

Análise Exploratória de Dados

processo de usar as ferramentas estatísticas (gráficos, medidas de tendência central e de variabilidade) para investigar o conjunto de dados, de maneira a compreender suas características mais importantes.

Outliers

- ❖ **um valor localizado muito distante de quase todos os outros valores;**
- ❖ **um valor extremo;**
- ❖ **pode ter um efeito dramático na média, no desvio-padrão e na escala do histograma, distorcendo a verdadeira natureza da distribuição.**

Boxplots

(Diagrama em Caixas)

Revela:

- ❖ o centro dos dados;
- ❖ a dispersão dos dados;
- ❖ a distribuição dos dados;
- ❖ a presença de 'outliers';

Excelente para comparar dois ou mais conjuntos de dados.

Boxplots

Resumo de cinco números:

- ❖ **Mínimo**
- ❖ **1º Quartil (Q_1)**
- ❖ **Mediana (Q_2)**
- ❖ **3º Quartil (Q_3)**
- ❖ **Máximo**

Boxplots

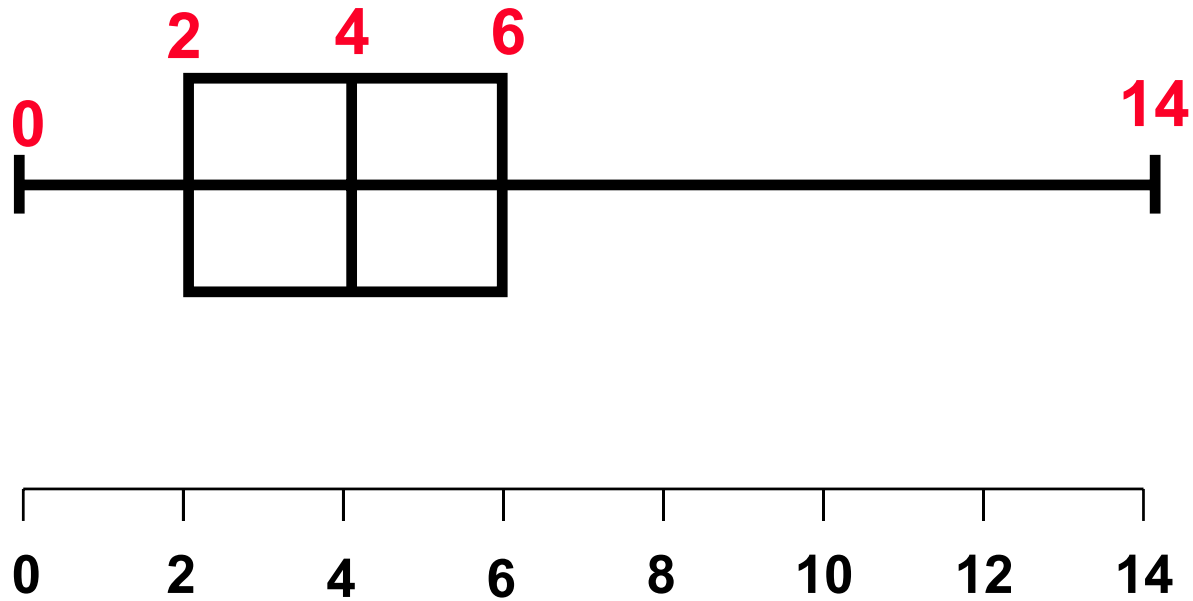
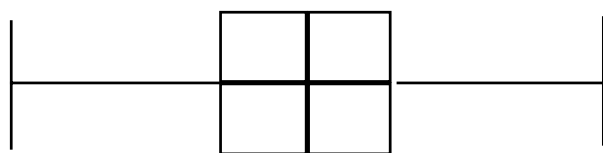
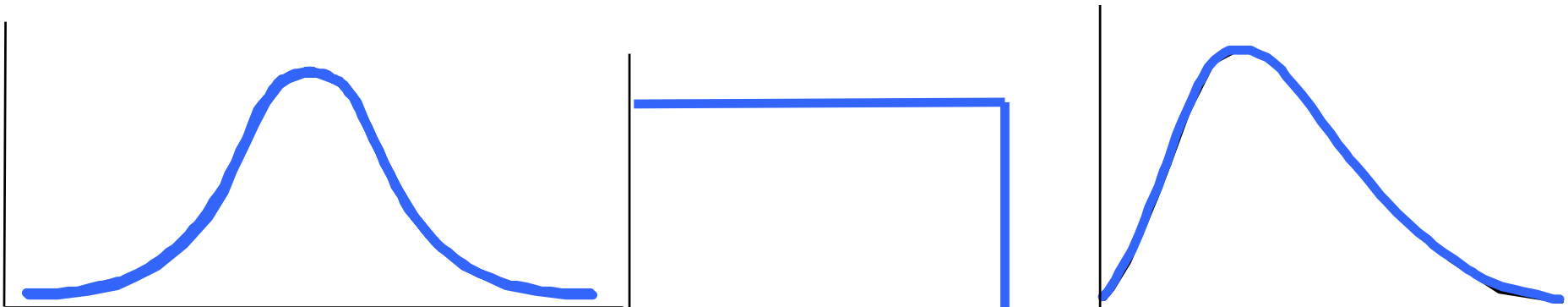


