Lista 06a de Estatística II (aula 06)

1. O modelo de regressão linear simples a ser estimado $y_t = \beta_1 + \beta_2 x_t + \mu_t$ é composto de quatro partes básicas, qual o significado de cada uma delas?

a) y_t variável explicativa; x_t variável explicada; β_1 e β_2 parâmetros; e μ_t variável aleatória normalmente distribuída com média zero e variância constante. ( )

b) y_t variável explicada; x_t variável explicativa; β_1 e β_2 parâmetros; e μ_t variável aleatória normalmente distribuída com média zero e variância constante. ( )

c) y_t variável explicativa; x_t variável explicada; β_1 e β_2 parâmetros; e μ_t variável aleatória com média diferente de zero e variância constante. ( )

d) y_t variável explicada; x_t variável explicativa; β_1 e β_2 parâmetros; e μ_t variável estocástica com média zero e variância pequena e não constante. ( ).

Resp. b

2. O economista-chefe de uma cafeteria resolveu calcular a correlação linear entre a quantidade consumida de açúcar (quilos) em relação à quantidade vendida de café (número de xícaras) e em relação à quantidade vendida de sucos naturais (litros) a partir de dados diários levantados durante os três primeiros meses do ano. Os resultados encontrados foram:

Correlação entre quantidade consumida de açúcar e quantidade vendida de café: 0,70
Correlação entre quantidade consumida de açúcar e quantidade vendida de sucos naturais 0,35

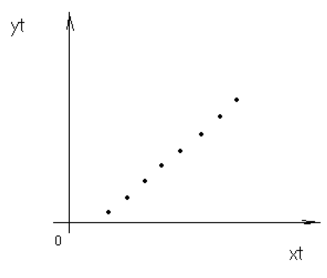
Explique:
a) o significado das correlações encontradas
b) o fato de a correlação linear entre quantidade consumida de açúcar e quantidade vendida de café ser maior que a correlação linear entre quantidade consumida de açúcar e a quantidade vendida de sucos naturais tem algum significado especial?

Resp.
2. a) o coeficiente de correlação é uma medida que varia entre [-1, 1] que, assim como a covariância (o coeficiente de correlação é calculado a partir dela), mede o grau de associação linear entre duas variáveis.

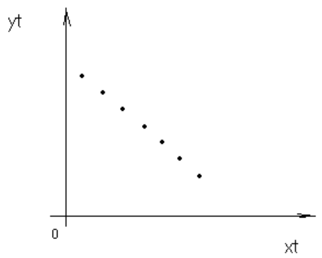
$$\rho = \frac{Cov[X,Y]}{\sqrt{Var[X].Var[Y]}} .$$

Se $\rho = 1$ ou $\rho = -1$, tem-se que as duas variáveis estudadas y_t e x_t possuem um grau de associação linear perfeito. Ou seja, se eu traçar uma reta no plano cartesiano de (y_t , x_t) todos os pares ordenados de y_t e x_t observados estão contidos exatamente nesta reta.

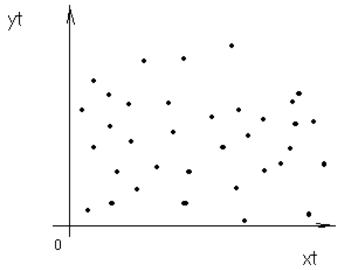
Se $\rho = 1$



Se $\rho = -1$



Se $\rho = 0$ não existe associação linear entre y_t e x_t.



b)
Se o coeficiente de correlação entre café e açúcar é de 0,70 significa que existe um grau de associação linear positiva relativamente alto entre estas variáveis observadas. Fazendo-nos crer que café e açúcar são bens complementares.
No que se refere à correlação entre suco natural e açúcar, como o grau de associação linear é baixo, 0,30, tendemos a crer que não existe relação de complementariedade para estes dois bens nesta cafeteria.

