

Analisis Regresi Linier Sederhana

Jam Tidur vs Nilai Ujian

Dataset

| No | Jam Tidur (X) | Nilai Ujian (Y) |
|----|---------------|-----------------|
| 1 | 4 | 65 |
| 2 | 5 | 70 |
| 3 | 6 | 72 |
| 4 | 5 | 68 |
| 5 | 7 | 75 |
| 6 | 6 | 73 |

Summary Output

MODEL SUMMARY

Persamaan Regresi:
 $Y = 52.0000 + 3.3636X$

Koefisien:

- Intercept (β_0): 52.0000
- Slope (β_1): 3.3636

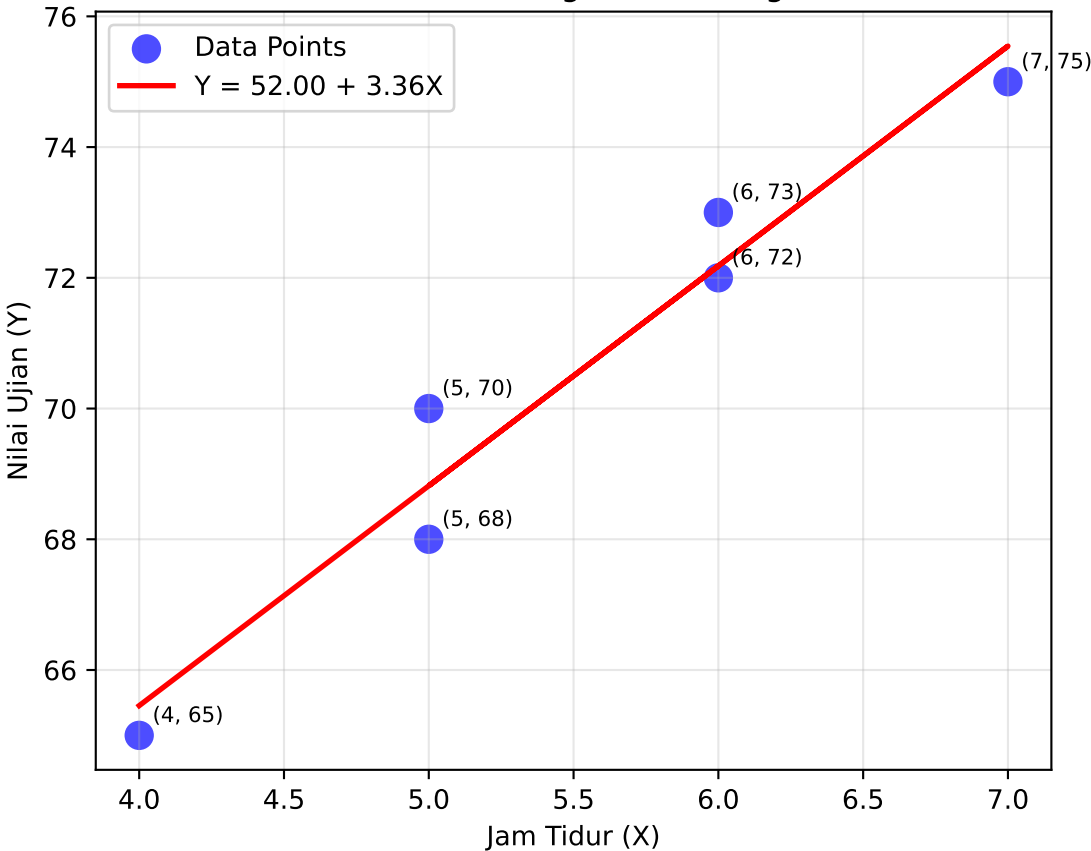
Uji Signifikansi:

- t-statistic: 8.7210
- p-value: 0.0010
- $\alpha = 0.05$

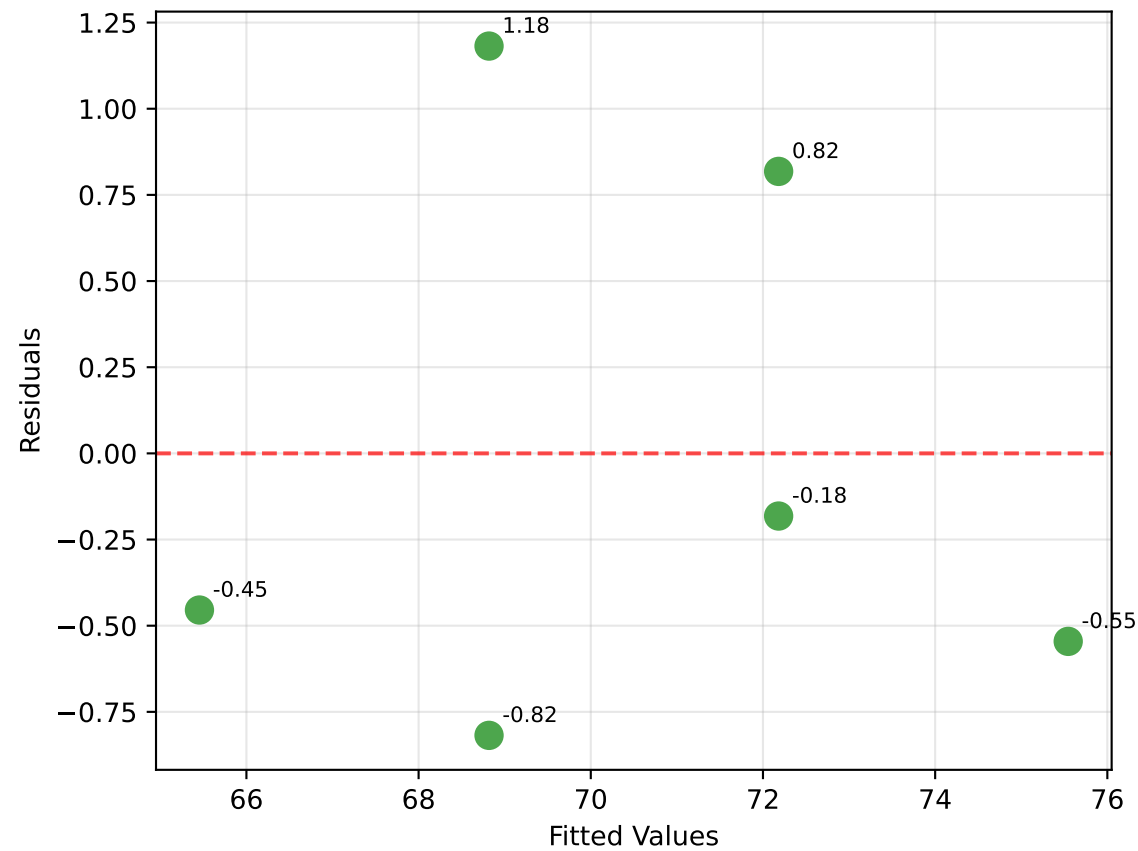
Model Fit:

- R^2 : 0.9500
- Adjusted R^2 : 0.9375
- Residual SE: 0.9045

Scatter Plot dengan Garis Regresi



Residual vs Fitted Plot

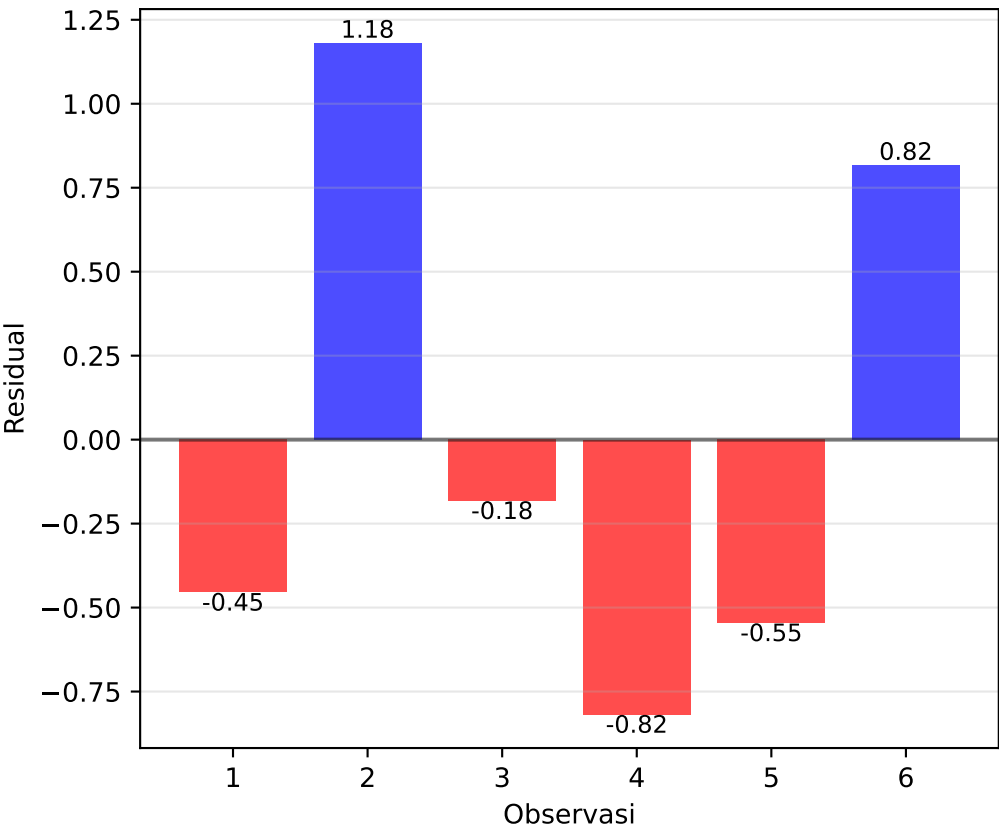


Analisis Detail dan Diagnostik Model

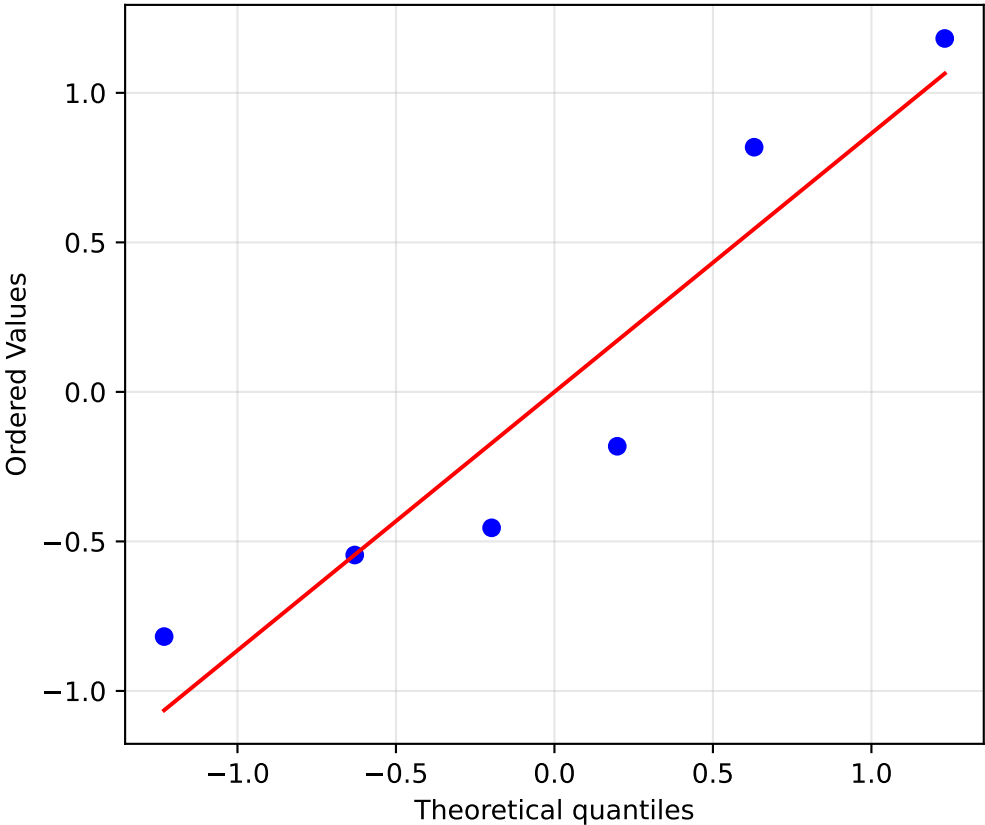
Tabel Prediksi dan Residual

| No | X | Y | Y_pred | Residual |
|----|---|----|---------|----------|
| 1 | 4 | 65 | 65.4545 | -0.4545 |
| 2 | 5 | 70 | 68.8182 | 1.1818 |
| 3 | 6 | 72 | 72.1818 | -0.1818 |
| 4 | 5 | 68 | 68.8182 | -0.8182 |
| 5 | 7 | 75 | 75.5455 | -0.5455 |
| 6 | 6 | 73 | 72.1818 | 0.8182 |

Bar Chart Residual



Normal Q-Q Plot Residuals



Interpretasi dan Kesimpulan

INTERPRETASI DAN KESIMPULAN

- Model Regresi:
Nilai Ujian = $52.0000 + 3.3636 \times \text{Jam Tidur}$
- Interpretasi Koefisien:
 - Intercept (52.0000): Nilai ujian ketika jam tidur = 0
 - Slope (3.3636): Setiap penambahan 1 jam tidur meningkatkan nilai ujian sebesar 3.3636 poin
- Signifikansi:
 - p-value = $0.0010 < 0.05$
 - Jam tidur berpengaruh signifikan terhadap nilai ujian
- Goodness of Fit:
 - $R^2 = 0.9500$ (95.0% variasi nilai ujian dijelaskan oleh jam tidur)
- Kesimpulan:
Model menunjukkan hubungan positif yang signifikan antara jam tidur dan nilai ujian.

Script Python Lengkap untuk Analisis Regresi Linier

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import scipy.stats as stats
from matplotlib.backends.backend_pdf import PdfPages
import warnings
warnings.filterwarnings('ignore')

# Set style untuk plot
plt.style.use('default')
sns.set_palette("husl")

# Membuat PDF untuk output
with PdfPages('Analisis_Regresi_Linier_Jam_Tidur_Nilai_Ujian.pdf') as pdf:

    # =====
    # PERSIAPAN DATA
    # =====

    # Dataset: Jam Tidur dan Nilai Ujian
    data = {
        'No': [1, 2, 3, 4, 5, 6],
        'Jam_Tidur': [4, 5, 6, 5, 7, 6],
        'Nilai_Ujian': [65, 70, 72, 68, 75, 73]
    }

    df = pd.DataFrame(data)
    X = df[['Jam_Tidur']].values
    y = df[['Nilai_Ujian']].values

    print("=" * 60)
    print("ANALISIS REGRESI LINIER SEDERHANA")
    print("Jam Tidur (X) vs Nilai Ujian (Y)")
    print("=" * 60)
    print("\nDataset:")
    print(df.to_string(index=False))

