19. Os jogadores *A* e *B* se encontram para jogar uma partida de tênis em no máximo cinco *sets*, na qual será vencedor aquele que primeiro ganhar três *sets*.

Por exemplo, partidas terminadas poderão ter como resultado: *AAA*, *AABA*, *BABAB*, etc. Então, o número de possíveis resultados para uma partida terminada é:

(A) 4.
(B) 10.
(C) 6.
(D) 20.
(E) 8.

RESOLUÇÃO:

Para a partida terminar em 3 sets teremos apenas **2** possibilidades: AAA ou BBB;

Para a partida terminar em 4 sets serão **6** possibilidades: AABA, BBAB, ABAA, BABB, BAAA ou ABBB;

Para a partida terminar em 5 sets, serão **12** possibilidades: AABBA, BBAAB, ABBAA, BAABB, BBAAA, AABBB, ABABA, BABAB, BAABA, ABBAB, BABAA ou ABABB;

Número total de resultados possíveis: 2 + 6 + 12 = **20**.

Gabarito oficial - letra D.

20. Uma companhia utiliza um sistema de avaliação de desempenho de seus funcionários por meio de dois indicadores de *performance*: Qualidade das tarefas e a Tempestividade com que as tarefas são realizadas.

Os funcionários receberam, na última avaliação, as medidas indicadas na tabela a seguir:

Medidas	Indicador	
	Qualidade	Tempestividade
Média	50	25
Desvio-Padrão	10,0	6,0
Coeficiente de Variação (%)	20	24

Com base na tabela, é correto afirmar que:

(A) a média aritmética não é uma boa medida para representar a *performance* dos funcionários em face do elevado nível de dispersão das avaliações.
(B) as avaliações da Qualidade foram mais dispersas do que as avaliações da Tempestividade.
(C) as avaliações da Qualidade foram mais homogêneas do que as avaliações da Tempestividade.
(D) os funcionários demoram mais para realizar as tarefas, mas a qualidade das tarefas é melhor.
(E) nada se pode afirmar sem o conhecimento do tamanho da amostra.

RESOLUÇÃO:

Questão muito parecida com uma das questões do concurso anterior, sendo que nesta prova já foram dados, prontinhos, os percentuais correspondentes aos Coeficientes de Variação (CV's). Assim como na prova anterior, vamos relembrar a parte teórica à página 37 do livro "Estatística Básica para Concursos" da Editora Ferreira:

"Considera-se que um CV superior a 50% indica alto grau de dispersão e conseqüentemente pequena representatividade da Média, enquanto para um CV inferior a 50% a Média será tanto mais representativa quanto menor for o valor do CV, ou seja, quanto menor for o CV mais homogênea será considerada a série e quanto maior for o CV, mais heterogênea."

Observando-se a tabela dada, vemos que o CV da Qualidade é inferior ao da Tempestividade, indicando que as avaliações da Qualidade foram mais homogêneas do que as da Tempestividade.

Gabarito oficial - letra C.

21. Dentre as distribuições de probabilidades a seguir, aquela em que $E(X) = E(X - E(X))^2$ é:

(A) de densidade $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty$

(B) de densidade $f_X(x) = 1, \quad 0 < x < 1$

(C) $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots n$

(D) $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$

(E) $P(X = x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{N+M}{n}}, \quad x = 0, 1, 2, \ldots n$

RESOLUÇÃO:

O que o enunciado apresenta, nada mais é do que: E[X] = V[X], ou seja, a esperança é igual à variância.

Dentre as distribuições de probabilidade, a única em que isso ocorre é na Distribuição de Poisson (E[X] = V[X] = λ), cuja probabilidade de "x sucessos" é dada por: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$.

Gabarito oficial - letra D.

22. Considere uma Amostra Aleatória Simples de n unidades extraídas de uma população na qual a característica, X, estudada tem distribuição Normal com média μ e variância σ^2, ambas desconhecidas, mas finitas. Considere, ainda, as estatísticas média da amostra, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, e variância da amostra $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Então, é correto afirmar que:

(A) $\overline{X}$ e S^2 são, ambos, não tendenciosos para a estimação da média e da variância da população, respectivamente.

