

# Assignment Guidance: Descriptive Statistics

Data Science & Data Analyst Bootcamp

## Periode Pembelajaran

Descriptive Statistics

## Objectives

Setelah menyelesaikan tugas ini, student diharapkan mampu:

1. Melakukan perhitungan statistik deskriptif dasar menggunakan Python
2. Menafsirkan makna statistik seperti mean, standard deviation, covariance, correlation, dan expected value
3. Menggunakan Z-score untuk mengevaluasi posisi nilai dalam distribusi data
4. Mengembangkan keterampilan berpikir analitis dari data numerik nyata


## Deskripsi Assignment

Dalam tugas ini, student diminta untuk menerapkan konsep-konsep dasar **statistik deskriptif** menggunakan dataset nyata: **Wine Quality (Red Wine)**. Tugas ini dirancang agar student tidak hanya mampu melakukan perhitungan statistik, tetapi juga mampu **menginterpretasikan** hasilnya dalam konteks data yang tersedia.

Student akan melakukan analisis terhadap data wine berdasarkan **ukuran pemusatan** (mean, median, mode), **penyebaran** (variance, standard deviation), **hubungan antar variabel** (covariance & correlation), serta **analisis posisi nilai** menggunakan **Z-score**. Student juga akan menghitung **expected value** berdasarkan tabel probabilitas kualitas wine yang disediakan.

## Detail Assignment

**Dataset yang digunakan:**

- [Red Wine Quality](#)
-  Tabel Distribusi Probabilitas Kualitas Wine

Untuk menyelesaikan assignment ini, lakukan langkah-langkah berikut:

1. **Unduh** dan **load** dataset ke notebook.
2. **Lakukan analisis** sesuai dengan bagian-bagian berikut:
  - a. **Bagian 1: Statistik Deskriptif & Eksplorasi Awal**
    - i. Tampilkan 5 baris pertama, informasi dataset, dan ringkasan statistik
    - ii. Hitung dan interpretasikan:
      1. Mean, median, dan mode untuk `alcohol`, `residual sugar`, dan `quality`
      2. Variance dan standard deviation untuk ketiga kolom tersebut
    - iii. Hitung jumlah dan persentase data yang:
      1. `alcohol > mean(alcohol)`
      2. `residual sugar > mean(residual sugar)`
      3. `quality > mean(quality)`
    - iv. Analytical thinking prompts:
      1. Kolom mana yang nilai-nilainya paling menyebar?
      2. Apakah data lebih banyak yang berada di atas atau di bawah rata-rata?
      3. Apakah data simetris atau cenderung skewed?
      4. Tambahkan pertanyaan analisis lainnya dan tuliskan insight-nya!
  - b. **Bagian 2: Analisis Hubungan antar Variabel**
    - i. Hitung covariance dan correlation:
      1. `alcohol` dan `quality`
      2. `residual sugar` dan `quality`
      3. *Tambahkan fitur lainnya yang menurutmu menarik*
    - ii. Tambahkan perhitungan manual untuk 1 pasang data  $((x_i, y_i))$  sebagai ilustrasi
    - iii. Analytical thinking prompts:
      1. Apakah hubungan antar fitur tersebut positif atau negatif? Kuat atau lemah?
      2. Apa implikasinya terhadap kualitas wine?
      3. Tambahkan pertanyaan analisis lainnya dan tuliskan insight-nya!
  - c. **Bagian 3: Z-Score Analysis**
    - i. Hitung Z-score untuk seluruh data pada kolom `quality`.
    - ii. Identifikasi:
      1. Berapa banyak wine yang memiliki **Z-score > 1.5 (high quality wines)**?
      2. Berapa banyak yang **Z-score < -1.5 (low quality wines)**?
      3. Berapa persentasenya terhadap seluruh data?
    - iii. Ambil subset data dengan **High Quality Wines** dan **Low Quality Wines**

- iv. Bandingkan statistik deskriptif untuk masing-masing subset terhadap fitur-fitur berikut:
  1. alcohol
  2. volatile acidity
  3. residual sugar
  4. sulphates
  5. Tambahkan fitur lainnya yang menurutmu menarik
- v. Analytical thinking prompts:
  1. Apakah kadar alkohol cenderung lebih tinggi di wine berkualitas tinggi?
  2. Apakah volatile acidity lebih rendah di wine berkualitas bagus?
  3. Fitur apa yang paling konsisten membedakan kualitas? Apa yang membuat wine berkualitas tinggi atau rendah?
  4. Tambahkan pertanyaan analisis lain dan tuliskan insight-nya!

**d. Bagian 4: Expected Value dari Quality**

- i. Gunakan tabel probabilitas **quality** yang sudah disediakan
- ii. Hitung Expected Value dari **quality**
- iii. Bandingkan hasilnya dengan mean aktual **quality** di dataset
- iv. Analytical thinking prompts:
  1. Apa interpretasi dari expected value tersebut?
  2. Apakah hasil expected value mendekati mean aktual?
  3. Apa artinya jika keduanya berbeda?
  4. Tambahkan pertanyaan analisis lain dan tuliskan insight-nya!
3. **Tambahkan interpretasi** atau **insight** setelah **setiap bagian**. Gunakan pertanyaan pada analytical thinking prompts sebagai panduan untuk menafsirkan hasil analisis.
4. **Buat kesimpulan akhir** berdasarkan temuanmu.
5. Pastikan notebook-nya **rapi!** (akan mendapatkan nilai bonus)

## Pengumpulan Assignment

**Deadline :**

Maksimal H+7 Kelas (Pukul 23.30 WIB)

**Details :**

Dikumpulkan dalam bentuk Link Google Colaboratory, secara INDIVIDU, di LMS

## Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Statistik Deskriptif &	Mampu menampilkan statistik deskriptif,	25

	Eksplorasi Awal	perbandingan terhadap rata-rata, dan interpretasinya	
2	Analisis Hubungan antar Variabel	Mampu menghitung kovarians dan korelasi antar fitur, serta menjelaskan arah dan kekuatan hubungan antar variabel terhadap kualitas	20
3	Z-score	Mampu menghitung Z-score untuk quality, mengidentifikasi nilai ekstrem, dan membandingkan ciri-cirinya	25
4	Expected Value	Mampu menghitung expected value berdasarkan probabilitas, membandingkan dengan mean aktual, dan menginterpretasikannya	20
5	Insight & Kesimpulan Akhir	Mampu menyusun insight akhir berdasarkan keseluruhan hasil analisis statistik	10

## Referensi/Template

1. <https://faculty.washington.edu/otoomet/machinelearning-py/descriptive-analysis-with-pandas.html>
2. [https://snakebear.science/O8-Statistics/O8\\_1\\_Basic\\_Statistics\\_Descriptives.html](https://snakebear.science/O8-Statistics/O8_1_Basic_Statistics_Descriptives.html)
3. <https://medium.com/data-science/statistics-in-python-understanding-variance-covariance-and-correlation-4729b528db01>
4. <https://www.machinelearningplus.com/probability/expected-value/>
- 5.

### Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

### Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1 x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor