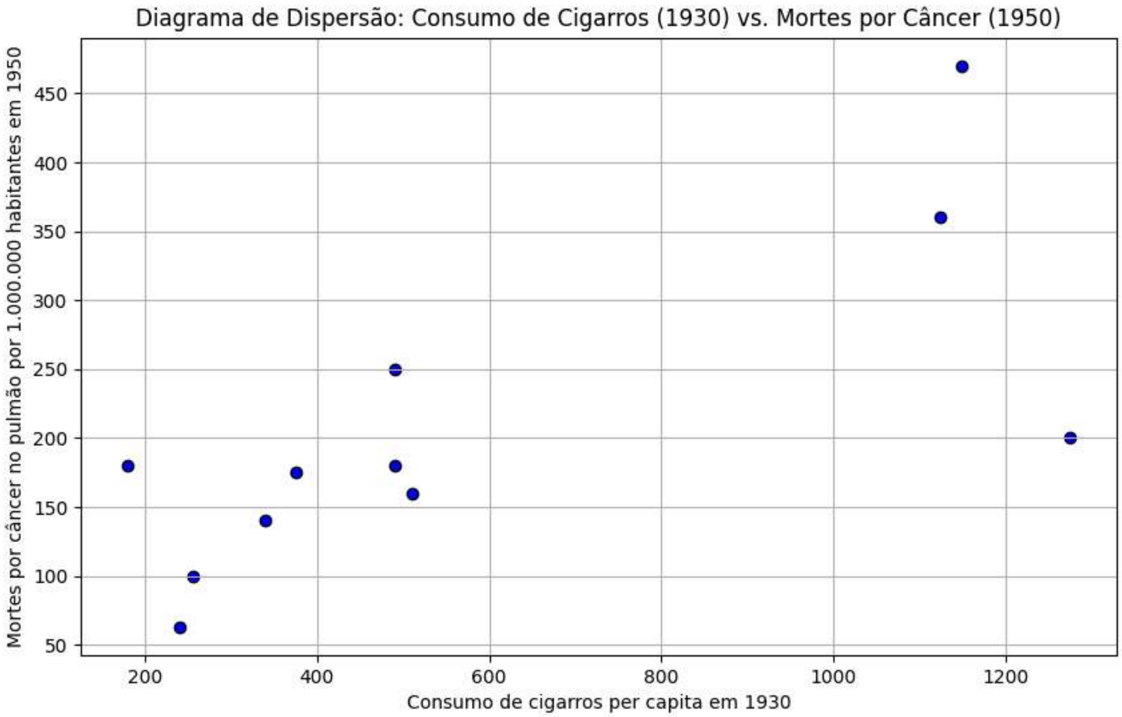
```python
import matplotlib.pyplot as plt

# Dados fornecidos pela atividadeÇ
paises = ["Islândia", "Noruega", "Suécia", "Dinamarca", "Canadá", "Austrália", "Holanda", "Suíça", "Finlândia", "Grã-Bretanha", "EUA"]
consumo_cigarros_1930 = [240, 255, 340, 375, 510, 490, 490, 180, 1125, 1150, 1275]  # Consumo per capita em 1930
mortes_cancer_1950 = [63, 100, 140, 175, 160, 180, 250, 180, 360, 470, 200]  # Mortes por câncer no pulmão por 1.000.000 habitantes em 1950

# (a) Escolha adequadamente X e Y.
X = consumo_cigarros_1930
Y = mortes_cancer_1950

print("(a) Escolha de X e Y:")
print(f"Variável Independente (X): Consumo de cigarros per capita em 1930 = {X}")
print(f"Variável Dependente (Y): Mortes por câncer no pulmão em 1950 = {Y}")
print("-" * 50)

# (b) Construa o diagrama de dispersão entre X e Y.
print("\n(b) Diagrama de Dispersão:")

plt.figure(figsize=(10, 6)) # Define o tamanho da figura do gráfico
plt.scatter(X, Y, color='blue', edgecolor='black') # Cria o gráfico de dispersão com pontos azuis e bordas pretas
plt.title('Diagrama de Dispersão: Consumo de Cigarros (1930) vs. Mortes por Câncer (1950)') # Título do gráfico
plt.xlabel('Consumo de cigarros per capita em 1930') # Rótulo do eixo X
plt.ylabel('Mortes por câncer no pulmão por 1.000.000 habitantes em 1950') # Rótulo do eixo Y
plt.grid(True) # Adiciona uma grade ao gráfico para melhor visualização
plt.show() # Exibe o gráfico
print("-" * 50)


# Função para calcular a média de uma lista de números
def calcular_media(lista):
    soma_lista = sum(lista) # Soma todos os elementos da lista
    n_elementos = len(lista) # Conta o número de elementos na lista
    media_calculada = soma_lista / n_elementos # Calcula a média
    return media_calculada, soma_lista, n_elementos

