Segunda avaliação (A2)

Disciplina: Inferência Estatística
Instrutor: Professor Carvalho

03 de Dezembro de 2020

- Por favor, entregue um único arquivo PDF;
- O tempo para realização da prova é de 3 horas, mais vinte minutos para upload do documento para o e-class;
- Leia a prova toda com calma antes de começar a responder;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo ou figura geométrica de sua preferência;
- A prova vale 80 pontos. A pontuação restante é contada como bônus;
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;
- Lembre-se de consultar o catálogo de fórmulas no fim deste documento.

Dicas

- Em uma regressão linear simples, temos:

$$\hat{\beta}_0 \sim \text{Normal}\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right)\right),$$
$$\hat{\beta}_1 \sim \text{Normal}\left(\beta_1, \frac{\sigma^2}{s_x^2}\right),$$
$$\text{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\bar{x}\sigma^2}{s_x^2},$$

  onde $s_x = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$ e $\hat{\beta}_0$ e $\hat{\beta}_1$ são os estimadores de máxima verossimilhança dos coeficientes.

- Um processo de Poisson com taxa λ por unidade de tempo é um processo estocástico que satisfaz:
  - O número de chegadas em um intervalo de tempo Δ_t tem distribuição Poisson com média $\lambda\Delta_t$.
  - Os números de chegadas em qualquer coleção de intervalos disjuntos são independentes.

- Se X tem distribuição Poisson com média λ, então

$$\Pr(X \le x) = Q(\lfloor x + 1 \rfloor, \lambda),$$

  onde

$$Q(x, s) = \frac{\Gamma(x, s)}{\Gamma(x)}$$

  é a função Gama regularizada superior e $\lfloor y \rfloor$ é maior inteiro menor ou igual a y – também chamado de *floor*. Ademais, temos

$$\frac{\partial}{\partial s} Q(x, s) = -\frac{e^{-s} s^{x-1}}{\Gamma(x)},$$

  onde $\Gamma(x) = (x - 1)!$ é a função Gamma.

1. Catando poesia.

Eu te vejo sumir por aí
Te avisei que a cidade era um vão
Dá tua mão, olha pra mim
Não faz assim, não vai lá, não
Os letreiros a te colorir
Embaraçam a minha visão
Eu te vi suspirar de aflição
E sair da sessão frouxa de rir
Já te vejo brincando gostando de ser
Tua sombra a se multiplicar
Nos teus olhos também posso ver
As vitrines te vendo passar
Na galeria, cada clarão
É como um dia depois de outro dia
Abrindo um salão
Passas em exposição
Passas sem ver teu vigia
Catando a poesia
Que entornas no chão

As Vitrines (Almanaque, 1981) de Chico Buarque (1944-).

O eu-lírico da canção, que vamos chamar aqui de Ivo, pensa em seu amado, Adão. Adão é poeta, e tem a estranha mania de deixar cair seus poemas ao passear pelo shopping. Ivo, muito solícito e perdidamente apaixonado, corre atrás do companheiro catando os papéis que o desastrado deixa cair. Sendo estatístico, Ivo sabe que pode modelar o processo de queda dos poemas como um processo de Poisson com média θ. Ivo quer saber se será capaz de acompanhar Adão na sua jornada sem perder nenhum poema. Para isso, julga que se $\theta \leq \theta_0$, ele será capaz de catar toda a poesia deixada por Adão antes de ser carregada pelo vento.

Suponha que Ivo observa o processo de queda dos poemas em n intervalos de exatamente t unidades de tempo e toma nota dos números $Y_1, Y_2, \ldots, Y_n$ de poemas caídos em cada intervalo. Ivo considera a estatística de teste $S = \sum_{i=1}^{n} Y_i$ e constrói o teste δ_c de modo que, se $S \geq c$, ele rejeita a hipótese $H_0 : \theta \leq \theta_0$.

a) (10 pontos) Encontre a função poder do teste de Ivo.

b) (10 pontos) Mostre que a função poder do item anterior é **não-decrescente** em θ;

c) (2,5 pontos) Encontre uma expressão para o tamanho α_0 do teste δ_c;

d) (2,5 pontos) O teste em questão é não-viesado? Justifique;

e) (5 pontos) Discuta se é possível atingir qualquer tamanho para δ_c e o que fazer se queremos um tamanho de, por exemplo, $\alpha_0 = 0.01$.

2.Temos que pegar!

Além de apaixonados um pelo outro, Joelinton e Valcicléia também amam Pokémon. Os dois jogam competitivamente na Liga Brasileira de Pokémon (LBP). Há, contudo, um pequeno incoveniente: Joelinton é *Team Magma* enquanto Valcicléia é *Team Aqua*. Durante uma conversa acalorada, Joelinton afirma que o *Team Magma* é melhor, em termos de *pokescores* médios, que o *Team Aqua*. Valcicléia propõe consultar o site da LBP para obter dados sobre o assunto. Ao consultar o site, eles obtem m valores de *pokescores* de integrantes do *Team Magma* e n valores de integrantes do *Team Aqua*.

