E aí pessoal? Vamos à resolução de nossa prova?

A CESPE pegou muito pesado, conforme eu tinha alertado em vídeos e no nosso pdf. Ela utilizou conceitos de Estatística muito avançados para uma matéria de "Noções de Estatística". Mas, vamos a elas.

Exercício 1

Com o objetivo de modelar a arrecadação anual do ICMS em municípios brasileiros (y), o modelo de regressão linear múltipla foi representado, na forma matricial, como $y = X\beta + \varepsilon$, em que y representa o vetor de respostas, X denota a matriz de delineamento, β é o vetor de parâmetros e ε é o vetor de erros aleatórios independentes e identicamente distribuídos. Considerando-se que $\boldsymbol{X'}$ representa a transposta da matriz de delineamento, apresenta-se a seguir a matriz inversa do produto matricial $\boldsymbol{X'X}$ produzida no modelo.

$$(X'X)^{-1} = \begin{bmatrix} 0{,}50 & -0{,}25 & -0{,}40 \\ -0{,}25 & 0{,}40 & 0{,}10 \\ -0{,}40 & 0{,}10 & 0{,}50 \end{bmatrix}$$

Com base nessas informações, e sabendo que $\boldsymbol{X'y} = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$, julgue o próximo item.

91 A estimativa do vetor de parâmetros produzida pelo método de mínimos quadrados ordinários é $\begin{bmatrix} 3{,}5 \\ 0 \\ -2 \end{bmatrix}$.

Resolução

Nós estudamos isso em nossa aula extra de Regressão Múltipla!

Eu já expliquei para vocês que uma Regressão Linear Múltipla, com k parâmetros, é dada, genericamente, por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \dots \beta_k x_{ki} + \varepsilon_i$$

De forma que i represente qual observação está sendo avaliada para cada uma das variáveis, de forma que i = 1, 2...n.

Assim, vocês vão pensar com bastante calma e enxergar que é possível dispor os dados da seguinte forma:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_k x_{k1} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_k x_{k2} + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{kn} + \varepsilon_n$$

E, portanto, também podemos escrever:

$$\begin{vmatrix} y_1 \\ \vdots \\ y_n \end{vmatrix} = \begin{vmatrix} 1 & \cdots & x_{k1} \\ \vdots & \dots & \vdots \\ 1 & \dots & x_{kn} \end{vmatrix} \cdot \begin{vmatrix} \beta_0 \\ \vdots \\ \beta_k \end{vmatrix} + \begin{vmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{vmatrix}$$

O que pode ser reduzido à notação matricial:

$$Y = X\beta + \varepsilon$$

Opa! Agora vocês estão entendendo o porquê de todo aquele falatório sobre matrizes? Esta última forma é a melhor forma de se avaliar uma regressão múltipla, pois simplifica, e muito, a descrição das operações realizadas.

Então, o estimador de MQO para a regressão múltipla é o seguinte:

$$\beta = (X'X)^{-1}(X'Y)$$

Olha só pessoal! Percebam a semelhança entre esta expressão e a expressão utilizada na regressão simples. No caso, $(X'Y)$ associa-se ao fator $\sum xy$, enquanto que $(X'X)^{-1}$ é análogo a $\frac{1}{\sum x^2}$.

Assim, temos de multiplicar estas duas matrizes. O que temos de fazer multiplicar as linhas pelas colunas e somar o resultado. No caso:

$$\begin{bmatrix} 0{,}50 & -0{,}25 & -0{,}40 \\ -0{,}25 & 0{,}40 & 0{,}10 \\ -0{,}40 & 0{,}10 & 0{,}50 \end{bmatrix}\begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix} = \begin{bmatrix} 0{,}5 \times 20 - 0{,}25 \times 10 - 0{,}40 \times 10 \\ -0{,}25 \times 20 + 0{,}4 \times 10 + 0{,}1 \times 10 \\ -0{,}4 \times 20 + 0{,}1 \times 10 + 0{,}5 \times 10 \end{bmatrix} = \begin{bmatrix} 3{,}5 \\ 0 \\ -2 \end{bmatrix}$$

Alternativa Correta.

