

Nome: \_\_\_\_\_ Nº: \_\_\_\_\_

## Espaço reservado para classificações

1a.(10)	2a.(15)	3a.(5)	4.(15)	5a.(10)	5e.(15)
1b.(15)	2b.(10)	3b.(15)		5b.(10)	5f.(20)
1c.(15)				5c.(15)	5g.(15)
				5d.(15)	

**Atenção: todas as questões devem ser devidamente formalizadas e justificadas.** Sempre que fizer um teste de hipóteses formule as hipóteses em teste e apresente a estatística de teste e sempre que fizer um intervalo de confiança apresente a variável fulcral.

- Um analista de uma instituição financeira pretende comparar as taxas de incumprimento de clientes que contraíram um crédito automóvel com as de clientes que contraíram crédito para férias. Para esse fim, analisou os registos de 1000 créditos para compra de automóvel e 1500 créditos para férias, escolhidos de forma aleatória. Verificou que no crédito automóvel houve 230 incumprimentos da contraparte, enquanto nos créditos para férias houve 195 incumprimentos.
  - [10] Construa um intervalo de confiança a 95% para a diferença entre as proporções de incumprimentos. O que pode o analista concluir?
  - [15] Mostre que o estimador de máxima verosimilhança para a taxa de incumprimento de clientes que contraíram um crédito automóvel é dado pela respetiva média amostral.
  - [15] Estude o enviesamento e consistência do estimador referido na alínea anterior.
- Admita que  $X$  é uma variável aleatória com distribuição normal cuja média e variância são iguais, mas desconhecidas, isto é:  $X \sim N(\theta, \theta)$ , onde  $\theta > 0$ . Para testar  $H_0: \theta = 25$  contra  $H_1: \theta > 25$ , considerou-se uma amostra casual de dimensão  $n = 51$  e definiram-se as seguintes regiões de rejeição:

$$W_1 = \{\bar{x}: \bar{x} > 26.3723\} \quad \text{e} \quad W_2 = \{s'^2: s'^2 > 35.71\}$$

- [15] Mostre que a probabilidade associada com o erro de 1ª espécie é igual para ambas as regiões de rejeição.
  - [10] Observada uma amostra em particular, obteve-se:  $\bar{x} = 26.94$  e  $s'^2 = 34.87$ . Calcule o valor-p associado com a estatística-teste  $\bar{X}$ . Sem efetuar cálculos, mas explicando brevemente, diga se o valor-p associado com a estatística-teste  $S'^2$  é superior, inferior ou igual ao baseado na estatística-teste  $\bar{X}$ .
- Um clínico pretende testar a eficácia de um novo suplemento de ferro no tratamento de baixos níveis de hemoglobina no sangue (anemia) por carência de ferro. Para esse fim, selecionou uma amostra casual de 25 mulheres com idades compreendidas entre os 15 e os 45 anos. Antes de iniciar o tratamento, o clínico avaliou a hemoglobina no sangue em cada paciente e, três meses após o tratamento, voltou a avaliá-la. Calculou então, para cada paciente, o ganho em termos de hemoglobina, isto é, a diferença entre os níveis de hemoglobina depois e antes do tratamento, tendo obtido uma média de 2.1 g/dL e um desvio-padrão corrigido de 1.8 g/dL. Sabe-se que o nível de hemoglobina segue uma distribuição normal.
    - [5] Explique as razões pelas quais em estudos deste tipo se devem usar amostras emparelhadas.
    - [15] Para um nível de significância de 1%, será que pode afirmar-se que o suplemento de ferro tem efeito benéfico sobre os níveis de hemoglobina?
  - [15] A relação\associação entre a vacinação e a intensidade dos efeitos do vírus COVID 19 tem sido um tema comentado. Para tentar trazer alguma fundamentação ao debate realizou-se um inquérito a uma amostra aleatória de 1000 indivíduos e obtiveram-se os seguintes resultados:

Intensidade\Vacinação	Não vacinados	Vacinação completa
Fraca	20	400
Média	130	90
Grande	350	10

Com base num teste adequado o que pode concluir.

