

SOLUÇÃO

PRO3200 – Estatística

Turma:

Prof:

Prova 3 - 2018

Nome (completo legível):

Teste 1: Em um experimento genético, os pesquisadores observaram 300 cromossomos de um tipo particular e contaram o número de cromátides irmãs trocadas em cada um. Um modelo de Poisson foi suposto para a distribuição do número de trocas.

Num. Trocas	0	1	2	3	4	5	6	7	8	9
Contagens observadas	6	24	42	59	62	44	41	14	6	2

- A De fato, pode-se afirmar a 5% de significância que o número de trocas segue uma distribuição de Poisson, pois a maior diferença entre a frequência observada e esperada é de 11,70.
- B Não se pode afirmar a 5% de significância que o número de trocas segue uma distribuição de Poisson, uma vez que o qui-quadrado observado é de 5,740 e o crítico é de 2,167.
- C O total de classes levadas em consideração no teste de aderência é de 10 (0 a 9).
- D Não é possível fazer este teste de aderência, uma vez que o parâmetro (a média) da distribuição proposta (Poisson) não foi fornecida.
- E De fato, pode-se afirmar a 5% de significância que o número de trocas segue uma distribuição de Poisson, uma vez que o qui-quadrado observado é de 5,740 e o crítico é de 14,067.

Teste 2: Os dados (à esquerda) se referem aos tempos de reação provenientes de um experimento realizado para investigar se o produto de um determinado processo químico depende da formulação de um insumo específico e/ou da velocidade do misturador. (Considere 5% de significância).

Formulação	Velocidade			Tabela da ANOVA						
	60	70	80	Fonte da variação	SQ	gl	Quadrado Médio	$F_{observado}$	p-value	$F_{tabelado}$
1	189,7	185,1	189,0	Formulação	2.253,44				<0,0001	
	188,6	179,4	193,0	Velocidade	230,81				0,0002	
	190,1	177,3	191,1	Interações	18,58				0,2516	
2	165,1	161,7	163,3	Residual/Erro						
	165,9	159,8	166,6	Total	2.574,70					
	167,6	161,6	170,3							

- A Considerando níveis de significância entre 1% a 20%, pode-se concluir que efeitos dos fatores Velocidade e Formulação estão ativos.
- B A soma de quadrados residual apresentada na tabela representa o efeito do fator “velocidade” e “formulação”, caso não houvessem replicações no experimento .
- C Pelos valores observados dos p-values (valores p ou níveis descritivos), pode-se afirmar que o fator “Formulação”, por possuir um p-value menor, apresenta um menor efeito, na sequência, em ordem crescente de efeito “velocidade” e por último, o de maior efeito é a “Interação”.
- D Pode-se afirmar que existe efeito de interação entre os fatores “Formulação” e “Velocidade”, uma vez que se trata de um experimento de dois fatores com 3 replicações em cada combinação dos níveis dos fatores em estudo).
- E Por ser um experimento com replicação, pode-se afirmar que existem formulações específicas que são privilegiadas por velocidades específicas, por exemplo: velocidade 80 funciona melhor para a formulação 1 e velocidade 60 seria melhor para formulação 2 .

SOLUÇÃO

Teste 3: Em um estudo de Ergonomia (que busca compreender a interação do homem com seu posto de trabalho), coletou-se o esforço necessário para nove funcionários levantarem-se de quatro diferentes tipos de assento, usando uma escala adequada. Utilize significância de 5%:

Assento	F1	F2	F3	F4	F5	F6	F7	F8	F9	Tabela ANOVA				
										Fonte da variação	SQ	gl	QM	F_{obs}
A1	12	10	7	7	8	9	8	7	9	Assentos				
A2	15	14	14	11	11	11	12	11	13	Funcionários	66,50			
A3	12	13	13	10	8	11	12	8	10	Resíduo/Erro	29,05			
A4	10	12	9	9	7	10	11	7	8	Total	176,75			

- A Dado que $F_{tabelado} = F_{(3,8,5\%)} = 4,066$, a hipótese nula de igualdade dos esforços exigidos entre os assentos é rejeitada.
- Embora não se possa indicar qual assento necessitou maior esforço, pode-se afirmar que os esforços exigidos não são iguais entre os assentos testados.
- C Existem funcionários que melhor se adaptaram a determinados tipos de assentos, devido ao efeito da interação.
- D Pode-se afirmar que os esforços exigidos são iguais entre os funcionários avaliados.
- E Rejeita-se a hipótese de igualdade de esforços exigidos entre os funcionários a nível 5% de significância visto que observou-se um valor $F_{calculado}$ de 6,86. Assim pode-se afirmar que $\mu_{F1} \neq \mu_{F2} \neq \dots \neq \mu_{F9}$, ou seja todos os funcionários apresentam diferenças.

