**TUGAS 1 STATISTIKA**

**ANALISA DESKRIPTIF DATA SATU DIMENSI**

**DENGAN MENGGUNAKAN HISTOGRAM DAN *BOXPLOT***

**PADA DATA *FACTBOOK (GDP REAL GROWTH RATE)***



Disusun oleh :

|  |  |
| --- | --- |
| Brama Hartoyo | 103052400030 |
| Muhammad Farhan Mappaewa | 103052400026 |
| Zaky Muhammad Fauzi | 103052400064 |

**TELKOM UNIVERSITY**

**BANDUNG**

**2024**

# **BAB I**

**PENDAHULUAN**

## **Latar Belakang**

Data Telah menjadi salah satu aset terpenting dalam berbagai disiplin ilmu, termasuk statistik, ekonomi, dan ilmu sosial. Analisis data yang tepat dapat memberikan wawasan yang mendalam dan mendukung pengambilan keputusan yang berbasis bukti data. Era informasi saat ini, kemampuan untuk menganalisis dan menginterpretasikan data dengan tepat sangat diperlukan untuk mendukung pengambilan keputusan yang berbasis bukti.

*Gross Domestic Product* (*GDP*) atau Produk Domestik Bruto (PDB) adalah salah satu indikator kunci dalam mengukur kesehatan ekonomi suatu negara. Pengukutan tersebut dari jumlah produk berupa barang dan jasa, yang dihasilkan oleh unit-unit produksi di dalam batas wilayah suatu negara (domestik) selama satu tahun. Pertumbuhan *GDP* yang positif menunjukkan pertumbuhan ekonomi, sementara penurunan *GDP* dapat menjadi indikasi masalah ekonomi yang lebih dalam. Oleh karena itu, analisis data *GDP* tidak hanya penting bagi ekonom, tetapi juga bagi masyarakat luas yang ingin memahami kondisi ekonomi negara mereka.

Melalui makalah ini, kami akan menjelaskan langkah-langkah analisis deskriptif pada data satu dimensi serta cara menginterpretasikan hasil analisis pertumbuhan PDB riil yang diperoleh dari dua metode yaitu histogram dan *boxplot*. Dengan demikian, diharapkan pembaca dapat memahami pentingnya analisis visual dalam menggali informasi dari data yang ada, serta meningkatkan kemampuan mereka dalam menganalisis dan menyajikan data secara efektif.

## **Tujuan**

1. Peran data sebagai aset penting dalam berbagai disiplin ilmu, contohnya dalam ekonomi kebutuhan akan analisis data yang akurat dalam pengambilan keputusan berbasis bukti.
2. Sebagai indikator utama kesehatan ekonomi, termasuk pengaruh pertumbuhan dan penurunan *GDP* terhadap kondisi ekonomi suatu negara.
3. Menjelaskan langkah-langkah dalam analisis deskriptif, khususnya dalam konteks *GDP real growth rate*, agar pembaca dapat mengikuti proses analisis dengan jelas.
4. Menginterpretasikan data pertumbuhan *GDP* menggunakan metode visual seperti *boxplot* dan histogram, sehingga pembaca dapat melihat bagaimana analisis visual dapat menggali informasi dari data.
5. Meningkatkan keterampilan dalam menganalisis dan menyajikan data secara efektif, serta memahami pentingnya analisis visual dalam menarik kesimpulan yang tepat.

## **Manfaat**

1. Memahami menginterpretasikan metode visualisasi data dengan *boxplot* dan histogram;
2. Dapat mengambil keputusan dan kesimpulan melalui data yang telah diolah;
3. Dapat mengerti tentang kondisi ekonomi suatu negara melalui *GDP real growth rate.*

# **BAB II**

**PEMBAHASAN**

## **Pengolahan Data**

Dilakukan dengan menggunakan histogram untuk melihat distribusi data dan *boxplot* untuk mendeteksi *outlier*. Data *GDP real growth rate* dianalisis secara deskriptif untuk mengidentifikasi karakteristik pusat (*mean,* median*, modus*) dan dispersi (*range, IQR*, simpangan baku). Berikut adalah pembahasan histogram dan *boxplot* dengan data *GPD real growth rate* yang diperoleh dari *factbook:*

