-- CONHECIMENTOS ESPECÍFICOS --

A tabela de frequência a seguir mostra dados coletados em uma pesquisa para se verificar o número de disciplinas que os estudantes de determinada universidade estão cursando por semestre.

disciplinas	2	3	4	5	6	7	8
estudantes	10	15	40	35	28	10	4

Considerando essas informações, julgue os itens seguintes.

51 Na pesquisa foram entrevistados 142 alunos.

52 Em média, os alunos cursam entre 4 e 5 disciplinas por semestre.

53 Os dados na tabela apresentam uma distribuição assimétrica.

54 Como os dados na tabela são discretos, eles podem ser modelados pela distribuição normal, que é uma distribuição discreta.

55 A proporção de alunos que cursam mais de 6 disciplinas é maior que a proporção de alunos que cursam 3 disciplinas.

56 O gráfico do tipo *pizza* é o mais apropriado para representar os dados apresentados na tabela, visto que a variável analisada é qualitativa ordinal.

57 Sabendo-se que a variância amostral dos dados é igual a 2, conclui-se que o coeficiente de variação é maior que 50%.

Considere que, em determinado curso de uma universidade, as notas dos alunos seguiu uma distribuição normal com média 4,5 e variância 9, e assuma que:

- $P(Z > 0) = 0{,}5$;
- $P(Z > 0{,}84) = 0{,}2$;
- $P(Z > 1{,}28) = 0{,}1$;
- $P(Z > 1{,}645) = 0{,}05$;
- $P(Z > 1{,}96) = 0{,}025$;
- $P(Z > 2{,}33) = 0{,}01$; e
- $P(Z > 2{,}575) = 0{,}005$.

Com base nessa situação, julgue os próximos itens.

58 A probabilidade de, em uma amostra aleatória, 9 alunos conseguirem média entre 5,34 e 7,075 é inferior a 20%.

59 A probabilidade de um aluno ter nota exatamente igual a 4,5 é superior ou igual a 50%.

60 Se apenas os 5% melhores alunos poderão concorrer a uma bolsa de iniciação científica no próximo semestre, então o aluno que pretenda concorrer a essa bolsa deve obter nota superior a 9,5.

61 Se os 20% piores alunos deverão refazer a disciplina no próximo semestre, então o aluno que tenha obtido nota inferior a 2 deverá refazer a disciplina.

62 Se uma amostra aleatória de 10 alunos for retirada, sem reposição, então a variância da média será inferior a 1.

63 Se a população é formada por 50 alunos, para calcular a distribuição amostral real, considerando uma amostra de 10 alunos, sem reposição, seriam necessários mais de 10 bilhões de amostras possíveis.

64 A probabilidade de um aluno qualquer conseguir nota superior a 8 é inferior a 10%.

65 Se, para ser aprovado, um aluno precisa de uma nota igual ou superior a 5, então a probabilidade de um aluno ser aprovado é superior a 50%.

Uma universidade está fazendo um estudo para verificar a distribuição dos tempos que os alunos do curso de mestrado levam até a defesa da dissertação. Os dados a seguir mostram a função de probabilidade desses tempos, em meses.

tempo de defesa (meses)	probabilidade
12	0,01
15	0,02
18	0,04
20	0,10
22	0,22
24	0,31
25	0,18
26	0,04
28	0,03
30	0,05

Considerando essas informações, julgue os itens subsequentes.

66 Os dados referentes ao tempo de defesa têm mediana igual a 24 meses.

67 Em média, os alunos levam mais de 24 meses para concluir o mestrado.

68 Se o prazo máximo recomendado para a defesa da dissertação de mestrado é de 24 meses, então a probabilidade de um aluno defender sua dissertação até 2 meses antes desse prazo é igual à probabilidade de um aluno defendê-la até 2 meses depois.

69 Assumindo-se que $E(X^2) = 552$, obtém-se um valor superior a 5 para o desvio padrão dos dados referentes ao tempo de defesa.

70 Se o prazo máximo de defesa recomendado é de 24 meses, então a probabilidade de um aluno defender sua dissertação no prazo é superior a 70%.

71 Caso, a partir dos dados em tela, fosse feita uma aproximação pela distribuição normal, então a probabilidade de um aluno defender sua dissertação em um prazo igual ou superior a 24 meses seria superior ao calculado pela distribuição original apresentada. Assuma que: $P(Z > 0) = 0{,}5$, $P(Z > 0{,}06) = 0{,}476$, $P(Z > 0{,}23) = 0{,}409$ e $P(Z > 0{,}4) = 0{,}3446$.