Teste 4: O declínio no fornecimento de água em determinadas regiões populosas criou a necessidade de se compreender melhor as relações entre fatores econômicos, como o volume da colheita, fatores hidrológicos e do solo. Os dados sobre a produção agrícola (Y em gramas / metro linear) e distância ascendente (X em metros) em uma bacia inclinada são apresentados a seguir.

X	0	10	20	30	45	50	70	80	100	120	140	160	170	190
Y	500	590	410	470	450	480	510	450	360	400	300	410	280	350

Seguem duas tabelas contendo: a) as estimativas dos coeficientes do modelo de regressão linear simples com os respectivos erros padrão e intervalos de confiança a 95% e b) Soma de Quadrados devido à regressão e dos resíduos .

	Estimativas	Erro padrão	P-Value	IC a 95%	
Coef linear	515.447	25.141	<0.00001	[460.669; 570.225]	
Coef angular	-1.060	0.241	0.00088	[-1.586; -0.534]	
	gl	SQ	QM	F	p-value
	Regressão				
	Resíduo		36036.911		
	Total		93942.857		

- A O p-value (valor p ou nível descritivo) relativo ao coeficiente linear pode ser interpretado de que a hipótese nula de coeficiente linear igual zero não foi rejeitada. Em decorrência desta observação, o modelo mais adequado deveria ser um modelo que passe na origem ou seja no ponto (0,0)
- B Embora os coeficientes do modelo de regressão linear simples sempre possam ser estimados, porém, este modelo de regressão linear simples não é adequado devido aos baixos valores de p-values associados aos coeficientes linear e angular
- A amplitude do intervalo de confiança do coeficiente angular da regressão construída a partir dos dados apresentados é de aproximadamente uma unidade.
- D O método de construção da curva de regressão baseia-se na busca da identificação de um polinômio de maior grau que rejeite a hipótese nula.
- E O coeficiente de determinação R^2 é 0,6164. Assim pode-se afirmar que a correlação entre os dados é positiva e igual a 0,7851.

SOLUÇÃO

Teste 5: Em um experimento para investigar o desempenho de quatro marcas diferentes de velas de ignição que seriam usadas em motos de 125 cc de dois tempos, cinco velas de cada marca foram testadas, e se observou o número de quilômetros percorridos (com velocidade constante) até o momento em que apresentaram falha (Nível de significância é de 5%).

Fonte	Graus de liberdade	SQ	QM	F
Marca			14.713,69	
Resíduo				
Total		310.500,76		

- A O p-value (valor p ou nível descritivo) desta análise é de 0,03, e, por este motivo, deve-se rejeitar a hipótese nula, indicando, na significância determinada, que existe diferença significativa entre as marcas de vela
- B A falta de informações na tabela apresentada torna impossível a análise solicitada
- C Não se pode afirmar, a 5% de significância, que as marcas de velas são diferenciadas pela distância percorrida até a primeira falha ($F_{crit} = 0,88$ e $F_{obs} = 3,23$)
- D Pode-se afirmar, a 5% de significância, que as marcas de velas não podem ser diferenciadas pela distância percorrida até a primeira falha ($F_{obs} = 0,88$ e $F_{crit} = 3,23$)
- E O grau de liberdade relacionado ao fator em investigação nesta análise pode ser calculado através da equação $gl = n - m - 1$, onde m indica o número de parâmetros da distribuição populacional esperada dos dados, que, no caso, por ser normal ($m = 2$)

Teste 6: Três marcas de cimento estão sendo testadas quanto à resistência oferecida pelo concreto fabricado com cada uma. Tabelas com resumo e Análise de variância dos resultados estão dadas a seguir:

Grupo	n	Média	Variância	ANOVA					
				Fonte da variação	SQ	gl	QM	F observ	F crítico
A	5	26,6	4,3	Entre grupos	50,533				
B	5	28,0	5,0	Dentro dos grupos					
C	5	23,6	5,8	Total					

Com base nesta tabela podemos afirmar que:

1. Existe diferença entre as marcas de cimento ao nível de 5% de significância
 2. Não podemos afirmar que existe diferença entre as marcas ao nível de 1% de significância
 3. Existe diferença entre as marcas de cimento ao nível de 1% de significancia
 4. Não podemos afirmar que existe diferença entre as marcas ao nível de 5% de significância
- A As afirmações 1 e 3 são corretas
 - B As afirmações 1 e 3 são incorretas
 - C As afirmações 2 e 4 são corretas
 - D As afirmações 1 e 2 são corretas.
 - E A afirmação 1 é correta e a afirmação 2 é incorreta

SOLUÇÃO

Teste 7: Considere os três conjuntos de dados a seguir, nos quais as variáveis de interesse são: X = distância e Y = tempo. Considere que um modelo de regressão linear simples foi ajustado (Y em função de X) em cada amostra.

