Introdução à Análise de dados em FAE **(19/09/2021)**

Lista 5: Estatística

Professores: Sandro Souza, Sheila Da Silva, Eliza Da Costa, Dilson Damião *Name:* Vitor Souza Ramos

EXERCICIO 1

Solução: Para $u = u(x, y)$, sabemos que a variância é dada por:

$$\sigma_u^2 = D_u^T V D_u \tag{1.1}$$

Onde D_u é a matriz das derivadas parciais e D_u^T é sua transposta e V é a matriz covariância:

$$D_u = \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{pmatrix}\Bigg|_{(\bar{x},\bar{y})} \tag{1.2}$$

$$D_u^T = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix}\Bigg|_{(\bar{x},\bar{y})} \tag{1.3}$$

$$V = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} \tag{1.4}$$

Assim, encontramos:

$$\sigma_u^2 = \left(\frac{\partial u}{\partial x}\right)^2\Bigg|_{(\bar{x},\bar{y})} \sigma_x^2 + \left(\frac{\partial u}{\partial y}\right)^2\Bigg|_{(\bar{x},\bar{y})} \sigma_y^2 + 2\left(\frac{\partial u}{\partial x}\right)\Bigg|_{(\bar{x},\bar{y})} \left(\frac{\partial u}{\partial y}\right)\Bigg|_{(\bar{x},\bar{y})} \sigma_{xy} \tag{1.5}$$

Para encontrar o erro da média, devemos dividir este valor por N:

$$\sigma_{\bar{u}}^2 = \frac{\sigma_u^2}{N} = \frac{\left(\frac{\partial u}{\partial x}\right)^2\Big|_{(\bar{x},\bar{y})} \sigma_x^2 + \left(\frac{\partial u}{\partial y}\right)^2\Big|_{(\bar{x},\bar{y})} \sigma_y^2 + 2\left(\frac{\partial u}{\partial x}\right)\Big|_{(\bar{x},\bar{y})} \left(\frac{\partial u}{\partial y}\right)\Big|_{(\bar{x},\bar{y})} \sigma_{xy}}{N} \tag{1.6}$$

Portando, contraindo N com as incertezas de x e de y:

$$\sigma_{\bar{u}}^2 = \left(\frac{\partial u}{\partial x}\right)^2\Bigg|_{(\bar{x},\bar{y})} \sigma_{\bar{x}}^2 + \left(\frac{\partial u}{\partial y}\right)^2\Bigg|_{(\bar{x},\bar{y})} \sigma_{\bar{y}}^2 + \frac{2}{N}\left(\frac{\partial u}{\partial x}\right)\Bigg|_{(\bar{x},\bar{y})} \left(\frac{\partial u}{\partial y}\right)\Bigg|_{(\bar{x},\bar{y})} \sigma_{xy} \tag{1.7}$$

EXERCICIO 2

Item 1: Para o caso $u = x \pm y$, deduzir:

$$\sigma_{\bar{u}} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 + 2r\sigma_{xy}} \tag{2.1}$$

Solução: Da equação do erro padrão:

$$\sigma_{\bar{u}}^2 = \left(\frac{\partial f}{\partial x}\right)^2\Bigg|_{\bar{x},\bar{y}} \sigma_{\bar{x}}^2 + \left(\frac{\partial f}{\partial y}\right)^2\Bigg|_{\bar{x},\bar{y}} \sigma_{\bar{y}}^2 + \frac{2}{N}\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right)\Bigg|_{\bar{x},\bar{y}} \sigma_{xy} \tag{2.2}$$

Para este caso temos $\frac{\partial f}{\partial x} = 1$, $\frac{\partial f}{\partial y} = \pm 1$, portanto

$$\sigma_{\bar{u}}^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 \pm \frac{2}{N}\sigma_{xy} \tag{2.3}$$

Do coeficiente de correlação linear de Pearson, temos:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \Rightarrow \sigma_{xy} = r\sigma_x\sigma_y \tag{2.4}$$

Assim, temos

$$\sigma_{\bar{u}}^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 \pm 2r\frac{\sigma_x}{\sqrt{N}}\frac{\sigma_y}{\sqrt{N}} \tag{2.5}$$

Sabendo que $\frac{\sigma_x}{\sqrt{N}} = \sigma_{\bar{x}}$, temos

$$\sigma_{\bar{u}} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 \pm 2r\sigma_{\bar{x}}\sigma_{\bar{y}}} \tag{2.6}$$

