

Tugas 1 Kapita Selekta Statistika

“Eksplorasi Data Longitudinal”



Disusun oleh :

Kelompok 3

1. Rosa Amalia Nursinta (11190940000041)
2. Elviana Saputri (11190940000043)
3. Meissy Astariva Putri (11190940000063)

FAKULTAS SAINS DAN TEKNOLOGI
UIN SYARIF HIDAYATULLAH JAKARTA
2022

A. Latar Belakang

Kadar gula merupakan adanya zat gula atau glukosa dalam darah tubuh manusia yang harus diperhatikan kestabilan kadarnya agar tetap di batas normal dengan cara tidak berlebihan dalam mengonsumsi makanan dan minuman yang mengandung gula seperti nasi yang dimana nasi merupakan makanan pokok pada sebagian besar negara di wilayah Asia. Sesuai dengan data yang diperoleh dari data WHO pada tahun 2017 mengenai fakta dan jumlah penderita diabetes di Indonesia, bahwa pada tahun 2012 sekitar 1 juta orang dewasa di wilayah Asia Tenggara meninggal disebabkan oleh kadar gula darah yang tinggi yang dapat menyebabkan komplikasi akut dan kronis, seperti gagal ginjal, penyakit jantung, dan pembuluh darah maupun tuberkulosis. Menurut WHO (2017) kematian tertinggi di Indonesia, disebabkan adanya komplikasi akibat diabetes.

B. Tujuan Penelitian

1. Mengetahui hubungan makanan pokok masyarakat suatu negara dengan tingkat kadar gula
2. Mengetahui hubungan pola pikir masyarakat negara maju dan berkembang terhadap tingkat kadar gula

C. Data

Data dalam penelitian ini merupakan data sekunder mengenai Kadar Gula Darah pada negara wilayah Asia, yang diperoleh dari website <https://www.gapminder.org/data/>. Terdapat 3 variabel yaitu, angka kadar gula darah, kategori negara maju atau berkembang dan makanan pokok dari masing-masing negara. Subjek dalam data yaitu 18 negara dengan 15 kali pengamatan dari 2004 sampai 2018. Dalam data ini, kami ingin mengetahui bagaimana hubungan makanan pokok masyarakat suatu negara dengan tingkat kadar gula dan hubungan pola pikir masyarakat negara maju dan berkembang terhadap tingkat kadar gula. Data yang kami gunakan termasuk data longitudinal karena mengamati bagaimana variabilitas antar subjek maupun dalam subjek dan variansi tiap individu cenderung tidak konstan maka variansi tersebut melanggar asumsi homoskedastisitas. Kami menggunakan metode longitudinal untuk mengolah data karena data ini terdapat pengamatan yang dilakukan sebanyak 15 kali pada 18 negara di Asia dan untuk menganalisis data dengan variansi yang tidak konstan serta respon yang memiliki korelasi pada setiap pengamatan.

Berikut data yang digunakan pada analisis ini <https://docs.google.com/spreadsheets/d/12gQ-ZvpW7r7iN3MHJghCq2Ho3FhjFFKNY4O3xCPMuJU/edit?usp=sharing>

D. Hasil dan Pembahasan

1. Manipulasi Data

- Transformasi struktur format data dari panjang menjadi lebar

```
> head(data.wide)
# A tibble: 6 x 18
  Negara      Kategori MakananPokok `2004` `2005` `2006` `2007` `2008` `2009` `2010` `2011`
  <chr>      <chr>      <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Afganistan Berkemb~ Gandum      15.5  18.6  19.7  20.8  23.4  23.3  24.2  24.8
2 Arab Saudi Berkemb~ Nasi       84.6  88.6  89.4  87.9  88.8  92.2  90.7  88.8
3 Bangladesh Berkemb~ Nasi       21.9  28.7  24.6  23.7  22.7  22.8  21.9  22.5
4 Filipina   Berkemb~ Nasi       75.9  70.1  67.5  68.2  70.6  62.7  63.3  65.7
5 Hong Kong  Maju      Telur       97.3  97.9  97.3  99    99.2  97.2  99.7  97.8
6 Indonesia  Berkemb~ Nasi       44.9  46.5  39.6  39.1  37.9  39    39.3  40.5
# ... with 7 more variables: 2012 <dbl>, 2013 <dbl>, 2014 <dbl>, 2015 <dbl>, 2016 <dbl>,
# 2017 <dbl>, 2018 <dbl>
```

- Transformasi struktur format data dari lebar menjadi panjang

```
> head(data)
# A tibble: 6 x 5
  Negara      Tahun KadarGula Kategori      MakananPokok
  <chr>      <dbl>    <dbl> <chr>      <chr>
1 Indonesia  2004      44.9 Berkembang Nasi
2 Indonesia  2005      46.5 Berkembang Nasi
3 Indonesia  2006      39.6 Berkembang Nasi
4 Indonesia  2007      39.1 Berkembang Nasi
5 Indonesia  2008      37.9 Berkembang Nasi
6 Indonesia  2009      39    Berkembang Nasi
```

- Dalam memanipulasi data longitudinal, kami menggunakan mean dan variansi karena variansi dan rata-rata sudah cukup mewakili bagaimana persebaran data. Ringkasan data yang menarik adalah varian karena pengamatan ini menghasilkan variansi yang sangat tidak konstan yang disebabkan oleh banyaknya subjek pengamatan dalam hal ini yaitu negara, khususnya negara yang menjadikan nasi sebagai makanan pokok.