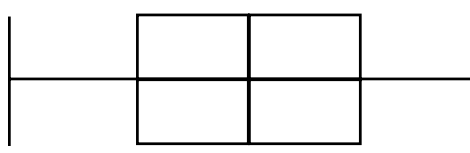
Figura 2-18

Figura 2-19

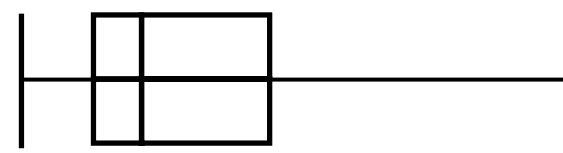
Boxplots



Forma de Sino



Uniforme



Assimétrico

Exploração de Dados

- ❖ **Medidas de tendência central:**
Média, mediana e moda.
- ❖ **Medidas de variação:**
Desvio-padrão e amplitude.
- ❖ **Medidas de posição:**
Valor mínimo, valor máximo e quartis.
- ❖ **Valores incomuns: 'Outliers'**
- ❖ **Distribuição:**
Histogramas, gráfico ramo-e-folhas e 'boxplots'.

Exemplo de Aplicação

❖ Descrição:

Empreendimento imobiliário, consistindo de dois edifícios residenciais, com apartamentos de 1 dormitório e área útil de 53m².

❖ Histórico:

Em uma 1ª fase, foi construído apenas um edifício, denominado “Bloco A”. Para a conclusão do Bloco B houve uma troca de empreiteiras.

❖ Objetivo:

Verificar inicialmente a suspeita de que os apartamentos foram construídos com metragem diferente daquela especificada na escritura.

❖ Fonte:

“Noções de Probabilidade e Estatística”, de Marcos Nascimento Magalhães e Antonio Carlos Pedroso de Lima, Edusp.

Exercício 3:

a)

b)

c) Área Útil Total

Análise Global (Blocos A e B):

Descriptive Statistics: Total

| Variable | N | Mean | Median | TrMean | StDev | SE Mean |
|----------|-----|--------|--------|--------|-------|---------|
| Total | 152 | 50,914 | 50,750 | 50,887 | 3,011 | 0,244 |

| Variable | Minimum | Maximum | Q1 | Q3 |
|----------|---------|---------|--------|--------|
| Total | 44,600 | 58,400 | 48,525 | 53,200 |

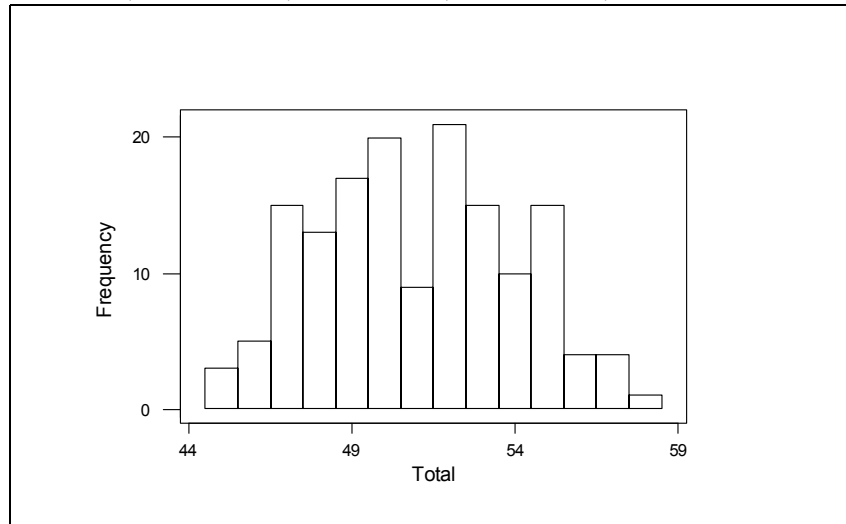


Figura 3 - Histograma Área Útil (Todos os Blocos)

Análise por Bloco:

Descriptive Statistics: Total by Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|----|--------|--------|--------|-------|
| Total | A | 76 | 53,422 | 53,200 | 53,396 | 1,772 |
| | B | 76 | 48,407 | 48,550 | 48,435 | 1,538 |

| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
|----------|-------|---------|---------|---------|--------|--------|
| Total | A | 0,203 | 48,800 | 58,400 | 52,200 | 54,700 |
| | B | 0,176 | 44,600 | 51,700 | 47,250 | 49,600 |

