

Final Assignment

Pelatihan Statistics For Finance

Deadline: ~ 1 bulan (akhir Juli 2024)

Proyek 1: Uji Hipotesa dengan Chi Square **(semua kelompok wajib mengerjakan proyek ini)**

Proyek dikerjakan dengan Jupyter Notebook dan dikumpulkan dengan format file ipynb (file Jupyter Notebook). Nama file adalah nama kelompok ditambah dengan “_proyek_1.ipynb” (misalnya “Group_A_proyek_1.ipynb”). Data yang digunakan adalah tabel yang mencatat insiden gangguan burung/unggas liar terhadap pesawat udara di Amerika Serikat sepanjang tahun 1990 – 2015.

Kerangka penulisan sebagai berikut:

1. Pendahuluan: tujuan dari proyek ini meneliti adanya keterkaitan antara satu variabel dengan variabel lainnya. Karenanya dilakukan Uji Hipotesa dengan metode Chi Square guna memastikan apakah dua variable tersebut independen atau terkait satu dengan lainnya.
2. Definisi: jabarkan variabel-variable yang akan kita uji keterkaitan/ketergantungan satu dengan yang lainnya
 - a) ‘flight impact’ versus ‘visibility’
 - b) ‘flight impact’ versus dua ‘Species Name’ yaitu ‘Mourning Dove’ dan ‘Gull’
 - c) ‘flight impact’ versus ‘operator’
 - d) ‘flight impact’ versus ‘make’Saran: gunakan metode Pandas Crosstab
3. Desain Studi/Penelitian: dengan kata-kata sendiri, jelaskan prosedur statistik yang akan ditempuh untuk memberikan jawaban terkait pertanyaan di atas
4. Hasil Descriptive Statistics: tabulasikan variabel-variabel yang akan diuji di poin 2 di atas dengan menggunakan Pandas Crosstab
5. Hasil Statistical Analysis: laporkan hasil analisa dalam bentuk parameter-parameter:
 - a) Hipotesa Null dan hipotesa Alternatif
 - b) Nilai dari χ^2 (Chi Square)
 - c) p-value terkait
 - d) Degrees of freedom (DoF)
6. Temuan: interpretasikan hasil analisa di poin 5 di atas dalam bahasa yang dapat dimengerti di lingkungan pekerjaan baik teman sejawat, atasan maupun team yang Bapak/Ibu pimpin.

Proyek 2 (dikerjakan kelompok genap) – Uji hipotesa dengan ANOVA

Proyek dikerjakan dengan Jupyter Notebook dan dikumpulkan dengan format file ipynb (file Jupyter Notebook). Nama file adalah nama kelompok ditambah dengan “_proyek_2.ipynb” (misalnya “Group_B_proyek_2.ipynb”). Data input memuat tabel penurunan berat badan setelah peserta eksperimen menyelesaikan salah satu dari 3 program diet yang hendak kita pelajari efektivitasnya.

Kerangka penulisan sebagai berikut:

