Final exam
January 25, 2007
Statistics II

1. (6 points) O Sr. Adalberto vende maquinaria agrícola. Dada a complexidade dos seus produtos, o processo de venda passa sempre por uma demonstração dos produtos, aos potenciais compradores, o que é feito pelo Sr. Adalberto, numa pequena propriedade agrícola, situada em Santarém. Na loja, está um empregado que tem como função receber os clientes, apresentar os produtos através de catálogos e marcar as demonstrações para os interessados. Sabe-se que:

   - Entram na loja, em média, 0.75 clientes por hora. O número de clientes que entram na loja segue uma distribuição de Poisson.
   - A probabilidade de um cliente marcar uma demonstração é de 50%.
   - O tempo que o Sr. Adalberto leva a fazer uma demonstração segue uma distribuição normal com média de 2 horas e desvio padrão de 3 horas. Todas as demonstrações são independentes entre si.
   - A probabilidade de um cliente que entra na loja efectivar a compra é de 20%.

   (a) Sabendo que a loja está aberta 8 horas por dia, qual a probabilidade de num determinado dia entrarem mais de 10 clientes na loja?

   Solution:

   (b) Sabendo que a loja está aberta 22 dias por mês, 12 meses por ano, qual o número médio de clientes que visitam a loja num ano? E desses, quantos, em média, realizarão uma compra? (Considere que cada cliente não entra na loja mais do que uma vez por ano.)

   Solution:

   (c) Numa dada manhã o empregado olha para o relógio e exclama "São 11h e 30m e só agora me entra o primeiro cliente na loja!". Qual a probabilidade de ainda ter de esperar mais de 1 hora pela chegada de um outro cliente?

   Solution:

   (d) Numa semana em que entrem 45 clientes na loja qual a probabilidade de conseguir vender produtos a pelo menos 12 destes clientes?

   (e) E qual a probabilidade de terem de entrar 7 clientes na loja para que o empregado consiga assegurar a marcação de 3 ou 4 demonstrações?

   (f) Num dia em que o Sr. Adalberto tenha que realizar quatro demonstrações qual a probabilidade de ter de trabalhar mais do que 9 horas nesse dia?

2. (8 points) Numa amostra aleatória de 148 directores de marketing, 75 classificaram o sentido de humor como uma característica muito importante no sucesso das suas carreiras. Da mesma opinião foram 81 directores financeiros numa amostra de 178 inquiridos.

(a) Teste a hipótese de **pelo menos** 50% dos directores financeiros considerarem o senso de humor como muito importante ($\alpha = 0.05$).

Solution: Seja X o número de directores financeiros que consideram o sentido de humor como muito importante $\Longrightarrow X \sim \mathrm{b}\left(N_X = 178, p_X\right)$.
Temos que testar a hipótese $H_0 : p_X = 0.5$ contra $H_1 : p_X < 0.5$ a nível de significância de $\alpha = 0.05$. A estatística para este teste é, pelo teorema do límite central,

$$Z = \frac{f_N - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \xrightarrow{d} \mathcal{N}(0,1)$$

$\mathbb{P}\left(\text{rejeitar } H_0 \mid H_0 \text{ é verdadeira}\right) = 0.05 \iff \mathbb{P}\left(f_N < a \mid p_X = 0.5\right) 0.05 \iff$

$$\mathbb{P}\left(Z < \frac{a - 0.05}{\sqrt{\frac{0.5 \cdot 0.5}{178}}}\right) = 0.05$$

pelo que $\dfrac{a - 0.05}{0.0374766} = -1.645 \iff a = 0.43835$. Então, $RC = (-\infty; 0.43835]$.
Como $f_N = \dfrac{81}{178} = 0.455 \notin RC \Longrightarrow$ não rejeitamos H_0.

(b) Tendo em conta a decisão que tomou na alínea anterior qual o erro que pode estar a cometer? Calcule a sua probabilidade se a verdadeira proporção de directores financeiros que considera o senso de humor muito importante for de 40%.

Solution: Posso estar a cometer o erro tipo II: não rejeitar H_0 sendo H_0 falsa.

$$\begin{aligned} \beta\left(p_X = 0.4\right) &= \mathbb{P}\left(f_N > 0.43835 \mid p_X = 0.4\right) = \mathbb{P}\left(Z > \frac{0.43835 - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{178}}}\right) = \\ &= \mathbb{P}\left(Z > 1.04\right) = 1 - \Phi\left(1.04\right) = 1 - 0.8508 = 0.1492 \end{aligned}$$

(c) Calcule o nível de significância mais baixo que o levaria a rejeitar a hipótese nula.

Solution: O que se pede é que calcule o p-value.

$$\begin{aligned} \text{p-value} &= \mathbb{P}\left(f_N < 0.455 \mid H_0\right) = \mathbb{P}\left(Z < \frac{0.455 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{178}}}\right) = \\ &= \mathbb{P}\left(Z < -1.2\right) = 1 - \Phi\left(1.2\right) = 1 - 0.8849 = 0.1151 \end{aligned}$$

(d) Teste a hipótese de ser igual a proporção de directores de marketing e directores financeiros que consideram o senso de humor como uma característica muito importante ($\alpha = 0.05$).