72 O gráfico de setores é adequado para representar a distribuição em questão.

73 Os valores da probabilidade de um aluno defender a dissertação em 13, 14, 16, 19, 21, 23, 27 ou 29 meses, somados, é igual à probabilidade de um aluno defender a dissertação em exatamente 31 meses.

No que diz respeito à inferência estatística, julgue os itens a seguir.

74 Na inferência estatística, o viés é um importante critério de julgamento de estimadores; um estimador que, em média, fornece a resposta correta é denominado estimador não enviesado.

75 Sabendo-se que a população brasileira, segundo o IBGE, soma 213,3 milhões de habitantes e que a população estimada de determinado município é de 42.561 ± 236 habitantes, então este é um exemplo de estimativa pontual, pois apresenta um valor exato do parâmetro de um único município.

76 Em uma amostra de tamanho 36, cujo valor da média amostral seja igual a 60, é possível que o valor real da média, com 95% de confiança, seja igual a 49, caso o desvio padrão populacional seja igual a 25.

77 Considere-se que tenham sido coletadas cinco amostras aleatórias do volume de vacina contra a covid-19 aplicadas em pacientes e que tenham sido obtidos, em mL, os valores 0,49; 0,44; 0,51; 0,52 e 0,46. Nesse caso, 0,490 mL é uma estimativa não tendenciosa e eficiente para a média do volume de vacina aplicada em pacientes.

78 Considere-se que, em uma pesquisa, a média amostral seja de 75,0 e que o erro padrão da média amostral — obtido com base no valor do desvio padrão populacional — seja de 4,0. Nesse caso, se o intervalo de confiança para a média amostral possui limite inferior de 67,16 e limite superior de 82,84, então o nível de confiança será de 95%.

Tendo como referência os testes de hipóteses, que são ferramentas auxiliares nas tomadas de decisão acerca de uma ou mais populações com base nas informações obtidas da amostra, julgue os itens seguintes.

79 Considere-se que, para analisar se uma moeda é viciada, ou não, defina-se como hipótese nula que a referida moeda é viciada. Nesse caso, para essa hipótese, deve-se admitir que a probabilidade p de sair cara em um lançamento da moeda será $p = 0,5$.

80 O poder do teste ou potência consiste na probabilidade de rejeitar a hipótese nula H_0, quando a hipótese alternativa H_1 for falsa.

81 Considere-se que, em um teste de hipótese para a análise do funcionamento de determinada máquina, seja admitido como hipótese nula o fato de a referida máquina estar funcionando perfeitamente. Nesse caso, se houver a ocorrência de um erro do tipo I, então a máquina não estará funcionando adequadamente.

82 Considerando-se, para certa hipótese, que a distribuição amostral de uma estatística S seja normal, com média μ_S e desvio padrão σ_S, então, caso se verifique, para a única amostra aleatória, que o escore z dessa estatística esteja fora do intervalo de –1,96 a 1,96, e se o tamanho desse teste bilateral for $\alpha = 5\%$, é correto concluir que z difere significativamente do que se pode esperar para essa hipótese, pois está fora da região de aceitação da hipótese.

Julgue os itens subsequentes, considerando oito pares de valores das variáveis X e Y, tais que $\Sigma X = 24$; $\Sigma Y = 49$; $\Sigma X \cdot Y = 181$; $\Sigma X^2 = 100$ e $\Sigma Y^2 = 343$.

83 Existe uma correlação forte entre as variáveis X e Y.

84 O coeficiente de correlação de Pearson para os valores apresentados será negativo, o que indica que a regressão linear será representada por uma reta decrescente.

85 A reta dos mínimos quadrados ordinários que representa a regressão linear simples de Y em X com intercepto não nulo terá coeficiente linear aproximado de 2,48.

86 Com base no coeficiente de correlação linear, é correto afirmar, em face dos dados apresentados, que se trata de uma correlação espúria.

87 Se o par (x_i, y_i) for um dos oito pares ordenador das variáveis X e Y, ampliando-se o valor de x_i na reta dos mínimos quadrados ordinários que representa a regressão linear simples de Y em X, o valor de Y encontrado será tal que $Y = y_i$.

Para uma determinada amostra, observou-se um conjunto de n eventos E_n, cujas frequências observadas e esperadas são, respectivamente, $o_1, o_2, o_3, o_4, ..., o_n$, e $e_1, e_2, e_3, e_4, ..., e_n$.

Tendo como referência essas informações, julgue os próximos itens.

