



SÃO PAULO TURISMO S.A.
PROCESSO SELETIVO Nº 001/2007



Cód 36 – Estatístico

1. Os dados estatísticos constituem a matéria-prima das pesquisas estatísticas. São utilizados quatro níveis de medida na sua obtenção: nominal, ordinal, intervalar e racional. Variáveis medidas no nível ordinal permitem:
 - A) categorizar e ordenar os dados, subtrair valores dos dados.
 - B) categorizar e ordenar os dados, determinar se o valor de um dado é múltiplo de outro.
 - C) categorizar e ordenar os dados.
 - D) categorizar e ordenar os dados, subtrair valores dos dados e determinar se o valor de um dado é múltiplo de outro.

2. As variáveis aleatórias podem ser divididas em dois grandes grupos: as discretas e as contínuas. Exemplos de variáveis aleatórias contínuas são aquelas que seguem distribuições:
 - A) normal, Poisson e uniforme.
 - B) uniforme, normal e exponencial.
 - C) binomial, normal e uniforme.
 - D) exponencial e binomial.

3. Estatística clássica se refere à análise de dados amostrais, excluindo qualquer juízo ou opinião pessoal. Esta se diferencia da análise *bayesiana* de decisão. A análise *bayesiana* de decisão:
 - A) incorpora na análise estatística o uso de juízo, com ênfase especial em possíveis ganhos ou perdas econômicas associadas a decisões alternativas.
 - B) incorpora, além das técnicas de inferência, o uso conjunto de gráficos, excluindo qualquer juízo ou opinião pessoal.
 - C) é utilizada apenas em levantamentos censitários.
 - D) é utilizada apenas em levantamentos censitários, com ênfase especial apenas nos ganhos econômicos associados a decisões alternativas.

4. De acordo com o IDB Brasil 2000, a taxa de crescimento da população é obtida pelo método:
 - A) geométrico. Em termos práticos, subtrai-se 1 da raiz enésima da população final dividida pela população no começo do período considerado, multiplicando-se o resultado por 100.
 - B) aritmético. Em termos práticos, divide-se por n a diferença entre a população final e a do começo do período considerado.
 - C) aritmético. Em termos práticos, divide-se por n a diferença entre a população final e a do começo do período considerado, multiplicando-se o resultado por 100.
 - D) geométrico. Em termos práticos, é a raiz enésima da população final dividida pela população no começo do período considerado, multiplicando-se o resultado por 100.

5. Uma amostra constituída por elementos da população que têm a mesma probabilidade de serem selecionados para a amostra constitui o que se chama de amostra:
 - A) sistemática.
 - B) estratificada simples.
 - C) aleatória múltipla ou conglomerada.
 - D) aleatória simples.

6. Um dos resultados do aumento do tamanho da amostra é:
 - A) a diminuição do poder dos testes estatísticos, pois a amostra maior representa melhor a população.
 - B) a perda de consistência da estatística calculada.
 - C) a redução do erro tipo II nos testes estatísticos.
 - D) a não necessidade de utilização de testes de aderência.

