Caríssimos,

O prof. Juci e eu estamos com um curso de estatística e financeira para o concurso do ICMS/RJ, organizado pela FGV.

Alguns de nossos alunos pediram pra gente resolver a recente prova da FGV, para o Governo do Estado do Amapá.

No nosso curso, já encerramos a parte de estatística. Entraremos esta semana em financeira. Por isso dá mais para incluir tais questões em nosso curso. Assim, estamos disponibilizando os comentários aqui na parte aberta do site.

Vamos lá!

Auditor Da Receita Estadual – Estado Amapá – 1° dia – Tipo 1

Questão 16.

Pretende-se usar um modelo de regressão linear $Y = \alpha + \beta X + \varepsilon$ para ajustar n pares de valores observados (x1, y1), (x2, y2), ..., (xn, yn). Supõe-se que os erros ε tenham média 0, sejam normalmente distribuídos, todos com mesma variância, e sejam não correlacionados. Se a e b são as estimativas de mínimos quadrados de α e β, respectivamente, avalie as afirmativas a seguir:

I - As estimativas de mínimos quadrados de α e β são os valores de a e b que resolvem as equações normais

$$an + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{e} \quad a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} x_i y_i$$

II - a e b são estimadores não viesados de α e β respectivamente.

III - O coeficiente de determinação, R^2, mede a porcentagem da variância total dos valores y’s que é explicada pela regressão e, desse modo, quanto maior o valor do coeficiente de determinação, melhor é o ajuste do modelo.

Assinale:

(A) se apenas a afirmativa I estiver correta.

(B) se apenas a afirmativa II estiver correta.

(C) se apenas as afirmativas I e II estiverem corretas.

(D) se apenas as afirmativas II e III estiverem corretas.

(E) se todas as afirmativas estiverem corretas.

Comentários.

O primeiro item cobra um aspecto que não estudamos em nosso curso. Nós vimos apenas as fórmulas dos coeficientes *a* e *b*. Dissemos que tais coeficientes são obtidos do seguinte modo: tomamos os erros do modelo, elevamos ao quadrado, e somamos todos eles. Feito isso, procuram-se coeficientes *a* e *b* que minimizam a soma dos quadrados destes desvios.

Mas nós não mostramos, matematicamente, quais são as equações que originam as fórmulas de *a* e *b*. E, sinceramente, não sei se vale a pena esquentar muito a cabeça com isso, apesar da cobrança desta questão.

Uma alternativa seria o candidato simplesmente resolver o sistema de equações fornecidos e verificar que ele, de fato, resulta nas expressões usuais para *a* e *b*.

Outra solução, que exige ferramentas de cálculo, é a que segue.

Queremos minimizar a seguinte expressão:

$$\sum_{i=1}^{n}\left(y_i - (a + bx_i)\right)^2$$

Para tanto, calculamos as derivadas parciais em relação a *a* e *b*, e igualamos a zero.

Derivando em relação a "*a*":

$$2\sum_{i=1}^{n}\left(y_i - (a + bx_i)\right)\times(-1) = 0$$

$$\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i = 0$$

E esta foi exatamente a primeira equação fornecida no item. E como foi obtida esta equação? Derivamos em relação a *a* e igualamos a zero, para determinamos o ponto de mínimo.

Resolvendo esta equação é que chega-se à expressão usual para o coeficiente *a*.

Por fim, derivando em relação a *b*, temos:

$$2\sum_{i=1}^{n}\left[\left(y_i - (a + bx_i)\right)\times(-x_i)\right] = 0$$

$$\sum_{i=1}^{n}\left[\left(y_i - (a + bx_i)\right)\times x_i\right] = 0$$

$$\sum_{i=1}^{n}\left(y_i x_i - ax_i - bx_i^{\,2}\right) = 0$$

$$\sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} bx_i^{\,2}$$

Que é a outra equação fornecida.

Resumo da ópera: não esquente a cabeça com este passo a passo. A maioria das questões cobram apenas o conhecimento das fórmulas de *a* e *b*. Nesta questão em específico, a cobrança caiu sobre o procedimento de obtenção de tais fórmulas, o que não é tão cobrado.

E, para marcar se o item estava certo ou errado, uma opção seria resolver as equações fornecidas para ver se elas fornecem as expressões corretas para *a* e *b*.

Segundo item: de fato, *a* e *b*, além de serem estimadores de mínimos quadrados, também são estimadores não viciados (ou não viesados, ou não tendenciosos).

Terceiro item: temos exatamente a definição de coeficiente de determinação.

Apesar do primeiro item mais complicadinho, os itens dois e três foram bem tranqüilos. Já permitiriam que o candidato, de cara, ficasse entre as alternativas "d e "e".

Todos os itens estão corretos.

Gabarito: E

Questão 17.

Os dados a seguir são as quantidades de empregados de cinco pequenas empresas: 6, 5, 8, 5, 6. A variância da quantidade de empregados dessas cinco empresas é igual a:

(A) 0,8.

(B) 1,2.

(C) 1,6.

(D) 2,0.

(E) 2,4.

Comentários:

O cálculo alternativo, em geral, é sempre mais fácil. Como ele é feito?

Calculamos a média dos quadrados. Calculamos o quadrado da média. E subtraímos um do outro.

Média dos quadrados:

$$\frac{36+25+64+25+36}{5}=37,2$$

Quadrado da média:

$$\left(\frac{6+5+8+5+6}{5}\right)^2=36$$

Diferença:

$$37,2-36=1,2$$

Gabarito: B

Questão 18.

