Resoluções comentadas das 5 questões de Estatística (números 71 a 75 da prova tipo 1) do concurso para Analista Legislativo – Atribuição Contador realizada pela FCC em 16.09.2007.

71. Numa pesquisa realizada com 300 famílias, levantaram-se as seguintes informações:

Número de filhos	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Proporção de famílias	**0,17**	**0,20**	**0,24**	**0,15**	**0,10**	**0,10**	**0,04**

Com base nestas informações a média e a mediana do número de filhos são dadas, respectivamente, por

(A) 2,27 e 3

(B) 3 e 2

(C) 2,27 e 2

(D) 2,5 e 3,5

(E) 2,5 e 3

Vamos criar uma tabela com 3 colunas, onde na 1ª coluna teremos a variável "número de filhos", que vamos designar como X, noutra coluna teremos f (freqüência relativa simples, dada pela proporção) e na 3ª coluna teremos a freqüência relativa acumulada (f_{AC}).

X	f	f_{AC}
0	0,17	0,17
1	0,20	0,37
2	0,24	0,61
3	0,15	0,76
4	0,10	0,86
5	0,10	0,96
6	0,04	1,00
Σ	1,00	-

Assim o fazendo, vemos que a freqüência relativa acumulada até o valor X = 1 é de apenas 0,37 (37%). Logo, só teremos 50% da distribuição quando X = **2**, <u>sendo este o valor da mediana</u>. Só com este raciocínio já eliminaremos 3 das 5 opções de resposta.

E podemos chegar à opção correta sem fazer nenhum cálculo, bastando raciocinar que, se a distribuição fosse simétrica (e neste caso: média, moda e mediana serão iguais) a média seria o valor central da distribuição (no caso, igual a 3). Mas basta olhar para a freqüência relativa simples e verificar que a distribuição <u>não é simétrica</u>. Portanto, a média é diferente de 3, sendo a 3ª opção de resposta a única opção possível (média = 2,27 e mediana = 2). Como poderíamos afirmar que a distribuição é simétrica?

Se a maior freqüência estivesse no valor central e as freqüências eqüidistantes deste fossem iguais. <u>Por exemplo</u>:

X	f
0	0,04
1	0,12
2	0,19
3	0,30
4	0,19
5	0,12
6	0,04
Σ	1,00

Maior freqüência, com freqüências eqüidistantes iguais.

Neste caso, poderíamos, sem fazer cálculo algum, afirmar que a média é igual a 3. Mas como isto não ocorre, então a média, <u>com certeza</u>, é diferente de 3.

Para quem não enxergasse esse raciocínio, a fórmula a ser utilizada para a média é: $\overline{X} = \sum X \cdot f$, pois a freqüência relativa está na forma unitária. Assim, teremos:

X	f	X · f
0	0,17	0
1	0,20	0,20
2	0,24	0,48
3	0,15	0,45
4	0,10	0,40
5	0,10	0,50
6	0,04	0,24
Σ	1,00	2,27

Logo, média = **2,27**.

72. Se a média e a variância da variável aleatória X são 12 e 80 respectivamente, então a média e a variância da variável aleatória Y = X/4 + 1 são dadas respectivamente por

(A) 4 e 20

(B) 4 e 5

(C) 3 e 20

(D) 4 e 21

(E) 3 e 5

Do enunciado, temos para a variável X: a média $\overline{X} = 12$ e a variância $\sigma^2_X = 80$.

Para responder com facilidade esta questão, vamos relembrar algumas importantes propriedades para a média e para a variância:

1) Quando <u>multiplicamos ou dividimos</u> todos os valores de uma variável (X) por uma constante (k):

a sua <u>média</u> fica <u>multiplicada ou dividida</u> pela constante;

a sua <u>variância</u> fica <u>multiplicada ou dividida</u> pelo <u>QUADRADO</u> da constante;

2) Quando <u>somamos ou subtraímos</u> uma constante (k) a todos os valores de uma variável (X):

a sua <u>média</u> fica <u>acrescida ou diminuída</u> dessa constante;