(B) $\overline{X}$ é não-tendencioso, mas S^2 é tendencioso para a estimação da média e da variância da população, respectivamente.

(C) $\overline{X}$ é tendencioso, mas S^2 é não-tendencioso para a estimação da média e da variância da população, respectivamente.

(D) $\overline{X}$ e S^2 são, ambos, tendenciosos para a estimação da média e da variância da população, respectivamente.

(E) $\overline{X}$ e S^2 são, ambos, não-tendenciosos para a estimação da média e da variância da população, mas apenas $\overline{X}$ é consistente.

<u>RESOLUÇÃO</u>:

$\overline{X}$, dado por $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é um estimador não tendencioso (não viciado) da média populacional μ, ou seja, $\overline{X}$ = μ.

Já a variância S^2, sendo dada por: $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ será um estimador viciado (tendencioso) da variância populacional σ^2. Justamente para corrigir essa tendência, é que precisamos usar o Fator de Correção de Bessel, dado por: $\frac{n}{n-1}$. Assim, o estimador não viciado de σ^2 é: $\frac{n \cdot S^2}{n-1}$.

Para os estimadores dados no enunciado:

$\overline{X}$, será <u>não tendencioso</u> para a estimação da média populacional μ;

S^2 será <u>tendencioso</u> para a estimação da variância populacional σ^2.

Gabarito oficial - letra B.

23. Sejam A, B e C três eventos quaisquer definidos em um espaço amostral S. Então,

$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$ refere-se à probabilidade da ocorrência de:

(A) um ou dois eventos.
(B) exatamente um dos eventos.
(C) pelo menos um dos eventos.
(D) no máximo dois eventos.
(E) pelo menos dois eventos.

<u>RESOLUÇÃO</u>:

Questão ardilosa. À primeira vista dá a impressão de que sobrará apenas um dos eventos: A, B ou C.

Mas basta saber que a união de 3 conjuntos (A, B e C) é dada por:

$A \cup B \cup C = A + B + C - (A \cap B) - (A \cap C) - (B \cap C) + (A \cap B \cap C)$.

Na presente questão, vemos que está sendo tirada, em cada conjunto, a interseção comum dois a dois <u>sem que seja incluída a interseção dos três</u>. Logo o enunciado refere-se à probabilidade da ocorrência de um ou dois dos eventos. A opção da letra D é uma armadilha, pois "no máximo dois eventos" inclui o zero, ou seja, nenhum dos três, o que não é verdade.

Gabarito oficial - letra A.

24. O coeficiente de determinação de um modelo de regressão linear serve como uma importante ferramenta para avaliar o grau de ajustamento do modelo aos dados.

A respeito desse coeficiente, assinale a afirmativa **incorreta**.

(A) Seu valor varia entre 0 e 1.
(B) É invariante a uma mudança de escala das variáveis independentes.
(C) É utilizado para escolher modelos com número de variáveis independentes diferentes.
(D) É uma função não decrescente no número de variáveis independentes no modelo.
(E) Representa a participação relativa da soma dos quadrados da regressão sobre a soma dos quadrados total.

RESOLUÇÃO:

O enunciado pede para apontar a alternativa INCORRETA.

A opção da letra A está correta, pois o coeficiente de determinação (R^2) é o quadrado do coeficiente de correlação e sendo este último um número real variando de –1 até +1, o coeficiente de determinação estará compreendido entre 0 e 1;

A opção da letra B também está correta, pois uma mudança na escala de valores das variáveis não altera o coeficiente de correlação e conseqüentemente não alterará o R^2;

A opção da letra D também está correta, pois ao aumentarmos o número de variáveis independentes o R^2 pode aumentar ou ficar igual, mas nunca decrescerá;

A opção da letra E também está correta, pois o coeficiente de determinação indica quantos % a variação explicada pela regressão representa da variação total e é obtido por: $R^2 = \frac{VE}{VT}$, onde VE é a variação explicada pelo modelo (soma dos quadrados da regressão) e VT é a variação total (soma total dos quadrados, ou seja, regressão + residuais);

A letra C está **incorreta**, pois diz que <u>é</u> utilizado para escolher modelos com número de variáveis aleatórias independentes diferentes, quando, na realidade, <u>pode ser</u> utilizado, mas não é essa a principal finalidade do R^2.

