Tugas Besar Probabilitas dan Statistik 2023

Nama : Aththariq Lisan Q. D. S.

NIM : 18222013

Kelas : K-01 Sistem dan Teknologi Informasi

IMPORT

Import Libraries

In [65]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as st
import math
import seaborn as sbs
```

Import Dataset

In [10]:
```
from google.colab import files

uploaded = files.upload()
for name, data in uploaded.items():
    with open(name, 'wb') as f:
        f.write(data)
!ls
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving 18222013.xlsx to 18222013.xlsx
18222013.xlsx  sample_data
```

In [117…
```
df = pd.read_excel("18222013.xlsx")
df
```

Out[117…

	Jenis Kelamin	Usia	Pendidikan Terakhir	Pekerjaan	Penghasilan per Bulan	Domisili	Durasi Penggunaan Internet per Hari (dalam Jam)	Aktivitas Online Meningkat	Aktivitas yang Meningkat dalam 3 Bulan Terakhir	layanan_aktif_1	...	cara_pembayaran_belanja_online_5	cara_pembayaran_belanja_online_6	cara_pembayaran_belanja_online_7	cara_pembayaran_belanja_online_8	keluhan_belanja_online_1	keluhan_belanja_online_2	keluhan_belanja_online_3	keluhan_belanja_online_4	keluhan_belanja_online_5	keluhan_belanja_online_6
0	Pria	19	S1	Pelajar / Mahasiswa	< Rp 2 juta	Depok	NaN	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	Jumlah barang yang diterima kurang	NaN	NaN
1	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	800%	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	Barang rusak/ salah tetapi tidak dapat dikemba...	Pembayaran sudah dilakukan; barang tidak tersedia	Jumlah barang yang diterima kurang	NaN	NaN
2	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	1000%	Sama saja	NaN	Mobile Banking	...	Transfer via ATM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Jakarta	700%	Ya	Akses media sosial	Mobile Banking	...	Transfer via ATM	Kartu Kredit / Debit Online	Melalui minimarket	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	Pembayaran sudah dilakukan; barang tidak tersedia	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	Saldo eMoney/ eWallet berkurang tanpa melakuka...
4	Wanita	20	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Medan	700%	Ya	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
280	Pria	21	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	800%	Ya	Streaming video/ film	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
281	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Bogor	1400%	Sama saja	NaN	Mobile Banking	...	NaN	Kartu Kredit / Debit Online	NaN	NaN	NaN	Barang rusak/ salah tetapi tidak dapat dikemba...	NaN	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	NaN
282	Pria	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	padang	1500%	Sama saja	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
283	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Bandung	NaN	Tidak	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	Jumlah barang yang diterima kurang	NaN	NaN
284	Pria	51	S1	Pengusaha	Rp 5 juta – Rp 10 juta	Semarang	800%	Sama saja	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

285 rows × 150 columns

In [67]:
```
print("Informasi Dataset:")
df.info()
```

```
Informasi Dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 285 entries, 0 to 284
Columns: 150 entries, Jenis Kelamin to keluhan_belanja_online_6
dtypes: float64(1), int64(1), object(148)
memory usage: 334.1+ KB
```

SOAL 1

Dimensi Data Set

In [68]:
```
print("Dimensi dataset (baris, kolom):", df.shape)
```

```
Dimensi dataset (baris, kolom): (285, 150)
```

Jumlah missing value per kolom

In [69]:
```
missing_values = df.isnull().sum()
print("\nJumlah missing value per kolom:")
print(missing_values)
```

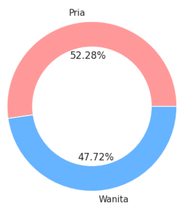
```
Jumlah missing value per kolom:
Jenis Kelamin                0
Usia                         0
Pendidikan Terakhir          0
Pekerjaan                    0
Penghasilan per Bulan        0
                          ...
keluhan_belanja_online_2   215
keluhan_belanja_online_3   209
keluhan_belanja_online_4   242
keluhan_belanja_online_5   253
keluhan_belanja_online_6   270
Length: 150, dtype: int64
```

SOAL 2: Visualisasi

Donutchart Perbandingan Proporsi Jenis Kelamin Responden

In [70]:
```
# Generate Dataframe
gender_freq = pd.DataFrame(df["Jenis Kelamin"].value_counts())

label = df["Jenis Kelamin"].unique()
label = label[~pd.isnull(label)]
# Plotting Dataframe
plt.pie(gender_freq.values.flatten(), labels=label, autopct='%.2f%%', colors=['#ff9999','#66b3ff'])

## Change Pie Chart to Donut Chart
my_circle = plt.Circle((0, 0), 0.7, color='white')
p = plt.gcf()
p.gca().add_artist(my_circle)

plt.title("Grafik 1: Proporsi Jenis Kelamin Responden", fontsize=9, loc="left", fontweight="bold")