7. As técnicas de amostragem denominadas amostragem acidental, amostragem intencional e amostragem por quotas, constituem métodos:
- A) probabilísticos de amostragem.
 - B) não probabilísticos de amostragem.
 - C) estocásticos de amostragem.
 - D) que garantem a representatividade da população.
8. No cálculo de tamanho de amostra para estimar a média de uma população infinita utiliza-se a fórmula $n = (Z \cdot s / d)^2$. A variável d da fórmula é:
- A) o nível de significância adotado no cálculo.
 - B) a mínima diferença que o pesquisador admite suportar entre o nível de significância e o erro escolhido.
 - C) a máxima diferença que o pesquisador admite suportar entre o escolhido e aquele que realmente permitiria estimar a média.
 - D) o erro amostral expresso na unidade de medida da variável.
9. A ogiva ou gráfico de porcentagens acumuladas de uma distribuição de freqüências dá uma noção da:
- A) fdp empírica dos dados.
 - B) função geratriz de momentos.
 - C) probabilidade de falhas nos dados.
 - D) função distribuição empírica dos dados.
10. Uma distribuição de freqüências que se apresente pontiaguda, com as observações concentradas em um pequeno intervalo de valores, pode ser classificada como:
- A) mesocúrtica.
 - B) leptocúrtica.
 - C) contínua.
 - D) centróide.
11. Considerando as estatísticas média (Me), moda (Mo) e mediana (Md), obtidas numa distribuição de freqüências, com ordem de grandeza $Me < Md < Mo$, pode-se afirmar que a referida distribuição de freqüências:
- A) apresenta assimetria positiva.
 - B) é simétrica.
 - C) apresenta assimetria negativa.
 - D) é platicúrtica.
12. Uma variável apresenta os seguintes valores: 20, 21, 19, 17, 20, 17, 18, 16, 18, 13, 22, 18. Pode-se afirmar que:
- A) a distribuição tem três modas, cujos valores são 17, 18 e 20.
 - B) a distribuição é amodal, pois não se pode caracterizar um único valor com freqüência superior a 1.
 - C) a distribuição tem apenas uma moda, cujo valor é 18.
 - D) a distribuição é bimodal, com dois valores, 17 e 20, que se repetem.
13. A representação em gráficos de pontos é bastante útil, principalmente quando se deseja ajustar modelos de regressão. Os gráficos de pontos também são conhecidos como gráficos:
- A) de dispersão XY.
 - B) de colunas ou barras.
 - C) de ramos e folhas.
 - D) auxiliares de modelos XY.
14. Mais recentemente diversas técnicas empregadas na análise exploratória de dados foram introduzidas. Uma muito difundida e utilizada é o diagrama de caixa, ou blox-plot, como é mais conhecida. Algumas informações são apresentadas no diagrama: a mediana, os quartis, além de indicações de valores que podem se caracterizar como:
- A) dependentes.
 - B) out-liers.
 - C) outside.
 - D) padronizados.

15. Outra técnica muito empregada em estatística descritiva é o diagrama ramo-e-folhas, constituindo-se em técnica alternativa, principalmente utilizada:
- no cálculo do número de classes.
 - na obtenção dos limites inferiores das classes.
 - no cálculo do número de classes quando a amostra é grande.
 - no resumo ou sistematização de um conjunto de dados.
16. Os gastos médios de dois grupos de pessoas, denominados A e B, são, respectivamente, R\$ 2.000,00 e R\$ 3.000,00. O desvio-padrão dos gastos do grupo A é de R\$ 450,00, assim como o do grupo B é de R\$ 450,00. Relativamente ao valor médio, o grupo de maior variabilidade:
- é o B.
 - não pode ser determinado, visto não se conhecer o número de componentes de cada grupo.
 - é o A.
 - não pode ser determinado, pois são exatamente iguais.
17. A probabilidade de que uma pessoa moradora de uma comunidade seja portadora de uma determinada doença é de 10%. Uma vez por semana a comunidade recebe a visita de um médico que atende a um grupo de 10 pessoas. A média e a variância do número de portadores da doença por grupo assistido são, respectivamente:
- 1 e 0,9.
 - 1 e 1.
 - 0,1 e 0,09.
 - 1 e 9.
18. Uma variável aleatória segue distribuição uniforme. A probabilidade de ocorrência de um valor menor ou igual a 120 é de 80%, enquanto que o de um valor maior ou igual a 130 é de 10%. A variância dessa variável é:
- 10.000.
 - 833,33.
 - 100.
 - 8,33
19. O valor médio de uma variável que se distribui de acordo com uma distribuição de Poisson é 4. A variância dessa variável é igual a:
- e^{-4} .
 - $e^{-4} / 4!$.
 - 24.
 - 4.
20. Um pesquisador tomou as estaturas, em cm, de um grupo de 100 pessoas es colhidas aleatoriamente. Ao calcular a média e a variância das cem observações chegou aos valores 160 e 100. Ao finalizar o relatório do trabalho efetuado, o pesquisador notou que deveria apresentar os valores das estatísticas calculadas na unidade m e não em cm. Dessa forma, no relatório ele deverá escrever:
- média = 1,6 m e variância = 1 m.
 - média = 1,6 e variância = 1.
 - média = 1,6 m e variância = 0,01 m².
 - média = 1,6 e variância = 0,1.
21. Um cruzamento de duas ruas é estudado pelo Departamento de Trânsito que anota o número de carros que o cruzam num intervalo de um minuto. Durante a manhã de um sábado foram efetuados 11 experimentos de 1 minuto cada, tendo sido anotados os seguintes números de carros por minuto: 4, 6, 3, 6, 8, 3, 4, 6, 7, 1, 7. Pode-se afirmar que:
- a distribuição desses dados é desconhecida, porém tem média igual à da mediana.
 - a amostra tem mais de uma moda.
 - os valores da média, da moda e da mediana desses dados são, respectivamente, 5, 6 e 6.
 - a amostra tem três modas: 3, 4 e 7, sendo a mediana igual a 6.
22. Uma empresa concedeu 5% de aumento de salário a todos os seus funcionários. O desvio-padrão dos salários, antes do aumento, era de R\$ 300,00. A variância dos novos salários será igual a:
- 99.225.
 - 300.
 - 90.000.
 - 315.

