**Estatística II – Licenciatura em Gestão**
**ER - 24 Junho 2014**

**Parte teórica**

**Nome:** ______________________________________________________________ **Nº**________

1. **Perguntas de Verdadeiro/Falso (1.5 valores) -** Para cada afirmação assinale se esta é verdadeira (V) ou falsa (F). Uma resposta certa vale 0.3 e uma resposta errada penaliza em idêntico valor.

| | **V** | **F** |
|---|---|---|
| Um estimador só pode ser o mais eficiente se for centrado | X | |
| Quando uma população não é normal a variância corrigida da amostra é um estimador enviesado da variância da população (assumindo que esta existe). | | X |
| Num teste de dimensão $\alpha = 0.03$ em que o valor-*p*=0.04, rejeita-se $H_0$. | | X |
| Quando se afirma que o estimador dos MQ, $b_j$ é, dado *X*, linear em $y$, está-se a afirmar que $b_j = \sum_{i=1}^{n} a_{ij} y_i$, sendo os $a_{ij}$ funções dos valores contidos na matriz *X* | X | |
| O método da máxima verosimilhança tem a vantagem de originar sempre estimadores centrados. | | X |

2. **Escolha Múltipla (2.25 valores) -** Para cada pergunta assinale com X a alternativa correcta. Uma resposta certa vale 0.75 valores e uma resposta errada penaliza em 0.25 valores.

a. Quando se afirma que o modelo de regressão linear $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ com $t = 1, 2, \ldots, n$ não sofre de autocorrelação condicionada, isto significa que, para $t, s = 1, 2, \ldots, n$ com $t \neq s$:

☐ $\text{cov}(u_t, u_s \mid X) \neq 0$

☐ $\text{var}(u_t \mid X) = \sigma^2$

X $\text{cov}(u_t, u_s \mid X) = 0$

☐ $\text{var}(u_t \mid X) = \sigma^2 \, x_{t2}$

b. Num teste do qui-quadrado à bondade do ajustamento, feito com base numa amostra de dimensão $n = 100$ e a partir de uma partição com 12 classes ($m = 12$) qual das seguintes afirmações lhe parece mais adequada

☐ O número observado de elementos em cada classe tem de ser $\geq 5$

X O número esperado de elementos em cada classe tem de ser $\geq 5$

☐ Os graus de liberdade da qui-quadrado são dados por $n - m$

☐ Este teste é feito com base na distribuição do *F-Snedecor*

c. Considere o MRL $y_t = \beta_1 + \beta_2 \, x_{t2} + \cdots + \beta_k \, x_{tk} + u_t$ com base no qual se obtiveram os resíduos OLS, $\hat{u}_t$.

☐ Cada resíduo $\hat{u}_t$ pode ser considerado como uma observação da variável aleatória $u_t$

X Os resíduos $\hat{u}_t$ não podem ser considerados como observações de $u_t$ porque foram obtidos utilizando as estimativas $b_j$ e não os parâmetros $\beta_j$

☐ Os resíduos $\hat{u}_t$ não podem ser considerados como observações de $u_t$ porque nunca são observáveis

☐ Embora as variáveis $u_t$ sejam observáveis achou-se preferível substituir uma observação directa dos valores observados pelos resíduos $\hat{u}_t$

3. **Perguntas de desenvolvimento (2.25 valores) –** alínea a) 1 valor; alínea b) 1.25 valores.

a. Considere o MRL $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$ a verificar as hipóteses habituais. Sabendo que $\sum_{t=1}^{n} \hat{u}_t = 0$ e que $\sum_{t=1}^{n} x_{tj} \hat{u}_t = 0$, $j = 2,3$, sendo $\hat{u}_t$ os resíduos MQ do modelo $(t = 1,2,\cdots,n)$, prove que $\sum_{t=1}^{n} \hat{y}_t \hat{u}_t = 0$.