# Menambahkan Insight di Bawah Visualisasi
plt.text(0, -1.3, "\nInsight: Proporsi responden pria lebih tinggi dari wanita", ha='center', fontsize=10, fontweight='bold')

plt.show()
```

Grafik 1: Proporsi Jenis Kelamin Responden



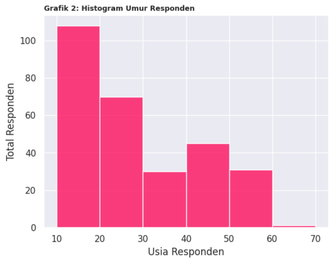
Insight: Proporsi responden pria lebih tinggi dari wanita

Histogram Sebaran Umur Responden

In [71]:
```
sbs.set(color_codes=True)
sbs.set(style="darkgrid", palette="muted")

histogram = sbs.histplot(data=df, x="Usia", bins=[10,20,30,40,50,60,70], color="#ff0054")
histogram.set(xlabel='Usia Responden', ylabel='Total Responden')
plt.title("Grafik 2: Histogram Umur Responden", fontsize=9, loc="left", fontweight="bold")

# Menambahkan Insight di Bawah Visualisasi
plt.text(35, -28, "Insight: Mayoritas responden berada di rentang usia 10-20 tahun", ha='center', fontsize=10, fontweight='bold')
```

```
plt.show()
```

Grafik 2: Histogram Umur Responden



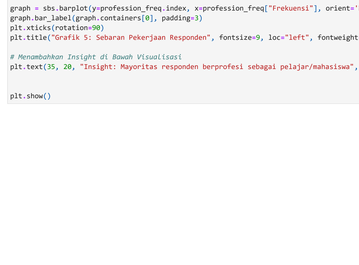
Insight: Mayoritas responden berada di rentang usia 10-20 tahun

Barchart Sebaran Pendidikan Terakhir Responden

In [72]:
```
# Cleaning baris yang memiliki nilai '0 Level' pada kolom 'Pendidikan Terakhir'
df = df[df['Pendidikan Terakhir'] != 'O LEVEL']
```

In [73]:
```
profession_freq = pd.DataFrame(df["Pendidikan Terakhir"].value_counts())
profession_freq.rename(columns={"Pendidikan Terakhir":"Frekuensi"}, inplace=True)
profession_freq.rename_axis("Pendidikan Terakhir", inplace=True)

graph = sbs.barplot(y=profession_freq.index, x=profession_freq["Frekuensi"], orient='h', palette="muted")
graph.bar_label(graph.containers[0], padding=3)
plt.xticks(rotation=90)
plt.title("Grafik 3: Perbandingan Frekuensi/Proporsi Pendidikan Terakhir Responden", fontsize=9, loc="left", fontweight="bold")

# Menambahkan Insight di Bawah Visualisasi
plt.text(80, 9.6, "Insight: Mayoritas responden menempuh pendidikan Terakhir SMA", ha='center', fontsize=10, fontweight='bold')

plt.show()
```

Grafik 3: Perbandingan Frekuensi/Proporsi Pendidikan Terakhir Responden

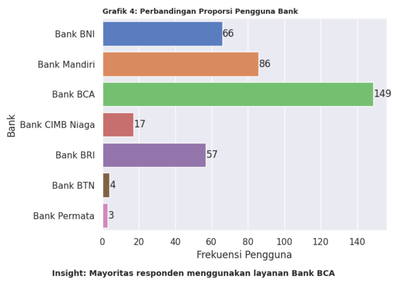


Insight: Mayoritas responden menempuh pendidikan Terakhir SMA

Barchart Perbandingan Proporsi Pengguna Bank (Kolom bank_1 sampai bank_7)

In [74]:
```
# Generate Dataframe
bank_1 = pd.DataFrame(df["bank_1"].value_counts())
bank_2 = pd.DataFrame(df["bank_2"].value_counts())
bank_3 = pd.DataFrame(df["bank_3"].value_counts())
bank_4 = pd.DataFrame(df["bank_4"].value_counts())
bank_5 = pd.DataFrame(df["bank_5"].value_counts())
bank_6 = pd.DataFrame(df["bank_6"].value_counts())
bank_7 = pd.DataFrame(df["bank_7"].value_counts())

bank_val = np.array(['bank_1.index[0]',bank_2.index[0],bank_3.index[0],bank_4.index[0],bank_5.index[0],bank_6.index[0],bank_7.index[0]])
bank_freq = np.array([bank_1["bank_1"][0],bank_2["bank_2"][0],bank_3["bank_3"][0],bank_4["bank_4"][0],bank_5["bank_5"][0],bank_6["bank_6"][0],bank_7["bank_7"][0]])
bank_data = {'Bank' : bank_val, "Frekuensi Pengguna" : bank_freq}
df_bank = pd.DataFrame(data=bank_data)

graph = sbs.barplot(y=df_bank['Bank'], x=df_bank["Frekuensi Pengguna"], orient="h",  palette="muted")
graph.bar_label(graph.containers[0])
plt.title("Grafik 4: Perbandingan Proporsi Pengguna Bank", fontsize=9, loc="left", fontweight="bold")