23. Uma empresa resolveu aumentar o salário de todos os seus funcionários em R\$ 150,00 por mês. Em relação ao mês anterior:
- o salário médio sobe R\$ 150,00, enquanto o desvio-padrão dos salários diminui, pois a média aumenta.
 - tanto o salário médio quanto o desvio-padrão dos salários aumentam.
 - o salário médio sobe R\$ 150,00, enquanto o desvio-padrão dos salários permanece igual.
 - o coeficiente de variação não se altera.
24. O coeficiente de correlação calculado para 20 pares de valores X e Y resultou no valor 0,64. Em seguida todos os valores de Y foram divididos por 4 e novamente foi calculado o coeficiente de correlação, resultando em:
- 0,32.
 - 0,64.
 - 0,16.
 - valor indeterminado, visto não serem conhecidos os 20 valores de X e Y.
25. A covariância entre duas variáveis com variâncias 160 e 1.000 é igual a -160. É possível afirmar que:
- há um erro na informação, pois duas variáveis com variâncias positivas não podem ter covariância negativa.
 - há um erro na informação, pois esses valores caracterizam problema de heterocedasticidade, não sendo possível, portanto, obter covariância negativa.
 - a variância da segunda variável, 1.000, não foi bem estimada, devendo sofrer a correção de Winston, dada por: $s^2 - \sqrt{x}/n$, onde n é o número de dados utilizados nos cálculos da variância e covariância.
 - o coeficiente de correlação entre as duas variáveis é igual a -0,4.
26. Os valores obtidos para pares de valores X e Y permitiu o ajuste de um modelo de regressão polinomial do segundo grau, com $r^2 = 0,9025$. Pode-se afirmar, com certeza, que o valor do coeficiente de correlação entre X e Y é:
- muito próximo de 0.
 - igual a 0,95.
 - igual a 0,9025.
 - igual a 0,0975.

27. Minimizar $S = \sum_{i=1}^n [y_i - (a + b.x_i)]^2$, derivando a função em relação a a e b é o princípio do método denominado:

- análise de variância.
- teste do qui-quadrado.
- mínimos quadrados.
- binômio de Newton.

28. O valor do coeficiente de correlação entre as variáveis:

X	10	20	30	40	50	60	70
Y	6	12	17	24	31	35	42

- 0,0046.
- 0,8900.
- 0,9986.
- 1,0000.

29. O valor calculado de b num ajuste de regressão linear do tipo $\hat{y} = a + b.x$, é igual a 5. Sabendo que a média dos valores x é igual a 20 e que a média dos valores y é 250, o valor estimado será:

- 230.
- 150.
- 350.
- 0.