1. Histogram

Merupakan grafik batang yang menggambarkan distribusi frekuensi dari suatu data. Histogram berguna untuk mengetahui bentuk distribusi data, seperti simetris, menceng kanan, atau menceng kiri. Menentukan histogram perlu mengetahui :

* Jumlah data (n) adalah total jumlah observasi atau data yang digunakan dalam analisis, perhitungan dapat dengan manual, ataupun dengan excel dengan rumus “=Count”.
* Maksimum adalah nilai terbesar dari dataset. Mengidentifikasi nilai maksimum membantu melihat batas atas dari data dengan manual, ataupun dengan excel dengan rumus “=Max”.
* Minimum adalah nilai terkecil dari dataset. Mengidentifikasi nilai ini penting untuk melihat batas bawah dari dataset dengan manual, ataupun dengan excel dengan rumus “=Min”.
* *Range* menunjukkan sebaran atau distribusi data. *Range* yang tinggi mengindikasikan data tersebar luas, perhitungan dapat dengan manual yaitu selisih nilai terbesar dengan nilai terkecil, ataupun dengan excel dengan rumus “=Max - Min”.
* Jumlah kelas (K) merupakan jumlah interval atau kelompok yang digunakan saat membuat distribusi frekuensi data, dapat dihitung dengan rumus “K = 1 + 3,3 x log(n)” (aturan Sturgess).
* Lebar kelas (P) adalah rentang nilai yang mencakup setiap kelas dalam histogram. Panjang kelas menentukan seberapa lebar setiap *bin* dalam histogram, dapat dihitung dengan rumus “P = *range*/K”
* *Lower Control Limit* (LCL) adalah batas bawah yang digunakan untuk tujuan yang sama, namun untuk data yang berada di bawah LCL, dapat dihitung dengan manual yaitu penjumlahan nilai terkecil dengan lebar kelas, ataupun dengan excel dengan rumus “=Min + P” dengan 8 kelas (diawali dengan nilai minimum).
* *Upper Control Limit* (UCL) adalah batas atas yang digunakan untuk mengidentifikasi nilai ekstrem dalam dataset. Jika data berada di atas UCL, maka dianggap sebagai *outlier* atau indikator variasi yang tidak wajar, dapat dihitung dengan manual yaitu penjumlahan nilai terkecil dengan lebar kelas lalu pengurangan 1, ataupun dengan excel dengan rumus “=(Min + P) - 1” dengan 8 kelas (tanpa diawali nilai minimum).
* *Lower Confidence Bound* (LCB) dan *Upper Confidence Bound* (UCB) biasanya digunakan untuk interval kepercayaan, memberikan konteks seberapa baik rata-rata sampel merepresentasikan populasi. LCB dapat dihitung dengan manual yaitu selisih LCL dengan 0,5, ataupun dengan excel dengan rumus “=(LCL – 0,5)” dengan 8 kelas. UCB dapat dihitung dengan manual yaitu penjumlahan UCL dengan 0,5, ataupun dengan excel dengan rumus “=(UCL + 0,5)” dengan 8 kelas.
* Bin adalah interval kelas yang digunakan dalam histogram, menggunakan *excel* dapat dengan *addons* “*XLMiner Analysis Toolpak Histogram*”.
* Frekuensi menunjukkan jumlah data yang termasuk dalam setiap bin, jika melebihi kelas (8 kelas pada data ini) maka dapat dimasukkan pada data “*more*”, jika menggunakan *excel* dapat dengan rumus “=Frequency”, atau dengan menggunakan “*XLMiner Analysis Toolpak Histogram*”.

1. *Boxplot*

Merupakan grafik yang menggambarkan ringkasan statistik dari suatu data, seperti nilai minimum, kuartil pertama, median, kuartil ketiga, dan nilai maksimum. *Boxplot* juga menampilkan nilai-nilai *outlier* dari data. Menentukan *boxplot* perlu mengetahui :