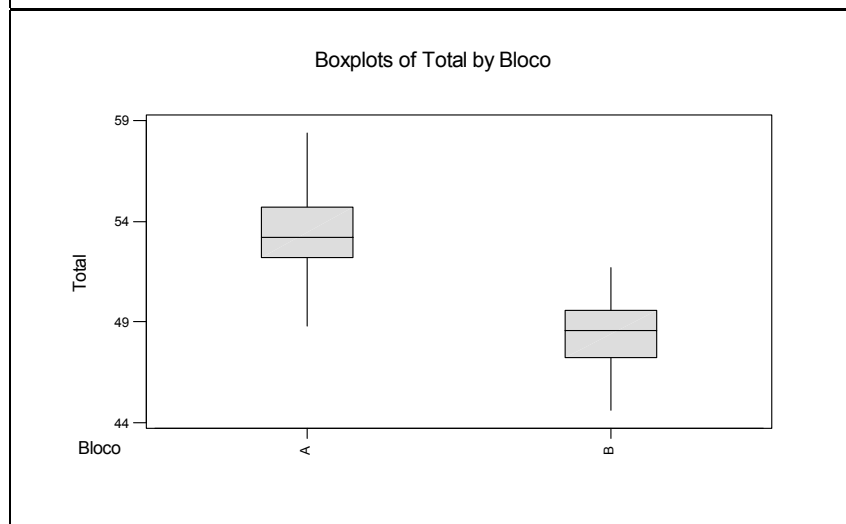
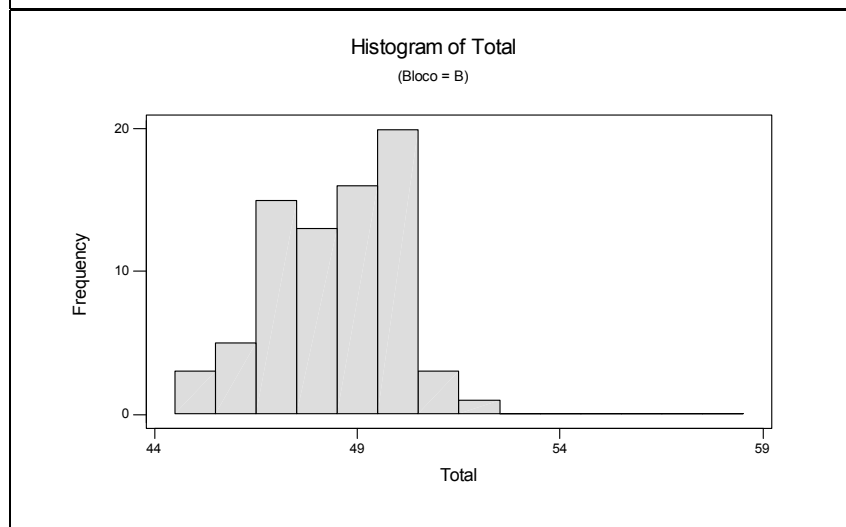
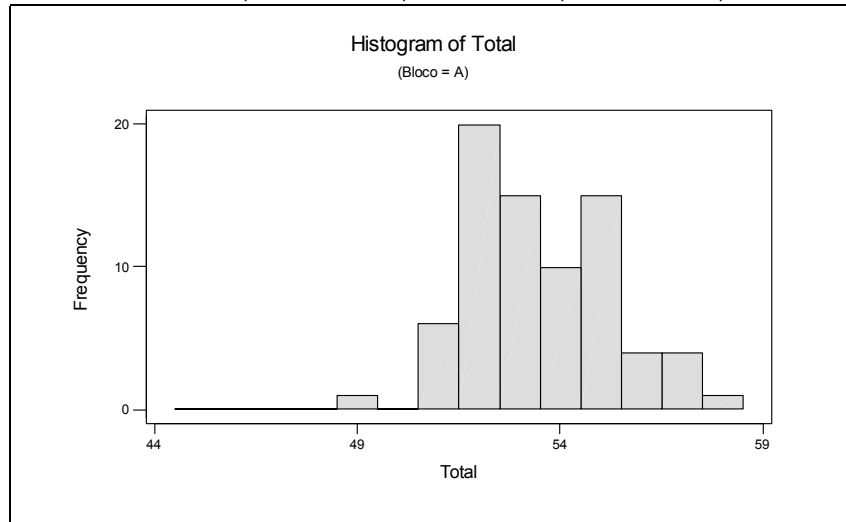


Figura 4 - Análise Gráfica Área Útil por Bloco

Análise por cômodo:

Descriptive Statistics: Sala by Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|---------|---------|---------|--------|--------|
| Sala | A | 76 | 28,093 | 27,950 | 28,091 | 1,083 |
| | B | 76 | 23,132 | 23,100 | 23,147 | 0,849 |
| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
| Sala | A | 0,124 | 25,600 | 30,900 | 27,500 | 28,900 |
| | B | 0,097 | 20,900 | 25,000 | 22,500 | 23,775 |

