

# Eksplorasi Peubah PDB per Kapita & Consumer Prices

OLeh: Kelompok 5

## Kelompok 5 Beranggotakan

1. Rizal Mujahiddan (G64190069)
2. Nisma Karmiahtun Fadilah (G64190036)
3. Dhiya Ulayya Tsabitah (G1401201013)
4. Muhammad Nachnoer Novatron Fitra Arss  
(G1401201014)

# PDB per Kapita

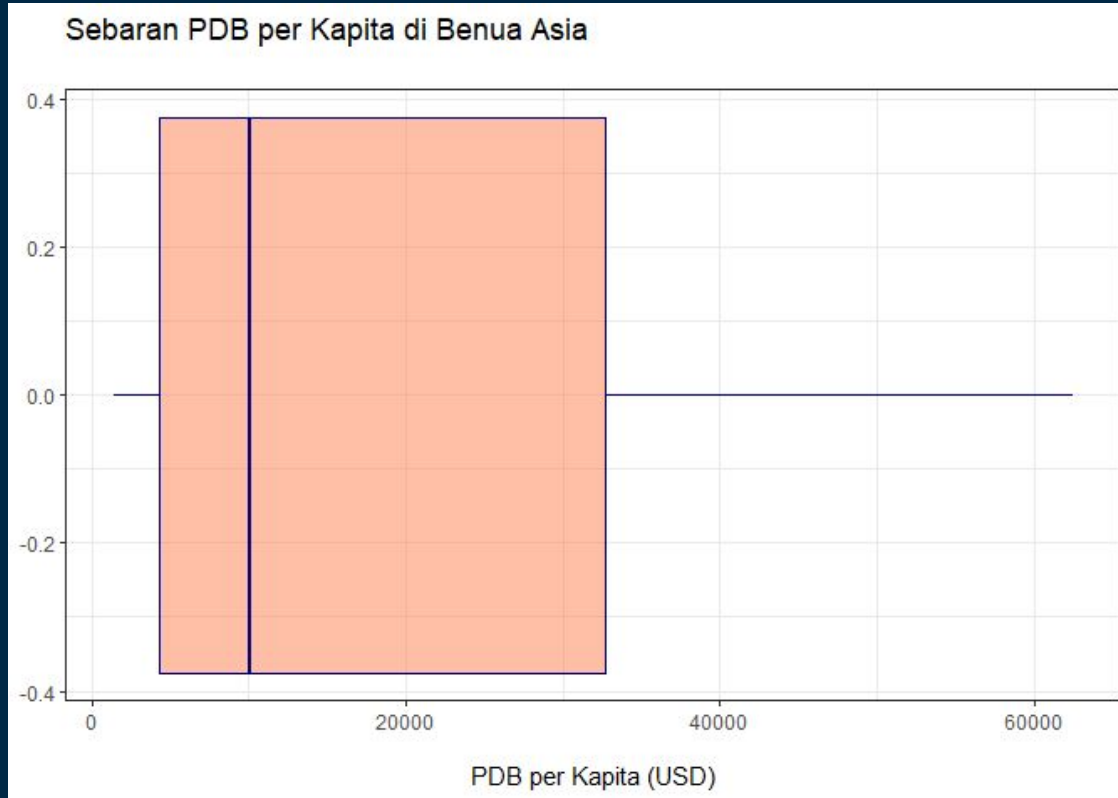
Pendapatan tiap individu pada  
suatu negara (USD)

01

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, orange, and teal, some of which are solid while others are hollow. Thin white vertical lines of varying lengths are scattered across the composition, creating a sense of depth and movement.