Uma urna contém 50 bolinhas idênticas numeradas de 1 a 50. Se quatro bolinhas são aleatoriamente sorteadas com reposição, a probabilidade de que, dos quatro números sorteados, dois sejam pares e dois sejam impares é igual a:

(A) 12,5%.

(B) 25,0%.

(C) 37,5%.

(D) 50,0%.

(E) 62,5%.

Resolução:

Há diversas formas de resolver este problema.

Você pode listar todos os casos possíveis, listar os casos favoráveis, contar cada um deles e dividir um pelo outro.

Ficaria assim:

Casos possíveis: (par, par, par, par), (ímpar, par, par, par), ...

Casos favoráveis: (par, par, ímpar, ímpar), (par, impar, par, impar)...

Ou então, você poderia usar análise combinatória para fazer a contagem acima indicada.

Outra opção: podemos analisar este problema usando a distribuição binomial. Temos 4 experimentos independentes. Consideramos sucesso se a bola sorteada, em cada extração, for par. Seja X a variável que designa o número de sucessos. Queremos calcular:

$$P(X = 2) = \binom{4}{2} \times 0{,}5^2 \times 0{,}5^2 = 0{,}375$$

Gabarito: C

Questão 19

Em relação à distribuição normal, assinale a afirmativa incorreta.

(A) a função de densidade de probabilidade é simétrica em relação à média.

(B) se X tem distribuição normal com média μ e variância σ^2 então a variável $Z = \frac{X - \mu}{\sigma^2}$ tem distribuição normal padrão.

(C) a probabilidade de que uma variável Z que tenha distribuição normal padrão seja maior do que 5 é aproximadamente igual a 0.

(D) a média de uma variável aleatória que tenha distribuição normal pode ser negativa.

(E) o valor da mediana é igual ao valor da média.

Comentários:

Letra A: correto. A distribuição normal é simétrica em torno da média. Por sinal, a média coincide com a mediana, que coincide com a moda.

Letra B: errado. Para conseguirmos a distribuição normal padrão, a divisão é feita pelo desvio padrão, não pela variância. Por sinal, esta transformação é muito importante. É usada em praticamente todos os exercícios de inferência.

Letra C: correto. Como a prova não trouxe uma tabela de áreas para a variável normal, a idéia era que o candidato tivesse noção de ordem de grandeza de probabilidades associadas à distribuição normal. De fato, se você notar em qualquer tabela fornecida em livros, provas, etc, as áreas fornecidas geralmente vão até próximo de 3,0, que já abrangem probabilidades de mais de 99%. Ou seja, para uma distribuição normal padrão, o valor 5,0 é muitíssimo afastado da média. É muito raro ocorrerem valores superiores a 5.

Letra D: correto. A média de uma variável normal qualquer pode ser negativa sim. Já uma variável normal padrão, esta tem sempre média zero.

Letra E: correto, como já mencionamos na letra A.

Gabarito: B

Questão 20

Para testar a hipótese de que uma média populacional μ de uma variável normalmente distribuída com variância igual a 64 é maior do que 200, uma amostra aleatória simples de tamanho 100 será observada. Ao nível de significância de 5%, o critério de decisão usual estabelece que a hipótese nula de que $\mu \leq 100$ deve ser rejeitada se o valor observado da média amostral for:

Dados: se Z tem distribuição normal padrão,

P[0 < Z < 0,45] = 1,64;

P[0 < Z < 0,475] = 1,96;

P[0 < Z < 0,49] = 2,33]

(A) maior do que 201,312.

(B) menor do que 198,788.

(C) maior do que 204,860.

(D) menor do que 196,348.

(E) maior do que 210,346.

Comentários.

Achei a questão meio confusa. Inicialmente, ela menciona que a hipótese nula era: média maior que 200. Depois, é indicado que a hipótese nula é: média menor ou igual a 100.

Pelas alternativas, que apresentam valores próximos a 200, dá para perceber que a hipótese nula, de fato, é “média maior que 200”. Apesar disso, caberia pedir a anulação da questão, pela falha no enunciado.

Feitos estes comentários, vamos ao teste.

O teste é unilateral, com região crítica formada pelos valores significativamente menores que 200.

O valor crítico de Z, fornecido pela questão, é de -1,64.

O valor do desvio padrão da média amostral é:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{10} = 0,8$$

O valor de $\overline{X}$ correspondente é:

$$-1,64 = \frac{\overline{X} - 200}{0,8} \Rightarrow \overline{X} = 198,68$$

Não há alternativa que contemple este valor.

O gabarito fornecido foi a letra "A".

Na verdade, o teste que resulta nesta resposta seria o seguinte.

Hipótese nula: média menor que 200

Hipótese alternativa: média maior que 200.

Aí o valor crítico seria de 1,64, e teríamos:

$$1,64 = \frac{\overline{X} - 200}{0,8} \Rightarrow \overline{X} = 201,312$$

Bom, na minha opinião, não dá para interpretar a questão e concluir que este era o teste pretendido.

Diante de tanta confusão sobre a hipótese a ser testada (média menor que 100? média menor que 200? média maior que 200?), acho que daria para pedir a anulação da questão.

Gabarito : A

É isso.

No próximo ponto resolveremos a prova de fiscal.

Abraços

Vítor.