



**LABORATORIUM KOMPUTASI STATISTIKA DAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS GADJAH MADA  
Sekip Utara BLS 21, Yogyakarta 55281**

---

**KETENTUAN DAN PERATURAN**

1. Jangan lupa berdoa sebelum mengerjakan.
2. Jawaban diketik dengan font Times New Roman ukuran 12pt, line spacing 1.15, margin normal, dan ukuran kertas A4.
3. Lembar jawaban terdiri dari soal, jawab, sintaks, output, interpretasi output, tabel (apabila diminta)
4. **Dilarang meng-copypaste atau menjiplak hasil jawaban teman (plagiarisme).** Apabila ketahuan meng-copypaste atau menjiplak, maka **nilai akan langsung 0.**
5. **Dilarang bertanya pada Asisten Praktikum mengenai jawaban.**
6. Jangan lupa mencantumkan screenshot dari output program R dan interpretasi.
7. Yang perlu dicantumkan di dalam bagian Pembahasan adalah ketikan ulang syntax, screenshot output, dan interpretasi output.
8. **Jawaban dikumpulkan dalam bentuk .pdf dan .txt (Sintaks ) disimpan melalui app notepad) dengan format nama file: RESPONSI\_NIF (contoh : RESPONSI\_19777).**
9. Submit file di Google Classroom paling lambat hari **Senin, 12 Desember 2022 pukul 23.59 WIB.** Keterlambatan pengumpulan akan ada pengurangan nilai.
10. **Keterlambatan pengumpulan akan ada pengurangan nilai akhir (-1 per menit) .**

**Format pengerjaan laporan :**

Cover

Bab I (Permasalahan)

Bab II (Pembahasan)

Bab III (Penutup)

- o Kesimpulan
- o Kritik dan saran

**Asisten Praktikum :**

Muhammad Zaki Nurkholis    19733

Ahmad Habib Hasan Zein    20277

**Dosen Pengampu :**

Prof. Dr.rer.nat. Dedi Rosadi, S.Si., M.Sc.

---

**Nomor 1 [Skor 25]**

*Bayesian analysis is performed by combining the prior information ( $\pi(\theta)$ ) and the sample information ( $x$ ) into what called the posterior distribution of  $\theta$  given  $x$ , from which all decision and inference are made. The posterior distribution of  $\theta$  given  $x$  (or posterior for short) will be denoted  $\pi(\theta|x)$ , and as the notation indicates, is defined to be the conditional distribution of  $\theta$  given the sample observation  $x$ . Nothing that  $\theta$  and  $X$  have joint (subjective) density :*

$$h(x, \theta) = \pi(\theta)f(\theta|x),$$

*and that  $X$  has marginal (unconditional) density :*

$$m(x) = \int_{\theta} f(x|\theta) dF^{\pi}(\theta)$$

$$m(x) = \int_{\theta} \pi(\theta)f(\theta|x) d\theta$$

*It is clear that providing  $m(x) \neq 0$ , and then the posterior distribution density :*

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)}$$

*The name “posterior distribution” is indicative of the role of  $\pi(\theta|x)$ . Just as the prior distribution reflects beliefs about  $\theta$  prior to experimentation, so  $\pi(\theta|x)$  reflects the update beliefs about  $\theta$  after (posterior to) observing the sample  $x$ . In other words, the posterior distribution combines the prior beliefs about  $\theta$  with the information about  $\theta$  contained in the sample,  $x$ , to give a composite picture of the final beliefs about  $\theta$ . Note that the Likelihood Principle is implicitly assumed in the above statement, in that there is felt to be no sample information about  $\theta$  other than that contained in  $f(\theta|x)$  (for the given  $x$ ).*

**Lemma 1.** *Assume  $m(t)$  (the marginal density of  $t$ ) is greater than zero, and that the factorization theorem holds. Then if  $T(x) = t$ ,*

$$\pi(\theta|x) = \pi(\theta|t) = \frac{\pi(\theta)g(t|\theta)}{m(t)}$$

*The reason for determining  $\pi(\theta|x)$  from a sufficient statistics  $T$  (if possible) is that  $g(t|\theta)$  and  $m(t)$  are usually much easier to handle than  $f(x|\theta)$  and  $m(x)$*