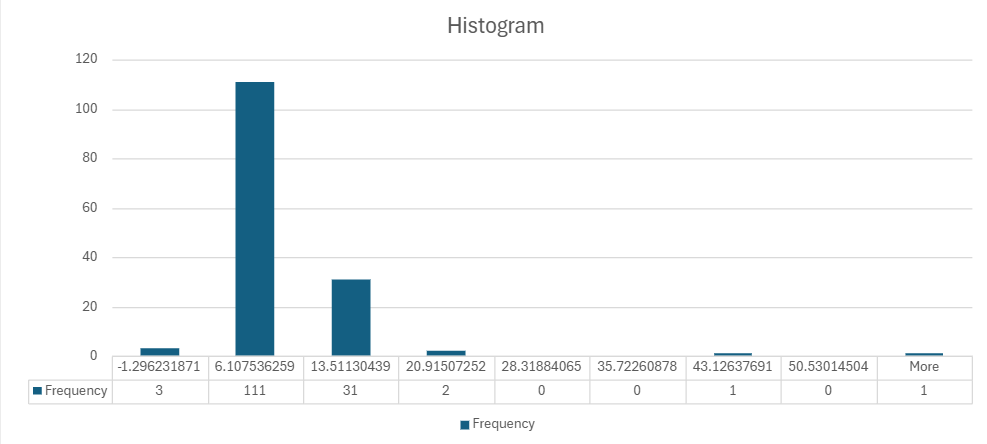
* Maksimal, minimal, dan *range.*
* Kuartil membagi data menjadi empat bagian yang sama. Q1 (kuartil pertama) adalah 25% data pertama, Q2 (median) adalah titik tengah data, dan Q3 (kuartil ketiga) adalah 75% dari data. Menghitung kuartil dapat dengan manual dengan rumus “Qi = (i x (n+1))/4” lalu “Kuartil i = (Qi bulat) + (Qi desimal) x (((Qi bulat) + 1) - (Qi bulat))”, jika dengan menggunakan *excel* “=Quartile.Exc”.
* Median adalah nilai tengah dari dataset yang diurutkan. Menggunakan median lebih baik daripada rata-rata jika data mengandung *outlier*. Menghitung median dapat dengan manual dengan rumus “(((urut ke-n)/2) + (urut ke-n+1)/2))/2” atau sama dengan kuartil 2, jika dengan menggunakan *excel* “=Median”.
* *Mean* adalah nilai rata-rata dari dataset. Ini menunjukkan ukuran pusat data secara umum. Menghitung mean dapat dengan manual dengan rumus “Σx/n”, jika dengan menggunakan *excel* “=Average”.
* Modus adalah nilai yang paling sering muncul dalam dataset. Data dengan modus lebih dari satu disebut multimodal. Menentukan modus dapat dengan manual dengan mencari frekuensi data terbanyak, jika dengan menggunakan *excel* “=Mode.Sngl”.
* *Range* antar kuartil (IQR) menunjukkan sebaran data di sekitar median, dan sering digunakan untuk mendeteksi *outlier*. Menghitung IQR dapat dengan manual dengan selisih kuartil 3 dengan kuartil 1, jika dengan menggunakan *excel* “=Q3 - Q1”.
* Pagar dalam dan luar digunakan untuk mendeteksi outlier. Data di luar pagar dalam dianggap outlier ringan, sedangkan data di luar pagar luar dianggap outlier ekstrim. Pagar dalam dan luar terdapat 2 bagian yaitu atas dan bawah. Menghitung PDA dan PDB dapat dengan rumus “PDA = Q3 + 1,5 x IQR” dan “PDB = Q1 - 1,5 x IQR”. Menghitung PLA dan PLB dapat dengan rumus “PLA = Q3 + 3 x IQR” dan “PLB = Q1 - 3 x IQR”.
* Simpangan baku mengukur seberapa jauh data menyebar dari rata-rata. Semakin besar simpangan baku, semakin tersebar datanya. Menghitung SB dapat dengan manual dengan rumus “S = ✓(Σ((x - mean(x))^2) / (n-1))”, jika dengan menggunakan *excel* “=STDV.S”.
* Varians adalah kuadrat dari simpangan baku, yang menunjukkan seberapa jauh penyebaran data dari rata-rata. Menghitung V dapat dengan manual dengan rumus “S^2 = (Σ((x - mean(x))^2) / (n-1))”, jika dengan menggunakan *excel* “=VAR.S”.

