

Simulação de Estimadores

ANÁLISE DE GASTOS MÉDIOS COM LAZER E ENTRETENIMENTO PARA ABERTURA DO MUSEU ESPORTIVO INTERATIVO

Análise da População

Com a *seed* de número 523, utilizamos o R Studio para desenvolver nossas análises, que se iniciaram com a análise da população, ou seja, 35.000 registros. Nesse primeiro momento, descobrimos a média dos gastos dessa população, cerca de 1.709,061; também descobrimos a variância dos dados, cerca de 36.800,400, posteriormente descobrimos o Desvio Padrão dos dados, cerca de 606,637. Para nos ajudar em nossas análises, usamos histogramas e *boxplots*, assim poderíamos enxergar como os dados se comportavam em relação à média.

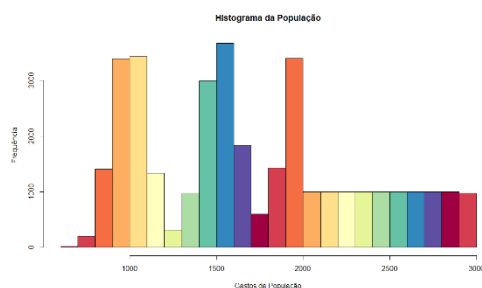


Figura 1.0 – histograma da amostra

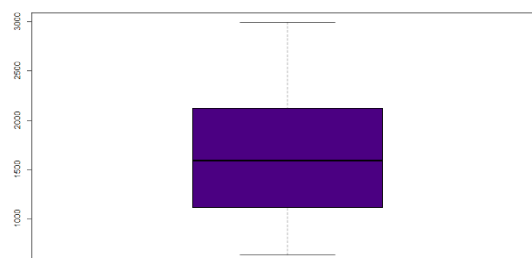


Figura 1.1 – boxplot comparativo

O histograma nos permite visualizar que certos intervalos de valores aparecem com mais frequência do que outros, fazendo com que a média fique mais centralizada, visto que muitos picos de frequência são as classes cujo os centros de classe têm valores de 1000, 1500 e 2000. Por sua vez, o *boxplot* nos permite visualizar o comportamento dos dados em relação à média. Caso só dispuséssemos do *boxplot*, poderíamos facilmente deduzir que os dados estão mais dispersos em relação à média, visto que o primeiro e o terceiro quartil estão mais afastados da média. No entanto, como dispomos das medidas de dispersão, podemos agora, quantificar essa dispersão dos dados; sendo a variância de 36.800,400 e o Desvio Padrão de 606,637.

Explicação do Código – Metodologia aplicada

Utilizamos em nossas análises, primordialmente a ferramenta R studio. Nossas análises começaram pela população, na qual utilizamos comandos como o *mean*, *var*, *sd*, *hist* e *boxplot*. O próximo passo foi a análise das amostras, algo mais trabalhoso, visto que tivemos de utilizar um pequeno *loop* de 5.000 iterações; em cada iteração um *sample* de tamanho $c(10, 30, 50, 100, 200, 500, 1.000, 5.000 \text{ e } 20.000)$, guardava os valores sorteados aleatoriamente; o comando $media[i] = c(mean(amostra))$, guardava as médias das amostras. Portanto no fim da execução do *loop*, tínhamos um vetor com 5.000 médias, objeto no qual foi utilizado para fazermos as análises.

Código do loop

```
for (i in 1:5000) { #controla o número de amostras
  amostra = c(sample(tab$gasto, 10)); #controla no número de elementos amostrais
  media[i] = c(mean(amostra)); #tira a média das amostras e salva num vetor
}
```

Demais códigos

```
mean(x);var(x);sd(x)
hist.(main="", x, xlab="", ylab="", col=vetorCores)
boxplot(main="", xlab="", col=vetorCores, arquivo)
```

Análise das Amostras

Utilizamos a mesma metodologia para todas as 9 amostras colhidas, o que mudava de uma para outra era o número de elementos amostrais colhidos e as cores finais do histograma e *boxplot*.