88 A medida denominada *qui-quadrado*, representada por x^2, define a discrepância existente entre as frequências observadas e esperadas.

89 Se, para 50 lançamentos de uma moeda, forem observadas 30 caras e 20 coroas, então o valor de x^2 será inferior a 3,0.

90 A utilização do *qui-quadrado* como teste de aderência objetiva analisar a adequação dos modelos teóricos de distribuição à distribuição amostral.

Julgue os itens que se seguem, em relação à análise de variância, técnica estatística utilizada para a comparação das médias de uma variável aleatória numérica em mais de duas populações.

91 A obtenção de uma razão F significante é condição para a utilização do teste DHS de Tukey para a comparação múltipla de médias.

92 A variável numérica cujas médias sejam comparadas em três ou mais populações é chamada de variável de tratamento, ou fator, ou variável explanatória, ou variável independente.

93 Para a utilização da razão F, definida como a divisão entre a média quadrática entre grupos (MQentre) e a média quadrática dentro de grupos (MQdentro), é necessário que as amostras sejam extraídas aleatoriamente de determinada população de escores.

Uma regressão linear de Y sobre X consiste em obter a equação de uma reta, ou uma função linear, como o modelo que irá melhor representar a relação entre as variáveis; a determinação dos parâmetros dessa reta é denominada ajustamento.

Considerando essas informações, julgue os seguintes itens.

94 Suponha-se que, em uma pesquisa, o coeficiente de correlação entre duas variáveis X e Y tenha gerado um valor para o coeficiente de correlação de Pearson de 0,9200. Nesse caso, considerando-se X a variável independente e Y a variável dependente, o percentual da variância de Y explicado por X será de 84,64%.

95 Para quaisquer valores das variáveis X e Y, a existência de um coeficiente de correlação diferente de zero é garantia para que haja uma relação entre X e Y.

96 Um coeficiente de determinação entre as variáveis X e Y de 95% implica necessariamente a obtenção de uma reta dos mínimos quadrados crescente, ou seja, em uma correlação positiva.

```
R Console
> require(tidyverse)
> idade  <- c( 20, 30, NA, 30, 40, 50)
> altura <- c(1.6,1.7,1.5,1.8,1.7,1.7)
> dados  <- data.frame(idade,altura)
> dados
  idade altura
1    20    1.6
2    30    1.7
3    NA    1.5
4    30    1.8
5    40    1.7
6    50    1.7
>
```

Considerando que um usuário inicie uma sessão R escrevendo um código na janela R Console conforme mostra a figura apresentada, julgue os itens subsequentes.

97 A aplicação do código `dados %>% filter(idade==30)` proporciona o resultado conforme mostra a figura a seguir.

```
R Console
> dados %>% filter(idade==30)
[1] 30 30
```

98 O código `write_csv(dados, "D:/dados.csv")` permite salvar o *data frame* dados em "`D:/dados.csv`".

99 Os códigos `dados %>% na.omit()` e `dados %>% drop_na()` proporcionam o mesmo resultado.

100 O código `dados %>% mutate(alt = altura*100)` retorna um *data frame* conforme mostra a figura que se segue.

```
R Console
> dados %>% mutate(alt = altura*100)
  idade altura  alt
1    20    1.6  160
2    30    1.7  170
3    NA    1.5  150
4    30    1.8  180
5    40    1.7  170
6    50    1.7  170
>
```

101 Os códigos `dados %>% arrange(desc(altura))` e `dados %>% order(altura, decreasing = TRUE)` retornam o mesmo resultado.

Com respeito aos comandos de repetição da linguagem R, julgue os itens que se seguem.

102 Ao final do *loop* `while` mostrado na figura a seguir, o valor da variável `x` será igual a `4`.

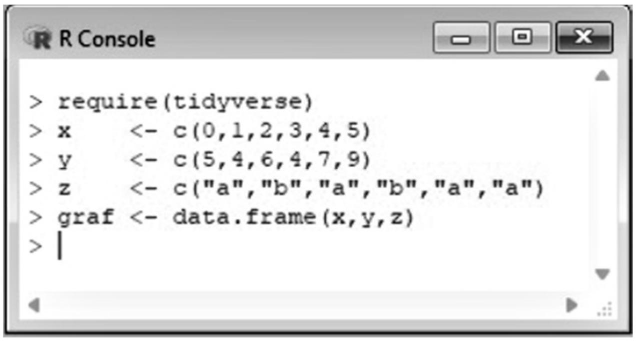
```
R Console
> x <- 1
> while (x < 5)
+ {
+ x = x+1
+ }
>
```

103 A variável `x`, que resulta da aplicação do *loop* `for` mostrado na figura que se segue, é igual à variável `y` obtida mediante aplicação do seguinte código: `y <- 1:5`.

```
R Console
> x <- NULL
> for (i in 1:5)
+ {
+ x[i] = i
+ }
>
```

104 Com respeito ao código mostrado na figura a seguir, é correto afirmar que a variável `x` resultante da aplicação do *loop* `repeat` é igual ao vetor `c(4,3,2,1,0)`.

