FUJA DE ESTATÍSTICA · P2 (NA SEMANA DA P3!)

Para a P2, a matéria é o resto do Costa Neto que não caiu na P1, ou seja, a partir do Capítulo 6.

Acredito que a matéria dessa prova é mais "mecânica" do que a da P1, por isso, os fujás serão exercícios misturados c/ explicação.

P2 - 2015

__Questão 1)__ (2 pontos) Um experimento foi realizado visando avaliar o efeito de 2 fatores na conversão (produtividade) de um processo químico: temperatura e pressão. Por falta de tempo, não foi possível fazer o experimento com repetição. Um funcionário descuidado deixou cair café sobre a tabela, perdendo alguns de seus valores. Felizmente, alguns cálculos já haviam sido realizados. Os resultados obtidos foram:

	1 atm		2 atm		3 atm		Soma dos X_{ij} valores da linha	Soma dos $(X_{ij})^2$ valores da linha
220 °C	5,4	29				36	17,1	92,65
250 °C	3,2					38,44	13,8	68,04
270 °C	3,8					34,81	13,9	66,89
300 °C	4,6	21,16	5,2	27,04	6,1	37,21	15,9	85,41
Soma dos X_{ij} valores da coluna	17		19,5		24,2			
Soma dos $(X_{ij})^2$ valores da coluna		75		96,53		146,46		

a) Ao nível de significância de 5%, existem diferenças significativas entre os fatores pressão e temperatura? (1,0)

b) Para responder à questão anterior, que hipóteses você teve que assumir? (1,0)

Resolução

Veja que a ideia do exercício não é que você saiba fazer infinitas contas, certo? Por isso, o enunciado já deu os valores das somatórias. Esse exercício, por tratar de duas "classificações" sem repetição (7.4 - pg 156 do Costa Neto), é similar ao exemplo da pg 160 do C.N. (Costa Neto, para não ficar repetindo ü)). Vamos dar uma olhada

Exemplo (CN - pg 160)

Numa experiência agrícola, foram usados 6 diferentes fertilizantes em 2 variedades de milho, tendo sido obtidas as colheitas dadas a seguir, em sacos, p/ os vários canteiros de mesma área que foram plantados.

Fertilizantes	A	B	C	D	E	F
Variedade 1	5,4	3,2	3,8	4,6	5,0	4,4
Variedade 2	5,7	4,0	4,2	4,5	5,3	5,0

Utilizar a $\boxed{\text{Análise de Variância}}$ para verificar se

___/___/___

existem diferenças significativas entre os fertilizantes e entre as variedades ao nível de 1% de significância.

Viu a semelhança? Sim, essa é a matéria do Capítulo 7, comparação de várias médias. Como o próprio nome já diz, iremos comparar várias médias através de uma técnica chamada <u>Análise de Variância</u>.

Para realizarmos esse estudo, partimos de uma Hipótese inicial, Ho, de que as médias são iguais:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

em oposição à hipótese alternativa, que é de pelo menos uma das médias serem diferentes.

Hipóteses do MODELO

Para usar a Análise de Variância, adotamos as seguintes hipóteses implícitas básicas:

* As populações tenham a mesma variância (homocedasticidade)
  em outras palavras...
  Ho é a hipótese de que todos os valores experimentais são igualmente distribuídos

* A variável de interesse seja normalmente distribuída em todas as populações

---

Notamos que a ideia é analisar se há efeitos na aplicação de diferentes tratamentos na população!

6ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

___/___/___

Notação:

- $T_i = \sum_{j=1}^{n} x_{ij}$ = soma dos valores da i-nésima amostra;
- $Q_i = \sum_{j=1}^{n} x_{ij}^2$ = soma dos quadrados dos valores da i-nésima amostra;
- $T = \sum_{i=1}^{k} T_i = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}$ = soma total dos valores;
- $Q = \sum_{i=1}^{k} Q_i = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2$ = " " dos quadrados;
- $\bar{x}_i = \frac{T_i}{n}$ = média da amostra i
- $\bar{\bar{x}} = T/nk$: média de todos os valores
- k = nº de amostras
- n = tamanho de uma amostra
- SQT = soma de quadrados total
- s^2_T = estimativa total
- s^2_E = estimativa entre amostras
- SQE = soma de quadrados entre amostras
- s^2_R = (estimativa) residual
- SQR = soma de quadrados residual

- SQC = soma de quadrados entre colunas
- SQL = " " " " linhas
- $SQTr$ = " " " " tratamentos

$\Big\}$ Aparece só em alguns!

___/___/___

TIPOLOGIA

*Uma Classificação - Amostras de mesmo tamanho

- $s^2_T = \dfrac{Q - (T^2/nk)}{nk-1}$ • $SQT = Q - (T^2/nk)$

- $s^2_E = \dfrac{\dfrac{\sum_{i=1}^{k} T_i^2}{n} - \dfrac{T^2}{nk}}{k-1}$ • $SQE = \dfrac{\sum_{i=1}^{k} T_i^2}{n} - \dfrac{T^2}{nk}$

- $s^2_R = \dfrac{Q - \left(\sum_{i=1}^{k} T_i^2/n\right)}{k(n-1)}$ • $SQR = Q - \left(\sum_{i=1}^{k} T_i^2/n\right)$

$$\boxed{SQT = SQE + SQR}$$

- $F \text{ calculado} = \dfrac{S^2_E}{S^2_R}$ • $F \text{ crítico} = F_{k-1,\, kn-1,\, \alpha}$

ANOVA

Fonte de Variação	Soma dos quadrados	Graus de liberdade	Quadrado médio	F cal	F α
Entre amostras	SQE	$k-1$	S^2_E	F cal	F crit
Residual	SQR	$k(n-1)$	S^2_R		
Total	SQT	$nk-1$			

6ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

___/___/___

*Uma Classificação - Amostras de tamanho diferentes

- $\boxed{s^2_T = \dfrac{Q - \left(T / \sum_{i=1}^{k} n_i\right)}{\sum_{i=1}^{k} n_i - 1}}$ • $\boxed{SQT = Q - \left(T / \sum_{i=1}^{k} n_i\right)}$

CANSEI de copiar de novo (haha)

- $\boxed{s^2_E = \dfrac{\sum_{i=1}^{k} (T_i^2/n_i) - \left(T^2 / \sum_{i=1}^{k} n_i\right)}{k-1} = \dfrac{SQE}{k-1}}$

- $\boxed{s_R^2 = \dfrac{Q - \sum_{i=1}^{k} (T_i^2/n_i)}{\left(\sum_{i=1}^{k} n_i\right) - k} = \dfrac{SQR}{\left(\sum_{i=1}^{k} n_i\right) - k}}$

$\boxed{SQT = SQE + SQR}$

- $\boxed{F_{calculado} = \dfrac{S^2_E}{S^2_R}}$ • $\boxed{F_{crítico} = F_{k-1, (\sum n_i) - k, \alpha}}$

ANOVA

Fonte da variação	Soma dos Quadrados	Graus de liberdade	Quadrado médio	F_{cal}	F_{crit}
Entre as amostras	SQE	$k-1$	S^2_E	F_{cal}	F_{crit}
Residual	SQR	$(\sum n_i) - k$	S^2_R		
Total	SQT	$(\sum n_i) - 1$			

*Duas classificações - SEM REPETIÇÃO

↳ classificações cruzadas, tipo ao Ex 1 da P2 de 2015!