Amostras com 10 Elementos amostrais

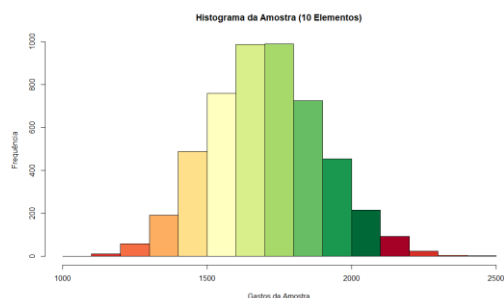


Figura 1.2 – histograma da amostra

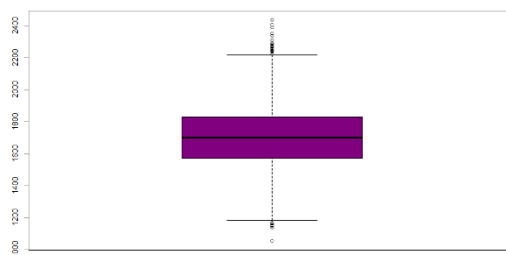


Figura 1.3 – boxplot comparativo

Média: 1.704,563

Variância: 37.439,530

Desvio Padrão: 193,493

Amostras com 30 Elementos amostrais

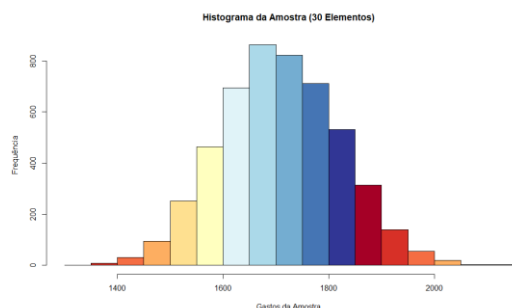


Figura 1.4 – histograma da amostra

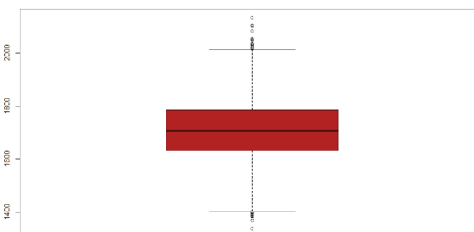


Figura 1.5 – boxplot comparativo

Média: 1.708,559

Variância: 12.508,00

Desvio Padrão: 111,839

Amostras com 50 Elementos amostrais

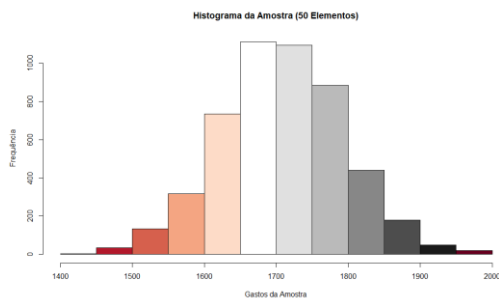


Figura 1.6 – histograma da amostra

Média: 1.708,824

Variância: 7.291,196

Desvio Padrão: 85,885

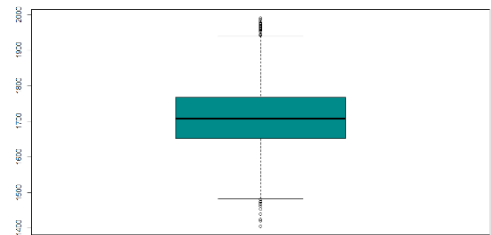


Figura 1.7 – *boxplot* comparativo

Amostras com 100 Elementos amostrais

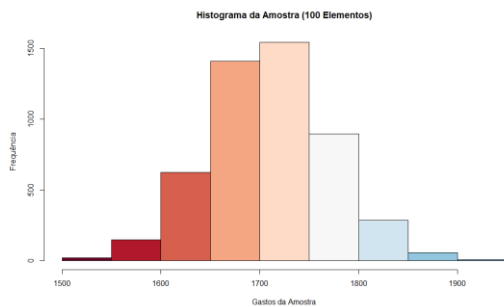


Figura 1.8 – histograma da amostra

Média: 1.709,926

Variância: 3.697,495

Desvio Padrão: 60,807

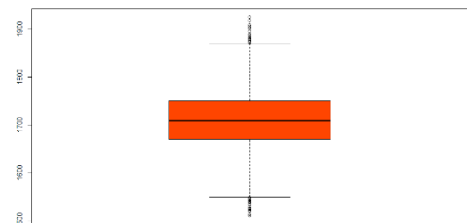


Figura 1.9 – *boxplot* comparativo

Amostras com 1000 Elementos amostrais

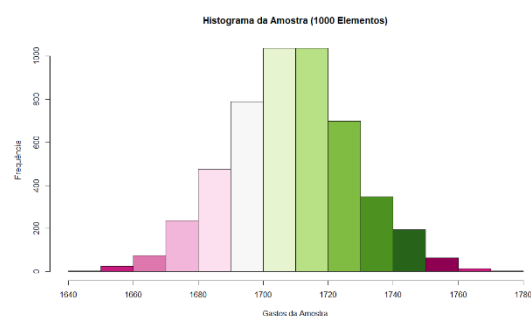


Figura 1.10 – histograma da amostra

Média: 1.708,580

Variância: 362,290

Desvio Padrão: 19,033

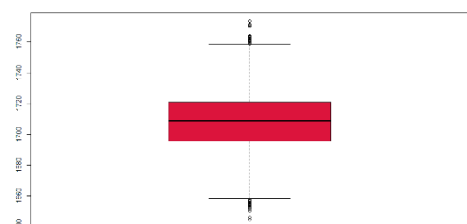


Figura 1.11 – *boxplot* comparativo

Mesmo embora, nem todas amostras de dados estejam dispostas nesse relatório, é possível visualizar o comportamento dos dados em relação a sua dispersão da média. Ao observarmos os histogramas, podemos perceber que os dados passam a estruturar no formato de uma