

**MENGANALISIS DAMPAK MENTAL HEALTH TERHADAP
IPK MAHASISWA MENGGUNAKAN TWO-WAY ANOVA
ANALISA STATISTIKA TERAPAN**

Dosen Pengampu : Fitrah Maharani Humaira M.Kom



Disusun Oleh:

Kadek Savita Dyutianaya

(3324600033)

Sains Data Terapan B

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER
PROGRAM STUDI SAINS DATA TERAPAN**

1. Tinjauan Pustaka

a. ANOVA (Analysis of Variance)

ANOVA merupakan metode statistik yang digunakan untuk menganalisis perbedaan rata-rata lebih dari dua kelompok. Dalam penerapannya, ANOVA didasarkan asumsi bahwa setiap populasi yang dianalisis bersifat independen, berdistribusi normal, dan memiliki variansi yang sama. Untuk memastikan asumsi-asumsi tersebut terpenuhi, dapat dilakukan pemeriksaan terhadap residual data. Residual didefinisikan sebagai selisih antara nilai observasi suatu data dalam populasi dengan nilai rata-rata yang diestimasi dari populasi tersebut.

Tujuan utama dari analisis ragam atau ANOVA adalah sebagai prediksi dan sebagai eksplanasi. Dalam konteks prediksi, ANOVA digunakan sebagai alat eksplorasi untuk menemukan kombinasi prediktor respon yang paling tepat. Metode ini mampu membedakan pengaruh berbagai sumber variasi yang berbeda secara bersamaan, serta mengidentifikasi faktor-faktor yang saling berinteraksi. Oleh karena itu, ANOVA ini dianggap efektif dalam menjawab pertanyaan tentang kausalitas. (Doncaster dan Davey, 2007). Selain itu, disebutkan juga bahwa tujuan ANOVA adalah membandingkan rata-rata antar faktor/kelompok, menganalisis variabilitas data termasuk Between Sum Square dan Within Sum Square, serta mencegah kesalahan tipe 1 (false positif), di mana hipotesis nol yang sebenarnya benar, tapi di tolak.

Terdapat dua jenis ANOVA, yakni One-way ANOVA yang digunakan untuk mengetahui apakah rata-rata dependen berbeda secara signifikan antar kategori, dan Two-way ANOVA yang digunakan untuk mengetahui bagaimana rata-rata suatu variabel kuantitatif dipengaruhi oleh dua variabel kategori sekaligus.

b. Two-way ANOVA

Seperti definisi di poin a, Two-way ANOVA melibatkan satu variabel dependen (y) dan dua variabel independen (x). Two-way ANOVA digunakan ketika terdapat dua variabel independen dan ingin mengetahui apakah interaksi keduanya berpengaruh terhadap variabel dependen. Variabel dependen pada Two-way ANOVA harus berskala numerik. Sedangkan variabel independen kategorikal.

c. P-Value

p-value adalah probabilitas yang menunjukkan seberapa kuat bukti terhadap hipotesis nol (H_0). Jika $p\text{-value} < \alpha$ (biasanya 0.05), maka terdapat cukup bukti untuk menolak H_0 dan menyimpulkan bahwa terdapat perbedaan yang signifikan secara statistik. Sebaliknya, jika $p\text{-value} \geq \alpha$, maka H_0 tidak dapat ditolak, artinya tidak ada bukti yang cukup untuk menyatakan adanya perbedaan yang signifikan.

Dengan demikian, p-value membantu peneliti dalam pengambilan keputusan: apakah hasil yang diperoleh hanya karena kebetulan (random) atau benar-benar menunjukkan pengaruh nyata.

d. Uji Asumsi

Sebelum melakukan ANOVA, ada beberapa asumsi penting yang harus dipenuhi agar hasil analisis valid:

- Normalitas residual:

Data dalam setiap kelompok atau perlakuan harus mengikuti distribusi normal. Ini berarti distribusi frekuensi data harus simetris di sekitar nilai rata-rata. Asumsi ini biasanya diuji dengan Shapiro-Wilk. Jika tidak normal, data dapat dilakukan transformasi dengan log atau sqrt atau menggunakan uji non-parametrik sebagai alternatif.

- **Homogenitas Varians**

Varians dari kelompok-kelompok yang dibandingkan harus sebanding, artinya variasi dalam setiap kelompok sekitar nilai rata-rata yang sama (seimbang). Asumsi ini biasanya diuji menggunakan Levene's Test. Jika asumsi tidak terpenuhi, hasil ANOVA bisa bias.

- **Independensi**

Data antar kelompok harus independen (tidak saling memengaruhi). Biasanya dijamin oleh desain penelitian, misalnya sampel diambil secara acak. Ini sangat penting, karena kalau ada dependensi maka Two-Way ANOVA tidak valid.

2. Data Understanding

Dataset yang digunakan berjudul "*Student Mental health: A STATISTICAL RESEARCH ON THE EFFECTS OF MENTAL HEALTH ON STUDENTS' CGPA dataset*".
<https://www.kaggle.com/datasets/shariful07/student-mental-health?resource=download>

Kolom	Keterangan
<i>Timestamp</i>	Waktu pengisian survey
<i>Choose your gender</i>	Jenis kelamin mahasiswa (Female/Male)
<i>Age</i>	Usia mahas
<i>What is your course</i>	Jurusan mahasiswa
<i>Your current year of study?</i>	Tahun studi mahasiswa berkuliah (1-4 tahun)
<i>What is your CGPA</i>	Nilai IPK mahasiswa
<i>Marital Status</i>	Status pernikahan
<i>Do you have Depression?</i>	Apakah mahasiswa mengalami depresi (Yes/No)
<i>Do you have Anxiety?</i>	Apakah mahasiswa mengalami gangguan kecemasan (Yes/No)
<i>Do you have Panic Attack?</i>	Apakah mahasiswa mengalami serangan panik (Yes/No)
<i>Did you seek any specialist for a treatment?</i>	Apakah mahasiswa mencari bantuan profesional untuk penanganan <i>mental health</i> ? (Yes/No)

Dataset ini dikumpulkan melalui survei yang dilakukan dengan menggunakan Google formulir dari mahasiswa IIUM (International Islamic University Malaysia) untuk memeriksa situasi akademis dan kesehatan mental mereka saat ini. Dari Judul, diketahui bahwa dataset ini digunakan oleh penulis untuk penelitian statistik mengenai dampak kesehatan mental terhadap CGPA (IPK) Mahasiswa. Dari informasi ini, dapat disimpulkan bahwa CGPA merupakan variabel dependen, dan sisanya dapat dikategorikan independen. Dari preview kaggle, diketahui bahwa 74% dari pengisi survei ini adalah perempuan, dan 26% sisanya adalah Pria.

Pada kasus ini, karena membahas mengenai kesehatan mental dan nilai, maka variabel utama yang digunakan adalah:

- CGPA (dependent / Y), sebelumnya dalam bentuk rentang sehingga ditransformasikan menjadi angka midpoint (CGPA_cat)
- Anxiety (independent / X1, kategori: Yes/No)
- Gender (independent / X2, kategori: Female/Male)

Dilansir dari artikel kementerian kesehatan <https://ayosehat.kemkes.go.id/jenis-gangguan-mental>, terdapat beberapa jenis gangguan mental yang perlu diketahui, salah satunya adalah gangguan kecemasan (Anxiety). Melanjutkan hal tersebut, dilansir dari artikel halodoc <https://www.halodoc.com/artikel/ini-alasan-wanita-lebih-rentan-alami-gangguan-kecemasan?>, diketahui bahwa wanita lebih rentan mengalami gangguan kecemasan akibat perbedaan fungsi otak dan hormon yang dimilikinya. Hal ini membawa saya menuju keputusan akhir dalam pemilihan dua variabel independen yaitu Anxiety dan Gender terhadap variabel dependen yaitu CGPA atau nilai IPK mahasiswa. Dengan harapan, outputnya benar-benar memberikan insight bahwa asumsi awal saya berdasarkan literatur adalah benar.

Perlu diketahui bahwa Jumlah data adalah 100 baris (mahasiswa).

3. Penerapan pada Python dan Hasil Analisis

a. Memuat data:

Code:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm
from statsmodels.stats.multicomp import pairwise_tukeyhsd

from scipy.stats import shapiro, levene

mental_health = pd.read_csv('student_mental_health.csv')
mental_health = mental_health.rename(columns={
    "Choose your gender" : "Gender",
    "What is your course?" : "Course",
    "Your current year of Study" : "Year_of_Study",
    "What is your CGPA?": "CGPA",
    "Marital status" : "Marital_status",
    "Do you have Depression?": "Depression",
    "Do you have Anxiety?": "Anxiety",
    "Do you have Panic attack?": "Panic_Attack",
    "Did you seek any specialist for a treatment?": "Treatment"
})

mental_health
```

✓ 0.0s

Ouput:

	Timestamp	Gender	Age	Course	Year_of_Study	CGPA	Marital_status	Depression	Anxiety	Panic_Attack	Treatment
0	8/7/2020 12:02	Female	18.0	Engineering	year 1	3.00 - 3.49	No	Yes	No	Yes	No
1	8/7/2020 12:04	Male	21.0	Islamic education	year 2	3.00 - 3.49	No	No	Yes	No	No
2	8/7/2020 12:05	Male	19.0	BIT	Year 1	3.00 - 3.49	No	Yes	Yes	Yes	No
3	8/7/2020 12:06	Female	22.0	Laws	year 3	3.00 - 3.49	Yes	Yes	No	No	No
4	8/7/2020 12:13	Male	23.0	Mathematics	year 4	3.00 - 3.49	No	No	No	No	No
...
96	13/07/2020 19:56:49	Female	21.0	BCS	year 1	3.50 - 4.00	No	No	Yes	No	No
97	13/07/2020 21:21:42	Male	18.0	Engineering	Year 2	3.00 - 3.49	No	Yes	Yes	No	No
98	13/07/2020 21:22:56	Female	19.0	Nursing	Year 3	3.50 - 4.00	Yes	Yes	No	Yes	No
99	13/07/2020 21:23:57	Female	23.0	Pendidikan Islam	year 4	3.50 - 4.00	No	No	No	No	No
100	18/07/2020 20:16:21	Male	20.0	Biomedical science	Year 2	3.00 - 3.49	No	No	No	No	No

101 rows x 11 columns

Analisis:

Sebelum melakukan ANOVA, saya melakukan load dataset yang telah didownload sebelumnya terlebih dahulu, mengimport semua library yang diperlukan, dan mengubah nama kolom yang sebelumnya deskriptif menjadi lebih singkat agar mempermudah kodingan kedepannya. Setelah dicetak, terlihat bahwa kolom dataset telah direname.

b. EDA

Code:

```

# One-way ANOVA
import statsmodels.api as sm
from statsmodels.formula.api import ols

# depression
model_oneway = ols("CGPA_num ~ C(Depression)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Depression vs CGPA):")
print(anova_oneway)

# gender
model_oneway = ols("CGPA_num ~ C(Gender)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Gender vs CGPA):")
print(anova_oneway)

# anxiety
model_oneway = ols("CGPA_num ~ C(Anxiety)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Anxiety vs CGPA):")
print(anova_oneway)

# treatment
model_oneway = ols("CGPA_num ~ C(Treatment)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Treatment vs CGPA):")
print(anova_oneway)

# Year_of_Study
model_oneway = ols("CGPA_num ~ C(Year_of_Study)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Year_of_Study vs CGPA):")
print(anova_oneway)

# Panic_Attack
model_oneway = ols("CGPA_num ~ C(Panic_Attack)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Panic_Attack vs CGPA):")
print(anova_oneway)

# Age
model_oneway = ols("CGPA_num ~ C(Age)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Age vs CGPA):")
print(anova_oneway)

# Marital_status
model_oneway = ols("CGPA_num ~ C(Marital_status)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Marital_status vs CGPA):")
print(anova_oneway)

# Course
model_oneway = ols("CGPA_num ~ C(Course)", data=clean_mental_health).fit()
anova_oneway = sm.stats.anova_lm(model_oneway, typ=2)

print("\nHasil One-Way ANOVA (Course vs CGPA):")
print(anova_oneway)

```

✓ 0.0s

Output:

Hasil One-Way ANOVA (Depression vs CGPA):

	sum_sq	df	F	PR(>F)
C(Depression)	0.012314	1.0	0.073678	0.786621
Residual	16.546090	99.0	NaN	NaN

Hasil One-Way ANOVA (Gender vs CGPA):

	sum_sq	df	F	PR(>F)
C(Gender)	0.357855	1.0	2.18682	0.142371
Residual	16.200549	99.0	NaN	NaN

Hasil One-Way ANOVA (Anxiety vs CGPA):

	sum_sq	df	F	PR(>F)
C(Anxiety)	0.433625	1.0	2.662291	0.10593
Residual	16.124779	99.0	NaN	NaN

Hasil One-Way ANOVA (Treatment vs CGPA):

	sum_sq	df	F	PR(>F)
C(Treatment)	0.000477	1.0	0.002852	0.95752
Residual	16.557927	99.0	NaN	NaN

Hasil One-Way ANOVA (Year_of_Study vs CGPA):

	sum_sq	df	F	PR(>F)
C(Year_of_Study)	0.586887	6.0	0.575685	0.748804
Residual	15.971517	94.0	NaN	NaN

Hasil One-Way ANOVA (Panic_Attack vs CGPA):

	sum_sq	df	F	PR(>F)
C(Panic_Attack)	0.027058	1.0	0.162038	0.688155
Residual	16.531346	99.0	NaN	NaN

Hasil One-Way ANOVA (Age vs CGPA):

	sum_sq	df	F	PR(>F)
C(Age)	0.349665	6.0	0.365443	0.899064
Residual	14.830826	93.0	NaN	NaN

Hasil One-Way ANOVA (Marital_status vs CGPA):

	sum_sq	df	F	PR(>F)
C(Marital_status)	0.000060	1.0	0.000356	0.984984
Residual	16.558344	99.0	NaN	NaN

Hasil One-Way ANOVA (Course vs CGPA):

	sum_sq	df	F	PR(>F)
C(Course)	8.947667	48.0	1.273636	0.19646
Residual	7.610737	52.0	NaN	NaN

Analisis:

Pada tahap Exploratory Data Analysis ini, dilakukan uji signifikansi antara variabel (CGPA) dengan masing-masing variabel independen secara terpisah menggunakan **One-Way ANOVA** untuk sekali lagi meyakinkan saya terhadap variabel utama yang telah dijelaskan sebelumnya, untuk digunakan dalam analisis lanjutan. Output ini menunjukkan bahwa sebagian besar variabel independen memiliki

nilai p-value yang relatif besar ($p > 0,05$), sehingga tidak signifikan. Namun terlihat bahwa variabel Gender dan Anxiety memiliki p-value yang paling mendekati batas signifikansi. Hal ini menandakan bahwa terdapat kemungkinan kedua faktor tersebut lebih berpotensi memengaruhi nilai CGPA/IPK mahasiswa.

Code:

```
print(mental_health.info())
```

✓ 0.0s

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101 entries, 0 to 100
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Timestamp              101 non-null   object 
1   Gender                 101 non-null   object 
2   Age                   100 non-null   float64
3   Course                101 non-null   object 
4   Year_of_Study         101 non-null   object 
5   CGPA                  101 non-null   object 
6   Marital_status        101 non-null   object 
7   Depression            101 non-null   object 
8   Anxiety               101 non-null   object 
9   Panic_Attack         101 non-null   object 
10  Treatment              101 non-null   object 
dtypes: float64(1), object(10)
memory usage: 8.8+ KB
None
```

Analisis:

Selanjutnya, dilakukan pengecekan info detail terhadap dataset. Output menunjukkan bahwa kolom dependen yang seharusnya numerik masih bertipe data object. Selain itu, output juga menunjukkan bahwa terdapat 100 baris data dengan 11 kolom, di mana tidak terdapat data yang Nan (None)

Code:

```
print("\nJumlah data tiap grup (Gender x Anxiety):")
mental_health.groupby(["Gender", "Anxiety"]).describe()
```

✓ 0.0s

Output:

Jumlah data tiap grup (Gender x Anxiety):

		Age							
		count	mean	std	min	25%	50%	75%	max
Gender	Anxiety								
Female	No	51.0	20.568627	2.539724	18.0	18.0	19.0	23.00	24.0
	Yes	24.0	20.458333	2.519130	18.0	18.0	19.5	23.25	24.0
Male	No	15.0	21.133333	2.356349	18.0	19.0	21.0	23.00	24.0
	Yes	10.0	19.600000	2.503331	18.0	18.0	18.0	20.50	24.0

Analisis:

pada tahap ini, dilakukan pengecekan data tiap grup (Gender x Anxiety) untuk mengetahui jumlah, rata-rata, standar deviasi, minimal, kuartil, dan maksimal. Di mana ini memberikan informasi bahwa:

- 51 mahasiswa perempuan yang tidak mengalami anxiety
- 24 mahasiswa perempuan yang mengalami anxiety
- 15 mahasiswa pria yang tidak mengalami anxiety
- 10 mahasiswa pria yang mengalami anxiety

Tapi karena informasi awal terdapat 74% perempuan yang mengisi survey ini, maka dapat dikatakan data ini cukup bias

c. Preprocessing

Code:

```

# normalisasi sebelum mapping (hilangkan spasi berlebih)
mental_health["CGPA"] = mental_health["CGPA"].str.strip()
mental_health["Gender"] = mental_health["Gender"].str.strip()
mental_health["Anxiety"] = mental_health["Anxiety"].str.strip()

# mapping
mental_health["CGPA_cat"] = mental_health["CGPA"].map({
    "0 - 1.99": 0.995,
    "2.00 - 2.49": 2.245,
    "2.50 - 2.99": 2.745,
    "3.00 - 3.49": 3.245,
    "3.50 - 4.00": 3.75
})

mental_health["Gender_cat"] = mental_health["Gender"].map({
    "Female": 0,
    "Male": 1
})

mental_health["Anxiety_cat"] = mental_health["Anxiety"].map({
    "No": 0,
    "Yes": 1
})

# cek
print(mental_health[["CGPA", "CGPA_cat"]].head(5))
print()
print(mental_health[["Gender", "Gender_cat"]].head(5))
print()
print(mental_health[["Anxiety", "Anxiety_cat"]].head(5))

```

✓ 0.0s

Output:

	CGPA	CGPA_cat
0	3.00 - 3.49	3.245
1	3.00 - 3.49	3.245
2	3.00 - 3.49	3.245
3	3.00 - 3.49	3.245
4	3.00 - 3.49	3.245

	Gender	Gender_cat
0	Female	0
1	Male	1
2	Male	1
3	Female	0
4	Male	1

	Anxiety	Anxiety_cat
0	No	0
1	Yes	1
2	Yes	1
3	No	0
4	No	0

Analisis:

Pada tahap ini, dilakukan normalisasi pada kolom yang akan digunakan, yaitu menghilangkan spasi berlebih, untuk menghindari error. Setelah dinormalisasi, dilakukan mapping untuk ketiga kolom tersebut. Di mana kolom CGPA yang sebelumnya berbentuk rentang akan diubah menjadi nilai tengah, lalu gender yang sebelumnya berupa kategori female-male diubah menjadi 0 dan 1, serta Anxiety yang sebelumnya kategori No-Yes diubah menjadi 0 dan 1.

Code:

```
# Drop missing value hanya di kolom CGPA_cat
clean_mental_health = mental_health.dropna(subset=["CGPA_cat"]).copy()

# Pastikan CGPA_cat numeric (bukan string/label)
clean_mental_health["CGPA_cat"] = clean_mental_health["CGPA_cat"].astype(float)
```

✓ 0.0s

Analisis:

Pada tahap ini dilakukan penghapusan missing value yang berguna untuk memastikan bahwa dataset aman untuk dianalisis. Selain itu, perlu dipastikan bahwa variabel dependen bertipe numerik dikarenakan dalam two-way ANOVA, variabel dependennya harus berupa numerik.

Code:

```
# Timpa kolom lama dengan hasil cleaning
clean_mental_health["CGPA"] = clean_mental_health["CGPA_cat"]
clean_mental_health["Gender"] = clean_mental_health["Gender_cat"]
clean_mental_health["Anxiety"] = clean_mental_health["Anxiety_cat"]

# Hapus kolom tambahan
clean_mental_health = clean_mental_health.drop(columns=["CGPA_cat"])
clean_mental_health = clean_mental_health.drop(columns=["Gender_cat"])
clean_mental_health = clean_mental_health.drop(columns=["Anxiety_cat"])

# print(mental_health.info())
clean_mental_health
```

✓ 0.0s

Output:

