

Tugas Praktikum Sesi UTS STA561 Pemrograman Statistika

Ketentuan

- Tugas dikerjakan secara berkelompok beranggotakan 4 orang. Pembentukan kelompok diserahkan ke PJ masing-masing. Khusus mahasiswa sinergi tidak boleh berada di kelompok yang sama bersama sesama mahasiswa sinergi.
- Tugas dikumpulkan oleh salah satu anggota kelompok (tidak perlu setiap orang submit) via LMS paling lambat 1 November 2020 pukul 23.59 WIB.
- Konten jawaban yang disusun adalah **nama anggota kelompok**, **syntax program**, **deskripsi singkat** mengenai program yang disusun, dan **preview hasil**. Syntax program harus dituliskan dalam bentuk teks (bukan gambar) agar dapat langsung disalin ke R.
- Format file jawaban boleh berupa file word, power point, pdf, ataupun Rmarkdown (html/pdf). Format nama file jawaban adalah "**Tugas UTS STA561 Par [no paralel] Kel [no kelompok]**", contoh: "Tugas UTS STA561 Par 1 Kel 5.docx".
- Ada apresiasi tambahan bagi kelompok yang submit jawabannya lebih awal. Jika ada lebih dari satu anggota kelompok yang submit jawaban, jawaban yang akan dinilai adalah yang di-submit paling akhir.
- Penilaian didasarkan pada kesesuaian program dan output yang diminta, orisinalitas, kreativitas, dan urutan pengumpulan jawaban.

Soal

1. Buatlah program R untuk melakukan proses berikut ini!
 - Bangkitkanlah data **populasi** yang terdiri atas 1000 amatan dengan peubah:
 - **No** : 1 sd 500, 9001 sd 9500 (integer)
 - **ID** : G0001 sd G0500, G9001 sd G9500 (diperoleh berdasarkan kolom **No**, dengan penambahan huruf "G" didepan **No** dan jumlah digitnya harus 4)
 - **Kelas** : "Paralel 1" jika **ID** genap, dan "Paralel 2" jika **ID** ganjil (factor)
 - **UTS** : Diperoleh dengan membangkitkan bilangan acak yang menyebar $N(\mu = 70, \sigma^2 = 25)$
 - **Indeks** : Diperoleh dengan membangkitkan bilangan acak yang menyebar $\text{exponential}(\theta = 1)$
 - Hilangkan data peubah **Indeks** yang bernilai lebih dari 1 (menjadi missing value)
 - Hitunglah rata-rata dari **UTS** yang **Indeks**-nya ada
 - Buatlah data **sampel** berukuran 100 amatan dari **Kelas** "Paralel 1" yang **Indeks**-nya ada
 - Hitunglah rata-rata dari **UTS** pada data **sampel** yang terambil, lalu bandingkan dengan **populasi**

2. Buatlah fungsi di R yang bernama **RKU** yang dapat digunakan untuk menghitung regresi komponen utama (*principal component regression*) dengan output utama adalah koefisien penduga parameter regresi komponen utama yang sudah ditransformasi balik untuk peubah X !

Petunjuk dan ketentuan:

- Untuk menghindari kesalahpahaman konsep yang dimaksud, berikut gambaran umum regresi komponen utama yang dimaksud beserta output yang diminta
 - Terdapat peubah Y, X_1, X_2, \dots, X_p
 - Transformasi ke komponen utama PC_1, PC_2, \dots, PC_m

$$PC_j = \sum_{i=1}^p a_{ij}X_i = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{pj}X_p; j = 1, \dots, m$$
 - Pemodelan regresi komponen utama

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \dots + \beta_m PC_m + \varepsilon$$
 - Transformasi balik model regresi komponen utama ke peubah X

$$Y = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \dots + \beta_p^* X_p + \varepsilon^*$$
 - Output utama yang diminta: penduga koefisien β_i^*
 - Jika kurang jelas, dipersilakan merujuk ke referensi tertentu
- Fungsi yang dapat digunakan adalah yang terdapat di R base tanpa menggunakan fungsi “tambahan” yang sudah terdapat pada package tertentu di R (tanpa mengaktifkan package tertentu), seperti **lm**, **eigen**, **%*%**, dsb.
- Format bentuk argumen input dan output/return fungsi tidak ditentukan secara spesifik tipe dan banyaknya objek, dapat dikreasikan sesuai keinginan masing-masing kelompok.
- Beberapa contoh variasi yang dapat digunakan (variasi diluar contoh berikut diizinkan untuk digunakan):
 - Argumen input hanya 1 berupa matriks $[Y|X]$ (peubah Y gabung/gandeng peubah-peubah X)
 - Argumen input ada 2, yaitu data frame peubah-peubah X dan vektor peubah Y
 - Argumen input ada 3, yaitu matriks peubah-peubah X , vektor peubah Y , skalar k (banyaknya peubah komponen utama yang digunakan dalam pemodelan regresi)
 - Output berupa matriks koefisien RKU untuk X
 - Output berupa list yang berisi vektor koefisien RKU untuk X , vektor koefisien RKU untuk PC , matriks skor komponen utama, dan list akar ciri-vektor ciri
 - Output berupa list yang berisi seluruh proses algoritme mulai dari akar ciri hingga vektor koefisien RKU untuk X , kemudian dibuat fungsi generik sehingga ketika dipanggil hanya muncul tampilan tertentu walaupun sebenarnya elemennya banyak (seperti output dari fungsi **lm**)
 - dsb

3. Bangkitkan peubah Y, X_1, X_2, X_3 sebanyak 1000 amatan berdasarkan model regresi linear berganda berikut ini:

$$Y = 10 + 3X_1 + 5X_2 + 7X_3 + \varepsilon$$

dengan mengasumsikan bahwa $\varepsilon \sim N(0,1)$ dan antara peubah bebas terjadi **multikolinearitas** (tidak ada batasan fungsi/package yang digunakan)! Kemudian hitung koefisien penduga parameter regresi komponen utama yang sudah ditransformasi balik menggunakan fungsi **RKU** yang sudah dibuat di nomor 2!