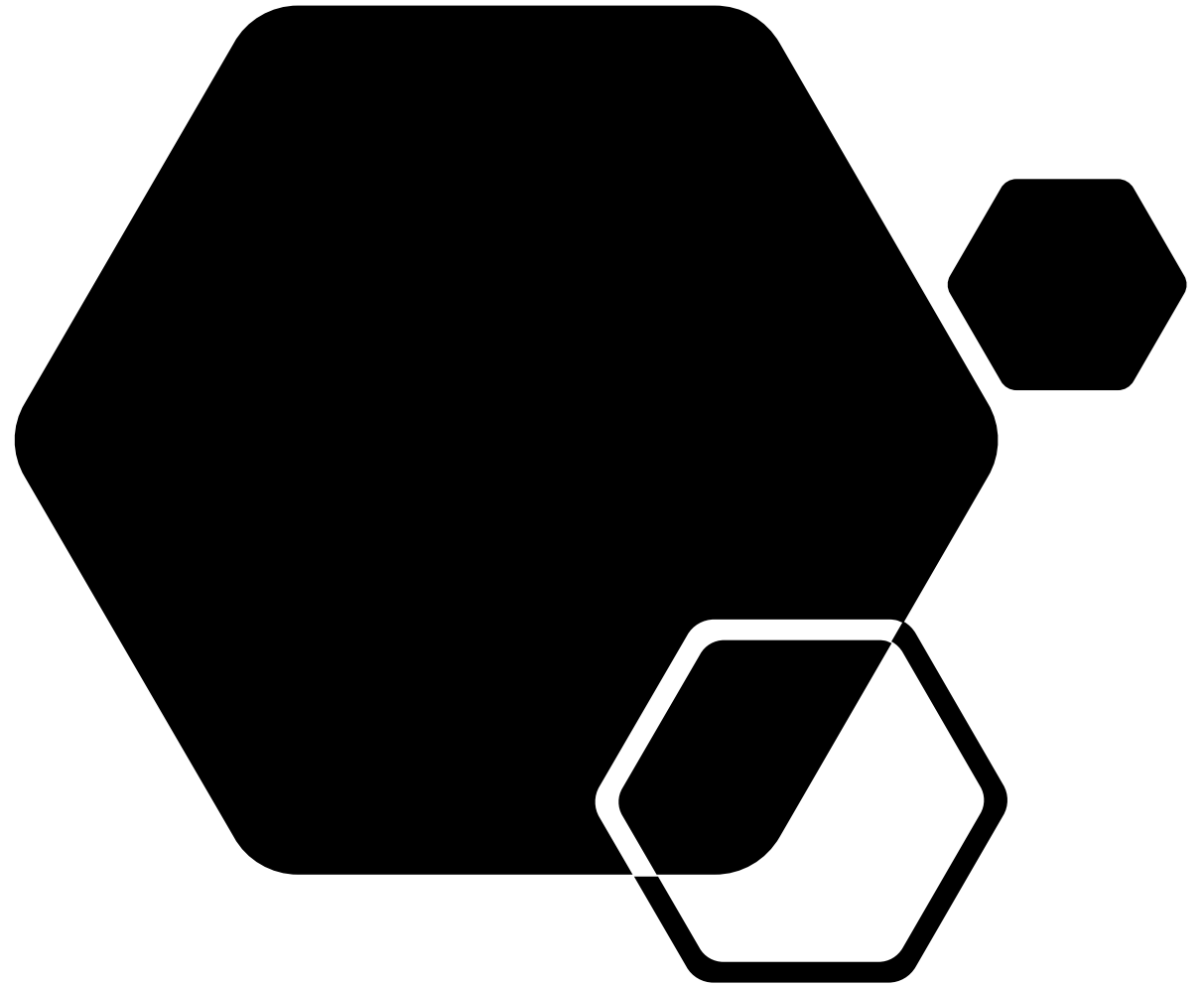


Analisis Data Eksploratif

Pertemuan 6

Transformasi Angkatan Data

Transformasi Angkatan Data



Transformasi Data

- Transformasi angkatan merupakan cara bagaimana membentuk angkatan menjadi “lebih baik” dalam arti memenuhi anggapan yang digunakan dalam statistik inferensi atau statistik konfirmasi yaitu suatu angkatan berdistribusi normal.
- Angkatan yang memiliki puncak tunggal dan tidak simetris dibuat menjadi simetris sehingga dapat mendekati distribusi normal yang lebih baik daripada data asli.

Beberapa Jenis Transformasi Data

1. Transformasi Square Root (Akar),
2. Transformasi Logaritma,
3. Transformasi Arcsin,
4. Transformasi Square (Kuadrat),
5. Transformasi Cubic (Pangkat Tiga),
6. Transformasi Inverse (Kebalikan),
7. Transformasi Inverse Square Root (Kebalikan Akar),
8. Transformasi Inverse Square (Kebalikan Kuadrat),
9. Transformasi Inverse Cubic (Kebalikan Pangkat Tiga), dsb

Transformasi Logaritma

- Transformasi logaritma suatu angkatan didapat dengan mengambil logaritma setiap observasi dalam Angkatan.
- Transformasi Logaritma digunakan apabila data tidak memenuhi asumsi pengaruh aditif.
- Misalkan X adalah data asli dan X' adalah data hasil transformasi sehingga $X' = \log X$.

- Ada beberapa hal yang perlu diperhatikan dalam penggunaan transformasi logaritma, yaitu:

1. Apabila **data asli** menunjukkan sebaran nilai **kurang dari 10** atau nilai **mendekati nol**, maka digunakan transformasi

$$\log X + 1$$

2. Apabila **data asli banyak mengandung nilai nol**, maka sebaiknya digunakan **transformasi** yang **lain**, misalnya transformasi akar, dsb