	Amostra 1		Amostra 2		Amostra 3	
	X	Y	X	Y	X	Y
	15	42	5	16	5	8
	16	35	10	32	10	16
	17	45	15	44	15	22
	18	42	20	45	20	23
	19	49	25	63	25	31
	20	46	30	115	30	60
Coef Ang	1,6857		2,142		1,127	
Coef Linear	13,666		7,868		3,197	
SQT	114,83		5897,5		1627,30	
SQRegressão	49,728		4956,0		1337,65	

- Ao se analisar os modelos de regressão linear simples, o modelo construído utilizando os dados da amostra 2 apresenta o maior coeficiente de determinação (R^2).
- Observou-se uma forte correlação entre X e Y superior a 0.7 em todas as amostras.
- Ao se analisar os modelos de regressão linear, o modelo construído utilizando os dados da amostra 1 apresenta o maior coeficiente de determinação (R^2).
- Todos os modelos de regressão simples são significativos, uma vez que todos apresentaram F_{obs} inferior ao F_{tab} .
- Em cada modelo de regressão simples, os graus de liberdade associado à soma de quadrados devido a regressão é igual a 2, visto que envolvem duas variáveis (X e Y).

Teste 8: Os dados a seguir se referem a uma investigação sobre um processo de fundição onde: X_1 =temperatura da fornalha, X_2 =tempo de moldagem da matriz e Y =diferença de temperatura na superfície de moldagem da matriz.

X_1	1250	1300	1350	1250	1300	1250	1300	1350	1350
X_2	6	7	6	7	6	8	8	7	8
Y	80	95	101	85	92	87	96	106	108

As três tabelas seguintes referem-se respectivamente às somas de quadrados de um modelo de regressão simples de Y com X_1 , às somas de quadrados um modelo de regressão linear múltipla de Y com X_1 e X_2 e as estimativas dos coeficientes do modelo de regressão múltipla com os respectivos erros padrão. (Utilize um nível de significância de 5%)

	g.l.	SQ	QM	F	p-value		G.L.	SQ	QM	F	p-value
Regressão X_1	1	661.5			< 0.001	Regressão X_1 e X_2		715.5			< 0.001
Residual						Residual					
Total	8					Total	8	722.222			

	Estimativas	Erro padrão	valor-p
Coef linear	-199.556	11.641	<0.001
Coef X_1 -Temperatura	0.210	0.009	<0.001
Coef X_2 - Tempo	3.000	0.432	<0.001

- Pode-se afirmar que o modelo de regressão múltipla de Y com X_1 e X_2 é significativo ao nível proposto.
- O coeficiente linear negativo indica uma correlação negativa entre temperatura e tempo. Ou seja, a temperatura e tempo apresentam efeitos inversos (enquanto um aumenta, outro diminui com Y).
- O modelo ajustado correto e significativo é: $Y = 199,5555 + 0,21 \times X_1 + 3 \times X_2$.
- O p-value na Tabela das soma de quadrados do modelo de regressão múltipla indica que a hipótese nula não foi rejeitada, consequentemente o modelo de regressão múltipla não é válido.
- O coeficiente de determinação (R^2) do modelo de regressão múltipla é 0,990. Desta forma, pode-se afirmar que as variáveis X_1 e X_2 possuem forte correlação entre elas, podendo ser positivo ou negativo, em função da covariância entre elas.

SOLUÇÃO

Teste 9: Os dados a seguir foram reproduzidos de um estudo de emissões em uma megacidade (mais que 10 milhões de habitantes) onde CO= monóxido de carbono e NOy=óxido nitroso.

ΔCO	50	60	95	108	135	210	214	315	720
ΔNOy	2,3	4,5	4,0	3,7	8,2	5,4	7,2	13,8	32,1

- A A covariância entre estas duas variáveis é obtida pelo produto dos desvios absolutos em relação à média.
- B Existe uma forte correlação entre as emissões de CO e NOy em um ambiente urbano.
- C As variáveis emissões de CO e NOy são independentes.
- D Na construção de um modelo de regressão linear simples entre estas variáveis, seria observada um coeficiente angular negativo.
- E Os dados indicam que um aumento de NOy tem como consequência um aumento também de emissões de CO.