Agora são duas hipóteses iniciais

H_{01}: $\mu_{1.} = \mu_{2.} = \dots = \mu_{k.}$ (linhas)
H_{02}: $\mu_{.1} = \mu_{.2} = \dots = \mu_{.n}$ (colunas)
para nk observações

A aceitação de uma não influencia na aceitação da outra

Agora fica mais difícil! Devemos considerar a interação entre as classificações.

$$x_{ij} = \mu + \delta_i + \gamma_j + \psi + h_{ij}$$

x_{ij}: um elemento
μ: média teórica
δ_i: efeito da i-ésima linha
γ_j: efeito da j-ésima coluna
ψ: variação aleatória
h_{ij}: efeito da interação

Veja que a interação influencia na parte residual. Por isso, é válido saber do que estamos tratando:

• MODELO FIXO: k são populações → hipótese de inexistência de interação

• MODELO ALEATÓRIO: k são amostras aleatórias → sem hipóteses adicionais

____/____/____

Agora será preciso adotar mais algumas notações:

"NOTAÇÃO ADICIONAL"

- $T_i = \sum_{j=1}^{n} x_{ij}$ = soma dos valores da i-ésima linha
- $T_j = \sum_{i=1}^{k} x_{ij}$ = " " " j-ésima coluna
- $Q_i = \sum_{j=1}^{n} (x_{ij})^2$ = soma dos quadrados dos valores da i-ésima linha
- $Q_j = \sum_{i=1}^{k} (x_{ij})^2$ = " " " da j-ésima coluna
- $T = \sum_{i=1}^{k} T_i = \sum_{j=1}^{n} T_j$ • $\bar{x}_i = T_i / n$ = média da i-ésima linha
- $Q = \sum_{j=1}^{n} Q_j = \sum_{i=1}^{k} Q_i$ • $\bar{x}_j = T_j / k$ = média da j-ésima coluna
- $\bar{\bar{x}} = T / nk$ = média de todos os valores.

Agora sim podemos ver o esquema de fórmulas e ANOVA deste caso

- s^2_L = estimativa entre linhas
- s^2_C = " " colunas

- $S^2_T = \dfrac{Q - (T^2/nk)}{nk-1} = \dfrac{SQT}{nk-1}$

- $S^2_L = \dfrac{\sum_{i=1}^{k} T_i^2/n - T^2/n.k}{k-1} = \dfrac{SQL}{k-1}$

- $S^2_C = \dfrac{\sum_{j=1}^{n} T_j^2/k - T^2/nk}{n-1} = \dfrac{SQC}{n-1}$

- $S^2_R = \dfrac{Q - \dfrac{\sum_{i=1}^{k} T_i^2}{n} - \dfrac{\sum_{j=1}^{n} T_j^2}{k} + \dfrac{T^2}{nk}}{nk - n - k + 1}$

$nk - n - k + 1 \rightarrow (n-1)(k-1)$

$$\boxed{SQT = SQL + SQC + SQR}$$

- F calc de linha $= F_L = \dfrac{S^2_L}{S^2_R}$

- F calc de coluna $= F_C = \dfrac{S^2_C}{S^2_R}$

- F crítico para linha $= F_{crit,L} = F_{k-1,(k-1)(n-1),\alpha}$
- F crítico para coluna $= F_{crit,c} = F_{n-1,(k-1)(n-1),\alpha}$

6ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

MODELO FIXO
AQUI CONSIDERAMOS sem interação!

ANOVA:

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado médio	F	F_α
Entre linhas	SQL	$k-1$	s^2_L	F_L	$F_{crit,L}$
Entre colunas	SQC	$n-1$	s^2_c	F_c	$F_{crit,c}$
Residual	SQR	$(k-1)(n-1)$	s^2_R		
Total	SQT	$nk-1$			

*Duas classificações - COM REPETIÇÃO

Agora o total de observações é igual a rkn, para r sendo as repetições.

Para cada um dos (n.k) tratamentos, é possível ter uma estimativa de σ^2. Logo, surge uma nova parcela a ser considerada, $SQTr$ como a soma dos quadrados e s^2_{Tr} como estimativa

$$\begin{cases} SQT = SQT_r + SQR \\ nkr - 1 = (nk-1) + nk(r-1) \end{cases}$$

Para testar a igualdade de todos os nk tratamentos, usar o modelo de "Uma classificação - Amostras de mesmo tamanho". Caso Ho aceito, prova-se que S^2_{Tr} e S^2_R são independentes.

Decompondo em tratamentos por linhas e colunas:

- $SQL = \dfrac{\left(\sum_{i=1}^{k} T_i^2\right)}{nr} - \dfrac{T^2}{nkr}$

- $SQC = \dfrac{\left(\sum_{j=1}^{n} T_j^2\right)}{kr} - \dfrac{T^2}{nkr}$

- $SQTr = \dfrac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n} T_{ij}^2\right)}{r} - \dfrac{T^2}{nkr}$

Através dessa fórmula que se acha SQI!

$$\boxed{SQTr = SQL + SQC + SQI}$$

- F_I = Fcalculado da interação $= \dfrac{S^2_I}{S^2_R}$

- F_{Tr} = Fcalculado entre tratamentos $= \dfrac{S^2_{Tr}}{S^2_R}$

- $F_{crit,I}$ = Fcrítico da interação $= F_{(k-1)(n-1),\, nk(r-1),\, \alpha}$

- $F_{crit,Tr}$ = Fcrítico entre tratamentos $= F_{nk-1,\, nk(r-1),\, \alpha}$

___/___/___

Aqui testamos a interação!