2. Plot Profil Subjek

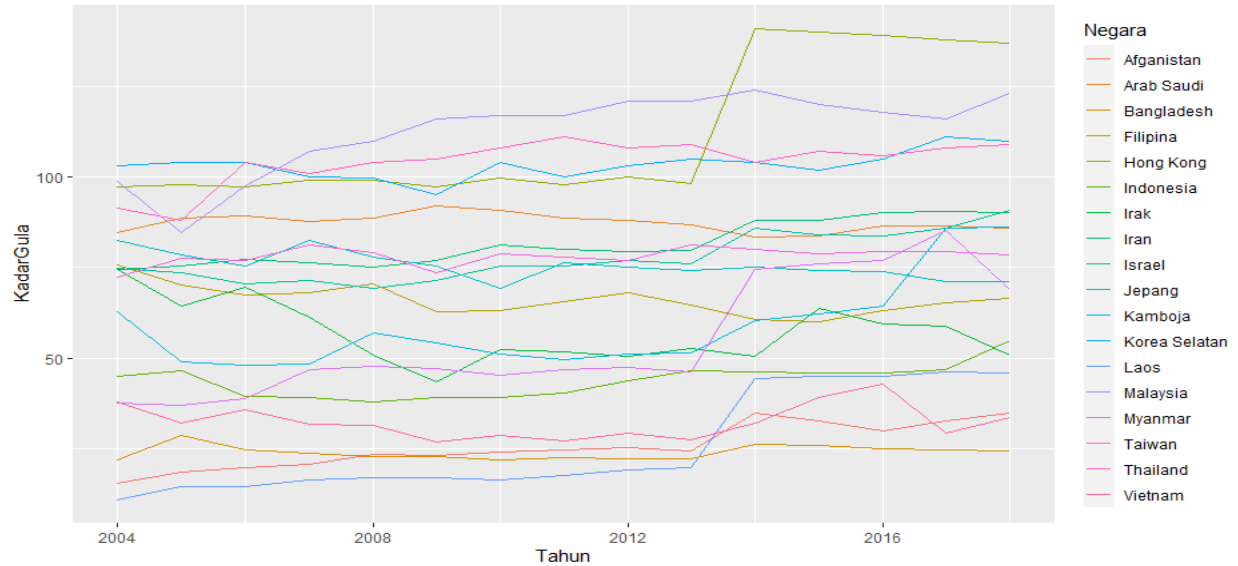
- Plot profil merupakan alat yang penting untuk mengeksplorasi data longitudinal untuk menunjukkan perubahan tingkat kadar gula pada setiap negara sepanjang tahun.
- Terlihat bahwa profil plot bergantung pada nilai kovariat yang berbeda karena berdasarkan grafik terlihat bahwa pada 12 negara dengan jenis makanan pokok nasi tidak memiliki tingkat kadar gula yang relatif sama dan memiliki perbedaan yang cukup jauh, maka secara subjektif ini menunjukkan bahwa tingkat kadar gula tidak bergantung pada jenis makanan pokok.

3. Korelasi

- Eksplorasi korelasi data longitudinal penting karena data time series respon yang didapat sangat berpotensi memiliki korelasi antar pengamatan.
- Scatter plot yang pertama yang menunjukkan hubungan antara respon untuk tahun 2004 dengan tahun 2005. Terlihat bahwa trend korelasi pada scatter plot tersebut naik yang berarti korelasinya besar atau korelasinya positif memiliki nilai 0.978, selanjutnya untuk tahun 2004 dengan tahun 2006 korelasinya sedikit menurun yaitu sebesar 0.978 dan seterusnya sampai tahun 2004 dengan tahun 2018 korelasinya sebesar 0.849 Hal ini menunjukkan bahwa respon pengamatan 2004 sampai respon pengamatan 2018 memiliki korelasi yang semakin menurun setiap tahunnya.
- Berdasarkan scatterplot terhadap data pengamatan dapat mengkonfirmasi hasil perhitungan matrik korelasi. Sebagai ilustrasi Scatter plot yang pertama yang menunjukkan hubungan antara respon untuk tahun 2004 dengan tahun 2005 terlihat bahwa trend korelasi pada scatter plot tersebut naik yang berarti korelasinya besar atau korelasinya positif memiliki nilai 0.978 sedangkan hasil perhitungan matrik korelasi sebesar 0.978 sehingga benar bahwa scatterplot dapat mengkonfirmasi hasil perhitungan matrik korelasi.