Exercício 2

Uma empresa publicou um relatório acerca das previsões para sua receita operacional nos próximos meses. Essas previsões foram obtidas com base em um modelo estacionário de séries temporais na forma $X_t = 2 + 0{,}5X_{t-1} + a_t$, em que X_t representa a receita operacional no mês t e a_t é um ruído branco (no instante t) que possui média nula e variância igual a 3.

A partir dessas informações, julgue o item abaixo.

O modelo apresentado é um processo autorregressivo de primeira ordem, AR(1), em que a média e o desvio padrão de X_t são, respectivamente, iguais a 4 e 2.

Resolução

Para resolver esta questão precisamos dos operadores Esperança e Variância, lembra-se?

Para encontrar a média do processo, aplique o operador esperança na equação:

$E(X_t) = E(2 + 0{,}5X_{t-1} + a_t)$

Ora, pelas propriedades da esperança sabemos que:

$E(X_t) = E(2) + E(0{,}5X_{t-1}) + E(a_t)$

A primeira coisa a perceber é que, para calcularmos a esperança incondicional:

$$E(X_t) = E(X_{t-1})$$

Ou seja, queremos saber qual a média do processo, que no caso, seria constante ao longo do processo.

A média de uma constante (2) é igual à constante e a média dos erros é zero, segundo o enunciado:

$$E(X_t) = 2 + 0{,}5E(X_t) \rightarrow 0{,}5E(X_t) = 2 \rightarrow E(X_t) = 4$$

E a variância?

Mesmo jeito! Aplique o operador variância:

$$Var(X_t) = Var(2 + 0{,}5X_{t-1} + a_t)$$

Lembre-se das propriedades da variância e saiba que somar uma constante em uma variável não altera a variância, portanto:

$$Var(X_t) = Var(0{,}5X_{t-1} + a_t)$$

Como trata-se de um ruído branco, sabemos que o mesmo respeita as hipóteses básicas do modelo de regressão, portanto:

$$Cov(X_t, a_t) = 0$$

Assim, a soma da variância destas duas variáveis será:

$$Var(X_t) = Var(0{,}5X_{t-1}) + Var(a_t)$$

Da mesma forma que na aplicação anterior, para encontrar a variância incondicional do processo:

$$Var(X_t) = Var(X_{t-1})$$

Substituindo:

$$Var(X_t) = Var(0{,}5X_t) + Var(a_t)$$

Ao substituir o valor da variância do ruído branco e tirando a constante que multiplica X ao quadrado (lembre-se da propriedade de multiplicação de constante nas propriedades da variância):

$$Var(X_t) = 0{,}25Var(X_t) + 3 \rightarrow Var(X_t) = \frac{3}{0{,}75} = 4$$

Se a variância é 4, o desvio padrão é 2.

Alternativa correta.

Exercício 3

Em dado relatório de avaliação da qualidade do transporte aéreo, considerou-se a relação entre o nível de estresse de controladores de tráfego aéreo e a ocorrência de incidentes aeronáuticos. Para o estudo, foram selecionados ao acaso 251 controladores de tráfego aéreo, que foram separados em dois grupos, de acordo com seus níveis de estresse. A tabela a seguir mostra a quantidade de incidentes registrados dentro de cada grupo.

nível de estresse	ocorrência de incidentes		
	sim	não	total
baixo	1	40	41
alto	10	200	210

Tendo como referência as informações acima, julgue o item a seguir, considerando que o logaritmo natural da razão de chances (*odds ratio*) é representado por $\ln(\hat{\omega})$, e que sua distribuição amostral é gaussiana.

Julgue a afirmativa abaixo:

O erro padrão de $\ln(\hat{\omega})$ é inferior a 1.