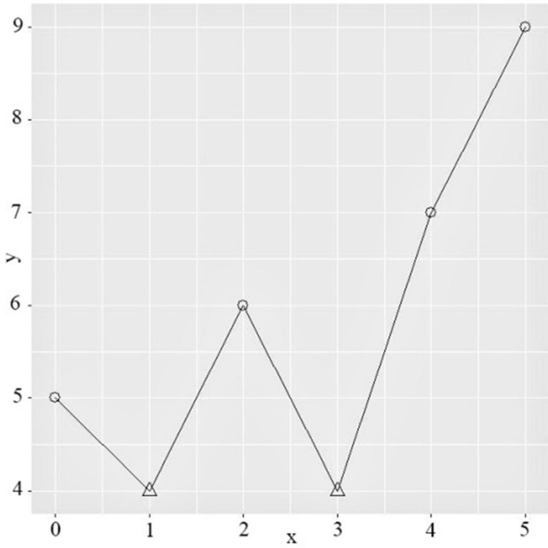
```
R Console
> x <- 4
> repeat {
+ print(x)
+ x = x-1
+ if (x < 0){
+ break
+ }
+ }
[1] 4
[1] 3
[1] 2
[1] 1
[1] 0
>
```



Considerando o código mostrado na figura apresentada, julgue os próximos itens.

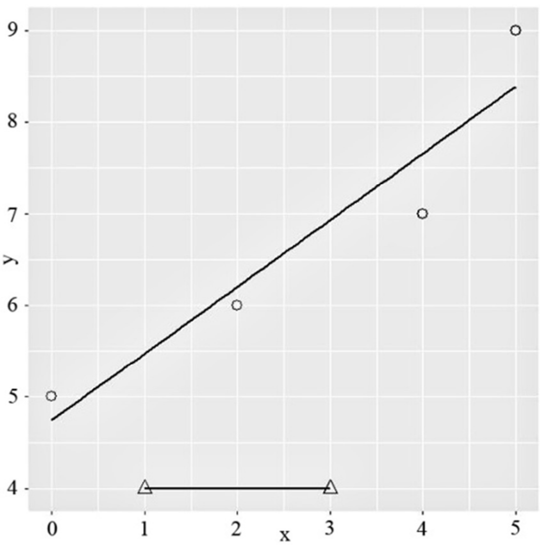
105 A figura a seguir é obtida mediante aplicação do seguinte código R:

```
ggplot(data=graf,mapping=aes(x=x, y=y))+
 geom_line()+
 geom_point()+
 xlab("x")+
 ylab("y")
```



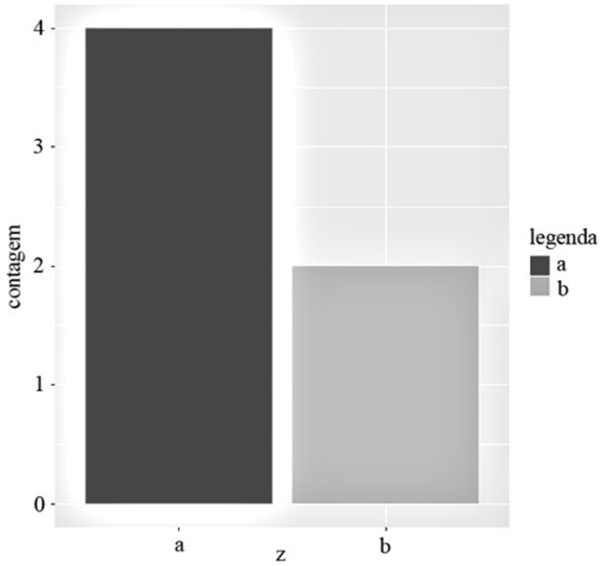
106 A figura subsequente é obtida mediante aplicação do seguinte código R:

```
ggplot(data=graf, mapping=aes(x=x,y=y))+
 geom_point(shape = factor(z),size=4) +
 geom_smooth(aes(group=z))+
 xlab("x")+
 ylab("y")
```