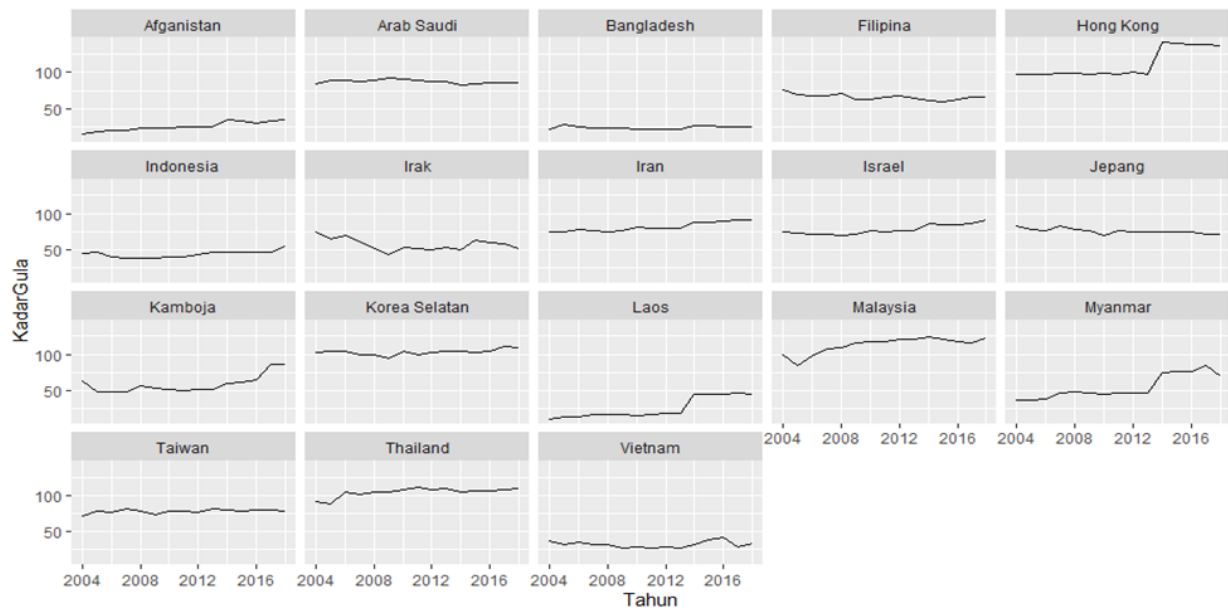
E. Visualisasi Data

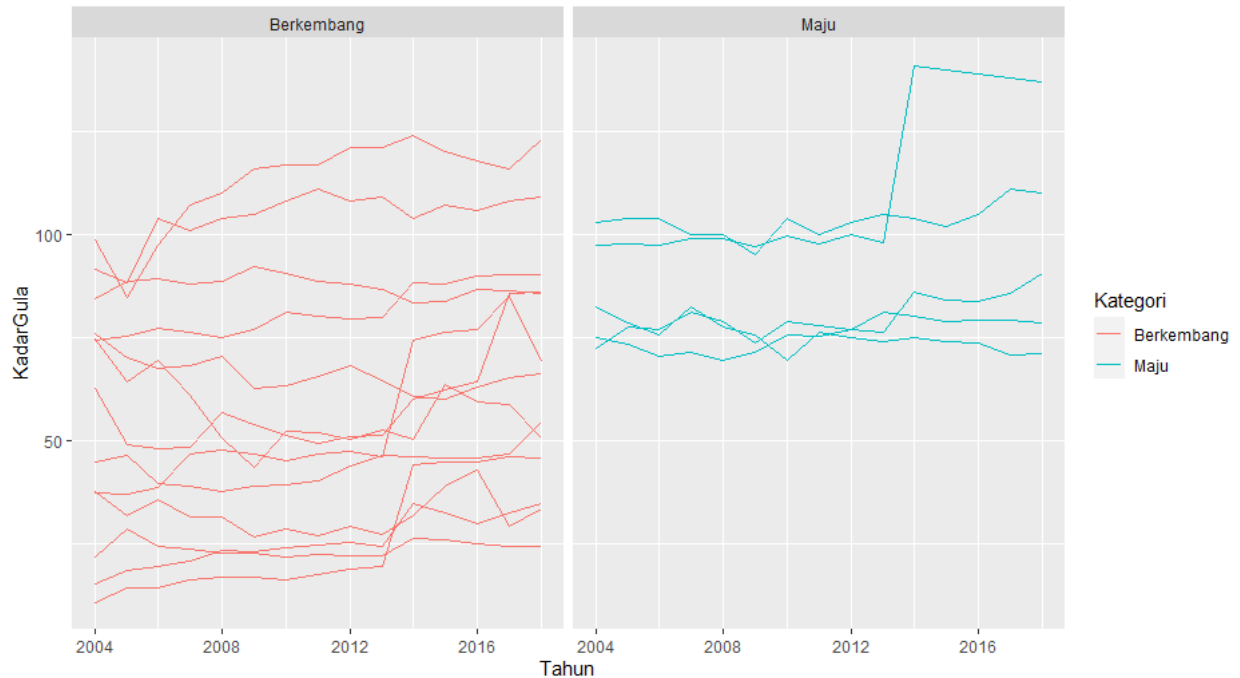
Berikut adalah grafik yang menggambarkan perubahan kadar gula darah secara keseluruhan pada 18 negara di wilayah Asia dalam periode 2004-2018



1) Profil Individu

Berikut merupakan grafik yang menggambarkan perubahan kadar gula darah pada tiap negara di wilayah Asia dalam periode 2004-2018.