5. Com o propósito de modelar o preço de um bilhete de avião a partir de um grande aeroporto europeu, estimou-se o seguinte modelo

$$\widehat{Preço} = 41.785 + 0.0805 Dist - 0.00911 NPass + 1.166 QMerc + 0.0178 Dest - 68.722 LowCost$$

(17.125)
(0.0051)
(0.0042)
(0.239)
(0.0044)
(8.127)

$$n = 320 \quad R^2 = 0.5299$$

Onde

- Preço – valor mediano do preço de uma viagem “one-way” em classe turística numa determinada unidade monetária
- Dist – Distância percorrida pelo voo (em milhas)
- NPass – Número médio diário de passageiros naquela rota
- QMerc – Quota de mercado do maior operador na rota em percentagem
- Dest – nº de habitantes (em milhares) na cidade servida pelo aeroporto de destino
- LowCost – variável que assume o valor 1 se o voo é de uma companhia low-cost e 0 no caso contrário

Assuma que o modelo verifica as hipóteses habituais e recorra, sempre que necessário, aos outputs apresentados em anexo.

- a) **[10]** Teste a significância estatística global da regressão e teste também a significância individual do parâmetro associado com a variável NPass para  $\alpha = 0.05$ .
- b) **[10]** Interprete as estimativas obtidas para os coeficientes associados com as variáveis Dist e QMerc.
- c) **[15]** Com base no valor-p, teste se é aceitável concluir que, tudo o resto constante, o aumento da distância percorrida encarece o preço do voo. A sua conclusão é válida para todos os casos ou apenas em média. Justifique.
- d) **[15]** Construa um intervalo de confiança a 90% para o parâmetro associado com a variável Dest. Será possível concluir, a partir deste intervalo, sobre a significância estatística do parâmetro? Justifique.
- e) **[15]** Determinado técnico afirma que um acréscimo de 200 milhas na distância do voo tem o mesmo efeito do que um aumento de 15 pontos percentuais na quota de mercado da maior operadora naquela rota. Teste a afirmação do técnico para  $\alpha = 0.05$ .
- f) **[20]** Com o propósito de construir um intervalo de previsão a 90% para o preço de um voo específico feito numa companhia “low cost”, com uma distância de 1000 milhas para um destino com um milhão de habitantes e com um número médio diário de 1500 passageiros e onde a quota de mercado do maior operador é de 50%, estimou-se a regressão apresentada no anexo 3. Explique a que correspondem as variáveis explicativas da regressão e construa o referido intervalo.
- g) **[15]** Um técnico, não totalmente satisfeito com os resultados obtidos decidiu estimar a regressão auxiliar  $\hat{u}^2 = \gamma_0 + \gamma_1 Npass + \varepsilon$ , onde  $\hat{u}^2$  corresponde ao quadrado dos resíduos obtidos na regressão inicial. Qual o problema que o técnico pretende analisar? Sabendo que nesta regressão auxiliar se obteve  $R^2 = 0.0078$  qual a conclusão a tirar?

**ANEXOS – Em todas as regressões a variável dependente é o Preço**

**Anexo 1**

<i>Regression Statistics</i>					
Multiple R	0.72794				
R Square	0.52990				
Adjusted R Square	0.52241				
Standard Error	51.10079				
Observations	320				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	924234.75	184846.95	70.787	1.9939E-49
Residual	314	819945.24	2611.29		
Total	319	1744179.99			

  

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	41.7854	17.1252	2.43999	0.01524
Dist/200	-1.4010	3.3336	-0.42028	0.67457
Npass	-0.0091	0.0042	-2.15304	0.03208
Qmerc+15*Dist/200	1.1663	0.2388	4.88394	1.66E-06
Dest	0.0178	0.0044	4.03098	6.98E-05
LowCost	-68.7221	8.1268	-8.45622	1.06E-15

**Anexo 2**

<i>Regression Statistics</i>					
Multiple R	0.72794				
R Square	0.52990				
Adjusted R Square	0.52241				
Standard Error	51.10079				
Observations	320				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	924234.75	184846.95	70.78758	1.99E-49
Residual	314	819945.24	2611.29		
Total	319	1744179.99			