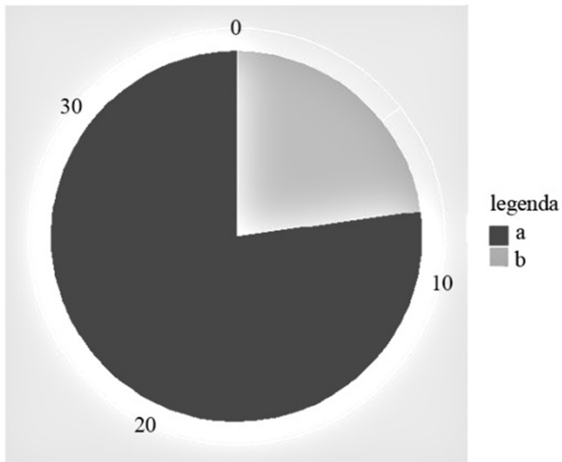
107 A figura que se segue é obtida mediante aplicação do seguinte código R:

```
ggplot(data=graf,mapping=aes(factor(z)))+
 geom_bar(aes(fill=factor(z)))+
 scale_fill_manual("legenda", values= c("a" =
"black", "b" = "gray"))+
 xlab("z")+
 ylab("contagem")
```



108 A figura subsecutiva é obtida mediante aplicação do seguinte código R:

```
ggplot(data=graf,mapping=aes(factor(z)))+
 geom_piechart(aes(fill=factor(z)))+
 scale_fill_manual("legenda", values= c("a" =
"black", "b" = "gray"))
```



Com respeito a dados abertos, julgue os seguintes itens.

109 Diz-se que os dados são abertos se qualquer pessoa tiver acesso a eles livremente, podendo utilizá-los, modificá-los e compartilhá-los sem necessidade de uma licença formal.

110 Os dados devem ser disponibilizados em um formato de arquivo não proprietário, como txt, csv e ods.

111 Entre os princípios que regem os dados abertos governamentais, encontra-se aquele que estabelece que os dados devem ser publicados conforme foram coletados da fonte e, preferencialmente, na forma não estruturada.

Um levantamento estatístico foi realizado entre os estudantes de graduação de três diferentes cursos no país para se estimar o percentual populacional P desses alunos que estavam otimistas quanto ao seu futuro profissional. Para isso, considerou-se que havia 12.000 estudantes matriculados nesses cursos no país na ocasião do levantamento. O quadro a seguir mostra a distribuição desses alunos conforme o curso de graduação.

curso de graduação	total de alunos
A	4.000
B	6.000
C	2.000

As quantidades de estudantes dos cursos A, B e C que participaram do levantamento bem como os respectivos percentuais de alunos otimistas observados nessas amostras e suas estimativas dos erros padrão encontram-se no seguinte quadro.

curso de graduação	tamanho da amostra	estimativa do percentual de alunos que estão otimistas quanto ao seu futuro profissional	erro padrão
A	200	80%	2,5%
B	100	65%	4,7%
C	100	95%	2,0%

A respeito dessa situação hipotética, julgue os itens subsecutivos.

112 A técnica descrita no texto para a estimação do percentual populacional P refere-se à amostragem aleatória simples.

113 Nesse levantamento, cada estudante representa uma unidade amostral.

114 O erro padrão da estimativa do percentual populacional foi superior a 2% e inferior a 4,7%.

115 A estimativa do percentual populacional P foi igual a 75%.

Suponha que uma população de tamanho N seja constituída pelos elementos $y_1, \dots, y_N$, de modo que a média populacional é representada como

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$$

e a variância populacional é definida como

$$V^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2,$$

tal que $V > 0$. Denotando-se uma amostra aleatória simples de tamanho n retirada dessa população como $Y_1, \dots, Y_n$, e considerando que a média amostral possa ser escrita como

$$\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{1}{n}\sum_{i=1}^{N} \pi_i y_i,$$

em que $\pi_i \sim \text{Binomial}\left(n, \frac{1}{N}\right)$ e $\sum_{i=1}^{N} \pi_i = n$, julgue os itens seguintes.

116 Se $i \neq j$, a covariância entre π_i e π_j é negativa.

117 O valor esperado de $\bar{Y}$ é igual a μ.

118 A variância de $\bar{Y}$ é igual a $\frac{V^2}{n} \times \left(1 - \frac{n}{N}\right)$.

119 Se o estimador da variância populacional for $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y})^2$, então o valor esperado de S é igual a V.

120 $\text{Var}(\pi_i) = \frac{n}{N} \times \left(1 - \frac{1}{N}\right)$.

Espaço livre