	Timestamp	Gender	Age	Course	Year_of_Study	CGPA	Marital_status	Depression	Anxiety	Panic_Attack	Treatment
0	8/7/2020 12:02	0	18.0	Engineering	year 1	3.245	No	Yes	0	Yes	No
1	8/7/2020 12:04	1	21.0	Islamic education	year 2	3.245	No	No	1	No	No
2	8/7/2020 12:05	1	19.0	BIT	Year 1	3.245	No	Yes	1	Yes	No
3	8/7/2020 12:06	0	22.0	Laws	year 3	3.245	Yes	Yes	0	No	No
4	8/7/2020 12:13	1	23.0	Mathematics	year 4	3.245	No	No	0	No	No
...
96	13/07/2020 19:56:49	0	21.0	BCS	year 1	3.750	No	No	1	No	No
97	13/07/2020 21:21:42	1	18.0	Engineering	Year 2	3.245	No	Yes	1	No	No
98	13/07/2020 21:22:56	0	19.0	Nursing	Year 3	3.750	Yes	Yes	0	Yes	No
99	13/07/2020 21:23:57	0	23.0	Pendidikan Islam	year 4	3.750	No	No	0	No	No
100	18/07/2020 20:16:21	1	20.0	Biomedical science	Year 2	3.245	No	No	0	No	No

101 rows x 15 columns

Analisis:

Setelah tahap preprocessing diatas, data yang telah bersih dan siap ditimpa ke kolom lama agar tidak duplikat. Setelah menimpa kolom lama, kolom baru dapat dihapus karena sudah tidak digunakan lagi.

d. Cek Asumsi (Normalitas residual: Shapiro-Wilk & Homogenitas varians: Levene)

Code:

```
# model
model = ols("CGPA ~ C(Gender) * C(Anxiety)", data=clean_mental_health).fit()
residuals = model.resid

# Uji normalitas residual
shapiro_test = shapiro(residuals)
print(f"\nShapiro-Wilk Test: Statistik {shapiro_test.statistic:.4f}, p-value = {shapiro_test.pvalue:.4f}")

# Uji homogenitas varians
levene_test = levene(
    *[group["CGPA"].values for name, group in clean_mental_health.groupby(["Gender", "Anxiety"])]
)
print(f"Levene Test: Statistik = {levene_test.statistic:.4f}, p-value = {levene_test.pvalue:.4f}")

✓ 0.0s
```

Output:

```
Shapiro-Wilk Test: Statistik 0.7356, p-value = 0.0000
Levene Test: Statistik = 1.4696, p-value = 0.2276
```

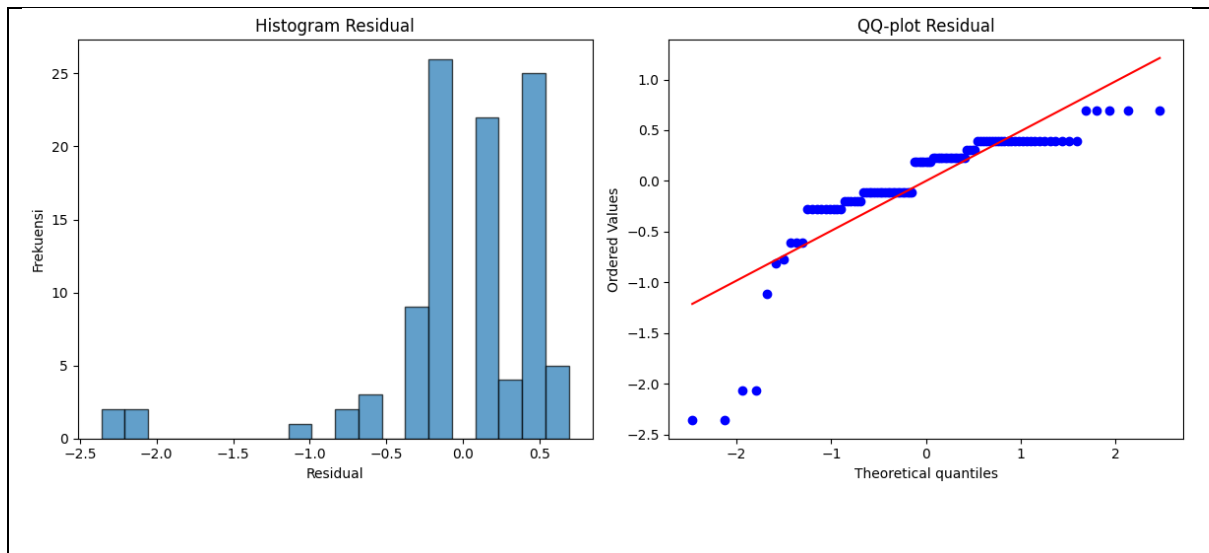
Analisis dan Interpretasi:

Uji asumsi terbagi menjadi dua, yaitu Normalitas residual (Shapiro-Wilk) dan Homogenitas varians (Levene). Uji normalitas residual (shapiro-wilk test) adalah pengujian yang menguji apakah residual dari model (selisih antara nilai prediksi dan nilai aktual) berdistribusi normal. Jika p-valuenya > 0.05 , maka residual berdistribusi normal. Pada kasus ini, $p\text{-value} < 0.05$, sehingga residual dikatakan tidak berdistribusi normal. Ini mengindikasikan bahwa asumsi normalitas untuk ANOVA tidak terpenuhi. Sedangkan uji homogenitas varians (levene test) adalah pengujian yang menguji apakah varians antar kelompok (Gender x Anxiety) homogen/sama besar. jika p-valuenya > 0.05 , maka tidak ada perbedaan yang signifikan, sehingga bisa dianggap homogen. Pada kasus ini, $p\text{-value} > 0.05$ sehingga dapat dikatakan homogen.

Uji asumsi ini menunjukkan bahwa hanya homogenitas varian yang terpenuhi, sedangkan normalitas residual gagal. Sehingga dapat disimpulkan bahwa two-way ANOVA masih bisa dilakukan, dengan syarat perlu kehati-hatian akibat normalitas yang tidak terpenuhi. Alternatif yang ditawarkan adalah melakukan transformasi (log, sqrt) untuk mendekatkan distribusi ke normal. Selain itu dapat menggunakan metode non-parametrik juga.

Secara keseluruhan, dapat disimpulkan bahwa tahap ini menunjukkan bahwa data cukup stabil dari sisi varians, namun distribusinya bias/tidak normal.

Visualisasi Normalitas residual:



e. Transformasi data (opsional)

Code:

```
import numpy as np
from scipy.stats import boxcox

# Transformasi log
clean_mental_health["CGPA_log"] = np.log(clean_mental_health["CGPA_num"])

# Transformasi sqrt
clean_mental_health["CGPA_sqrt"] = np.sqrt(clean_mental_health["CGPA_num"])

✓ 0.0s
```

Analisis dan Interpretasi:

Dikarenakan hasil uji asumsi menunjukkan bahwa residual tidak berdistribusi normal, maka dilakukan alternatif transformasi data. Transformasi yang digunakan adalah **logaritmik (log)** dan **akar kuadrat (sqrt)**, karena keduanya paling umum digunakan untuk memperbaiki distribusi yang skewed. Pada tahap ini, transformasi tersebut dipilih karena paling relevan terhadap data yang ada, meskipun hasilnya hanya cukup membantu mendekati signifikansi, namun belum sepenuhnya menjadikan variabel menjadi signifikan.

