Project 2

Pengujian Hipotesis

**Tujuan:**

Tugas ini bertujuan untuk menerapkan estimasi non-parametrik dalam merangkum data survival, dan melakukan prosedur pengujian hipotesis statistika.

**Sub CPMK yang dinilai:**

**[Sub-CPMK 6]** Mahasiswa mampu melakukan pengujian hipotesis, dan mampu melakukan perbandingan antar metode pengujian untuk menentukan metode yang terbaik untuk suatu data. (C4)

**Due date**: Hari Sabtu di Pekan perkuliahan ke-12 (3 Mei 2025) pukul 23.59 WIB

**Anggota kelompok:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Nama | NPM | Kontribusi | Tingkat kontribusi |
| 1 | Mhs 1 |  | Misal: terlibat aktif diskusi, menuliskan ringkasan hasil diskusi | 100% |
| 2 | Mhs 2 |  | Misal: Mencari sumber lain untuk klarifikasi pemahaman topik, menulis ringkasan hasil diskusi | 100% |
| 3 | Mhs 3 |  | Misal: pasif dalam diskusi, menuliskan ringkasan hasil diskusi, mengerjakan Latihan 2.1 no.1 &2, dst…. | 40% |
|  |  |  |  |  |
|  |  |  |  |  |

**Latar Belakang**

Perusahaan telekomunikasi *TelcoX* menghadapi tantangan meningkatnya jumlah pelanggan yang berhenti berlangganan (*customer churn*). Untuk meningkatkan retensi, perusahaan ingin memahami faktor-faktor yang memengaruhi **lama pelanggan bertahan** dalam layanan.

Tim analis diminta melakukan analisis **survival** atas durasi pelanggan tetap berlangganan (tenure, dalam bulan), menggunakan data historis. Mereka ingin mengetahui apakah terdapat **perbedaan survival** antara berbagai jenis layanan, apakah **tren churn** berubah seiring waktu, serta apakah perlu dilakukan **analisis stratifikasi** berdasarkan profil pelanggan.

**Tujuan Project**

Mahasiswa diminta untuk:

1. **Membuat kurva Kaplan-Meier** untuk menggambarkan survival pelanggan.
2. **Melakukan uji hipotesis survival**:
   * Uji log-rank untuk membandingkan survival antar dua atau lebih grup.
   * Uji tren untuk melihat kecenderungan pola survival.
   * Uji stratifikasi untuk mengontrol variabel *confounding*.

**Deskripsi Data Survival**

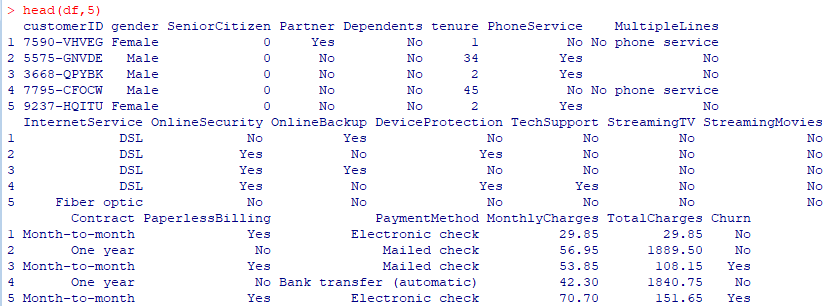
Data berisi informasi pelanggan seperti:

* tenure: lama waktu (dalam bulan) hingga pelanggan berhenti
* Churn: status (1 = churn, 0 = masih aktif → censored)
* Contract: jenis kontrak (Month-to-month, One year, Two year)
* InternetService: tipe layanan internet (DSL, Fiber optic, None)
* SeniorCitizen: 1 jika pelanggan usia lanjut, 0 jika tidak
* PaymentMethod: metode pembayaran

**Data lengkap** dapat diunduh di: [Telco-Customer-Churn/Datasets/telco\_churn.csv at master · KimathiNewton/Telco-Customer-Churn · GitHub](https://github.com/KimathiNewton/Telco-Customer-Churn/blob/master/Datasets/telco_churn.csv?plain=1), atau juga tersedia pada tautan di emas2.ui.ac.id