  

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	41.7854	17.1252	2.43999	0.01524
Dist-200*Qmerc/15	0.0805	0.0051	15.80373	8.41E-42
Npass	-0.0091	0.0042	-2.15304	0.03208
Qmerc/15	33.5869	4.0770	8.23822	4.78E-15
Dest	0.0178	0.0044	4.03098	6.98E-05
LowCost	-68.7221	8.1268	-8.45622	1.06E-15

### Anexo 3

---

<i>Regression Statistics</i>	
Multiple R	0.72794
R Square	0.52990
Adjusted R Square	0.52241
Standard Error	51.10079
Observations	320

---

ANOVA

---

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	924234.75	184846.95	70.78758	1.99E-49
Residual	314	819945.24	2611.29		
Total	319	1744179.99			

---

---

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	115.9597	8.5312	13.59241	2.07E-33
Dist1	0.0805	0.0051	15.80373	8.41E-42
Npass1	-0.0091	0.0042	-2.15304	0.03208
Qmerc1	1.1663	0.2388	4.883936	1.66E-06
Dest1	0.0178	0.0044	4.03098	6.98E-05
LowCost1	-68.7221	8.1268	-8.45622	1.06E-15

---

Nome: \_\_\_\_\_ Nº: \_\_\_\_\_

### Espaço reservado para classificações

1a.(10)	2a.(15)	3a.(5)	4.(15)	5a.(10)	5e.(15)
1b.(15)	2b.(10)	3b.(15)		5b.(10)	5f.(20)
1c.(15)				5c.(15)	5g.(15)
				5d.(15)	

**Atenção: todas as questões devem ser devidamente formalizadas e justificadas.** Sempre que fizer um teste de hipóteses formule as hipóteses em teste e apresente a estatística de teste e sempre que fizer um intervalo de confiança apresente a variável fulcral.

- Um analista de uma instituição financeira pretende comparar as taxas de incumprimento de clientes que contraíram um crédito automóvel com as de clientes que contraíram crédito para férias. Para esse fim, analisou os registos de 1000 créditos para compra de automóvel e 1500 créditos para férias, escolhidos de forma aleatória. Verificou que no crédito automóvel houve 230 incumprimentos da contraparte, enquanto nos créditos para férias houve 195 incumprimentos.
  - [10]** Construa um intervalo de confiança a 95% para a diferença entre as proporções de incumprimentos. O que pode o analista concluir?

População: Indicador de incumprimento	Amostra
Crédito automóvel: $X_1 \sim B(1, \theta_1)$	$m = 1000, \bar{x}_1 = 230/1000 = 0.23$
Crédito para férias: $X_2 \sim B(1, \theta_2)$	$n = 1500, \bar{x}_2 = 195/1500 = 0.13$

**IC a 95% para  $\theta_1 - \theta_2$**

$$VF: Z = \frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1 - \bar{X}_1)}{m} + \frac{\bar{X}_2(1 - \bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$$

$$IC: \left( \bar{x}_1 - \bar{x}_2 \pm z_{0.025} \times \sqrt{\frac{\bar{x}_1(1 - \bar{x}_1)}{m} + \frac{\bar{x}_2(1 - \bar{x}_2)}{n}} \right) =$$

$$= \left( 0.23 - 0.13 \pm 1.96 \times \sqrt{\frac{0.23(1 - 0.23)}{1000} + \frac{0.13(1 - 0.13)}{1500}} \right) = (0.069, 0.131)$$

Com uma confiança de 95% pode afirmar-se que a diferença entre as proporções de incumprimentos de crédito automóvel e de crédito para férias se situa entre 0.069 e 0.131. Porque o IC engloba valores estritamente positivos, existe evidência estatística de que a proporção de incumprimentos no crédito automóvel é superior à do crédito para férias.

