ESTATÍSTICA II (Eco+Fin) - 2º Ano/2º sem.
Exame Época Normal **29.Maio. 2023**

2 horas. (20 valores)

Nome:____ **nº**__________

Espaço reservado para classificações				
1 (20)	3 a) (10)	4 a) (20)	*5 a) (20)*	*5 d) (15)*
2 a) (15)	3 b) (15)	4 b) (10)	*5 b) (10)*	*5 e) (20)*
2 b) (15)	3 c) (20)		5 c) (10)	
T:				

1. Perguntas de Verdadeiro/Falso [20 pontos] Para cada afirmação assinale se esta é verdadeira (V) ou falsa (F). Uma resposta certa vale 2 pontos e uma resposta errada penaliza em 1 ponto.

	V	F
Sejam θ e ψ dois parâmetros desconhecidos, tais que $\theta = g(\psi)$, sendo g uma função. Se $\hat{\psi}$ é um estimador de máxima verosimilhança para ψ, então é necessariamente um estimador de máxima verosimilhança para θ.		X
Sejam 2 estimadores T_1 e T_2 para o parâmetro θ desconhecido tais que $E(T_1) = \theta$ e $E(T_2) \neq \theta$. Se $Var(T_1) < Var(T_2)$, então T_1 é relativamente mais eficiente que T_2.		X
Intervalos de confiança são função da distribuição da população, amostra e desvio padrão.	X	
Um intervalo de confiança a 95% para θ contém, com probabilidade 0.95, o verdadeiro valor de θ.		X
Num teste de hipóteses sobre o parâmetro $\theta \in \Theta$, $H_{0:}\ \theta \in \Theta_0$ contra $H_{1:}\ \theta \in \Theta_1$, $\Theta_0 \cup \Theta_1 = \Theta, \Theta_0 \cap \Theta_1 = \emptyset$.	X	
Num teste de hipóteses é cometido um erro de segunda espécie quando não se rejeita a hipótese nula quando esta é falsa.	X	
No teste de independência entre dois atributos de uma população, o número observado de elementos em cada célula deverá ser de pelo menos 5.		X
No modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, as variáveis explicativas tem uma distribuição de probabilidade.		X
No modelo de regressão linear $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ não é usual testar a significância individual do termo independente.	X	
Um modelo de regressão linear $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ é homocedástico se $var(u\|x)$ é constante.	X	

Versão B

	V	F
Sejam θ e ψ dois parâmetros desconhecidos, tais que $\theta = g(\psi)$, sendo g uma função biunívoca de ψ . Se $\hat{\psi}$ é um estimador de máxima verosimilhança para ψ, então $g(\hat{\psi})$ é um estimador de máxima verosimilhança para θ.	X	
Sejam 2 estimadores T_1 e T_2 dois estimadores não enviesados para o parâmetro θ. Se $Var(T_1) < Var(T_2)$, então T_1 é relativamente mais eficiente que T_2.	X	
Intervalos de confiança são função de Informação fornecida pela amostra, nível de confiança e a dimensão da amostra.	X	
Probabilidade de o verdadeiro valor do parâmetro na população θ pertencer a um intervalo confiança a $(1-\alpha) * 100\%$ para θ é igual a 0 ou 1.	X	
Num teste de hipóteses é cometido um erro de tipo 1 quando rejeita uma hipótese nula verdadeira.	X	
Num teste de hipóteses sobre o parâmetro $\theta \in \Theta$, $H_{0:}\ \theta \in \Theta_0$ contra $H_{1:}\ \theta \in \Theta_1$, os espaços paramétricos satisfazem $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 \neq \emptyset$.		X
No teste de independência entre dois atributos de uma população, a região de rejeição é bilateral.		X
No modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, as variáveis explicativas não são variáveis aleatórias	X	
De acordo com as hipóteses do modelo de regressão linear simples, o valor médio do efeito das variáveis não observadas é nulo.	X	
Um modelo de regressão linear, $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ é homocedástico se $var(u_i) = \sigma_i^2$.		X