Item 2: Para os casos $u = xy$ e $u = x/y$, deduzir:

$$\frac{\sigma_{\bar{u}}}{|\bar{u}|} = \sqrt{\left(\frac{\sigma_{\bar{x}}}{\bar{x}}\right)^2 + \left(\frac{\sigma_{\bar{y}}}{\bar{y}}\right)^2 \pm 2r\left(\frac{\sigma_{\bar{x}}}{\bar{x}}\right)\left(\frac{\sigma_{\bar{y}}}{\bar{y}}\right)} \tag{2.7}$$

Para o caso $u = xy$, partindo da equação (2.2) e usando $\frac{\partial f}{\partial x} = y$, $\frac{\partial f}{\partial y} = x$, temos

$$\sigma_{\bar{u}}^2 = \bar{y}^2\sigma_{\bar{x}}^2 + \bar{x}^2\sigma_{\bar{y}}^2 \pm \frac{2}{N}\sigma_{xy}\bar{x}\bar{y} \tag{2.8}$$

dividindo esta equação por $(\bar{x}\bar{y})^2 = \bar{u}^2$ e usando a equação (2.4), temos:

$$\frac{\sigma_{\bar{u}}^2}{\bar{u}^2} = \left(\frac{\sigma_{\bar{x}}}{\bar{x}}\right)^2 + \left(\frac{\sigma_{\bar{y}}}{\bar{y}}\right)^2 + 2r\frac{\sigma_{\bar{x}}}{\bar{x}}\frac{\sigma_{\bar{y}}}{\bar{y}} \tag{2.9}$$

Tirando a raiz desta equação encontramos a equação do problema.

Para o caso $u = x/y$, partindo da equação (2.2) e usando $\frac{\partial f}{\partial x} = \frac{1}{y}$, $\frac{\partial f}{\partial y} = \frac{-x}{y^2}$, temos

$$\sigma_{\bar{u}}^2 = \frac{1}{\bar{y}^2}\sigma_{\bar{x}}^2 + \left(\frac{-\bar{x}}{\bar{y}^2}\right)^2\sigma_{\bar{y}}^2 \pm \frac{2}{N}\sigma_{xy}\frac{-\bar{x}}{\bar{y}^3} \tag{2.10}$$

dividindo esta equação por $(\bar{x}/\bar{y})^2 = \bar{u}^2$ e usando a equação (2.4), temos:

$$\frac{\sigma_{\bar{u}}^2}{\bar{u}^2} = \left(\frac{\sigma_{\bar{x}}}{\bar{x}}\right)^2 + \left(\frac{\sigma_{\bar{y}}}{\bar{y}}\right)^2 - 2r\frac{\sigma_{\bar{x}}}{\bar{x}}\frac{\sigma_{\bar{y}}}{\bar{y}} \tag{2.11}$$

Tirando a raiz desta equação encontramos a equação do problema.

EXERCICIO 3

3.7.1 - Solução:
a) Este intervalo é entre x e $x \pm \sigma$. Portanto, representa 68,3% dos dados totais.
b) Este intervalo é entre x e $x - 2\sigma$. Portanto, 47,8% dos dados.
c) Este intervalo é entre $x + \sigma$ e $x + 2\sigma$, portanto representa $(47,8 - 34,2) = 13,6\%$ dos dados.

3.7.2 - Solução: A estimativa padrão é dada por $\bar{x} \pm \sigma_{\bar{x}}$ (média mais ou menos erro da média). Assim, a melhor estimativa para os dados é: $g = (9,734 \pm 0.050)$ m/s^2. Cálculos disponíveis no notebook.

3.7.3 - Solução: Seguindo a solução do exercício anterior, a estimativa padrão é de $(1,714 \pm 0,028)$ V. Cálculos disponíveis no notebook.

3.7.4 - Solução: Para encontrar o erro padrão de uma bateria de medidas, dividimos o desvio padrão pelo número de medidas, seguindo a equação:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \tag{3.1}$$

Assim, isolando N temos:

$$N = \left(\frac{\sigma_x}{\sigma_{\bar{x}}}\right)^2 = 11,111... \tag{3.2}$$

Portanto, precisamos de aproximadamente 11 medidas para que o erro seja da ordem de 3 m/s.

3.7.5 - Solução: As medidas são compatíveis quando $|x_{\text{ref}} - \bar{x}| < 2\sigma_{\bar{x}}$. Assim, para os valores informados, somente a terceira estimativa é satisfatória. Cálculos disponíveis no notebook.

3.7.6 - Solução: Para que as estimativas sejam compatíveis entre si, seus intervalos de confiança de 65% devem se interceptar. Assim, será compatível se:

$$|m_1 - m_2| < \sigma_1 + \sigma_2 \tag{3.3}$$

Para este caso, temos $|7,8 - 7,0|kg > (0,2 + 0,3)kg$, portanto as medidas não são compatíveis entre si, e não devem representar a mesma partícula.