Resolução

Questão puramente decorativa. Dada uma tabela de razão de chances:

	ocorrência de incidentes	
Nível de estresse	Sim	Não
baixo	a	c
alto	b	d

O erro padrão do logaritmo natural da razão de chances é dado por:

$$erro\ padrão = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Assim, substitua os valores do enunciado:

$$erro\ padrão = \sqrt{\frac{1}{1} + \frac{1}{40} + \frac{1}{10} + \frac{1}{200}}$$

Aplicando um método de simplificação:

$$erro\ padrão = \sqrt{\frac{200 + 5 + 20 + 1}{200}} = \sqrt{\frac{226}{200}}$$

Não precisa nem resolver. Como esta fração é maior do que 1, a raiz também é.

Alternativa falsa.

Exercício 4

Considerando que uma amostra aleatória simples X_1, X_2, X_3, X_4 tenha sido retirada de uma distribuição X cuja função de probabilidade é definida como $P(X=k)=\binom{10}{k}p^k(1-p)^{10-k}$, em que $0 \le p \le 1$, $k \in \{0, 1, 2, ..., 10\}$, sendo p o parâmetro desconhecido, e que os valores observados na amostra tenham sido 0, 4, 6 e 2, julgue o item a seguir.

A estimativa de máxima verossimilhança para a variância populacional é igual a 2,1.

Resolução

Você percebeu fácil que se trata de uma binomial, certo?

Pessoal, não tente encontrar a função e verossimilhança e derivar! Vamos ao mais fácil. Você aprendeu no nosso curso que:

Só para finalizarmos, como seria o estimador MLE para a probabilidade de sucesso de um evento em uma distribuição binomial.

Lembram-se da distribuição binomial? É aquela em que há dois eventos possíveis, um considerado "sucesso", com probabilidade p, e outro, mutuamente exclusivo, considerado "fracasso", com probabilidade $1-p$.

Então suponha que em uma amostra com n elementos, x apresentam o atributo sucesso. Você consegue adivinhar qual o estimador MLE para p?

$$p = \frac{x}{n}$$

Ou seja, é a própria proporção deste elemento na amostra como um todo, tal como no caso da média!

Bom, O estimador de Máxima Verossimilhança (MLE) será dado pela própria proporção de ocorrência.

No caso, a nossa variável X tem a ver com a quantidade de sucessos de um determinado experimento. Veja que o total de experimentos é de 10.

Como eu sei isso?

Veja a fórmula da binomial:

$$P(X = k) = C_{n,k} \times p^{k} \times p^{n-k}$$

No caso, onde devia estar ***n*** está o número 10!

A nossa quantidade de sucessos nos 4 experimentos foram:

$$sucessos: 0(1^{º}experimento); 4(2^{º}experimento); 6(3^{º}experimento); 2(4^{º}experimento)$$

Qual a proporção de sucessos média?

$$p = \frac{total\ de\ sucessos}{total\ de\ experimentos} = \frac{0+4+6+2}{10+10+10+10} = \frac{12}{40} = 0{,}3$$

Esta é a estimativa de verossimilhança para *p*.

Qual a fórmula da variância de uma binomial?

$$Variância = np - np^{2} = 3 - 0{,}9 = 2{,}1$$

Alternativa correta.

Exercício 5

Considerando duas variáveis aleatórias independentes X e Y que seguem distribuições normal padrão, julgue o próximo item.

A diferença $X - Y$ segue uma distribuição normal cuja variância é igual ou inferior a 1.

Resolução

Qual é a variância da operação de diminuição de duas variáveis?

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

As variáveis são independentes, portanto:

$$Cov(X, Y) = 0$$

Assim:

$$Var(X - Y) = Var(X) + Var(Y)$$

Qual é o valor da variância destas variáveis? Tratam-se de variáveis normais padronizadas, portanto sabemos que este é um caso em que as variáveis tem média zero e variância igual a 1. Portanto:

$$Var(X - Y) = Var(X) + Var(Y) = 1 + 1 = 2$$

Alternativa falsa.

Boa pessoal! Espero que tenham ido bem.

Estou sempre à disposição para qualquer dúvida de Estatística\Economia\Econometria.

Um abraço

jeroymobj@hotmail.com