Final exam
June 25, 2007
Statistics II

1. (7 points) No Hospital "Medecis", o número de doentes que recorrem ao serviço de urgências pediátricas e o número de doentes que recorrem ao serviço de urgências para adultos, no período da noite (20h até às 8h do dia seguinte), são variáveis com distribuição de Poisson e independentes entre si.

   Sabe-se que, em média, entram 4 crianças doentes por noite nas urgências pediátricas. Sabe-se, ainda, que, em média, entram 6 doentes adultos por noite nas urgências de adultos.

   (a) Qual a probabilidade de serem atendidos mais de 50 adultos no serviço de urgências numa semana (7 noites)?

   Solution: Seja A o número de doentes adultos que recorrem as serviço de urgências durante a noite $\Longrightarrow A \sim \mathcal{P}(6)$.
   Seja C o número de doentes crianças que recorrem as serviço de urgências durante a noite $\Longrightarrow C \sim \mathcal{P}(4)$.
   Seja CA o número de doentes adultos que recorrem as serviço de urgências durante a noite durante 7 dias $\Longrightarrow A \sim \mathcal{P}(6 \cdot 7 = 42)$.
   Como $\lambda = 42 > 5$, aproximamos a distribuição Poisson com a distribuição normal $\Longrightarrow CA \overset{a}{\sim} \mathcal{N}(42, 42)$. Então, tendo em conta a correcção de continuidade,

   $$\begin{aligned} \mathbb{P}(CA > 50) &= 1 - \mathbb{P}(CA \leq 50) = 1 - \mathbb{P}\left(Z < \frac{50.5 - 42}{\sqrt{42}}\right) = \\ &= 1 - \mathbb{P}(Z < 1.31) = 1 - \Phi(1.31) = 1 - 0.9049 = 0.0951 \end{aligned}$$

   (b) Qual a probabilidade de numa determinada noite a enfermeira do serviço de triagem ter de esperar mais de uma hora entre a chegada consecutiva de quaisquer dois doentes?

   Solution: Seja D o número de doentes (adultos e crianças) que chegam à noita às urgências $\Longrightarrow D = C + A \sim \mathcal{P}(6 + 4 = 10)$
   T = tempo de espera entre chegados de dois doentes consecutivos $\Longrightarrow T \sim \exp(10)$ para 12 horas.
   Então,

   $$\mathbb{P}\left(T > \frac{1}{12}\right) = e^{-\frac{10}{12}} = 0.4346$$

   (c) Qual a probabilidade de numa noite serem atendidas 5 crianças nas urgências sabendo que entraram 9 doentes para as urgências?

Solution:

$$\mathbb{P}(C=5 \mid D=9)=\frac{\mathbb{P}(C=5\cap D=9)}{\mathbb{P}(D=9)}=\frac{\mathbb{P}(C=5\cap A=4)}{\mathbb{P}(D=9)}$$

Pela independência de A e C,

$$\mathbb{P}(C=5\cap A=4)=\mathbb{P}(C=5)\cdot\mathbb{P}(A=4)=0.1563\cdot 0.1339=0.0209$$

Além disso, como $D\sim\mathcal{P}(10)$, $\mathbb{P}(D=9)=0.1251$. Assim,

$$\mathbb{P}(C=5 \mid D=9)=\frac{0.0209}{0.1251}=0.1673$$

Sabe-se também que a probabilidade de um doente entrar para as urgências, durante a noite, em estado considerado grave é de 5%.

(d) Sabendo que num determinado fim-de-semana foram atendidos durante a noite 25 doentes nas urgências, qual a probabilidade de os casos considerados graves não terem ultrapassado os 2 casos?

Solution: X: número de doentes em caso grave em 25 doentes $\Longrightarrow X\sim \mathrm{b}(25, 0.05)$

$$\mathbb{P}(X\leq 2)=0.8729$$

Sabe-se, ainda, que a probabilidade de um doente que recorre às urgências durante a noite não ter que ficar internado é de 70%.

(e) Qual a probabilidade de, numa determinada noite, o sétimo doente a ser observado nas urgências ser o primeiro a necessitar de internamento?