3.7.7 - Solução:
a) A estimativa padrão é de $1,90 \pm 0,028$ g/cm^3. Cálculos disponíveis no notebook.
b) As medidas são compatíveis. Cálculos disponíveis no notebook.

3.7.8 - Solução: Como no exercício 3.7.6, os intervalos de confiança devem se interceptar para que as medidas sejam compatíveis. Temos $|75 - 60|MeV > (3 + 9)MeV$, portanto as medidas não são compatíveis, e a discrepância é significativa.
3.7.9 - Solução: A combinação de resultados é feita seguindo as equações:

$$\bar{x} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{3.4}$$

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}} \tag{3.5}$$

Assim, o resultado combinado é de $m_{top} = (177,916 \pm 4,036)$ GeV/c^2. Cálculos disponíveis no notebook.

3.7.10 - Solução: O valor encontrado pelo estudante não é compatível com a referência. Cálculos no notebook.

3.7.11 - Solução: A estimativa padrão para os valores encontrados é de $(1,0186 \pm 0.0037)$ V. Cálculos no notebook.

EXERCICIO 4

Problema resolvido no notebook.

EXERCICIO 5

Exercício não resolvido

EXERCICIO 6

Exercício não resolvido

EXERCICIO 7

Solução:
Existem 2587 alunos e 2852 alunas, totalizando 5419 alunos. Nos cursos de tecnologia existem 1291 alunos e 547 alunas. Assim a probabilidade de se escolher um aluno do grupo de tecnologia é de $\frac{1291}{5419}$, em torno de 24%.

EXERCICIO 8

Solução:

Temos:

$$P(A) = 0,15 \Rightarrow \text{ Prob. do táxi ser azul} \tag{8.1}$$
$$P(V) = 0,85 \Rightarrow \text{ Prob. do táxi ser verde} \tag{8.2}$$
$$P(T|A) = 0,8 \Rightarrow \text{ Prob. da testemunha dizer que é azul dado que o táxi é azul} \tag{8.3}$$
$$\tag{8.4}$$

Pode-se inferir que:

$$P(T|V) = 0,2 \Rightarrow \text{ Prob. da testemunha dizer que é azul dado que o táxi é verde} \tag{8.5}$$

Assim, usando o fator de Bayes:
$$\frac{P(A|T)}{P(V|T)} = \frac{P(T|A)}{P(T|V)}\frac{P(A)}{P(V)} \tag{8.6}$$

Encontramos:
$$\frac{P(A|T)}{P(V|T)} = \frac{0,8}{0,2}\frac{0,15}{0,85} = \frac{12}{17} \tag{8.7}$$

Sabendo que $P(A|T) + P(V|T) = 1$, temos:

$$P(A|T) = \frac{12}{12+17} \approx 41\%. \tag{8.8}$$

EXERCICIO 9

Solução:
Temos:

$$P(C) = 0,008 \Rightarrow \text{ Prob. de haver câncer} \tag{9.1}$$
$$P(S) = 0,992 \Rightarrow \text{ Prob. de ser saudável} \tag{9.2}$$
$$P(P|S) = 0,07 \Rightarrow \text{ Prob. de teste positivo dado que é saudável} \tag{9.3}$$
$$P(N|C) = 0,1 \Rightarrow \text{ Prob. de teste negativo dado que há câncer} \tag{9.4}$$
$$\tag{9.5}$$

Assim, usando o fator de Bayes:
$$\frac{P(C|P)}{P(S|P)} = \frac{P(P|C)}{P(P|S)}\frac{P(C)}{P(S)} \tag{9.6}$$

Encontramos:
$$\frac{P(A|T)}{P(V|T)} = \frac{0,9}{0,07}\frac{0,008}{0,992} = \frac{45}{434} \tag{9.7}$$

Sabendo que $P(C|P) + P(S|P) = 1$, temos:

$$P(C|P) = \frac{45}{45+434} \approx 9,4\%. \tag{9.8}$$

EXERCICIO 10

Solução:

$$P(III) = \frac{8}{29} \Rightarrow \text{ Prob. de vir da urna III} \tag{10.1}$$
$$P(P) = \frac{16}{29} \Rightarrow \text{ Prob. de ser preta} \tag{10.2}$$
$$P(P|III) = \frac{4}{8} \Rightarrow \text{ Prob. de ser preta dado que veio da urna III} \tag{10.3}$$
$$\tag{10.4}$$

Usando o teorema o Teorema de Bayes, temos

$$P(III|P) = \frac{P(P|III) \times P(III)}{P(P)} = \frac{\frac{4}{8}\frac{8}{29}}{\frac{16}{29}} = 25\% \tag{10.5}$$