**Keterangan lengkap variabel** dapat dilihat di: [Telco-Customer-Churn/Telco\_Customer\_Churn.ipynb at master · KimathiNewton/Telco-Customer-Churn · GitHub](https://github.com/KimathiNewton/Telco-Customer-Churn/blob/master/Telco_Customer_Churn.ipynb)

**Contoh subset data**:



**0. [10 poin] Deskripsi data**

Tuliskan secara singkat deskripsi dari data yang digunakan (berupa ringkasan numerik, tabel, atau grafik yang sesuai), event yang akan dianalisis dan apakah ada censoring atau tidak, jenis censoring (jika ada), juga faktor-faktor yang akan digunakan dalam pengujian hipotesis di bawah ini. Tujuan deskripsi ini adalah untuk mendapatkan gambaran umum mengenai data dan faktor-faktor terkait yang akan dianalisis.

**1. [30 poin] Log-Rank Test: Pengaruh Jenis Kontrak**

Apakah ada perbedaan waktu survival antar jenis kontrak (Month-to-month, One year, Two year)?

**Tugas Mahasiswa:**

* Plot kurva Kaplan-Meier untuk setiap jenis kontrak.
* Lakukan uji log-rank: tuliskan prosedur pengujian dengan lengkap, berikan interpretasi hasil pengujian, lampirkan codes (file codes pengerjaan atau tautannya)

Termasuk di dalamnya, lakukan hal berikut:

1. Gambarkan kurva Kaplan-Meier per grup.
2. Jelaskan hasil pengujian hipotesis survival (termasuk nilai p dan interpretasinya).
3. Jika mungkin, berikan saran kebijakan berdasarkan hasil: misal strategi retensi untuk jenis kontrak tertentu atau segmen usia tertentu.

Tuliskan prosedur pengujian dengan lengkap, lampirkan codes (file codes pengerjaan atau tautannya).

**2. [30 poin] Uji Tren: Churn Berdasarkan Kelompok Tagihan Penggunaan**

Apakah ada tren peningkatan/penurunan survival seiring bertambahnya tagihan biaya per bulan (MonthlyCharges)?

**Tugas Mahasiswa:**

* Kelompokkan MonthlyCharges: <40, 40–80, ≥80.
* Gunakan uji tren: tuliskan prosedur pengujian dengan lengkap, berikan interpretasi hasil pengujian, lampirkan codes (file codes pengerjaan atau tautannya)

Termasuk di dalamnya, lakukan hal berikut:

1. Gambarkan kurva Kaplan-Meier per grup.
2. Jelaskan hasil pengujian hipotesis survival (termasuk nilai p dan interpretasinya).
3. Jika mungkin, berikan saran kebijakan berdasarkan hasil: misal strategi retensi untuk jenis kontrak tertentu atau segmen usia tertentu.

Tuliskan prosedur pengujian dengan lengkap, lampirkan codes (file codes pengerjaan atau tautannya).

**3. [30 poin] Uji Stratifikasi: Kontrak vs SeniorCitizen**

Apakah ada perbedaan survival antara jenis kontrak setelah dikontrol oleh status usia lanjut?

**Tugas Mahasiswa:**

Lakukan uji **stratified log-rank test** berdasarkan Contract, dengan stratifikasi pada SeniorCitizen.

Termasuk di dalamnya, lakukan hal berikut:

1. Gambarkan kurva Kaplan-Meier per grup.
2. Jelaskan hasil pengujian hipotesis survival (termasuk nilai p dan interpretasinya).
3. Jika mungkin, berikan saran kebijakan berdasarkan hasil: misal strategi retensi untuk jenis kontrak tertentu atau segmen usia tertentu.

Tuliskan prosedur pengujian dengan lengkap, lampirkan codes (file codes pengerjaan atau tautannya).