



— UNIVERSITAS —
Nusa Putra

Laporan ini merupakan bagian dari tugas mata kuliah

Data Science

Dengan Dosen Pengampu

GINA PURNAMA INSANY, S.ST., M.Kom



Nama : SAILA JULIA

Kelas : TI22A

NIM : 20220040082

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
UNIVERSITAS NUSA PUTRA SUKABUMI**

DATA SCIENCE SESI 3

1. Seorang data scientist sedang menganalisis data penjualan produk dari dua cabang toko yang berbeda. Dia ingin mengetahui apakah terdapat perbedaan signifikan dalam rata-rata penjualan harian antara dua cabang tersebut.

Pertanyaan:

Jelaskan bagaimana uji hipotesis dapat digunakan dalam kasus ini, dan sebutkan langkah-langkahnya!

2. Sebuah perusahaan teknologi ingin memahami apakah pengenalan fitur baru dalam aplikasi mereka meningkatkan kepuasan pengguna. Perusahaan melakukan survei terhadap 100 pengguna sebelum fitur baru diperkenalkan dan 100 pengguna setelah fitur tersebut diperkenalkan.

Pertanyaan:

Bagaimana perusahaan dapat menggunakan interval kepercayaan untuk mengukur perbedaan rata-rata kepuasan pengguna, dan apa interpretasinya?

JAWABAN

1. Uji hipotesis digunakan untuk menentukan apakah terdapat perbedaan signifikan dalam rata-rata penjualan harian antara dua cabang toko berdasarkan data yang tersedia. Langkah-langkah uji hipotesis yang dapat digunakan dalam kasus ini:

1. Menyusun Hipotesis

Langkah pertama adalah menyusun dua hipotesis:

- Hipotesis nol (H_0): Rata-rata penjualan harian di kedua cabang sama. Tidak ada perbedaan signifikan.
- Hipotesis alternatif (H_1): Ada perbedaan signifikan dalam rata-rata penjualan harian antara kedua cabang.

2. Menentukan Tingkat Signifikansi

Tingkat signifikansi adalah peluang melakukan kesalahan tipe I (menolak hipotesis nol yang benar). Biasanya ditetapkan pada 5% atau 0.05.

3. Memilih Uji Statistik yang Tepat

Untuk membandingkan rata-rata dua kelompok, biasanya digunakan Uji T dua sampel independen, jika data berdistribusi normal. Jika asumsi normalitas tidak terpenuhi, dapat digunakan Uji Mann-Whitney.

Asumsi untuk uji T dua sampel independen:

- Data harus independen (penjualan harian dari dua cabang tidak boleh saling terkait).
- Data diambil dari distribusi normal (bisa diuji dengan uji normalitas seperti Shapiro-Wilk).
- Varians antara kedua kelompok data harus sama (homogenitas varians).

4. Mengumpulkan Data

Kumpulkan data penjualan harian dari kedua cabang untuk periode tertentu. Misalnya, data dari cabang A dan cabang B.

5. Melakukan Uji Statistik

- Uji t dua sampel independen

Jika data tidak berdistribusi normal, gunakan Uji Mann-Whitney, yang merupakan uji non-parametrik untuk dua kelompok independen.

6. Menghitung dan Menginterpretasi Nilai

- Setelah uji statistik dilakukan, periksa nilai p .
- Jika nilai (p) lebih kecil dari α (misalnya ($p < 0.05$)), tolak hipotesis nol dan simpulkan bahwa terdapat perbedaan signifikan dalam rata-rata penjualan antara dua cabang.
- Jika ($p < 0.05$), gagal menolak hipotesis nol, yang berarti tidak ada cukup bukti untuk menyimpulkan bahwa rata-rata penjualan kedua cabang berbeda.

7. Mengambil Keputusan

Berdasarkan hasil uji statistik:

- Jika hipotesis nol ditolak: Ada perbedaan signifikan dalam penjualan harian antara cabang A dan cabang B.
- Jika hipotesis nol tidak ditolak: Tidak ada perbedaan signifikan antara rata-rata penjualan harian dari dua cabang.

8. Kesimpulan dan Tindakan Lanjutan

Berdasarkan keputusan uji hipotesis, data scientist dapat menarik kesimpulan dan memberikan rekomendasi untuk pengelolaan cabang lebih lanjut. Misalnya, jika ada perbedaan signifikan, toko yang memiliki penjualan lebih rendah mungkin perlu mengkaji ulang strategi penjualannya.

