

Aula 07: Amplitude Total e Boxplot

O Boxplot é a representação gráfica dos valores principais abaixo:

1. **Linha Inferior (Mínimo)**: O menor valor da distribuição que **não é um outlier**.
2. **Limite Inferior (LI)**: O valor calculado de $Q1 - (1.5 \text{ vezes IQR})$.
3. **Q1 (25%)**: O primeiro quartil.
4. **Mediana (Q2/50%)**: O valor central.
5. **Q3 (75%)**: O terceiro quartil.
6. **Linha Superior (Máximo)**: O maior valor da distribuição que **não é um outlier**.
7. **Limite Superior (LS)**: O valor calculado de $Q3 + (1.5 \text{ vezes IQR})$.
8. **Pontos (Outliers)**: Valores que caem fora dos "braços" (acima do Limite Superior ou abaixo do Limite Inferior).

Painel de Visualização

O código abaixo gera um painel com o Boxplot dos preços e um resumo de todas as medidas calculadas.

```
def calcular_medidas_descritivas(data_array):
```

```
    if data_array is None or len(data_array) == 0:
```

```
        return None
```

```
# Medidas de Tendência Central
```

```
media = np.mean(data_array)
```

```
mediana = np.median(data_array)
```

```
# Medidas de Posição (Quartis e IQR)
```

```
Q1 = np.percentile(data_array, 25)
```

```
Q3 = np.percentile(data_array, 75)
```

```
IQR = Q3 - Q1
```

```
# Limites de Outliers
```

```
limite_superior = Q3 + (1.5 * IQR)
```

```
limite_inferior = Q1 - (1.5 * IQR)
```

```
# Valores Extremos

min_valor = np.min(data_array)

max_valor = np.max(data_array)

return {

'media': media,
'mediana': mediana,
'Q1': Q1,
'Q3': Q3,
'IQR': IQR,
'limite_superior': limite_superior,
'limite_inferior': limite_inferior,
'min_valor': min_valor,
'max_valor': max_valor,
}

def gerar_painel_boxplot(data_array, medidas, titulo_boxplot='Boxplot da Distribuição de Dados', caminho_salvar=None):

if medidas is None:
    print("Erro: Medidas estatísticas não fornecidas ou inválidas.")
    return

fig, axes = plt.subplots(1, 2, figsize=(16, 8))

# --- POSIÇÃO 1: BOXPLOT ---

sns.boxplot(y=data_array, ax=axes[0])
```

```
axes[0].set_title(titulo_boxplot)

axes[0].set_ylabel(data_array.name if hasattr(data_array, 'name') else 'Valores')

# --- POSIÇÃO 2: CENÁRIO DE MEDIDAS (plt.text) ---

axes[1].axis('off')

axes[1].set_title('Medidas Estatísticas Calculadas')

# Preparando o texto formatado (usando as chaves do dicionário 'medidas')

resumo = (
    f"Medidas de Tendência Central:\n"
    f"  Média: R$ {medidas['media']:.2f}\n"
    f"  Mediana (Q2): R$ {medidas['mediana']:.2f}\n"
    f"\n"
    f"Medidas de Posição/Dispersão:\n"
    f"  Q1: R$ {medidas['Q1']:.2f}\n"
    f"  Q3: R$ {medidas['Q3']:.2f}\n"
    f"  IQR: R$ {medidas['IQR']:.2f}\n"
    f"\n"
    f"Limites e Extremos:\n"
    f"  Limite Superior (LS): R$ {medidas['limite_superior']:.2f}\n"
    f"  Limite Inferior (LI): R$ {medidas['limite_inferior']:.2f}\n"
    f"  Valor Máximo: R$ {medidas['max_valor']:.2f}\n"
    f"  Valor Mínimo: R$ {medidas['min_valor']:.2f}\n"
)
# Adicionando o texto
axes[1].text(0.1, 0.95, resumo,
```

```
transform=axes[1].transAxes,  
fontsize=12,  
verticalalignment='top',  
bbox=dict(boxstyle="round,pad=0.5", alpha=0.1, color='lightgray'))  
  
plt.tight_layout()  
plt.show()
```

Análise

Com base no painel, vamos debater sobre alguns pontos:

1. **Centro dos Dados:** Onde está a mediana?
2. **Assimetria:** A mediana está mais próxima de Q1 ou Q3? (Isso indica se os preços baixos ou altos têm maior concentração).
3. **Dispersão:** A caixa é muito longa? Se sim, o **IQR** é alto, indicando grande variabilidade nos 50% centrais.
4. **Outliers:** Existem pontos acima ou abaixo dos "bigodinhos" (outliers)? Se sim, isso justifica a alta diferença entre Média e Mediana, e a grande **Amplitude Total**.

Atividade Prática de Revisão

Atualize seu script com o conteúdo de hoje, construindo um resumo mais estruturado da amostra que vem trabalhando nas últimas aulas.