

# Soal & Pembahasan

---

## 0.0.3 Berikan ulasan Anda tentang model tersebut

**Keterangan mengenai definisi, komponen, dan pengujian signifikansi** (pada tugas sebelumnya)

**Persamaan regresi linear:**

$$Y = a + bX$$

ket.

- $a$  = intersep (nilai  $Y$  ketika  $X = 0$ ) => intercept
- $b$  = koefisien regresi (kemiringan/gradien) => slope
- $Y$  = variabel dependen (perkembangan penyakit)
- $X$  = variabel independen (usia)

**p-value:**

*nilai probabilitas* yang digunakan pada pengujian hipotesis. Nilai ini digunakan untuk mengukur apakah hubungan tersebut signifikan secara statistik atau hanya terjadi karena kebetulan.

- $p < 0.05$ : Hubungan antara variabel  $X$  dan  $Y$  \*signifikan secara statistik\*. Artinya, peluang hubungan tersebut terjadi secara kebetulan adalah kurang dari 5%.
- $p > 0.05$ : Tidak cukup bukti untuk menyatakan hubungan signifikan. Artinya ada indikasi bahwa lebih dari 5% hubungan yang ada terjadi secara kebetulan.

**stderr (standard error):**

*ukuran variabilitas* dalam estimasi parameter, sebagai contoh nilai slope ( $b$ ). Nilai stderr ini memberi gambaran seberapa luas kemungkinan estimasi parameter bisa berubah jika data yang digunakan berbeda (misalnya, dalam pengambilan sampel data ulang).

- Jika nilai stderr kecil, berarti estimasi parameter stabil dan dapat dipercaya (karena variasi parameternya kecil).
- Jika nilai stderr besar, berarti estimasi parameter kurang akurat, sehingga hasil regresi menjadi diragukan.

Didapatkan **persamaan regresinya** adalah:

$$Y = -94.10 + 8.96X$$

persamaan ini didapatkan dari penggabungan beberapa variabel independen (yaitu **BMI**, **AGE**, **BP**, dan **S1**) untuk memprediksi variabel dependen (yaitu **Y**) menggunakan multiple linear regression. Hasil persamaan regresi yang sebenarnya adalah:

$$Y = -94.10 + 8.96 * BMI + (-0.0748 * AGE) + 0.8582 * BP + (-0.1886 * S1)$$

$$Y = -94.10 + 8.96 * BMI - 0.0748 * AGE + 0.8582 * BP - 0.1886 * S1$$

$$Y = -94.10 + 8.96 * BMI$$

$$Y = -94.10 + 8.96 * X$$

Hanya saja, karena variabel **BMI** menjadi satu-satunya variabel yang memberikan pengaruh yang signifikan (dilihat dari nilai intercept nya), maka variabel sisanya (**AGE**, **BP**, dan **S1**) tidak ikut dimasukkan. Hal ini dikarenakan ketiganya tidak memberikan pengaruh signifikan meskipun tetap digunakan dalam persamaan.

Keterangan:

#### 1. Intercept dan Koefisien (Koefisien Regresi)

- **Intercept (-94.10):** Nilai ini menunjukkan bahwa jika semua variabel independen (**AGE**, **BMI**, **BP**, dan **S1**) adalah 0, maka nilai perkembangan penyakit (**Y**) diperkirakan -94.10 (meskipun nilai ini tidak realistis, karena kita tidak mungkin memiliki nilai 0 untuk variabel seperti **BMI**).

- **Koefisien BMI (8.96):** Koefisien untuk **BMI** menunjukkan bahwa setiap kenaikan satu unit pada **BMI** akan meningkatkan perkembangan penyakit (**Y**) sebesar **8.96** satuan. Berarti semakin tinggi **BMI**, semakin tinggi perkembangan penyakitnya (menunjukkan pengaruh positif **BMI** terhadap **Y**, yang ).