ANOVA

Fonte de Variação	Soma de Quadrados	Graus de liberdade	Quadrado médio	F	F_α
Entre linhas	SQL	$k-1$	S_L^2	F_L	
Entre colunas	SQC	$n-1$	S_C^2	F_C	
Interação	SQI	$(k-1)(n-1)$	S^2_I	F_I	$F_{crit. I}$
Entre tratamentos	SQT_r	$nk-1$	$S^2_{T_r}$	F_{T_r}	$F_{crit. T_r}$
Residual	SQR	$nk(r-1)$	S^2_R		
Total	SQT	$nkr-1$			

- $\boxed{SQR = SQT = SQT_r}$
- $\boxed{SQT = Q - \dfrac{T^2}{nkr}}$
- $\boxed{s_R^2 = \dfrac{SQR}{nk(r-1)}}$

Atenção: O teste para interação deve ser realizado quando for modelo aleatório em "duas classificações - sem repetição"!

O teste anterior foi realizado para a verificação de interação e entre os tratamentos!

___/___/___

Se o resultado desse teste der que há interação, para testar se há influência de linhas e colunas, utilizamos:

$F_C = \boxed{F_{\text{calculado entre linhas}} = \frac{S^2_L}{S^2_I}}$

$F_L = \boxed{F_{\text{calculado entre colunas}} = \frac{S^2_C}{S^2_I}}$

$F_{crit,L} \boxed{F_{\text{crítico entre linhas}} = F_{K-1,(K-1)(n-1),\alpha}}$

$F_{crit,C} \boxed{F_{\text{críticos entre colunas}} = F_{n-1,(K-1)(n-1),\alpha}}$

Se NÃO haver interação, usamos no denominador o S^2_R mesmo.

Veja que, caso haja influência da interação, não faz sentido testar a influência global da classificação entre linhas e entre colunas!

Agora que já vimos os casos com a aplicação da ANOVA, podemos resolver o exercício 1 e o exemplo!

Q1 - P1 - 2015)

a) $\alpha = 5\%$, verificar se existe diferenças significativas entre pressão e temperatura?

Veja que este é um exercício de 2 classificações cruzadas sem repetição (FOLHA 4 deste RESUMO/RESOLUÇÃO DA P2 - 2015)! Veja que NÃO é preciso ficar fazendo contas eternas de somatória! Porquê? Simples! A ideia da prova foi avaliar o conceito da questão, e não se sua calculadora é boa ou se você tem velocidade e atenção em fazer infinitas contas.

Na tabela dada, os números que usaremos estão intactos. São eles:

TANTO FAZ!

$Q = 75 + 96{,}53 + 146{,}6 = 317{,}99$

$T = 17 + 19{,}5 + 24{,}2 = 60{,}7$

$Q = 97{,}65 + 68{,}04 + 66{,}89 + 85{,}41 = 317{,}99$

$T = 17{,}1 + 13{,}8 + 13{,}9 + 15{,}9 = 60{,}7$

Veja que, para $(T_j)^2$ e $(T_i)^2$, temos que elevar ao quadrado os valores de T_i e T_j da tabela dada. Logo.

PARA $(T_j)^2$:

	1atm	2atm	3atm	Σ
T_j	17	19,5	24,2	60,7
$(T_j)^2$	289	380,25	585,64	1254,89

PARA $(T_i)^2$:

	220°C	250°C	270°C	300°C	Σ
T_i	17,1	13,8	13,9	15,9	60,7
$(T_i)^2$	292,41	190,44	193,21	252,81	928,87

Calculando as estimativas:

- $$s^2_T = \frac{Q - (T^2/n.k)}{nk-1} = \frac{SQT}{nk-1} = \frac{317,99 - (60,7^2/3.4)}{3.4-1}$$

$$s^2_T = 10,949167$$ ← SQT

(11) → grau de liberdade

- $$s^2_L = \frac{\sum_{i=1}^{k} Ti^2/n - T^2/n.k}{k-1} = \frac{SQL}{k-1}$$

$$= \frac{928,87/3 - (60,7)^2/3.4}{4-1} = 2,5825$$ ← SQL

(3) → grau de liberdade

$$s^2_L = 0,8608$$

- $$s^2_c = \frac{\sum_{j=1}^{4} T_j^2/k - T^2/n.k}{n-1} = \frac{SQC}{n-1}$$

$$= \frac{1254,89/4 - (60,7)^2/3.4}{3-1} = 6,6817$$ ← SQC

(2) → grau de liberdade

$$s_c^2 = 3,3408$$

$$\bullet\ s_R^2 = \frac{Q - \frac{\sum_{i=1}^{k} T_i^2}{n} - \frac{\sum_{j=1}^{n} T_j^2}{k} + \frac{T^2}{nk}}{(n-1)(k-1)}$$

$$= \frac{SQR}{(n-1)(k-1)}$$

Dá para resolver o SQR desse jeito ou do jeito "traquinas". Temos SQT, SQC e SQL. Sabemos que:

$$SQT = SQL + SQC + SQR$$
$$10,949167 = 2,5825 + 6,6817 + SQR$$
$$SQR \cong 1,685$$

Logo:

$$s_R^2 = \frac{1,685}{(3-1)(3-1)} = \frac{1,685}{2 \cdot 3} = 0,2808$$

→ 6 → grau de liberdade

Construindo nossa ANOVA:

Fonte de variação	Soma de quadrados	Graus de liberdade	Quadrado médio	F	Fα
Entre linhas	SQL = 2,5825	3	0,8608 (s_L^2)	3,07	4,76
Entre colunas	SQC = 1,68	2	3,3408 (s_C^2)	11,90	5,14
Residual	SQR = 1,685	6	0,2808 (s_R^2)		
Total	SQT = 10,949167	11			

veja como encontramos na próx. pg!

CALCULADOS

F entre Linhas $= \frac{S^2_L}{S^2_R} = \frac{0,8608}{0,2808} \cong 3,07$

F entre colunas $= \frac{S^2_C}{S^2_R} = \frac{3,3408}{0,2808} \cong 11,90$

CRÍTICOS

F crítico entre Linhas $= F_{3,\ 6,\ 5\%} = 4,76$

nível de significância

grau de liberdade de SQL

grau de liberdade do SQR

F crítico entre colunas $= F_{2,\ 6,\ 5\%} = 14,54$

nível de significância

grau de liberdade de SQC

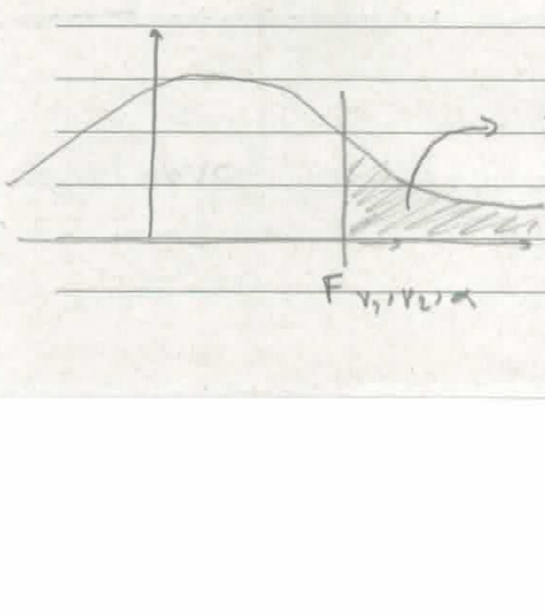
grau de liberdade do SQR

COMO VER UMA TABELA DE F de Snedecor

Para um $F_{\nu_1, \nu_2, \alpha}$

1º passo ⇒ Achar a tabela referente ao seu nível de significância (α).
$P = \alpha$

2º passo ⇒ ν_1 é a coluna e ν_2 é a linha. Você achou um valor. Esse é seu valor!