Versão C

	V	F
A função de verosimilhança é função do parâmetro desconhecido θ dada uma amostra particular $(x_1, x_2, \cdots, x_n)$.	X	
Sejam T_1 e T_2 dois estimadores enviesados para o parâmetro θ. Se $Var(T_1) < Var(T_2)$, então T_1 é relativamente mais eficiente que T_2.		X
Intervalos de confiança são sempre função da amostra, variável de interesse e graus de liberdade.		X
A probabilidade de o verdadeiro valor do parâmetro θ na população pertencer a um intervalo confiança a $(1-\alpha) * 100\%$ para θ calculado para uma amostra particularé igual a $1-\alpha$.		X
Num teste de hipóteses a potência do teste é a probabilidade de se rejeitar uma hipótese nula falsa.	X	
Num teste de hipóteses sobre o parâmetro $\theta \in \Theta$, $H_{0:}\ \theta \in \Theta_0$ contra $H_{1:}\ \theta \in \Theta_1$, os espaços paramétricos satisfazem $\Theta_0 \cup \Theta_1 \neq \Theta, \Theta_0 \cap \Theta_1 = \emptyset$.		X
Nos testes não paramétricos estudados, a região de rejeição pode não ser bilateral.		X
No modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, a variável residual u_i é uma variável aleatória.	X	
De acordo com as hipóteses do modelo de regressão linear simples, não pode existir correlação entre as variáveis explicativas x_j e a variável residual.	X	
Num modelo de regressão linear, $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, os regressores não devem estar correlacionados com as variáveis residuais u_i.	X	

2. De uma população X com média $\theta + 4$ e função densidade de probabilidade:

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x-4}{\theta}} \quad x > 4, \ \theta > 0$$

foi obtida uma amostra aleatória de dimensão n.

a) Encontre o estimador pelo método da máxima verosimilhança para o parâmetro θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x-4}{\theta}} = \left(\frac{1}{\theta}\right)^n e^{-\sum_i^n \left(\frac{x_i-4}{\theta}\right)}$$

$$l(\theta) = lnL(\theta) = nln\left(\frac{1}{\theta}\right) - \sum_i^n \left(\frac{x_i-4}{\theta}\right) = -nln(\theta) - \sum_i^n \left(\frac{x_i-4}{\theta}\right)$$

$$\underbrace{\frac{dl(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum_i^n (x_i - 4)}{\theta^2} = 0} \Leftrightarrow \underbrace{-n\theta + \sum_i^n (x_i - 4) = 0} \Leftrightarrow \underbrace{\theta = \frac{\sum_i^n (x_i - 4)}{n} = \bar{X} - 4}$$

$$\underbrace{\frac{d^2l(\theta)}{d\theta^2} = \frac{n}{\theta^2} - 2\frac{\sum_i^n (x_i - 4)}{\theta^3} = \frac{-n}{\theta^2} < 0 \quad \rightarrow \hat{\theta} = \bar{X} - 4}_{maximizante}$$

$b)$ Analise o enviesamento e a consistência de $\hat{\theta}$.

$\underbrace{E(\hat{\theta}) = E(\bar{X} - 4) = E(\bar{X}) - 4 = E(X) - 4 = \theta + 4 - 4 = \theta}$, então $\hat{\theta}$ é um estimador não enviesado para θ.

$$Var(\hat{\theta}) = \underbrace{Var(\bar{X} - 4) = Var(\bar{X})} = \underbrace{\frac{Var(X)}{n}}$$

$\lim_{n\to\infty} E\left(\hat{\theta}\right) = \lim_{n\to\infty} \theta = \theta$ e $\lim_{n\to\infty} Var\left(\hat{\theta}\right) = \lim_{n\to\infty} \frac{Var(X)}{n} = 0$ pelo que $\boldsymbol{\hat{\theta}}$ **cumpre as condições necessárias e suficientes de consistência como estimador para $\boldsymbol{\theta}$**

3. Um inquérito a 500 portugueses com mais de 18 anos sobre a utilização da internet na compra de bens ou serviços permitiu resumir os resultados no seguinte quadro:

Utilização de Internet / Idade	compra online	não compra online	Total
18 - 30	92	38	130
30 - 50	80	80	160
>50	50	160	210
Total	222	278	

a) Com base num intervalo de confiança a 95 % diga se concorda com a seguinte afirmação "Cerca de 50% dos portugueses efectuam compras de bens ou serviços online". Justifique.