## **Hasil Analisis dan Visualisasi Data**

Berikut hasil analisis histogram dengan menggunakan perangkat lunak statistik (*Microsoft Excel*) untuk menghitung rumus GDP *real growth rate* negara-negara di dunia:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LCL** | **UCL** | **LCB** | **UCB** | **Bin** | **Frequency** |
| -8.2 | -1.796231871 | -8.7 | -1.296231871 | -1.296231871 | 3 |
| -0.796231871 | 5.607536259 | -1.296231871 | 6.107536259 | 6.107536259 | 111 |
| 6.607536259 | 13.01130439 | 6.107536259 | 13.51130439 | 13.51130439 | 31 |
| 14.01130439 | 20.41507252 | 13.51130439 | 20.91507252 | 20.91507252 | 2 |
| 21.41507252 | 27.81884065 | 20.91507252 | 28.31884065 | 28.31884065 | 0 |
| 28.81884065 | 35.22260878 | 28.31884065 | 35.72260878 | 35.72260878 | 0 |
| 36.22260878 | 42.62637691 | 35.72260878 | 43.12637691 | 43.12637691 | 1 |
| 43.62637691 | 50.03014504 | 43.12637691 | 50.53014504 | 50.53014504 | 0 |
|  |  |  |  | More | 1 |

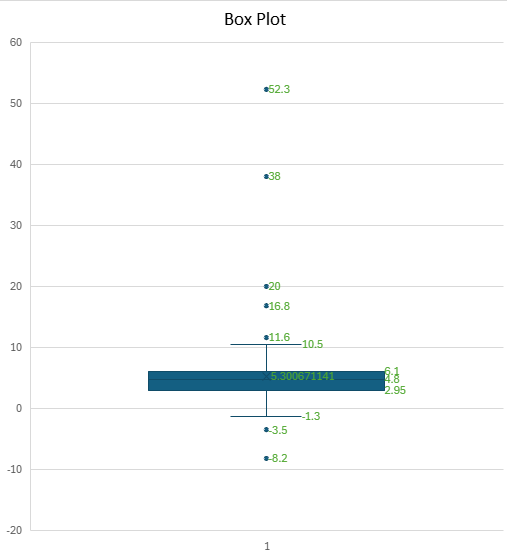
Visualisasi disajikan dalam bentuk tabel distribusi frekuensi, histogram untuk memudahkan pemahaman.



Berikut hasil analisis *boxplot* dengan menggunakan perangkat lunak statistik (*Microsoft Excel*) untuk menghitung rumus GDP *real growth rate* negara-negara di dunia:

|  |  |  |
| --- | --- | --- |
| **Fungsi Rumus** | **Keterangan** | **Nilai** |
| "=Minimal" | Min | -8.2 |
| "=Maximal" | Max | 52.3 |
| "=Maximal - Minimal" | Range | 60.5 |
| "=Quartile.Exc(Q1)" | Q1 | 2.95 |
| "=Quartile.Exc(Q2)" | Q2 | 4.8 |
| "=Quartile.Exc(Q3)" | Q3 | 6.1 |
| "=Median" | median | 4.8 |
| "=Count" | n data | 149 |
| "=1+3.3\*log(n data)" | K | 8.171514686 |
| "=Average" | Rata Rata | 5.300671141 |
| "=Range/K" | P | 7.403768129 |
| "=Mode.Sngl" | Modus | 6 |
| "Q3-Q1" | Range Antar Kuartil | 3.15 |
| "Q3+1.5\*(RAK)" | Pagar Dalam Atas | 10.825 |
| "Q1-1.5\*(RAK)" | Pagar Dalam Bawah | -1.775 |
| "Q3+3\*(RAK)" | Pagar Luar Atas | 15.55 |
| "Q1-3\*(RAK)" | Pagar Luar Bawah | -6.5 |
| "=STDV.S" | Simpangan Baku | 5.705420569 |
| "=VAR.S" | Varians | 32.55182387 |

Visualisasi data disajikan dalam bentuk *Boxp & Whiskers plot* rotasi vertikal.



