**MATA KULIAH PROBABILISTIK STATISTIK**



Disusun Oleh:

Muhammad Firdaus 2109106052

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK UNIVERSITAS MULAWARMAN**

**2022**

# **KATA PENGANTAR**

# 

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nyalah saya dapat menyelesaikan laporan ini. Laporan ini disusun untuk memenuhi salah satu tugas mata kuliah Probabilistik Statistikpada Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman. Saya berharap semoga laporan ini dapat bermanfaat bagi para pembacanya. Saya menyadari penyusunan laporan ini tidak akan selesai tanpa bantuan dari berbagai pihak. Karena itu, pada kesempatan ini kami ingin mengucapkan terima kasih kepada Dr. Fahrul Agus, MT selaku Dosen Probabilistik Statistik.

Meskipun telah berusaha menyelesaikan laporan ini sebaik mungkin, saya sebagai penulis menyadari bahwa di dalam laporan ini masih terdapat kekurangan. Oleh karena itu, saya mengharapkan kritik dan saran yang membangun dari para pembaca guna menyempurnakan segala kekurangan dalam penyusunan laporan ini.

Soal :

**Tugas II**

**Pentunjuk Soal:**

1.  Kerjakan secara individu, dilarang *copy-paste coding*, jika melanggar dikenakan sangsi ‘*plagiator*’

2.      Download dataset berikut: StudentsPerformance.csv (Terlampir)

3.      Tentukan jenis/type data dari setiap variable yang ada di data set tersebut

4.      Lakukan analisis berikut dengan pemrograman Python/R (*laporkan scriptnya*)

a.Buat Tabel Distribusi Frekuensi dan Histogram Variabe *reading\_score*dan*writing\_score* dengan jumlah dan lebar interval sebesar 10.

b. Tentukan nilai Tendensi Central: Rata-rata, Median, Modus, Min dan Max data kedua variable tersebut.

c. Tentukan nilai Variasi: Jangkauan Range, Kisaran Interkuartil, Varians, Standar Deviasi data *reading\_score*dan*writing\_score.*

d.  Buat Grafik data Variabel *reading\_score*dan*writing\_score*: Histogram, BOX PLOT, dan Tentukan Nilai Skewness dan Kurtosis. Buat interpretasi saudara tentang perbedaan distribusi kedua variable tersebut, terutama pada gambar BOX PLOT dan nilai Skewness - Kurtosisnya.

e.Buat scater plot antara *writing\_score* (dependent variable-Y) versus *reading\_score* (independent variable-X). Tulis interpretasi saudara apakah ada indikasi hubungan liear antar keduanya.

Buat laporan saudara tentang analisis data terhadap persoalan di atas pada dokumen Word, dan kumpulkan (upload) pada aplikasi mols.unmul.ac.id sesuai waktu yang ditentukan (Jumat , 07-10-2022 pkl 07.30).

Jawaban:

|  |
| --- |
| from matplotlib import pyplot as plt  import pandas as pd  import numpy as np  import seaborn as sns  from prettytable import PrettyTable |

|  |
| --- |
| df = pd.read\_csv("StudentsPerformance.csv")  df.head(20) |

|  |
| --- |
|  |

|  |
| --- |
| df.info() |

|  |
| --- |
|  |

|  |
| --- |
| Buat Tabel Distribusi Frekuensi dan Histogram Variabe reading\_score dan writing\_score dengan jumlah dan lebar interval sebesar 10.[¶](http://localhost:8888/lab/tree/muhammad_052/Tugas_Probas/2109106052_Muhammad_Firdaus.ipynb#Buat-Tabel-Distribusi-Frekuensi-dan-Histogram-Variabe-reading_score-dan-writing_score-dengan-jumlah-dan-lebar-interval-sebesar-10.) |

|  |
| --- |
| label = np.vectorize(lambda x: f"{x}-{x + 10}")(np.arange(0, 100, 10)) |

