Final exam
January 19, 2008
Statistics II

1. (6 points) Imagine um pequeno país servido por dois aeroportos — o aeroporto A e o aeroporto B. O número médio de aviões que aterra por hora, aos fins de semana, no aeroporto A é 4. O número médio de aviões que aterra por dia, aos fins de semana no aeroporto B é 48.

   Sabe-se que o número de aviões que aterram, seja no aeroporto A seja no aeroporto B, obedece a uma lei de Poisson. Sabe-se ainda que o número de aviões que aterra em A é independente do número de aviões que aterra em B.

   (a) Hoje, sábado, a família Feliz está no aeroporto A, com o seu filho mais novo, para lhe mostrar os aviões. Sobem ao terraço panorâmico e decidem aguardar pela chegada do próximo avião. Qual a probabilidade de terem de esperar mais de 20 minutos para verem um avião aterrar?

   Solution:

   (b) E qual a probabilidade de aterrarem em todo o país, e durante o próximo fim de semana (Sábado e Domingo) mais de 300 aviões?

   Solution:

   No terraço panorâmico de aeroporto existe um pequeno bar explorado pelo Sr. Gordo. A probabilidade de uma qualquer pessoa que se dirija ao bar pedir uma "tosta aérea", especialidade da casa, é sempre de 10%. Chega ao bar um grupo de 20 pessoas — todas as pessoas do grupo começam a fazer os seus pedidos ao mesmo tempo o que gera uma grande confusão. O Sr. Gordo propõe então que se organizem e escrevam num papel o seu pedido. O grupo aceita a sugestão mas diz: "- Sr Gordo pode desde já começar a fazer 2 "tostas aéreas" porque pelo menos duas vão ser pedidas."

   (c) Qual a probabilidade de serem pedidas ao todo 4 "tostas aéreas"?

   Solution:

   Entretanto o grupo vai-se embora e o Sr. Gordo suspira de alívio. Mas eis que mais clientes se aproximam do bar.

   (d) Qual a probabilidade do oitavo cliente a ser a atendido ser o segundo a pedir uma "tosta aérea"?

   Solution:

2. (5 points) Considere uma variável aleatória X com média μ e variância σ^2. Considere os seguintes estimadores para μ obtidos a partir de uma amostra aleatória de dimensão N:

$$\hat{\mu}_1 = \overline{X} + \frac{80}{N} \qquad \hat{\mu}_2 = \overline{X}\frac{N}{N-1}.$$

(a) Mostre que ambos os estimadores são enviesados para μ.

Solution:

$$\begin{aligned}\mathbb{E}(\hat{\mu}_1) &= \mathbb{E}\left(\overline{X}\right) + \frac{80}{N} = \mu + \frac{80}{N} \neq \mu \\ \mathbb{E}(\hat{\mu}_2) &= \mathbb{E}\left(\overline{X}\right)\frac{N}{N-1} = \mu\frac{N}{N-1} \neq \mu\end{aligned}$$

(b) Qual o estimador que tem um maior enviesamento (bias) se $N = 20$ e $\mu = 70$?

Solution:

$$\begin{aligned}\text{bias}(\hat{\mu}_1) &= \mathbb{E}(\hat{\mu}_1) - \mu = \frac{80}{N} = 4 \\ \text{bias}(\hat{\mu}_2) &= \mathbb{E}(\hat{\mu}_2) - \mu = \mu\left(\frac{N}{N-1} - 1\right) = \frac{\mu}{N-1} = 3.6842\end{aligned}$$

Assim, $\text{bias}(\hat{\mu}_2) < \text{bias}(\hat{\mu}_1)$.

(c) Obtenha o erro quadrático médio para ambos os estimadores. Se $N = 20$, $\mu = 70$ e $\sigma^2 = 100$, por qual dos dois estimadores optaria?