# Menambahkan Insight di Bawah Visualisasi
plt.text(50, 8, "Insight: Mayoritas responden menggunakan layanan Bank BCA", ha='center', fontsize=10, fontweight='bold')

plt.show()
```

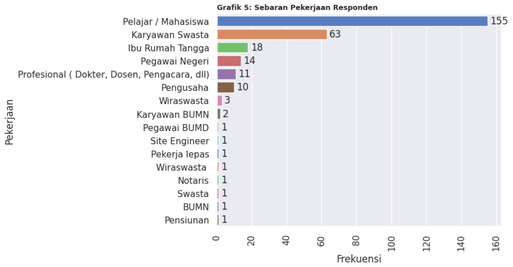
Grafik 4: Perbandingan Proporsi Pengguna Bank



Insight: Mayoritas responden menggunakan layanan Bank BCA

Bar chart sebaran Profesi responden.

In [75]:
```
profession_freq = pd.DataFrame(df["Pekerjaan"].value_counts())
profession_freq.rename(columns={"Pekerjaan":"Frekuensi"}, inplace=True)
profession_freq.rename_axis("Pekerjaan",inplace=True)

graph = sbs.barplot(y=profession_freq.index, x=profession_freq["Frekuensi"], orient='h', palette="muted")
graph.bar_label(graph.containers[0], padding=3)
plt.xticks(rotation=90)
plt.title("Grafik 5: Sebaran Pekerjaan Responden", fontsize=9, loc="left", fontweight="bold")

# Menambahkan Insight di Bawah Visualisasi
plt.text(35, 20, "Insight: Mayoritas responden berprofesi sebagai pelajar/mahasiswa", ha='center', fontsize=10, fontweight='bold')


plt.show()
```



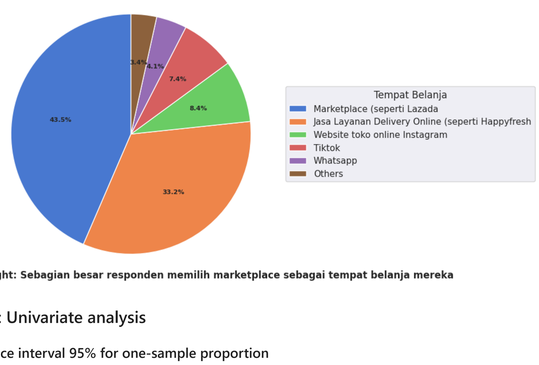
Insight: Mayoritas responden berprofesi sebagai pelajar/mahasiswa

Pie Chart Proporsi Preferensi Tempat Belanja Online

In [116…

```
tempat_belanja_counts = df[['tempat_belanja_online_1', 'tempat_belanja_online_2', 'tempat_belanja_online_3', 'tempat_belanja_online_4', 'tempat_belanja_online_5', 'tempat_belanja_online_6', 'tempat_belanja_online_7', 'tempat_belanja_online_8']].stack().value_counts()

threshold = 25

small_parts = tempat_belanja_counts[tempat_belanja_counts < threshold]
large_parts = tempat_belanja_counts[tempat_belanja_counts >= threshold]

other_parts = small_parts.sum()
large_parts['Others'] = other_parts

plt.figure(figsize=(10, 10))
wedges, texts, autotexts = plt.pie(large_parts, autopct='%1.1f%%', startangle=90)

plt.legend(wedges, large_parts.index, title="Tempat Belanja", loc="center left", bbox_to_anchor=(1, 0, 0.5, 1))
plt.setp(autotexts, size=8, weight="bold")

plt.title("Grafik 6: Preferensi Tempat Belanja Online", fontsize=20, loc="center", fontweight="bold")
plt.tight_layout()

# Menambahkan Insight di Bawah Visual(sasa
plt.text(0.7, -1.2, "Insight: Sebagian besar responden memilih marketplace sebagai tempat belanja mereka", ha='center', fontsize=12, fontweight='bold')

plt.show()
```

Grafik 6: Preferensi Tempat Belanja Online



Insight: Sebagian besar responden memilih marketplace sebagai tempat belanja mereka

SOAL 3: Univariate analysis

Confidence interval 95% for one-sample proportion

In [77]:

```
bca_data = df_bank[df_bank["Bank"] == "Bank BCA"]["Frekuensi Pengguna"].sum()
total_bank = df_bank["Frekuensi Pengguna"].sum()
prop_bca = bca_data/total_bank

alpha = 0.05
z_alphaper2 = st.norm.ppf(1-alpha/2)

lower_bound = prop_bca - z_alphaper2*np.sqrt(prop_bca*(1-prop_bca)/total_bank)
upper_bound = prop_bca + z_alphaper2*np.sqrt(prop_bca*(1-prop_bca)/total_bank)

print("Confidence Interval dari proporsi Bank BCA adalah {:.4f} < Proporsi Bank BCA < {:.4f}".format(lower_bound, upper_bound))
```

```
Confidence Interval dari proporsi Bank BCA adalah 0.3411 < Proporsi Bank BCA < 0.4390
```

In [78]:

```
emoney_1 = pd.DataFrame(df["emoney_pilihan_1"].value_counts())
emoney_2 = pd.DataFrame(df["emoney_pilihan_2"].value_counts())
emoney_3 = pd.DataFrame(df["emoney_pilihan_3"].value_counts())
emoney_4 = pd.DataFrame(df["emoney_pilihan_4"].value_counts())
emoney_5 = pd.DataFrame(df["emoney_pilihan_5"].value_counts())
emoney_6 = pd.DataFrame(df["emoney_pilihan_6"].value_counts())
emoney_7 = pd.DataFrame(df["emoney_pilihan_7"].value_counts())

emoney_val = np.array([emoney_1.index[0],emoney_2.index[0],emoney_3.index[0],emoney_4.index[0],emoney_5.index[0],emoney_6.index[0],emoney_7.index[0]])
emoney_freq = np.array([emoney_1["emoney_pilihan_1"][0],emoney_2["emoney_pilihan_2"][0],emoney_3["emoney_pilihan_3"][0],emoney_4["emoney_pilihan_4"][0],emoney_5["emoney_pilihan_5"][0],emoney_6["emoney_pilihan_6"][0],emoney_7["emoney_pilihan_7"][0]])
emoney_data = {'E-money' : emoney_val, "Frekuensi Pengguna" : emoney_freq}
df_emoney = pd.DataFrame(data=emoney_data)

total_emoney = df_emoney["Frekuensi Pengguna"].sum()

prop_gopay = df_emoney[df_emoney["E-money"] == "GoPay"]["Frekuensi Pengguna"].sum()/total_emoney

alpha = 0.05
z_alphaper2 = st.norm.ppf(1 - alpha / 2)

lower_bound = prop_gopay - z_alphaper2 * np.sqrt(prop_gopay * (1 - prop_gopay) / total_emoney)
upper_bound = prop_gopay + z_alphaper2 * np.sqrt(prop_gopay * (1 - prop_gopay) / total_emoney)

print("Confidence Interval dari proporsi GoPay adalah {:.4f} < Proporsi GoPay < {:.4f}".format(lower_bound, upper_bound))
```

```
Confidence Interval dari proporsi GoPay adalah 0.2196 < Proporsi GoPay < 0.2780
```

In [79]:

```
alpha = 0.05

ecommerce_pilihan_1 = pd.DataFrame(df["ecommerce_pilihan_1"].value_counts())
ecommerce_pilihan_2 = pd.DataFrame(df["ecommerce_pilihan_2"].value_counts())
ecommerce_pilihan_3 = pd.DataFrame(df["ecommerce_pilihan_3"].value_counts())
ecommerce_pilihan_4 = pd.DataFrame(df["ecommerce_pilihan_4"].value_counts())
ecommerce_pilihan_5 = pd.DataFrame(df["ecommerce_pilihan_5"].value_counts())
ecommerce_pilihan_6 = pd.DataFrame(df["ecommerce_pilihan_6"].value_counts())
ecommerce_pilihan_7 = pd.DataFrame(df["ecommerce_pilihan_7"].value_counts())
ecommerce_pilihan_8 = pd.DataFrame(df["ecommerce_pilihan_8"].value_counts())
ecommerce_pilihan_9 = pd.DataFrame(df["ecommerce_pilihan_9"].value_counts())
ecommerce_pilihan_10 = pd.DataFrame(df["ecommerce_pilihan_10"].value_counts())
ecommerce_pilihan_11 = pd.DataFrame(df["ecommerce_pilihan_11"].value_counts())
ecommerce_pilihan_12 = pd.DataFrame(df["ecommerce_pilihan_12"].value_counts())