Solution: Seja gora Y o número de directores de marketing que consideram o sentido de humor como muito importante $\Longrightarrow Y \sim \text{b}\left(N_Y = 148, p_Y\right)$
Temos que testar o parâmetro $p_X - p_Y$,

$$\begin{aligned} H_0 &: p_X - p_Y = 0 \\ H_1 &: p_X - p_Y \neq 0 \end{aligned}$$

a nível de significância $\alpha = 0.05$. A estatística do teste é

$$Z = \frac{\left(f_{N_X} - f_{N_Y}\right) - \left(p_X - p_Y\right)}{\sqrt{\frac{f_{N_X}\left(1-f_{N_X}\right)}{N_X} + \frac{f_{N_Y}\left(1-f_{N_Y}\right)}{N_Y}}} \xrightarrow{d} \mathcal{N}\left(0,1\right)$$

$$\mathbb{P}\left(Z < -z_{\alpha/2} \vee Z > z_{\alpha/2} \mid p_X - p_Y = 0\right) = 0.05$$

onde $z_{\alpha/2} = z_{0.025} = 1.96$. Assim, $RC = (-\infty, -1.96] \cup [1.96, \infty)$. Neste exercício, $f_{N_Y} = \dfrac{75}{148} = 0.507$. Então,

$$Z = \frac{(0.455 - 0.507) - 0}{\sqrt{\frac{0.455 \cdot 0.545}{178} + \frac{0.507 \cdot 0.493}{148}}} = -0.93 \notin RC$$

$\Longrightarrow$ não rejeitamos H_0.

Para a mesma amostra de 148 directores de marketing acima considerada, verificou-se que relativamente ao salário mensal

$$\sum_{i=1}^{148} X_i = 592 \qquad \sum_{i=1}^{148} X_i^2 = 3600$$

em que X_i está em 1000 euros e a distribuição da população pode ser considerada normal.

(e) Construa um intervalo de confiança a 99% para o verdadeiro salário médio mensal dos directores de marketing.

Solution: Seja X o salário mensal dos directores de marketing $\Longrightarrow X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, onde μ é o parâmetro a testar.
A variável fulcral é

$$T = \frac{\overline{X} - \mu}{S/\sqrt{N}} \sim \text{t}_{N-1}$$

onde $N = 148$ e $\alpha = 0.01$. Então, $t_{N-1;\alpha/2} = t_{N-1;0.005} = 2.61$. Assim, o intervalo de confiança é

$$[\mu]_{99\%} = \left[\overline{X} \pm t_{\alpha/2} \frac{S'}{\sqrt{N}}\right]$$

A média amostral é

$$\overline{X} = \frac{\sum_{i=1}^{148} X_i}{148} = \frac{592}{148} = 4$$

e o desvio padrão amostral é

$$S' = \sqrt{\frac{\sum_{i=1}^{148} X_i^2 - N \cdot \overline{X}^2}{N-1}} = \sqrt{\frac{36.. - 148 \cdot 16}{147}} = 2.895$$

Então, o intervalo de confiança é

$$[\mu]_{99\%} = \left[4 \pm 2.61 \frac{2.895}{\sqrt{148}}\right] = [3.379; 4.621]$$

(f) Assumindo que a verdadeira variância do salário mensal é igual a 9 e para um nível de confiança de 99%, qual o número de directores a incluir na amostra para que a amplitude do intervalo não seja superior a 1.

Solution: Agora sabemos que $\sigma^2 = 9$. Então, a variável fulcral é

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

e $\alpha = 0.01 \implies z_{\alpha/2} = z_{0.005} = 2.576$. Então, o intervalo de confiança é

$$[\mu]_{99\%} = \left[\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}\right]$$

A amplitude do intervalo é

$$\Delta = 2 z_{\alpha/2} \frac{\sigma}{\sqrt{N}} = 2 \cdot 2.576 \frac{3}{\sqrt{N}} = 1$$

pelo que $N = 238.8 \approx 239$.

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

(a) O custo (em milhares de euros) que uma empresa de seguros deve assumir quando há N acidentes numa fábrica é $1+2N$. Se o número de acidentes que ocorrem num mês é uma variável aleatória com distribuição Poisson de parâmetro 0.25, então o custo médio que a empresa de seguros terá que assumir num mês é de 1500 euros, com variância de 1500 euros ao quadrado.

> **Solution:** Seja C a variável aleatória que representa o custo. Então, $\mathbb{E}(C) = \mathbb{E}(1+2N) = 1+2\mathbb{E}(N) = 1+2\cdot 0.25 = 1.5$, 1500 euros. Por outro lado, $\mathbb{V}\text{ar}(C) = \mathbb{V}\text{ar}(1+2N) = 4\mathbb{V}\text{ar}(N) = 4\cdot 0.25 = 1$, 1000 euros ao quadrado. Assim, a afirmação é **FALSA**.