Solution:

$$\text{MSE}(\hat{\mu}_1) = \mathbb{V}\text{ar}(\hat{\mu}_1) + [\text{bias}(\hat{\mu}_1)]^2$$

A variância de $\hat{\mu}_1$ é

$$\mathbb{V}\text{ar}(\hat{\mu}_1) = \mathbb{V}\text{ar}\left(\overline{X}\right) = \frac{\sigma^2}{N} = \frac{100}{20} = 5$$

Então, $\text{MSE}(\hat{\mu}_1) = 5 + 4^2 = 21$.

$$\mathbb{V}\text{ar}(\hat{\mu}_2) = \frac{N^2}{(N-1)^2}\mathbb{V}\text{ar}\left(\overline{X}\right) = \frac{20^2}{19^2} \cdot 5 = 5.54$$

Então, $\text{MSE}(\hat{\mu}_2) = 5.54 + 3.6842^2 = 19.11$.
Optaria pelo estimador $\hat{\mu}_2$ pois, $\text{MSE}(\hat{\mu}_2) < \text{MSE}(\hat{\mu}_1)$.

3. (5 points) Nos EUA uma cadeia de *fast-food* testa diariamente se o peso médio dos seus *hamburgers*, conhecidos por *"two-pounders"*, é de **pelo menos** 907.17 gramas. Se esta hipótese for rejeitada, então haverá a necessidade ajustar o processo de produção. O peso destes *hamburgers* segue uma distribuição normal com um desvio-padrão de 85.05

gramas. A regra de decisão adoptada é rejeitar a hipótese nula se o peso médio de cada um destes *hamburgers* na amostra for inferior a 873.15 gramas.

(a) Se for seleccionada uma amostra de dimensão 16, qual a probabilidade de rejeitar incorrectamente a hipótese nula?

Solution: X: peso dos hamburgers (em gramas) $\Longrightarrow X \sim \mathcal{N}\left(\mu, \sigma^2 = 85.05^2\right)$
O objectivo é calcular a probabilidade de cometer o erro tipo I (α)

$$\begin{aligned} H_0 &: \mu = 907.17 \\ H_1 &: \mu < 907.17 \end{aligned}$$

Sabemos que

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$$

A região crítica é $RC = \left\{\overline{X} : \overline{X} < 873.15\right\}$. Então,

$$\begin{aligned} \alpha &= \mathbb{P}\left(\text{rej. } H_0 \mid H_0 \text{ verd.}\right) = \mathbb{P}\left(\overline{X} < 873.15 \mid \mu = 907.17\right) = \\ &= \mathbb{P}\left(Z < \frac{873.15 - 907.17}{85.05/\sqrt{16}}\right) = \mathbb{P}\left(Z < -1.6\right) = \\ &= 1 - 0.9452 = 0.0548 \end{aligned}$$

(b) Suponha que o verdadeiro peso médio dos *hamburgers* é de 878.82 gramas. Se for seleccionada uma amostra aleatória de 36 *"two-pounders"*, qual é a probabilidade de tomarmos uma decisão errada?

Solution: O objectivo é agora calcular a probabilidade de cometer o erro tipo II (β).

$$\begin{aligned} \beta &= \mathbb{P}\left(\text{não rej. } H_0 \mid H_0 \text{ falsa}\right) = \mathbb{P}\left(\overline{X} > 873.15 \mid \mu = 878.82\right) = \\ &= \mathbb{P}\left(Z > \frac{873.15 - 878.82}{85.05/\sqrt{36}}\right) = \mathbb{P}\left(Z > -0.4\right) = 0.6554 \end{aligned}$$

(c) Admita, agora, que o verdadeiro valor da variância do peso dos *hamburgers* não é conhecido e que para uma amostra de dimensão 9 se obteve $\sum x_i = 7920$ e $\sum x_i^2 = 7034400$. Com base nestes valores alguém afirmou que "o verdadeiro peso médio dos *hamburgers* se situa entre 810 e 949 gramas". Concorda? Justifique.

Solution: A variável fulcral é

$$T = \frac{\overline{X} - \mu}{S'/\sqrt{N}} \sim \mathrm{t}_{N-1}$$

Intervalo de confiança a 95% para μ é

$$[\mu]_{95\%} = \left[\overline{X} \pm t_{N-1;\alpha/2} \frac{S'}{\sqrt{N}}\right]$$

Agora $\alpha = 0.05 \iff t_{8;0.025} = 2.306$, $N = 9$,

$$\begin{aligned}
\overline{X} &= \frac{\sum_{i=1}^{9} X_i}{9} = \frac{7920}{9} = 880 \\
S' &= \sqrt{\frac{\sum_{i=1}^{9} X_i^2 - N \cdot \overline{X}^2}{N-1}} = 90
\end{aligned}$$

Então,

$$[\mu]_{95\%} = \left[880 \pm 2.306 \frac{90}{\sqrt{9}}\right] = [810.82; 949.18]$$

4. (4 points) Diga se as seguintes afirmações são verdadeiras ou falsas. Justifique as suas respostas.