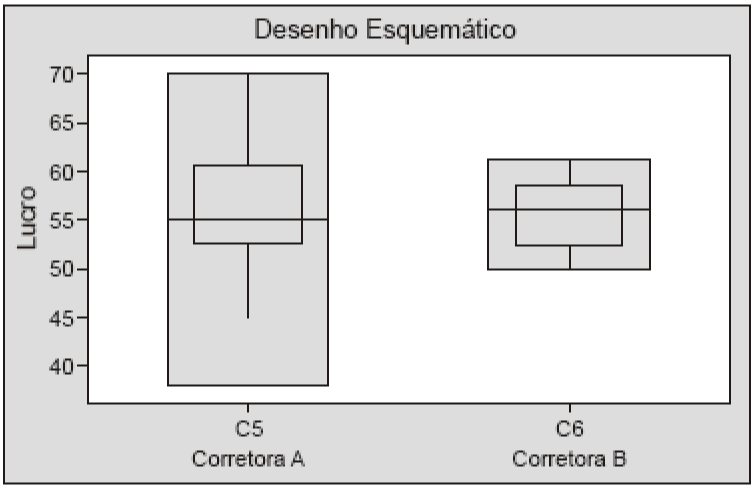
a sua <u>variância</u> fica <u>INALTERADA</u>, pois a variância de uma constante é igual a zero;

Portanto, sendo a variável $Y = \frac{X}{4} + 1$, ao aplicarmos as propriedades supracitadas, teremos:

$$\overline{Y} = \frac{\overline{X}}{4} + 1 \Rightarrow \overline{Y} = \frac{12}{4} + 1 \Rightarrow \overline{Y} = 3 + 1 \Rightarrow \boxed{\overline{Y} = 4};$$

$$\sigma^2_Y = \frac{\sigma^2_X}{4^2} \Rightarrow \sigma^2_Y = \frac{80}{16} \Rightarrow \boxed{\sigma^2_Y = 5}.$$

73. Para se estudar o desempenho das corretoras de ações A e B, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada computou-se a porcentagem de lucro apresentada durante o período de um ano. Os gráficos a seguir apresentam os desenhos esquemáticos relativos à porcentagem de lucro das amostras de A e B durante o período citado.



Relativamente à porcentagem de lucro obtida por essas corretoras pode-se afirmar que

(A) exatamente 25% dos valores de A são inferiores a 55.

(B) menos de 50% dos valores de B são superiores a 55.

(C) o maior valor de A é 60.

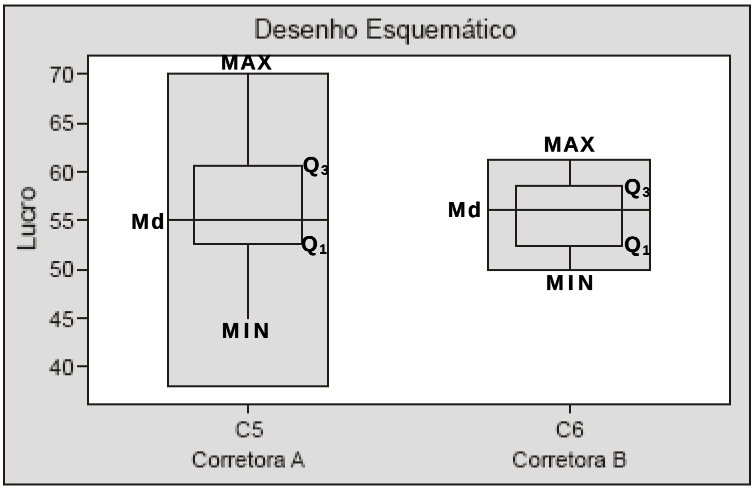
(D) os valores de A apresentam maior variabilidade que os de B.

(E) os valores de B apresentam assimetria positiva.

É apresentado na questão o desenho esquemático chamado Diagrama de Caixa (Box Plot), que utiliza o "esquema dos cinco números" a saber: Mínimo, 1º Quartil, Mediana, 3º Quartil e o Máximo da distribuição, onde os quartis são chamados de "juntas" da Caixa.

A distância entre as juntas (d_j) corresponde à amplitude interquartílica, pois será dada pela diferença entre o 3º Quartil (Q_3) e o 1º Quartil (Q_1), ou seja: $d_j = Q_3 - Q_1$. Essa medida, serve para a detecção de outiliers (valores atípicos) de uma distribuição. Isto não é cobrado na presente questão, mas é útil saber para futuros concursos: são considerados outliers os valores inferiores a $Q_1 - 1{,}5d_j$ ou superiores a $Q_3 + 1{,}5d_j$.