# EKSPLORASI DATA

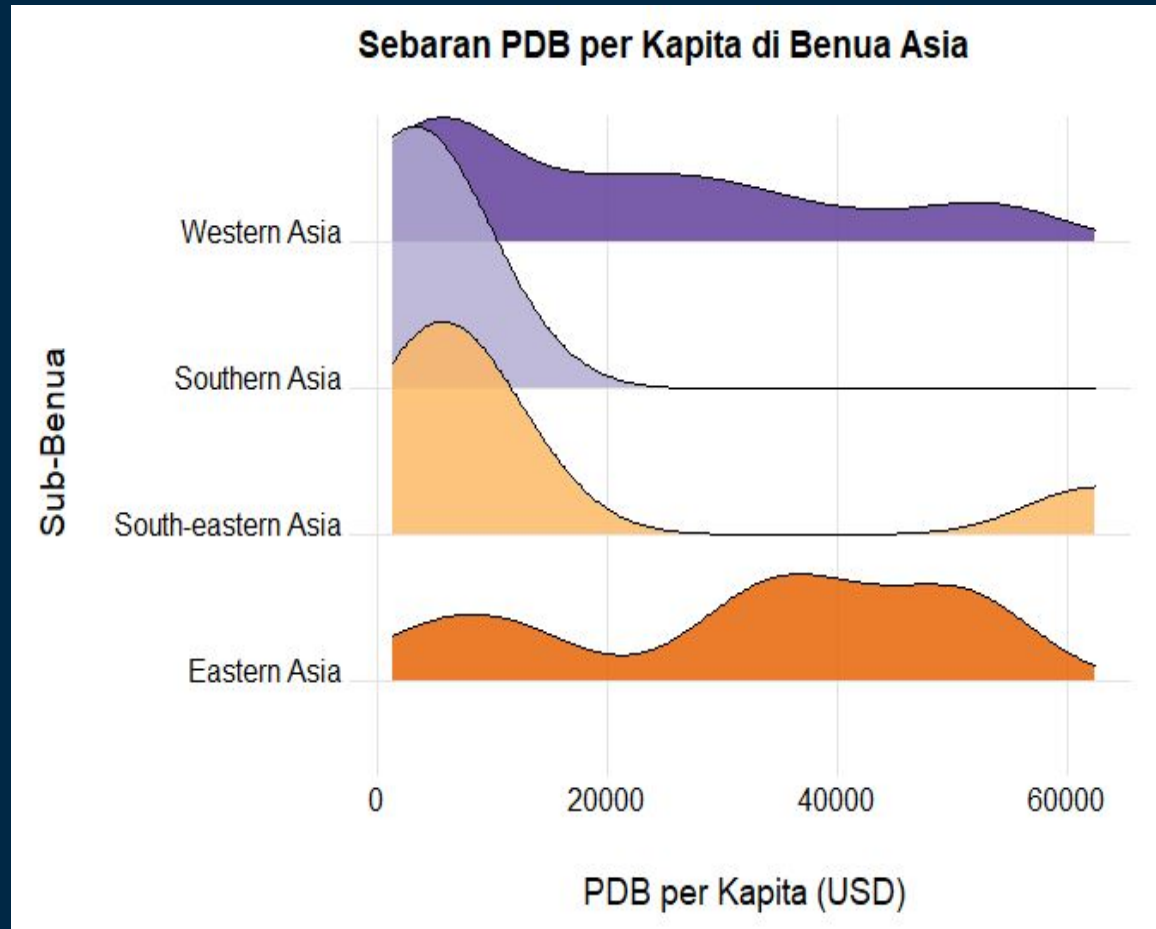
# BOXPLOT



Dari boxplot disamping, terlihat bahwa sebaran data menjulur ke kanan, dengan kata lain sebagian besar negara di Asia memiliki besaran PDB per Kapita yang dekat dengan 25% negara dengan nilai terendah, yaitu sekitar 668 USD.

# DENSITY PLOT

Walaupun secara umum terlihat menjulur ke kanan, tetapi pada Asia Barat dan Asia Timur, negara dengan PDB per Kapita yang cenderung tinggi, yaitu di atas 20 ribu USD, jumlah dan keragamannya cukup besar. Bahkan, sebaran terlihat seragam dengan dua puncak pada Asia Timur.



# Uji Formal dengan K-Smirnov Test

GDP per cap. (USD)

```
> ##Uji Formal Kolmogorov-Smirnov Test  
> ## Normal  
> set.seed(42)  
> ks.test(B, "pnorm", mean = mean(B), sd = sd(B))
```

One-sample Kolmogorov-Smirnov test

```
data: B  
D = 0.22915, p-value = 0.0473  
alternative hypothesis: two-sided
```

```
> ## Exponential  
> set.seed(42)  
> ks.test(log(B), "pexp", rate = 1/mean(B))
```

One-sample Kolmogorov-Smirnov test

```
data: log(B)  
D = 0.99942, p-value = 2.22e-16  
alternative hypothesis: two-sided
```

```
> ## Chi-Squared  
> set.seed(42)  
> ks.test(B, "pchisq", df = mean(B))
```

One-sample Kolmogorov-Smirnov test

```
data: B  
D = 0.61765, p-value = 5.057e-13  
alternative hypothesis: two-sided
```

```
> ## Lognormal  
> set.seed(42)  
> ks.test(B, "plnorm", meanlog = mean(log(B)), sdlog = sd(log(B)))
```

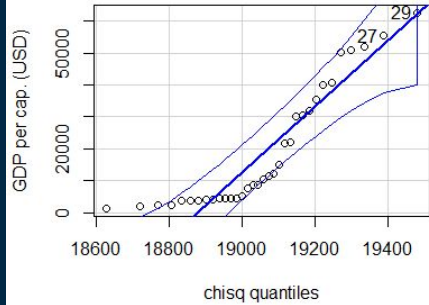
One-sample Kolmogorov-Smirnov test

```
data: B  
D = 0.16192, p-value = 0.3012  
alternative hypothesis: two-sided
```

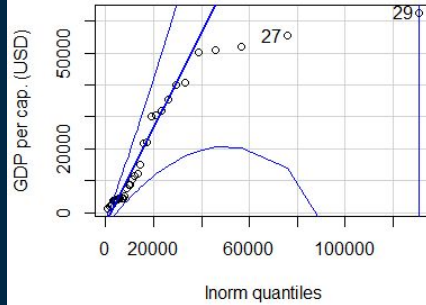
Berdasarkan hasil Uji Formal K-Smirnov, diperoleh hanya P-Value sebaran Lognormal yang lebih besar dibandingkan nilai alpha (0.05) atau Terima  $H_0$ . Sehingga dapat disimpulkan sebaran data GDP mendekati sebaran Lognormal.

# QQPLOT – GDP per cap. (USD)

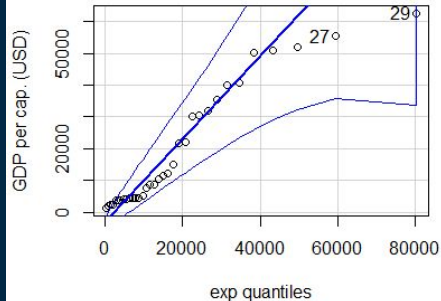
QQ-plot Terhadap Sebaran Chi-Squared



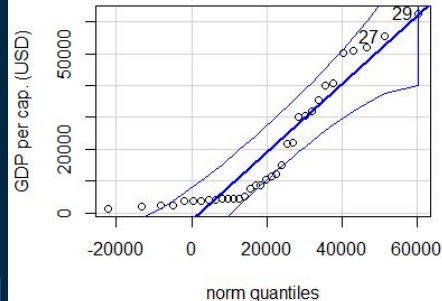
QQ-plot Terhadap Sebaran Lognormal



QQ-plot Terhadap Sebaran eksponensial



QQ-plot Terhadap Sebaran normal



Hasil dari Uji Formal K-Smirnov diperoleh bahwa sebaran data GDP mendekati sebaran Lognormal. Hal ini selaras dengan hasil QQPLOT pada sebaran Lognormal yang cenderung paling mendekati garis diagonal sebaran Lognormal.







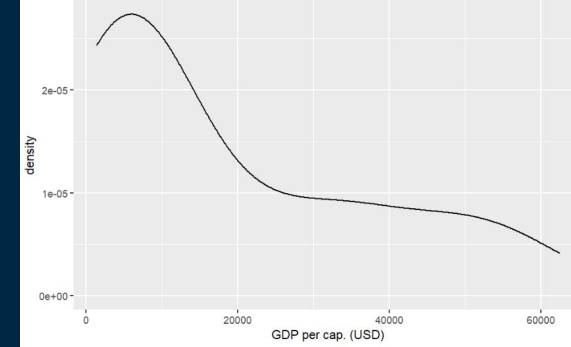
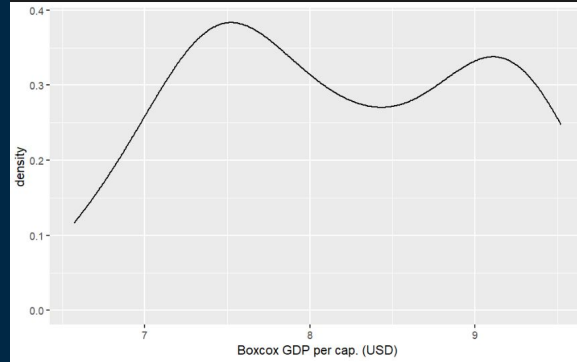
# Transformasi DATA