- b. [15] Mostre que o estimador de máxima verosimilhança para a taxa de incumprimento de clientes que contraíram um crédito automóvel é dado pela respetiva média amostral.

$$L(\theta_1) = \prod_{i=1}^m f(x_{1i}|\theta_1) = \prod_{i=1}^m [\theta_1^{x_{1i}}(1-\theta_1)^{1-x_{1i}}] = \theta_1^{\sum_{i=1}^m x_{1i}}(1-\theta_1)^{m-\sum_{i=1}^m x_{1i}} = \theta_1^{m\bar{x}_1}(1-\theta_1)^{m-m\bar{x}_1} \quad (0 < \theta_1 < 1)$$

$$l(\theta_1) = \ln[L(\theta_1)] = \ln[\theta_1^{m\bar{x}_1}(1-\theta_1)^{m-m\bar{x}_1}] = m\bar{x}_1 \ln(\theta_1) + (m-m\bar{x}_1) \ln(1-\theta_1), \quad 0 < \theta_1 < 1$$

$$l'(\theta_1) = 0 \Leftrightarrow m\bar{x}_1 \frac{1}{\theta_1} + (m-m\bar{x}_1) \left(-\frac{1}{1-\theta_1}\right) = 0 \Leftrightarrow \frac{m\bar{x}_1(1-\theta_1) - (m-m\bar{x}_1)\theta_1}{\theta_1(1-\theta_1)} = 0 \Leftrightarrow$$

$$\Leftrightarrow m\bar{x}_1(1-\theta_1) - (m-m\bar{x}_1)\theta_1 = 0 \Leftrightarrow m\bar{x}_1 - m\bar{x}_1\theta_1 - m\theta_1 + m\bar{x}_1\theta_1 = 0 \Leftrightarrow m\bar{x}_1 - m\theta_1 = 0 \Leftrightarrow$$

$$\Leftrightarrow \theta_1 = \bar{x}_1$$

$$l''(\theta_1) = m\bar{x}_1 \left(-\frac{1}{\theta_1^2}\right) + (m-m\bar{x}_1) \left[-\frac{1}{(1-\theta_1)^2}\right] = -\left\{m \left[\frac{\bar{x}_1}{\theta_1^2} + \frac{1-\bar{x}_1}{(1-\theta_1)^2}\right]\right\} < 0, \quad \forall \theta_1 \in (0, 1)$$

Logo, mostra-se que:  $\hat{\theta}_1 = \bar{X}_1$ .

- c. [15] Estude o enviesamento e consistência do estimador referido na alínea anterior.

$$E(\hat{\theta}_1) = E(\bar{X}_1) = E(X_1) = \theta_1. \text{ Logo, } \hat{\theta}_1 \text{ é um estimador centrado para } \theta_1.$$

As condições suficientes para a consistência verificam-se:

$$\lim_{m \rightarrow \infty} E(\hat{\theta}_1) = \lim_{m \rightarrow \infty} \theta_1 = \theta_1$$

$$\lim_{m \rightarrow \infty} Var(\hat{\theta}_1) = \lim_{m \rightarrow \infty} Var(\bar{X}_1) = \lim_{m \rightarrow \infty} \frac{Var(X_1)}{m} = \lim_{m \rightarrow \infty} \frac{\theta_1(1-\theta_1)}{m} = 0$$

Logo,  $\hat{\theta}_1$  é um estimador consistente para  $\theta_1$ .

2. Admita que  $X$  é uma variável aleatória com distribuição normal cuja média e variância são iguais, mas desconhecidas, isto é:  $X \sim N(\theta, \theta)$ , onde  $\theta > 0$ . Para testar  $H_0: \theta = 25$  contra  $H_1: \theta > 25$ , considerou-se uma amostra casual de dimensão  $n = 51$  e definiram-se as seguintes regiões de rejeição:

$$W_1 = \{\bar{x}: \bar{x} > 26.3723\} \quad \text{e} \quad W_2 = \{s'^2: s'^2 > 35.71\}$$

- a. **[15]** Mostre que a probabilidade associada com o erro de 1ª espécie é igual para ambas as regiões de rejeição.

$$\alpha_1 = P(\bar{X} > 26.3723 | \theta = 25) = P\left(\frac{\bar{X} - \theta}{\sqrt{\theta}/\sqrt{n}} > \frac{26.3723 - 25}{\sqrt{25}/\sqrt{51}}\right) = P(Z > 1.96) = 0.025,$$

onde  $Z \sim N(0,1)$ .

$$\alpha_2 = P(S'^2 > 35.71 | \theta = 25) = P\left(\frac{(n-1)S'^2}{\theta} > \frac{50 \times 35.71}{25}\right) = P(Q > 71.42) = 0.025,$$

onde  $Q \sim \chi^2(50)$ .