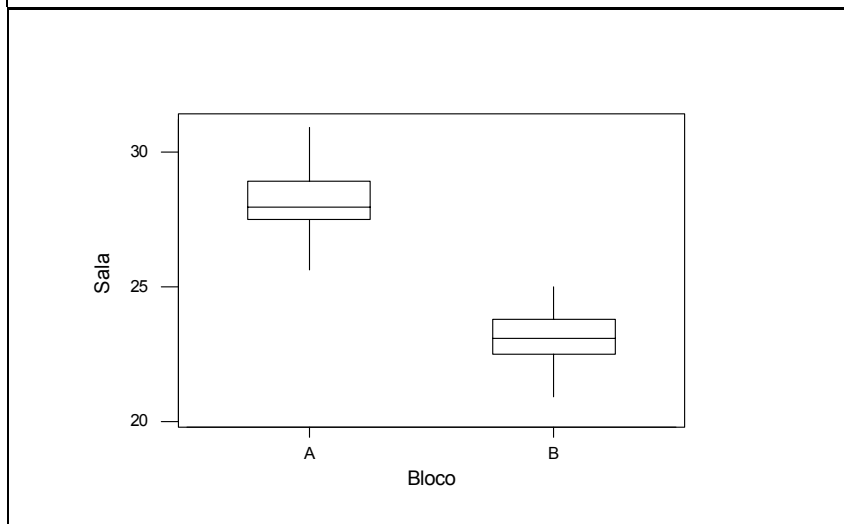
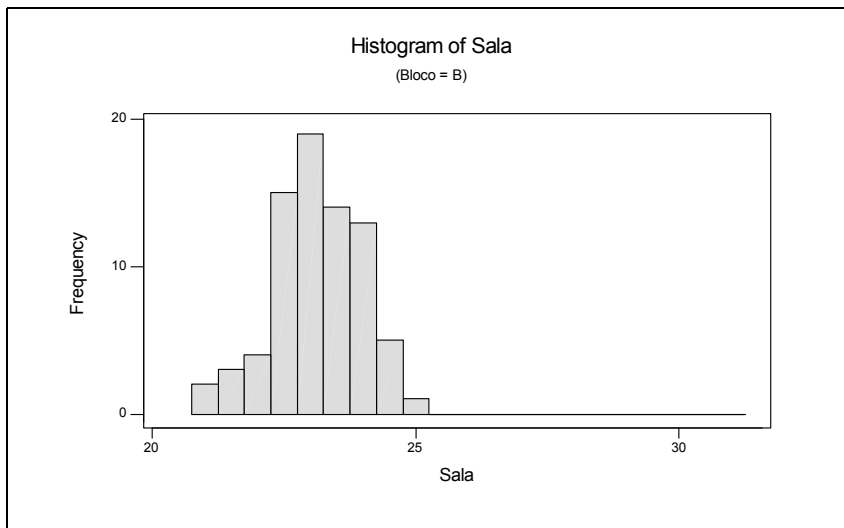
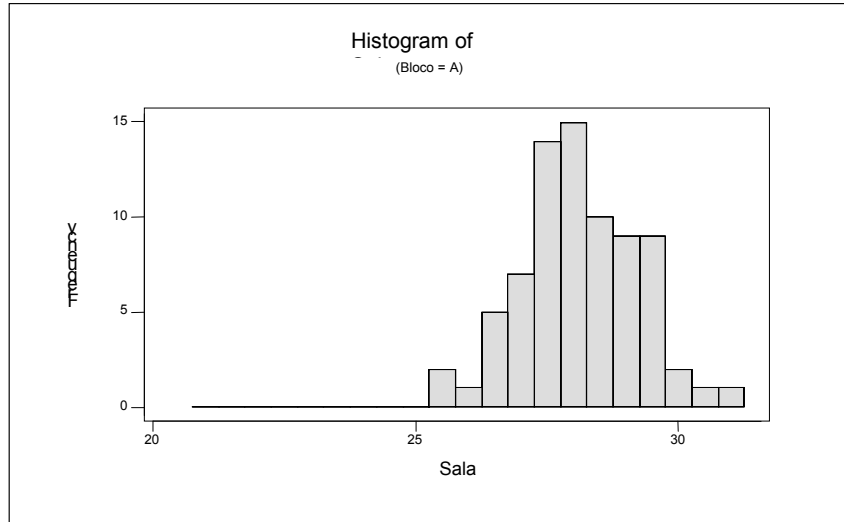
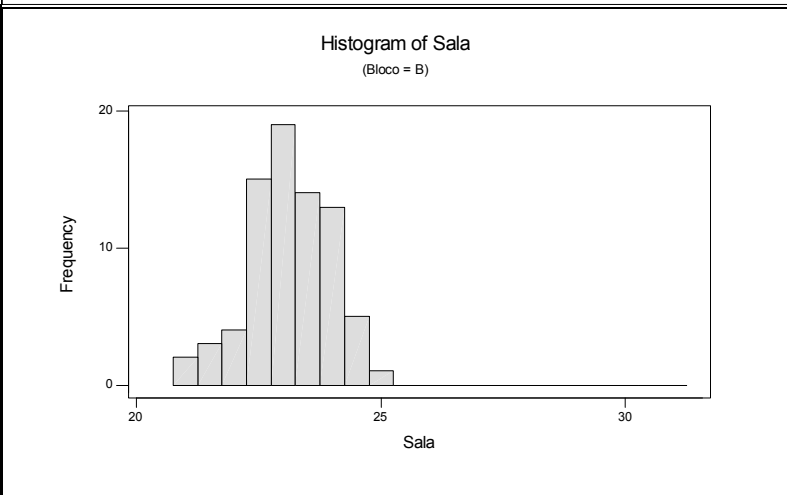
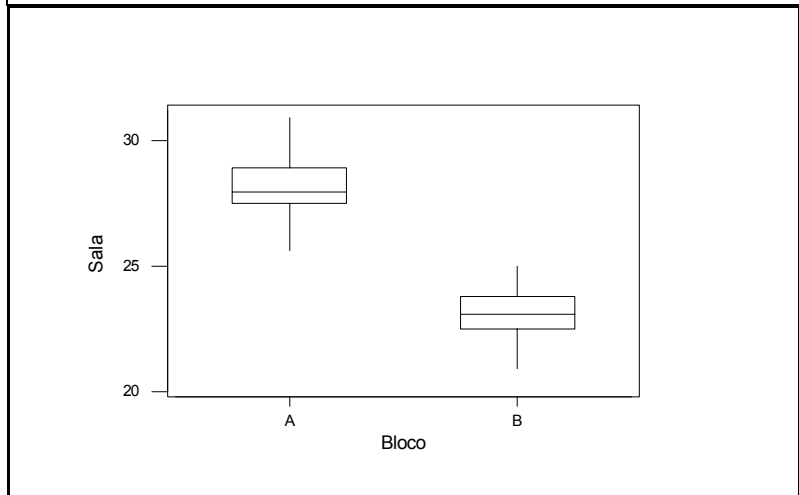
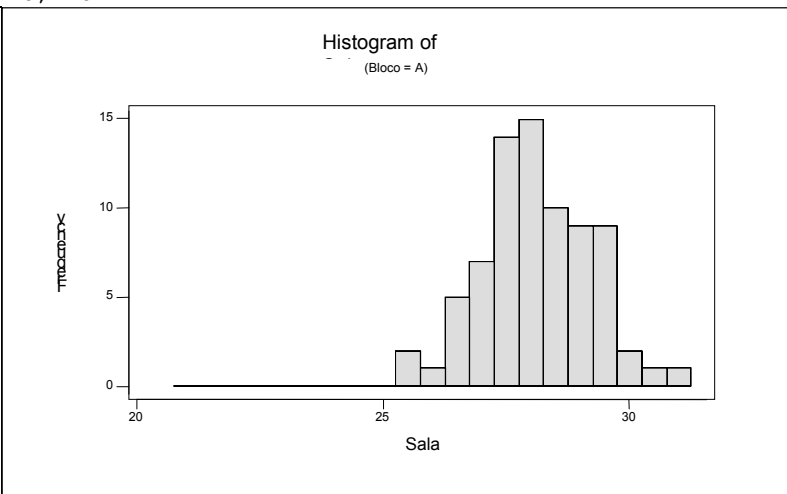
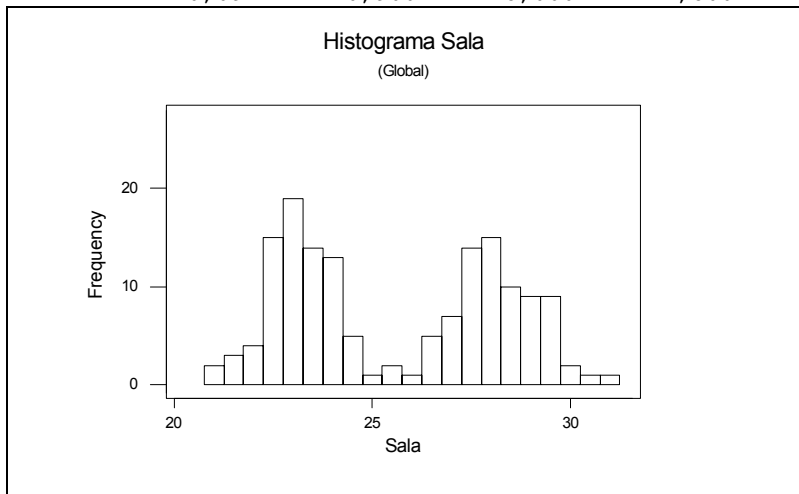