Teste 10: Abaixo observam-se dados sobre a resistência à flexão (psi) em amostras do concreto não-inchado agregado com argila com 7 dias:

257	327	317	300	340	340	343	374	377	386	383	393	407	407	434
427	440	407	450	440	456	460	456	476	480	490	497	526	546	700

Ao se realizar o teste Kolmogorov-Smirnov (K-S) através de um software, o resultado foi a seguinte mensagem (considere 5% de significância):

Kolmogorov-Smirnov test: p-value = 0.9572

- A Para as 30 observações acima, tem-se um grau de liberdade de $\theta=27$, uma vez que $\theta = n - m - 1$ (onde n: número de observações e m: número de parâmetros estimados. Para o caso, $m=2$, pois estimaram-se média e desvio padrão)
- B O valor crítico da distribuição utilizada baseia-se no cálculo da estatística Z, da distribuição normal
- C No teste realizado comprova-se, ao se comparar o p-value (valor p ou nível descritivo) com o nível de significância considerado, que os dados seguem uma distribuição normal, uma vez que o teste "Kolmogorov-Smirnov" (K-S) é utilizado somente para este fim.
- D O teste K-S se baseia no somatório das diferenças entre os dados observados e esperados, para posterior comparação com os valores críticos extraídos da tabela.
- E Pode-se afirmar que houve aderência entre os dados e o modelo de distribuição proposto.

SOLUÇÃO

Distribuição Normal - valores de $P(0 \leq Z \leq z_0)$

Table with columns z_0 from 0.0 to 3.0 and rows for probabilities from 0.00 to 0.99. Includes a small table for z_0 between 1.9 and 2.0.

Chi-Quadrado - valores $\chi^2_{v,p}$ em que $p = P(\chi^2_v \geq \chi^2_{v,p})$

Table with columns p from 0.995 to 0.005 and rows for degrees of freedom from 1 to 60.

t de Student - valores $t_{v,p}$ em que $p = P(t_v \geq t_{v,p})$

Table with columns p from 0.100 to 0.010 and rows for degrees of freedom from 1 to 30.

Distribuição F - valores de $F_{v_1,v_2,p}$ em que $p = P(F_{v_1,v_2} \geq F_{v_1,v_2,p})$; $p = 0,05$

Table with columns p from 0.10 to 0.01 and rows for degrees of freedom from 1 to 30.

Distribuição F - valores de $F_{v_1,v_2,p}$ em que $p = P(F_{v_1,v_2} \geq F_{v_1,v_2,p})$; $p = 0,01$

Table with columns p from 0.10 to 0.01 and rows for degrees of freedom from 1 to 30.

graus de liberdade do denominador (v2)

graus de liberdade do denominador (v2)

graus de liberdade do numerador (v1)

graus de liberdade (v)

graus de liberdade (v)

SOLUÇÃO

PRO3200 – Estatística

Turma:

Prof:

Prova 3 - 2018

Nome (completo legível):

INSIRA SEU NÚMERO SEU USP nas caixas abaixo.

Se seu número USP tem apenas 7 dígitos, a primeira coluna (i.e., mais à esquerda) deve ter o 0 pintado/preenchido.

0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9

LEIA ATENTAMENTE AS SEGUINTE INSTRUÇÕES:

- 1) Use caneta azul ou preta para marcar as caixas e preencha a caixa totalmente para correta interpretação. Exemplo: . Não use .
- 2) As respostas devem ser assinaladas exclusivamente na folha de resposta.
- 3) A prova tem duração de 120 minutos. Não haverá tempo adicional.
- 4) O aluno deve comprovar sua identidade com documento oficial.
- 5) Após a entrega da prova por um aluno, nenhum outro aluno poderá entrar na sala.
- 6) Alunos só podem sair da sala de prova após entregarem a prova.
- 7) É permitido o uso de calculadoras.
- 8) Não é permitido o uso de telefones celulares, tablets, equipamentos móveis de qualquer natureza e computadores. Esses equipamentos devem ser guardados nas bolsas/mochilas e colocados na frente da sala.
- 9) É permitido o uso de um formulário pessoal, com nome. Apenas uma folha A4 será permitida.
- 10) Favor não retirar o grampo da prova.

FOLHA DE RESPOSTA

As respostas devem ser assinaladas exclusivamente nesta página.

Teste 1: A B C D

Teste 2: B C D E

Teste 3: A C D E

Teste 4: A B D E

Teste 5: A B C E

Teste 6: A B C E

Teste 7: B C D E

Teste 8: B C D E

Teste 9: A C D E

Teste 10: A B C D