   (a) O número de acidentes num fim de semana de 2 dias numa determinada estrada segue uma distribuição de Poisson com média 3. Perante esta informação, a probabilidade do número médio de acidentes num ano (52 fins de semana) ser maior que 4 é aproximadamente 0.25.

Solution: X = número de acidentes num fim de semana de 2 dias numa determinada estrada

$X \sim \mathcal{P}(3)$

$Y = \sum\limits_{i=1}^{52} X_i = 52 \cdot \overline{X}$ número de acidentes num ano

$Y \sim \mathcal{P}(52 \cdot 3 = 156) \Longrightarrow Y \overset{a}{\sim} \mathcal{N}(156, 156)$

$$\begin{aligned}
\mathbb{P}\left(\overline{X} > 4\right) &= \mathbb{P}(Y > 4 \cdot 52) = 1 - \mathbb{P}(Y \leq 208) = \\
&= 1 - \Phi\left(\frac{208 - 156 + 0.5}{\sqrt{156}}\right) = 1 - \Phi(4.2) \approx 0 \neq 0.25
\end{aligned}$$

Assim, a afirmação é **FALSA**.

(b) Um departamento de polícia municipal registou ao longo de 60 dias o número de acidentes de tráfego nos quais houve danos pessoais. Os resultados obtidos foram os seguintes:

Número de acidentes	0	1	2	≥ 3
Frequência observada	17	17	16	10

Então, num determinado dia o número de acidentes de tráfego nos quais houve danos pessoais é uma variável aleatória com distribuição Poisson. Use um nível de significância de 5%.

Solution: X = número de acidentes de tráfego

Para calcular os valores esperados, precisamos da média da distribuição de Poisson. Sabemos que o estimador de máxima verosimilhança (e do método de momentos) de λ é a média amostral. Então,

$$\hat{\lambda} = \overline{X} = \frac{0 \cdot 17 + 1 \cdot 17 + 2 \cdot 16 + 3 \cdot 10}{60} = 1.32$$

$H_0 : X \sim \mathcal{P}\,(1.32)$

H_1 : não H_0

$$\begin{aligned}
\mathbb{P}\,(X = 0) &= e^{-1.32}\frac{1.32^0}{0!} = 0.2671 \implies \hat{E}_0 = 60 \cdot 0.2671 = 16.03 \\
\mathbb{P}\,(X = 1) &= e^{-1.32}\frac{1.32^1}{1!} = 0.3526 \implies \hat{E}_1 = 60 \cdot 0.3526 = 21.16 \\
\mathbb{P}\,(X = 2) &= e^{-1.32}\frac{1.32^2}{2!} = 0.2327 \implies \hat{E}_2 = 60 \cdot 0.2327 = 13.96 \\
\mathbb{P}\,(X \geq 3) &= 1 - 0.2671 - 0.3526 - 0.2327 = 0.1475 \Rightarrow \hat{E}_{\geq 3} = 8.85
\end{aligned}$$

Então, a tabela de valores observados e esperados é

Número de acidentes	0	1	2	≥ 3
O_i	17	17	16	10
$\hat{E}_i$	16.03	21.16	13.96	8.85

A estatística de teste é

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 1.3241$$

O valor crítico é $\chi^2_{4-1-1;0.05} = 5.99 > 1.3241 \implies$ não se rejeita H_0

Assim, a afirmação é **VERDADEIRA**.

(c) Num teste de hipótese, com região crítica bilateral, o p-value foi 0.0653. Então, podemos rejeitar a hipótese nula para um nível de significância de 5%, mas não rejeitamos para um nível de significância de 10%.

Solution: p-value $= 0.0653 > 0.05 \Longrightarrow$ não se rejeita H_0 a 5%
p-value $= 0.0653 < 0.10 \Longrightarrow$ rejeita-se H_0 a 10%
Assim, a afirmação é **FALSA**.