30. Uma regressão linear $\hat{y} = a + b.x$ foi estimada para um conjunto de 15 pares de dados. A variação total calculada, dada pela soma dos quadrados das variações de y em relação à sua média \bar{y} , é igual a 500. A variação explicada calculada, dada pela soma dos quadrados das variações de \hat{y} em relação à sua média $\bar{\hat{y}}$, foi igual a 240. Sabe-se que o valor de t com 13 graus de liberdade e nível de significância de 5%, bilateral, é igual a 2,16. Com esses dados pode-se afirmar que:
- a regressão é significativa no nível de significância de 5%.
 - y é independente de x .
 - a linha de regressão passa pela origem.
 - a regressão é não significativa no nível de significância de 5%.
31. Uma imobiliária está desenvolvendo um modelo de regressão que tem como variável dependente o aluguel mensal do imóvel e como independentes o número de quartos, de banheiros, a área do imóvel e sua localização, expressa pelo bairro, tendo (a localização) valor aditivo. Supondo que a região de atendimento da imobiliária seja composta por 3 bairros e que será ajustado um único modelo, o número de variáveis *dummys* a utilizar será:
- três.
 - uma.
 - nenhuma, pois será utilizado o nome do bairro.
 - duas.
32. Sazonalidade é um fator estudado em séries temporais. A sazonalidade explica:
- as variações ou movimentos aleatórios durante certo período de tempo.
 - as variações impróprias que denunciam, por si, erros nas coletas de dados.
 - as variações ou movimentos cíclicos durante certo período de tempo.
 - os períodos sem nenhuma variabilidade de valores da série temporal.
33. Média móvel de ordem k de uma série ordenada no tempo é:
- a mediana de k elementos sucessivos da série, atribuída ao elemento final dos k elementos utilizados no cálculo.
 - a média de k elementos sucessivos da série, atribuída ao elemento central dos k elementos utilizados no cálculo.
 - a média de k elementos sucessivos da série, atribuída ao elemento inicial dos k elementos utilizados no cálculo.
 - a média harmônica de k elementos sucessivos da série, atribuída ao elemento inicial dos k elementos utilizados no cálculo.
34. A probabilidade de não rejeitar a hipótese nula, quando a hipótese nula é falsa, é chamada de:
- poder do teste.
 - $1-\beta$.
 - erro tipo II.
 - nível de significância.
35. _____ é a diferença consistente e persistente da estatística em relação ao _____ que se deseja estimar.
- Tendência, parâmetro.
 - Variância mínima, valor.
 - Consistência, desvio-padrão.
 - Mínimos quadrados, estimador.
36. Uma variável aleatória se comporta de acordo com uma distribuição normal de média 500 e desvio-padrão 100. Se 10.000 determinações dessa variável forem coletadas e classificadas em ordem ascendente, formando um vetor, é de se esperar que o elemento de ordem 1.600^{a} desse vetor de dados seja:
- aproximadamente igual a 304.
 - maior que 500.
 - aproximadamente igual a 696.
 - aproximadamente igual a 400.

37. Uma amostra de 144 indivíduos foi tomada aleatoriamente, obtendo-se média 200 e desvio-padrão 48. Pode-se afirmar que:
- A) $P(192,16 = X = 207,84) = 0,95$.
 B) $P(192,16 = \mu = 207,84) = 0,95$.
 C) $P(104 = \mu = 296) = 0,95$.
 D) $P(152 < X < 248) = 0,99$.
38. O desvio-padrão de uma população é conhecido e igual a 60. Foi tomada uma amostra de 25 elementos, tendo sido obtido o valor 180 para a média. Sabendo que $P(z = 1,645) = 0,05$, que $P(z = 1,96) = 0,025$, pode-se afirmar que:
- A) a média populacional é inferior a 196, no nível de 2,5% de significância.
 B) a média populacional é superior a 150, no nível de 5 % de significância.
 C) a média populacional é inferior a 120, no nível de 2,5% de significância.
 D) a população se distribui como normal.
39. Numa linha de produção, dez por cento das peças produzidas apresentam algum tipo de defeito. A empresa comercializa sua produção em caixas contendo 3 peças. A probabilidade de que uma caixa escolhida ao acaso não apresente peças defeituosas é igual a:
- A) 70,0%.
 B) 90,0%.
 C) 72,9%.
 D) Aproximadamente 96,7%.
40. A probabilidade de que uma variável com distribuição normal assumira valores maiores que 100 é de 69,15%. A probabilidade de que assumira valores menores que 190 é de 97,72%. A média e a variância dessa variável são, respectivamente:
- Dados:
- | Z | P(z = Z) |
|-----|----------|
| 0,0 | 0,5000 |
| 0,5 | 0,3085 |
| 1,0 | 0,1587 |
| 1,5 | 0,0668 |
| 2,0 | 0,0228 |
| 2,5 | 0,0062 |
- A) 118 e 1.296.
 B) 90 e 60.
 C) 70 e 1.600.
 D) 90 e 3.600.
41. O número de graus de liberdade para o teste de qui-quadrado numa tabela de contingência 2 x 2 é igual a:
- A) 2.
 B) 3.
 C) 4, pois os valores da tabela são populacionais.
 D) 1.
42. O cálculo das freqüências esperadas nos testes para verificação de associação ou dependência entre duas variáveis, tem por fundamento que:
- A) se X e Y são dependentes então a distribuição conjunta de probabilidades (X,Y) é igual ao produto das distribuições marginais de probabilidades X e Y.
 B) se X e Y são independentes então a distribuição conjunta de probabilidades (X,Y) é igual ao produto das distribuições marginais de probabilidades X e Y.
 C) se X e Y são independentes então a distribuição conjunta de probabilidades (X,Y) é igual ao produto das distribuições integrais (levam em conta a covariância positiva) de probabilidades X e Y.
 D) se X e Y são independentes então a distribuição conjunta de probabilidades (X,Y) é igual ao produto das distribuições marginais de probabilidades X e Y, adicionado à covariância, necessariamente não nula e positiva, entre X e Y.