Como $\hat{y}_t = b_1 + b_2 x_{t2} + b_3 x_{t3}$, vem

$$\begin{aligned}\sum_{t=1}^{n} \hat{y}_t \hat{u}_t &= \sum_{t=1}^{n} (b_1 + b_2 x_{t2} + b_3 x_{t3}) \hat{u}_t = \sum_{t=1}^{n} b_1 \hat{u}_t + \sum_{t=1}^{n} b_2 x_{t2} \hat{u}_t + \sum_{t=1}^{n} b_3 x_{t3} \hat{u}_t \\ &= b_1 \sum_{t=1}^{n} \hat{u}_t + b_2 \sum_{t=1}^{n} x_{t2} \hat{u}_t + b_3 \sum_{t=1}^{n} x_{t3} \hat{u}_t \\ &= 0\end{aligned}$$

Já que $\sum_{t=1}^{n} \hat{u}_t = 0$, $\sum_{t=1}^{n} x_{t2} \hat{u}_t = 0$ e $\sum_{t=1}^{n} x_{t3} \hat{u}_t = 0$

b. Sendo $T$ um estimador para $\theta$ obtido com base numa amostra de dimensão $n$, defina o erro quadrático médio de $T$ e prove que $EQM(T) = \text{var}(T) + \left(E(T) - \theta\right)^2$

Definição de EQM(T): $EQM(T) = E(T - \theta)^2$, $\theta \in \Theta$

Definindo $X = T - \theta$ e recordando que $\text{var}(X) = E(X^2) - \left(E(X)\right)^2$ vem $\text{var}(T - \theta) = E(T - \theta)^2 - \left(E(T - \theta)\right)^2$

Ora como $\text{var}(T - \theta) = \text{var}(T)$ e $E(T - \theta) = E(T) - \theta$ obtém-se sem dificuldade

$$\text{var}(T) = EQM(T) - \left(E(T) - \theta\right)^2 \Leftrightarrow EQM(T) = \text{var}(T) + \left(E(T) - \theta\right)^2$$

**Estatística II – Licenciatura em Gestão**
**ER - 24 Junho 2014**

**Parte Prática**

**Nome:** ______________________________________________ **Nº**________

| **Espaço reservado para classificações** | | | | |
|---|---|---|---|---|
| | | | | **T:** |
| 1a. (10) | 3. (15) | 5a. (15) | 5d. (15) | **P:** _____ |
| 1b. (15) | 4.( 15) | 5b. (10) | 5e. (15) | |
| 2. (15) | | 5c. (15) | | |

**Em todos os testes de hipóteses que fizer, formule as hipóteses em teste, indique a estatística de teste e a sua distribuição. Se nada for dito em contrário utilize $\alpha = 0.05$. Para os intervalos de confiança proceda de forma semelhante referindo a variável fulcral.**
**Se necessitar de espaço dispõe de uma folha em branco no fim do enunciado**

1) Considere uma variável aleatória $X$ com distribuição de Poisson de parâmetro $\lambda$. Admita que se retirou uma amostra aleatória de dimensão $n$. Considere o seguinte estimador para $\lambda$:

$$T_1 = \frac{4X_1 - 2X_n + \sum_{i=2}^{n-1} X_i}{n}$$

a) Que pode concluir acerca do enviesamento e consistência deste estimador?

$$E(T_1) = E\left(\frac{4X_1 - 2X_n + \sum_{i=2}^{n-1} X_i}{n}\right) = \frac{1}{n}E\left(4X_1 - 2X_n + \sum_{i=2}^{n-1} X_i\right) = \frac{1}{n}\left[4E(X_1) - 2E(X_n) + \sum_{i=2}^{n-1} E(X_i)\right]$$

$$= \frac{1}{n}[4\lambda - 2\lambda + (n-2)\lambda] = \frac{1}{n}n\lambda = \lambda$$

Logo, conclui-se que o estimador $T_1$ é **não enviesado** (ou centrado) para $\lambda$.

$$Var(T_1) = Var\left(\frac{4X_1 - 2X_n + \sum_{i=2}^{n-1} X_i}{n}\right) = \frac{1}{n^2}Var\left(4X_1 - 2X_n + \sum_{i=2}^{n-1} X_i\right) =$$

$$= \frac{1}{n^2}\left[16Var(X_1) + 4Var(X_n) + \sum_{i=2}^{n-1} Var(X_i)\right] = \frac{1}{n^2}[16\lambda + 4\lambda + (n-2)\lambda] = \frac{1}{n^2}(18\lambda + n\lambda) = \frac{18\lambda}{n^2} + \frac{\lambda}{n}$$