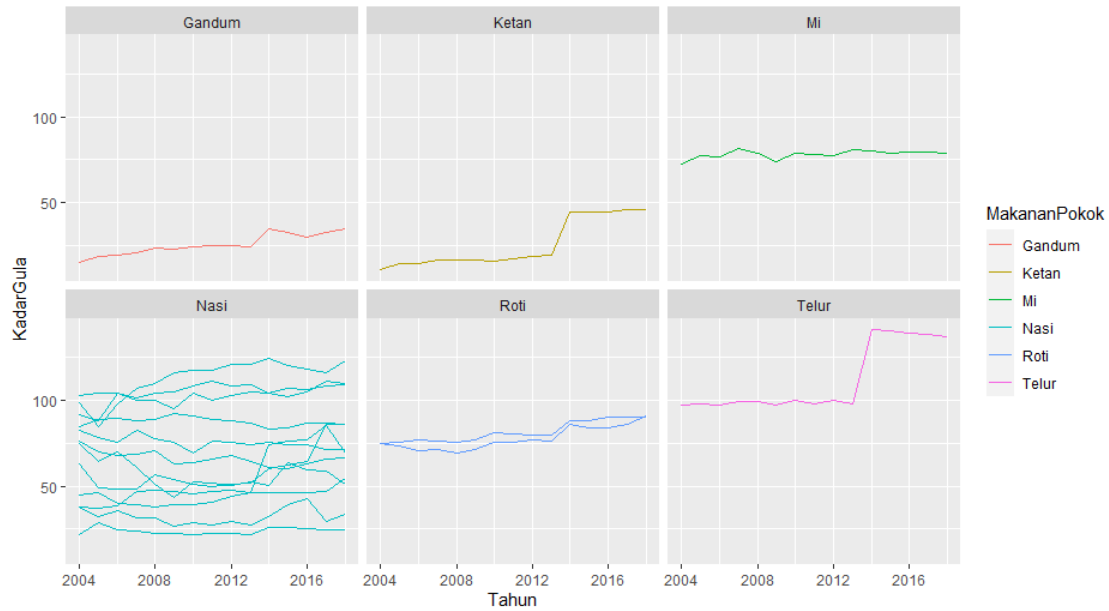




Interpretasi :

Berdasarkan visualisasi grafik plot, terlihat bahwa pada 18 negara Asia yang kami analisis, pada setiap tahunnya cenderung mengalami kenaikan, dan Hong Kong memiliki kadar gula darah tertinggi dibanding 17 negara lainnya dan mencapai nilai tertingginya pada tahun 2014 sebesar 140.0 g/hari/orang. Berdasarkan kategori negara, negara maju cenderung memiliki tingkat kadar gula yang tinggi, selain itu terdapat pula beberapa negara berkembang yang memiliki tingkat kadar gula tinggi.

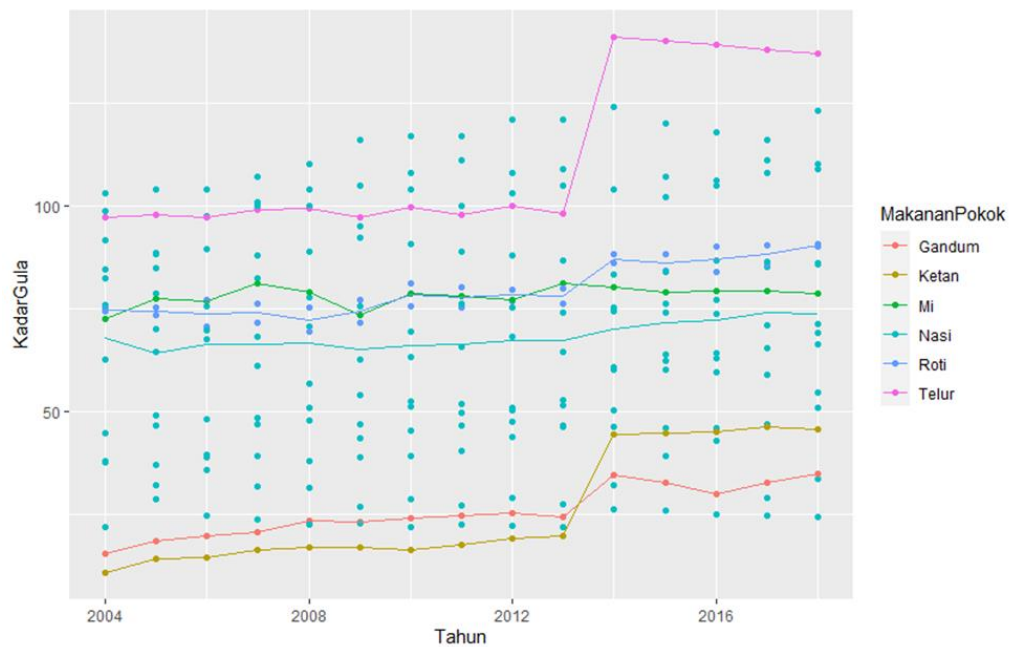
Kemudian untuk menganalisis lebih lanjut mengenai beberapa faktor yang menyebabkan tingginya kadar gula, kami menambahkan variabel makanan pokok tiap negara untuk mengetahui adanya hubungan dengan kadar gula darah.



Interpretasi :

Berdasarkan visualisasi grafik, terlihat bahwa pada 6 jenis makanan pokok yang kami analisis, terdapat beberapa negara yang menjadikan nasi sebagai bahan pokok memiliki kadar gula darah tinggi, selain itu Hong Kong yang menjadikan telur sebagai makanan pokok juga memiliki kadar gula darah tertinggi dibanding 17 negara lainnya. Sedangkan negara dengan makanan pokok gandum dan ketan cenderung memiliki kadar gula yang rendah.