## (Case Study)

*Diabetes mellitus is a chronic disease that requires long treatment therapy to reduce the incidence of complications (ADA, 2017). One of the influential factors in the rise and fall of blood pressure is blood sugar. According to Tanto and Hustrini (2014) hyperglycemia is one of the risk factors for hypertension. Hyperglycemia is often accompanied by the emergence of metabolic syndrome, namely hypertension, dyslipidemia, obesity, endothelial dysfunction and prothrombotic factors, all of which will trigger and exacerbate cardiovascular complications (Tanto and Hustrini, 2014).*

*One of the complications of diabetic macroangiopathy can occur due to changes in blood sugar levels, high blood sugar will stick to the walls of blood vessels. After that, an oxidation process occurs where blood sugar reacts with proteins from the walls of blood vessels which causes AGEs. Advanced Glycosylated Endproducts (AGEs) are substances formed from excess sugar and protein that bind together. This situation damages the inner walls of blood vessels, and attracts saturated fat or cholesterol to stick to the walls of blood vessels, so that an inflammatory reaction occurs. White blood cells (leukocytes) and blood clotting cells (thrombocytes) as well as other materials join together to form a plaque clot, which makes the walls of blood vessels hard, stiff and eventually a blockage arises which results in a change in blood pressure which is called hypertension (Tanto and Hustrini). , 2014). According to the American Diabetes Association (ADA) (2017) two out of three people with diabetes have high blood pressure.*

*High blood pressure is often only discovered when routine blood pressure checks are carried out by a doctor or undergo medical examinations in connection with work or insurance (Huda, 2016). People with high blood pressure are three times more likely to have a heart attack, five times more likely to have heart failure, and eight times more likely to have a heart attack, have a stroke compared to people with normal blood pressure (Huda, 2016).*

*In a study conducted by the American Diabetes Association (ADA), glucose levels and blood pressure were checked in 5,000 people. it was found that the test the glucose levels of the sample approached a distribution denoted by  $X \sim T(\theta, W)$ . Where  $\theta$  is unknown but  $W$  is known. It is also known that the properties of the random variable  $X$  have the  $n^{\text{th}}$  moment with the following function :*

$$M_X(t) = e^{\theta t + \frac{2401}{8}t^2}$$

From this study it was also found that blood pressure was normally distributed as the prior distribution of glucose levels denoted by  $\pi(\theta) \sim N(\mu, \tau^2)$  where  $\mu$  and  $\tau^2$  are known from the data.

**Instruction :**

- Determine **with the steps** the probability density function of the test results on glucose levels !
- Estimate the parameters (**with MLE**) of the distribution of blood pressure test from data “**Data Responsi No. 1.xlsx**” !
- The American Diabetes Association (ADA) also conducts an HbA1c test or commonly called a hemoglobin test to measure the average amount of hemoglobin that binds to glucose. It was found that the results of the HbA1c test were the marginal distribution of the glucose level test. Show **with the steps** the probability density function of HbA1c test result and notate it (example :  $C \sim J(A, B)$ )
- Estimate the parameters (**with MLE**) of the distribution of HbA1c test from data !

**Bonus :**

- Indicate the probability density function of the posterior distribution of the research above and notate it (example :  $C \sim J(A, B)$ )

**\* caution beware in performing algebraic operations in determining marginal distributions**  
**\* you must attach the process in determining each probability density function**

**Nomor 2 [Skor 25]**

Diketahui  $X$  berdistribusi geometri, yakni  $P(X = j) = p(1 - p)^{j-1}, j = 1, 2, \dots$ . Variabel acak  $X$  dalam berdistribusi geometri menyatakan banyaknya kegagalan hingga munculnya sukses pertama. Berikut merupakan fungsi distribusi kumulatif (CDF) dari distribusi geometri,

$$F(x) = P(X \leq x) = 1 - (1 - p)^{x+1}$$

Misalkan suatu percobaan mengikuti distribusi GEOM(0,2). Buatlah suatu program untuk membangkitkan 10 data acak berdistribusi geometri dengan metode *inverse transform*. Lalu lakukan *Uji Goodness of Fit Test* menggunakan *Kolmogorov-Smirnov Test* untuk membuktikan sampel yang dibangkitkan benar-benar berdistribusi Geometri dengan  $\alpha = 5\%$ .