ecommerce_val = np.array([ecommerce_pilihan_1.index[0],ecommerce_pilihan_2.index[0],ecommerce_pilihan_3.index[0],ecommerce_pilihan_4.index[0],ecommerce_pilihan_5.index[0],ecommerce_pilihan_6.index[0],ecommerce_pilihan_7.index[0],ecommerce_pilihan_8.index[0],ecommerce_pilihan_9.index[0],ecommerce_pilihan_10.index[0],ecommerce_pilihan_11.index[0]])
ecommerce_freq = np.array([ecommerce_pilihan_1["ecommerce_pilihan_1"][0],ecommerce_pilihan_2["ecommerce_pilihan_2"][0],ecommerce_pilihan_3["ecommerce_pilihan_3"][0],ecommerce_pilihan_4["ecommerce_pilihan_4"][0],ecommerce_pilihan_5["ecommerce_pilihan_5"][0],ecommerce_pilihan_6["ecommerce_pilihan_6"][0],ecommerce_pilihan_7["ecommerce_pilihan_7"][0],ecommerce_pilihan_8["ecommerce_pilihan_8"][0],ecommerce_pilihan_9["ecommerce_pilihan_9"][0],ecommerce_pilihan_10["ecommerce_pilihan_10"][0],ecommerce_pilihan_11["ecommerce_pilihan_11"][0]])
ecommerce_data = {'ecommerce' : ecommerce_val, "Frekuensi Pengguna" : ecommerce_freq}
df_ecommerce = pd.DataFrame(data=ecommerce_data)

total_ecommerce = df_ecommerce["Frekuensi Pengguna"].sum()

prop_shopee = df_ecommerce[df_ecommerce["ecommerce"] == "Shopee"]["Frekuensi Pengguna"].sum()/total_ecommerce

alpha = 0.05
z_alphaper2 = st.norm.ppf(1 - alpha / 2)

lower_bound = prop_shopee - z_alphaper2 * np.sqrt(prop_shopee * (1 - prop_shopee) / total_ecommerce)
upper_bound = prop_shopee + z_alphaper2 * np.sqrt(prop_shopee * (1 - prop_shopee) / total_ecommerce)

print("Confidence Interval dari proporsi Shopee adalah {:.4f} < Proporsi Shopee < {:.4f}".format(lower_bound, upper_bound))
```

```
Confidence Interval dari proporsi Shopee adalah 0.2320 < Proporsi Shopee < 0.2907
```

In [80]:

```
#cek isi
print(df["ecommerce_pilihan_1"].value_counts())
print(df["ecommerce_pilihan_2"].value_counts())
print(df["ecommerce_pilihan_3"].value_counts())
print(df["ecommerce_pilihan_4"].value_counts())
print(df["ecommerce_pilihan_5"].value_counts())
```

```
print(df["ecommerce_pilihan_6"].value_counts())
print(df["ecommerce_pilihan_7"].value_counts())
print(df["ecommerce_pilihan_8"].value_counts())
print(df["ecommerce_pilihan_9"].value_counts())
print(df["ecommerce_pilihan_10"].value_counts())
print(df["ecommerce_pilihan_11"].value_counts())
print(df["ecommerce_pilihan_12"].value_counts())
```

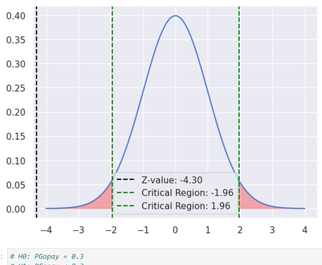
```
Shopee    225
Name: ecommerce_pilihan_1, dtype: int64
Gojek (GoFood    208
Name: ecommerce_pilihan_2, dtype: int64
Tokopedia    188
Name: ecommerce_pilihan_3, dtype: int64
Grab (GrabFood    89
Name: ecommerce_pilihan_4, dtype: int64
Traveloka    56
Name: ecommerce_pilihan_5, dtype: int64
Lazada    30
Name: ecommerce_pilihan_6, dtype: int64
Tiket.com    31
Name: ecommerce_pilihan_7, dtype: int64
Bukalapak    19
Name: ecommerce_pilihan_8, dtype: int64
Blibli    5
Name: ecommerce_pilihan_9, dtype: int64
JD.id    9
Name: ecommerce_pilihan_10, dtype: int64
Zalora    1
Name: ecommerce_pilihan_11, dtype: int64
Series([], Name: ecommerce_pilihan_12, dtype: int64)
```

One sample Z-test for proportion dengan alpha = 5%

In [81]:
```
# H0: pbca = 0.5
# H1: pbca ≠ 0.5
alpha = 0.05

pBCA0 = 0.5
z =(total_bank*prop_bca - total_bank*pBCA0)/np.sqrt(total_bank*pBCA0*(1-pBCA0))
z_alphaper2 = st.norm.ppf(1-alpha/2)

print("Critical Region: Z < -{:.2f} | Z > {:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2 or z > z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.8f}".format(pvalue))
```