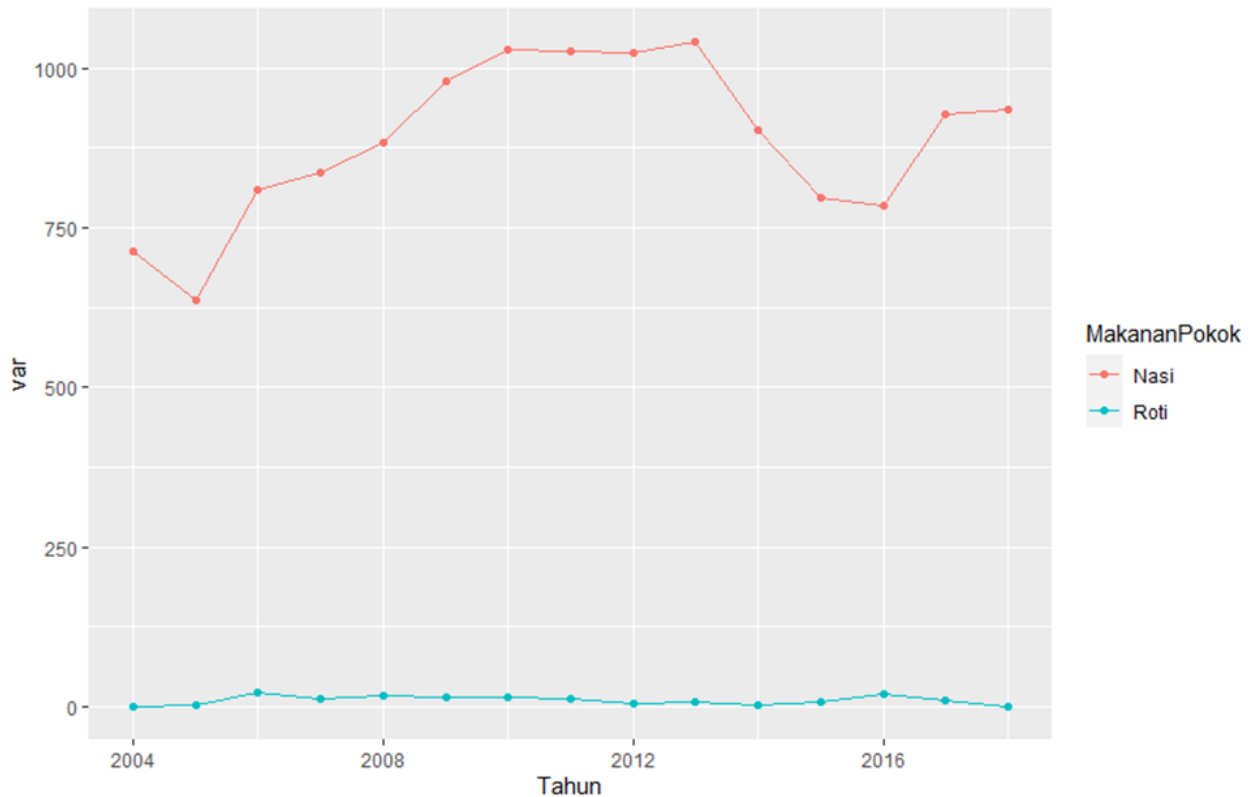
2) Struktur Mean



Interpretasi :

Perubahan rata-rata kadar gula cenderung linear untuk jenis makanan pokok nasi, mi, dan roti. Pada tahun 2004-2012 perubahan tingkat kadar gula meningkat perlahan sedangkan dari tahun 2012-2018 beberapa negara mengalami peningkatan yang cukup signifikan.