Averiguando a consistência,

$$\lim_{n\to\infty} E(T_1) = \lim_{n\to\infty} \lambda = \lambda$$

$$\lim_{n\to\infty} Var(T_1) = \lim_{n\to\infty}\left(\frac{18\lambda}{n^2} + \frac{\lambda}{n}\right) = \lim_{n\to\infty}\frac{18\lambda}{n^2} + \lim_{n\to\infty}\frac{\lambda}{n} = 0 + 0 = 0$$

Podemos também concluir que o estimador $T_1$ é consistente para a média.

b) Qual o estimador de $\lambda$ pelo método dos momentos? Compare a eficiência do estimador do método dos momentos com a do estimador $T_1$.

Sabe-se que o estimador MM para a média populacional, $\lambda$, é a média amostral, $\bar{X}$, uma vez que igualando os primeiros momentos, populacional e amostral, se obtém $\tilde{\lambda} = \bar{X}$.

Seja então o estimador MM definido por $T_2 = \bar{X}$. A comparação em termos de eficiência baseia-se na comparação das variâncias de $T_1$ e de $T_2$, já que ambos os estimadores são centrados. Tem-se então,

$$Var(T_1) = \frac{18\lambda}{n^2} + \frac{\lambda}{n}$$

$$Var(T_2) = Var(\bar{X}) = \frac{\lambda}{n}$$

Pelo que $Var(T_1) \geq Var(T_2)$, o que nos leva a concluir que o estimador MM é mais eficiente do que o estimador apresentado.

2) Represente-se a duração de anúncios publicitários, em segundos, por uma variável aleatória $X$. Para testar a adequação de uma distribuição exponencial de média 10 segundos, recolheu-se uma amostra de 100 tempos, registando-se a seguinte informação:

| Duração (Segundos) | $(0,5]$ | $(5,10]$ | $(10,15]$ | $(15,20]$ | $\geq 20$ |
|---|---|---|---|---|---|
| Frequência Observada | 25 | 30 | 15 | 20 | 10 |
| Frequência Esperada | ***A*** | ***B*** | 14.475 | 8.779 | 13.534 |

Calcule as duas primeiras frequências esperadas em falta no quadro (***A*** e ***B***) e efetue o teste em questão, considerando um nível de 1%. Que pode concluir?

$$H_0: X \sim Ex(0.1) \quad vs \quad H_1: H_0\ falsa$$

$$Q = \sum_{j=1}^{m} \frac{(N_j - fe_j)^2}{fe_j} \overset{a}{\sim} \chi^2(m-1)$$

$$\boldsymbol{A} = nP(X < 5) = 100\int_0^5 0.1e^{-0.1x}\,dx = 100(1 - e^{-0.5}) \approx 100 \times 0.39347 = 39.347$$

$$\boldsymbol{B} = nP(5 \leq X < 10) = 100\int_5^{10} 0.1e^{-0.1x}\,dx = 100(e^{-0.5} - e^{-1}) \approx 100 \times 0.23865 = 23.865$$

Todas $fe_j \geq 5 \Rightarrow m = 5 \Rightarrow Q \overset{a}{\sim} \chi^2(4)$

$$q_{obs} = \frac{(25 - 39.347)^2}{39.347} + \frac{(30 - 23.865)^2}{23.865} + \cdots + \frac{(10 - 13.534)^2}{13.534} \approx 22.093$$

Valor-p: $p_{obs} = P(Q \geq q_{obs}|H_0) = P(Q \geq 22.093) \approx 0.0002 < \alpha = 0.01$

Região de rejeição: $W_{0.01} = \{q_{obs}: q_{obs} > q_{0.01} = 13.277\} \Rightarrow q_{obs} \in W_{0.01}$

Rejeita-se assim $H_0$ ao nível de 1%. Evidência estatística desfavorável a $X \sim Ex(0.1)$.