Solution: Y: número de doentes a observar até obter um internamento $\Longrightarrow Y\sim \mathrm{bn}(1, 0.3)$

$$\mathbb{P}(Y=7)=0.7^6\cdot 0.3=0.0353$$

2. (6 points) Há alguns anos foi realizado um estudo no qual se apurou que 62% dos passageiros que entram na estação ABC do metro tem como destino o centro da cidade. Um dos administradores do Metro, tendo dúvidas quanto à actualidade deste valor, resolveu fazer um inquérito na referida estação. De um total de 240 passageiros inquiridos, cerca de 126 tinham como destino o centro da cidade.

   (a) Obtenha um intervalo de confiança a 90% para a percentagem de passageiros com origem na estação ABC e destino no centro da cidade. Interprete o resultado obtido.

Solution: X: número de passageiros que entram na estação ABC e têm como destino o centro da cidade, em 240 $\Longrightarrow X \sim \mathrm{b}(240, p)$
O parâmetro a estimar é p. A variável fulcral é, pelo teorema do límite central,

$$Z = \frac{f_N - p}{\sqrt{\frac{p(1-p)}{N}}} \xrightarrow{d} \mathcal{N}(0,1)$$

e $1-\alpha = 0.9 \iff \alpha = 0.1 \iff z_{\alpha/2} = z_{0.05} = 1.645$, e $f_N = \dfrac{126}{240} = 0.525$.
O intervalo de confiança é

$$\begin{aligned} [p]_{90\%} &= \left[f_N \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{N}} \right] = \left[0.525 \pm 1.645\sqrt{\frac{0.525 \cdot 0.475}{240}} \right] = \\ &= [0.4719; 0.5780] \end{aligned}$$

Com um nível de confiança de 90%, a verdadeira proporção de passageiros que entram na estação ABC e têm como destino o centro da cidade fica entre 47.19% e 57.8%. O valor apurado no passado parece, de facto, estar desactualizado.

(b) Mais uma vez para um nível de confiança de 90%, quantos passageiros deveriam ser inquiridos para que a percentagem de passageiros referida em (a) fosse estimada com uma margem de erro máxima de 2%.

Solution: A amplitude do intervalo é

$$\Delta = 2 \cdot 1.645\sqrt{\frac{p(1-p)}{N}}$$

com $1 - \alpha = 0.9$.
A margem de erro é $e = \Delta/2$. Queremos que $e = 0.02$ e, fazendo $p = 0.5$,

$$N = \frac{1.645^2 \cdot 0.5 \cdot 0.5}{0.02^2} = 1691.266 \approx 1692$$

Teriam de ser inquiridos 1692 passageiros.

Seja X uma variável aleatória com distribuição normal com variância igual a 4. Para testar as hipóteses

$$\begin{aligned} H_0 &: \mu = 1 \\ H_1 &: \mu = 2 \end{aligned}$$

decidiu-se rejeitar a hipótese nula se $\overline{X} > c$.

(a) Para uma amostra de dimensão 25 determine c de modo a que $\alpha = 0.1$.

Solution: Sabemos que $X \sim \mathcal{N}\left(\mu, \sigma^2 = 4\right)$, e o parâmetro a testar é μ.

$$\begin{aligned} H_0 &: \mu = 1 \\ H_1 &: \mu = 2 \end{aligned}$$

Além disso, $\alpha = \mathbb{P}\,(\text{erro tipo I}) = 0.1$, e

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

Então,

$$\mathbb{P}\left(\overline{X} > c \mid H_0\right) = 0.1 \quad \Longleftrightarrow \quad \mathbb{P}\left(Z > \frac{c-1}{2/\sqrt{25}}\right) = 0.1$$

pelo que $\dfrac{c-1}{2/5} = 1.282 \iff c = 1.5128.$

(b) Determine a dimensão da amostra, N, e c de modo a que $\alpha = 0.05$ e $\beta = 0.1$.

Solution: Agora temos

$$\begin{aligned} \alpha &= \mathbb{P}\left(\overline{X} > c \mid \mu = 1\right) = 0.05 \\ \beta &= \mathbb{P}\left(\overline{X} < c \mid \mu = 2\right) = 0.1 \end{aligned}$$

Então,

$$\begin{array}{l} \mathbb{P}\left(Z > \frac{c-1}{2/\sqrt{N}}\right) = 0.05 \\ \mathbb{P}\left(Z < \frac{c-2}{2/\sqrt{N}}\right) = 0.1 \end{array} \quad \Longleftrightarrow \quad \begin{array}{l} \frac{c-1}{2/\sqrt{N}} = 1.645 \\ \frac{c-2}{2/\sqrt{N}} = -1.282 \end{array}$$

Resolvendo, obtêm-se $c = 1.562$ e $N = 34.26 \approx 35$.