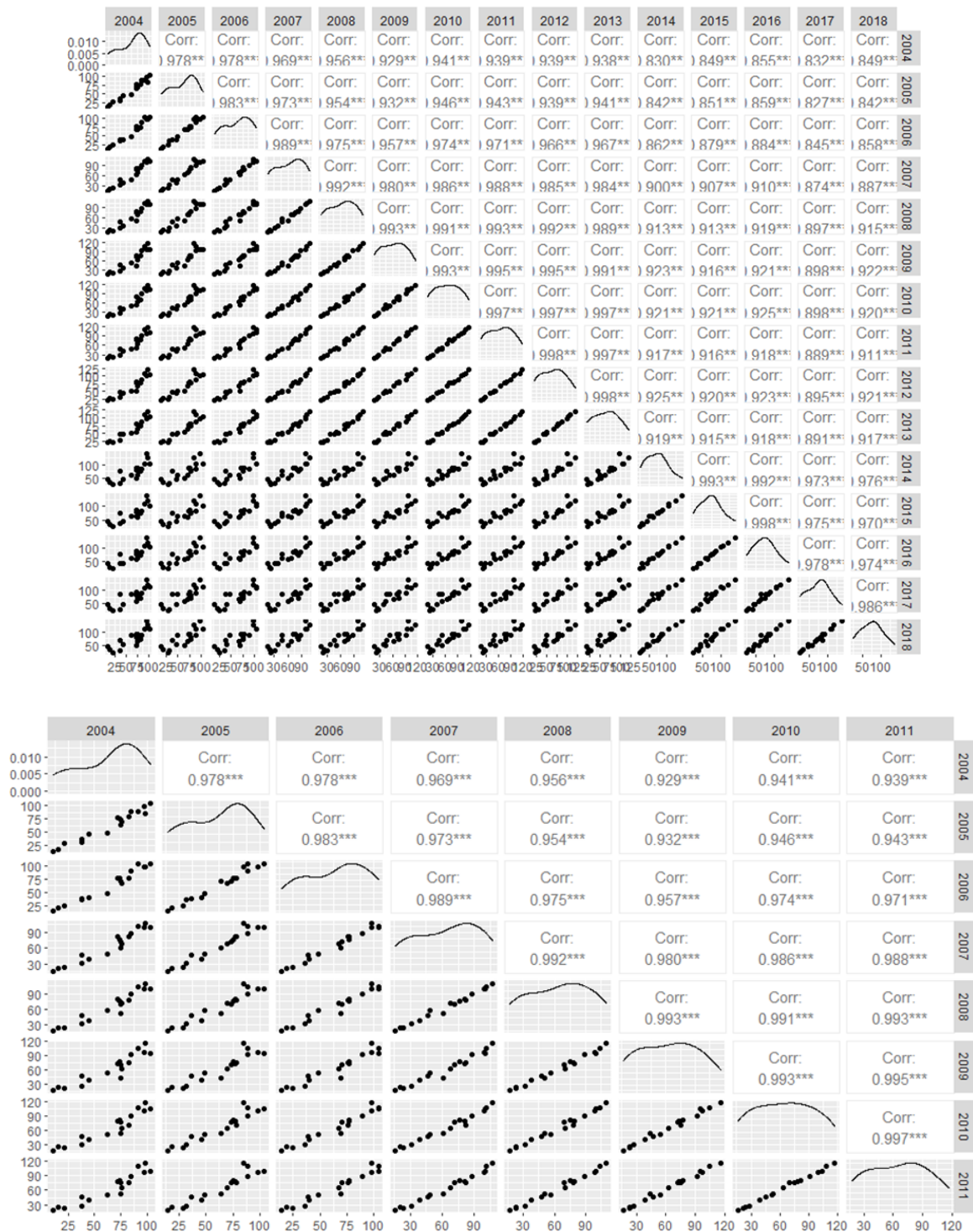
3) Struktur Variansi

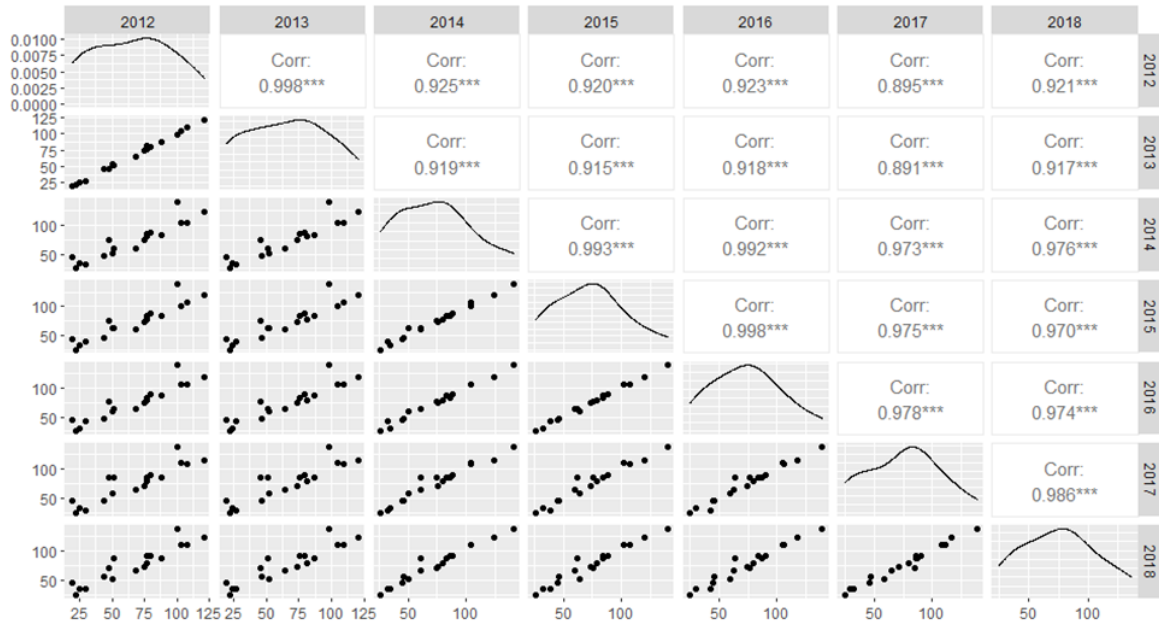


Interpretasi :

Variansi dari negara yang menjadikan nasi sebagai makanan pokok cenderung tidak konstan sehingga variansi dapat dikatakan mengalami heterokedastisitas, sedangkan variansi dari negara yang menjadikan roti sebagai makanan pokok cenderung konstan pada setiap tahunnya sehingga variansi terbebas dari asumsi heterokedastisitas.

