Segunda avaliação (A2)

Disciplina: Inferência Estatística
Professor: Luiz Max Carvalho

29 de Novembro de 2021

- 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;
- Lembre-se de consultar o catálogo de fórmulas no fim deste documento.

Dicas

- Se $Y \sim \text{Poisson}(\lambda)$ então, para λ grande o suficiente, temos
  - $E[\sqrt{Y} + \sqrt{Y+1}] \approx \sqrt{4\lambda + 1}$;
  - $\text{Var}(\sqrt{Y} + \sqrt{Y+1}) \approx 1$.

  Ademais, temos
  $$\frac{Y - \lambda}{\sqrt{\lambda}} \to \text{Normal}(0, 1),$$
  onde a seta representa convergência em distribuição.

- Se X tem distribuição Poisson com média λ, então
  $$\Pr(X \leq x) = Q(\lfloor x + 1 \rfloor, \lambda),$$
  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.

- Se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória, então
  $$\Pr(\max(X_1, X_2, \ldots, X_n) \leq m) = \prod_{i=1}^{n} \Pr(X_i \leq m).$$

- Em uma regressão linear simples, temos:
  $$\begin{aligned} \hat{\beta_0} &\sim \text{Normal}\left(\beta_0, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right)\right), \\ \hat{\beta_1} &\sim \text{Normal}\left(\beta_1, \frac{\sigma^2}{s_x^2}\right), \\ \text{Cov}\left(\hat{\beta_0}, \hat{\beta_1}\right) &= -\frac{\bar{x}\sigma^2}{s_x^2}, \end{aligned}$$
  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.

1. A duck row

Tio Patinhas e seus sobrinhos continuam sua jornada no aprendizado da Estatística e consideram $X_1, X_2, \ldots, X_n \sim \text{Uniforme}(0, \theta)$, $\theta > 0$, uma amostra aleatória com $M := \max(X_1, X_2, \ldots, X_n)$. Várias dúvidas e discordâncias surgiram entre os patinhos, no entanto. No que se segue, você deve utilizar seus conhecimentos de Estatística para esclarecer as coisas.

a) (5 pontos) Huguinho diz que $Y = \theta^2 M$ é pivotal, Zezinho discorda. Mostre que Huguinho está equivocado e encontre uma quantidade pivotal.

b) (5 pontos) Mostre ao Pato Donald como utilizar a quantidade obtida no item anterior para obter um intervalo de confiança unilateral exato de $100\gamma\%$ para θ.

c) (5 pontos) Defina[1] as hipóteses $H_0 : \theta = 1/3$ e $H_1 : \theta > 1/3$. Suponha que propomos um procedimento de teste δ_c que rejeita H_0 quando $M > c$. Deduza a função poder do teste e encontre c de modo que o teste tenha tamanho α_0.

d) (10 pontos) Luisinho cismou que o este teste tem razão de verossimilhanças monótona. Ele está certo? Justifique sua resposta.

e) (10 pontos) Tio Patinhas jura que este teste é uniformemente mais poderoso. Explique para Zezinho o que isso significa e depois discuta se o Tio Patinhas tem razão. Justifique sua resposta.

2. The God(damn) particle

O bóson de Higgs[2], que confere massa a partículas que de outra forma não teriam massa, foi a última partícula do chamado Modelo Padrão (*standard model*) da Física Quântica a ter sua existência confirmada experimentalmente.

A ideia fundamental da descoberta experimental de novas partículas é a de comparar o que se espera sob um modelo base sem a nova partícula e um modelo proposto que a inclui. Uma maneira de fazer isso é detectar o número X de ocorrências (eventos) de decaimento de bósons e comparar esse número com o que é esperado sob o modelo mais simples. Como o medidor é imperfeito, considere o seguinte modelo:

$$X \sim \text{Poisson}(\beta + \kappa\mu),$$

com $\mu > 0$, onde vamos assumir que a taxa base $\beta > 0$ e a taxa esperada do bóson de Higgs $\kappa > 0$ são conhecidas. Desta forma, o modelo com $\mu = 0$ é o modelo de base e aquele onde $\mu = 1$ é o modelo com o bóson de Higgs.

a) (10 pontos) Encontre o EMV de μ e comente sobre a sua adequação; o EMV sempre produz estimativas válidas?