3) Uma grande empresa de venda ao retalho tem, naturalmente, um departamento de "reclamações". Observada uma amostra de 500 reclamações escolhidas aleatoriamente, obtiveram-se os seguintes elementos acerca do tempo de resolução das reclamações (em minutos):

| Mínimo | Máximo | Média | Desvio-padrão corrigido |
|---|---|---|---|
| 4.5 | 121.3 | 19.5 | 25.3 |

Construa um intervalo de confiança a 90% para o tempo médio necessário à resolução de uma reclamação, não se esquecendo de referir a variável fulcral que utilizar.

$X$ — tempo (em minutos) necessário à resolução, em minutos

Grandes Amostras

$$Z = \frac{\bar{X} - \mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

IC a 90% para $\mu$:

$$\left(\bar{x} \pm z_{0.05}\frac{s'}{\sqrt{n}}\right) = \left(19.5 \pm 1.645\frac{25.3}{\sqrt{500}}\right) = (17.64, 21.36)$$

De acordo com o intervalo obtido pode concluir-se, com uma confiança de 90%, que o tempo médio necessário à resolução de uma reclamação se situa entre 17.64 e 21.36 minutos.

4) A gerência de uma padaria decidiu implementar um novo processo de fabrico, o qual originou uma redistribuição de tarefas e a necessidade de rever os parâmetros salariais. Recolhida uma amostra de 20 trabalhadores calculou-se a diferença entre o salário atual e o antigo para cada trabalhador, diferença essa que se denominou por $d_i$, $i = 1, 2, \cdots, 20$. ,Posteriormente calculou-se a média e o desvio-padrão das diferenças individuais observadas, tendo-se obtido $\bar{d} = 25$ euros e $s'_D$=50 euros respectivamente. Admita que o salário de um trabalhador segue uma distribuição normal. Com base num teste a 5% diga se o novo processo resultou num aumento do salário médio.

Amostras emparelhadas (1 - Novo salário; 2 - Antigo salário)

$$H_0: \mu_1 = \mu_2 \quad vs \quad H_1: \mu_1 > \mu_2$$

$$T = \frac{\bar{D}}{S'_D/\sqrt{n}} \sim t(n-1) = t(19)$$

$$t_{obs} = \frac{25}{50/\sqrt{20}} \approx 2.236$$

$$p_{obs} = P(T \geq t_{obs}|H_0) = P(T \geq 2.236) \approx 0.019$$

Desta forma, rejeita-se $H_0$ ao nível de 5%, ou seja, para $\alpha = 0.05$, o aumento do salário médio é estatisticamente significativo.

5) Encontramo-nos perante um estudo sobre que variáveis afetam o consumo de peras produzidas ecologicamente, o que levou à formulação do seguinte modelo:

$$ecokg = \beta_1 + \beta_2 ecoprc + \beta_3 normprc + \beta_4 rendfam + \beta_5 nind + u$$

Onde:

- $ecokg$ — quantidade, em kg, de peras ecológicas adquiridas semanalmente;
- $ecoprc$ — preço das peras ecológicas, em euros/kg;
- $normprc$ — preço das peras "normais", em euros/kg;
- $rendfam$ — rendimento familiar anual, em milhares de euros;
- $nind$ — número de indivíduos na família.

Tendo tido acesso a uma amostra de 300 famílias, procedeu-se à estimação deste modelo, cujos resultados são apresentados na **Equação 1**, em anexo. Em anexo também se encontram outputs de algumas regressões que poderá utilizar sempre que necessário.

a) Interprete o efeito estimado do preço das peras ecológicas e analise também a significância estatística desta variável explicativa, considerando um nível de significância de 1%.

$b_2 = -1.8714 \Rightarrow$ Tudo o resto constante, se o preço/kg das peras ecológicas aumentar 1 euro, estima-se que, em média, a quantidade semanal de peras ecológicas adquirida diminui cerca de 1.87 quilos.