# (c) Calcule o coeficiente de correlação linear de Pearson. Interprete o valor obtido.
print("\n(c) Coeficiente de Correlação Linear de Pearson (r):")

# Calcular médias de X e Y:
media_X, soma_X, n = calcular_media(X)
media_Y, soma_Y, _ = calcular_media(Y) # n é o mesmo para X e Y

print(f"Soma dos valores de X: {soma_X}")
print(f"Número de observações (n): {n}")
print(f"Média de X (consumo de cigarros): {media_X:.2f}")

print(f"Soma dos valores de Y: {soma_Y}")
print(f"Média de Y (mortes por câncer): {media_Y:.2f}")

# Calcular covariância entre X e Y:
# Covariância(X,Y) = Σ((X_i - media_X) * (Y_i - media_Y)) / n
termos_covariancia = []
for i in range(n):
    termo = (X[i] - media_X) * (Y[i] - media_Y)
    termos_covariancia.append(termo)

soma_prod_desvios_XY = sum(termos_covariancia)
cov_XY = soma_prod_desvios_XY / n
print(f"Soma dos produtos dos desvios (Σ((X_i - media_X) * (Y_i - media_Y))): {soma_prod_desvios_XY:.2f}")
print(f"Covariância entre X e Y: {cov_XY:.2f}")

# Calcular variância de X:
# Variância(X) = Σ((X_i - media_X)^2) / n
termos_variancia_X = []
for x_i in X:
    termo = (x_i - media_X) ** 2
    termos_variancia_X.append(termo)
    # print(f"  Termo para variância de X ((x_i - media_X)^2): ({x_i} - {media_X:.2f})^2 = {termo:.2f}")
soma_quadrados_desvios_X = sum(termos_variancia_X)
var_X = soma_quadrados_desvios_X / n
print(f"Soma dos quadrados dos desvios de X (Σ((X_i - media_X)^2)): {soma_quadrados_desvios_X:.2f}")
print(f"Variância de X: {var_X:.2f}")

# Calcular variância de Y:
# Variância(Y) = Σ((Y_i - media_Y)^2) / n
```

```
termos_variancia_Y = []
for y_i in Y:
    termo = (y_i - media_Y) ** 2
    termos_variancia_Y.append(termo)
    # print(f"  Termo para variância de Y ((y_i - media_Y)^2): ({y_i} - {media_Y:.2f})^2 = {termo:.2f}")
soma_quadrados_desvios_Y = sum(termos_variancia_Y)
var_Y = soma_quadrados_desvios_Y / n
print(f"Soma dos quadrados dos desvios de Y (Σ((Y_i - media_Y)^2)): {soma_quadrados_desvios_Y:.2f}")
print(f"Variância de Y: {var_Y:.2f}")

# Calcular desvio padrão de X e Y:
# DesvioPadrão(X) = sqrt(Variância(X))
std_X = var_X ** 0.5
std_Y = var_Y ** 0.5
print(f"Desvio Padrão de X: {std_X:.2f}")
print(f"Desvio Padrão de Y: {std_Y:.2f}")

# Calcular coeficiente de correlação de Pearson (r)
# r = Covariância(X,Y) / (DesvioPadrão(X) * DesvioPadrão(Y))
r = cov_XY / (std_X * std_Y)
print(f"Coeficiente de Correlação de Pearson (r): {r:.4f}")

# Interpretação de r:
# - O valor de r está entre -1 e 1.
# - Próximo de 1: correlação linear positiva forte.
# - Próximo de -1: correlação linear negativa forte.
# - Próximo de 0: correlação linear fraca ou inexistente.
# - Sinal positivo: indica que quando X aumenta, Y tende a aumentar.
# - Sinal negativo: indica que quando X aumenta, Y tende a diminuir.
if r > 0.7:
    interpretacao_r = "Correlação linear positiva forte."
elif r > 0.4:
    interpretacao_r = "Correlação linear positiva moderada."
elif r > 0:
    interpretacao_r = "Correlação linear positiva fraca."
elif r < -0.7:
    interpretacao_r = "Correlação linear negativa forte."
elif r < -0.4:
    interpretacao_r = "Correlação linear negativa moderada."
elif r < 0:
    interpretacao_r = "Correlação linear negativa fraca."
else:
    interpretacao_r = "Correlação linear muito fraca ou inexistente."