1. Pendahuluan: tujuan dari proyek ini meneliti program diet mana yang paling efektif dalam menurunkan berat badan. Karenanya dilakukan Uji Hipotesa dengan metode (One-Way) ANOVA guna memastikan diet mana yang paling efektif atau, ternyata ketiga jenis diet tersebut memiliki efek yang sama.
2. Definisi: jabarkan variabel-variabel yang kita perlukan untuk melakukan ANOVA. Pertama adalah variabel “Diet”. Kemudian variabel “BMI_change” yang kita tambahkan di tabel dan didapatkan dari “BMI_end” – “BMI_start”. Kita tambah kolom “BMI_start” dimana “BMI_start” = “pre.weight” / (“Height”/100)² dan “BMI_end” dimana “BMI_end” = “weight6weeks” / (“Height”/100)².
3. Desain Studi/Penelitian: dengan kata-kata sendiri, jelaskan prosedur statistik yang akan ditempuh untuk memberikan jawaban terkait pertanyaan di atas.
4. Hasil Descriptive Statistics: plot keterkaitan rata-rata “BMI_change” untuk setiap jenis “Diet” seperti yang dicontohkan pada catatan pelatihan kita di https://github.com/digital-analyst/statistics/blob/main/Module_6/01-anova.ipynb
5. Hasil Statistical Analysis: laporkan hasil analisa dengan menggunakan F test dalam bentuk parameter-parameter:
 - a) Hipotesa Null dan hipotesa Alternatif
 - b) Nilai dari F test
 - c) p-value terkait
 - d) Degrees of freedom (DoF)
6. Post-hoc Statistical Analysis: laporkan hasil analisa perbandingan dari setiap pasangan diet dengan menggunakan Pairwise Student-T test
 - a) Hipotesa Null dan hipotesa Alternatif
 - b) Nilai dari T test untuk setiap pasangan diet
 - c) p-value dari setiap pasangan diet
 - d) Degrees of freedom (DoF)
7. Temuan: interpretasikan hasil analisa statistik di poin 5 dan 6 di atas dalam bahasa yang dapat dimengerti di lingkungan pekerjaan baik teman sejawat, atasan maupun team yang Bapak/Ibu pimpin.

Proyek 3 (dikerjakan kelompok ganjil) – Uji hipotesa dengan Multiple Linear Regression

Proyek dikerjakan dengan Jupyter Notebook dan dikumpulkan dengan format file ipynb (file Jupyter Notebook). Nama file adalah nama kelompok ditambah dengan “_proyek_3.ipynb” (misalnya “Group_A_proyek_3.ipynb”). Data input memuat tabel kinerja akademis para pelajar di suatu sekolah relatif terhadap variabel-variabel lainnya.

Kerangka penulisan sebagai berikut:

1. Pendahuluan: tujuan dari proyek ini meneliti apakah model regresi linear cukup baik untuk melihat keterkaitan antara prestasi akademis dan berbagai variabel pendukung, dan apabila demikian halnya maka variabel-variabel mana dominan mempengaruhi prestasi seorang pelajar.
2. Definisi: jabarkan variabel-variabel yang kita perlukan untuk melakukan Multiple Linear Regression. Pertama adalah variabel outcome, yang kedua adalah variabel-variabel predictors.
3. Data Cleaning: check adanya baris/row duplikat dengan Pandas `df.duplicated.sum()`. Kalau ada semua row duplikat di buang dengan `df.drop_duplicates()`. Kemudian kolom “Extracurricular Activities” masih menggunakan “Yes” atau “No”. Kita ganti “Yes” dan “No” dengan “1” dan “0” dengan `df['Extracurricular Activities'] = df['Extracurricular Activities'].replace({'Yes': 1, 'No': 0})`
4. Desain Studi/Penelitian: dengan kata-kata sendiri, jelaskan prosedur statistik yang akan ditempuh untuk memberikan jawaban terkait pertanyaan di atas.
5. Hasil Descriptive Statistics: tabulasi mean, std, min dll dengan menggunakan `df.describe()`.
6. Hasil Statistical Analysis: laporkan hasil analisa apakah model regresi linear layak digunakan dengan menggunakan F test dalam bentuk parameter-parameter:
 - a) Hipotesa Null dan hipotesa Alternatif
 - b) Nilai dari F test
 - c) p-value terkait
7. Post-hoc Statistical Analysis: laporkan hasil analisa dari setiap koefisien setiap variabel predictor dengan menggunakan Pairwise Student-T test
 - a) Hipotesa Null dan hipotesa Alternatif
 - b) Nilai dari T test untuk setiap koefisien
 - c) p-value dari setiap koefisien
8. Temuan: interpretasikan hasil analisa statistik di poin 6 dan 7 di atas dalam bahasa yang dapat dimengerti di lingkungan pekerjaan baik teman sejawat, atasan maupun team yang Bapak/Ibu pimpin. Jelaskan variabel predictor mana yang dominan mempengaruhi prestasi seorang pelajar.