3. Sobre parâmetros e estimadores, podemos dizer que a única alternativa que não está correta é:
a) Os parâmetros são características da população e os estimadores são características amostrais. ( ).
b) Os parâmetros são características da amostra e os estimadores são características populacionais. ( ).
c) Os $\hat{\beta}_1 = \dfrac{\sum_{i=i}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ e $\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$ calculados por Mínimos quadrados a partir de um modelo de regressão linear simples são estimadores dos parâmetros verdadeiros β_1 e β_2. ( )

d) A média populacional é um parâmetro e a média amostral $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ é um estimador deste parâmetro populacional com distribuição normal para o caso de amostras grandes. ( )

e) A variância populacional é um parâmetro e a variância amostral $V(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ é um estimador deste parâmetro populacional. ( )

b Pois, parâmetro é uma característica da população, como por exemplo média e variância, que por ser, às vezes impossível medi-los, usamos amostras da população para obter estimadores destes parâmetros. De modo que estimador é uma característica de amostra, como média e variância, que é usada para avaliar os parâmetros verdadeiros da população,

4. Qual o significado dos estimadores da regressão linear simples quando se utiliza o logaritmo das variáveis x_t e y_t ao invés das mesmas em nível?
Resp. Relações percentuais entre x_t e y_t.

5. (ver lista6aapendice.xls). Suponha que a Sanduichov, uma nova cadeia de lanchonetes em Moscou, esteja indecisa quanto a sua política de preços. Cada mês a firma modifica ligeiramente o preço do produto, utilizando vários artifícios. A tabela mostra as quantidades vendidas e os preços correspondentes. Suponha que a equação de demanda que relaciona preço (p_t) com quantidade vendida (q_t) seja $\ln(q_t) = \beta_1 + \beta_2 \ln(p_t) + \mu_t$. Note que esta equação pode ser escrita de maneira mais familiar $y_t = \beta_1 + \beta_2 x_t + \mu_t$, definindo $y_t = \ln q_t$ e $x_t = \ln p_t$.

Lanchonete Sanduichov		
mês	Quant. Vendida	Preço pt
1	892	1,23
2	1012	1,15
3	1060	1,1
4	987	1,2
5	680	1,35
6	739	1,25
7	809	1,28
8	1275	0,99
9	946	1,22
10	874	1,25
11	720	1,3
12	1096	1,05

a) Calcule $y_t = \ln q_t$ e $x_t = \ln p_t$. (ver lista6apendice.xls).

b) Suponha que as hipóteses de regressão simples sejam válidas para os valores transformados x_t e y_t. Então ache as estimativas de mínimos quadrados de $\hat{\beta}_0$ e $\hat{\beta}_1$ para β_1 e β_2.

c) Dê uma interpretação para $\hat{\beta}_1$.

Resp. Trata-se da elasticidade preço da demanda. Pois, mede quanto a variação percentual no preço afeta em termos percentuais o valor da quantidade demandada.

(ver lista6apendice.xls).

6. Calcule as estimativas de mínimos quadrados de $\hat{\beta}_0$ e $\hat{\beta}_1$ para β_1e β_2, para a equação de consumo explicada pela renda, usando os dados abaixo. O sinal dos coeficientes estão de acordo com o esperado? (use: lista6aapendice.xls)

3,1 Despesa com Alimentaçào e Renda		
t	yt	xt
1	52,25	258,3
2	58,32	343,1
3	81,79	425
4	119,9	467,5
5	125,8	482,9
6	100,46	487,7
7	121,51	496,5
8	100,08	516,4
9	127,75	543,3
10	104,94	548,7
11	107,48	564,6
12	98,48	588,3
13	181,21	591,3
14	122,23	607,3
15	129,57	611,2
16	92,84	631
17	117,92	659,6
18	82,13	664
19	182,28	704,2
20	139,13	704,8
21	98,14	98,14
22	123,94	123,94
23	126,31	126,31
24	146,47	146,47
25	115,98	115,98
26	207,23	207,23
27	119,8	119,8
28	151,33	151,33
29	169,51	169,51
30	108,03	108,03
31	168,9	168,9
32	227,11	227,11
33	84,94	84,94
34	98,7	98,7
35	141,06	141,06
36	215,4	215,4
37	112,89	112,89
38	166,25	166,25
39	115,43	115,43
40	269,03	269,03