4) Struktur Korelasi





```
> cor(data.wide[,4:18])
```

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 2004 | 1.0000000 | 0.9784048 | 0.9781119 | 0.9686301 | 0.9555395 | 0.9294654 | 0.9412374 | 0.9392083 | 0.9388393 |
| 2005 | 0.9784048 | 1.0000000 | 0.9832670 | 0.9726713 | 0.9543543 | 0.9315331 | 0.9460487 | 0.9426274 | 0.9394602 |
| 2006 | 0.9781119 | 0.9832670 | 1.0000000 | 0.9886444 | 0.9749004 | 0.9570370 | 0.9736490 | 0.9714068 | 0.9656460 |
| 2007 | 0.9686301 | 0.9726713 | 0.9886444 | 1.0000000 | 0.9918007 | 0.9803882 | 0.9856403 | 0.9883549 | 0.9849282 |
| 2008 | 0.9555395 | 0.9543543 | 0.9749004 | 0.9918007 | 1.0000000 | 0.9928013 | 0.9913300 | 0.9927818 | 0.9918825 |
| 2009 | 0.9294654 | 0.9315331 | 0.9570370 | 0.9803882 | 0.9928013 | 1.0000000 | 0.9929148 | 0.9951313 | 0.9948360 |
| 2010 | 0.9412374 | 0.9460487 | 0.9736490 | 0.9856403 | 0.9913300 | 0.9929148 | 1.0000000 | 0.9971617 | 0.9968344 |
| 2011 | 0.9392083 | 0.9426274 | 0.9714068 | 0.9883549 | 0.9927818 | 0.9951313 | 0.9971617 | 1.0000000 | 0.9981921 |
| 2012 | 0.9388393 | 0.9394602 | 0.9656460 | 0.9849282 | 0.9918825 | 0.9948360 | 0.9968344 | 0.9981921 | 1.0000000 |
| 2013 | 0.9381119 | 0.9405503 | 0.9668983 | 0.9842125 | 0.9888596 | 0.9913096 | 0.9966733 | 0.9971225 | 0.9982487 |
| 2014 | 0.8302023 | 0.8421983 | 0.8623072 | 0.9001931 | 0.9130197 | 0.9226324 | 0.9214205 | 0.9174131 | 0.9246493 |
| 2015 | 0.8491453 | 0.8511074 | 0.8794881 | 0.9065597 | 0.9125096 | 0.9160169 | 0.9208594 | 0.9155015 | 0.9199693 |
| 2016 | 0.8553296 | 0.8591121 | 0.8835572 | 0.9098705 | 0.9192965 | 0.9212051 | 0.9246202 | 0.9178289 | 0.9233103 |
| 2017 | 0.8323431 | 0.8271110 | 0.8452523 | 0.8741945 | 0.8971363 | 0.8981637 | 0.8977227 | 0.8890498 | 0.8949385 |
| 2018 | 0.8492533 | 0.8421187 | 0.8583939 | 0.8866440 | 0.9145600 | 0.9221027 | 0.9197998 | 0.9112505 | 0.9209643 |
| 2013 | | 2014 | 2015 | 2016 | 2017 | 2018 | | | |
| 2004 | 0.9381119 | 0.8302023 | 0.8491453 | 0.8553296 | 0.8323431 | 0.8492533 | | | |
| 2005 | 0.9405503 | 0.8421983 | 0.8511074 | 0.8591121 | 0.8271110 | 0.8421187 | | | |
| 2006 | 0.9668983 | 0.8623072 | 0.8794881 | 0.8835572 | 0.8452523 | 0.8583939 | | | |
| 2007 | 0.9842125 | 0.9001931 | 0.9065597 | 0.9098705 | 0.8741945 | 0.8866440 | | | |
| 2008 | 0.9888596 | 0.9130197 | 0.9125096 | 0.9192965 | 0.8971363 | 0.9145600 | | | |
| 2009 | 0.9913096 | 0.9226324 | 0.9160169 | 0.9212051 | 0.8981637 | 0.9221027 | | | |
| 2010 | 0.9966733 | 0.9214205 | 0.9208594 | 0.9246202 | 0.8977227 | 0.9197998 | | | |
| 2011 | 0.9971225 | 0.9174131 | 0.9155015 | 0.9178289 | 0.8890498 | 0.9112505 | | | |
| 2012 | 0.9982487 | 0.9246493 | 0.9199693 | 0.9233103 | 0.8949385 | 0.9209643 | | | |
| 2013 | 1.0000000 | 0.9190224 | 0.9153779 | 0.9176014 | 0.8912731 | 0.9174858 | | | |
| 2014 | 0.9190224 | 1.0000000 | 0.9928048 | 0.9918647 | 0.9728017 | 0.9763402 | | | |
| 2015 | 0.9153779 | 0.9928048 | 1.0000000 | 0.9975348 | 0.9746070 | 0.9695151 | | | |
| 2016 | 0.9176014 | 0.9918647 | 0.9975348 | 1.0000000 | 0.9778021 | 0.9736160 | | | |
| 2017 | 0.8912731 | 0.9728017 | 0.9746070 | 0.9778021 | 1.0000000 | 0.9857583 | | | |
| 2018 | 0.9174858 | 0.9763402 | 0.9695151 | 0.9736160 | 0.9857583 | 1.0000000 | | | |

Interpretasi :

Pada diagonal digambarkan bentuk distribusi masing-masing respon untuk setiap periodenya. Respon yang kita amati yaitu kadar gula pada tahun 2004 – 2018. Pada diagonal bawah terdapat scatter plot untuk melihat hubungan antara 2 variabel. Scatter plot yang pertama yang menunjukkan hubungan antara respon untuk tahun 2004 dengan tahun 2005. Terlihat bahwa trend korelasi pada scatter plot tersebut naik yang berarti korelasinya besar atau korelasinya positif memiliki nilai 0.978 dan seterusnya sampai dengan tahun 2018.