|  |
| --- |
| interval = np.arange(0, 110, 10)  print("==||reading score||==")  reading\_total = df.groupby(pd.cut(df["reading\_score"], interval, labels=label))["reading\_score"].count().to\_list()  reading\_df = pd.DataFrame({ 'Interval': label, 'Total': reading\_total })  print(reading\_df) |

|  |
| --- |
|  |
| fig, (plt\_reading)=plt.subplots(figsize=(20,8))  plt\_reading.bar(label,reading\_total, width=1.0, color="crimson", edgecolor='black',linewidth=2)  plt\_reading.grid()  plt\_reading.bar\_label(reading\_bar)  plt\_reading.set\_xlabel('Interval', fontsize=12)  plt\_reading.set\_ylabel('Total', fontsize=12)  plt\_reading.set\_title('Histogram Reading Score')  plt\_reading.set\_facecolor("paleturquoise")  plt.show() |

|  |
| --- |
|  |

|  |
| --- |
| interval = np.arange(0, 110, 10)  print("==||writing score score||==")  writing\_total = df.groupby(pd.cut(df["writing\_score"], interval, labels=label))["writing\_score"].count().to\_list()  writing\_df = pd.DataFrame({ 'Interval': label, 'Total': writing\_total })  print(writing\_df) |

|  |
| --- |
|  |

|  |
| --- |
| fig, (plt\_writing)=plt.subplots(figsize=(20,8))  plt\_writing.bar(label,writing\_total, width=1.0,color="darkslateblue",edgecolor='black',linewidth=2)  plt\_writing.grid()  plt\_writing.bar\_label(writing\_score\_bar)  plt\_writing.set\_title('Histogram Writing Score')  plt\_writing.set\_xlabel('Interval', fontsize=12)  plt\_writing.set\_ylabel('Total', fontsize=12)  plt\_writing.set\_facecolor("paleturquoise")  plt.show() |

|  |
| --- |
|  |

|  |
| --- |
| Tentukan nilai Tendensi Central: Rata-rata, Median, Modus, Min dan Max data kedua variable tersebut. |

|  |
| --- |
| mean = df["reading\_score"].mean(axis = 0)  median = df["reading\_score"].median()  modus = df["reading\_score"].mode()[0]  min = df["reading\_score"].min()  max = df["reading\_score"].max()  meanz = df["writing\_score"].mean(axis = 0)  medianz = df["writing\_score"].median()  modusz = df["writing\_score"].mode()[0]  minz = df["writing\_score"].min()  maxz = df["writing\_score"].max() |

|  |
| --- |
| tableku.add\_row(["Rata-rata",mean])  tableku.add\_row(["Median",median])  tableku.add\_row(["Modus",modus])  tableku.add\_row(["Min",min])  tableku.add\_row(["Max",max])  print("==||reading score||==")  print(tableku) |

|  |
| --- |
|  |

|  |
| --- |
| tablemu.add\_row(["Rata-rata",meanz])  tablemu.add\_row(["Median",medianz])  tablemu.add\_row(["Modus",modusz])  tablemu.add\_row(["Min",minz])  tablemu.add\_row(["Max",maxz])  print("==||writing score||==")  print(tablemu) |

|  |
| --- |
|  |

|  |
| --- |
| Tentukan nilai Variasi: Jangkauan Range, Kisaran Interkuartil, Varians, Standar Deviasi data reading\_score dan writing\_score. |

|  |
| --- |
| tables = PrettyTable(["Nilai", "Hasil"])  tables.align = "l"  tablex = PrettyTable(["Nilai", "Hasil"])  tablex.align = "l" |

|  |
| --- |
| jarak = maxim - minim  Q1 = (np.percentile(df["reading\_score"], 75) + np.percentile(df["reading\_score"], 25)) / 2  varians = df["reading\_score"].var()  standards = df["reading\_score"].std()  Q2 = (np.percentile(df["writing\_score"], 75) + np.percentile(df["writing\_score"], 25)) / 2  varian = df["writing\_score"].var()  standard = df["writing\_score"].std() |