EXERCICIO 11

Solução:
Assumindo uma região de dimensão a entre $-a/2$ e $a/2$, o valor médio da posição é dado por:

$$< x >= \int_{-a/2}^{a/2} x \sin^2\left(\frac{\pi x}{a}\right) dx = 0 \tag{11.1}$$

O valor médio quadrático da posição é dado por:

$$< x^2 >= \int_{-a/2}^{a/2} x^2 \sin^2\left(\frac{\pi x}{a}\right) dx = \frac{6+\pi^2}{24\pi^2} a^3 \tag{11.2}$$

Assim, o desvio padrão é dado por:

$$\sigma_x = \sqrt{< x^2 > - < x >^2} = \sqrt{\frac{6+\pi^2}{24\pi^2} a^3} \tag{11.3}$$

EXERCICIO 12

Solução:
Usando a função massa de probabilidade para uma distribuição binomial, temos:

$$B(k|n;p) = \frac{n!}{k!(n-k)!} \times (p)^k \times (1-p)^{n-k} \tag{12.1}$$

Assim, para o caso de 3 sensores com 3 acertos, temos:

$$B(3|3;0,6) = \frac{3!}{3!(3-3)!} \times (0,6)^3 \times (0,4)^{3-3} = 1 \times 0,216 \times 1 = 21,6\% \tag{12.2}$$

Esta equação dá a probabilidade de que 3 sensores sejam atingidos quando existem 3 sensores. Para que haja uma detecção são necessários **ao menos** 3 sensores, então para o caso de 4 sensores, devemos somar as chances de se atingirem 3 sensores e de se atingirem 4:

$$P(4) = B(3|4;0,6) + B(4|4;0,6) \tag{12.3}$$

Temos:

$$P(4) = \frac{4!}{3!(4-3)!} \times (0,6)^3 \times (0,4)^{4-3} + \frac{4!}{4!(4-4)!} \times (0,6)^4 \times (0,4)^{4-4} = 0,3456 + 0,1296 \approx 48\% \tag{12.4}$$

Igualmente, para o caso de 5 sensores, temos:

$$P(3) = B(3|5;0,6) + B(4|5;0,6) + B(5|5;0,6) \tag{12.5}$$

Assim:

$$P(5) = \frac{5!}{3!(5-3)!} \times (0,6)^3 \times (0,4)^{5-3} + \frac{5!}{4!(5-4)!} \times (0,6)^4 \times (0,4)^{5-4} + \frac{5!}{5!(5-5)!} \times (0,6)^5 \times (0,4)^{5-5} \approx 68\% \tag{12.6}$$

EXERCICIO 13

Solução: Para que duas pessoas façam aniversário no mesmo dia, a primeira pode fazer em qualquer dia do ano, e a segunda só pode fazer em um dos 365 disponíveis. Assim, a probabilidade é 1/365. Em um conjunto de N pessoas, o número de pares de pessoas é:

$$\binom{N}{2} = \frac{N!}{2!(N-2!)} \tag{13.1}$$

Pela distribuição binomial, a probabilidade de duas dentre N pessoas fazerem aniversário no mesmo dia é dada por:

$$B(2|N;1/365) = \frac{N!}{2!(N-2)!} \times (1/365)^2 \times (364/365)^{N-2} \tag{13.2}$$

Podemos também usar a aproximação de Poisson para a distribuição binomial usando $\lambda = N \times 1/365$:

$$f(2;N/365) = e^{-N/365}\frac{(N/365)^2}{2!} \tag{13.3}$$

Assim, usando $N = 720$ nas duas equações acima, temos:

$$B(2|720;1/365) = 0,27995 \tag{13.4}$$
$$f(2;720/365) = 0,27062 \tag{13.5}$$

EXERCICIO 14

Exercício feito no notebook.

EXERCICIO 15

Exercício feito no notebook.

EXERCICIO 16

Solução: Assumindo uma distribuição gaussiana com $\mu = 170$ e $\sigma = 5$, a função densidade de probabilidade tem a forma:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{16.1}$$

Assim, ao integrar esta densidade de probabilidade de $-\inf$ a x' e forçar que o resultado seja 0,9, encontramos o valor que divide os 10% mais altos dos 90% mais baixos. Integrando, temos:

$$\frac{1}{2}\left[1 + \text{erf}\left(\frac{x' - 170}{5\sqrt{2}}\right)\right] = 0,9 \tag{16.2}$$

Assim, encontramos $x' = 176,41$.

EXERCICIO 17

Exercício não resolvido

EXERCICIO 18

Exercício feito no notebook.