Tambahkan fungsi `set.seed(NIF)` sebelum melakukan pemanggilan fungsi agar hasil pengacakan setiap di run menghasilkan hasil yang sama (menggunakan NIF masing-masing), contohnya sebagai berikut :

*Hint :*

- *Buatlah vektor  $X$  yang berisi angka 0 sepanjang  $n$  sebagai inisiasi.*
- *Percobaan (trial) akan terus dijalankan hingga munculnya sukses pertama, sehingga  $X = X + 1$ .*
- *Karena  $X$  menyatakan jumlah kegagalan hingga sukses pertama dalam percobaan Bernoulli dengan probabilitas keberhasilan ( $p$ ), maka untuk bilangan acak bernoulli ( $p$ ), dimana gagal ( $Y = 0$ ), jika  $U \leq 1 - p$*

### Nomor 3 [Skor 25]

Sebuah toko reparasi menjual 2 jenis komputer bekas, yakni komputer Desktop dan Laptop. Toko tersebut dikelola oleh dua orang karyawan. Karyawan 1 bernama Budi bekerja selama 40 jam tiap minggunya dengan gaji Rp 50.000,00 per jam. Karyawan 2 bernama Ilham bekerja selama 50 jam tiap minggunya dengan gaji Rp 60.000,00 per jam. Setiap jenis komputer membutuhkan sumber daya tertentu untuk dapat diperbaiki dan dijual, semua ditampilkan dalam tabel berikut.

	<b>Desktop</b>	<b>Laptop</b>
<b>Harga jual</b>	Rp 250.000,00	Rp 220.000,00
<b>Kebutuhan</b>	Alat : Rp 50.000,00	Alat : Rp 40.000,00
	Budi : 1 jam	Budi : 2 jam
	Ilham : 2 jam	Ilham : 1 jam

Tabel di atas menunjukkan bahwa untuk memperbaiki sebuah Desktop agar dapat dijual, diperlukan alat seharga Rp 50.000,00. Selain itu, Budi juga harus meluangkan waktu sebanyak 1 jam dan Ilham sebanyak 2 jam. Kemudian, harga jualnya sebesar Rp 250.000,00.

Selesaikan permasalahan program linear di atas untuk memaksimalkan keuntungan yang dapat diperoleh toko tersebut! Berapa banyak Desktop dan Laptop yang harus diproduksi oleh toko?

### Nomor 4 [Skor 25]

(Pentingnya Memahami Data)

Di masa sekarang, data sudah menjadi bagian dari keseharian kita. Data dapat membantu kita untuk mengambil keputusan yang lebih baik. Tentunya, kita harus dapat memahami data yang akan kita olah. Misal, sebuah musisi Dewo 20 ingin memasarkan album barunya. Tim pemasaran dari musisi tersebut dapat memaksimalkan jumlah penjualan album baru dengan memperhatikan data tren *genre* yang sedang viral sekarang, geografi umur yang sering mendengarkan lagu-lagu Dewo 20, dan data-data lainnya. Dengan memahami data yang kita punya, kita dapat mengambil keputusan yang menguntungkan stakeholder atau bisnis yang kita miliki.

(Case Study)

Pada kasus ini, kita memiliki data aktivitas internet banyak orang di dunia dan analisis yang membuat beberapa kode analisis tentang perusahaannya untuk dipublikasikan. Dari kode

yang dibuat oleh analis, beberapa orang merasa terbantu dan membagikan kembali kode yang telah dibuat oleh analis tersebut. Penjelasan dari masing-masing kolom pada data (“**Data Responsi No 4.csv**”) dapat dilihat sebagai berikut:

Variabel	Deskripsi
<b>Date</b>	Tanggal kode dibagikan
<b>Company_code</b>	Nama perusahaan
<b>Language</b>	Bahasa pemrograman perusahaan
<b>Last_published</b>	Waktu terakhir kali kode dipublikasikan
<b>Stars</b>	Banyaknya orang yang menganggap kode analisis sangat bermanfaat
<b>Copy</b>	Banyaknya orang yang menyalin kode analisis
<b>watchers</b>	Banyaknya orang yang melihat kode analisis

Dari data di atas, kalian sebagai analis diperintahkan untuk mencari informasi:

- 3 bahasa pemrograman perusahaan dengan rata-rata banyaknya orang yang menganggap kode analisis sangat **bermanfaat** paling tinggi.
- 3 bahasa pemrograman perusahaan dengan rata-rata banyaknya orang yang **menyalin** kode analisis paling rendah.
- 3 bahasa pemrograman perusahaan dengan rata-rata banyaknya orang yang **melihat** kode analisis paling tinggi.

“Jangan lupa ucapkan terima kasih kepada diri sendiri yang telah berjuang hingga di titik ini”

Semangat dan sukses mengerjakan 😊