# Transformasi PDB per Kapita

Untuk Keseluruhan Data

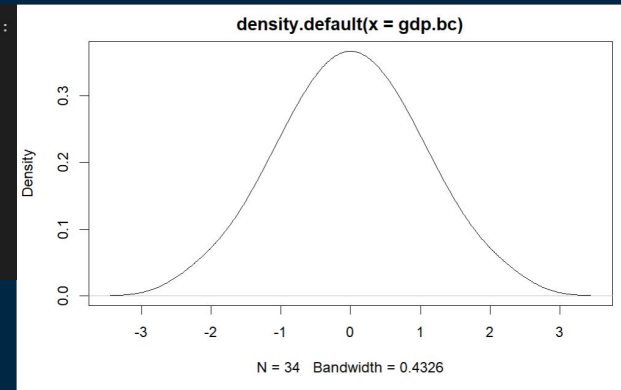
jika menggunakan tukey dan boxcox , akan menghasilkan distribusi yang cenderung bimodal distribusi ya

Untuk Transformasi ini,  
Bisa normal dengan  
orderNorm ya atau quantile  
transform istilahnya ya



```
Best Normalizing transformation with 34 Observations
Estimated Normality Statistics (Pearson P / df, lower => more normal):
- arcsinh(x): 1.4133
- Box-Cox: 1.4533
- Center+scale: 1.9867
- Log_b(x+a): 1.4133
- orderNorm (ORQ): 1.3067
- sqrt(x + a): 1.68
- Yeo-Johnson: 1.4533
Estimation method: Out-of-sample via CV with 10 folds and 5 repeats

Based off these, bestNormalize chose:
orderNorm Transformation with 34 nonmissing obs and no ties
- Original quantiles:
0%      25%      50%      75%      100%
1406.130 4256.246 9622.804 31548.284 62432.995
```



# Transformasi PDB per Kapita dengan Sub-Benua

Central Asia ditiadakan dikarenakan kurang cocok untuk buat distribusi. Pada Data western Asia inilah yang menyebabkan bimodal dikarenakan banyak data di Western Asia

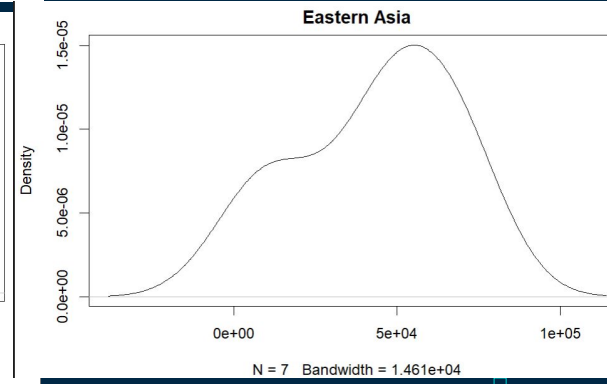
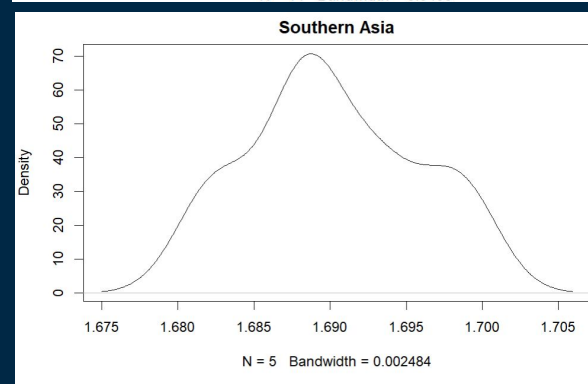
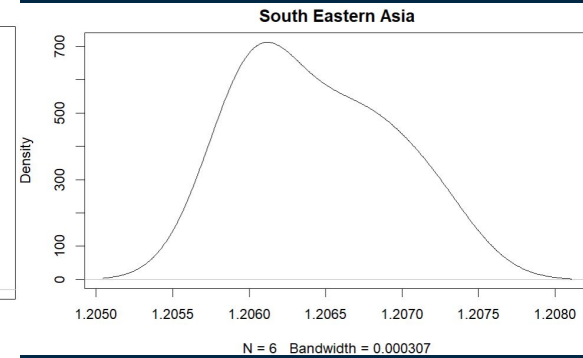
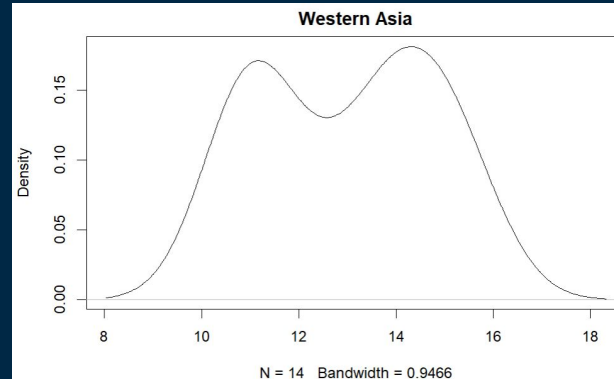
Ini boxcox Transformation per sub region

$\lambda$  western = 0.0606060

$\lambda$  eastern = 1.0303030

$\lambda$  south = -0.5858586

$\lambda$  south eastern = -0.8282828



# Transformasi PDB per Kapita dengan Sub-Benua

Shapiro-Wilk normality test

```
data: DA[DA["sub-region"] == "Southern Asia", "GDP per cap. (USD)"]  
W = 0.9776, p-value = 0.9214
```

Shapiro-Wilk normality test

```
data: DA[DA["sub-region"] == "Western Asia", "GDP per cap. (USD)"]  
W = 0.87086, p-value = 0.04315
```

Shapiro-Wilk normality test

```
data: DA[DA["sub-region"] == "Eastern Asia", "GDP per cap. (USD)"]  
W = 0.91535, p-value = 0.4342
```

Shapiro-Wilk normality test

```
data: DA[DA["sub-region"] == "South-eastern Asia", "GDP per cap. (USD)"]  
W = 0.90324, p-value = 0.3934
```