Code:

```
# pengecekan
model_log = ols("CGPA_log ~ C(Gender) * C(Anxiety)", data=clean_mental_health).fit()
model_sqrt = ols("CGPA_sqrt ~ C(Gender) * C(Anxiety)", data=clean_mental_health).fit()

from scipy.stats import shapiro

for name, model in [("Log", model_log), ("Sqrt", model_sqrt)]:
    stat, p = shapiro(model.resid)
    print(f"{name} Transform - Shapiro-Wilk: Statistik={stat:.4f}, p-value={p:.4f}")

✓ 0.1s
```

Output:

Log Transform - Shapiro-Wilk: Statistik=0.8258, p-value=0.0000
Sqrt Transform - Shapiro-Wilk: Statistik=0.8462, p-value=0.0000

Analisis dan Interpretasi:

Setelah dilakukan transformasi, $p\text{-value} < 0.05$, di mana menunjukkan bahwa data residual masih tidak berdistribusi normal meskipun telah dilakukan transformasi log dan sqrt. Sedangkan nilai statistik shapiro-wilk semakin mendekati 1, ini membuktikan bahwa transformasi memang membantu untuk memperbaiki distribusi (mengurangi skewness), namun ini belum cukup untuk memperbaiki masalah normalitas.

Alternatif lainnya adalah uji non-parametrik, namun pada penugasan Analisa Statistika Terapan kali ini, fokusnya hanya uji ANOVA di Python.

f. Two-way ANOVA

Code:

```
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Two-Way ANOVA
model2 = smf.ols("CGPA ~ C(Gender) + C(Anxiety) + C(Gender):C(Anxiety)", data=clean_mental_health).fit()
anova2 = sm.stats.anova_lm(model2, typ=3)
print(anova2)
```

✓ 0.0s

Output:

	sum_sq	df	F	PR(>F)
Intercept	574.157929	1.0	1734.397105	1.077665e-63
C(Gender)	1.068755	1.0	3.228460	7.548109e-02
C(Anxiety)	0.436035	1.0	1.317160	2.539251e-01
C(Gender):C(Anxiety)	0.225191	1.0	0.680249	4.115245e-01
Residual	32.111054	97.0	NaN	NaN

Analisis dan Interpretasi:

Pada tahap ini dilakukan pengujian Two-way ANOVA setelah melalui proses cleaning dan transformasi. Output menunjukkan bahwa:

- Intercept: menunjukkan nilai rata-rata umum yang signifikan
- Gender ($F = 3.228$, $p = 0.075$): karena $p > 0.05$, maka tidak ada perbedaan/pengaruh Gender secara signifikan terhadap CGPA
- Anxiety ($F = 1.317$, $p = 0.254$): karena $p > 0.05$, maka tidak ada perbedaan/pengaruh Anxiety secara signifikan terhadap CGPA
- Interaksi Gender & Anxiety ($F = 0.680$, $p = 0.412$): karena $p > 0.05$, maka tidak signifikan di mana efek Gender terhadap variabel dependen tidak bergantung pada tingkat Anxiety, dan sebaliknya.
- Residual ($df = 97$, $sum_sq = 32.111$): ini merupakan variasi dalam data yang tidak dijelaskan oleh faktor-faktor diatas.

Dari tabel ini dapat disimpulkan bahwa secara umum tidak ada efek signifikan dari Gender, Anxiety, maupun interaksi dari keduanya terhadap CGPA (karena semuanya $p > 0.05$). namun, diantara semua kolom yang dimiliki dataset ini, hanya faktor ini yang mendekati signifikan. Dibuktikan dengan variabel Gender yang mendekati signifikan ($p = 0.075$) sehingga ada indikasi kecil bahwa mungkin terdapat perbedaan, namun tidak cukup kuat secara statistik pada level signifikansi 0.05.

g. Post Hoc (Tukey HSD)

Code:

- meandiff = selisih rata-rata CGPA.
- p-adj = p-value hasil Tukey.
- reject = True artinya signifikan beda, False artinya tidak beda signifikan.

```
import pandas as pd
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Asumsi: mental_health sudah bersih dan CGPA sudah numerik (float)
# Post-hoc Tukey HSD untuk faktor Gender
# menggunakan CGPA nilai tengah
tukey1 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA"],
                           groups=clean_mental_health["Gender"],
                           alpha=0.05)
tukey2 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA"],
                           groups=clean_mental_health["Anxiety"],
                           alpha=0.05)

print(tukey1)
print(tukey2)
print()
print('><'*25)
print()

# menggunakan CGPA yang sudah di transformasi sqrt
tukey11 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA_sqrt"],
                             groups=clean_mental_health["Gender"],
                             alpha=0.05)
tukey22 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA_sqrt"],
                             groups=clean_mental_health["Anxiety"],
                             alpha=0.05)

print(tukey11)
print(tukey22)
print()
print('><'*25)
print()

# menggunakan CGPA yang sudah di transformasi log
tukey111 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA_log"],
                              groups=clean_mental_health["Gender"],
                              alpha=0.05)
tukey222 = pairwise_tukeyhsd(endog=clean_mental_health["CGPA_log"],
                              groups=clean_mental_health["Anxiety"],
                              alpha=0.05)

print(tukey111)
print(tukey222)
```

68]

✓ 2.2s

Output, Analisis dan Interpretasi:

- # 1. menggunakan CGPA nilai tengah


```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
      0      1  -0.1993 0.1352 -0.4619 0.0633  False

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
      0      1   0.2131 0.0837 -0.0289 0.4551  False

```

Tukey HSD by Gender:

Output ini menunjukkan rata-rata CGPA pria (1) lebih rendah 0.20 poin (dilihat dari meandiff) dibanding perempuan (0), tapi **tidak signifikan** ($p=0.1352 > 0.05$). untuk interval kepercayaannya (lower & upper) mencakup 0, artinya memang tidak bisa dipastikan ada beda nyata

Tukey HSD by Anxiety:

Output ini menunjukkan rata-rata CGPA pria (1) lebih tinggi 0.21 poin dibanding perempuan (0), lebih dekat ke signifikan ($p = 0.0837$), tapi tetap **tidak signifikan** karena lebih dari 0.05. untuk interval kepercayaannya masih melewati 0 jadi tidak bisa dianggap perbedaan nyata

- # 2. menggunakan CGPA yang sudah di transformasi sqrt

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
      0      1  -0.0402 0.1447 -0.0945 0.0141  False

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
      0      1   0.0423 0.0965 -0.0077 0.0924  False

```

Tukey HSD by Gender:

Output ini menunjukkan rata-rata CGPA pria (1) lebih rendah 0.04 poin (dilihat dari meandiff) dibanding perempuan (0), tapi **tidak signifikan** ($p = 0.1447 > 0.05$).

Tukey HSD by Anxiety:

Output ini menunjukkan rata-rata CGPA pria (1) lebih tinggi 0.04 poin dibanding perempuan (0), lebih dekat ke signifikan $p = 0.096$ di mana > 0.05 . sehingga **tidak signifikan**. namun angkanya lebih mendekati signifikan dibandingkan sebelum melakukan transformasi.

- # 3. menggunakan CGPA yang sudah di transformasi log

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
      0      1  -0.0479 0.1474 -0.113 0.0172 False
-----

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
      0      1   0.052 0.0884 -0.008 0.112 False
-----

```

Tukey HSD by Gender:

Output ini menunjukkan rata-rata CGPA pria (1) lebih rendah 0.047 poin (dilihat dari meandiff) dibanding perempuan (0), tapi **tidak signifikan** ($p = 0.1474 > 0.05$).