|  |
| --- |
| tables.add\_row(["Jangkaun Range", jarak])  tables.add\_row(["Kisaran Kuartil", Q1])  tables.add\_row(["Varians", varians ])  tables.add\_row(["Standard Deviasi", standards])  print("==||reading score||==")  print(tables) |

|  |
| --- |
|  |

|  |
| --- |
| tablex.add\_row(["Jangkaun Range", jarak])  tablex.add\_row(["Kisaran Kuartil", Q2])  tablex.add\_row(["Varians", varian ])  tablex.add\_row(["Standard Deviasi", standard])  print("==||writing score||==")  print(tablex) |

|  |
| --- |
|  |

|  |
| --- |
| Buat Grafik data Variabel reading\_score dan writing\_score: Histogram, BOX PLOT, dan Tentukan Nilai Skewness dan Kurtosis. Buat interpretasi saudara tentang perbedaan distribusi kedua variable tersebut, terutama pada gambar BOX PLOT dan nilai Skewness - Kurtosisnya |

|  |
| --- |
| tableD = PrettyTable(["nilai","Reading\_score","Writing\_score"]) |

|  |
| --- |
| skewness1 = df['reading\_score'].skew()  kurtosis1 = df['reading\_score'].kurt()  skewness2 = df['writing\_score'].skew()  kurtosis2 = df['writing\_score'].kurt() |

|  |
| --- |
| tableD.add\_row(["skewness", skewness1, skewness2])  tableD.add\_row(["kurtosis", kurtosis1, kurtosis2])  print(tableD) |

|  |
| --- |
|  |

|  |
| --- |
| fig, (hisx,hisz)=plt.subplots(1,2,figsize=(10,8))  hisx.hist(df['reading\_score'], color="crimson", edgecolor='black',linewidth=2)  hisx.set\_xlabel('Interval', fontsize=12)  hisx.set\_ylabel('total', fontsize=12)  hisx.set\_title('Reading Score')  hisx.set\_facecolor("paleturquoise")  hisx.grid()  hisz.hist(df['writing\_score'], color="darkslateblue",edgecolor='black',linewidth=2)  hisz.set\_xlabel('Interval', fontsize=12)  hisz.set\_ylabel('total', fontsize=12)  hisz.set\_title('Writing Score')  hisz.set\_facecolor("paleturquoise")  hisz.grid()  plt.show() |

|  |
| --- |
|  |

|  |
| --- |
| fig, (boxplot) = plt.subplots(figsize=(10,10))  sns.boxplot([df['reading\_score'], df['writing\_score']], color="darkslateblue")  plt.ylabel('total', fontsize=12)  plt.xticks([0,1],['Boxplot Reading Score','Boxplot Writing Score'])  plt.grid(color='red')  plt.show() |

|  |
| --- |
|  |

|  |
| --- |
| Buat scater plot antara writing\_score (dependent variable-Y) versus reading\_score (independent variable-X). Tulis interpretasi saudara apakah ada indikasi hubungan liear antar keduanya. |

|  |
| --- |
| plt.figure(figsize=(12,10))  plt.scatter(df["reading\_score"], df["writing\_score"],color="crimson")  plt.xlabel('Reading Score')  plt.ylabel('Writing Score')  plt.title('Scatter Plot')  plt.show() |

|  |
| --- |
|  |

# **PENUTUP**

## **Kesimpulan**

Untuk pengerjaan Tugas Probabilitas Statistik ini saya menggunakan jupyter-lab sebagai alat untuk pengamplikasiannya dan visualisasinya. Library yang dibutuhkan adalah pandas, numpy, matploylib, dan seaborn.

## Saran

Saya menyadari ini masih jauh dari kata sempurna. Maka dari itu saya mohon maaf apabila terdapat kesalahan kata-kata yang kurang berkenan dan saya mengharapkan kritik dan saran dari pembaca yang membangun demi perbaikan di masa depan. Saya berharap semoga laporan ini dapat dipahami bagi siapapun yang membacanya, dan dapat berguna bagi para pembaca dan pihak-pihak lain yang berkepentingan.