3. Apabila **data asli banyak yang mendekati nol** (misalnya bilangan desimal), maka semua data dikalikan 10 sebelum dijadikan ke logaritma. Jadi, gunakan transformasi

$$X' = \log(10X)$$

Contoh

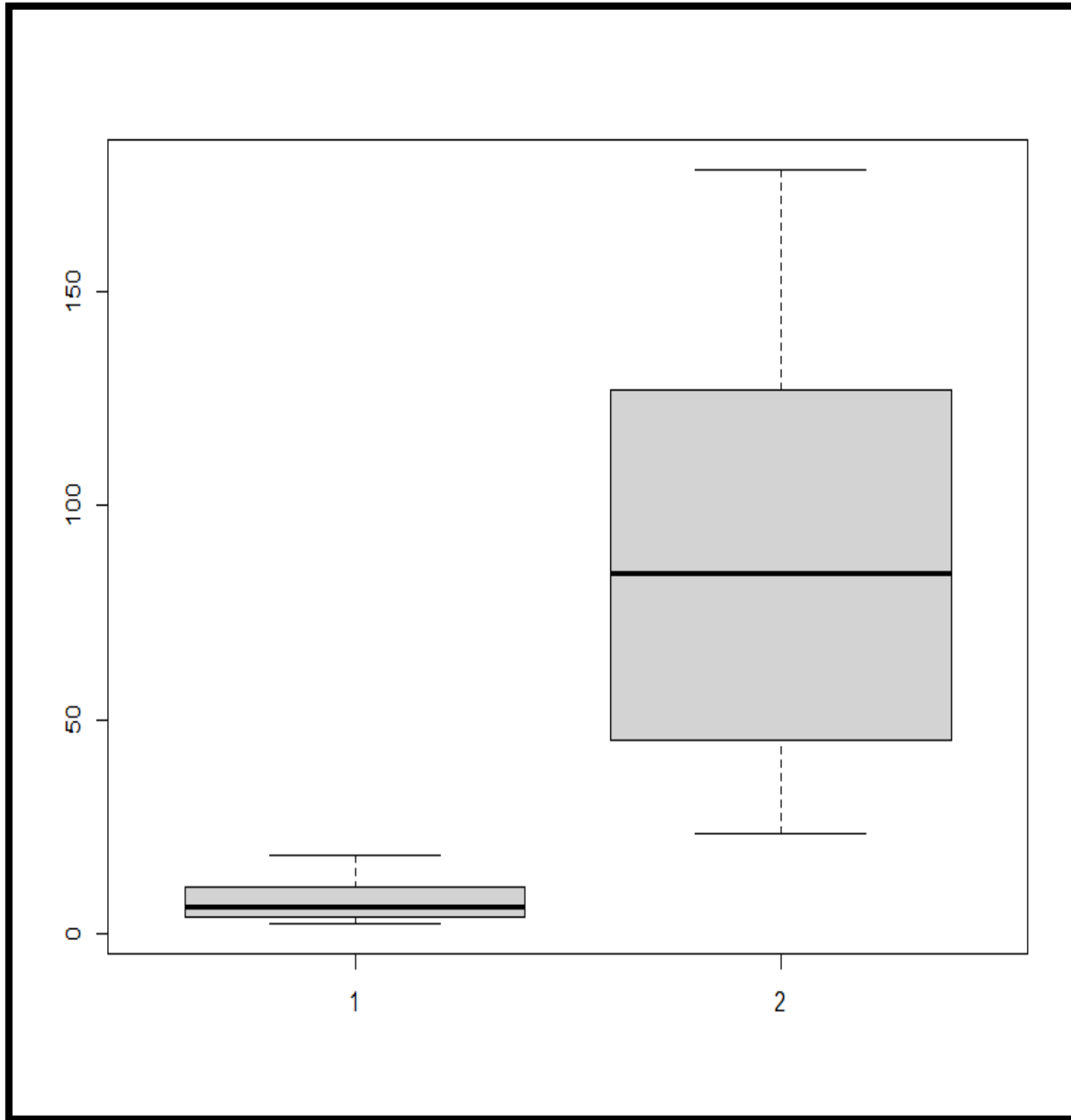
Berikut diberikan data jumlah penduduk Amerika Serikat (AS) dan Kanada pada 12 sensus yang berturutan. Kolom pertumbuhan yang berisi perbedaan antara setiap sensus dengan sensus sebelumnya, terlihat besar pertumbuhan penduduk.

Penduduk Kanada			Penduduk AS		
Tahun	Jumlah (jutaan)	Pertumbuhan	Tahun	Jumlah (jutaan)	Pertumbuhan
1850	2,44		1850	23,2	
1860	3,23	0,79	1860	31,4	8,2
1870	3,69	0,46	1870	39,8	8,4
1880	4,32	0,63	1880	50,2	10,4
1890	4,83	0,51	1890	62,9	12,7
1900	5,37	0,54	1900	76,0	13,1
1910	7,21	1,84	1910	92,0	16
1920	8,79	1,58	1920	105,7	13,7
1930	10,38	1,59	1930	122,8	17,1
1940	11,51	1,13	1940	131,7	8,9
1950	14,01	2,50	1950	150,7	19
1960	18,24	4,23	1960	178,5	27,8