Berdasarkan scatterplot terhadap data pengamatan dapat mengkonfirmasi hasil perhitungan matrik korelasi. Sebagai ilustrasi Scatter plot yang pertama yang menunjukkan hubungan antara respon untuk tahun 2004 dengan tahun 2005 terlihat bahwa trend korelasi pada scatter plot tersebut naik yang berarti korelasinya besar atau korelasinya positif memiliki nilai 0.978 sedangkan hasil perhitungan matrik korelasi sebesar 0.978 sehingga benar bahwa scatterplot dapat mengkonfirmasi hasil perhitungan matrik korelasi.

1) Struktur Kovarian

```
> cov(data.wide[,4:18])
      2004      2005      2006      2007      2008      2009      2010      2011      2012
2004 858.6833 794.3330 847.9065 846.7397 838.4147 837.7776 875.5222 863.9707 862.8083
2005 794.3330 767.6002 805.9013 803.9131 791.7187 793.8618 832.0176 819.8383 816.3050
2006 847.9065 805.9013 875.1566 872.4860 863.5687 870.8647 914.3170 902.1207 895.9162
2007 846.7397 803.9131 872.4860 889.9199 885.9181 899.6065 933.3518 925.5693 921.4814
2008 838.4147 791.7187 863.5687 885.9181 896.5766 914.3976 942.2442 933.1858 931.4520
2009 837.7776 793.8618 870.8647 899.6065 914.3976 946.1459 969.4882 960.9041 959.7035
2010 875.5222 832.0176 914.3170 933.3518 942.2442 969.4882 1007.6344 993.6597
992.3869
2011 863.9707 819.8383 902.1207 925.5693 933.1858 960.9041 993.6597 985.4650 982.7460
2012 862.8083 816.3050 895.9162 921.4814 931.4520 959.7035 992.3869 982.7460 983.5877
2013 866.6189 821.4981 901.7387 925.5957 933.4377 961.2700 997.3816 986.7931 986.9663
2014 778.1371 746.3415 815.9444 858.9479 874.4391 907.7441 935.5458 921.1731 927.5542
2015 765.1800 725.1319 800.0884 831.6431 840.2263 866.4582 898.8970 883.7812 887.2480
2016 770.2721 731.4953 803.2889 834.1597 845.9476 870.8224 902.0052 885.4754 889.9148
2017 775.3116 728.4315 794.8527 828.9741 853.9049 878.1971 905.8392 887.1644 892.1895
2018 789.3147 740.0096 805.4266 838.9213 868.5651 899.6112 926.0646 907.3083 916.1062
      2013      2014      2015      2016      2017      2018
2004 866.6189 778.1371 765.1800 770.2721 775.3116 789.3147
2005 821.4981 746.3415 725.1319 731.4953 728.4315 740.0096
2006 901.7387 815.9444 800.0884 803.2889 794.8527 805.4266
2007 925.5957 858.9479 831.6431 834.1597 828.9741 838.9213
2008 933.4377 874.4391 840.2263 845.9476 853.9049 868.5651
2009 961.2700 907.7441 866.4582 870.8224 878.1971 899.6112
2010 997.3816 935.5458 898.8970 902.0052 905.8392 926.0646
2011 986.7931 921.1731 883.7812 885.4754 887.1644 907.3083
2012 986.9663 927.5542 887.2480 889.9148 892.1895 916.1062
2013 993.8344 926.6992 887.4064 889.0072 893.1517 917.3875
2014 926.6992 1023.0842 976.5280 974.9948 989.0937 990.4974
2015 887.4064 976.5280 945.6496 942.7300 952.6909 945.6190
2016 889.0072 974.9948 942.7300 944.4704 955.2181 949.0265
2017 893.1517 989.0937 952.6909 955.2181 1010.4500 993.8580
2018 917.3875 990.4974 945.6190 949.0265 993.8580 1005.9884
```

Interpretasi :

Melihat trend secara keseluruhan dapat mengeksplor struktur korelasi dan struktur kovarian, secara umum didalam subjek atau within subjek biasanya pengamatan berulang berkorelasi. Menghitung korelasi antara pengamatan berulang ke-1 dengan pengamatan berulang lainnya, kita melihat korelasinya untuk setiap periode waktu yang diamati.

Dalam data ini priode yang kita gunakan yaitu tahun dari 2004-2018, dengan cara mengubah datanya menjadi bentuk wide atau lebar menggunakan perintah spread, agar memudahkan menghitung kovariansi dan korelasinya. Kami menghitung mulai dari kolom keempat sampai ke-18 mulai dari tahun 2004 sampai 2018 dengan pengamatan 15 kali atau 15 tahun. Diagonalnya merupakan nilai kovarian sedangkan diluar diagonal merupakan nilai kovariansi pada variabel yang berbeda untuk periode pengamatan yang berbeda. Elemen dibawah diagonal dan diatas diagonal nilainya pasti sama karena simetris. Sebagai ilustrasi kovariansi untuk tahun 2004 dengan tahun 2005 sama saja dengan kovariansi tahun 2005 dengan 2004

F. Kesimpulan

Berdasarkan analisa kami dapat disimpulkan bahwa jenis makanan pokok dan kategori negara tidak memiliki hubungan yang kuat dengan tingkat gula darah.