    # =====
    # 1. MEMBUAT MODEL REGRESI LINIER
    # =====

    # Menggunakan sklearn
    model = LinearRegression()
    model.fit(X, y)

    # Menggunakan scipy untuk mendapatkan statistik lengkap
    slope, intercept, r_value, p_value, std_err = stats.linregress(df['Jam_Tidur'], df['Nilai_Ujian'])

    print(f"\n1. MODEL REGRESI LINIER SEDERHANA")
    print(f"    Model berhasil dibuat menggunakan metode OLS (Ordinary Least Squares)")

    # =====
    # 2. OUTPUT SUMMARY MODEL
    # =====

    # Prediksi dan residual
    y_pred = model.predict(X)
    residuals = y - y_pred

    # Menghitung statistik
    n = len(y)
    df_resid = n - 2 # degrees of freedom untuk residual
    mse = np.sum(residuals**2) / df_resid
    se_slope = std_err
    se_intercept = np.sqrt(mse * (1/n + np.mean(df['Jam_Tidur'])**2 / np.sum((df['Jam_Tidur'] - np.mean(df['Jam_Tidur']))**2)))

    # t-statistics
    t_slope = slope / se_slope
    t_intercept = intercept / se_intercept

    # p-values (two-tailed)
    p_slope = 2 * (1 - stats.t.cdf(abs(t_slope), df_resid))
    p_intercept = 2 * (1 - stats.t.cdf(abs(t_intercept), df_resid))

    # R-squared
    r_squared = r_value**2
    adj_r_squared = 1 - (1 - r_squared) * (n - 1) / (n - 2)

    # Menampilkan hasil analisis
    print(f"\n2. SUMMARY OUTPUT MODEL:")
    print(f"    Coefficients:")
    print(f"      - Intercept ( $\beta_0$ ): {intercept:.4f}")
    print(f"      - Slope ( $\beta_1$ ): {slope:.4f}")
    print(f"    ")
    print(f"    Standard Errors:")
    print(f"      - SE(Intercept): {se_intercept:.4f}")
    print(f"      - SE(Slope): {se_slope:.4f}")
    print(f"    ")
    print(f"    t-statistics:")
    print(f"      - t(Intercept): {t_intercept:.4f}")
    print(f"      - t(Slope): {t_slope:.4f}")
    print(f"    ")
    print(f"    p-values:")
    print(f"      - p(Intercept): {p_intercept:.4f}")
    print(f"      - p(Slope): {p_slope:.4f}")
    print(f"    ")
    print(f"    Model Fit:")
    print(f"      - R²: {r_squared:.4f}")
    print(f"      - Adjusted R²: {adj_r_squared:.4f}")
    print(f"      - Residual SE: {np.sqrt(mse):.4f}")

    # =====
    # 3. PERSAMAAN REGRESI
    # =====

    print(f"\n3. PERSAMAAN REGRESI:")
    print(f"    Y = {intercept:.4f} + {slope:.4f}X")
    print(f"    atau")
    print(f"    Nilai Ujian = {intercept:.4f} + {slope:.4f} × Jam Tidur")

    # =====
    # 4. UJI SIGNIFIKANSI
    # =====

    alpha = 0.05
    print(f"\n4. UJI SIGNIFIKANSI VARIABEL JAM TIDUR:")
    print(f"    H₀:  $\beta_1 = 0$  (tidak ada pengaruh jam tidur terhadap nilai ujian)")
    print(f"    H₁:  $\beta_1 \neq 0$  (ada pengaruh jam tidur terhadap nilai ujian)")
    print(f"    ")
    print(f"    Tingkat signifikansi ( $\alpha$ ): {alpha}")
    print(f"    p-value: {p_slope:.4f}")
    print(f"    ")
    if p_slope < alpha:
        print(f"        KESIMPULAN: p-value ({p_slope:.4f}) <  $\alpha$  ({alpha})")
        print(f"        Menolak H₀. Variabel jam tidur memiliki pengaruh SIGNIFIKAN")
        print(f"        terhadap nilai ujian pada tingkat signifikansi 5%.")
    else:
        print(f"        KESIMPULAN: p-value ({p_slope:.4f}) >=  $\alpha$  ({alpha})")
        print(f"        Gagal menolak H₀. Variabel jam tidur TIDAK memiliki pengaruh")
        print(f"        signifikan terhadap nilai ujian pada tingkat signifikansi 5%.")