Tukey HSD by Anxiety:

Output ini menunjukkan rata-rata CGPA pria (1) lebih tinggi 0.052 poin dibanding perempuan (0), lebih dekat ke sedangkan $p = 0.088$ di mana > 0.05 . sehingga **tidak signifikan**. angkanya tidak lebih baik dari transformasi sqrt.

Analisis dan Interpretasi secara umum:

pada tahap ini, dilakukan Post hoc teset menggunakan Tukey HSD untuk:

- mengidentifikasi kelompok mana yang berbeda signifikan. Ini dikarenakan ANOVA hanya memberitahu ada tidaknya pengaruh signifikan, sehingga belum tahu mana yang signifikan.
- mengontrol error tipe 1 (false positif)
- memberi ukuran perbedaan antar kelompok yaitu p-value, selisih rata-rata, confidence interval, dan apakah perbedaan itu signifikan.

setelah membandingkan Tukey HSD menggunakan ketiga CGPA yang berbeda, tidak ada perbedaan signifikan antara kelompok baik Gender maupun Anxiety pada semua kondisi (data CGPA asli, sqrt, dan log) karena semua $p\text{-adj} > 0.05$.

dalam konteks arah perbedaan (meandiff), nilai Gender selalu negatif (artinya rata-rata CGPA kelompok perempuan sedikit lebih rendah dari pria). Untuk Anxiety, nilainya selalu positif (artinya rata-rata CGPA perempuan lebih sedikit tinggi dari pria) namun karena tidak signifikan, perbedaan ini hanya bersifat kecenderungan, bukan bukti statistik.

Dalam konteks transformasi data (sqrt & log) mereka belum bisa mengubah kesimpulan utama menjadi signifikan. Efek transformasi hanya membuat selisih mean terlihat lebih kecil, dan ini wajar karena transformasi hanya mengubah skala saja. namun dapat disimpulkan bahwa CGPA dengan transformasi sqrt yang memberikan hasil mendekati signifikan, walaupun belum benar-benar signifikan. transformasi ini hanya memperbaiki distribusi untuk memenuhi asumsi ANOVA, tapi tidak mengubah hasil uji signifikan.

4. Summary Statistik dan Visualisasi

Model ini tidak menunjukkan adanya interaksi antara kedua faktor, di mana keduanya berarti tidak saling memengaruhi efeknya pada variabel dependen. Ini dapat diakibatkan oleh dataset yang kurang banyak, atau memang tidak ada efek dari interaksi kedua faktor tersebut.