```
Critical Region: Z < -1.96 | Z > 1.96
Z-value: -4.2978
Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak
P value : 0.00001725
```

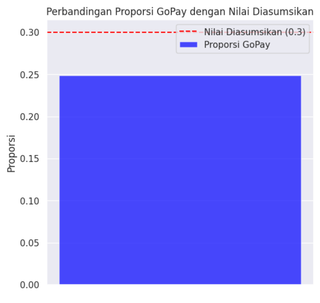
In [82]:
```
x = np.linspace(-4, 4, 1000)
y = st.norm.pdf(x, 0, 1)
plt.plot(x, y)

plt.fill_between(x, 0, y, where=(x < -z_alphaper2) | (x > z_alphaper2), color='red', alpha=0.3)
plt.axvline(z, color='black', linestyle='--', label=f'Z-value: {z:.2f}')
plt.axvline(-z_alphaper2, color='green', linestyle='--', label=f'Critical Region: -{z_alphaper2:.2f}')
plt.axvline(z_alphaper2, color='green', linestyle='--', label=f'Critical Region: {z_alphaper2:.2f}')

plt.legend()
plt.show()
```



In [83]:
```
# H0: PGopay = 0.3
# H1: PGopay ≠ 0.3
alpha = 0.05

pGopay0 = 0.3
z =(total_emoney*prop_gopay - total_emoney*pGopay0)/np.sqrt(total_emoney*pGopay0*(1-pGopay0))
z_alphaper2 = st.norm.ppf(1-alpha/2)

print("Critical Region: Z < -{:.2f} | Z > {:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2 or z > z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.8f}".format(pvalue))
```

```
Critical Region: Z < -1.96 | Z > 1.96
Z-value: -3.2449
Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak
P value : 0.00117487
```

In [84]:
```
import matplotlib.pyplot as plt

# Proporsi GoPay
proporsi_GoPay = 0.2488
nilai_diasumsikan = 0.3


plt.figure(figsize=(6, 6))
plt.bar(0, proporsi_GoPay, color='blue', alpha=0.7, label='Proporsi GoPay')
plt.axhline(y=nilai_diasumsikan, color='red', linestyle='--', label='Nilai Diasumsikan (0.3)')

plt.ylabel('Proporsi')
plt.title('Perbandingan Proporsi GoPay dengan Nilai Diasumsikan')
plt.xticks([])
plt.legend()
plt.show()
```

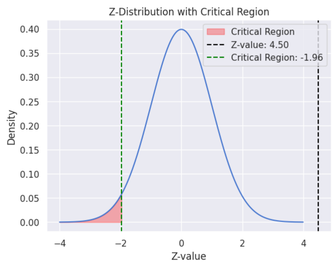


In [85]:
```
# H0: Shopee = 0.2
# H1: Shopee < 0.2
alpha = 0.05

pShopee0 = 0.2
z =(total_ecommerce*prop_shopee - total_ecommerce*pShopee0)/np.sqrt(total_ecommerce*pShopee0*(1-pShopee0))
z_alphaper2 = st.norm.ppf(1-alpha/2)

print("Critical Region: Z < -{:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
```

```
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.8f}".format(pvalue))
```

```
Critical Region: Z < -1.96
Z-value: 4.4985
Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak
P value : 0.00000684
```

In [86]:
```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as st

alpha = 0.05
pShopee0 = 0.2
z = 4.4985

x = np.linspace(-4, 4, 1000)
y = st.norm.pdf(x, 0, 1)

plt.plot(x, y)
plt.fill_between(x, 0, y, where=(x < -1.96), color='red', alpha=0.3, label='Critical Region')
plt.axvline(z, color='black', linestyle='--', label=f'Z-value: {z:.2f}')
plt.axvline(-1.96, color='green', linestyle='--', label='Critical Region: -1.96')

plt.legend()
plt.title('Z-Distribution with Critical Region')
plt.xlabel('Z-value')
plt.ylabel('Density')
plt.show()
```



Chi-squared test for goodness-of-fit test dengan alpha = 5%

In [87]:
```
#H0: Distribusi Pendidikan Terakhir = uniform distribution
#H1: Distribusi Pendidikan Terakhir ≠ uniform distribution
alpha = 0.05

pend_akhir = df['Pendidikan Terakhir'].value_counts().reset_index()
pend_akhir.columns = ['Pendidikan Terakhir', 'Oi']
pend_akhir['Ei'] = 1/8 * len(df)
pend_akhir['(Oi - Ei)^2 / Ei'] = pend_akhir.apply(lambda x: (x['Oi'] - x['Ei'])**2 / x['Ei'], axis = 1)

pend_akhir
```

Out[87]:

	Pendidikan Terakhir	Oi	Ei	(Oi - Ei)^2 / Ei
0	SMA	149	35.5	362.880282
1	S1	90	35.5	83.669014
2	S2	23	35.5	4.401408
3	D3	15	35.5	11.838028
4	D4	3	35.5	29.753521
5	SMK	2	35.5	31.612676
6	SMP	1	35.5	33.528169
7	S3	1	35.5	33.528169

In [88]:
```
# test statistic (chi2)
chi2 = pend_akhir['(Oi - Ei)^2 / Ei'].sum()