X – português efectuar compras online $\sim B(1, \theta)$

$\alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$ $\qquad \bar{x} = \frac{222}{500} = 0.444$

$$IC_\theta^{95\%} = \left(\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}};\ \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right)$$

$$= \left(0.444 - 1.96\sqrt{\frac{0.444(1-0.444)}{500}};\ 0.444 + 1.96\sqrt{\frac{0.444(1-0.444)}{500}}\right)$$

$$= (0.4004; 0.4876).$$

Como 0.5 não pertence ao intervalo a afirmação não é verdadeira a um nível de significância de 5%.

b) Se usasse um teste de hipóteses para testar a veracidade da afirmação da alínea anterior como formularia as hipóteses a testar e qual a estatística teste a utilizar? A conclusão a que chegaria seria idêntica à que chegou na alínea anterior? Explique porquê. [Considere um nível de significância de 5%]

$\underbrace{H_0: \theta = 0.5 \quad contra \quad H_1: \theta \neq 0.5}$ Estatística teste: $\frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \dot{\sim} N(0,1)$

Sim, **seria idêntica** pois o $\boldsymbol{IC_\theta^{95\%}}$ **corresponde à região de não rejeição da hipótese $\boldsymbol{H_0}$ no teste Formulado .**

c) Ao nível de 5%, pode afirmar-se que a utilização da internet para compras de bens ou serviços é independente da idade?

$H_0: \hat{p}_{ij} = \hat{p}_{i.} * \hat{p}_{.j} \ \forall (i,j) \ i = 1,2,3; j = 1,2 \quad contra \ H_1: \exists \hat{p}_{ij} \neq \hat{p}_{i.} * \hat{p}_{.j}$

Estatística teste - $Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(\boldsymbol{n_{ij}} - n.\widehat{p_i}\,\widehat{p}_j\right)^2}{n\,\widehat{p}_i\,\widehat{p}_j} \dot{\sim} \boldsymbol{\chi}^2_{[(r-1)(s-1)]}$

Utilização de internet / Idade	compra online	não compra online	Total
18 - 30	92	38	130
30 - 50	80	80	160
>50	50	160	210
Total	222	278	500

Matriz das frequências esperadas

Utilização de internet / Idade	compra online	não compra online	Total
18 - 30	57.72	72.28	130
30 - 50	71.04	88.96	160
>50	93.24	116.76	210
Total	222	278	500

Resolução usando a região de rejeição:

$\underbrace{q_{obs} = 74.715}$, $\underbrace{q_{0.05}: P(Q > q_{0.05}) = 0.05 \Leftrightarrow \overbrace{P\left(\chi^2_{[(3-1)(2-1)]} > q_{0.05}\right) = 0.05} \Rightarrow \overbrace{q_{0.05} = 5.991}}$

então $\underbrace{q_{obs} \in W = \{q: q > 5.991\}}$

Ou

Resolução usando valor-p:

$\underbrace{q_{obs} = 74.715}$, $\underbrace{valor - p = P(Q > q_{obs}|H_0) = P\left(\chi^2_{[(3-1)(2-1)]} > 74.715\right) \approx 0}$

pelo que se **rejeita hipótese de independência entre idade e compra online**.

4. Com o objetivo de promover o desporto na infância, foram recolhidos dados sobre duas variáveis aleatórias: o peso de crianças que praticam desporto (variável X) e o peso de crianças que não praticam desporto (variável Y). Considere que ambas as variáveis aleatórias são normalmente distribuídas.

As amostras recolhidas de cada população apresentam os seguintes valores:

$$\sum_{i=1}^{25} x_i = 873.4\,, \sum_{i=1}^{25} x_i^2 = 30641\,, \sum_{i=1}^{21} y_i = 780.7\,, \sum_{i=1}^{21} y_i^2 = 29134$$

a) Com base num teste estatístico adequado, com 5% de significância, mostre que não se rejeita a igualdade de variâncias das duas variáveis aleatórias.