Code:

```
print (model2.summary())
```

✓ 0.0s

Output:

```

=====
                        OLS Regression Results
=====
Dep. Variable:          CGPA      R-squared:                0.062
Model:                  OLS      Adj. R-squared:            0.033
Method:                 Least Squares      F-statistic:          2.144
Date:                   Mon, 15 Sep 2025    Prob (F-statistic):    0.0996
Time:                   16:18:38           Log-Likelihood:        -85.444
No. Observations:       101             AIC:                  178.9
Df Residuals:           97              BIC:                  189.3
Df Model:                3
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept                3.3553      0.081     41.646     0.000      3.195      3.515
C(Gender)[T.1]          -0.2962      0.165    -1.797     0.075     -0.623      0.031
C(Anxiety)[T.1]          0.1635      0.142     1.148     0.254     -0.119      0.446
C(Gender)[T.1]:C(Anxiety)[T.1]  0.2245      0.272     0.825     0.412     -0.316      0.765
=====
Omnibus:                 69.835   Durbin-Watson:          2.091
Prob(Omnibus):            0.000   Jarque-Bera (JB):       309.444
Skew:                    -2.412   Prob(JB):               6.38e-68
Kurtosis:                 10.089   Cond. No.               5.96
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Analisis:

pada tahap ini, ditunjukkan hasil summary dari OLS Regression yang setara dengan analisis ANOVA yang dijalankan sebelumnya.

- Variabel dependen adalah CGPA
- R-squared = 0.062, Adj.R² = 0.033: bahwa model hanya menjelaskan sekitar 6.2% variasi pada variabel dependen
- F-statistic = 2.144, p = 0.0996: bahwa secara keseluruhan model tidak signifikan pada alpha = 0.05

Interpretasi:

- Intercept = 3.355 (p < 0.001): ini merupakan nilai rata-rata perempuan tanpa anxiety
- Gender [Male] = -0.296 (p = 0.075): CGPA lebih rendah sekitar 0.29 poin dibanding perempuan. Ini tidak signifikan, tapi mendekati (ada indikasi perbedaan)
- Anxiety [Yes] = 0.163 (p = 0.254): CGPA mahasiswa dengan Anxiety lebih tinggi 0.16 poin. ini tidak signifikan
- Interaksi (Male x Anxiety) = 0.225 (p = 0.412): Efek gabungan laki-laki + anxiety meningkatkan CGPA sekitar 0.22 poin. Tidak signifikan

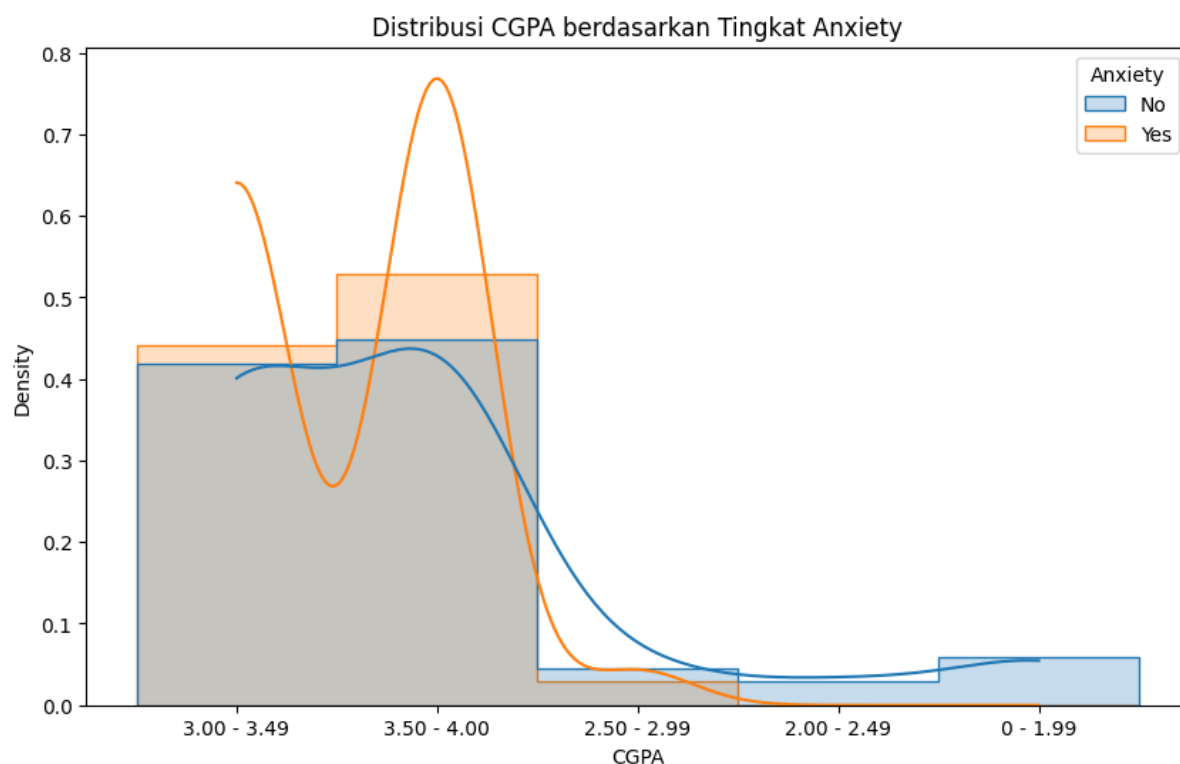
Uji asumsi:

- Omnibus & Jarque-Bera = 0.000: menunjukkan residu tidak normal
- Skew = -2.412: menunjukkan distribusi miring ke kiri

- Kurtosis = 10.089: data memiliki ekor berat. Artinya data punya lebih banyak nilai outlier dibanding normal
- Durbin-Watson = 2.091: tidak ada autokorelasi signifikan (nilai mendekati 2, artinya bagus)

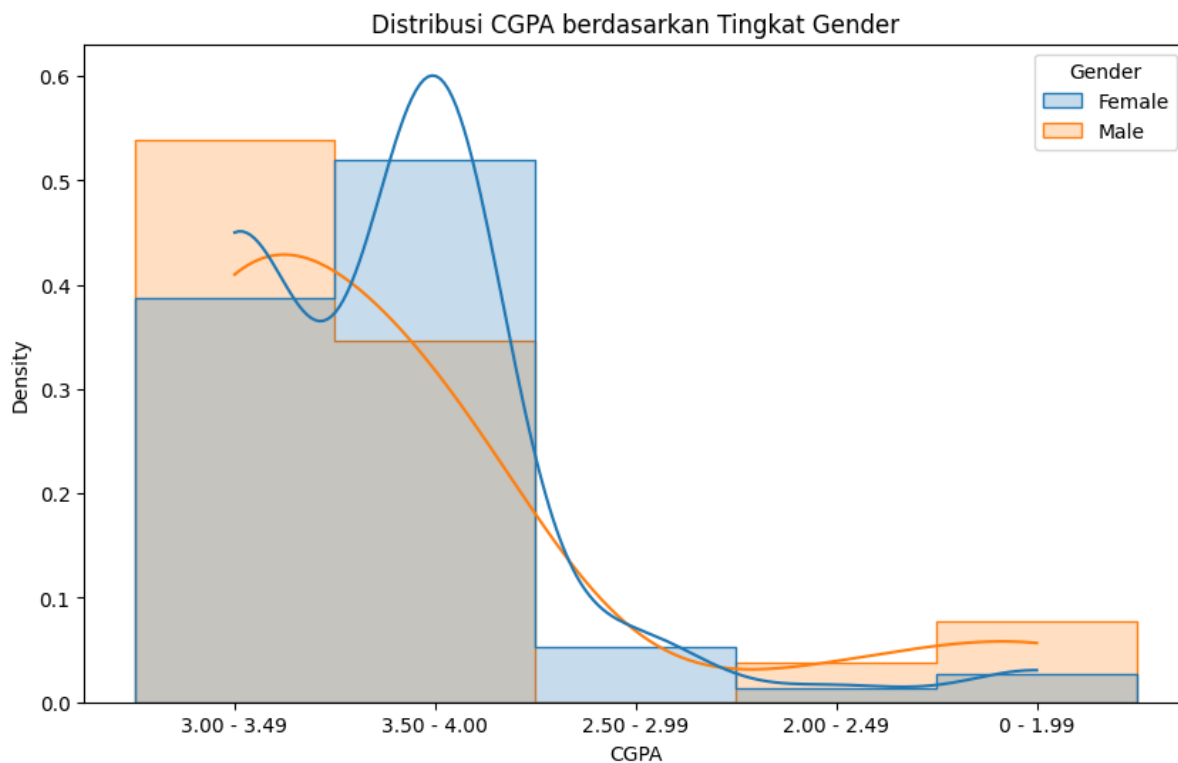
Dari informasi tersebut, dapat disimpulkan bahwa model tidak signifikan secara keseluruhan ($p = 0.0996$). artinya faktor-faktor Gender dan Anxiety dan interaksinya tidak mampu menjelaskan variasi CGPA secara kuat. Disamping itu, Gender menunjukkan pengaruh mendekati signifikan ($p = 0.075$), di mana laki-laki cenderung memiliki CGPA yang lebih rendah. selain itu, Anxiety dan Interaksi kedua variabel independen tidak signifikan sehingga tidak cukup memengaruhi CGPA. Dan terakhir, uji asumsi residu dapat dikatakan gagal karena normalitas tidak terpenuhi walaupun telah dilakukan transformasi data sekalipun. Alternatif lain adalah uji non-parametrik, namun tidak selaras dengan pembahasan tugas ini.

Visualisasi:



1. Kelompok Anxiety = Yes (garis oranye):
 - Distribusi memuncak pada CGPA 3.50 - 4.00, artinya sebagian besar mahasiswa dengan anxiety memiliki IPK tinggi.
 - Namun, ada juga variasi cukup besar di sekitar rentang 3.00 - 3.49, walau tidak sebesar kelompok No.
 - Hampir tidak ada mahasiswa dengan anxiety yang IPK-nya di bawah 3.0.
2. Kelompok Anxiety = No (garis biru):
 - Distribusi lebih menyebar dibanding yang Anxiety = Yes.
 - Puncak tertinggi juga di 3.50 - 4.00, tapi masih ada mahasiswa tanpa anxiety yang nilainya lebih rendah (bahkan sampai < 2.0).
 - Artinya, mahasiswa tanpa anxiety punya variasi CGPA yang lebih lebar (ada yang sangat tinggi, tapi ada juga yang rendah).
3. Perbandingan Kedua Kelompok:
 - Sama-sama punya puncak di 3.50 - 4.00, tapi:

- Anxiety Yes: cenderung terkonsentrasi pada IPK tinggi.
- Anxiety No: lebih tersebar, ada yang tinggi, ada yang rendah.
- Ini bisa menunjukkan bahwa tingkat kecemasan tidak serta merta membuat mahasiswa punya IPK rendah, justru mereka mungkin lebih termotivasi untuk menjaga performa akademik.



1. Mahasiswa Perempuan (biru):
 - Distribusi memuncak di rentang 3.50 - 4.00, artinya sebagian besar perempuan memiliki IPK tinggi.
 - Hampir tidak ada perempuan dengan IPK sangat rendah (< 2.0).
 - Polanya lebih terfokus (tidak terlalu menyebar), sehingga bisa dibilang perempuan cenderung lebih konsisten dalam capaian akademik.
2. Mahasiswa Laki-laki (oranye):
 - Distribusi relatif lebih menyebar dibanding perempuan.
 - Ada puncak kecil di 3.00 - 3.49, tapi jumlah yang mencapai 3.50 - 4.00 lebih sedikit dibanding perempuan.
 - Ada sebagian mahasiswa laki-laki dengan IPK sangat rendah (< 2.0), yang tidak tampak pada kelompok perempuan.
3. Perbandingan Kedua Gender:
 - Perempuan: lebih terkonsentrasi pada IPK tinggi (3.5 - 4.0).
 - Laki-laki: lebih variatif, ada yang tinggi tapi juga ada yang sangat rendah.
 - Secara umum, grafik menunjukkan perempuan memiliki distribusi CGPA yang lebih baik dan stabil dibanding laki-laki.