$P(F_{\nu_1, \nu_2} \geq F_{\nu_1, \nu_2, \alpha}) = \alpha$

Logo:

F crítico < F calc → Rejeita Ho → fator apresenta diferença para o α especificado

(Temp) 4,76 > 3,07 → Aceita Ho
(pres.) 5,14 < 11,90 → Rejeita Ho

À um nível de significância de 5%, é possível averiguar que o fator pressão apresenta diferença.

b) • Populações com mesma variância (homocedasticidade)
• Variável de interesse seja normalmente distribuída em todas as populações.

Essas duas garantem que Ho seja a hipótese de que os valores experimentais são igualmente distribuídos!

Questão 2) (3 pontos) A 1ª prova de PRO 3200 ocorreu no dia 19/10/15, às 7:30 da manhã (Nota do Autor do resumo/resolução de P2-2015: de fato ocorreu e, infelizmente, não foi pouco...). "Acesso" aqui quer dizer um clique de algum aluno online em um dos recursos do AVA.

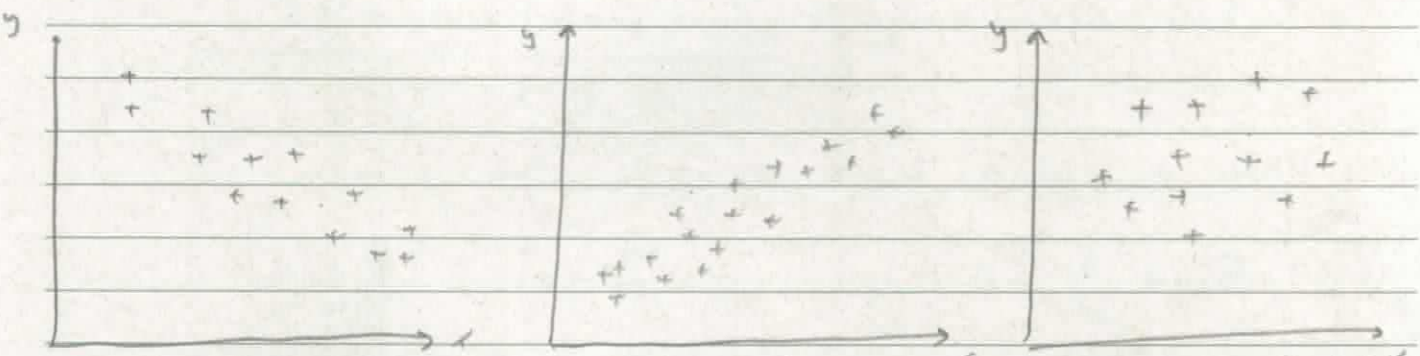
Nesse hora, percebi que era mais fácil simplesmente anexar uma cópia da prova do que copiá-la... Enfim, melhor não comentarmos... Os enunciados estão em anexo.

Resolução

Sim, agora o assunto é Regressão e Correlação (CN-Capítulo 8).

Este é um dos assuntos mais fáceis de visualizar em exemplos do nosso cotidiano, como os gráficos de dispersão de pontos e análises gráficas de Laboratório Experimental de Física.
A ideia é, com um conjunto de dados, analisar um comportamento médio.

• Correlação Linear



Correlação linear negativa — Correlação linear positiva — Correlação linear nula.

Veja os gráficos acima. A ideia é que, para um conjunto de pares de dados, é possível averiguar uma correlação, ou seja, a tendência de comportamento das variáveis consideradas. Positiva se a reta a ser considerada na correlação é ascendente, ou seja, x cresce com o y, ou negativa, se a reta é descendente, ou seja, x decresce com o y.

Você já usou a função gráfico do Excel? É possível traçar a linha de tendência (sim, a correlação linear seria a linha linear) com um coeficiente r. Sim, o r é o coeficiente de correlação linear de Pearson.

6ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

___/___/___

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{cov(x,y)}{\underbrace{S_x \cdot S_y}}$$

→ covariância entre x e y

→ produto dos desvios padrões de x e y (amostral)

$r =$

- $S_x = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$ • $S_y = \sqrt{\dfrac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$

- $S_{xy} = cov(x,y) = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Vale a pena lembrar que:

IMPORTANTE!

- $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum x_i^2 - \dfrac{(\sum x_i)^2}{n}$
- $S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum y_i^2 - \dfrac{(\sum y_i)^2}{n}$
- $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}$

Então, r pode ser

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{S_{xy}/n-1}{\sqrt{\frac{S_{xx}}{n-1}} \cdot \sqrt{\frac{S_{yy}}{n-1}}} = \frac{(1/n-1)}{(1/n-1)} \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

Acho legal saber usar as diferentes notações, porque o exercício pode te dar um dado que só usado em um dos métodos ou por questão de tempo mesmo.

___/___/___

• TESTES DO r

Na pegada da P1, o r é o coef. de correlação amostral. Mas, e se quisermos testar a sua validade para uma população? Por isso, há os testes para ρ, o coeficiente de correlação populacional.