$$\underbrace{H_0: \sigma_X^2 = \sigma_Y^2 \quad contra \quad H_1: \sigma_X^2 \neq \sigma_Y^2 \vee H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1 \quad contra \quad H_1: \frac{\sigma_X^2}{\sigma_Y^2} \neq 1}$$

$$Estatística\ teste: \frac{s'^2_1}{s'^2_2} \sim F_{(24,20)}$$

$$s_X^2 = \frac{\sum_{i=1}^{21} x_i^2}{21} - \left(\frac{\sum_{i=1}^{21} x_i}{21}\right)^2 = \frac{30641}{25} - \left(\frac{873.4}{25}\right)^2 = 5.1159$$

$$\Rightarrow s'^2_X = \frac{25}{24} * 5.1159 = 5.33$$

$$s_Y^2 = \frac{\sum_{i=1}^{21} y_i^2}{m} - \left(\frac{\sum_{i=1}^{21} y_i}{m}\right)^2 = \frac{29134}{21} - \left(\frac{780.7}{21}\right)^2 = 5.2642$$

$$\Rightarrow s'^2_Y = \frac{21}{20} * 5.2642 = 5.53$$

Resolução usando valor-p:

$$\underbrace{f_{obs} = \frac{s'^2_X}{s'^2_Y} = 0.96} \Rightarrow \underbrace{valor - p = 2P\left(F_{(24,20)} < 0.96\right) = 2\times 0.46 = 0.92}$$

$$\Rightarrow \underbrace{não\ se\ rejeita\ H_0 \Rightarrow \sigma_X^2 = \sigma_Y^2}$$

$$ou$$

Resolução usando a região de rejeição:

Pela tabela 9, $\boldsymbol{F_{\frac{\alpha}{2}} = 2.41}$ **e** $\boldsymbol{F_{\frac{\alpha}{2}}^{*} = 2.33 \quad \Rightarrow \frac{1}{F_{\frac{\alpha}{2}}^{*}} = 0.429}$

Logo $\boldsymbol{\underbrace{W_{0.05} = \{f_{obs}: f_{obs} < 0.429 \vee f_{obs} > 2.41\}} \Rightarrow \underbrace{f_{obs}\ não\ pertence\ a\ W}}$

Então, **não se rejeita $\boldsymbol{H_0}$, e conclui-se pela igualdade de variâncias**.

b) Usando o resultado na alínea anterior, construa um teste estatístico adequado, para comentar a seguinte afirmação: “as crianças que praticam desporto têm um peso médio inferior às crianças que não praticam desporto”. [Considere um nível de significância de 5%] [Nota: Se o valor dos graus de liberdade não se encontrar na tabela, use o valor mais próximo]

b)

$$H_0: \mu_X \geq \mu_Y \quad contra \quad H_1: \mu_X < \mu_Y \vee H_0: \mu_X - \mu_Y \geq 0 \quad contra \quad H_1: \mu_X - \mu_Y < 0$$

OU

$$H_0: \mu_X \leq \mu_Y \quad contra \quad H_1: \mu_X > \mu_Y \vee H_0: \mu_X - \mu_Y \leq 0 \quad contra \quad H_1: \mu_X - \mu_Y > 0$$

$$Estatística\ teste: T = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}\sqrt{\frac{(m-1)S_X'^2 + (n-1)S_Y'^2}{m+n-2}}} \sim t(44)$$

$$t_{obs} = \frac{34.936 - 37.176}{\sqrt{\left(\frac{1}{25} + \frac{1}{21}\right)}\sqrt{\frac{24 \times 5.33 + 20 \times 5.54}{44}}} = -3.25$$

Resolução usando valor-p:

$$valor - p = P\left(T_{(44)} < -3.25\right) \quad valor - p < 0.05 \Rightarrow Rejeita - se\ H_0$$

$$ou$$

Resolução usando a região de rejeição:

$$t_{0.05}: P\left(T_{(44)} > t_{0.05}\right) = 0.05 \Rightarrow t_{0.05} = 1.68 \text{ e } W = \{t_{obs}: t_{obs} < -1.68\} \Rightarrow t_{obs} \in W$$

$$\Rightarrow Rejeita - se\ H_0$$

para um nível de significância de 5%. Existe evidência estatística a favor da afirmação.Ou, caso tenha feito o teste de forma inversa, não rejeitar H_0, concluindo o mesmo.

5. Com o intuito de estudar o salário dos gestores empresariais, foi recolhida uma amostra de 177 gestores, tendo-se estimado o seguinte modelo:

$$lsal_i = \beta_0 + \beta_1 lucros_i + \beta_2 lvendas_i + \beta_3 exper_i + \beta_4 exper_i^2 + u_i$$

Onde: • lsal – Logaritmo do salário do gestor;
• $lucros$ – Lucros da empresa em milhões de euros;
• $lvendas$ – Logaritmo das vendas;
• $exper$ – Número de anos como gestor da empresa;
• $exper^2$ – Quadrado de $exper$.