43. A correção de Yates é utilizada no cálculo do qui-quadrado para tabelas 2 x 2 quando:
- as freqüências esperadas são muito grandes.
 - as freqüências observadas são muito pequenas.
 - as freqüências esperadas são muito pequenas.
 - as freqüências observadas são muito grandes.
44. Mensalmente uma instituição bastante conhecida divulga um dos índices mais importantes da economia brasileira, o IGP. Ele é um índice composto pelas contribuições dos índices de preços por atacado, de construção civil e de preços ao consumidor, utilizando, respectivamente, os pesos:
- 40%, 40% e 20%.
 - 33,4%, 33,3% e 33,3%.
 - 60%, 10% e 30%.
 - 30%, 35% e 35%.
45. Para estimar a evolução da renda real do consumidor num período podem ser utilizados os índices de Laspeyres e de Paasche. Porém, ocorrem problemas:
- o de Laspeyres subavalia e o de Paasche superavalia a renda.
 - o de Laspeyres superavalia e o de Paasche subavalia a renda.
 - o de Laspeyres superavalia a renda, enquanto que o de Paasche é indiferente.
 - o de Laspeyres avalia exatamente e o de Paasche superavalia a renda.
46. Num conceito amplo, número índice é a metodologia estatística idealizada para _____, quantitativamente, as _____ de um fenômeno complexo _____.
- comparar, variações, da população regional.
 - entender, igualdades, global.
 - entender, práticas, econômico.
 - comparar, variações, no tempo.
47. Na operação da planilha Excel com a intenção de ajustar modelos de regressão pode ser utilizada a ferramenta de análise regressão. Se a ferramenta for utilizada para ajustar um modelo de regressão polinomial de terceiro grau, do tipo $y = a + b.X + c.X^2 + d.X^3$, será necessário:
- informar no intervalo X de entrada, o intervalo de células da planilha que conterá, além dos valores da variável X, os valores de X^2 e X^3 .
 - informar no intervalo X de entrada, o intervalo de células da planilha que contém apenas os valores da variável X.
 - informar no intervalo X de entrada, o intervalo de células da planilha que contém os valores da variável X e o grau do polinômio a ser ajustado, no caso 3.
 - informar o grau do polinômio a ser ajustado, no caso 3, pois o Excel localiza automaticamente os valores da variável X.
48. As funções Desvpad(núm1; núm2; ...; númn) e Desvpadp(núm1; núm2; ...; númn) do Excel retornam, respectivamente:
- os desvios padronizados da média da amostra de valores e o desvios padronizados da média da população dos valores.
 - O desvio-padrão da amostra de valores e o desvio-padrão da média da amostra de valores.
 - o desvio-padrão da amostra de valores e o desvio-padrão da população dos valores.
 - O desvio-padrão da amostra de valores e os desvios padronizados da média da população dos valores.
49. As estatísticas e medidas de associação do procedimento Crosstabs do SPSS:
- são calculadas para tabelas de dupla ou múltipla entrada.
 - oferece 22 testes e medidas de associação para tabelas de múltipla (mais de duas) entrada.
 - são calculadas apenas em tabelas de dupla entrada.
 - podem ser utilizadas somente quando há pelo menos três grupos -controle.
50. A limpeza dos dados pode mostrar valores incomuns, extremos, descontinuidade ou outras peculiaridades. Pode também ajudar a determinar se a técnica estatística considerada para análise é apropriada. Pode ainda indicar a necessidade de transformação de dados. O procedimento mais indicado para limpeza de dados do pacote SPSS é o:
- Descriptive Statistics ... Descriptives.
 - Descriptive Statistics ... Explore.
 - Descriptive Statistics ... Frequencies.
 - Descriptive Statistics ... Clearing.