

■ Python Statistics Cheat Sheet

1. Descriptive Statistics

Metode	Library / Fungsi	Tipe Data	Contoh Kasus
Mean (Rata-rata)	numpy.mean(), pandas.DataFrame.mean()	Numerik	Rata-rata tinggi badan siswa
Median	numpy.median(), pandas.DataFrame.median()	Numerik	Nilai tengah pendapatan per bulan
Mode (Modus)	scipy.stats.mode()	Numerik / Kategorikal	Nilai paling sering muncul dalam data penjualan
Standard Deviation	numpy.std(), pandas.DataFrame.std()	Numerik	Sebaran tinggi badan
Variance	numpy.var(), pandas.DataFrame.var()	Numerik	Variabilitas berat badan
Min / Max / Sum / Count	numpy.min()/max()/sum(), pandas.min()/max()/sum()/count()	Numerik	Total penjualan, nilai minimum/maximum

2. Probability Distributions

Diskret

Distribusi	Library / Fungsi	Contoh Kasus
Binomial	scipy.stats.binom	Peluang sukses 7 kepala dari 10 lemparan koin
Poisson	scipy.stats.poisson	Jumlah email yang masuk per jam
Geometric	scipy.stats.geom	Percobaan pertama sukses

Kontinu

Distribusi	Library / Fungsi	Contoh Kasus
Normal / Gaussian	scipy.stats.norm	Distribusi tinggi badan, berat badan
Eksponensial	scipy.stats.expon	Waktu tunggu kedatangan bus berikutnya
Uniform	scipy.stats.uniform	Semua nilai dalam rentang sama kemungkinannya
t-Student	scipy.stats.t	Uji rata-rata sampel kecil
Chi-Square	scipy.stats.chi2	Membandingkan distribusi kategori
F-Distribution	scipy.stats.f	Uji variansi antar kelompok

3. Inferential Statistics

Metode	Library / Fungsi	Contoh Kasus
t-test (1 sample, 2 sample)	scipy.stats.ttest_1samp() / ttest_ind()	Membandingkan rata-rata grup
ANOVA	scipy.stats.f_oneway()	Membandingkan rata-rata 3 grup atau lebih
Chi-square test	scipy.stats.chi2_contingency()	Uji hubungan antar kategori
Correlation	pandas.DataFrame.corr(), scipy.stats.pearsonr()	Hubungan tinggi badan & berat badan
Regression (Linear)	sklearn.linear_model.LinearRegression	Prediksi harga rumah dari luas bangunan

4. Sampling & Resampling

Metode	Library / Fungsi	Contoh Kasus
Random sampling	<code>numpy.random.choice()</code> , <code>pandas.sample()</code>	Ambil sampel acak dari populasi
Bootstrapping	<code>scipy.stats.bootstrap()</code>	Estimasi rata-rata & confidence interval dari sampel

5. Central Limit Theorem (CLT)

- CLT menjelaskan: rata-rata sampel \rightarrow distribusi normal walau data asli tidak normal. - Digunakan untuk memperkirakan probabilitas atau persentase menggunakan distribusi normal dari sampel.