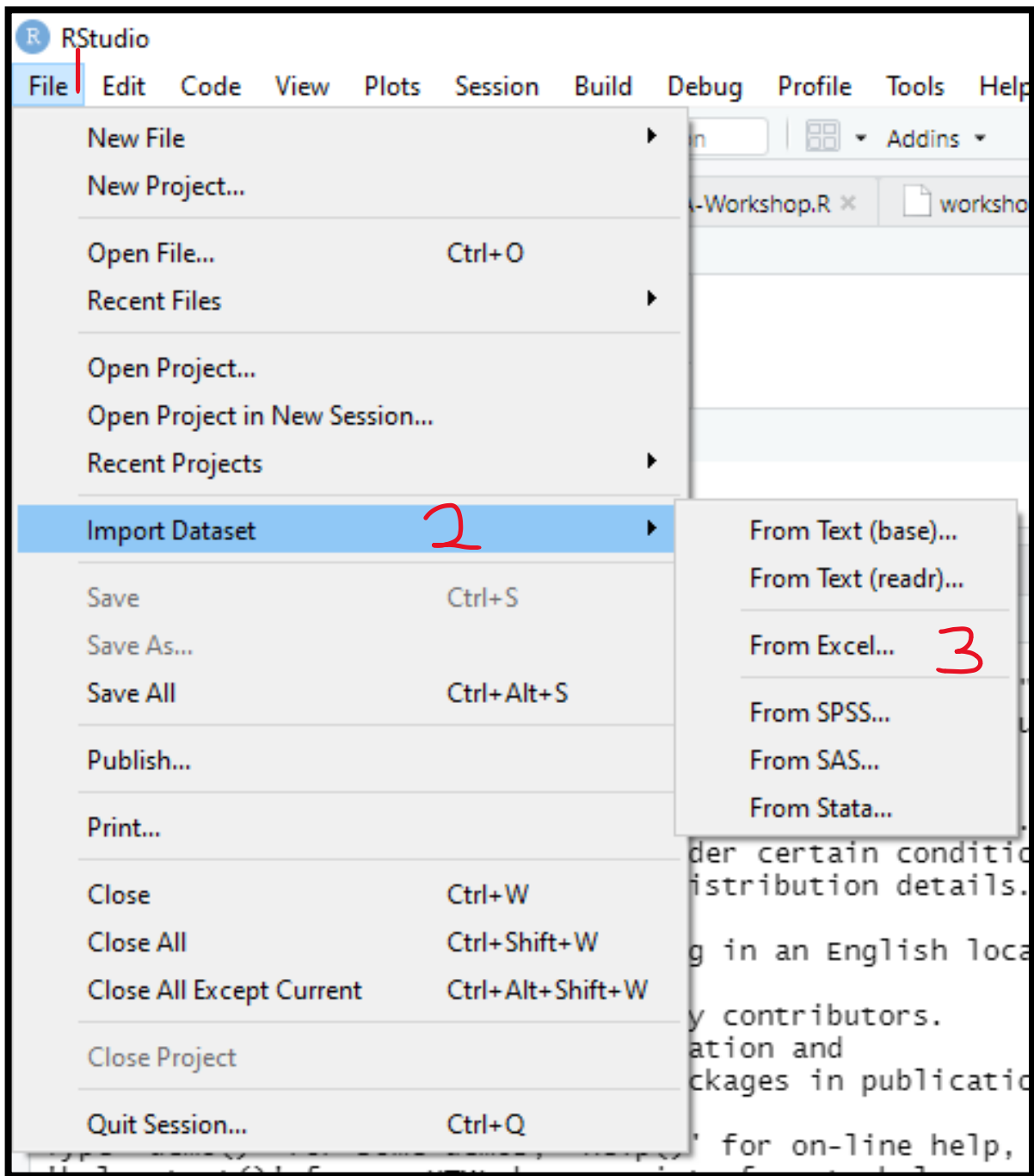
- Tampak bahwa besarnya pertumbuhan penduduk pada tahun terakhir lebih besar daripada tahun sebelumnya.
- Penduduk AS mempunyai pusat dan sebaran yang jauh lebih besar.

```
> summary(Data_1$`Jumlah Penduduk Kanada (jutaan)`)  
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
 2.440  4.162   6.290   7.835 10.662  18.240  
  
> summary(Data_1$`Jumlah Penduduk AS (Jutaan)`)  
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
 23.20  47.60   84.00   88.74 125.03  178.50
```

```
> boxplot(Data_1$`Jumlah Penduduk Kanada (jutaan)`, Data_1$`Jumlah Penduduk AS  
(Jutaan)`)
```

- Diagram kotak dan titik pada Gambar di samping tampak bahwa median lebih dekat dengan kuartil bawah atau nilai tinggi lebih menyebar daripada nilai rendah.
- Keadaan angkatan yang seperti ini disebut **“menjurai ke atas”**.
- Sebaliknya bila median dekat dengan kuartil atas atau nilai rendah lebih menyebar dibandingkan dengan nilai tinggi, maka angkatan seperti ini disebut **“menjurai ke bawah”**.
- Diagram kotak dan titik menunjukkan bahwa pusat dan sebaran berubah pada arah sama (pusat lebih besar mempunyai sebaran yang lebih besar).



Import Dataset Dari File Excel

1. Klik Menu File – Import Dataset – From Excel
2. Klik Browse – cari file excel yang akan diimport
3. Klik Import

Import Excel Data

File/URL:

Browse...

4

Data Preview:

Import Options:

Name: dataset

Max Rows:

☒ First Row as Names

Sheet: Default

Skip: 0

☒ Open Data Viewer

Range: A1:D10

NA:

Code Preview:

```
library(readxl)
dataset <- read_excel(NULL)
view(dataset)
```

5

? Reading Excel files using readxl

Import

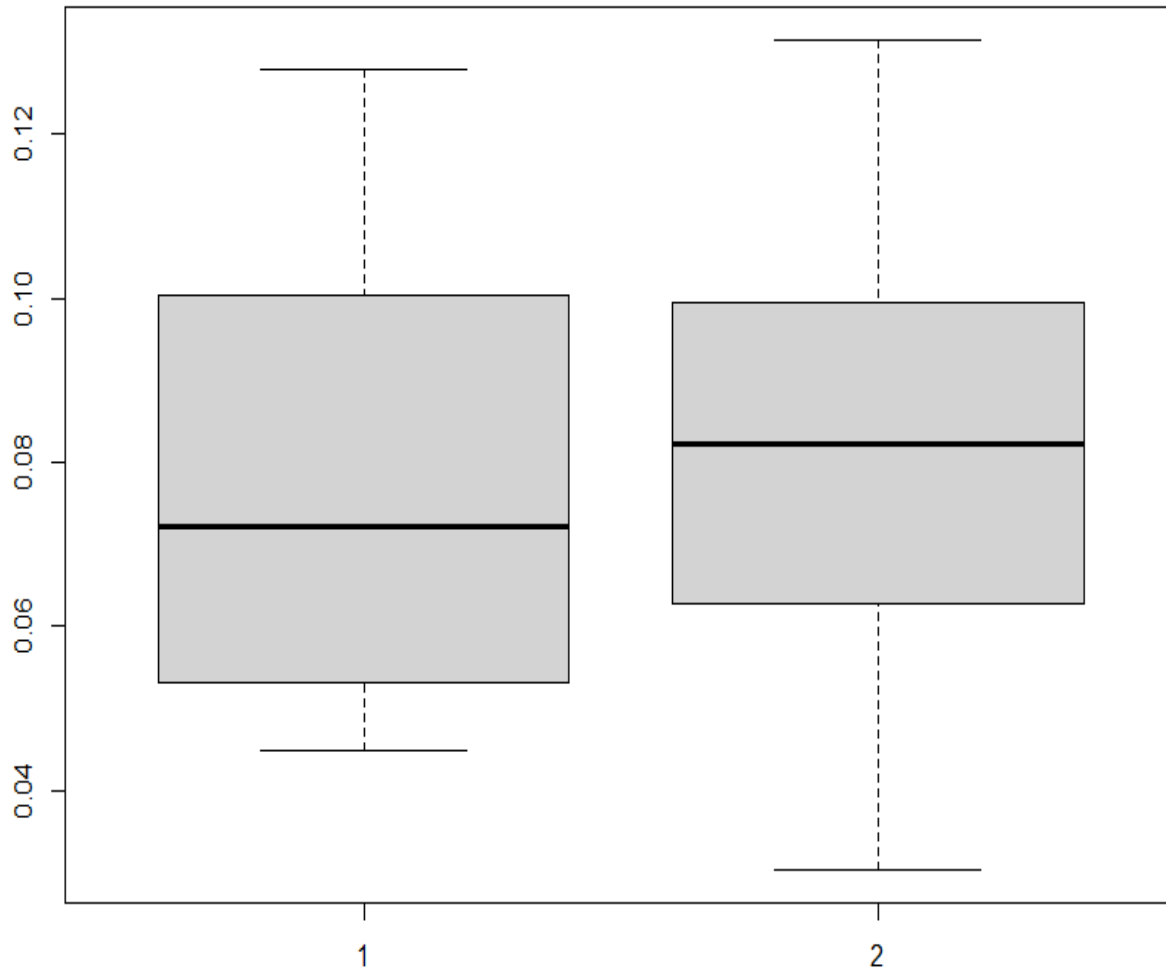
Cancel

- Tabel di samping berisi nisbah (hasil bagi) jumlah penduduk pada suatu sensus dengan sensus sebelumnya.
- Nisbah ini menunjukkan semacam laju pertumbuhan.
- Dari tahun ke tahun kedua negara mempunyai besar nisbah yang hampir sama.
- Hal ini menunjukkan bahwa pertumbuhan penduduk di kedua negara tetap.

Tahun	Nisbah	Tahun	Nisbah
1860/1850	1.32	1860/1850	1.35
1870/1860	1.14	1870/1860	1.27
1880/1870	1.17	1880/1870	1.26
1890/1880	1.12	1890/1880	1.25
1900/1890	1.11	1900/1890	1.21
1910/1900	1.34	1910/1900	1.21
1920/1910	1.22	1920/1910	1.15
1930/1920	1.18	1930/1920	1.16
1940/1930	1.11	1940/1930	1.07
1950/1940	1.22	1950/1940	1.14
1960/1950	1.30	1960/1950	1.18

- **Keuntungan menggunakan nisbah:** Mudah, sering digunakan dan jelas menggunakan ide laju pertumbuhan penduduk.
- **Kerugian menggunakan nisbah:** Menghitung nisbah tanpa bantuan kalkulator akan membuang waktu. Dari nisbah tidak dapat diperoleh bilangan semula.
- Untuk itu dicari transformasi yang tidak mengesampingkan keuntungan nisbah, sekaligus menghilangkan kerugiannya.
- Transformasi tersebut adalah transformasi logaritma.

Tahun	Pertumbuhan	log Nisbah	Tahun	Pertumbuhan	log Nisbah
1860/1850	1.32	0.12	1860/1850	1.35	0.13
1870/1860	1.14	0.06	1870/1860	1.27	0.10
1880/1870	1.17	0.07	1880/1870	1.26	0.10
1890/1880	1.12	0.05	1890/1880	1.25	0.10
1900/1890	1.11	0.05	1900/1890	1.21	0.08
1910/1900	1.34	0.13	1910/1900	1.21	0.08
1920/1910	1.22	0.09	1920/1910	1.15	0.06
1930/1920	1.18	0.07	1930/1920	1.16	0.07
1940/1930	1.11	0.04	1940/1930	1.07	0.03
1950/1940	1.22	0.09	1950/1940	1.14	0.06
1960/1950	1.30	0.11	1960/1950	1.18	0.07



- Apabila logaritma menangkap ide laju pertumbuhan yang tetap, angkatan seharusnya tidak menjurai ke atas setelah transformasi logaritma.
- Hal ini akan anda lihat dari Gambar yang menyajikan diagram kotak dan titik angkatan baru.
- Dari Tabel tampak bahwa kedua negara juraian ke atas telah hilang. Untuk Kanada kedua ekstrim dan kuartil berimbang diseitar median sedang untuk AS angkatan agak menjurai ke bawah.
- Tetapi secara keseluruhan harga yang lebih tinggi dan lebih rendah daripada median tampak menyebar merata.

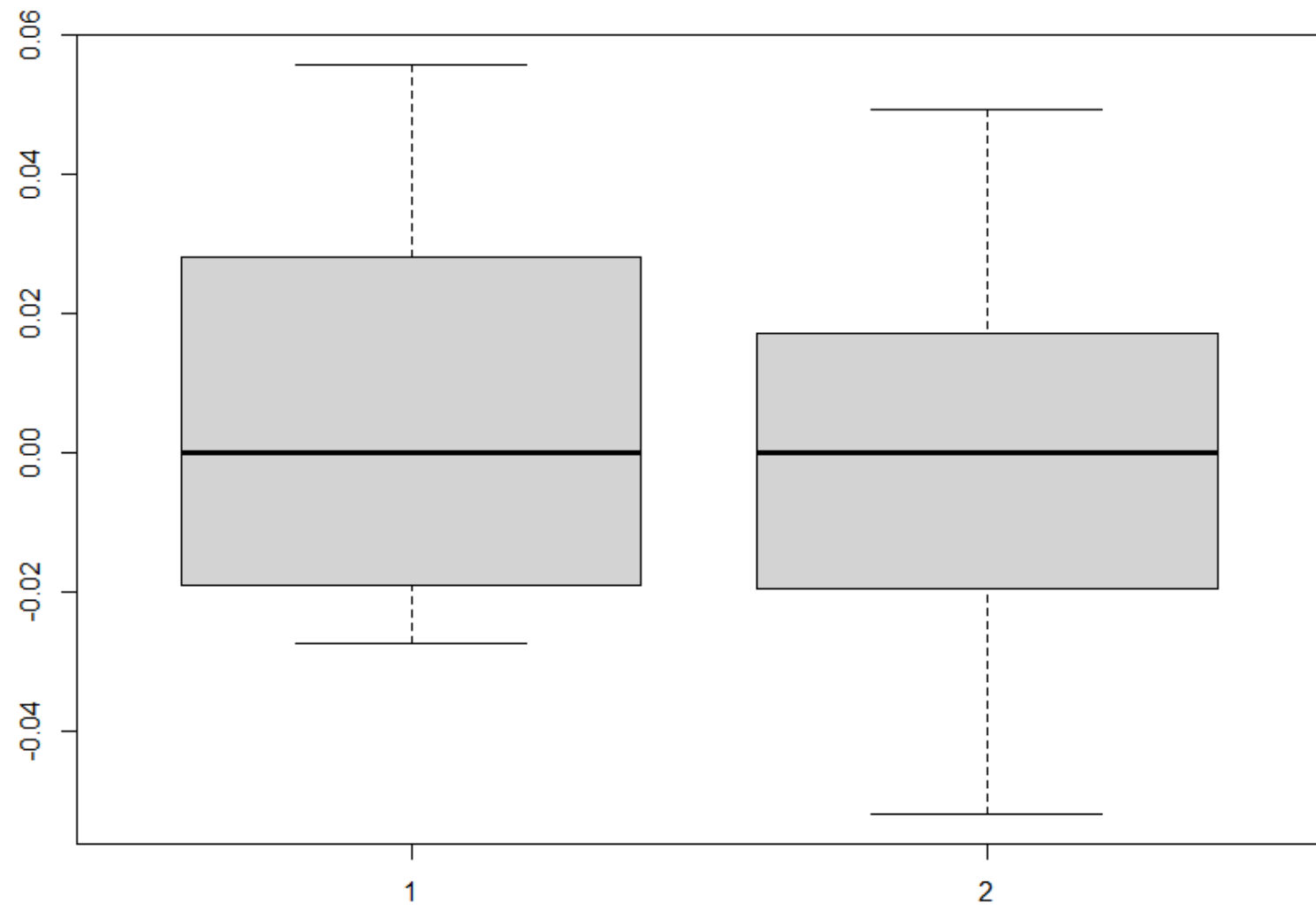
```
> summary(Nisbah_Data_1$`log Nisbah...3` )
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.04488 0.05314 0.07221 0.07942 0.10032 0.12796

> summary(Nisbah_Data_1$`log Nisbah...6` )
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03039 0.06271 0.08216 0.08056 0.09938 0.13144
```