5. Perbandingan beberapa kelompok variabel

- **Var Independen Gender & Anxiety**

	sum_sq	df	F	PR(>F)
Intercept	574.157929	1.0	1734.397105	1.077665e-63
C(Gender)	1.068755	1.0	3.228460	7.548109e-02
C(Anxiety)	0.436035	1.0	1.317160	2.539251e-01
C(Gender):C(Anxiety)	0.225191	1.0	0.680249	4.115245e-01
Residual	32.111054	97.0	NaN	NaN

Gender memiliki p-value 0.075

Anxiety memiliki p-value 0.254

Interaksi keduanya memiliki p-value 0.412

- **Var Independen Depression & Anxiety**

	sum_sq	df	F	PR(>F)
Intercept	533.827812	1.0	1567.278958	1.119866e-61
C(Depression)	0.057301	1.0	0.168230	6.825941e-01
C(Anxiety)	1.041482	1.0	3.057714	8.351820e-02
C(Depression):C(Anxiety)	0.176032	1.0	0.516818	4.739303e-01
Residual	33.038980	97.0	NaN	NaN

Depression memiliki p-value 0.682

Anxiety memiliki p-value 0.084

Interaksi keduanya memiliki p-value 0.473

- **Var Independen Depression & Gender**

	sum_sq	df	F	PR(>F)
Intercept	535.199544	1.0	1556.285972	1.544697e-61
C(Depression)	0.001358	1.0	0.003948	9.500285e-01
C(Gender)	0.796087	1.0	2.314911	1.313919e-01
C(Depression):C(Gender)	0.100862	1.0	0.293292	5.893606e-01
Residual	33.357851	97.0	NaN	NaN

Depression memiliki p-value 0.95

Gender memiliki p-value 0.131

Interaksi keduanya memiliki p-value 0.589

- **Var Independen Depression & Treatment**

	sum_sq	df	F	PR(>F)
C(Depression)	5.984727e-02	1.0	1.716133e-01	0.679587
C(Treatment)	-2.792735e-15	1.0	-8.008226e-15	1.000000
C(Depression):C(Treatment)	4.629901e-03	1.0	1.327634e-02	0.908504
Residual	3.417586e+01	98.0	NaN	NaN

Depression memiliki p-value 0.680

Treatment memiliki p-value 1

Interaksi keduanya memiliki p-value 0.908

Sehingga dapat disimpulkan bahwa Gender x Anxiety dan Depression x Anxiety merupakan kombinasi variabel yang menarik untuk dianalisa Two-way ANOVA akibat pengaruh salah satu variabelnya yang hampir signifikan.

6. Kesimpulan

(kesimpulan step-by-step)

Dataset mental health mahasiswa berhasil dimuat dan dieksplorasi, di mana Variabel dependen yang digunakan adalah CGPA atau IPK dalam bentuk rentang, sementara variabel independen utama adalah Gender dan Anxiety dalam bentuk kategori. Dikarenakan Two-way ANOVA memerlukan variabel dependen dalam bentuk numerik dan independen dalam bentuk bukan kategorikal, maka dilakukan normalisasi data dengan menghapus spasi berlebih dan preprocessing dengan mapping variabel, yaitu:

- CGPA (rentang) : nilai tengah (numerik).
- Gender : binary (0 = perempuan, 1 = pria).
- Anxiety : binary (0 = cemas, 1 = tidak cemas).

Dalam proses EDA, statistik deskriptif menunjukkan distribusi CGPA cenderung normal ke kanan (positively skewed). Selain itu, variabel Gender dan Anxiety memiliki kecenderungan berbeda pada nilai rata-rata CGPA, tetapi selisihnya relatif kecil. Setelah dilakukan uji asumsi, hasil menunjukkan bahwa distribusi data tidak normal. Oleh karena itu dilakukan transformasi data sqrt dan log untuk membantu distribusi. Hasilnya, transformasi membuat distribusi CGPA lebih baik, namun tidak membuat hasil menjadi signifikan.

Setelah dilakukan Two-way ANOVA, diketahui bahwa faktor Gender dan Anxiety, serta interaksi keduanya tidak berpengaruh signifikan terhadap CGPA dikarenakan p-value dari semua uji > 0.05 , sehingga H_0 (tidak ada perbedaan rata-rata) diterima.

Setelah Two-way ANOVA, dilakukan Post Hoc Test menggunakan Tukey HSD. Pada kasus ini saya mencoba menggunakan data asli, dan data transformasi sqrt dan log untuk melihat mana yang lebih mendekati signifikan. ternyata CGPA hasil transformasi sqrt lebih mendekati signifikansi. Hasil Post hoc dengan data transformasi sqrt menunjukkan bahwa tidak ada perbedaan signifikan antar kelompok (Gender maupun Anxiety). Namun terlihat pola bahwa mahasiswa dengan Anxiety cenderung memiliki CGPA sedikit lebih rendah, dan perbedaan antara gender juga terlihat kecil meskipun secara statistik tidak signifikan.

(kesimpulan akhir)

- Gender dan Anxiety tidak memberikan pengaruh signifikan terhadap CGPA mahasiswa dalam dataset ini.
- Transformasi data memperbaiki distribusi, tetapi tidak mengubah kesimpulan utama.
- Meskipun tidak signifikan secara statistik, terdapat kecenderungan pola bahwa mahasiswa dengan anxiety memiliki CGPA lebih rendah, sehingga insight ini masih relevan untuk penelitian lebih lanjut dengan data yang lebih besar atau metode analisis lain.
- Pengaruh ukuran data mempengaruhi signifikansi (pada kasus ini, hanya terdapat 100 baris data, di mana responden female/male tidak seimbang)

DAFTAR ISI

- Kaggle. (2022). *Student Mental Health*. Kaggle. <https://www.kaggle.com/datasets/shariful07/student-mental-health?resource=download>
- Kementerian Kesehatan Republik Indonesia. (2025, January 13). *Jenis gangguan mental*. Ayo Sehat. <https://ayosehat.kemkes.go.id/jenis-gangguan-mental>
- Rangga020716. (2023). *Analisis data student mental health* [RPods]. RPods. <https://rpubs.com/Rangga020716/1046744>
- Halodoc. (2021, July 24). *Ini alasan wanita lebih rentan alami gangguan kecemasan*. Halodoc. <https://www.halodoc.com/artikel/ini-alasan-wanita-lebih-rentan-alami-gangguan-kecemasan>
- Majumder, S. (2025, March 14). *Mastering ANOVA: A Data Scientist's Guide to Comparing Group Means with Python — Part I*. Medium. <https://medium.com/@sohom.majumder21/mastering-anova-a-data-scientists-guide-to-comparing-group-means-with-python-one-way-anova-feddc20fdf9b>