Southern Asia = Normal

Western Asia = tidak Normal

Eastern Asia = Normal

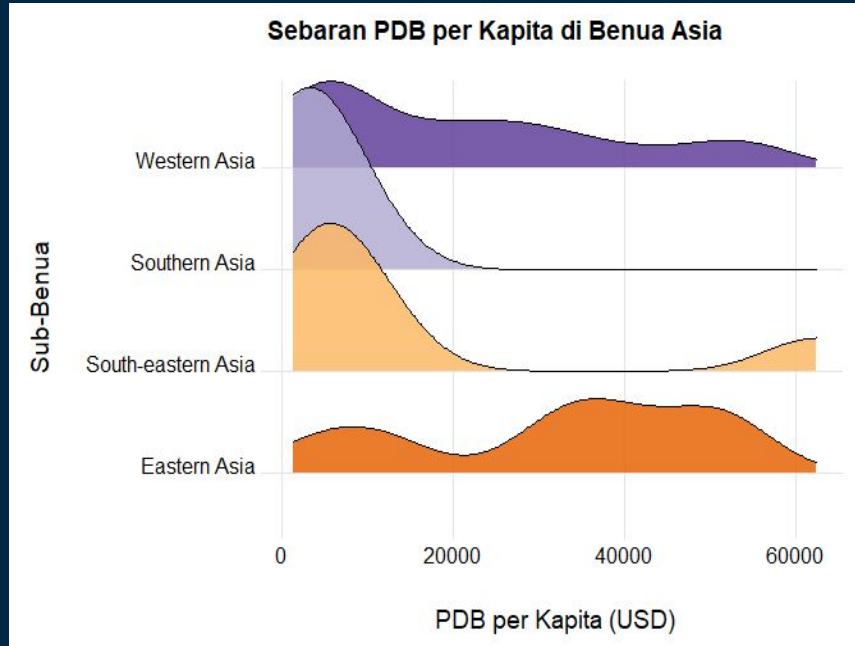
South-Eastern Asia = Normal

The background is a dark blue gradient. It features several thin, vertical white lines of varying lengths. Scattered throughout are small squares in three colors: light blue, pink, and orange. Some squares are solid, while others are outlined. The text is centered and consists of two lines: 'VISUALISASI' in white and 'POST-TRANSFORMATION' in orange.