Para resolver a presente questão vamos posicionar, no diagrama, as cinco medidas citadas:



Observando o desenho esquemático, agora com as cinco medidas posicionadas nas caixas correspondentes às Corretoras A e B, vemos que:

Para a Corretora A:

O valor mínimo (MIN) é aproximadamente igual a 45;

O valor do 1º Quartil (Q_1) está entre 50 e 55;

O valor da Mediana (Md) é aproximadamente igual a 55;

O valor do 3º Quartil (Q_3) é aproximadamente igual a 60;

O valor máximo (MAX) é aproximadamente igual a 70.

Para a Corretora B:

O valor mínimo (MIN) é aproximadamente igual a 50;

O valor do 1º Quartil (Q_1) está entre 50 e 55;

O valor da Mediana (Md) está entre 55 e 60;

O valor do 3º Quartil (Q_3) está entre 55 e 60;

O valor máximo (MAX) está entre 60 e 65;

Analisando agora as opções de resposta, vemos que:

A 1ª opção está errada porque não é exatamente 25% dos valores da Corretora A que são inferiores a 55 e sim aproximadamente 50%, pois é a Mediana (e não o 1º Quartil) que está próxima do valor 55;

A 2ª opção está errada, porque a Mediana de B é um valor maior do que 55. Então entre este valor e a Mediana teremos X% e acima de 55 teremos: 50% + X%, ou seja, mais de 50% serão superiores a 55;

A 3ª opção está errada porque o maior valor de A não é igual a 60, mas sim superior a 60. Como podemos notar no desenho esquemático, o máximo de A é aproximadamente igual a 70;

A 4ª opção de resposta está correta, porque a Amplitude Total (diferença entre o máximo e o mínimo) da distribuição A é bem maior do que a da distribuição B (basta olhar para o tamanho das caixas). Então a variabilidade de A, com certeza, será maior;

A 5ª opção de resposta está errada, pois na distribuição B a mediana está mais próxima do 3º quartil e assim a assimetria é negativa (e não positiva). Haverá assimetria positiva na distribuição A, pois naquela a mediana está mais próxima do 1º quartil. Quando a mediana está exatamente no meio (da amplitude entre os dois quartis), a distribuição será simétrica. Basta lembrar do Coeficiente Quartílico de Assimetria de Pearson, dado por: $AS = \dfrac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1}$.

Repare que o denominador, dado pela amplitude interquartílica (diferença entre os quartis) será sempre positivo, pois o 3º quartil sempre será maior do que o 1º quartil. Logo a assimetria (positiva, negativa ou nula) depende do resultado do numerador, que variará em função da posição da mediana.

Para facilitar o entendimento, vamos arbitrar valores para as medidas, por exemplo, suponha que $Q_3 = 60$, $Q_1 = 50$ e que a mediana esteja exatamente entre estes valores, ou seja, Md = 55. Neste caso, a assimetria será nula, pois $Q_3 + Q_1 - 2Md = 0$. Agora suponha que, a mediana é igual a 57 (está mais próxima do 3º quartil). Neste caso, a assimetria será negativa (caso da distribuição B), pois $Q_3 + Q_1 - 2Md = 60 + 50 - 114 = -4$. Agora suponha que a mediana é igual a 52 (mais próxima do 1º quartil). Então teremos uma assimetria positiva (caso da distribuição A), pois $Q_3 + Q_1 - 2Md = 60 + 50 - 104 = 6$.

74. Uma rede local de computadores é composta por um servidor e 2 (dois) clientes (Z e Y). Registros anteriores indicam que dos pedidos de certo tipo de processamento, cerca de 30% vêm de Z e 70% de Y. Se o pedido não for feito de forma adequada, o processamento apresentará erro. Sabendo-se que 2% dos pedidos feitos por Z e 1% dos feitos por Y apresentam erro, a possibilidade do sistema apresentar erro é

