Título

Autor

Data

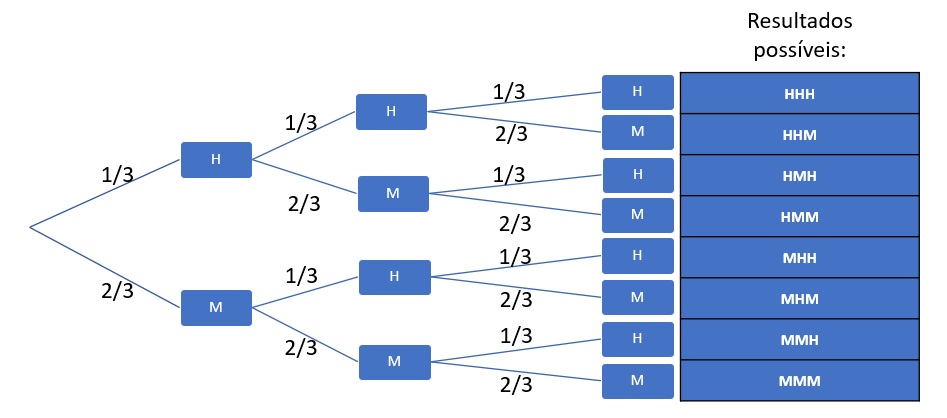
## Questão 1

### a)

Dado que 20 dos 60 alunos são homens, a probabilidade de um aluno homem ser selecionado ao acaso, considerando que todos os alunos tenham chances iguais de serem selecionados em um sorteio, é de , isto é, . Por outro lado, a chance de o aluno ser mulher, ou seja, não ser homem, é de . Para seleção de 3 representantes em 3 sorteios com reposição, nos três sorteios as probabilidades para seleção de homens e mulheres permanecem as mesmas. Nesse caso, a probabiliade de não haver nenhum homem entre os representantes será de . Por outro lado, se os sorteios forem feitos sem reposição, as probabilidades de seleção de um homem ou uma mulher serão diferentes entre o primeiro, segundo e o terceiro sorteio, já que a quantidade total de alunos, quantidade de homens e quantidade de mulheres serão alteradas pela retirada do indivíduo sorteado a cada rodada. Se uma mulher é sorteada na primeira rodada, a probabilidade de seleção de uma outra mulher no segundo sorteio deixa de ser de e passa a ser . Se uma mulher volta a ser sorteada na segunda rodada, a probabilidade do sorteio de uma terceira mulher entre os 58 alunos restantes passa a ser . Dessa forma, em sorteios sem reposição, considerando a regra dos produtos de probabilidades, a probabilidade de apenas mulheres serem selecionadas, isto é, nenhum homem ser selecionado, é de .

### b)

Na árvore de decisão abaixo, a letra H representa o sorteio de um homem, e a letra M o sorteio de uma mulher. Os resultados possíveis desses três sorteios, ou espaço amostral, será de .



### c)

O sorteio de um dos 60 alunos da turma corresponde a um experimento de Bernoulli, uma vez que, em cada sorteio, pode haver um sucesso (no caso, não ser sorteado um homem), ou um fracasso (ser sorteado um homem). A repetição desse experimento por três vezes, com reposição, ou seja, garantindo que todos os sorteios sejam independentes e que a probabilidade de sucesso (não ser sorteado um homem) sempre seja igual em todos os sorteios, caracteriza essa situação como um experimento binomial. Uma variável X cujos resultados correspondam ao número de sucessos nesse experimento terá uma distribuição binomial com a função de probabilidade abaixo:

No exemplo, (amostra de tamanho 3), (queremos saber a chance de que não haja homens na amostra) e :

Portanto, mesmo resultado obtido na resposta ao item A.

## Questão 2

### a)

*Situação I*: A variável dependente é a remoção dos prefeitos de seus cargos durante determinados mandatos. Essa variável é discreta, deverá indicar se prefeitos foram ou não removidos do cargo em um determinado mandato, com dois valores possíveis: 0 para mandatos sem remoção do cargo, 1 para mandatos com remoção do cargo.

*Situação II*: A variável dependente seria a quantidade de projetos apresentados por cada parlamentar por ano. A variável seria discreta e poderia assumir valores inteiros e positivos a partir de 0.

*Situação III*: A variável dependente é o tempo médio, em meses, que um prefeito cassado permanece no cargo. Dado que a duração do mandato utilizada para criar a variável dependente é medida em dias e existe a possibilidade de que o período de permanência após a cassação não dure meses inteiros, a variável dependente seria contínua. Poderia assumir valores reais positivos a partir do 0 até, teoricamente, a duração total do mandato do candidato em meses.