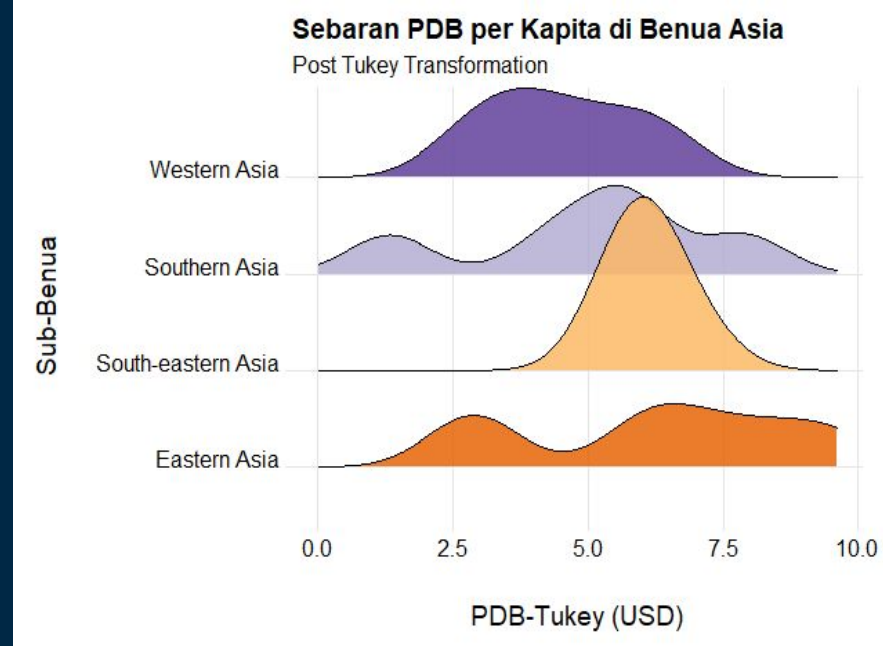
# VISUALISASI POST-TRANSFORMATION

# Transformasi Tukey~PDB per Kapita

Sebelum transformasi

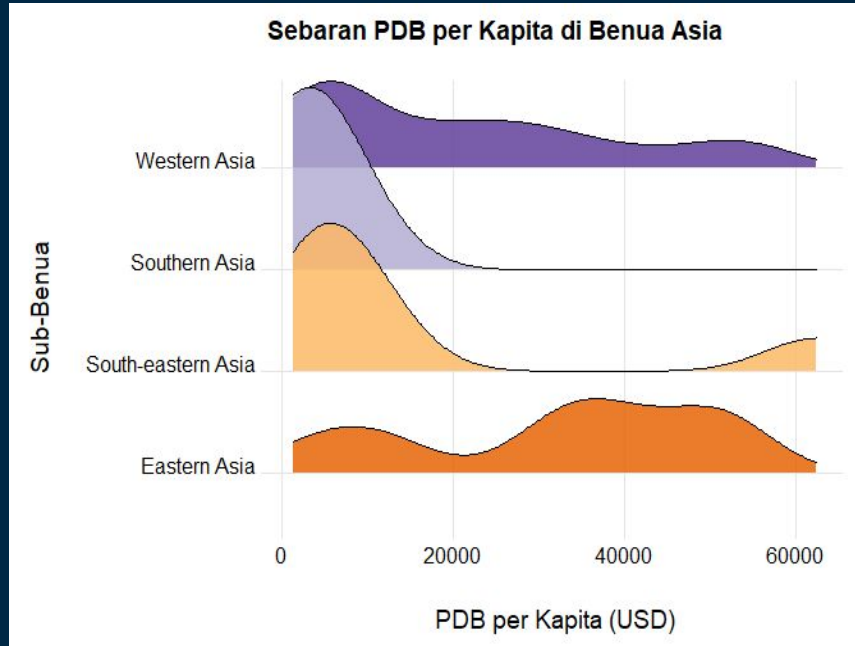


Sesudah transformasi

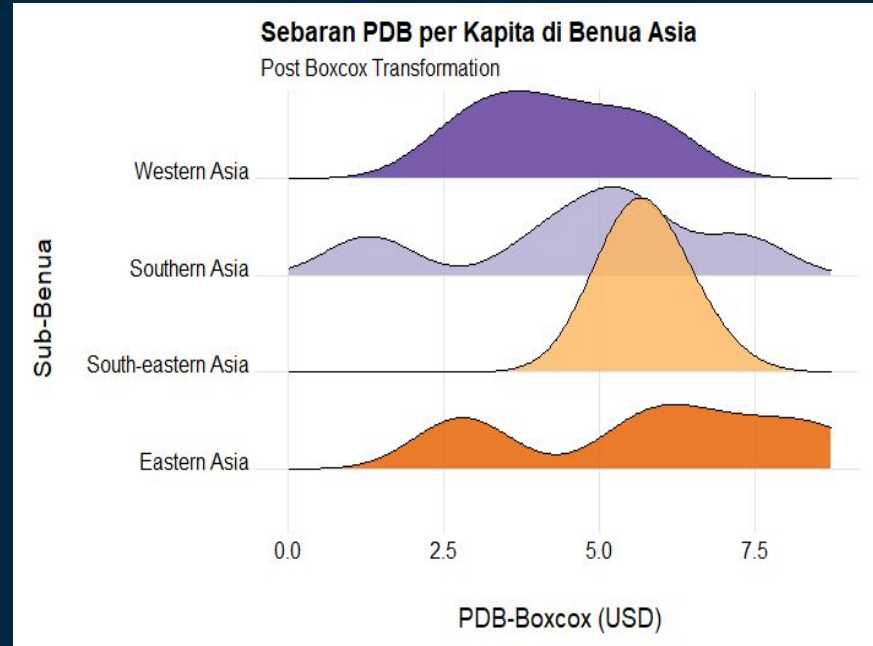


# Transformasi Boxcox~PDB per Kapita

Sebelum transformasi



Sesudah transformasi



# Uji Formal Setelah Transformasi Data

```
> var1 <- DA$`GDP per cap. (USD)`  
> ks.test(var1, "pnorm", mean = mean(var1), sd = sd(var1))
```

One-sample kolmogorov-Smirnov test

```
data: var1  
D = 0.22023, p-value = 0.07642  
alternative hypothesis: two-sided
```

```
> var1.bc <- (var1^lambda1-1)/lambda1  
> ks.test(var1.bc, "pnorm", mean=mean(var1.bc), sd=sd(var1.bc))
```

One-sample kolmogorov-Smirnov test

```
data: var1.bc  
D = 0.16905, p-value = 0.2863  
alternative hypothesis: two-sided
```

Nilai P-Value K-Smirnov Test setelah transformasi data jauh lebih besar dari nilai P-Value sebelum dilakukan transformasi data.

Data hasil transformasi menjadi data menyebar normal dan lebih simetris





# KESIMPULAN PDB

# Kesimpulannya

Pada data awal:

Data terlihat menjulur ke kanan (menyerupai sebaran lognormal) dengan beberapa kelompok data pada tiap sub-benua serta memiliki bimodus (2 puncak).

Pada Transformasi :

Distribution half t untuk gabungan data, sekilas ya.

Boxcox → bimodal distribution,

Boxcox per subregion → Western Asia yang kuat bimodal

Visualisasi Akhir: Data sudah simetris atau menyebar normal dengan ciri bimodal dan berkelompok



# Harga Konsumen

Rataan tahunan dari pertumbuhan  
harga konsumen (%) dalam 5 tahun  
terakhir

02

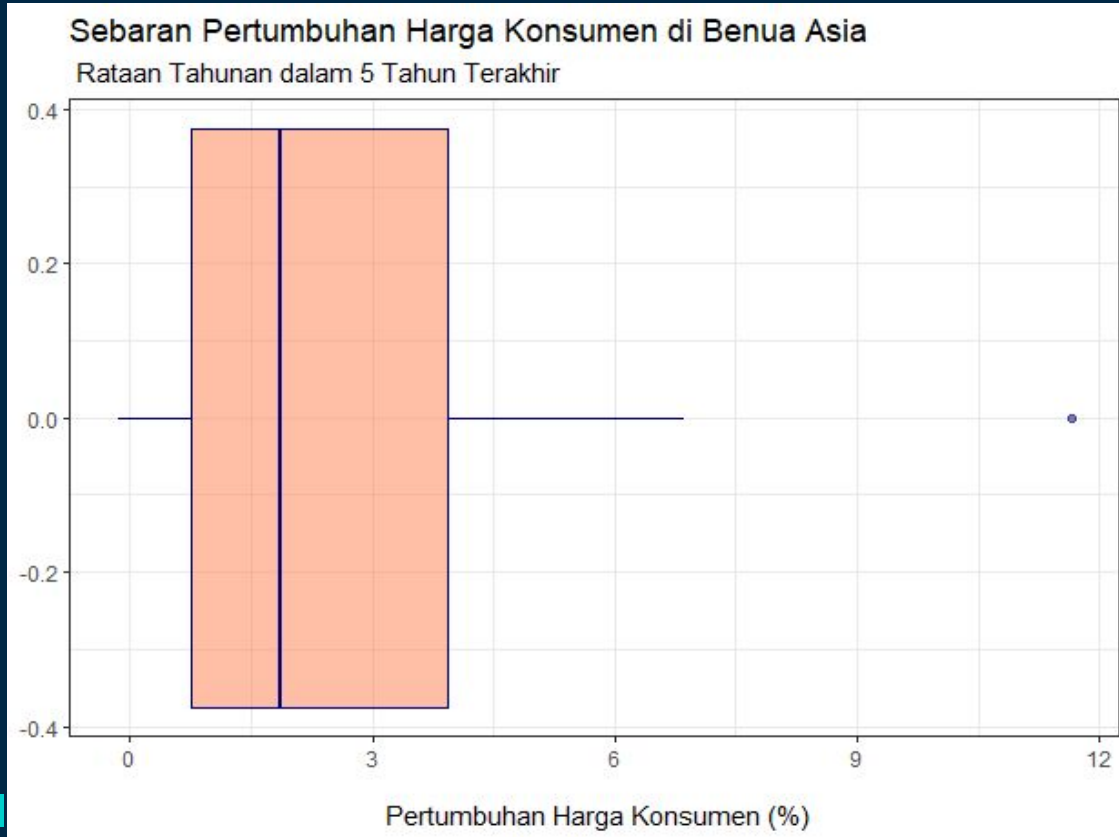
## ■ Harga konsumen

■ Indeks Harga Konsumen (IHK) adalah ukuran yang memeriksa rata-rata tertimbang dari harga sekeranjang barang dan jasa konsumen, seperti transportasi, makanan, dan perawatan medis. Ini dihitung dengan mengambil perubahan harga untuk setiap item dalam keranjang barang yang telah ditentukan dan merata-ratakannya.