Para verificar se há correlação linear:

$$\left.\begin{array}{l} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{array}\right\} t_{cal} = r\sqrt{\frac{n-2}{1-r^2}}$$

$t_{crit} = t_{n-2}$ (pode ser feito unilateralmente)

Para a hipótese de ρ igual a outro valor, há uma fórmula de Fisher, mas acho pouco provável de cair.

$$\mathcal{L} = \frac{1}{2} \ln \frac{1+r}{1-r}, \text{ sendo } \mu(\mathcal{L}) = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$$

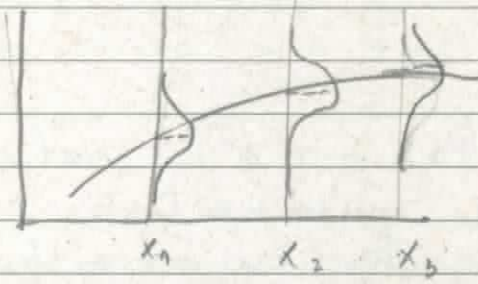
$$\sigma(\mathcal{L}) = \sqrt{\frac{1}{n-3}}$$

Vantagem dessa abordagem é que a distribuição é próxima da normal.
Essa abordagem é usada p/ intervalo de confiança.

$$P(\mathcal{L} - z_{\alpha/2} \cdot \sigma(\mathcal{L}) < \mu(\mathcal{L}) < \mathcal{L} + z_{\alpha/2} \cdot \sigma(\mathcal{L})) = 1-\alpha$$

• Regressão

Determinar uma função que exprima o relacionamento entre as variáveis (graficamente, a linha de regressão).

Analisar que  há variações residuais em torno da regressão teórica

Aqui, iremos tratar da regressão linear

$$\underbrace{y = \alpha + \beta x}_{\text{EQUAÇÃO TEÓRICA DA REGRESSÃO LINEAR}} + \psi$$ componente aleatório de variação de y

$\Rightarrow$ $\left.\begin{matrix}\alpha \\ \beta\end{matrix}\right\}$ coeficientes de regressão linear teóricos

Na prática, testamos $\underset{z}{a}$ e $\underset{z}{b}$ como coeficientes que são parâmetros de α e β, respect. vamente ($\hat{y} = a + bx$), para a estimativa

- Métodos PARA DETERMINAR a e b.

1º - "O MAIS COXA" - TRAÇAR UMA RETA
↳ Sim, simplesmente traçar uma reta no gráfico de dispersão a fim de reduzir as distâncias desta em relação aos pontos. (NÃO CAI - óbvio ...)

2º - Mínimos quadrados - reduzir ao máximo os valores das variâncias.

$$\min \sum d_i^2 = \min \sum (y_i - \hat{y}_i)^2 = \min \sum (y_i - a - bx_i)^2$$

derivando em a e b e igualando a 0:

$$-2 \sum (y_i - a - bx_i) = 0$$
$$-2 \sum x_i (y_i - a - bx_i) = 0$$

Desenvolvendo o sistema, chegamos em:

$$\begin{cases} b = \dfrac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \dfrac{S_{xy}}{S_{xx}} \\ a = \bar{y} - b\bar{x} \end{cases}$$

IMPORTANTE

Veja que saber calcular S_{xy}, S_{xx} e S_{yy} é essencial para esta parte.

Para reta passando pela origem, basta estimar b como $\dfrac{\sum x_i y_i}{\sum x_i^2}$.

• INDUÇÃO QUANTO AOS PARÂMETROS DA RETA

$H_0 : \beta = 0$ (não há regressão)
$H_1 : \beta \neq 0$ } hipótese iniciais

▪ CONSIDERAÇÕES

• $S_R^2 = \dfrac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}$ ↳ (α e β)

• $\hat{y}_i = \bar{y} - b\bar{x} + bx_i$

• $\sum (\hat{y}_i - \bar{y})^2 = b^2 \sum (x_i - \bar{x})^2$

• $\sum (\hat{y}_i - \bar{y})^2 = b^2 . S_{xx} = b S_{xy} = \dfrac{S_{xy}^2}{S_{xx}}$

- $\sum (y_i - \hat{y}_i)^2 = S_{yy} - b S_{xy}$

importante!

$$\boxed{s_R^2 = \frac{S_{yy} - bS_{xy}}{n-2}} = \boxed{\frac{S_{yy} - b^2 S_{xx}}{n-2}}$$

$$= \boxed{\frac{S_{yy} - S_{xy}^2 / S_{xx}}{n-2}}$$

$$= \boxed{\frac{S_{yy}(1-r^2)}{n-2}}$$

- $S_{yy} = \sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$

- $r^2 = \frac{b \cdot S_{xy}}{S_{yy}} = \frac{S_{xy}}{S_{xx}} \frac{S_{xy}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}} =$ coeficiente de determinação

- $(1-r^2)$ = coeficiente de indeterminação

↳ em geral, para $|r| \geq 0,9$, a regressão explica a maioria da variação.

Estimação e teste de parâmetros da reta de regressão

- Estimador b

$\mu(b) = \beta$ $\qquad \sigma^2(b) = \frac{\sigma_R^2}{S_{xx}}$ → Não conhecemos

↳ estima-se ⇒ $s^2(b) = \frac{S_R^2}{S_{xx}}$

___/___/___

- Testando $H_0: \beta = \beta_0$ — distribuição normal

$$\boxed{t_{calc} = \frac{b - \beta_0}{S_R/\sqrt{S_{xx}}} \qquad t_{crit} = t_{n-2}}$$

↳ p/ $\beta = 0$ na hipótese

$$t_{calc} = r\sqrt{\frac{n-2}{1-r^2}}$$

- Analogamente, para a:

$$\mu(a) = \alpha \qquad \underbrace{\sigma^2(a)}_{\text{estima-se}} = \frac{\overbrace{\sigma_R^2}^{\text{Não sabemos}} \sum x_i^2}{n\,S_{xx}}$$

↳ estima-se: $S^2(a) = \frac{S_R^2}{S_{xx}} \cdot \frac{\sum x_i^2}{n}$

- Testando $H_0 = \alpha = \alpha_0$

$$t_{calc} = \frac{a - \alpha_0}{\sqrt{S^2(a)}}$$

distribuição normal

Intervalos de confianças p/ $\alpha + \beta x'$ e y'

- IC p/ $\alpha + \beta x'$ c/ base na estimativa $\hat{y}'$

$$\left.\begin{aligned} \mu(\hat{y}') &= \alpha + \beta x' \\ \sigma^2(\hat{y}') &= \underbrace{\sigma_R^2}_{\text{estimar p/ } S_R^2}\left[\frac{1}{n} + \frac{(x' - \bar{x})^2}{S_{xx}}\right] \end{aligned}\right\}$$

x' = um valor de x
$\hat{y}'$ = valor de y para x'

↳ estimar p/ S_R^2

___/___/___

$$\hat{y}' \pm t_{n-2,\alpha/2} \cdot S_R \sqrt{\frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}}$$

$\hat{y}' = a + bx'$

Se variarmos x', podemos determinar a "região de confiança" para $\alpha + \beta x'$

• Para o intervalo de previsão

$$\hat{y}' \pm t_{n-2,\alpha/2} \; S_R \sqrt{1 + \frac{1}{n} + \frac{(x'-\bar{x})^2}{S_{xx}}}$$

Analogamente, variando x', teremos a "região de previsão para y'"!