- Tampak bahwa untuk kedua negara mediannya hampir sama. Hal ini menunjukkan bahwa kedua negara mempunyai laju pertumbuhan penduduk yang sama.
- Bila laju pertumbuhan tetap, setelah pertumbuhan pusat (median) dikeluarkan pertumbuhan log penduduk mempunyai sisa nol.

Tahun	Pertumbuhan	log Nisbah	log Nisbah - Median	Tahun	Pertumbuhan	log Nisbah	log Nisbah - Median
1860/1850	1.32	0.12	0.05	1860/1850	1.35	0.13	0.05
1870/1860	1.14	0.06	-0.01	1870/1860	1.27	0.10	0.02
1880/1870	1.17	0.07	0.00	1880/1870	1.26	0.10	0.02
1890/1880	1.12	0.05	-0.02	1890/1880	1.25	0.10	0.02
1900/1890	1.11	0.05	-0.03	1900/1890	1.21	0.08	0.00
1910/1900	1.34	0.13	0.06	1910/1900	1.21	0.08	0.00
1920/1910	1.22	0.09	0.01	1920/1910	1.15	0.06	-0.02
1930/1920	1.18	0.07	0.00	1930/1920	1.16	0.07	-0.02
1940/1930	1.11	0.04	-0.03	1940/1930	1.07	0.03	-0.05
1950/1940	1.22	0.09	0.01	1950/1940	1.14	0.06	-0.02
1960/1950	1.30	0.11	0.04	1960/1950	1.18	0.07	-0.01

- Bila laju pertumbuhan tetap, setelah pertumbuhan pusat (median) dikeluarkan pertumbuhan log penduduk mempunyai sisa nol.
- Dari Tabel di atas tampak bahwa sisa tidak seluruhnya nol.
- Ada kalanya penduduk bertambah lebih dari 0.09 (memberikan sisa positif) atau kurang dari 0,08 (memberikan sisa negatif).



- Dengan mendapatkan sisa, selain kita dapat melihat pola pertumbuhan log penduduk yang hampir konstan tetapi data asli tidak terlihat dengan jelas.
- Penggunaan utama transformasi adalah untuk membuat angkatan berbentuk standar, yaitu bentuk yang simetri.
- Hal ini ada hubungannya dengan statistika inferensi (konfirmasi) yang sering membutuhkan pengandaian bahwa angkatan berdistribusi normal.
- Suatu angkatan yang simetris dan berpuncak tunggal mungkin tidak begitu normal, tetapi bisa cukup mendekati normal (terutama bila jumlah observasi angkatan besar), sehingga teknik konfirmasi yang menganggap berbentuk normal dapat digunakan dengan aman

Memilih Transformasi Yang Terbaik

- Memilih transformasi yang unggul merupakan cara yang digunakan untuk menemukan transformasi yang sesuai bagi suatu angkatan agar bentuknya mendekati simetri.
- Berikut ini merupakan data pembangunan perumahan swasta bukan petani di negara Amerika Serikat pada tahun 1966.

Bulan	1966
Januari	78500
Februari	74800
Maret	115900
April	138600
Mei	126700
Juni	118200
Juli	97600
Agustus	99600
September	86900
Oktober	74400
November	71400
Desember	58900
Seluruh Tahun	1141500