Para analisar a significância estatística da variável "preço das peras ecológicas" efetua-se o seguinte teste de hipóteses

$$H_0: \beta_2 = 0 \quad vs \quad H_1: \beta_2 \neq 0$$

Recorrendo à estatística de teste

$$t_2 = \frac{b_2}{s_{b_2}} \sim t(n-k) = t(295)$$

De acordo com o output da **EQUAÇÃO 1**, o valor-p associado a este teste é de 0.0002. Considerando $\alpha = 0.01$, rejeita-se a hipótese nula ao nível de 1%. Ou seja, podemos concluir que o preço das peras ecológicas é um fator estatisticamente relevante sobre a quantidade adquirida de peras ecológicas.

b) As variáveis associadas à família ("$rendfam$" e "$nind$") são, no seu conjunto, fatores estatisticamente relevantes na aquisição de peras ecológicas? Efetue um teste a 10%.

$$H_0: \beta_4 = \beta_5 = 0 \quad vs \quad H_1: \exists \beta_j \neq 0 \ (j = 4, 5)$$

$$F = \frac{(R^2 - R_0^2)/m}{(1 - R^2)/(n-k)} \sim F(m, n-k) = F(2, 295)$$

Recorrendo aos outputs das **EQUAÇÕES 1** e **2**, obtém-se:

$$f_{obs} = \frac{(0.0644 - 0.0466)/2}{(1 - 0.0644)/(300 - 5)} \approx 2.806$$

$$p_{obs} = P(F \geq f_{obs} | H_0) = P(F \geq 2.806) \approx 0.062 < \alpha = 0.10$$

Logo, rejeita-se $H_0$ ao nível de 10%. Isto significa que as variáveis associadas à família (rendimento familiar e número de indivíduos) são, no seu conjunto, estatisticamente significativas na determinação da quantidade adquirida de peras ecológicas, devendo, portanto, permanecerem no modelo em análise.

c) Teste se, em média, o efeito de um aumento de 1 euro/kg nas peras ecológicas é anulado por um aumento no mesmo montante/kg nas peras normais. Para tal, admita que $\widehat{Cov}(b_2, b_3) = -0.2443$.

$$H_0: \beta_2 + \beta_3 = 0 \quad vs \quad H_1: \beta_2 + \beta_3 \neq 0$$

Definindo $\delta = \beta_2 + \beta_3$, quer-se testar

$$H_0: \delta = 0 \quad vs \quad H_1: \delta \neq 0$$

$$t_{\hat{\delta}} = \frac{\hat{\delta}}{s_{\hat{\delta}}} \sim t(n-k) = t(295)$$

Para o cálculo do valor observado da estatística de teste, tem-se pela **EQUAÇÃO 1**:

$\hat{\delta} = b_2 + b_3 = -1.8714 + 1.7277 = -0.1437$

$s_{\hat{\delta}} = \sqrt{\widehat{Var}(\hat{\delta})} = \sqrt{\widehat{Var}(b_2 + b_3)} = \sqrt{\widehat{Var}(b_2) + \widehat{Var}(b_3) + 2\widehat{Cov}(b_2, b_3)} =$

$= \sqrt{s_{b_2}^2 + s_{b_3}^2 + 2\widehat{Cov}(b_2, b_3)} = \sqrt{0.4909^2 + 0.5992^2 + 2 \times (-0.2443)} \approx 0.3338$

Assim,

$t_{obs} = \frac{-0.1437}{0.3338} \approx -0.4305$

$p_{obs} = 2P\left(t_{\hat{\delta}} \geq |t_{obs}| \middle| H_0\right) = 2P\left(t_{\hat{\delta}} \geq 0.4305\right) \approx 0.667$

O resultado do valor-p conduz à incontestável decisão de não rejeitar $H_0$, ou seja, de o aumento do preço nas peras ecológicas não ter impacto significativo na sua quantidade adquirida quando o preço das peras normais aumenta no mesmo montante.

d) A família Araújo gosta imenso de fazer as suas compras no supermercado Hiperpreço. Considere que numa certa semana nesse supermercado, o preço das peras ecológicas é de 1.25 euros/kg e o preço das peras normais é de 1 euro/kg. Tendo em conta o anexo, indique o número de elementos e qual o rendimento familiar anual da família Araújo e, por fim, obtenha um intervalo de previsão a 80% para a quantidade de peras que esta família irá adquirir nessa semana.