# crit region
alpha = 0.05
df = len(pend_akhir) - 1
chi2_alpha = st.chi2.ppf(1 - 0.05, df)

# kesimpulan
if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0,\nDistribusi Pendidikan Terakhir tidak mengikuti distribusi seragam'

#p-value
pval = (1 - st.chi2.cdf(chi2, df))

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

```
Hasil chi2 test:
chi2: 591.21
crit region: chi2 > 14.07
kesimpulan: chi2 di dalam crit region, reject H0,
Distribusi Pendidikan Terakhir tidak mengikuti distribusi seragam
p-value: 0.00
```

In [118…
```
#H0: Distribusi Jenis Kelamin = uniform distribution
#H1: Distribusi Jenis Kelamin ≠ uniform distribution
alpha = 0.05

jeniskelamin = df['Jenis Kelamin'].value_counts().reset_index()
jeniskelamin.columns = ['Jenis Kelamin', 'Oi']
jeniskelamin['Ei'] = 1/2 * len(df)
jeniskelamin['(Oi - Ei)^2 / Ei'] =  jeniskelamin.apply(lambda x: (x['Oi'] - x['Ei'])**2 / x['Ei'], axis = 1)

jeniskelamin
```

Out[118…

	Jenis Kelamin	Oi	Ei	(Oi - Ei)^2 / Ei
0	Wanita	149	142.5	0.296491
1	Pria	136	142.5	0.296491

In [119…
```
# test statistic (chi2)
chi2 = jeniskelamin['(Oi - Ei)^2 / Ei'].sum()

# crit region
alpha = 0.05
df = len(jeniskelamin) - 1
chi2_alpha = st.chi2.ppf(1 - 0.05, df)

# kesimpulan
if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0, \nDistribusi Jenis Kelamin dapat dianggap sebagai distribusi seragam (uniform).'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0'

#p-value
pval = (1 - st.chi2.cdf(chi2, df))

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

```
Hasil chi2 test:
chi2: 0.59
crit region: chi2 > 3.84
kesimpulan: chi2 di luar crit region, fail to reject H0,
Distribusi Jenis Kelamin dapat dianggap sebagai distribusi seragam (uniform).
p-value: 0.44
```

SOAL 4: Bivariate Analysis

Confidence interval 95% for two-sample proportion difference

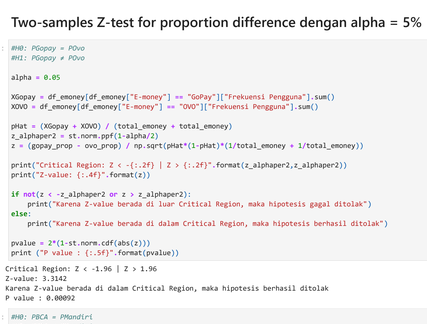
In [ ]:
```
alpha = 0.05
total_emoney = df_emoney["Frekuensi Pengguna"].sum()
gopay_prop = df_emoney[df_emoney["E-money"] == "GoPay"]["Frekuensi Pengguna"].sum()/total_emoney
ovo_prop = df_emoney[df_emoney["E-money"] == "OVO"]["Frekuensi Pengguna"].sum()/total_emoney
difference = gopay_prop - ovo_prop
z_alphaPer2 = st.norm.ppf(1-alpha/2)
lower_bound = difference - z_alphaPer2 * np.sqrt(gopay_prop*(1-gopay_prop)/total_emoney + ovo_prop*(1-ovo_prop)/total_emoney)
upper_bound = difference + z_alphaPer2 * np.sqrt(gopay_prop*(1-gopay_prop)/total_emoney + ovo_prop*(1-ovo_prop)/total_emoney)
print ("Confidence Interval Selisih Proporsi penggunaan Gopay dan OVO:")
print("{:.5f} < P Gopay - P OVO  < {:.5f}".format(lower_bound,upper_bound))
```

```
Confidence Interval Selisih Proporsi penggunaan Gopay dan OVO:
0.02724 < P Gopay - P OVO  < 0.10546
```

In [43]:
```
alpha = 0.05
total_bank = df_bank["Frekuensi Pengguna"].sum()
bca_prop = df_bank[df_bank["Bank"] == "Bank BCA"]["Frekuensi Pengguna"].sum()/total_bank
mandiri_prop = df_bank[df_bank["Bank"] == "Bank Mandiri"]["Frekuensi Pengguna"].sum()/total_bank
difference = bca_prop - mandiri_prop
z_alphaPer2 = st.norm.ppf(1-alpha/2)
lower_bound = difference - z_alphaPer2 * np.sqrt(bca_prop*(1-bca_prop)/total_bank + mandiri_prop*(1-mandiri_prop)/total_bank)
upper_bound = difference + z_alphaPer2 * np.sqrt(bca_prop*(1-bca_prop)/total_bank + mandiri_prop*(1-mandiri_prop)/total_bank)
print ("Confidence Interval Selisih Proporsi penggunaan BCA dan Bank Mandiri:")
print("{:.5f} < P BCA - P Bank Mandiri  < {:.5f}".format(lower_bound,upper_bound))
```