(c) Suponha que para amostras de dimensão 2 da população acima referida se usa a seguinte regra de decisão: rejeita-se H_0 se $\overline{X} > 1.5$. Calcule a potência do teste.

Solution: A região crítica agora é $RC = \left\{\overline{X} : \overline{X} > 1.5\right\}$

$$\begin{aligned} \pi\,(\mu = 2) &= \mathbb{P}\left(\overline{X} > 1.5 \mid \mu = 2\right) = \mathbb{P}\left(Z > \frac{1.5-2}{2/\sqrt{2}}\right) = \\ &= \mathbb{P}\,(Z > -0.35) = \mathbb{P}\,(Z < 0.35) = 0.6368 \end{aligned}$$

3. (6 points) Diga se as seguintes afirmações são verdadeiras ou falsas. Justifique as suas

respostas.

(a) As idades dum grupo de executivos que participam numa convenção seguem uma distribuição uniforme entre os 35 e os 65 anos. Entre os executivos que participam na convenção estão os directores da empresa. A única informação adicional disponível é que todos têm mais de 50 anos. Perante estas informações, se seleccionarmos um executivo aleatoriamente e ele for um director, a probabilidade de esse director ter mais de 60 anos é 1/3.

> **Solution:** X = idades dos executivos que participam numa convenção $\Longrightarrow X \sim \mathrm{U}(35, 65)$
>
> $$\begin{aligned}\mathbb{P}(X > 60 \mid X > 50) &= \frac{\mathbb{P}(X > 60)}{\mathbb{P}(X > 50)} = \frac{\int_{60}^{65} \frac{1}{65-35} \mathrm{d}x}{\int_{50}^{65} \frac{1}{65-35} \mathrm{d}x} = \\ &= \frac{5/30}{15/30} = \frac{1}{3}.\end{aligned}$$
>
> Assim, a afirmação é **VERDADEIRA**.

(b) Um analista está a considerar dois investimentos alternativos. Em ambos os casos tem dúvidas sobre o rendimento (em percentagem) que poderia obter. No entanto, crê que para o primeiro investimento o rendimento é uma variável aleatória com distribuição $\mathcal{N}(10.4, 1.2^2)$, enquanto que para o segundo investimento a distribuição do rendimento é $\mathcal{N}(11, 4^2)$. Se o analista decide investir na opção que tem maior probabilidade de produzir um rendimento superior a 10%, deveria escolher o primeiro investimento.

> **Solution:** R_i = rendimento do investimento i, $i = 1, 2$
> $R_1 \sim \mathcal{N}(10.4, 1.2^2)$, $R_2 \sim \mathcal{N}(11, 4^2)$
>
> $$\begin{aligned}\mathbb{P}(R_1 > 10) &= \mathbb{P}\left(Z > \frac{10 - 10.4}{1.2}\right) = \mathbb{P}\left(Z > -\frac{1}{3}\right) = 0.6293 \\ \mathbb{P}(R_2 > 10) &= \mathbb{P}\left(Z > \frac{10 - 11}{4}\right) = \mathbb{P}\left(Z > -\frac{1}{4}\right) = 0.5987\end{aligned}$$
>
> Como $0.6293 > 0.5987$, a afirmação é **VERDADEIRA**.

(c) Tem-se observado o número de defeitos produzidos por uma máquina ao longo de 100 semanas. Os resultados estão na seguinte tabela:

A hipótese nula do número de defeitos produzidos pela máquina numa semana seguir uma distribuição de Poisson com média 0.4 pode ser rejeitada para um nível de significança de 5%.