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, orange, and teal, some of which are solid while others are hollow. Thin white vertical lines of varying lengths are scattered across the composition, creating a sense of depth and movement.

# EKSPLORASI DATA

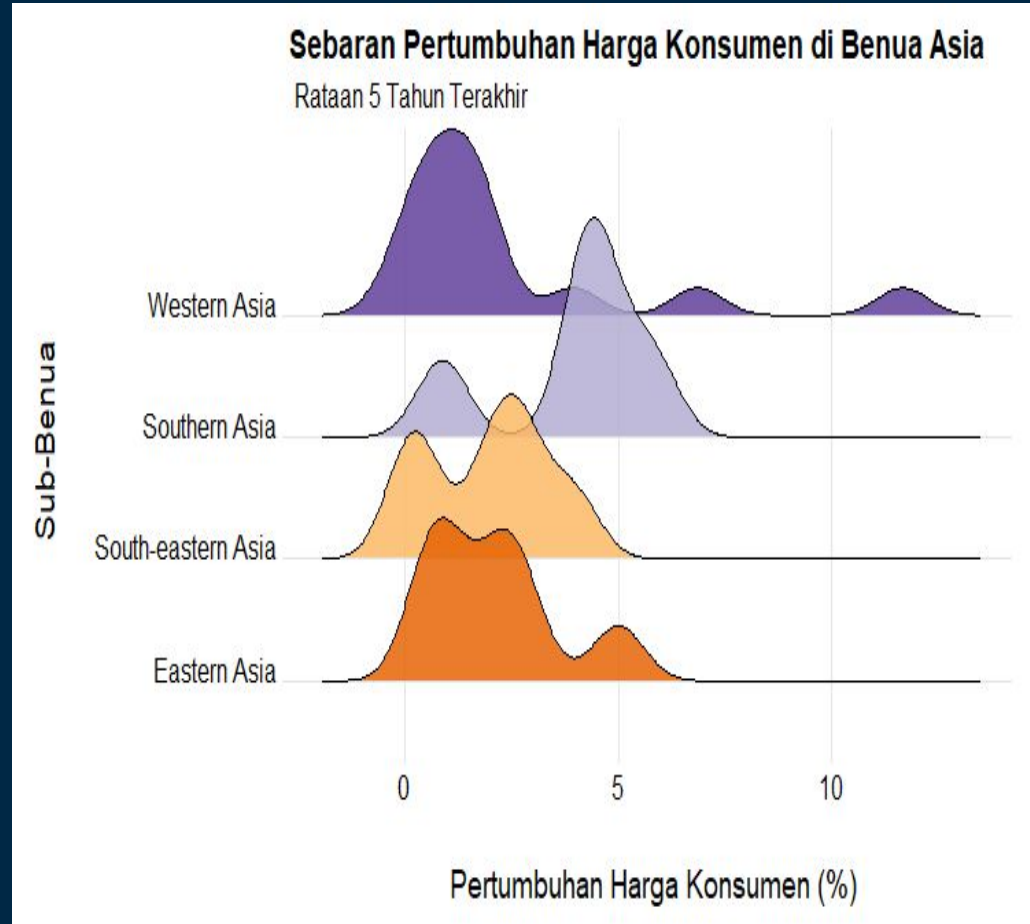
# BOXPLOT



Dari boxplot disamping, terlihat bahwa sebaran data menjulur ke kanan, dengan kata lain sebagian besar negara di Asia memiliki rata-rata kenaikan harga konsumen tahunan yang dekat dengan 25% negara dengan kenaikan terendah, yaitu sekitar 1.8 %.

# DENSITY PLOT

Pada tiap Sub-Benua nampak ada beberapa data yang berkelompok membentuk bimodus, dengan sebagian besar negara memiliki kisaran pertumbuhan harga konsumen 0-6% yang didominasi oleh negara-negara di Asia Barat. Namun, kenaikan tertinggi yaitu sekitar 10-12% juga terjadi pada beberapa negara di sub-benua ini.



# Uji Formal dengan K-Smirnov Test

Consumer prices ( avg annual avg. % growth 5yrs)

```
> ##Uji Formal Kolmogorov-Smirnov Test  
> ## Normal  
> set.seed(42)  
> ks.test(C, "pnorm", mean = mean(C), sd = sd(C))
```

One-sample Kolmogorov-Smirnov test

```
data: C  
D = 0.19469, p-value = 0.1326  
alternative hypothesis: two-sided
```

```
> ## Exponential  
> set.seed(42)  
> ks.test(log(C), "pexp", rate = 1/mean(C))
```

One-sample Kolmogorov-Smirnov test

```
data: log(C)  
D = 0.43286, p-value = 3.884e-06  
alternative hypothesis: two-sided
```

```
> ## Chi-Squared  
> set.seed(42)  
> ks.test(C, "pchisq", df = mean(C))
```

One-sample Kolmogorov-Smirnov test

```
data: C  
D = 0.14481, p-value = 0.4332  
alternative hypothesis: two-sided
```

```
> ## Lognormal  
> set.seed(42)  
> ks.test(C, "plnorm", meanlog = mean(log(C)), sdlog = sd(log(C)))
```