Logo, mostra-se que  $\alpha_1 = \alpha_2$ .

- b. **[10]** Observada uma amostra em particular, obteve-se:  $\bar{x} = 26.94$  e  $s'^2 = 34.87$ . Calcule o valor-p associado com a estatística-teste  $\bar{X}$ . Sem efetuar cálculos, mas explicando brevemente, diga se o valor-p associado com a estatística-teste  $S'^2$  é superior, inferior ou igual ao baseado na estatística-teste  $\bar{X}$ .

$$p_{1,obs} = P(\bar{X} \geq 26.94 | \theta = 25) = P\left(\frac{\bar{X} - \theta}{\sqrt{\theta}/\sqrt{n}} \geq \frac{26.94 - 25}{\sqrt{25}/\sqrt{51}}\right) = P(Z \geq 2.77) = 1 - \Phi(2.77) =$$

$$= 1 - 0.9972 = 0.0028$$

Seja  $p_{2,obs}$  o valor-p associado com a estatística-teste  $S'^2$  e  $\alpha$  o nível de significância comum a  $W_1$  e  $W_2$ .

Porque  $\bar{x} = 26.94 \in W_1$  rejeita-se  $H_0$ , pelo que  $p_{1,obs} < \alpha$ . Por outro lado,  $s'^2 = 34.87 \notin W_2$ , pelo que não se rejeita  $H_0$  e  $p_{2,obs} > \alpha$ . Logo,  $p_{2,obs} > p_{1,obs}$ .

3. Um clínico pretende testar a eficácia de um novo suplemento de ferro no tratamento de baixos níveis de hemoglobina no sangue (anemia) por carência de ferro. Para esse fim, selecionou uma amostra casual de 25 mulheres com idades compreendidas entre os 15 e os 45 anos. Antes de iniciar o tratamento, o clínico avaliou a hemoglobina no sangue em cada paciente e, três meses após o tratamento, voltou a avaliá-la. Calculou então, para cada paciente, o ganho em termos de hemoglobina, isto é, a diferença entre os níveis de hemoglobina depois e antes do tratamento, tendo obtido uma média de 2.1 g/dL e um desvio-padrão corrigido de 1.8 g/dL. Sabe-se que o nível de hemoglobina segue uma distribuição normal.

- a. [5] Explique as razões pelas quais em estudos deste tipo se devem usar amostras emparelhadas.

Em estudos deste tipo devem usar-se amostras emparelhadas pois submetendo os mesmos elementos da amostra a estudo nos dois momentos distintos permite isolar o efeito do fator principal (neste caso, o novo suplemento) ao reduzir ou eliminar potenciais efeitos de outros fatores não relevantes para o estudo.

- b. [15] Para um nível de significância de 1%, será que pode afirmar-se que o suplemento de ferro tem efeito benéfico sobre os níveis de hemoglobina?

Seja  $X$  o nível de hemoglobina depois do tratamento e  $Y$  o nível de hemoglobina antes do tratamento.

$$H_0: \mu_X - \mu_Y = 0 \quad vs \quad H_1: \mu_X - \mu_Y > 0$$

$$ET: T = \frac{\bar{Z}}{S'_Z/\sqrt{25}} \sim t(24) \quad W_{0.01} = \{t_{obs}: t_{obs} > 2.492\} \quad t_{obs} = \frac{2.1}{1.8/\sqrt{25}} \approx 5.83$$

Porque  $t_{obs} \in W_{0.01}$ , rejeita-se  $H_0$  ao nível de 1% e conclui-se que o novo suplemento de ferro tem efeito benéfico sobre os níveis de hemoglobina.