Uji hipotesis memberikan landasan statistik untuk menentukan apakah perbedaan rata-rata penjualan yang terlihat dalam data cukup signifikan atau hanya akibat variasi acak.

2. Perusahaan dapat menggunakan interval kepercayaan untuk mengukur perbedaan rata-rata kepuasan pengguna sebelum dan sesudah fitur baru diperkenalkan, guna memahami apakah fitur baru tersebut secara signifikan meningkatkan kepuasan pengguna. Langkah-langkah untuk menggunakan interval kepercayaan dalam kasus ini, beserta interpretasinya:

1. Menghitung Rata-rata dan Simpangan Baku (Standar Deviasi) dari Kepuasan Pengguna

- Pertama, perusahaan harus menghitung rata-rata (mean) kepuasan pengguna sebelum dan sesudah fitur diperkenalkan, berdasarkan survei.
- Lalu, hitung simpangan baku (standard deviation) dari kedua kelompok pengguna (sebelum dan sesudah fitur).

Misalnya:

- Rata-rata kepuasan sebelum fitur baru (\bar{x}_1) = 7.2
- Rata-rata kepuasan setelah fitur baru (\bar{x}_2) = 8.0
- Simpangan baku sebelum (s_1) = 1.5
- Simpangan baku setelah (s_2) = 1.2
- Ukuran sampel untuk masing-masing kelompok $n_1 = n_2 = 100$

2. Menghitung Selisih Rata-rata

Selisih rata-rata ($\Delta\mu$) adalah:

$$\Delta\mu = \bar{x}_2 - \bar{x}_1 = 8.0 - 7.2 = 0.8$$

Selisih ini adalah estimasi kasar peningkatan kepuasan pengguna setelah fitur baru diperkenalkan.

3. Menghitung Simpangan Baku Gabungan

Untuk menghitung interval kepercayaan, kita perlu mengetahui simpangan baku gabungan dari kedua kelompok:

$$s_{gabungan} = \sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}$$

Menggunakan nilai simpangan baku dan ukuran sampel di atas:

$$S_{gabungan} \sqrt{\frac{1.5^2}{100} + \frac{1.2^2}{100}} = \sqrt{\frac{2.25}{100} + \frac{1.44}{100}} \sqrt{0.0225 + 0.0144} = \sqrt{0.0369} = 0.192$$

4. Menentukan Tingkat Kepercayaan (Confidence Level)

Umumnya, perusahaan akan memilih tingkat kepercayaan 95% ($\alpha = 0.05$). Ini berarti perusahaan ingin 95% yakin bahwa interval tersebut mencakup perbedaan rata-rata kepuasan yang sebenarnya.

Untuk tingkat kepercayaan 95%, nilai z (dari tabel distribusi normal) adalah 1.96.

5. Menghitung Interval Kepercayaan

$$CI = (\Delta\mu - Z \cdot S_{gabungan}, \Delta\mu + Z \cdot S_{gabungan})$$

Menghitung intervalnya:

$$\{CI\} = (0.8 - 1.96 \cdot 0.192, 0.8 + 1.96 \cdot 0.192)$$

$$\{CI\} = (0.8 - 0.376, 0.8 + 0.376) = (0.424, 1.176)$$

6. Interpretasi Interval Kepercayaan

- Interval kepercayaan 95% untuk perbedaan rata-rata kepuasan adalah (0.424, 1.176).
- Ini berarti perusahaan 95% yakin bahwa peningkatan rata-rata kepuasan pengguna berada di antara 0.424 dan 1.176 poin.
- Karena seluruh interval berada di atas 0, ini menunjukkan bahwa ada peningkatan signifikan dalam kepuasan pengguna setelah fitur baru diperkenalkan. Jika interval tersebut mencakup nol, maka kita tidak dapat menyimpulkan bahwa fitur baru meningkatkan kepuasan secara signifikan.

7. Kesimpulan dan Tindakan Lanjutan

Berdasarkan hasil interval kepercayaan ini, perusahaan dapat menyimpulkan bahwa fitur baru kemungkinan besar meningkatkan kepuasan pengguna. Namun, seberapa besar peningkatannya dapat dilihat dari nilai interval (antara 0.424 hingga 1.176). Perusahaan dapat menggunakan informasi ini untuk memutuskan langkah lebih lanjut, seperti memperkenalkan fitur ke lebih banyak pengguna atau melakukan perbaikan tambahan. Jika interval kepercayaan tersebut mencakup angka negatif atau nol, perusahaan mungkin harus mengevaluasi lebih lanjut apakah fitur tersebut benar-benar memberikan dampak positif.