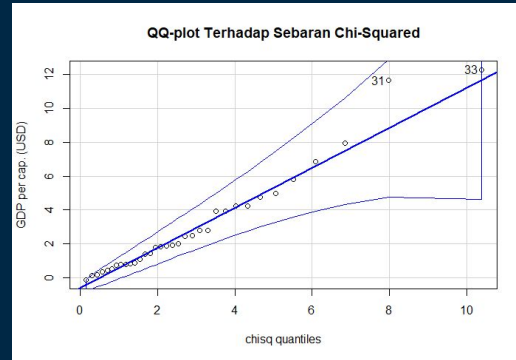
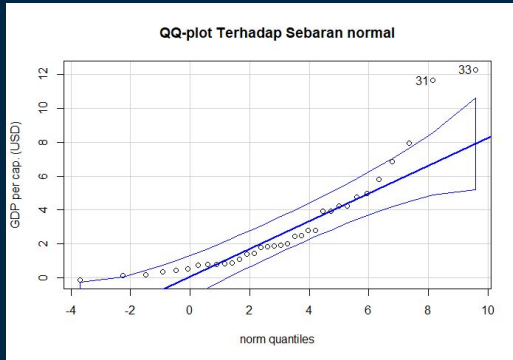
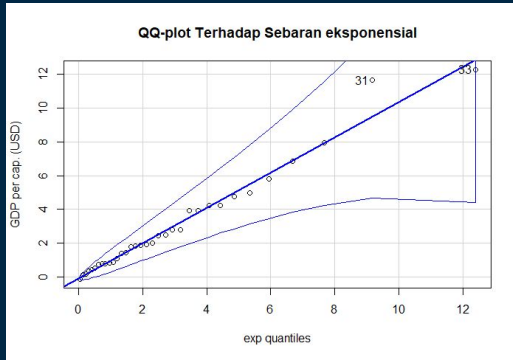
One-sample Kolmogorov-Smirnov test

```
data: C  
D = NA, p-value = NA  
alternative hypothesis: two-sided
```

Berdasarkan hasil Uji Formal K-Smirnov, diperoleh hanya P-Value sebaran Normal dan Chi-Square lebih besar dibandingkan nilai alpha (0.05) atau Terima  $H_0$ . Sehingga dapat disimpulkan sebaran data Harga Konsumen mendekati sebaran Normal dan Chi-Square (lebih baik).



# QQPLOT – Consumer prices ( avg annual avg. % growth 5yrs)



Hasil dari Uji Formal K-Smirnov diperoleh bahwa sebaran data GDP mendekati sebaran Normal dan Chi-Square. Hal ini selaras dengan hasil QQPLOT pada sebaran Normal dan Chi-Square yang cenderung mendekati garis diagonal sebarannya. Namun dapat dilihat sebaran Chi-Square memiliki sebaran lebih mendekati garis diagonalnya dibandingkan dengan sebaran Normal

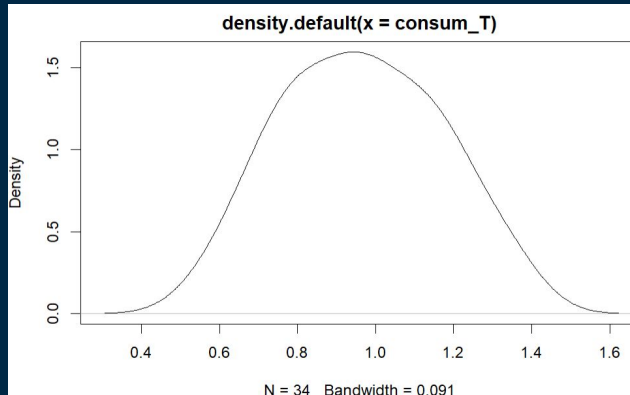
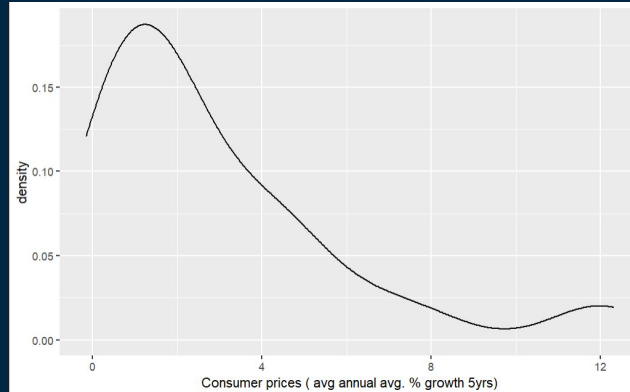
The background is a dark blue gradient. It is decorated with various geometric elements: thin white vertical lines of varying lengths, small squares in teal, orange, and pink, and larger squares in teal and orange. The text 'TRANSFORMASI DATA' is centered in the middle of the image.

# TRANSFORMASI DATA

# Transformasi consumer Price avg 5Year

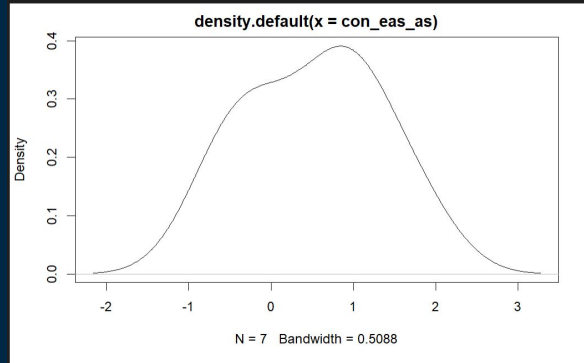
Dengan menggunakan boxcox,  
maka hasilnya adalah Seperti  
berikut ya