<u>Previsão Pontual</u>

De acordo com a **EQUAÇÃO 3**, podemos concluir que a família Araújo é composta por 4 pessoas e tem um rendimento familiar anual de 50 mil euros.

Quer-se então um IP a 80% para $ecokg_0$ quando $ecoprc = 1.25$, $normprc = 1.25$, $rendfam = 50$ e $nind = 4$.

$$t_d = \frac{ecokg_0 - \widehat{ecokg}_0}{s_d} \sim t(n-k) = t(295)$$

IP: $\left(\widehat{ecokg}_0 \pm z_{0.1} \times s_d\right)$, onde $z_{0.1} = 1.282$

Pela **EQUAÇÃO 3**, obtém-se:

$\widehat{ecokg}_0 = 2.2951$ (estimativa para o termo independente)

$s_d = \sqrt{s^2 + s_{\hat{\theta}}^2} = \sqrt{1.3684^2 + 0.1107^2} \approx 1.3729$

Logo, conclui-se que o IP a 80% para a quantidade de peras ecológicas adquirida é:

$(2.2951 \pm 1.282 \times 1.3729) = (0.535, 4.055)$

Isto significa que, nessa semana, há uma confiança de 80% de que a família Araújo irá adquirir uma quantidade de peras ecológicas entre cerca de meio quilo e quatro quilos.

e) Outra informação que se tem acerca das famílias inquiridas é se fazem ou não a reciclagem do seu lixo doméstico. Em anexo, na **Equação 4**, encontra o modelo estimado apenas para as famílias que fazem reciclagem, enquanto na **Equação 5** encontra o modelo estimado apenas para famílias que não fazem reciclagem. Qual o propósito desta metodologia? Através de um teste adequado, diga o que pode concluir.

Com esta metodologia pretendemos averiguar se a relação entre quantidade adquirida de peras ecológicas e as variáveis explicativas é ou não semelhante para os dois grupos de famílias considerados (as que fazem reciclagem e as que não fazem). Ou seja, quer-se efetuar um teste de alteração de estrutura, mais especificamente o **teste de Chow**.

$$H_0: \beta_{11} = \beta_{12}, \beta_{21} = \beta_{22}, \dots, \beta_{51} = \beta_{52} \quad vs \quad H_1: H_0 \; falsa$$

$$F = \frac{(VR_0 - VR_1)/k}{VR_1/(n-2k)} \sim F(k, n-2k) = F(5, 290)$$

Para o cálculo do valor observado da estatística de teste, tem-se:

$VR_0 = 552.4173$ (VR da **EQUAÇÃO 1**)

$VR_1 = 103.8044 + 437.4979 = 541.3023$ (Soma das VR das **EQUAÇÕES 4** e **5**)

Assim,

$$f_{obs} = \frac{(552.4173 - 541.3023)/5}{541.3023/(300 - 2 \times 5)} \approx 1.191$$

$$p_{obs} = P(F \geq f_{obs} | H_0) = P(F \geq 1.191) \approx 0.314$$

Pode concluir-se que existe margem suficiente para não rejeitar $H_0$. A evidência estatística é favorável à permanência de estrutura do modelo para ambos os grupos de famílias.

## ANEXO

**EQUAÇÃO 1 –** variável dependente: *ecokg*

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.2537 |
| R Square | 0.0644 |
| Adjusted R Square | 0.0517 |
| Standard Error | 1.3684 |
| Observations | 300 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 38.0035 | 9.5009 | 5.0736 | 0.0006 |
| Residual | 295 | 552.4173 | 1.8726 | | |
| Total | 299 | 590.4207 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | 2.3621 | 0.3839 | 6.1522 | 0.0000 | 1.6064 | 3.1177 |
| ecoprc | -1.8714 | 0.4909 | -3.8122 | 0.0002 | -2.8374 | -0.9053 |
| normprc | 1.7277 | 0.5992 | 2.8832 | 0.0042 | 0.5484 | 2.9070 |
| rendfam | 0.0023 | 0.0023 | 1.0226 | 0.3073 | -0.0021 | 0.0068 |
| nind | 0.1071 | 0.0546 | 1.9617 | 0.0507 | -0.0003 | 0.2146 |