Figura 5 - Análise Gráfica Área Sala por Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|-----|--------|--------|--------|-------|
| Global | | 152 | 25,613 | 25,300 | 25,593 | 2,671 |
| Sala | A | 76 | 28,093 | 27,950 | 28,091 | 1,083 |
| | B | 76 | 23,132 | 23,100 | 23,147 | 0,849 |

| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
|----------|-------|---------|---------|---------|--------|--------|
| Global | | 0,217 | 20,900 | 30,900 | 23,100 | 27,975 |
| Sala | A | 0,124 | 25,600 | 30,900 | 27,500 | 28,900 |
| | B | 0,097 | 20,900 | 25,000 | 22,500 | 23,775 |



Descriptive Statistics: Cozinha by Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|---------|---------|---------|--------|--------|
| Cozinha | A | 76 | 7,9289 | 8,0000 | 7,9441 | 0,4653 |
| | B | 76 | 7,8816 | 7,8000 | 7,8824 | 0,6048 |
| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
| Cozinha | A | 0,0534 | 6,5000 | 9,2000 | 7,7000 | 8,2000 |
| | B | 0,0694 | 6,4000 | 9,6000 | 7,5000 | 8,3750 |

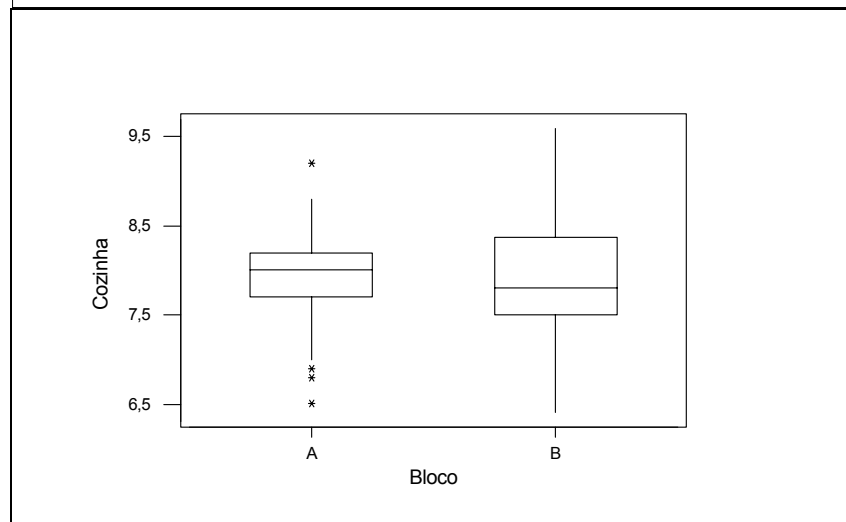
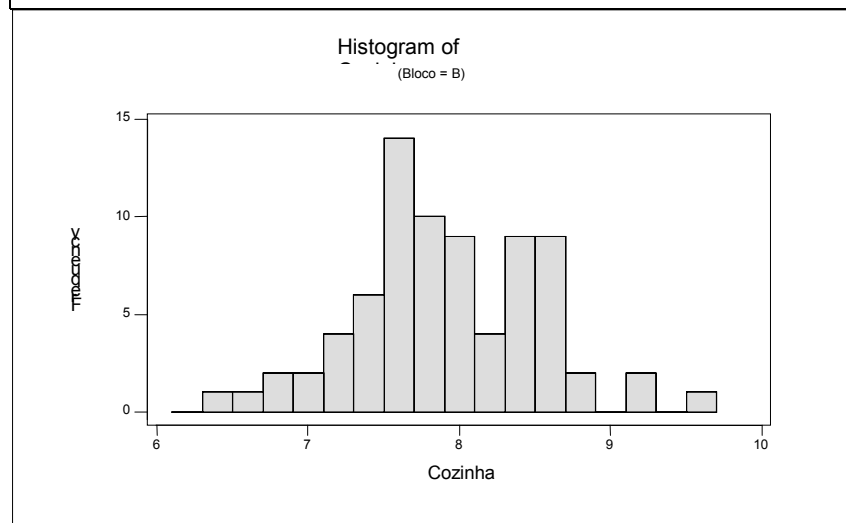
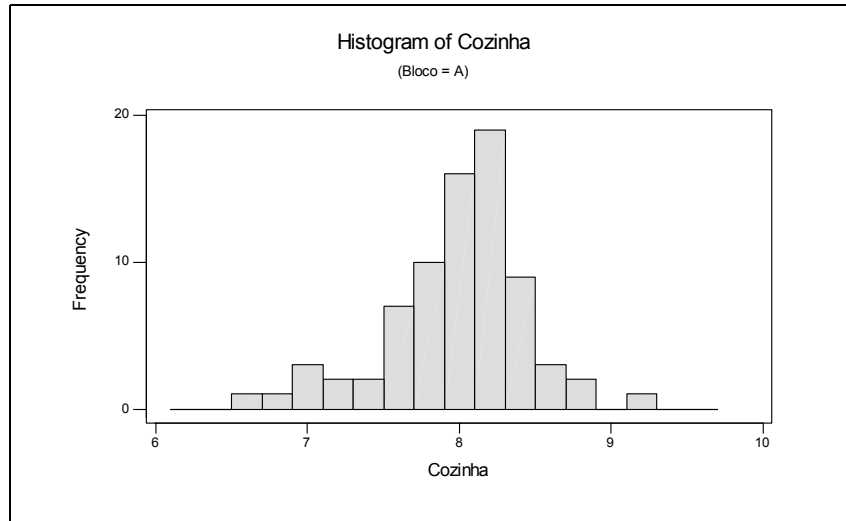


Figura 6 - Análise Gráfica Área Cozinha por Bloco

Descriptive Statistics: Banheiro by Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|---------|---------|---------|--------|--------|
| Banheiro | A | 76 | 4,4500 | 4,5000 | 4,4515 | 0,4623 |
| | B | 76 | 4,4645 | 4,5000 | 4,4647 | 0,5158 |
| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
| Banheiro | A | 0,0530 | 3,5000 | 5,4000 | 4,1000 | 4,7000 |
| | B | 0,0592 | 2,9000 | 5,8000 | 4,1000 | 4,8000 |

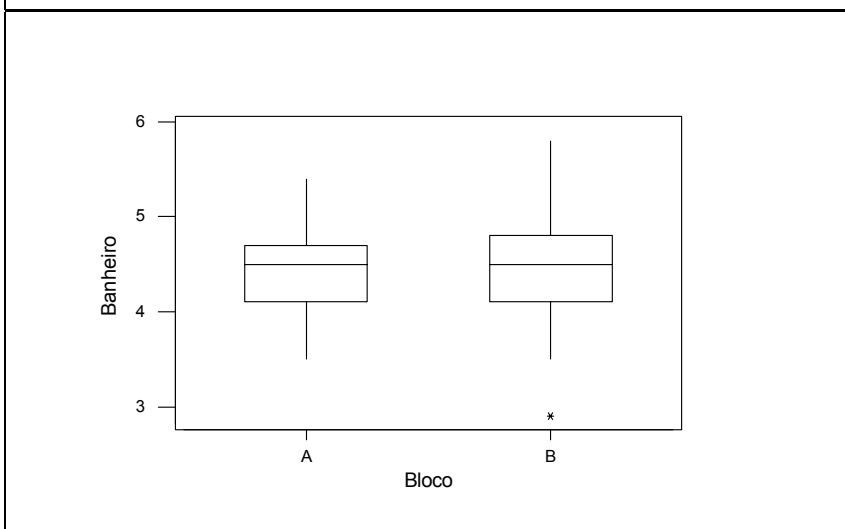
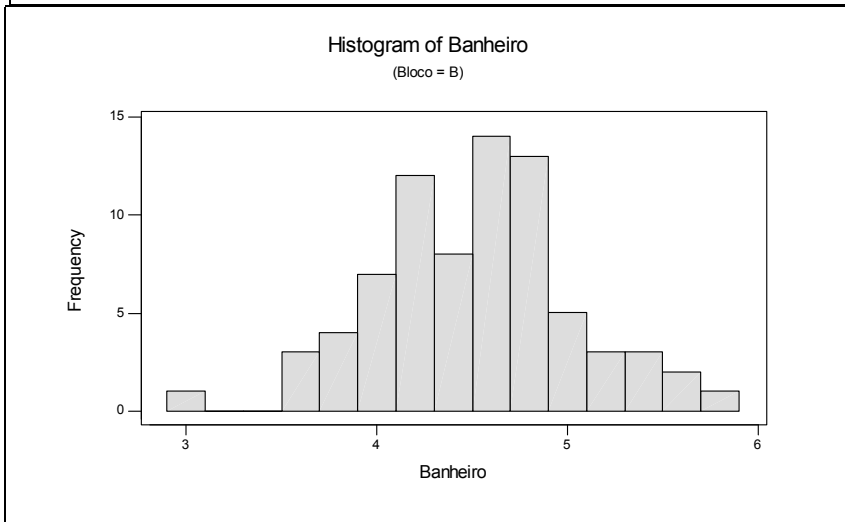
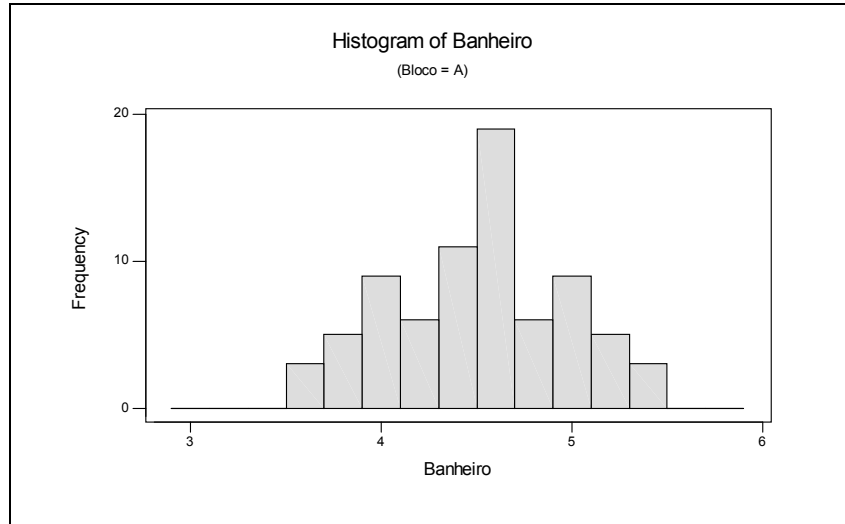