No Anexo 1 figuram os resultados obtidos na estimação deste modelo.

a) Interprete as estimativas obtidas para os coeficientes β1 e β2 e teste a significância estatística de cada um destes parâmetros.

$\hat{\beta}_1$= 0.0002: Estima-se que, **em média, ceteris paribus**, o aumento de 1 milhão de euros nos lucros da empresa levará a um aumento de, aproximadamente, 0.02% no salário do gestor.

$\hat{\beta}_2$= 0.196: Estima-se que, **em média, ceteris paribus**, o aumento de 1% das vendas da empresa levará a um aumento de, aproximadamente, 0.196% no salário do gestor.

$$\underbrace{H_0: \beta_j = 0 \quad contra \;\; H_1: \beta_j \neq 0 \quad j = 1, 2}$$

Estatística teste: $\frac{\hat{\beta}_j - \beta_j}{se_{\hat{\beta}_j}} \sim t_{(n-k-1)}$ com $n = 177, k = 4$

Resolução usando a região de rejeição:

$$t_{obs}(\hat{\beta}_1) \approx 1.566; \;\; t_{obs}(\hat{\beta}_2) \approx 5.925\;; \; t_{172}^{0.05} \approx 1.96$$

$$\underbrace{W = \{t_{obs}: |t_{obs}| > 1.96\}} \Rightarrow \underbrace{t_{obs}(\hat{\beta}_1) \notin W, t_{obs}(\hat{\beta}_2) \in W}$$

considerando um nível de significância de 5%, concluí-se que **a variável lucros não é estatisticamente significativa**, **a variável vendas é estatisticamente significativa**.

ou Resolução usando valor-p: $p_{obs}(\hat{\beta}_1) \approx 0.119\;; \;\; p_{obs}(\hat{\beta}_2) \approx 0.000\;; \alpha = 0.05$

$$\underbrace{p_{obs}(\hat{\beta}_1) > \alpha}, \underbrace{p_{obs}(\hat{\beta}_2) < \alpha}$$

considerando um nível de significância de 5%, concluí-se que **a variável lucros não é estatisticamente significativa, a variável vendas é estatisticamente significativa** .

b) Teste se é aceitável concluir que um aumento de 2% nas vendas origine um aumento de 0.5% no salário do gestor.

$$\underbrace{H_0: \beta_2 = 0.25 \quad contra \;\; H_1: \beta_2 \neq 0.25}$$

Estatística teste: $\frac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} \sim t_{(n-k-1)}$ $\quad t_{obs} = \frac{0.196-0.25}{0.033} = -1.636$

Resolução usando a região de rejeição:

$$\underbrace{W = \{t: t < -1.96 \;\cup t > 1.96\}} \Rightarrow \underbrace{t_{obs}\, não\ pertence\ a\ W}$$

Não se rejeita H_0 num teste com 5% de significância, logo, com os dados de que dispomos, **não temos razões para excluir a hipótese de que o verdadeiro valor do coeficiente (na população) seja 0.25**.

$$ou$$

Resolução usando valor-p:

$$\Rightarrow \underbrace{valor - p = P(|T_{(172)}| > |-1.636|) \approx 2 * 0.05 = 0.1} \Rightarrow \underbrace{valor - p > 0.05} \Rightarrow \underbrace{não\ se\ rejeita\ H_0}$$

Não se rejeita H_0 num teste com 5% de significância logo, com os dados de que dispomos, **não temos razões para excluir a hipótese de que o verdadeiro valor do coeficiente (na população) seja 0.25**.

c) Teste se o modelo é globalmente significativo.

$$\underbrace{H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \quad contra \;\; H_1: \exists\, \beta_j \neq 0 \;\; (j = 1, \cdots, 4)}$$

Estatística teste: $F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{SSE/k}{SSR/(n-k-1)} \sim F_{(k,\, n-k-1)}$

$$\underbrace{valor - p = P\left(F_{(4,\,177-4-1)} > 21.525\right) \approx 0}$$

$\underbrace{Rejeita - se\; H_0}$, logo o **modelo é globalmente significativo.**

d) Usando o "output" fornecido no Anexo 2, formalize o teste da hipótese de que a experiência não tem qualquer impacto no salário dos gestores e retire a sua conclusão.