Stem &
Leaf

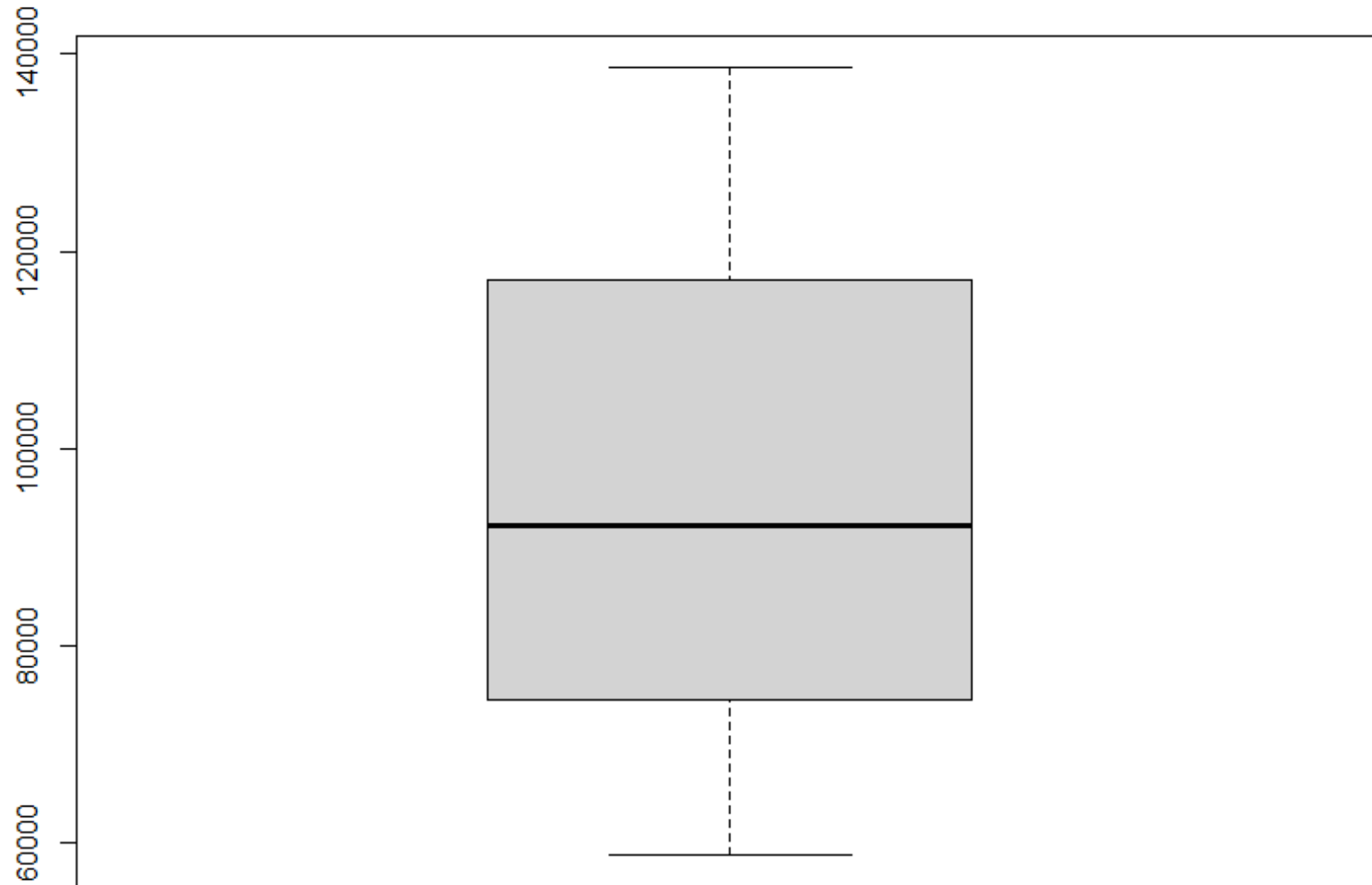
```
stem(Data_2$`Pembangunan Perumahan 1966`, scale = 5, 100, 10000)
```

The decimal point is 4 digit(s) to the right of the |

```

5 | 9
6 |
7 | 1459
8 | 7
9 | 8
10 | 0
11 | 68
12 | 7
13 | 9

```



Tampak bahwa angkatan sedikit menjurai ke atas.

Transformasi yang cocok untuk angkatan yang menjurai ke atas adalah akar pangkat dua, logaritma dan kebalikan negative (tanda negatif untuk menjamin supaya urutan angkatan tetap).

Bulan	Pembangunan	Dibulatkan	Akar	Log	$-1/x$
Januari	78500	79	280	4,89	-1,27E-05
Februari	74800	75	273	4,87	-1,34E-05
Maret	115900	116	340	5,06	-8,63E-06
April	138600	139	372	5,14	-7,22E-06
Mei	126700	127	356	5,1	-7,89E-06
Juni	118200	118	344	5,07	-8,46E-06
Juli	97600	98	312	4,99	-1,02E-05
Agustus	99600	100	316	5	-1E-05
September	86900	87	295	4,94	-1,15E-05
Oktober	74400	74	273	4,87	-1,34E-05
November	71400	71	267	4,85	-1,4E-05
Desember	58900	59	243	4,77	-1,7E-05

akar dua		Log X		$-1/x$	
3	70,60	5,1	40	-7	82
3	40,40,20,10	5,0	760	-8	64
2	94,82,76,70,64	4,9	940	-9	
2	46	4,8	875	-10	20
batang : Ratusan		4,7	7	-11	6
		batang : satuan dan Persepuluhan		-12	8
				-13	26
				-14	0
				-15	
				-16	8
				batang : Persepuluhan Juta	
q _A	340	q _A	5,07	q _A	-86
Md	320	Md	4,97	Md	-109
q _b	276	q _b	4,88	q _b	-134

- Tampak bahwa untuk angkatan ini akar pangkat dua bukan transformasi yang cukup kuat, karena tampak bahwa angkatan tetap menjurai ke atas.
- Logaritma tampak baik karena juraian telah hilang.
- Kebalikan negatif tampak terlalu kuat karena menjadi menjurai ke bawah.

- Secara umum transformasi mempunyai pengaruh yang berbeda untuk angkatan yang berbeda.
- Untuk angkatan yang menjurai ke atas transformasi yang cocok adalah akar pangkat dua (\sqrt{x}), logaritma ($\log x$) dan kebalikan negatif ($-\frac{1}{x}$).
- Untuk angkatan yang menjurai ke bawah x^2, x^3, x^4 , atau pangkat yang lain merupakan transformasi yang mungkin sesuai.

- Transformasi antilogaritma juga dapat mengoreksi juaraian ke bawah, tetapi apabila bilngan besar maka antilog menjadi sangat besar.
- Akibatnya transformasi ini jarang dipakai untuk mengembalikan transformasi log ke angkatan semula.
- Untuk angkatan yang menjurai ke bawah pangkat dua atau tiga biasanya sudah memadai. Sebaliknya, biasakan membuat lebih dari satu transformasi untuk dipilih mana yang lebih baik.

Transformasi Satu Angkatan

- Tujuan transformasi jelas yaitu membuat sedekat mungkin dengan bentuk standar yaitu puncak tunggal, simetris, mengecil dengan mulus di kedua sisinya.
- Perlu diingat bahwa tengah dari data lebih penting dari yang lain.
- Jika ada dua transformasi, pertama-tama diperhatikan bagian tengahnya.
- Bila bagian tengah sama-sama simetris, diperhatikan bagian di luar kuartil.

- Seperti contoh di awal, logaritma dan $-\frac{1}{x}$ sama-sama mempunyai bagian tengah simetris.
- Apabila diperhatikan bagian luar, logaritma tampak lebih baik (karena ekstrim atas dan bawah terletak simetris terhadap median), sehingga logaritma terpilih menjadi transformasi yang paling cocok.
- Apabila transformasi yang dipilih tidak cocok maka cara coba-coba ini akan sangat memakan waktu.