4. [15] A relação\associação entre a vacinação e a intensidade dos efeitos do vírus COVID 19 tem sido um tema comentado. Para tentar trazer alguma fundamentação ao debate realizou-se um inquérito a uma amostra aleatória de 1000 indivíduos e obtiveram-se os seguintes resultados:

Intensidade\Vacinação	Não vacinados	Vacinação completa
Fraca	20	400
Média	130	90
Grande	350	10

Com base num teste adequado o que pode concluir.

Pretende-se efetuar um teste de independência.

$$H_0: \forall (i, j), p_{ij} = p_{i\cdot}p_{\cdot j} \quad vs \quad H_1: \exists (i, j), p_{ij} \neq p_{i\cdot}p_{\cdot j}$$

$$ET: Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - fe_{ij})^2}{fe_{ij}} \underset{a}{\sim} \chi^2\{(r-1)(s-1)\}$$

Cálculo das frequências esperadas:  $fe_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{n}$

Intensidade\Vacinação	Não vacinados	Vacinação completa
Fraca	210	210
Média	110	110
Grande	180	180

Como todas as frequências esperadas são não inferiores a 5, não é necessário qualquer agrupamento de classes. Logo,  $r = 3$  e  $s = 2$ , pelo que  $Q \stackrel{a}{\sim} \chi^2(2)$ .

$$q_{obs} = \frac{(20 - 210)^2}{210} + \frac{(400 - 210)^2}{210} + \dots + \frac{(10 - 180)^2}{180} \approx 672.19 \text{ e } p_{obs} = P(Q > 672.19) \approx 0$$

Logo, existe forte evidência estatística desfavorável a  $H_0$ , o que significa que existe relação entre a vacinação e a intensidade dos sintomas.

5. Com o propósito de modelar o preço de um bilhete de avião a partir de um grande aeroporto europeu, estimou-se o seguinte modelo

$$\widehat{Preço} = 41.785 + 0.0805 Dist - 0.00911 NPass + 1.166 QMerc + 0.0178 Dest - 68.722 LowCost$$

$$(17.125) \quad (0.0051) \quad (0.0042) \quad (0.239) \quad (0.0044) \quad (8.127)$$

$$n = 320 \quad R^2 = 0.5299$$

Onde

- Preço – valor mediano do preço de uma viagem “one-way” em classe turística numa determinada unidade monetária
- Dist – Distância percorrida pelo voo (em milhas)
- NPass – Número médio diário de passageiros naquela rota
- QMerc – Quota de mercado do maior operador na rota em percentagem
- Dest – nº de habitantes (em milhares) na cidade servida pelo aeroporto de destino
- LowCost – variável que assume o valor 1 se o voo é de uma companhia low-cost e 0 no caso contrário

Assuma que o modelo verifica as hipóteses habituais e recorra, sempre que necessário, aos outputs apresentados em anexo.

- a) **[10]** Teste a significância estatística global da regressão e teste também a significância individual do parâmetro associado com a variável NPass para  $\alpha = 0.05$ .

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{contra} \quad H_1: \exists \beta_j \neq 0 \quad (j = 1, \dots, 5)$$

$$ET: F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{(m, n-k-1)} \quad \text{isto é} \quad F = \frac{R^2/5}{(1-R^2)/314} \sim F_{(5, 314)}$$

$$F_{obs} = \frac{0.5299/5}{(1-0.5299)/314} = 70.7886 \quad \text{valor} - p = P(F_{(5, 314)} > 70.789) \cong 0$$

$$\text{Alternativamente } W_{0.05} = \{f: f > 2.21\}$$

Rejeita-se  $H_0$  o modelo é globalmente significativo.

$$H_0: \beta_2 = 0 \quad \text{contra} \quad H_1: \beta_2 \neq 0 \quad ET: T = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(n - k - 1) \quad \text{isto é} \quad T = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} \sim t(314)$$

$$t_{obs} = \frac{-0.00911}{0.0042} = -2.169$$

$$\text{valor} - p = 2 \times P(t_{(314)} \geq |-2.169|) = 2 \times P(t_{(314)} > 2.169) = 0.0308$$

$$\text{Alternativamente } W_{5\%} = \{t: t < -1.96 \text{ ou } t > 1.96\}$$

Rejeita-se  $H_0$  para um nível de significância de 5% logo considera-se que a variável Npass não é estatisticamente relevante

- b) **[10]** Interprete as estimativas obtidas para os coeficientes associados com as variáveis Dist e QMerc.