Selanjutnya untuk:

### 1. Signifikansi Koefisien

- a. **p-value untuk BMI = 0.0000:** Menunjukkan bahwa koefisien **BMI** sangat signifikan, karena p-value lebih kecil dari 0.05 (ada hubungan yang kuat dan signifikan antara **BMI** dan **Y**).
- b. **p-value untuk AGE (0.880):** Menunjukkan bahwa **AGE** tidak memiliki pengaruh signifikan terhadap **Y**.
- c. **p-value untuk BP (0.100):** Meskipun p-value untuk **BP** mendekati 0.05, hasil ini masih menunjukkan bahwa **BP** tidak secara signifikan mempengaruhi **Y** (mungkin masih memiliki pengaruh, tetapi tidak cukup kuat).
- d. **p-value untuk S1 (0.378):** Pengaruh **S1** juga tidak signifikan terhadap **Y**. Nilai p-value yang lebih besar dari 0.05 menunjukkan bahwa **S1** tidak memiliki hubungan yang cukup kuat dengan **Y** dalam analisis.

### 2. R-squared dan Nilai Adjusted R-squared

- a. **R-squared (0.270):** Nilai ini menunjukkan bahwa model ini hanya dapat menjelaskan sekitar **27%** variasi dalam perkembangan penyakit (**Y**). Artinya ada banyak faktor lain yang mempengaruhi **Y**, selain dari variabel yang digunakan dalam model.
- b. **Adjusted R-squared (0.239):** Nilai yang mengoreksi R-squared dengan mempertimbangkan jumlah variabel independen. Nilai yang lebih rendah dari R-squared menandakan bahwa model mungkin tidak terlalu baik dalam menjelaskan variasi dalam **Y**, dan variabel lain yang tidak dimasukkan dalam model bisa jadi berpengaruh.

### 3. Standar Error

**Standard Error (1.58):** Nilai ini mengukur sejauh mana estimasi koefisien regresi dapat bervariasi. Nilai yang lebih kecil menunjukkan estimasi koefisien yang lebih akurat. Maka dari itu model regresi ini bisa dibilang cukup akurat.

### 4. F-statistic dan p-value F-statistic

**F-statistic (8.768):** Ini menguji apakah model secara keseluruhan signifikan dalam memprediksi **Y**. Karena nilai **p-value untuk F-statistic (4.55e-06)** sangat kecil, berarti model regresi secara keseluruhan adalah signifikan, meski hanya variabel **BMI** yang memberi kontribusi signifikan terhadap model.

Kesimpulan:

Berdasarkan hasil analisis regresi linear berganda, diperoleh persamaan regresi sebagai berikut:

$$Y = -94.10 + 8.96 * X$$

Meskipun persamaan awal melibatkan variabel **AGE**, **BP**, dan **S1**, hanya variabel **BMI** saja yang memiliki pengaruh signifikan terhadap perkembangan penyakit (**Y**) karena p-value BMI (0.0000) lebih kecil dari 0.05 dan paling kecil dibandingkan variabel lain. Oleh karena itu, hanya **BMI** yang dipertahankan dalam persamaan akhir. Meskipun modelnya signifikan secara keseluruhan, tetap saja kemampuan model ini dalam menjelaskan variasi **Y** hanya sebesar **27%**, sehingga perlu mempertimbangkan faktor lain dalam model prediksi.

# Sumber Referensi

---

Sutikno., & Ratnaningsih, D. J. (2025). Metode Statistika I. Modul 01 & 02. Tangerang, Banten. Universitas Terbuka.

Wijaya, T., & Budiman, S. (2016). Analisis multivariat untuk penelitian manajemen. Yogyakarta: Pohon Cahaya.

<https://learn.nural.id/course/statistics/regresi-linier/scatterplot#:~:text=Scatterplot%20adalah%20sebuah%20grafik%20yang,merepresentasikan%20nilai%20x%20dan%20y.>

[https://www.investopedia.com/terms/m/mlr.asp#:~:text=Multiple%20linear%20regression%20\(MLR\)%20is,uses%20just%20one%20explanatory%20variable.](https://www.investopedia.com/terms/m/mlr.asp#:~:text=Multiple%20linear%20regression%20(MLR)%20is,uses%20just%20one%20explanatory%20variable.)

[https://www.w3schools.com/python/python\\_ml\\_multiple\\_regression.asp](https://www.w3schools.com/python/python_ml_multiple_regression.asp)