Gabarito oficial - letra C.

25. Sejam X e Y duas variáveis aleatórias quaisquer. Então:

(A) VAR (X – Y) = VAR (X) – VAR (Y).

(B) VAR (X – Y) = VAR (X) + VAR (Y) – COV (X, Y).

(C) VAR (X – Y) = VAR (X) + VAR (Y) – 2 COV (X, Y).

(D) VAR (X – Y) = VAR (X) + VAR (Y) + COV (X, Y).

(E) VAR (X – Y) = VAR (X) + VAR (Y) + 2 COV (X, Y).

RESOLUÇÃO:

Questão fácil, bastando o conhecimento das propriedades da variância e, no entanto, a FGV errou na divulgação do gabarito preliminar, pois a variância da soma ou diferença entre 2 variáveis será a soma das variâncias, acrescida ou subtraída do dobro da covariância, ou seja:

VAR (X ± Y) = VAR (X) + VAR (Y) ± 2 COV (X, Y).

Para VAR (X + Y) teremos: VAR (X + Y) = VAR (X) + VAR (Y) + 2 COV (X, Y).

Para VAR (X – Y) teremos: **VAR (X – Y) = VAR (X) + VAR (Y) – 2 COV (X, Y)**.

O gabarito preliminar apontou, incorretamente, para a opção da letra E, que seria a opção correta se fosse pedida a variância da soma, e não da diferença.

Gabarito preliminar - letra E.

Gabarito oficial definitivo - letra C.

26. Sejam X, Y e Z três variáveis com correlações de Pearson expressas pela matriz abaixo:

	X	Y	Z
X	1,000		
Y	0,800	1,000	
Z	0,000	–0,500	1,000

Pode-se, então, afirmar que:

(A) X e Z são independentes.

(B) a correlação parcial entre X e Y, após a correção para Z, é negativa.

(C) o coeficiente de determinação da regressão de Y em X é maior do que 60%.

(D) a correlação entre $V = a + b \cdot X$ e $W = c + d \cdot Z$, com $a \neq 0$, $c \neq 0$, $b > 0$ e $d < 0$ é negativa.

(E) a covariância entre X e Y é igual a 0,64.

RESOLUÇÃO:

Basta observar, na tabela, que a correlação entre X e Y é igual a 0,8.

O coeficiente de determinação da regressão de Y em X será o quadrado desse valor, ou seja:

$R^2 = (0,8)^2 = 0,64 = 64\%$, sendo, portanto, maior do que 60%, conforme a opção C.

A opção da letra A não é válida porque o fato de X e Z terem correlação igual a zero não implica, necessariamente, que X e Z sejam independentes, podem ser ou não. Já quando sabemos que duas variáveis são independentes, aí sim, podemos afirmar que a correlação entre elas será, com certeza, igual a zero;

A letra D está errada porque, se a correlação entre X e Z é nula, a correlação entre as variáveis V e W não será negativa, também será nula;

A afirmativa da letra E não pode ser verificada sem o conhecimento do desvio padrão de X e do desvio padrão de Y, pois a correlação é dada por $\rho_{xy} = \dfrac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}$. Portanto $\text{cov}(x,y) = \rho_{xy} \cdot \sigma_x \cdot \sigma_y$. No enunciado da questão é fornecido, apenas, o valor do coeficiente de correlação, ρ_{xy} = 0,8, não sendo fornecidos os valores dos desvios de X e de Y.

Gabarito oficial - letra C.

29. Em um país, a probabilidade de um contribuinte cometer erro na declaração anual de ajuste de rendimentos aumenta na medida em que o valor do imposto final também aumenta. Estudos indicam que a probabilidade de um contribuinte cometer erro na declaração anual de ajuste (Y = 1) é expressa por meio de:

$$P(Y = 1 | X) = \frac{e^{-0,048+0,02X}}{1+e^{-0,048+0,02X}},$$

onde X é um número real que representa o valor do ajuste do imposto (diferença entre o imposto pago ao longo do ano e o que deveria pagar de acordo com os rendimentos, retenções e abatimentos), em $1.000.