Suponha que modelamos os *pokescores* de cada jogador(a) em cada time como variáveis aleatórias normais com médias μ_M e μ_A e variâncias σ_M^2 e σ_A^2, respectivamente. Nos itens a seguir, **enuncie claramente** qual é a hipótese nula – e a hipótese alternativa – em cada caso, qual é a estatística de teste e qual o procedimento de teste.

a) (2,5 pontos) A partir do desenho experimental descrito, encontre quantidades pivotais para μ_M e μ_A, supondo σ_M^2 e σ_A^2 **desconhecidas**. Justifique;

b) (2,5 pontos) Utilize as quantidades do item anterior para construir intervalos de confiança exatos de 99% para μ_A e μ_M;

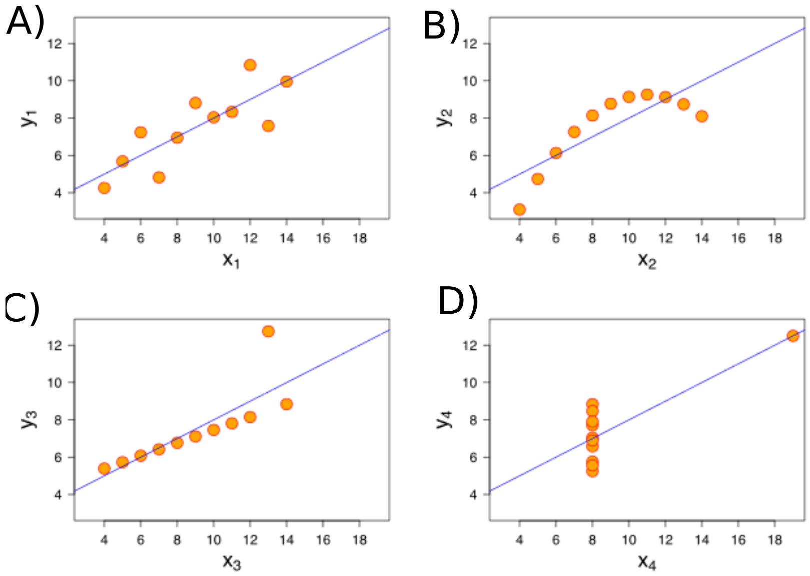
c) (5 pontos) Suponha que, no calor do momento, Valcicléia afirme que o *Team Magma* é tão ruim que não tem pokescore médio suficiente nem para competir na Liga Regional de Pokemon (LRP). Sabendo que o pokescore médio necessário para admissão na LRP é μ_0, mostre a Joelinton como utilizar o intervalo de confiança obtido no item anterior para testar a hipótese levantada por sua amada;

d) (10 pontos) Nossa dupla dinâmica está interessada em comparar as médias supondo que as variâncias são iguais. Proponha um teste de tamanho α_0 para avaliar a premissa de homogeneidade (variâncias iguais);

e) (10 pontos) Suponha que o teste do item anterior falhou em rejeitar H_0. Proponha um teste de tamanho α_0 para testar a hipótese inicial de Joelinton;

3. Regressão linear: o melhor modelo ruim que você já viu.

Considere a figura a seguir:



Em todos os painéis, $\bar{x} = 9$, $\bar{y} = 7,5$, $s_x^2 = 110$ e $\mathrm{Cor}(X, Y) = 0,816$. Isto é, todas as estatísticas sumárias relevantes atingem os mesmos valores. Disto resulta que $\hat{\beta}_0 = 3$ e $\hat{\beta}_1 = 0,5$ para todos os painéis.

a) (15 pontos) Comente sobre quais premissas básicas – ou nenhuma – da regressão linear aparentam estar sendo violadas em cada painel. Justifique.

b) (5 pontos) Os estimadores de máxima verossimilhança para os coeficientes no modelo

$$E[Y_i] = \beta_0 + \beta_1 X_i$$

são

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{y})(X_i - \bar{x})}{\sum_{i=1}^n (X_i - \bar{x})^2}.$$

Tais estimadores são viesados? Justifique.

Dica: Pode ser conveniente escrever

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{x}) Y_i}{s_x^2}.$$

Questão Bônus: uma transformação útil.

Muitas vezes na aplicação de modelos de regressão é conveniente aplicar uma transformação à(s) variável(is) independente(s) de modo a facilitar a computação e/ou a interpretação das estimativas.

a) (10 pontos) Considere uma regressão linear simples. Encontre uma transformação $X' = f(X)$ da variável independente de modo que $\hat{\beta}_0'$ e $\hat{\beta}_1'$ sejam independentes.

b) (5 pontos) Encontre o valor de $\hat{\beta}_0'$ e $\hat{\beta}_1'$ sob a transformação do item anterior.

c) (5 pontos) Como essa transformação muda a interpretação dos coeficientes estimados?

Fórmulas úteis

Como usar este catálogo: as fórmulas dadas aqui estão propositalmente privadas do seu contexto. O objetivo desta coleção é ajudar você a lembrar das expressões. Entretanto, saber quais expressões são utilizadas em que contexto é sua tarefa.

- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$;
- $\hat{\sigma}' = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$;
- $S_X^2 = \sum_{i=1}^{m}(X_i - \bar{X}_m)^2$;
- $S_Y^2 = \sum_{j=1}^{n}(Y_j - \bar{Y}_n)^2$;
- $U = \frac{\sqrt{m+n-2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)(S_X^2 + S_Y^2)}}$;
- $V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$;
- $\bar{x} = (1/n)\sum_{i=1}^{n} X_i$;
- $\bar{y} = (1/n)\sum_{i=1}^{n} Y_i$.