# **BAB III**

**PENUTUP**

## **Kesimpulan**

Analisis deskriptif data satu dimensi dengan menggunakan histogram dan *boxplot* pada data *GDP real growth rate* memberikan pemahaman yang komprehensif mengenai karakteristik data. Berbagai ukuran statistik deskriptif, seperti minimum, maksimum, rentang, kuartil, median, modus, rentang antar kuartil, pagar dalam, pagar luar, simpangan baku, dan varians, dapat digunakan untuk mendeskripsikan data secara detail. Visualisasi data menggunakan histogram dan *boxplot* juga membantu dalam mengidentifikasi pola distribusi, nilai *outlier*, serta memahami keseluruhan karakteristik data *GDP real growth rate*.

Histogram menunjukkan frekuensi data pada beberapa rentang (bin) tertentu dengan rincian frekuensinya: Bin 1 (-1.29 hingga 6.11) frekuensi tertinggi dengan 111 data. Ini menunjukkan bahwa sebagian besar data berada pada rentang ini. Bin 2 (6.11 hingga 13.51) frekuensi sebesar 31 data, jauh lebih rendah dibandingkan dengan bin pertama. Bin 3 (13.51 hingga 20.91) frekuensi sebesar 2 data. Bin 4 hingga Bin 6 (20.91 hingga 35.72) tidak ada data yang masuk dalam rentang ini (frekuensi 0). Bin 7 (43.13 hingga 50.53) hanya ada 1 data pada rentang ini. Bin terakhir (lebih dari 50.53) ada 1 data, merupakan nilai *outlier* atau nilai ekstrem.

Bin pertama (6.11 hingga 13.51) memiliki tinggi frekuensi yang jauh lebih dominan dibandingkan dengan bin lainnya. Perbedaan tinggi yang signifikan antara bin pertama dan bin-bin berikutnya menunjukkan bahwa distribusi data sangat terkonsentrasi pada kisaran nilai tersebut.

*Boxplot* memberikan visualisasi statistik deskriptif dari dataset, yaitu: nilai terendah pada data yang bukan *outlier* = -8.2. Kuartil Pertama (Q1) = 2.95. Median (Q2) garis tengah kotak, yang menunjukkan nilai tengah dataset = 5.3. Kuartil Ketiga (Q3) garis batas atas kotak = 6.1. Nilai tertinggi pada data yang bukan outlier = 11.6. IQR mengukur sebaran tengah data dan sering digunakan untuk mendeteksi outlier = 3,15. *Outlier* adalah data yang berada jauh di luar batas IQR, dapat mengindikasikan data ekstrem atau anomali dalam dataset. *Outlier* bawah = -8.2. *Outlier* atas = 20, 38, dan 52.3. Pagar dalam (*inner fences*) digunakan untuk mendeteksi outlier moderat, dengan PDA = 10.825 dan PDB = -1.775. Pagar luar (*outer fences*) digunakan untuk mendeteksi outlier ekstrem, dengan PLA = 15.55 dan PLB = -6.5. *Range* adalah selisih antara nilai maksimum dan minimum, yaitu 11.6 − (−8.2) = 19.8.

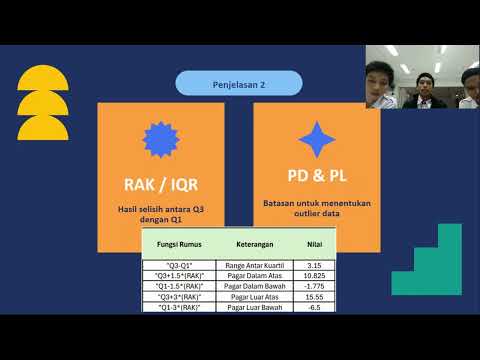
Hasil analisis deskriptif ini dapat memberikan wawasan yang berharga bagi pengambil keputusan dan peneliti dalam memahami kondisi ekonomi suatu negara berdasarkan pertumbuhan ekonomi riil (*GDP real growth rate*). Pemahaman yang mendalam terhadap data dapat membantu dalam membuat kebijakan yang tepat dan mengambil langkah-langkah strategis untuk meningkatkan pertumbuhan ekonomi.

# **DAFTAR PUSTAKA**

Banton, Caroline. *Real Economic Growth Rate: Definition, Calculation, and Uses. 2021*

# **VIDEO PRESENTASI YOUTUBE**

<https://youtu.be/Ny8hmxj2kCQ>

[](https://youtu.be/Ny8hmxj2kCQ)