(A) 5%

(B) 4,1%

(C) 3,5%

(D) 3%

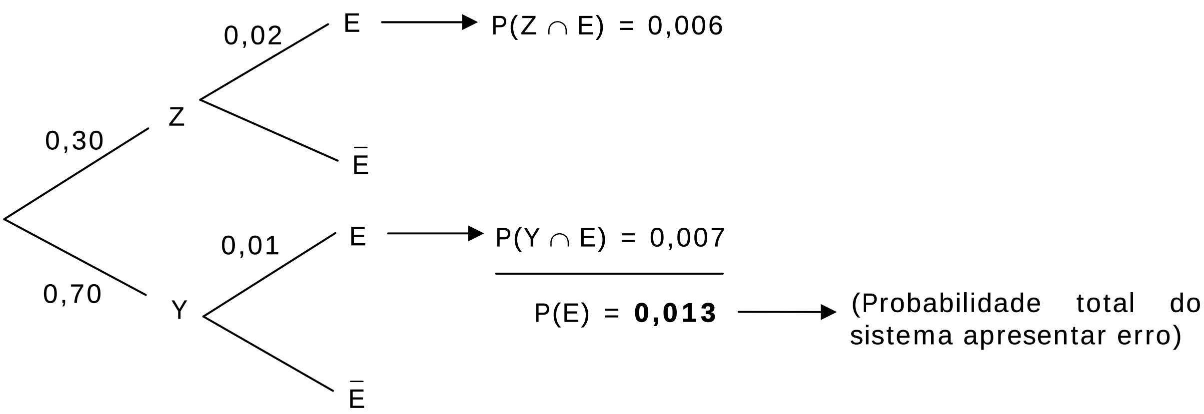
(E) 1,3%

Trata-se da aplicação do Teorema da Probabilidade Total, pois é pedida a probabilidade do sistema apresentar erro, seja o processamento oriundo do cliente Z ou do cliente Y.

Facilita muito a resolução construirmos a "árvore" de probabilidades, designando por:

E → o processamento apresentou erro;

$\bar{E}$ → o processamento não apresentou erro;



75. Uma pesquisa eleitoral foi realizada com uma amostra de 500 eleitores com o objetivo de estudar a influência da idade na preferência por dois candidatos presidenciais. Os resultados obtidos foram os seguintes:

Preferência Idade (anos)	**Candidato Alfa**	**Candidato Beta**	**Indecisos**	**Total**
20 |---- 30	**68**	**117**	**15**	**200**
30 |---- 50	**102**	**70**	**27**	**200**
50 |---- 80	**80**	**3**	**17**	**100**
Total	**250**	**190**	**60**	**500**

Duas pessoas serão selecionadas ao acaso e com reposição dentre os 500 eleitores.

A probabilidade de exatamente uma pertencer à faixa etária 50 |---- 80 e preferir o candidato Alfa é

(A) $\frac{168}{625}$

(B) $\frac{84}{625}$

(C) $\frac{64}{625}$

(D) $\frac{42}{625}$

(E) $\frac{21}{625}$

Devemos grifar duas partes importantes do enunciado:

1) A seleção das duas pessoas é com reposição, ou seja, a probabilidade da 2ª pessoa selecionada pertencer ou não à faixa etária especificada será a mesma da 1ª;
2) É pedida a probabilidade de exatamente uma pertencer à faixa etária e preferir o candidato Alfa.

Assim, designando os eventos por:

F → a pessoa selecionada pertence à faixa etária e prefere o candidato Alfa;

$\bar{F}$ → a pessoa selecionada não pertence à faixa etária e/ou não prefere o candidato Alfa.

As probabilidades desses eventos são: $P(F) = \frac{80}{500}$ e o seu complementar $P(\bar{F}) = \frac{420}{500}$.

Pode ocorrer exatamente um sucesso (a pessoa pertence à faixa e prefere o candidato) de duas formas: a 1ª pessoa selecionada sim e a 2ª não ou a 1ª pessoa selecionada não e a 2ª sim, ou seja, $P(F \cap \bar{F}) \cup P(\bar{F} \cap F)$.

Como a seleção é com reposição, $P(F \cap \bar{F}) = P(\bar{F} \cap F)$ e podemos fazer: $2 \cdot \frac{80}{500} \cdot \frac{420}{500}$.

Fazendo todas as simplificações possíveis restará: $\frac{8}{25} \cdot \frac{21}{25} = \boxed{\frac{168}{625}}$.

Mas a questão deverá ser anulada, pois a tabela dada está imperfeita. A coluna de Indecisos deveria totalizar 60, mas 15 + 27 + 17 = 59. A faixa de 30 a 50 anos deveria totalizar 200, mas 102 + 70 + 27 = 199.

Logo há um erro de digitação na tabela, que deveria trazer para o número de Indecisos na faixa de 30 a 50 anos a quantidade 28 e não 27. Essa pequena imperfeição não interfere na resolução da questão, mas toda questão de concurso deve ser elaborada tendo o seu enunciado bem claro, de modo a não gerar dúvida para os candidatos.

Assim, devido ao erro na digitação da tabela, a providência mais correta a ser tomada pela banca é a anulação da questão.

GABARITO

71. C 72. B 73. D 74. E 75. A