Se X > 0, o contribuinte tem imposto devido a pagar; se X < 0, tem imposto a ser restituído; e, se X = 0, o imposto retido ao longo do ano foi igual ao imposto total devido.

A esse respeito, é correto afirmar que:

(A) a cada acréscimo de $1.000 no imposto, a probabilidade de o contribuinte cometer erro na declaração de ajuste aumenta em 2%.

(B) a probabilidade de a declaração de ajuste apresentar erro (Y = 1) é maior do que a probabilidade de não haver erro (Y = 0), para todos os contribuintes com X > 0.

(C) essa função de probabilidade tem seu ponto de inflexão em X = 0.

(D) o logaritmo neperiano da razão entre a probabilidade de haver erro na declaração e a de não haver é uma função linear em X, expressa por $-0,048 + 0,02 \cdot X$.

(E) contribuintes com imposto devido têm probabilidade 0,5 de cometer erro na declaração.

RESOLUÇÃO:

A função, que é apresentada na questão, é a função LOGIT.

Para facilitar, vamos começar fazendo a seguinte transformação na função: $Z = -0,048 + 0,02X$.

Logo, a função fica: $P(Y = 1 | X) = \dfrac{e^Z}{1+e^Z}$.

A opção de resposta da letra A, pode ser rapidamente descartada, pois só seria verdadeira se tivéssemos uma função linear, o que não é o caso.

Também podemos descartar a opção da letra C, já que o ponto de inflexão da função não será zero e sim 2,4, pois: $-0,048 + 0,02X \geq 0 \Rightarrow 0,02X \geq 0,048 \Rightarrow X \geq 2,4$.

Idem para a opção da letra E, pois a probabilidade depende de X e a função não é uma função constante.

Para chegarmos à opção de resposta, vamos analisar que:

A probabilidade de haver erro, P(Y = 1), será dada por P_i;

Logo, a probabilidade de não haver erro, P(Y = 0), será dada por 1 – P_i, onde $P_i = \frac{e^{-0{,}048+0{,}02X}}{1+e^{-0{,}048+0{,}02X}}$;

Mas como fizemos a transformação Z = –0,048 + 0,02X fica:

$\boxed{P_i = \frac{e^Z}{1+e^Z}}$, que é a probabilidade de haver erro, P(Y = 1).

A probabilidade de não haver erro, P(Y = 0), dada por 1 – P_i, será:

$$1-P_i = 1-\frac{e^Z}{1+e^Z} = \frac{1+e^Z-e^Z}{1+e^Z} \Rightarrow \boxed{1-P_i = \frac{1}{1+e^Z}}$$

Se Z = 0, então $P_i = \frac{e^0}{1+e^0} = \frac{1}{2}$ = 0,5 (probabilidade de haver erro) e (1 – P_i) também será igual a 0,5;

A probabilidade de haver erro (Y = 1) começará a ser maior do que a probabilidade de não haver erro (Y = 0) a partir de Z > 0. Mas, como já vimos, para que X seja maior do que zero, Z deverá ser maior do que 2,4 (que é o ponto de inflexão da curva). Logo, a opção da letra B também está errada, pois não é com X > 0 e sim com X > 2,4 que P(Y = 1) > P(Y = 0).

Assim, temos a seguinte razão entre a probabilidade de haver erro e a de não haver erro:

$$\frac{P(Y=1)}{P(Y=0)} = \frac{P_i}{1-P_i} = \frac{\frac{e^Z}{1+e^Z}}{\frac{1}{1+e^Z}} = \boxed{e^Z}$$

Portanto, a razão é e^Z e o logaritmo neperiano da razão será: $\ln e^Z = \log_e e^Z$.

Pelas propriedades dos logaritmos, o logaritmo da potência fica sendo: $Z \cdot \log_e e$.

Mas $\log_e e = 1$ (o logaritmo de mesma base é unitário).

Logo, o logaritmo neperiano da razão é igual a Z, ou seja, $\boxed{\ln e^Z = Z}$

Mas Z = –0,048 + 0,02X, que é uma função linear em X, como afirma a opção de resposta da letra D.

Gabarito oficial - letra D.

GABARITOS:

19. D	20. C	21. D	22. B	23. A	24. C	25. C	26. C	29. D