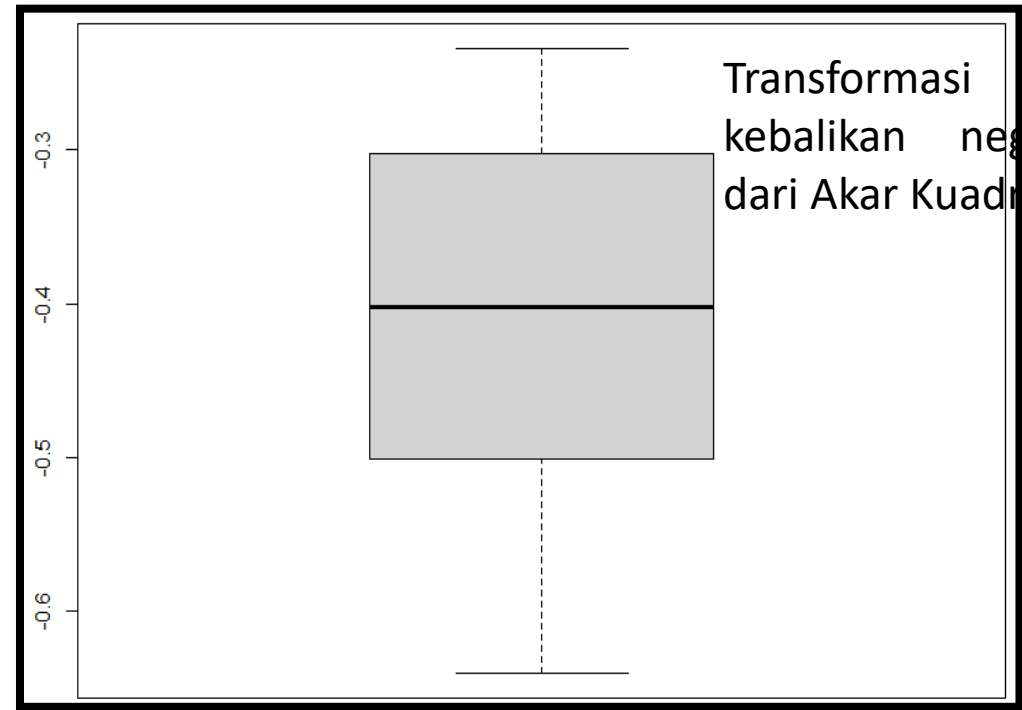
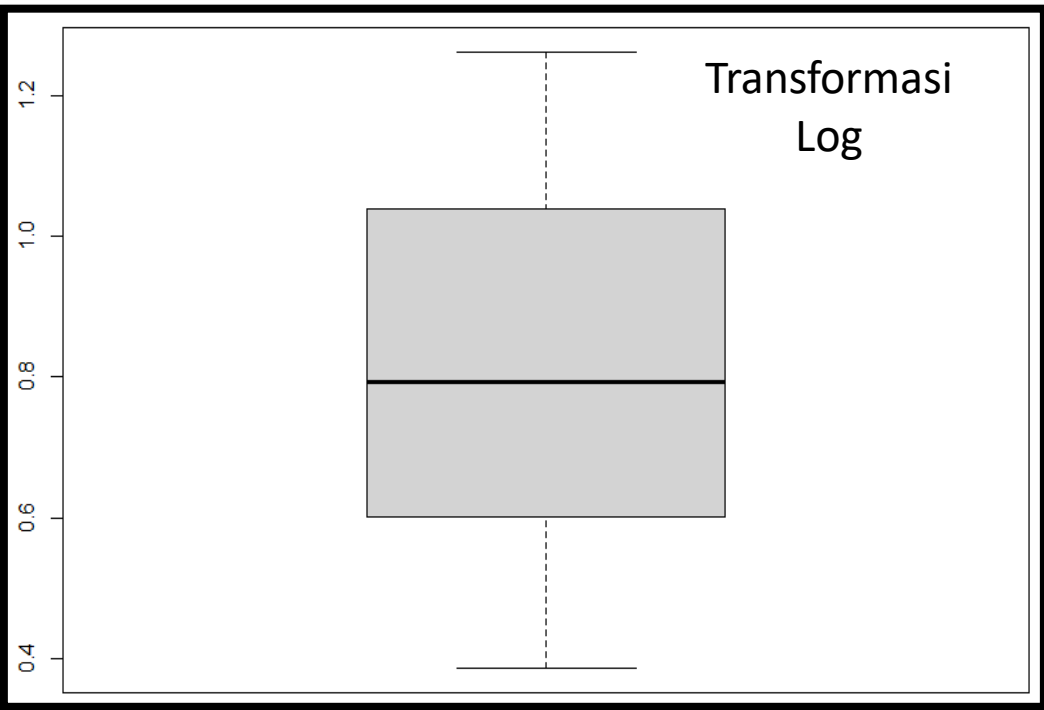
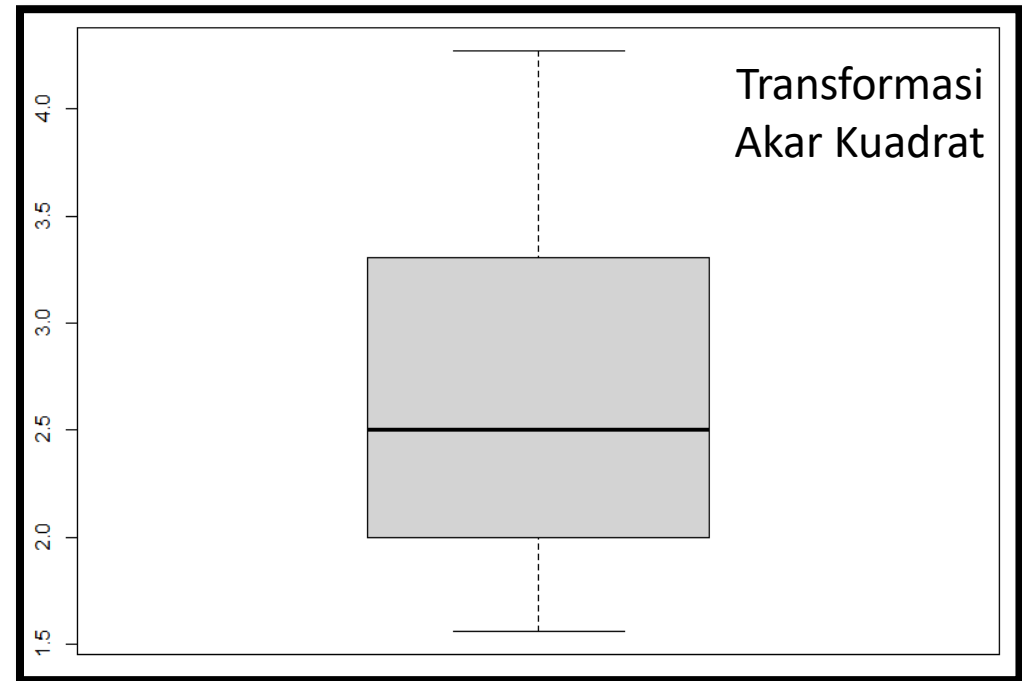
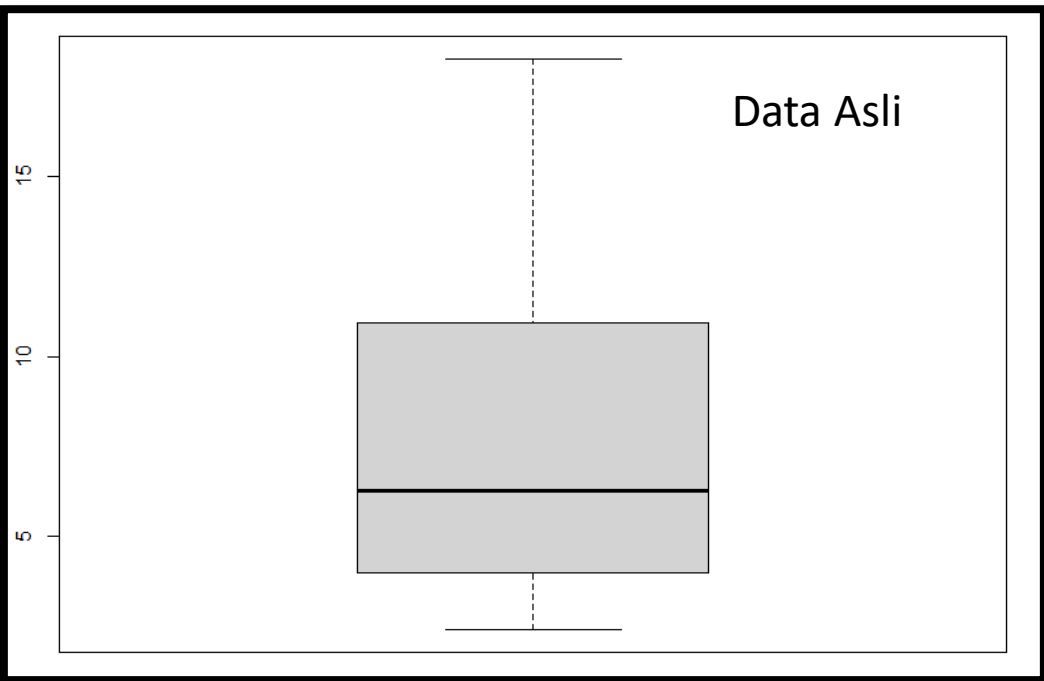
- Berdasarkan tangga transformasi tampak bahwa x^3 lebih kuat daripada x^2 .
- Semakin tinggi pangkat dari bilangan x yang lebih besar semakin disebarkan, sehingga jelas transformasi ini cocok untuk mengoreksi juraian ke bawah.
- Transformasi yang terletak disebelah kiri x pada tangga transformasi. Cocokan dengan suatu contoh, apakah betul transformasi tersebut dapat mengoreksi juraian ke atas.
- Hal ini dapat diatasi dengan memperhatikan diagram kotak dan titik yang hanya membutuhkan ringkasan lima angka untuk mengkonstruksinya.
- Dari diagram kotak dan titik, dapat dilihat apakah simetri atau tidak.