MAS QUAL É A DIFERENÇA??

- Intervalo de confiança de $\alpha + \beta x'$ ⇒
  ↳ você parte de um x' (determinado x) para estimar estimar $\hat{y}'$. (valor da reta de regressão estimada p/ um x').

- Intervalo de previsão de y'.
  ↳ vc determina o intervalo no qual o Y experimental se encontra p/ uma determinada significância vc usa

Q2 - P2 - 2015)

___/___/___

a) Determinar a reta de regressão

(tradução...) ⇒ determine os valores de a e b.

$b = \frac{S_{xy}}{S_{xx}} \Rightarrow \cdot\ S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

$\cdot\ S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$ $= 39934 - \frac{147 \cdot 4921}{14}$

$= 1771 - \frac{(147)^2}{14}$ $= -11.736,5$

$= 227,5$

$b = \frac{S_{xy}}{S_{xx}} = -\frac{11.736,5}{227,5} = -51,59$

$a = \bar{y} - b\bar{x} = 351,50 - (-51,59) \cdot (10,50)$

$a = 893,195$

$\therefore \hat{y} = 893,195 - 51,59x$

b) $\alpha = 5\%$, é possível averiguar correlação negativa?

___/___/___

Hipóteses iniciais

$H_0: \beta = 0$ $= \beta_0$
$H_1: \beta < 1$

$t_{calc} = \dfrac{b - \beta_0}{S_R / \sqrt{S_{xx}}}$

$* S_R = \dfrac{S_{yy} - b\,S_{xy}}{n-2}$

$\bullet\ S_{yy} = \sum y_i^2 - \dfrac{(\sum y_i)^2}{n} =$

$2360653 - \dfrac{(4921)^2}{14} = \boxed{630921,5}$

$\bullet\ S_{xy} =$ item anterior $= -11736,5$

$* S_R^2 = \dfrac{630921,5 - (-51,59).(-11.736,5)}{14-2}$

$\boxed{S_R^2 = 2119,62}$

$t_{calc} = \dfrac{-51,59}{\sqrt{\dfrac{2119,62}{227,5}}} = -16,902$

$t_{critico} = t_{n-2,\alpha} = t_{12,\,5\%} = -1,782$

$t_{calc} < t_{crit} \Rightarrow$ Rejeita H_0.

À nível de 5% de significância, é possível afirmar que há correlação negativa

c) Intervalo de confiança p/ α e β da reta teórica.

• β

$$b - t_{2,5\%,12} \cdot \sqrt{\frac{S_R^2}{S_{xx}}} \leq \beta \leq b + t_{2,5\%,12} \sqrt{\frac{S_R^2}{S_{xx}}}$$

• $b = -51,59$
• $t_{2,5\%,12} = 2,179$
• $S_R^2 = 2119,62$
• $S_{xx} = 227,50$

IC de β:

$$\boxed{-58,24 \leq \beta \leq -44,94}$$

ver folha 13

essas fórmulas são as mesmas da estimação, só que não testamos hipótese

• α

$$a - t_{2,5\%,12} \cdot \sqrt{S^2(a)} \leq \alpha \leq a + t_{2,5\%,12} \sqrt{S^2(a)}$$

$$S^2(a) = \frac{S_R^2}{S_{xx}} \cdot \frac{\sum x_i^2}{n} = \frac{2119,62}{227,50} \cdot \frac{1771}{14} = 1178,60$$

IC de α:

$$\boxed{818,39 \leq \alpha \leq 968,01}$$

15

___/___/___

Q3 · P2 · 2015) → continuação de matéria de Regressão e correlação linear.

• REGRESSÃO POLINOMIAL

• $y = \alpha + \beta x + \gamma x^2$ (função de regressão)

$\hat{y} = a + bx + cx^2$ (estimativa)

$$\begin{cases} \sum y_i = na + c\sum(x_i-\bar{x})^2 \\ \sum(x_i-\bar{x})y_i = b\sum(x_i-\bar{x})^2 \\ \sum(x_i-\bar{x})^2 y_i = a\sum(x_i-\bar{x})^2 + c\sum(x_i-\bar{x})^4 \end{cases}$$

Esses 4 tópicos são difíceis de cair para calcular por fórmula

Esse sistema provém das derivadas parciais do sistema original que considera as K+1 estimativas

• REGRESSÃO LINEAR MÚLTIPLA

• $\hat{y} = a + b_1x_1 + b_2x_2$

→ MMQ.

x^2 • x_{1j}

$$\begin{cases} \sum y_j = na + b_1\sum x_{1j} + b_2\sum x_{2j} \\ \sum x_{1j}y_j = a\sum x_{1j} + b_1\sum x_{1j}^2 + b_2\sum x_{1j}x_{2j} \\ \sum x_{2j}y_j = a\sum x_{2j} + b_1\sum x_{1j}x_{2j} + b_2\sum x_{2j}^2 \end{cases}$$

$S_{1y} = b_1S_{11} + b_2S_{12} + \dots + b_kS_{1k}$;

$S_{2y} = b_1S_{21} + b_2S_{22} + \dots + b_kS_{2k}$

$\vdots$

$S_{ky} = b_1S_{k1} + b_2S_{k2} + \dots + b_kS_{kk}$

$S_{iy} = \sum_{l=1}^{k} b_l S_{il} \quad (i = 1, 2, \dots, k)$

• Correlação linear múltipla

$$R = \sqrt{\frac{b_1 S_{1y} + b_2 S_{2y} + \dots + b_k S_{ky}}{S_{yy}}}$$ (coef. de correlação linear)

$$S_M^2 = \frac{S_{yy} - \sum_{i=1}^{k} b_i S_{iy}}{n-k-1}$$ (VARIÂNCIA RESIDUAL)

• Correlação parcial e variáveis fictícias.
↳ não contemplei esses assuntos pq me falaram que não foi passado...

ANÁLISE DE VARIÂNCIA APLICADA À REGRESSÃO

• teste de regressão linear

• $\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$

• $s_Y^2 = \dfrac{\sum(y_i - \bar{y})^2}{n-1} = \dfrac{S_{yy}}{n-1}$ (variância amostral de Y)

• $s_R^2 = \dfrac{\sum(y_i - \hat{y}_i)^2}{n-2} = \dfrac{S_{yy} - b^2 S_{xx}}{n-2}$

↳ considerando mais alguns desenvolvimentos...

$$F_{calc} = \frac{b^2 S_{xx}}{S_R^2}$$

TESTE SERÁ SEMPRE UNILATERAL

8ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

$F_{crit} = F_{v_1, v_2, \alpha}$.