Número de defeitos	0	1	2	3	4
Semanas	73	18	6	2	1

Solution:

$$\begin{aligned}
\mathbb{P}(X=0) &= e^{-0.4}\frac{0.4^0}{0!} = 0.6703 \Rightarrow \hat{E}_0 = 100 \cdot 0.6703 = 67.03 \\
\mathbb{P}(X=1) &= e^{-0.4}\frac{0.4^1}{1!} = 0.2681 \Rightarrow \hat{E}_1 = 100 \cdot 0.2681 = 26.81 \\
\mathbb{P}(X=2) &= e^{-0.4}\frac{0.4^2}{2!} = 0.0536 \Rightarrow \hat{E}_2 = 100 \cdot 0.0536 = 5.36 \\
\mathbb{P}(X \geq 3) &= 1 - 0.6703 - 0.2681 - 0.0536 = 0.008 \\
& \Longrightarrow \hat{E}_{\geq 3} = 100 \cdot 0.008 = 0.8
\end{aligned}$$

Como $\hat{E}_{\geq 3} < 5$, temos que reagrupar as observações.

$$\begin{aligned}
\mathbb{P}(X \geq 2) &= 1 - 0.6703 - 0.2681 = 0.0616 \\
\Longrightarrow \hat{E}_{\geq 2} &= 100 \cdot 0.0616 = 6.16
\end{aligned}$$

que é maior que 5, então a nova tabela é

Número de defeitos	0	1	≥ 2
O_i	73	18	9
$\hat{E}_i$	67.03	26.81	6.16

A estatística de teste é

$$\chi^2 = \sum_{i=1}^{3} \frac{\left(O_i - \hat{E}_i\right)^2}{\hat{E}_i} = 4.736$$

Valor crítico: $\chi^2_{3-1;0.05} = 5.99 > 4.736 \Longrightarrow$ não se rejeita H_0. Assim, a afirmação é **FALSA**.

(d) Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com distribuição binomial $\mathrm{b}(1,p)$. O estimador da variância do universo, $\hat{V} = \overline{X}\left(1 - \overline{X}\right)$, onde $\overline{X}$ é a média amostral, é não enviesado.

Solution: Como X é binomial, sabemos que $\mathbb{E}(X) = \mathbb{E}\left(\overline{X}\right) = p$ e $\mathbb{V}\mathrm{ar}(X) = p(1-p)$. Então, $\mathbb{V}\mathrm{ar}\left(\overline{X}\right) = \frac{p(1-p)}{n}$.

$$\begin{aligned}
\mathbb{E}\left(\hat{V}\right) &= \mathbb{E}\left(\overline{X} - \overline{X}^2\right) = \mathbb{E}\left(\overline{X}\right) - \mathbb{E}\left(\overline{X}^2\right) = \\
&= \mathbb{E}\left(\overline{X}\right) - \mathbb{V}\mathrm{ar}\left(\overline{X}\right) - \left[\mathbb{E}\left(\overline{X}\right)\right]^2 = \\
&= p - \frac{p(1-p)}{n} - p^2 = \frac{n-1}{n} p(1-p) \neq p(1-p)
\end{aligned}$$

Assim, a afirmação é **FALSA**.

(e) Se o *p-value* dum teste é 0.0368, então podemos rejeitar a hipótese nula para um nível de significância de 1%, mas não de 5%.

Solution: $0.0368 > 0.01 \Longrightarrow$ não se rejeita a 1%
$0.0368 < 0.05 \Longrightarrow$ rejeita-se a 5%
Assim, a afirmação é **FALSA**.

(f) Um sindicato pretende analisar o grau de sinistralidade laboral nos sectores da construção civil e da metalurgia. Para isso, analisa uma amostra aleatória de 200 empresas e observa o sector a que pertencem e o grau de sinistralidade laboral (baixo ou alto). A tabela seguinte mostra os resultados obtidos:

	Baixo	Alto
Construção civil	98	27
Metalurgia	54	21

Perante estes resultados, há evidência estatística para rejeitar a hipótese do grau de sinistralidade laboral ter a mesma distribuição nos dois sectores.

Solution: Valores observados:

	Baixo	Alto	*Total*
Construção civil	98	27	125
Metalurgia	54	21	75
Total	152	48	200

Valores esperados:

	Baixo	Alto	*Total*
Construção civil	95	30	125
Metalurgia	57	18	75
Total	152	48	200

onde $95 = \frac{125 \cdot 152}{200}$, etc.
A estatística de teste é

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(O_{ij} - \hat{E}_{ij}\right)^2}{\hat{E}_{ij}} = 1.053$$

O valor crítico é $\chi_{(2-1)(2-1);0.05} = 3.84 > 1.053 \Longrightarrow$ não se rejeita H_0
Assim, a afirmação é **FALSA**.