(b) Se X segue uma distribuição exponencial com função densidade de probabilidade $f(x) = 3e^{-3x}$ se $x \geq 0$ ou 0 caso contrário, e $\mathbb{E}(X) = \mu$ e $\mathbb{V}\text{ar}(X) = \sigma^2$, então $\mathbb{P}(|X-\mu| \leq 2\sigma) = 0.95$.

> **Solution:** O parâmetro da distribuição exponencial é $\lambda = 3$, assim, pelas propriedades da distribuição, $\mu = \sigma = 3$. Portanto, o acontecimento $|X-\mu| \leq 2\sigma$ equivale a $X \in [\mu - 2\sigma; \mu + 2\sigma] = \left[-\frac{1}{3}; 1\right]$. Então, $\mathbb{P}(|X-\mu| \leq 2\sigma) = \mathbb{P}(X \leq 1) - \mathbb{P}\left(X \leq -\frac{1}{3}\right) = 1 - e^{-3} = 0.95$. Assim, a afirmação é **VERDADEIRA**.

(c) Seja X_1, X_2 uma amostra aleatória de dimensão 2 de uma variável aleatória X com média μ e variância 9, onde μ é um parâmetro desconhecido. Então o erro quadrático médio do estimador $\hat{\mu} = \frac{2}{3}X_1 + \frac{1}{3}X_2$ é igual a 5.

> **Solution:** O erro quadrático médio é definido como $\text{MSE}(\hat{\mu}) = \mathbb{V}\text{ar}(\hat{\mu}) + \text{bias}(\hat{\mu})^2$. O valor esperado de $\hat{\mu}$ é $\mathbb{E}(\hat{\mu}) = \frac{2}{3}\mathbb{E}(X_1) + \frac{1}{3}\mathbb{E}(X_2) = \frac{2}{3}\mu + \frac{1}{3}\mu = \mu$. Assim $\hat{\mu}$ é um estimador não-enviesado, portanto, o bias é zero. A variância de $\hat{\mu}$ é
>
> $$\mathbb{V}\text{ar}(\hat{\mu}) = \text{MSE}(\hat{\mu}) = \frac{4}{9}\mathbb{V}\text{ar}(X_1) + \frac{1}{9}\mathbb{V}\text{ar}(X_2) = \frac{4}{9}\cdot 9 + \frac{1}{9}\cdot 9 = 5.$$
>
> Assim, a afirmação é **VERDADEIRA**.

(d) Se não conhecemos a variância de um universo normal, podemos calcular um intervalo de confiança aproximado para a média da população, mas não um intervalo de confiança exacto.

> **Solution:** Sem conhecer a variância de um universo normal, podemos calcular um intervalo de confiança exacto usando a distribuição de t-Studen. Assim, a afirmação é **FALSA**.

(e) No teste de hipóteses $H_0 : \mu \leq 1$ versus $H_1 : \mu > 1$ concluímos que $\mu > 1$. Então se o verdadeiro valor de μ é -2, teremos cometido um erro tipo I, enquanto se o verdadeiro valor de μ é 2, teremos cometido um erro tipo II.

> **Solution:** Da conclusão é claro que se rejeita a hipótese nula. Se o verdadeiro valor de μ é -2, então a hipótese nula é verdadeira, mas rejeitamo-la, tendo assim, um erro tipo I. Se o verdadeiro valor de μ é 2, então a hipótese nula é falsa e rejeitamo-la, pelo que não cometemos um erro. Por tanto, a afirmação é **FALSA**.

(f) Dadas duas amostras aleatórias independentes de um universo normal com variâncias iguais, se no teste da igualdade de médias com região de rejeição **unilateral** e nível de significância 0.05 rejeitamos H_0, então no teste da igualdade de médias com região de rejeição **bilateral** e nível de significância 0.10 também rejeitamos H_0.

> **Solution:** A estística do teste é
>
> $$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\left(\frac{1}{N_X} + \frac{1}{N_Y}\right) S^2}} \sim t_{N_X+N_Y-2}.$$
>
> Para um nível de significância 0.05 a região crítica para o teste unilateral é $R_C = \{T > t_{N_X+N_Y-2;0.05}\}$ ou $R_C = \{T < -t_{N_X+N_Y-2;0.05}\}$, dependendo da alternativa ser $\mu_X > \mu_Y$ ou $\mu_X < \mu_Y$. Para o teste bilateral, com um nível de significância 0.10 a região crítica é $R_C = \{|T| > t_{N_X+N_Y-2;0.05}\}$. Se rejeitamos a hipótese nula para um teste unilateral a 5% é porque $T > t_{N_X+N_Y-2;0.05}$ ou $T < -t_{N_X+N_Y-2;0.05}$, e portanto $|T| > t_{N_X+N_Y-2;0.05}$. Assim, para o teste bilateral pode-se afirmar que H_0 é rejeitada a 10%. Portanto, a afirmação é **VERDADEIRA**.