Figura 7 - Análise Gráfica Área Banheiro por Bloco

Descriptive Statistics: Dorm by Bloco

| Variable | Bloco | N | Mean | Median | TrMean | StDev |
|----------|-------|---------|---------|---------|--------|--------|
| Dorm | A | 76 | 12,950 | 13,000 | 12,947 | 1,173 |
| | B | 76 | 12,929 | 12,900 | 12,910 | 1,177 |
| Variable | Bloco | SE Mean | Minimum | Maximum | Q1 | Q3 |
| Dorm | A | 0,135 | 10,000 | 15,900 | 12,200 | 13,675 |
| | B | 0,135 | 10,600 | 15,700 | 12,200 | 13,600 |

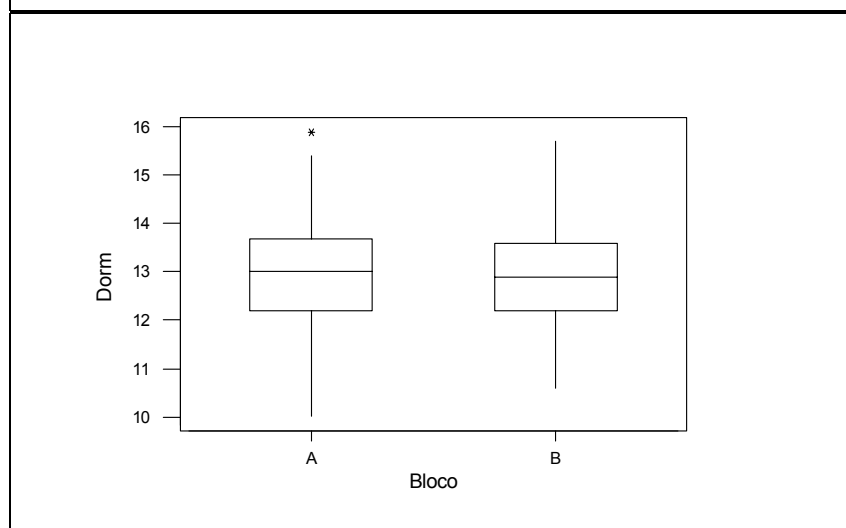
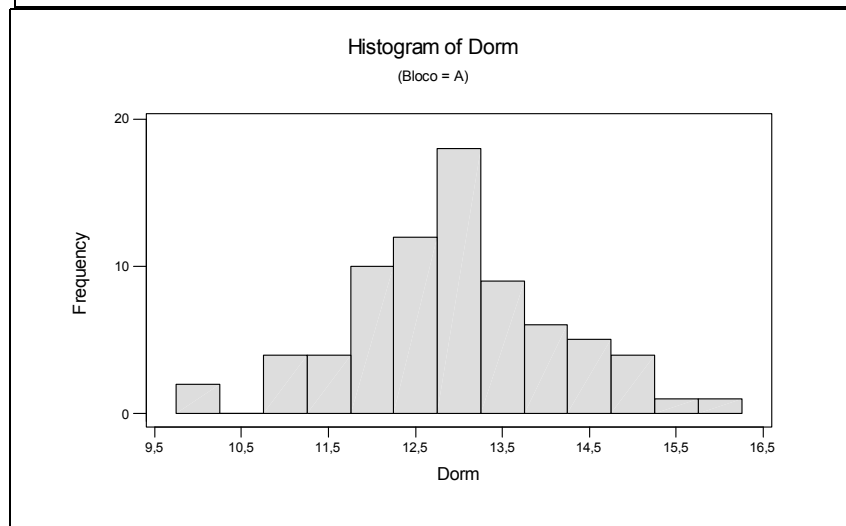
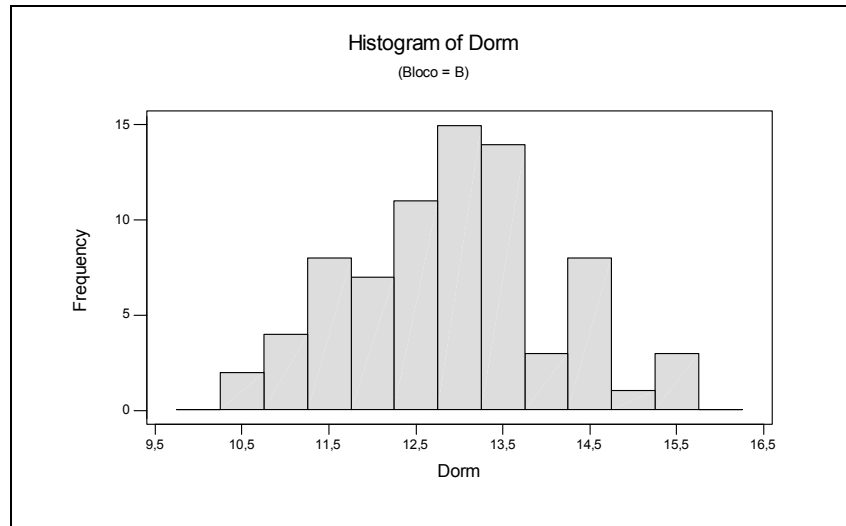