### b)

*Situação I*: Por se tratar de uma situação de análise de amostras de experimentos com resultados binários de sucesso ou fracasso (prefeitos foram removidos do cargo ou não), e havendo a premissa de que processos de cassação acontecem de forma independente entre os municípios, ou seja, que uma remoção de prefeito em um município não afeta a chance do prefeito de um segundo município ser removido do cargo, a distribuição apropriada é a binomial. Com ela, se tivermos uma estimativa de probabilidade da remoção de cargo ocorrer, podemos obter a probabilidade de haver casos de remoção de prefeitos em amostras de prefeitos com tamanho .

*Situação II*: Como argumentam Bussab e Morenttin (2010), para análise de variáveis discretas em casos em que se deseja obter a probabilidade da ocorrência de eventos em um intervalo de tempo, a função de probabilidade de variáveis aleatórias com distribuição de Poisson pode ser usada. No exemplo da questão, usar a função de probabilidade de uma variável com distribuição de Poisson possibilitaria, por exemplo, obter a probabilidade de um deputado apresentar 1 projeto no período de um ano. Vale lembrar ainda que a utilização da distribuição de Poisson depende da avaliação de que, no fenômeno de interesse, quanto maior o período de tempo, maior a probabilidade da ocorrência de um evento, e que cada evento ocorre de forma independente dos anteriores. Para a utilização na variável do exemplo, temos que supor que a chance de ocorrer um novo projeto apresentado pelo parlamentar é maior quanto maior o período de tempo sem apresentar um projeto. Ao mesmo tempo, supor que a apresentação de cada projeto é independente entre si. As duas suposições são, em geral, plausíveis, portanto a distribuição poderia ser usada.

Bussab, Wilton e Morettin, Pedro A. Estatística Básica. 6 ed. São Paulo: Saraiva, 2010.

*Situação III*: Podemos supor que a variável dependente tem distribuição simétrica em torno da média e, portanto, sendo uma variável contínua, a distribuição adequada seria a distribuição normal. A partir de estimativas de média e desvio padrão do tempo em que o prefeito se mantém no cargo obtida a partir da base de prefeitos cassados nos último 20 anos, podemos obter a probabilidade de que um prefeito fique no cargo por um determinado período.

### c)

O que ocorre com um prefeito em particular pode ser interpretado como uma variável aleatória de Bernoulli, pois pode haver apenas dois resultados na observação: o prefeito é removido (valor 1) ou não (valor 0). Para uma variável aleatória de Bernoulli X, a função de probabilidade *(x, p(x))* é:

Desse modo, no exemplo, se a probabilidade de afastamento do prefeito do cargo em decorrência de uma cassação é igual a 0,68, podemos dizer que , portanto, a probabilidade da não ocorrência do afastamento seria .

O mesmo resultado pode ser obtido usando a função de probabilidade de variáveis aleatórias com distribuição binomial:

No exemplo, (amostra de 1 prefeito), (não cassado) e :

## Questão 3

Para a decisão, podemos considerar a prova como um experimento binomial em que cada resposta pode ser considera correta ou incorreta. Levando em conta que cada questão tem duas opções possíveisde resposta (verdadeiro ou falso), podemos estimar que o aluno terá probabilidade de 1/2 de acertar cada resposta. Dessa forma, a partir da função de probabilidade de variáveis aleatórias com distribuição binomial, podemos estimar a probabilidade de acerto de ao menos 2/3 das questões. Em uma prova com três questões, os parâmetros seriam , (já que é necessário acertar 2/3 das perguntas para ser aprovado), e , de modo que:

Em uma prova com 36 questões, os parâmetros seriam , , ou seja, 2/3 das 36 questões, e , de modo que:

Portanto, a probabilidade de aprovação em uma prova com três questões (0,375) é consideravelmente maior do que em uma prova com 36 questões (0,0182). Em ambos os casos a probabilidade de não aprovação chutando todas as questões é mais alta do que a de aprovação, mas com 3 questões a probabilidade de aprovação é maior. Dessa forma, a preferência seria pela prova com 3 questões.