[1] Preste atenção à formulação das hipóteses!
[2] Assim nomeado em homenagem ao físico inglês Peter Higgs (1929-).

b) (10 pontos) Encontre um intervalo de confiança bilateral de $100\gamma\%$ **aproximado** para μ;

   Dica: Você pode assumir que β é grande o suficiente. Use as dicas no começo da prova.

c) (10 pontos) Considere um teste em que assumimos que $X \sim \text{Poisson}(\theta)$ e construímos um teste δ_c de modo que, se $X \geq c$, rejeitamos $H_0 : \theta \leq \theta_0$.

   - Mostre que o teste em questão é não-viesado.
   - Escreva θ_0 em função de β, κ e μ de modo a testar um desvio em relação ao modelo base.

     Dica: Existem várias respostas certas.

d) (10 pontos) Formule um teste para $H_0 : \mu = 0$ vs $H_1 : \mu > 0$ como um teste de razão de verossimilhanças;

3. Gosto muito de você, linearzinho.

O modelo linear (de regressão) é um dos cavalos de batalha da Estatística, sendo aplicado em problemas de Finanças, Medicina e Engenharia. Vamos agora estudar como utilizar as propriedades deste modelo para desenhar experimentos com garantias matemáticas de desempenho e obter estimadores de quantidades de interesse.

a) (5 pontos) Uma prática comum em regressão é a de **centrar** a variável independente (covariável), isto é subtrair a média; isto facilita a interpretação do intercepto e também simplifica alguns cálculos importantes. Mostre que no caso com a covariável centrada, $\hat{\beta}_0$ e $\hat{\beta}_1$ são independentes;

b) (7,5 pontos) Mais uma vez considerando o caso centrado, mostre como obter o número de observações n que faz com que a variância do estimador de máxima verossimilhança do intercepto seja menor que $v > 0$;

c) (5 pontos) Mostre como obter um estimador não-viesado da quantidade $\theta = a\beta_0 + b\beta_1 + c$, com $a, b, c \neq 0$, e encontre o seu erro quadrático médio.

d) (7,5 pontos) Quando $x_{\text{pred}} = \bar{x}$, mostre como obter o número de observações n necessário para que o intervalo de predição de $100(1 - \alpha_0)\%$ para a variável-resposta (Y) tenha largura menor ou igual a $l > 0$ com probabilidade pelo menos γ.

   Dicas:(i) A expressão dependerá *também* da variância dos resíduos, σ^2 e (ii) Você não precisa calcular n, apenas mostrar o procedimento para obtê-lo.

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.

- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$;
- $\hat{\sigma}' = \sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}$;
- $S_X^2 = \sum_{i=1}^m (X_i - \bar{X}_m)^2$;
- $S_Y^2 = \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$;
- $U = \frac{\sqrt{m+n-2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)(S_X^2+S_Y^2)}}$;
- $V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$;
- $\bar{x} = (1/n)\sum_{i=1}^n X_i$;
- $\bar{y} = (1/n)\sum_{i=1}^n Y_i$;
- $$E\left[\left(\hat{Y} - Y\right)^2\right] = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_{\text{pred}} - \bar{x})^2}{s_x^2}\right).$$
- $$\hat{\sigma}'_r := \sqrt{\frac{\sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{n-2}}.$$
- $$\hat{\beta}_0 \pm \hat{\sigma}' c\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}} \quad \text{e} \quad \hat{\beta}_1 \pm c\frac{\hat{\sigma}'}{s_x},$$
  $$\hat{\beta}_0 + \hat{\beta}_1 x_{\text{pred}} \pm c\hat{\sigma}'\sqrt{\frac{1}{n} + \frac{(x_{\text{pred}} - \bar{x})^2}{s_x^2}},$$
- $$\hat{Y} \pm c\hat{\sigma}'_r\sqrt{\left[1 + \frac{1}{n} + \frac{(x_{\text{pred}} - \bar{x})^2}{s_x^2}\right]},$$

  onde $c = T^{-1}(1 - \frac{\alpha_0}{2}; n-2)$.