$\hat{\beta}_{Dist} = \hat{\beta}_2 = 0.0805$  – por cada milha adicional percorrida pelo voo, tudo o resto constante, o preço mediano esperado aumenta de 0.0805 unidades monetárias

$\hat{\beta}_{Qmerc} = \hat{\beta}_3 = 1.166$  – um acréscimo de 1 p.p. na quota de mercado do maior operador na rota contribui, tudo o resto igual, para um aumento esperado do preço mediano em 1.166 unidades monetárias

- c) **[15]** Com base no valor-p, teste se é aceitável concluir que, tudo o resto constante, o aumento da distância percorrida encarece o preço do voo. A sua conclusão é válida para todos os casos ou apenas em média. Justifique.

$$H_0: \beta_{Dist} \leq 0 \text{ contra } H_1: \beta_{Dist} > 0 \text{ ET: } T = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(n - k - 1) \text{ isto é } T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t(314)$$

valor - p =  $P(t_{(314)} \geq 15.7843) \cong 0$ , logo rejeita-se  $H_0$  o que significa que existe evidência empírica que suporta a afirmação de que a distância percorrida encarece o preço do voo.

O resultado é válido apenas em média porque no MRLM, o parâmetro  $\beta_{Dist}$  mede o impacto de uma variação unitária na distância no valor esperado do preço mediano.

- d) **[15]** Construa um intervalo de confiança a 90% para o parâmetro associado com a variável Dest. Será possível concluir, a partir deste intervalo, sobre a significância estatística do parâmetro? Justifique.

$$VF: T = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(314)$$

$$IC_{\beta_j}^{90\%} = (\hat{\beta}_j - t_{\alpha/2} \hat{\sigma}_{\hat{\beta}_j}; \hat{\beta}_j + t_{\alpha/2} \hat{\sigma}_{\hat{\beta}_j}) \text{ Com } t_{\alpha/2}: P(T_{(n-k-1)} > t_{\alpha/2}) = \alpha/2$$

$$t_{\alpha/2}: P(T_{(314)} > t_{0.1/2}) = 0.05 \text{ logo } t_{0.05} = 1.645$$

$$IC_{90\%}(\beta_{Dest.}) = (0.0178 - 1.645 * 0.0044; 0.0178 + 1.645 * 0.0044) \\ = (0.0106, 0.0250)$$

Como o intervalo não contém o valor 0, é possível concluir, a um nível de significância de 10%, que o parâmetro é estatisticamente significativo.

- e) **[15]** Determinado técnico afirma que um acréscimo de 200 milhas na distância do voo tem o mesmo efeito do que um aumento de 15 pontos percentuais na quota de mercado da maior operadora naquela rota. Teste a afirmação do técnico para  $\alpha = 0.05$ .

A afirmação é  $200\beta_{Dist} = 15\beta_{Qmer}$  ou seja definindo  $\theta = 200\beta_{Dist} - 15\beta_{Qmer}$  vai-se testar  $H_0: \theta = 0$  contra  $H_1: \theta \neq 0$  com base numa regressão auxiliar.

$$\text{Sol 1: } \beta_{Dist} = \frac{\theta + 15\beta_{Qmer}}{200}$$

$$\text{Preço} = \beta_0 + \frac{\theta + 15\beta_{Qmer}}{200} \text{ Dist} + \beta_2 \text{ NPass} + \beta_3 \text{ QMerc} + \beta_4 \text{ Dest} + \beta_5 \text{ LowCost} + u$$

$$\text{Preço} = \beta_0 + \theta \frac{\text{Dist}}{200} + \beta_2 \text{ NPass} + \beta_3 \left( \text{QMerc} + \frac{15}{200} \text{ Dist} \right) + \beta_4 \text{ Dest} + \beta_5 \text{ LowCost} + u$$

$$\text{ET: } T = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t(314) \quad t_{obs} = -0.420 \quad \text{valor-p} = 0.67457 \quad (\text{ver anexo 1})$$

Não se rejeita  $H_0$  e existe evidência empírica a favor da afirmação do técnico.