**EQUAÇÃO 2 –** variável dependente: *ecokg*

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.2160 |
| R Square | 0.0466 |
| Adjusted R Square | 0.0402 |
| Standard Error | 1.3767 |
| Observations | 300 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 2 | 27.5375 | 13.7688 | 7.2650 | 0.0008 |
| Residual | 297 | 562.8832 | 1.8952 | | |
| Total | 299 | 590.4207 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | 2.8623 | 0.3128 | 9.1502 | 0.0000 | 2.2467 | 3.4780 |
| ecoprc | -1.8415 | 0.4927 | -3.7377 | 0.0002 | -2.8111 | -0.8719 |
| normprc | 1.6285 | 0.6011 | 2.7092 | 0.0071 | 0.4456 | 2.8115 |

**EQUAÇÃO 3 –** variável dependente: *ecokg*

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.2537 |
| R Square | 0.0644 |
| Adjusted R Square | 0.0517 |
| Standard Error | 1.3684 |
| Observations | 300 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 38.0035 | 9.5009 | 5.0736 | 0.0006 |
| Residual | 295 | 552.4173 | 1.8726 | | |
| Total | 299 | 590.4207 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | 2.2951 | 0.1107 | 20.7263 | 0.0000 | 2.0772 | 2.5130 |
| ecoprc-1.25 | -1.8714 | 0.4909 | -3.8122 | 0.0002 | -2.8374 | -0.9053 |
| normprc-1 | 1.7277 | 0.5992 | 2.8832 | 0.0042 | 0.5484 | 2.9070 |
| rendfam-50 | 0.0023 | 0.0023 | 1.0226 | 0.3073 | -0.0021 | 0.0068 |
| nind-4 | 0.1071 | 0.0546 | 1.9617 | 0.0507 | -0.0003 | 0.2146 |

**EQUAÇÃO 4 –** variável dependente: *ecokg*

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.2609 |
| R Square | 0.0681 |
| Adjusted R Square | 0.0242 |
| Standard Error | 1.1051 |
| Observations | 90 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 7.5833 | 1.8958 | 1.5524 | 0.1945 |
| Residual | 85 | 103.8044 | 1.2212 | | |
| Total | 89 | 111.3876 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper |
|---|---|---|---|---|---|---|
| Intercept | 2.7236 | 0.6142 | 4.4341 | 0.0000 | 1.5023 | 3.9449 |
| ecoprc | -1.2439 | 0.6970 | -1.7847 | 0.0779 | -2.6298 | 0.1419 |
| normprc | 0.9405 | 0.8659 | 1.0862 | 0.2805 | -0.7811 | 2.6620 |
| rendfam | 0.0044 | 0.0033 | 1.3553 | 0.1789 | -0.0021 | 0.0109 |
| nind | -0.0565 | 0.0819 | -0.6900 | 0.4921 | -0.2194 | 0.1063 |

**EQUAÇÃO 5 –** variável dependente: *ecokg*

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.2929 |
| R Square | 0.0858 |
| Adjusted R Square | 0.0679 |
| Standard Error | 1.4609 |
| Observations | 210 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 41.0492 | 10.2623 | 4.8086 | 0.0010 |
| Residual | 205 | 437.4979 | 2.1341 | | |
| Total | 209 | 478.5471 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | 2.1704 | 0.4857 | 4.4687 | 0.0000 | 1.2128 | 3.1280 |
| ecoprc | -2.0318 | 0.6468 | -3.1414 | 0.0019 | -3.3070 | -0.7566 |
| normprc | 1.9685 | 0.7807 | 2.5216 | 0.0124 | 0.4293 | 3.5077 |
| rendfam | 0.0013 | 0.0029 | 0.4472 | 0.6552 | -0.0045 | 0.0071 |
| nind | 0.2030 | 0.0742 | 2.7340 | 0.0068 | 0.0566 | 0.3493 |