ANOVA

Fonte de variação	Soma de quadrados	Graus de liberdade	Quadrado médio	F	F_α
Devido à Regressão	$b^2 S_{xx}$	1	$b^2 S_{xx}$	$F = \frac{b^2 S_{xx}}{S_R^2}$	$F_{1, n-2, \alpha}$
Residual	$S_{yy} - b^2 S_{xx}$	$n-2$	$\frac{S_{yy} - b^2 S_{xx}}{n-2}$ (S_R^2)		
Total	S_{yy}	$n-1$			

Se $F_{calc} > F_{crit}$, rejeita-se H_0.

ANÁLISE DE MELHORIA

A ideia central é analisar se ocorreu uma melhoria significativa perdendo-se um grau de liberdade e assim tornando um grau acima.

- Reta X Parábola

ANOVA

Fonte de variação	Soma dos quadrados	Grau de liberdade	Quadrado médio	F	F_α
Melhoria do ajuste	① $= \sum (\hat{y}_{Pi} - \hat{y}_i)^2$	1	①	$F = \frac{①}{S_p^2}$	$F_{1, n-3, \alpha}$
Residual s/ a parábola	② $= \sum (y_i - \hat{y}_{Pi})^2$	$n-3$	$S_p^2 = \frac{②}{n-3}$		
Residual s/ a reta	$S_{yy} - b^2 S_{xx}$	$n-2$			

Análise de VARIÂNCIA NA Regressão linear múltipla

ANOVA

Fonte de variação	Soma de quadrados	Graus de liberdade	Quadrado médio	F	F_α
Regressão	$① = \sum_{i=1}^{k} b_i S_{iy}$	k	$\frac{①}{k} = ②$	$F = \frac{①}{S_M^2}$	$F_{k, n-k-1, \alpha}$
Residual	$SQM = S_{yy} - ①$	$n-k-1$	$S_M^2 = \frac{SQM}{n-k-1}$		
Total	S_{yy}	$n-1$			

Resolução da Q3 - P2 - 2015

a) ANOVA

Fonte	Soma Quad.	Graus Lib.	Quad. Méd.	Fcalc.	F tab
Regressão	① 421200	2 ③	$\frac{①}{③} = ⑥$	$\frac{⑥}{⑦}$	⑧
Residual plano	② − ① = ④	7 − ③ = ⑤	$\frac{④}{⑤} = ⑦$		
Total	② 450521	7			

↳ não muda!

___/___/___

$④ = ② - ① = 450521 - 421200 = 29321$

$7 - ③ = ⑤ = 7 - 2 = 5$

$\frac{①}{③} = ⑥ = \frac{421200}{2} = 210600$

$\frac{④}{⑤} = ⑦ = \frac{29321}{5} = 5864,2$

$F_{calc} = \frac{⑥}{⑦} = \frac{210600}{5864,2} = 35,91$

$F_{crit.} = F_{③, ⑤, \alpha} = F_{2, 5, 5\%} = 5,79$

b) ANÁLISE DE MELHORIA

Fonte	Soma Quad.	Graus Lib.	Quad Méd	Fcalc	Ftab
Linear	353969 (dado no enunciado)	1	353969	④ 60,36	⑥ 6,61
Melhoria MCM	① 67231	1	② 67231	⑤ 11,46	④
Residual Puro	29321 (item anterior)	7-1-1=5	③ 5864,2		
Total	450521	7			

① → $SQT = SQ_{Linear} + \underbrace{SQ_{melhoria} + SQR_{plano}}_{\text{SQ Residual da reta}}$

① →, $450521 = 353969 + SQ_{melhoria} + 29321$

$SQ_{melhoria} = 67.231$

② $= \dfrac{67231}{1} = 67231$ ③ $= \dfrac{29321}{5} = 5864,2$

④ $= \dfrac{353964}{5864,2} = 60,36$ ⑤ $= \dfrac{67231}{5864,2} = 11,46$

PARA OS F's calculados, você divide pelo Residual (O único caso que isto não ocorre é para Análise de Variância com Dupla classificação c/ Repetição!)

⑥ $= F_{crít.}$ ($F_{tabelado}$) $= F_{v_1; v_2, \alpha} = F_{1, 5, 5\%} = 6,61$

Como o F calculado da melhoria (11,46) é maior que o Fcrítico, você rejeita H_0 e conclui que a melhoria é significativa

c) Coeficiente de Determinação $= R^2 = \dfrac{SQ_{reg}}{SQT} = \dfrac{421200}{450521} = 0,9349$

Sim, o mesmo do Excel!

a ideia dessa razão é expressar o quanto a regressão "explica" seus dados (SQ da regressão) sobre o total (SQT).

Q4) PAPEL DE PROBABILIDADE

Este é um dos chamados testes de aderência.
O papel de probabilidade é um procedimento relativamente simples e aproximado, ou seja, não possui um rigor muito grande.

Ele testa o quanto a sua amostra se aproxima de uma distribuição normal.

PASSOS

1) Dada uma amostra, ver qual o maior e o menor valor para achar a amplitude. A idea é que no eixo das abscissas, você utilize uma escala adequada a fim de caber completamente todos os valores e que seja fácil de traçar os pontos. No caso do exercício da prova, o menor é 13 e o maior 52.
Uma escala adequada no caso da questão é de 10 até 60.

2) Organize seus dados de forma crescente.
Ex. da Prova:

13 15 21 22 26 28 32 38 40 52

3) Dado que você já tem os valores de x, falta as probabilidades. A idea é que você distribua os pontos de sua amostra simetricamente, através da fórmula:

6ª OLIMPÍADA BRASILEIRA DE MATEMÁTICA DAS ESCOLAS PÚBLICAS
OBMEP 2010
Programa de Iniciação Científica

___/___/___

$\boxed{50(2i-1)/n}$ Sendo n o nº de elementos da amostra e i o índice do elemento já com a amostra ordenada de forma crescente.

Da prova: $n = 10$

x	13	15	21	22	26	28	32	38	40	52
i	1	2	3	4	5	6	7	8	9	10
y:	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%

esse valor nada mais é do que a frequência relativa acumulada!

4) Agora basta plotar seus pontos e traçar uma reta que aproxima ao máximo os pontos. Essa etapa é bem abstrata, indo do bom senso da pessoa. Veja na folha do exercício anexada no final.

5) Agora, o grande final! Veja que do lado direito, há μ (média) e -1σ (menos desvio padrão) e $+1\sigma$ (mais desvio padrão). Siga a reta horizontal de cada um dos três e, quando cruzar com a reta que você traçou, rebata para o eixo das abscissas. Assim você achará 3 valores

$\mu-\sigma$	μ	$\mu+\sigma$
≈ 18 e 17	≈ 29	$\approx 40{,}41$

$\underbrace{\quad}_{\approx 11 \text{ e } 12}$ $\underbrace{\quad}_{\approx 11 \text{ e } 12}$

$\Rightarrow \mu \approx 29$
$\sigma \approx 11$ ou 12

aqui não importa tanto a precisão!