print(f"Interpretação de r: {interpretacao_r} Indica que há uma tendência de aumento no número de mortes por câncer à medida que o consumo de
print("-" * 50)

# (d) Obtenha a equação de regressão que melhor descreve a associação entre X e Y. Interprete os valores de b0 e b1.
# Equação de regressão linear: Y_estimado = b0 + b1*X
# b1 (coeficiente angular) = Covariância(X,Y) / Variância(X)
# b0 (intercepto) = media_Y - b1 * media_X
print("\n(d) Equação de Regressão Linear (Y = b0 + b1*X):")

b1 = cov_XY / var_X
print(f"Coeficiente angular (b1): {b1:.4f}")

b0 = media_Y - (b1 * media_X)
print(f"Intercepto (b0): {b0:.4f}")

print(f"Equação de regressão linear estimada: Y_estimado = {b0:.2f} + {b1:.2f}*X")

# Interpretação de b0 e b1:
# b1: Indica a variação esperada em Y para cada unidade de aumento em X.
#     Neste caso, para cada unidade adicional no consumo de cigarros per capita em 1930,
#     espera-se um aumento de b1 no número de mortes por câncer por 1.000.000 de habitantes em 1950.
print(f"Interpretação de b1: Para cada unidade a mais no consumo de cigarros per capita em 1930, estima-se um aumento de {b1:.2f} mortes por
# b0: É o valor estimado de Y quando X é igual a 0.
#     Pode ou não ter interpretação prática dependendo do contexto (X=0 ser um valor factível e dentro do escopo dos dados).
#     Neste caso, seria o número estimado de mortes por câncer se o consumo de cigarros fosse zero.
print(f"Interpretação de b0: Se o consumo de cigarros per capita em 1930 fosse zero, o número estimado de mortes por câncer (por 1.000.000 hal
print("-" * 50)


# (e) Se, no ano de 1930, o consumo de cigarros per capita no Brasil foi 630, estime o número de mortes causadas por câncer de pulmão no ano
print("\n(e) Estimativa de mortes para o Brasil:")
consumo_brasil_1930 = 630
print(f"Consumo de cigarros per capita no Brasil em 1930 (X_brasil): {consumo_brasil_1930}")
```

```
# Usar a equação de regressão: Y_estimado_brasil = b0 + b1 * X_brasil
mortes_estimadas_brasil_1950 = b0 + b1 * consumo_brasil_1930
print(f"Número estimado de mortes por câncer de pulmão no Brasil em 1950 (Y_estimado_brasil): {mortes_estimadas_brasil_1950:.2f} por 1.000.00
print("-" * 50)

# (f) Calcule e interprete o coeficiente de determinação R² para a equação de regressão estimada no item (d).
# R² (R-quadrado) = (Coeficiente de Correlação de Pearson (r))^2
print("\n(f) Coeficiente de Determinação (R²):")
R2 = r ** 2
print(f"Coeficiente de Determinação (R²): {R2:.4f}")

# Interpretação de R²:
# R² varia de 0 a 1 e representa a proporção da variância total da variável dependente (Y)
# que é explicada pela variável independente (X) através do modelo de regressão linear.
# Um R² de 0.5531 significa que aproximadamente 55.31% da variação nas mortes por câncer
# pode ser explicada pelo consumo de cigarros, de acordo com este modelo.
print(f"Interpretação de R²: Aproximadamente {R2*100:.2f}% da variabilidade no número de mortes por câncer em 1950 pode ser explicada pelo co
print("-" * 50)

# Exibição dos resultados consolidados (semelhante ao resultado [3])
print("\nResultados Consolidados:")
print(f"  Coeficiente de Correlação de Pearson (r): {r:.4f}")
print(f"  Equação de Regressão: Y = {b0:.2f} + {b1:.2f}*X")
print(f"  Mortes Estimadas para o Brasil (consumo=630): {mortes_estimadas_brasil_1950:.2f}")
print(f"  Coeficiente de Determinação (R²): {R2:.4f} (ou {R2*100:.2f}%)")
```

```
(a) Escolha de X e Y:
Variável Independente (X): Consumo de cigarros per capita em 1930 = [240, 255, 340, 375, 510, 490, 490, 180, 1125, 1150, 1275]
Variável Dependente (Y): Mortes por câncer no pulmão em 1950 = [63, 100, 140, 175, 160, 180, 250, 180, 360, 470, 200]
--------------------------------------------------

(b) Diagrama de Dispersão:
```




```
--------------------------------------------------

(c) Coeficiente de Correlação Linear de Pearson (r):
Soma dos valores de X: 6430
Número de observações (n): 11
Média de X (consumo de cigarros): 584.55
Soma dos valores de Y: 2278
Média de Y (mortes por câncer): 207.09
Soma dos produtos dos desvios (Σ((X_i - media_X) * (Y_i - media_Y))): 347450.45
Covariância entre X e Y: 31586.40
Soma dos quadrados dos desvios de X (Σ((X_i - media_X)^2)): 1606672.73
Variância de X: 146061.16
Soma dos quadrados dos desvios de Y (Σ((Y_i - media_Y)^2)): 135840.91
Variância de Y: 12349.17
```