- Jadi, yang akan dikerjakan adalah transformasi untuk ringkasan lima angka, kemudian mengamati diagram kotak dan titiknya.
- Bila transformasi belum sesuai dicoba transformasi yang lain untuk ringkasan lima angka, demikian seterusnya sampai didapat transformasi yang unggul.
- Dengan demikian pekerjaan jauh lebih sedikit terutama untuk angkatan yang besar.

Contoh

Perhatikan kembali data pada contoh jumlah penduduk di Kanada Tahun 1860 – 1960. Hasil Transformasi Akar, Log dan negative kebalikan akar disajikan dalam Tabel di samping.

Kanada	Akar	Log	$-1/\sqrt{x}$
2,44	1,56	0,39	-0,6
3,23	1,8	0,51	-0,6
3,69	1,92	0,57	-0,5
4,32	2,08	0,64	-0,5
4,83	2,2	0,68	-0,5
5,37	2,32	0,73	-0,4
7,21	2,69	0,86	-0,4
8,79	2,96	0,94	-0,3
10,38	3,22	1,02	-0,3
11,51	3,39	1,06	-0,3
14,01	3,74	1,15	-0,3
18,24	4,27	1,26	-0,2



- Boxplot tersebut memperlihatkan diagram kotak dan titik untuk beberapa transformasi dari angkatan dalam Tabel Kanada.
- Terlihat bahwa angkatan asli menjurai ke atas, dengan transformasi Akar pangkat duanya (\sqrt{x}) lebih simetris daripada angkatan semula tapi masih menjurai ke atas.
- Transformasi Kebalikan negatif akar pangkat duanya ($-\frac{1}{\sqrt{x}}$) sedikit menjurai ke bawah.
- Transformasi Logaritma membuat angkatan menjadi simetris (lebih simetris dibandingkan \sqrt{x} dan $-\frac{1}{\sqrt{x}}$). Jadi transformasi \sqrt{x} mengoreksi sedang, $-\frac{1}{\sqrt{x}}$ mengoreksi berlebihan, sedang logaritma tepat mengoreksi jurai ke atas

Latihan

Diketahui satu Angkatan data dalam puluh ribuan sbb:

92, 50, 23, 2, 1, 2, 1, 3, 3, 10, 6, 5, 2

Tentukan transformasi yang terbaik untuk data tersebut!

Jelaskan!

Latihan

Kelompok Usia				
20-29	30-39	40-49	50-59	60-69
23,7	30,1	39,4	45,5	56,2
7,2	7,8	19,0	28,5	37,5
20,3	19,2	26,3	38,8	71,4
29,2	39,4	44,0	51,8	54,4
15,1	26,7	33,8	40,3	54,5
27,4	35,3	49,5	54,1	58,2
42,7	66,4	82,3	82,6	89,9

- Carilah Transformasi yang terbaik untuk kelompok Usia 20-29
- Carilah Transformasi yang terbaik untuk kelompok usia 30-39
- Carilah Transformasi yang terbaik untuk kelompok usia 40-49
- Carilah Transformasi yang terbaik untuk kelompok usia 50-59
- Carilah Transformasi yang terbaik untuk kelompok usia 60-69