Veja o que foi feito no anexo da prova, em cor azul!

19

6ª Olimpíada Brasileira de Matemática das Escolas Públicas
OBMEP 2010
Programa de Iniciação Científica

____/____/____

Bom, se você chegou até aqui, agradeço a paciência!
Mantenha a concentração e organize bem seu formulário.
Acredito em você! :)

Boa Prova!

Cassio Medeiros

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

PRO 3200 – ESTATÍSTICA – P2 (23/11/2015)

Aluno: ______________________________

NUSP: ______________________________

Assinatura: ______________________________

Professor: ______________________________

Turma: ______________________________

QUESTAO	NOTA
1	
2	
3	
4	
TOTAL	

Questão 01 (2 pontos)

Um experimento foi realizado visando avaliar o efeito de dois fatores na conversão (produtividade) de um processo químico: temperatura e pressão. Por falta de tempo, não foi possível fazer o experimento com repetição. Um funcionário descuidado deixou cair café sobre a tabela, perdendo alguns de seus valores. Felizmente, alguns cálculos já haviam sido realizados. Os resultados obtidos foram:

	1 atm		2 atm		3 atm		Soma dos Xij valores da linha	Soma dos $(Xij)^2$ valores da linha
220 °C	5,4	29[illegible]	[illegible]	[illegible]	[illegible]	36	17,1	97,65
250 °C	3,2	[illegible]	[illegible]	[illegible]	[illegible]	38,44	13,8	68,04
270 °C	3,8	14[illegible]	[illegible]	[illegible]	[illegible],9	34,81	13,9	66,89
300 °C	4,6	21,16	5,2	27,04	6,1	37,21	15,9	85,41
Soma dos Xij valores da coluna	17		19,5		24,2			
Soma dos $(Xij)^2$ valores da coluna		75		96,53		146,6		

a) Ao nível de significância de 5%, existem diferenças significativas entre os fatores pressão e temperatura? (1 Ponto)

b) Para responder à questão anterior, que hipóteses você teve que assumir? (1 Ponto)

Questão 02 (3 pontos)

A primeira prova de PRO3200 ocorreu no dia 19 de outubro de 2015 as 7:30 da manhã. A tabela com os acessos nos intervalos medidos está abaixo. "Acesso" aqui quer dizer um clique de algum aluno online em um dos recursos do AVA.

Período	Tempo Faltante para Prova (Horas)	Número de Acessos
18/10 - 14:00 a 14:59	17	37
18/10 - 15:00 a 15:59	16	96
18/10 - 16:00 a 16:59	15	[illegible]
18/10 - 17:00 a 17:59	[illegible]	[illegible]
18/10 - 18:00 a 18:59	[illegible]	[illegible]
18/10 - 19:00 a 19:59	[illegible]	[illegible]
18/10 - 20:00 a 20:59	[illegible]	[illegible]
18/10 - 21:00 a 21:59	[illegible]	[illegible]
18/10 - 22:00 a 22:59	9	490
18/10 - 23:00 a 23:59	8	557
19/10 - 00:00 a 00:59	7	542
19/10 - 01:00 a 01:59	6	558
19/10 - 02:00 a 02:59	5	621
19/10 - 03:00 a 03:59	4	703

Novamente o funcionário descuidado deixou cair café sobre a tabela, perdendo alguns de seus valores, mas felizmente, alguns cálculos já haviam sido realizados.

$$\sum x_i = 147 \qquad \sum y_i = 4921$$
$$\sum x_i^{\,2} = 1771 \qquad \sum y_i^{\,2} = 2360653$$
$$\bar{x} \cong 10{,}50 \qquad \bar{y} \cong 351{,}50$$
$$n = 14 \qquad \sum x_i \cdot y_i = 39934$$

a) Determine a reta de regressão. (1 Ponto)
b) Ao nível de significância de 5%, é possível verificar correlação negativa entre proximidade da prova e número de acessos? (1 Ponto)
c) Determine o Intervalo com 95% de confiança para os parâmetros Alfa e Beta da reta teórica. (1 Ponto)

Questão 03 (3 pontos)

O consumo de energia por um ser humano para se manter vivo é denominado Taxa Metabólica Basal ou TBM. Reflete o consumo mínimo diário de kcal para o corpo se manter em repouso. Um pesquisador estudou o TBM e quer correlacionar o TBM com a Massa Corporal Magra e a Massa de Gordura, medidas por bioimpedância. Obteve os seguintes dados:

Y	1745,3	1684,3	2095,1	1645,2	1803,3	1774,5	2395,3	1985,4
MG	10	5	21	16	18	7,5	44	16
MCM	60	55	72	45	62	63	56	67

Y-TBM kcal; MG-massa gorda kg; MCM-massa magra kg.

Avaliando a relação entre Y e MG encontrou a reta média $\bar{Y} = 1573 + 18,5\ MG$ com a seguinte tabela de ANOVA:

Fonte	Soma Quad.	Graus Lib.	Quad. Médio	F calc	F tab
Regressão	353969	1	353969	22,00	5,99
Residual reta	96552	6	16092		
Total	450521	7			

Para melhorar o modelo resolveu incluir a variável MCM e encontrou o modelo (plano) para média $\bar{Y} = 853 + 11,9\ MCM + 18,7\ MG$ com a tabela ANOVA.

Fonte	Soma Quad.	Graus Lib.	Quad. Médio	F calc	F tab
Regressão	421200	2			
Residual plano					
Total	450521				

a) Complete a tabela de ANOVA preenchendo os valores faltantes nos espaços vazios. (1 ponto)
b) Faça a análise de melhoria e verifique se a introdução da variável MCM no modelo traz melhoria significativa. (1 ponto)
c) Qual o valor do coeficiente de determinação para o modelo de Regressão Linear Múltipla (plano)? (1 ponto)

Adote α=5%.

Questão 04 (2 pontos)

Dada a seguinte amostra:

13 – 28 – 15 – 32 – 21 – 38 – 22 – 40 – 26 – 52

a) Mediante o uso do papel de probabilidade normal (fornecido na próxima página), verificar se uma distribuição normal se ajusta bem a estes dados. (1 ponto)
b) Obter **através deste mesmo papel** uma estimativa para a média e o desvio-padrão da população. (1 ponto)

. Ao traçar a reta, tenha bom senso!

Papel de Probabilidade Normal