curva normal, ou seja, a maior frequência de intervalos de valores, ocorre próximo a média. Os histogramas, nos permitem visualizar de forma mais clara, o quão dispersos da média os dados se encontram. As medidas de dispersão em relação à média, nos ajudam a quantificar essa dispersão, quanto maior for a amostra, mais precisa ela será, visto que as medidas de dispersão começam a tender a 0, indicando a maior precisão dos dados.

Conclusão – Comparação de *Boxplots*

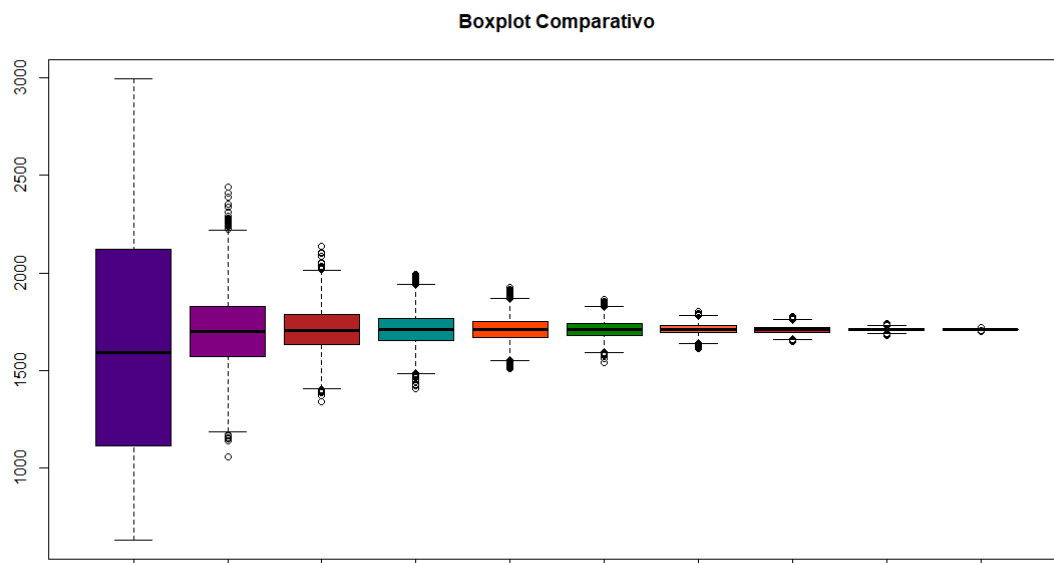


Figura 1.12 – *boxplot* comparativo

Na imagem acima, comparamos os *boxplots* da população com os *boxplots* das 9 amostras colhidas. No *boxplot* da população, podemos observar duas coisas interessantes, a primeira delas é a grande dispersão dos dados, visto que o primeiro e o terceiro quartil estão afastados da média; a segunda delas é o fato de não termos *outliers*. Isso se deve ao fato de estarmos considerando a população total de *Ursk*, e não somente uma parcela conhecida da população (amostra).

A partir do segundo *boxplot*, observamos a ocorrência de *outliers*, isso se deve ao fato de estarmos considerando apenas uma parcela da população, mesmo que sorteada aleatoriamente. Também podemos observar a partir do segundo *boxplot*, que os dados estão cada vez menos dispersos, significando que a precisão está tendendo a aumentar. Quando observamos o último *boxplot*, vemos que ainda há poucos *outliers*, entretanto, os dados estão mais precisos, visto que mantendo a escala, o intervalo interquartílico é quase imperceptível.