$$\underbrace{H_0: \beta_3 = \beta_4 = 0 \quad contra \;\; H_1: \exists \beta_j \neq 0 \;\; j = 3{,}4}$$

Estatística teste: $F = \frac{(\boldsymbol{SSR}_r - \boldsymbol{SSR}_{nr}) \Big/ \left(\overbrace{k-p}^{m}\right)}{\boldsymbol{SSR}_{nr}/(n-k-1)} \sim F_{(m,\, n-k-1)}$ com $m = 2, n = 177, k = 4$

$$\underbrace{f_{obs} = \frac{(\mathbf{45.909} - \mathbf{43.081})/2}{\mathbf{43.081}/172} = 5.645}$$

Ou

Estatística teste: $F = \frac{(R - R_*)/m}{(1-R^2)/(n-k-1)} \sim F_{(m,\, n-k-1)}$

$$f_{obs} = \frac{(\mathbf{0.3336} - \mathbf{0.29})/2}{(\mathbf{1 - 0.336})/172} = 5.62665$$

Resolução usando a região de rejeição:

$$\underbrace{f_{0.05}: P\left(F_{(2,\,172)} > f_{0.05}\right) = 0.05} \Rightarrow \underbrace{f_{0.05} = 3} \Rightarrow \underbrace{W = \{f : f > 3\}} \Rightarrow \underbrace{f_{obs}\; pertence\; a\; W}$$

Ou

Resolução usando valor-p:

$\underbrace{valor - p = P\left(F_{(2,\,172)} > 5.645\right) \approx 0}$ ou

$$\underbrace{0.025 < valor - p = P\left(F_{(2,\,172)} > 3.662\right) < 0.05}$$

Então, **rejeita-se $\boldsymbol{H_0}$** e **conclui-se que existe evidência estatística a favor da significância conjunta das variáveis $\boldsymbol{exper}$ e $\boldsymbol{exper^2}$.**

e) Qual o objectivo de incluir a variável $exper^2$ no modelo, dado que já tínhamos $exper$? Apresente o valor a partir do qual a experiência como gestor é, de acordo com o modelo, negativa para o salário.

R: Ao incluir a variável $exper^2$ o efeito da experiência do gestor no seu salário deixa de ser linear. Como $\hat{\beta}_3 > 0$ tem um valor positivo, tal indica que o salário aumenta quando exper aumenta mas $\hat{\beta}_4 < 0$ diz-nos que esse aumento se faz a uma taxa decrescente, pelo que a partir de um certo número de anos de experiência o impacto da variável $\boldsymbol{exper}$ passa de positivo a negativo.

Este valor calcula-se a partir de $\underbrace{\frac{\partial \widehat{lsal}}{\partial exper} = \hat{\beta}_3 + 2\hat{\beta}_4 exper}.$

$$= \underbrace{0{,}04537 + [2 * (-0{,}00123)]exper \leq 0} \Leftrightarrow \underbrace{exper \geq 18.44}$$

Assim, **estima-se que em média,**
a partir do momento em que o gestor passe, aproximadamente, dos 18 anos e 5 meses de experiência, o efeito desta variável no seu salário comece a ser negativo .

Anexo 1 – Variável dependente lsal

Regression Statistics	
R Square	0.3336
R Square Adjusted	0.3181
Standard Error	0.5005
Observations	177

ANOVA					
	df	SS	MS	F	Significance F
Regression	4	21.566	5.391	21.525	0.000
Residual	172	43.081	0.250		
Total	176	64.646			

	Coefficients	Standard Error	t Stat	P-value
Intercept	4.90417	0.24397	20.458	0.000
lucros	0.00018	0.00012	1.566	0.119
lvendas	0.19636	0.03314	5.925	0.000
exper	0.04537	0.01422	3.191	0.002
$exper^2$	-0.00123	0.00048	-2.572	0.011

Anexo 2 – Variável dependente lsal

Regression Statistics	
R Square	0.290
R Square Adjusted	0.282
Standard Error	0.514
Observations	177

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	18.737	9.369	35.507	0.000
Residual	174	45.909	0.264		
Total	176	64.646			

	Coefficients	Standard Error	t Stat	P-value
Intercept	5.14509	0.23473	21.919	0.000
lucros	0.00018	0.00012	1.483	0.140
lvendas	0.19370	0.03400	5.697	0.000