Figura 8 – Análise Gráfica Área Dormitório por Bloco

d) Problemas Estruturais

Tabulated Statistics: Rachadura; Bloco

| Rows: Rachadur | Columns: Bloco | | |
|----------------|----------------|-------------|---------------|
| A | B | All | |
| 0 | 43 50,59 | 42 49,41 | 85 100,00 |
| 1 | 33 49,25 | 34 50,75 | 67 100,00 |
| All | 76 50,00 | 76 50,00 | 152 100,00 |

Tabulated Statistics: Infiltr; Bloco

| Rows: Infiltr | Columns: Bloco | | |
|---------------|----------------|-------------|---------------|
| A | B | All | |
| 0 | 53 49,07 | 55 50,93 | 108 100,00 |
| 1 | 23 52,27 | 21 47,73 | 44 100,00 |
| All | 76 50,00 | 76 50,00 | 152 100,00 |

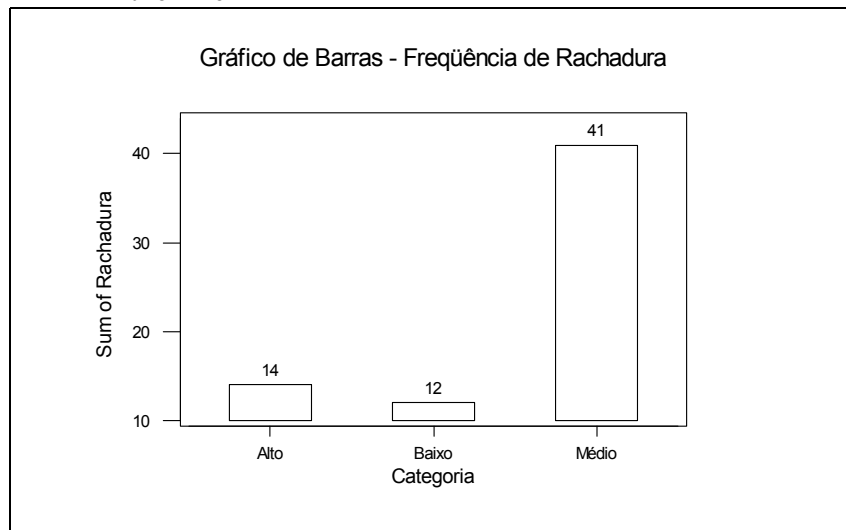
Cell Contents --
Count
% of Row

e) Ocorrência de Problemas Estruturais:

Tabulated Statistics: Rachadura; Categoria

| Rows: Rachadur | Columns: Categori | | | |
|----------------|-------------------|-------|-------|--------|
| | Alto | Baixo | Médio | All |
| 0 | 34 | 36 | 15 | 85 |
| | 40,00 | 42,35 | 17,65 | 100,00 |
| 1 | 14 | 12 | 41 | 67 |
| | 20,90 | 17,91 | 61,19 | 100,00 |
| All | 48 | 48 | 56 | 152 |
| | 31,58 | 31,58 | 36,84 | 100,00 |

Cell Contents --
Count
% of Row



Tabulated Statistics: Infiltr; Categoria

| Rows: Infiltr | Columns: Categori | | | |
|---------------|-------------------|-------|-------|--------|
| | Alto | Baixo | Médio | All |
| 0 | 20 | 41 | 47 | 108 |
| | 18,52 | 37,96 | 43,52 | 100,00 |
| 1 | 28 | 7 | 9 | 44 |
| | 63,64 | 15,91 | 20,45 | 100,00 |
| All | 48 | 48 | 56 | 152 |
| | 31,58 | 31,58 | 36,84 | 100,00 |

Cell Contents --
Count
% of Row

