  | **Ganho de Peso** |
|  | **registro** | **Diagnóstico** | **DST** | **MAC** | **por Semana** |
|  | 3173611E | 3 meses | abscesso ovariano | condon | 15 Kg em 40 semanas |
|  | 3296159D | pré natal | não | condon | 0 Kg em ? semanas |
|  | 3147820D1 | 2 anos | não | sem dados | 4 Kg em 37 semanas |
|  | 3274750K | 3 anos | não | condon | 8 Kg em 38 semanas |
|  | 3274447H | pré natal | sifílis com 3 meses | condon |  |
|  | 2960066D | 5 anos | não | ? | 13 Kg em 36 semanas |
|  | 3235727J | 7 anos | não | Condon | (-) 2 Kg em 38 semanas |
|  | 3264897E |  | condiloma | condon | nenhum Kg |
|  | 3044120J | 5 anos | HPV |  | 3 Kg em 39 semanas |

1. Preencha a ficha de inscrição do Centro de Estatística Aplicada ([www.ime.usp.br/~cea](http://www.ime.usp.br/~cea)) com as informações de um estudo em que você está envolvido.