7. Quando estimamos um modelo de regressão pelo método dos mínimos quadrados $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t$, buscamos descrever o modelo desconhecido $y_t = \beta_1 + \beta_2 x_t + \mu_t$ do qual só observamos o y_t e x_t . Com base nisso assinale a alternativa incorreta:

a) $\hat{\mu}_t = y_t - \hat{y}_t$ é o erro estimado para o erro não observado μ_t t ( )

b) $\boxed{\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}(\hat{\mu}_t)^2}$ é outra forma de descrever como a variação da variável observada y_t (SQT) pode ser decomposta numa parte explicada pelo modelo estimado $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ (SQE) e na outra não explicada, obtida por resíduo (SQR). ( )

c) O coeficiente de determinação $R^2 = \dfrac{SQR}{SQT} = 1 - \dfrac{SQE}{SQT}$ é um valor entre 0 e 1 , ou seja $0 \le R^2 \le 1$ ( )

d) O coeficiente de determinação mede a variação de y_t não explicada pelo modelo estimado de regressão $\hat{y}_t$ ( )

e) O coeficiente de determinação mede a variação de y_t explicada pelo modelo estimado de regressão $\hat{y}_t$ ( )

Resp. d

8.Quando estimamos um modelo de regressão pelo método dos mínimos quadrados $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t$, buscamos descrever o modelo desconhecido $y_t = \beta_1 + \beta_2 x_t + \mu_t$ do qual só observamos o y_t e x_t . Com base nisso assinale F se for falso e V se for verdadeiro:

a) $\boxed{\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}(\hat{\mu}_t)^2}$ é a decomposição da Soma dos Quadrados Total (SQT)= $\sum_{t=1}^{T}(y_t - \bar{y})^2$ na Soma dos Quadrados Explicados pela Regressão (SQE)= $\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2$ mais a Soma dos Quadrados dos Resíduos ou não explicados pela regressão (SQR) = $\sum_{t=1}^{T}(\hat{\mu}_t)^2$. ( ) V

b) $\boxed{\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}(\hat{\mu}_t)^2}$ pode ser escrito como $SQT = SQE + SQR$ ( ) V

c) O coeficiente de determinacão pode ser escrito como $R^2 = \dfrac{SQE}{SQT} = \dfrac{SQT - SQR}{SQT} = 1 - \dfrac{SQR}{SQT}$ ( ) V

d) O coeficiente de determinação mede exatamente a correlação entre a variável y_t e a variável x_t somente quando se estima uma regressão linear simples. ( ) V

e) O R^2 é um valor entre 0 e 1. Se for próximo de 1 dizemos que o modelo de regressão estimado explica a maior parte do comportamento da variável observada y_t. Se for próximo de 0 dizemos que o modelo de regressão estimado não explica o comportamento da variável observada y_t , pois neste caso o erro ou resíduo da regressão está explicando a maior parte do comportamento de y_t . ( ) V

f) Quando estimamos um modelo de regressão desejamos que o coeficiente de determinação seja próximo de zero.( ) F

g) A estatística de teste $t = \dfrac{\hat{\beta}_1 - \beta_1}{dp(\hat{\beta}_1)}$, sendo o dp($\hat{\beta}_2$) o desvio padrão de $\hat{\beta}_2$, é uma variável com distribuição t-Student, por isso é possível fazer testes de hipóteses sobre os parâmetros do modelo de regressão com base numa tabela t-Student, ou seja, podemos testar hipóteses sobre o valor de β_1 . V

h) Em modelos de regressão linear o usual é testar que os parâmetros estimados são iguais a zero, o que descreve a hipótese nula do teste de hipóteses, ou seja, $H_0 : \beta_1 = 0$. ( ) V

i) Se o modelo de regressão estimado estiver bem especificado esperamos rejeitar a hipótese nula em favor da hipótese alternativa. ( ) V

j) Quando testamos a hipótese nula, $H_0 : \beta_1 = 0$, a estatística de teste $t = \dfrac{\hat{\beta}_1 - \beta_1}{dp(\hat{\beta}_1)}$ fica $t = \dfrac{\hat{\beta}_1}{dp(\hat{\beta}_1)}$ e usamos este valor para verificar se a hipótese nula pode ou não ser rejeitada numa tabela de valores de distribuição t-Student. ( ) V

k) Se a hipótese nula for rejeitada, isto significa que prevalece a hipótese alternativa, $H_A : \beta_1 \neq 0$, e a conclusão é que, de fato, a variável x_t é relevante para explicar y_t .( ) V

l) Quando estimamos um modelo de regressão esperamos aceitar (não rejeitar) a hipótese nula. ( ) F

m) Se a hipótese nula for rejeitada dizemos que β_1 é signicativamente diferente de zero ou significante. ( ) V