## Questão 4

### a)

A frequência simples e relativa da variável Tempo está abaixo. A partir dela, já é possível obter que a moda da variável é **20**:

## Tempo n Freq\_relativa  
## 15 1 3.125  
## 20 6 18.750  
## 22 1 3.125  
## 23 1 3.125  
## 25 3 9.375  
## 30 4 12.500  
## 35 2 6.250  
## 40 3 9.375  
## 45 4 12.500  
## 50 1 3.125  
## 55 1 3.125  
## 57 1 3.125  
## 60 3 9.375  
## 65 1 3.125  
## Total 32 100.000

Para a variável de Tempo em minutos, os dados de mínimo, máximo, mediana, média e quartis 1 e 3 são:

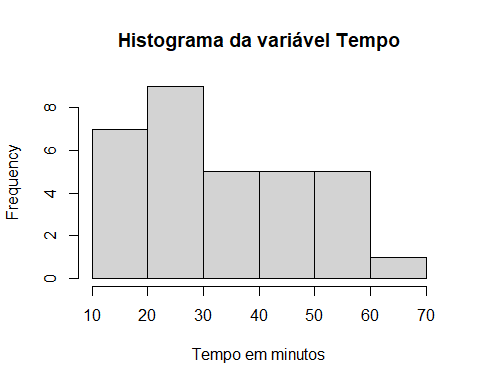
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 15.00 22.75 32.50 36.00 45.00 65.00

A variância da variável é de 219.03 e o desvio padrão de 14.8, enquanto a amplitude é de 50.

A partir desses resultados, podemos concluir que os tempos de viagem variavam sensivelmente, com um desvio padrão de 14.8 minutos, o que revela uma variação relevante em torno da média. A amplitude de 50 minutos também corrobora essa afirmação. Ao mesmo tempo, a menor diferença entre o primeiro quartil e a mediana se comparada à diferença entre a mediana e o terceiro quartil indica que há certa concentração dos alunos em tempos de viagem de até 35 minutos, intervalo que engloba mais da metade dos alunos. Nesse sentido, apesar de o tempo médio de viagem ter ficado em 36 minutos, é possível afirmar que essa média foi particularmente afetada por alunos que fazem viagens longas, o que aumentou o tempo médio de viagem em comparação à mediana de 32.5 minutos.

### b)

A partir do histograma abaixo, é possível observar que os dados são assimétricos à direita, característica que já poderia ser observada com as estatísticas descritivas. Isso pode ser visto, por exemplo, a partir da comparação do desvio padrão com a média, mediana, mínimo, máximo e quartis. No caso, um desvio padrão abaixo da média já chegaríamos a tempos de viagem próximos ao mínimo, enquanto entre os tempos de viagem acima da média seria necessário quase dois desvios padrão para alcançar o valor máximo, característica que indica maior concentração de casos abaixo da média. Além disso, as diferenças entre a mediana e os valores mínimo e máximo também demonstram onde há maior concentração de casos. Nesse caso, a diferença entre a mediana e o valor mínimo é menor do que a diferença entre o valor máximo e a mediana, o que indica maior dispersão (uma linha mais achatada) entre os dados acima da mediana.



### c)

A partir da região de procedência de um aluno não teremos o seu tempo exato de viagem até a universidade, afinal os tempos de viagem dentro de todas as regiões apresentam variação entre os apresentados na base . Porém, as estatísticas descritivas indicam que a região de procedência do indivíduo tem relação com o tempo de viagem. Na tabela 1, por exemplo, Observamos que as médias e medianas são mais baixas nas regiões Norte, Oeste e Centro se comparadas com essas medidas nas regiões Sul e Leste. Da mesma forma, entre alunos das regiões Sul e Leste encontramos os maiores valores mínimos e máximos de tempos de viagem.

A comparação dos histogramas por região (gráfico 1) também revela que os tempos de viagem nas zonas Leste e Sul estão mais deslocados à direita do que nas demais zonas, o que corrobora a interpretação de que os tempos de viagem são mais altos entre alunos residentes nessas localidades. Portanto, os dados indicam que alunos das regições Sul e Leste apresentam maior tempo de viagem em relação aos alunos de outras regiões. Para confirmação dessas análises, seria necessária a aplicação de outros testes, indo além das estatísticas descritivas.

#### Tabela 1 - Medidas de posição para a variável de tempo de viagem, por zona da cidade.

## NORTE OESTE SUL LESTE CENTRO  
## Min. 20.00 20.00 35.00 30.000 15.0  
## 1st Qu. 22.25 20.00 38.75 45.000 22.5  
## Median 26.50 20.00 42.50 52.500 27.5  
## Mean 25.75 21.50 45.00 50.875 27.5  
## 3rd Qu. 30.00 22.75 48.75 57.750 32.5  
## Max. 30.00 25.00 60.00 65.000 40.0  
## NA's 28.00 24.00 24.00 24.000 28.0

#### Gráfico 1 - Histogramas da variável de tempo de viagem por região:

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