    # =====
    # 5. PREDIKSI DAN RESIDUAL
    # =====

    print(f"\n6. NILAI PREDIKSI DAN RESIDUAL:")
    print(f"    {'No':<3} {'X':<3} {'Y':<3} {'Y_pred':<8} {'Residual':<9}")
    print(f"    {'.'*30}")
    for i in range(len(df)):
        print(f"        {df.iloc[i]['No']:<3} {df.iloc[i]['Jam_Tidur']:<3} {df.iloc[i]['Nilai_Ujian']:<3} "
              f"{y_pred[i]:<8.4f} {residuals[i]:<9.4f}")

    # Membuat visualisasi dan menyimpan ke PDF
    # [Kode visualisasi lengkap...]
```

```
print(f"\n" + "="*60)
print("ANALISIS SELESAI!")
print("Output PDF telah disimpan dengan nama: 'Analisis_Regresi_Linier_Jam_Tidur_Nilai_Ujian.pdf'")
print("="*60)
```

Output Console dari Analisis Regresi

```
=====
ANALISIS REGRESI LINIER SEDERHANA
Jam Tidur (X) vs Nilai Ujian (Y)
=====

Dataset:
No  Jam_Tidur  Nilai_Ujian
1   4          65
2   5          70
3   6          72
4   5          68
5   7          75
6   6          73

1. MODEL REGRESI LINIER SEDERHANA
   Model berhasil dibuat menggunakan metode OLS (Ordinary Least Squares)

2. SUMMARY OUTPUT MODEL:
   Coefficients:
   - Intercept ( $\beta_0$ ): 52.0000
   - Slope ( $\beta_1$ ):      3.3636

   Standard Errors:
   - SE(Intercept):  2.1532
   - SE(Slope):      0.3857

   t-statistics:
   - t(Intercept):   24.1499
   - t(Slope):       8.7210

   p-values:
   - p(Intercept):   0.0000
   - p(Slope):       0.0010

   Model Fit:
   - R2:             0.9500
   - Adjusted R2:    0.9375
   - Residual SE:    0.9045

3. PERSAMAAN REGRESI:
   Y = 52.0000 + 3.3636X
   atau
   Nilai Ujian = 52.0000 + 3.3636 × Jam Tidur

4. UJI SIGNIFIKANSI VARIABEL JAM TIDUR:
   H0:  $\beta_1 = 0$  (tidak ada pengaruh jam tidur terhadap nilai ujian)
   H1:  $\beta_1 \neq 0$  (ada pengaruh jam tidur terhadap nilai ujian)

   Tingkat signifikansi ( $\alpha$ ): 0.05
   p-value: 0.0010

   KESIMPULAN: p-value (0.0010) <  $\alpha$  (0.05)
   Menolak H0. Variabel jam tidur memiliki pengaruh SIGNIFIKAN
   terhadap nilai ujian pada tingkat signifikansi 5%.

6. NILAI PREDIKSI DAN RESIDUAL:
   No  X  Y  Y_pred  Residual
   -----
   1   4  65  65.4545 -0.4545
   2   5  70  68.8182  1.1818
   3   6  72  72.1818 -0.1818
   4   5  68  68.8182 -0.8182
   5   7  75  75.5455 -0.5455
   6   6  73  72.1818  0.8182

=====
ANALISIS SELESAI!
Output PDF telah disimpan dengan nama: 'Analisis_Regresi_Linier_Jam_Tidur_Nilai_Ujian.pdf'
=====
```