$$\lambda = -0.585858$$

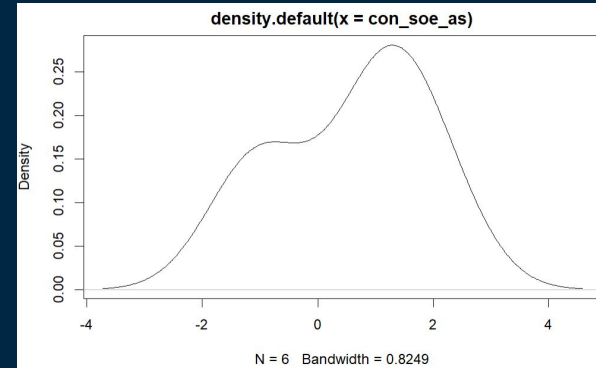


# Transformasi consumer Price avg 5 Year subregion

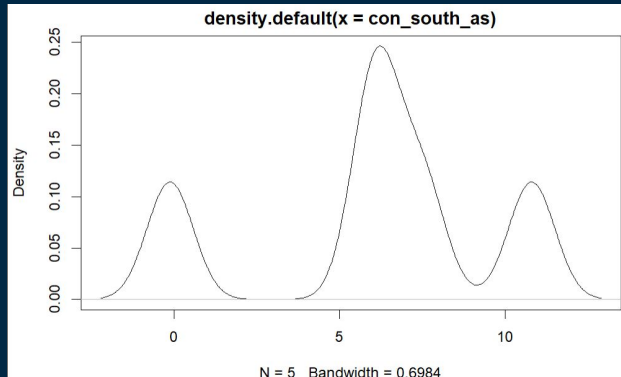
Eastern Asia



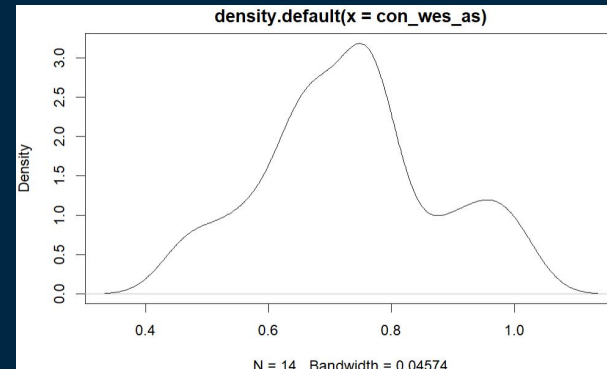
South eastern Asia



Southern Asia



Western Asia



# Transformasi consumer Price avg 5 Year subregion

Shapiro-wilk normality test

data: con\_eas\_as  
W = 0.96659, p-value = 0.8729

Eastern Asia = Normal

Shapiro-wilk normality test

data: con\_soe\_as  
W = 0.91107, p-value = 0.4435

South eastern Asia = Normal

Shapiro-wilk normality test

data: con\_south\_as  
W = 0.92189, p-value = 0.5422

South Asia = Normal

Shapiro-wilk normality test

data: con\_wes\_as  
W = 0.97542, p-value = 0.9395

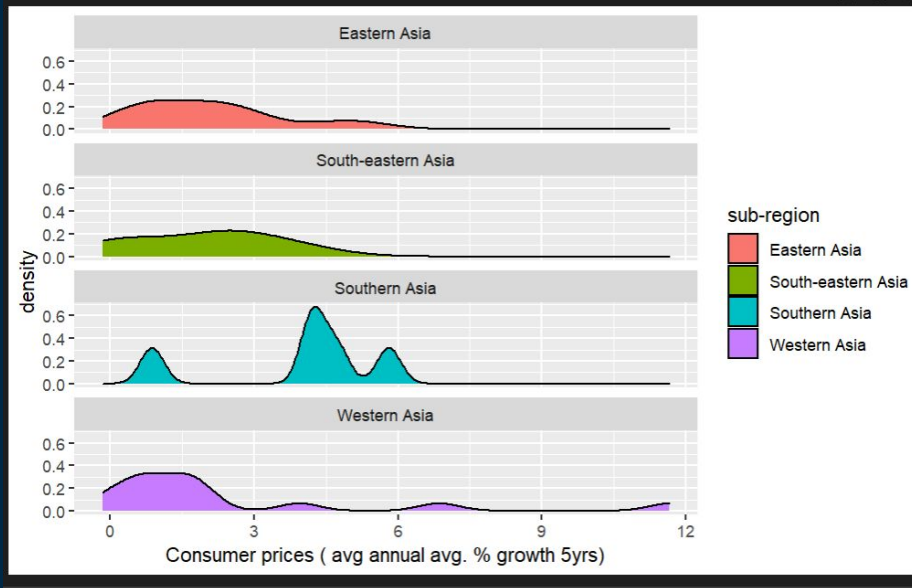
western Asia = Normal

The background is a dark blue gradient. It features several thin, vertical white lines of varying lengths. Scattered throughout are small squares in three colors: light blue, pink, and orange. Some squares are solid, while others are outlined. The text is centered and consists of two lines: 'VISUALISASI' in white and 'POST-TRANSFORMATION' in orange.

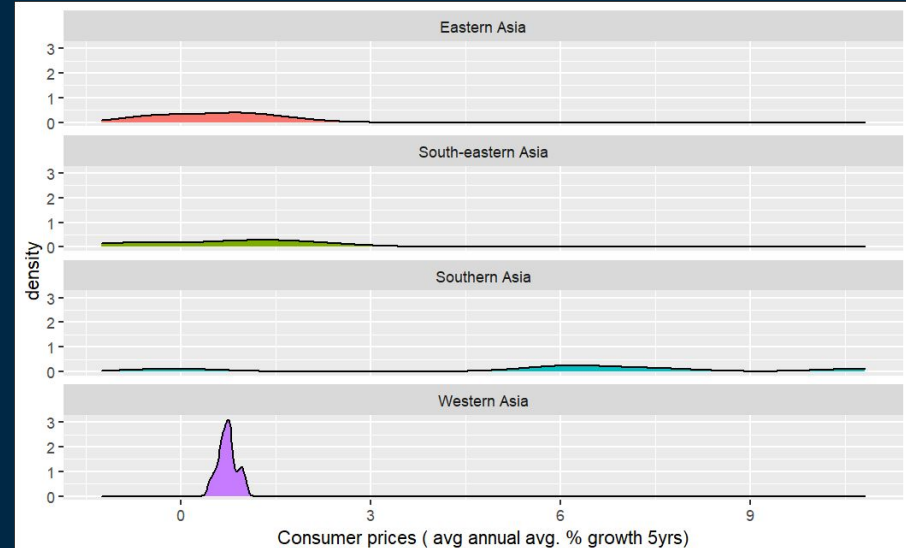
# VISUALISASI POST-TRANSFORMATION

# Transformasi consumer Price avg 5 Year subregion

Sebelum di transformasi



Sesudah Di transformasi



The background is a dark blue field decorated with various geometric elements. There are several thin white vertical lines of varying lengths. Scattered throughout are small squares in three colors: light blue, pink, and orange. Some of these squares are solid, while others are hollow outlines. The overall aesthetic is modern and minimalist.

# KESIMPULAN HARGA KONSUMEN



# Kesimpulannya

Pada data awal :

Data masih terlihat menjulur ke kanan pada boxplot, walaupun ketika divisualisasikan oleh density plot untuk tiap sub-benua dan hasil dari uji formal, data menyerupai sebaran normal ataupun chi-square.

Pada Transformasi : Data tersebut sangat lah bagus ditransformasi oleh boxcox, dan digrupkan sub region juga dominan ke normal setelah ditransformasi

Visualisasi akhir: Data sudah menyebar normal baik secara visual maupun menggunakan uji formal.



The background is a dark blue field decorated with various geometric elements. There are several thin white vertical lines of varying lengths. Scattered throughout are small squares in three colors: light blue, light orange, and light pink. Some of these squares are solid, while others are just outlines. The text 'TERIMA KASIH' is centered in the middle of the image. The word 'TERIMA' is in white, and 'KASIH' is in a light orange color, matching one of the background accents.

TERIMA KASIH