Ou

$$\text{Sol 2: } \beta_{Qmer} = \frac{200\beta_{Dist} - \theta}{15}$$

$$\text{Preço} = \beta_0 + \beta_1 \text{ Dist} + \beta_2 \text{ NPass} + \frac{200\beta_{Dist} - \theta}{15} \text{ QMerc} + \beta_4 \text{ Dest} + \beta_5 \text{ LowCost} + u$$

$$\text{Preço} = \beta_0 + \left( \text{Dist} + 200 \frac{\text{QMerc}}{15} \right) + \beta_2 \text{ NPass} + \theta \left( - \frac{\text{QMerc}}{15} \right) + \beta_4 \text{ Dest} + \beta_5 \text{ LowCost} + u$$

$$ET: T = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t(314) \quad t_{obs} = -0.420 \quad \text{valor-p}=0.67457 \quad (\text{ver anexo 2})$$

Não se rejeita  $H_0$  e existe evidência empírica a favor da afirmação do técnico.

- f) **[20]** Com o propósito de construir um intervalo de previsão a 90% para o preço de um voo específico feito numa companhia “low cost”, com uma distância de 1000 milhas para um destino com um milhão de habitantes e com um número médio diário de 1500 passageiros e onde a quota de mercado do maior operador é de 50%, estimou-se a regressão apresentada no anexo 3. Explique a que correspondem as variáveis explicativas da regressão e construa o referido intervalo.

As variáveis explicativas da regressão auxiliar apresentada no anexo 3 correspondem aos valores das variáveis originais menos os valores para os quais se quer fazer a previsão. Assim,  
 $DIST1=Dist-1000$ ,  $Npass1=Npass-1500$ ,  $Qmerc1=Qmerc-50$ ,  $Dest1=Dest-1000$ ,  $LowCost1=LowCost-1$ .

$$V.F. t = \frac{y^o - \hat{y}^o}{\sqrt{se(\hat{\theta})^2 + \hat{\sigma}^2}} \sim t(314)$$

Do anexo 3 vem  $\hat{y}^o = 115.9597$ ,  $se(\hat{\theta})=8.5312$  e  $\hat{\sigma}^2 = 2611.29$ . Como  $t_{\alpha/2} = 1.645$  o intervalo de previsão virá dado por  $115.9597 \pm 1.645 \sqrt{8.5312^2 + 2611.29}$ , isto é (30.74; 201.18)

- g) **[15]** Um técnico, não totalmente satisfeito com os resultados obtidos decidiu estimar a regressão auxiliar  $\hat{u}^2 = \gamma_0 + \gamma_1 Npass + \varepsilon$ , onde  $\hat{u}^2$  corresponde ao quadrado dos resíduos obtidos na regressão inicial. Qual o problema que o técnico pretende analisar? Sabendo que nesta regressão auxiliar se obteve  $R^2 = 0.0078$  qual a conclusão a tirar?

O técnico pretende testar a existência de heterocedasticidade ou melhor um padrão específico de heterocedasticidade onde esta dependa de  $Npass$ .

Estimada a regressão auxiliar ele irá testar  $H_0: \gamma_1 = 0$  contra  $H_1: \gamma_1 \neq 0$  com base no teste LM

$$E.T.: LM = n R^2 \sim \chi^2_{(1)} \quad \text{isto é } LM = 320 R^2 \sim \chi^2_{(1)}$$

$$LM_{obs} = 320 \times 0.0078 = 2.496$$

$$\text{valor} - p = P(\chi^2_{(1)} \geq 2.496) > 0.10 \quad \text{ou } W_{5\%} = \{q: q > 3.841\}$$

Não se rejeita  $H_0$ , isto é não se rejeita a ausência de heterocedasticidade.