```
Confidence Interval Selisih Proporsi penggunaan BCA dan Bank Mandiri:
0.10053 < P BCA - P Bank Mandiri  < 0.22932
```

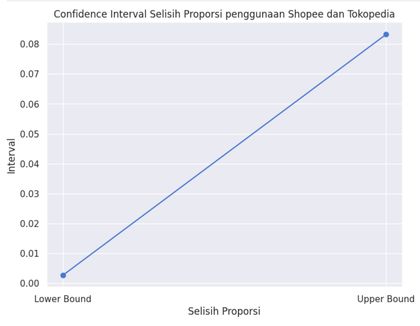
In [44]:
```
import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))

plt.plot([0, 1], [lower_bound, upper_bound], marker='o', linestyle='-')
plt.title('Confidence Interval Selisih Proporsi penggunaan BCA dan Bank Mandiri')
plt.xlabel('Selisih Proporsi')
plt.ylabel('Interval')
plt.xticks([0, 1], ['Lower Bound', 'Upper Bound'])

plt.grid(True)
plt.show()
```



In [45]:
```
alpha = 0.05
total_ecommerce = df_ecommerce["Frekuensi Pengguna"].sum()
shopee_prop = df_ecommerce[df_ecommerce["ecommerce"] == "Shopee"]["Frekuensi Pengguna"].sum()/total_ecommerce
tokopedia_prop = df_ecommerce[df_ecommerce["ecommerce"] == "Tokopedia"]["Frekuensi Pengguna"].sum()/total_ecommerce
difference = shopee_prop - tokopedia_prop
z_alphaPer2 = st.norm.ppf(1-alpha/2)
lower_bound = difference - z_alphaPer2 * np.sqrt(shopee_prop*(1-shopee_prop)/total_ecommerce + tokopedia_prop*(1-tokopedia_prop)/total_ecommerce)
upper_bound = difference + z_alphaPer2 * np.sqrt(shopee_prop*(1-shopee_prop)/total_ecommerce + tokopedia_prop*(1-tokopedia_prop)/total_ecommerce)
print ("Confidence Interval Selisih Proporsi penggunaan Shopee dan Tokopedia:")
print("{:.5f} < P Shopee - P Tokopedia < {:.5f}".format(lower_bound,upper_bound))
```

```
Confidence Interval Selisih Proporsi penggunaan Shopee dan Tokopedia:
0.00269 < P Shopee - P Tokopedia < 0.08326
```

In [46]:
```
import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))

plt.plot([0, 1], [lower_bound, upper_bound], marker='o', linestyle='-')
plt.title('Confidence Interval Selisih Proporsi penggunaan Shopee dan Tokopedia')
plt.xlabel('Selisih Proporsi')
plt.ylabel('Interval')
plt.xticks([0, 1], ['Lower Bound', 'Upper Bound'])

plt.grid(True)
plt.show()
```



Two-samples Z-test for proportion difference dengan alpha = 5%

In [ ]:
```
#H0: PGopay = POvo
#H1: PGopay ≠ POvo

alpha = 0.05

XGopay = df_emoney[df_emoney["E-money"] == "GoPay"]["Frekuensi Pengguna"].sum()
XOVO = df_emoney[df_emoney["E-money"] == "OVO"]["Frekuensi Pengguna"].sum()

pHat = (XGopay + XOVO) / (total_emoney + total_emoney)
z_alphaper2 = st.norm.ppf(1-alpha/2)
z = (gopay_prop - ovo_prop) / np.sqrt(pHat*(1-pHat)*(1/total_emoney + 1/total_emoney))

print("Critical Region: Z < -{:.2f} | Z > {:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2 or z > z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.5f}".format(pvalue))
```

```
Critical Region: Z < -1.96 | Z > 1.96
Z-value: 3.3142
Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak
P value : 0.00092
```

In [ ]:
```
#H0: PBCA = PMandiri
#H1: PBCA ≠ PMandiri

alpha = 0.05

XBCA = df_bank[df_bank["Bank"] == "Bank BCA"]["Frekuensi Pengguna"].sum()
XMandiri = df_bank[df_bank["Bank"] == "Bank Mandiri"]["Frekuensi Pengguna"].sum()

pHat = (XBCA + XMandiri) / (total_bank + total_bank)
z_alphaper2 = st.norm.ppf(1-alpha/2)
z = (bca_prop - mandiri_prop) / np.sqrt(pHat*(1-pHat)*(1/total_bank + 1/total_bank))
```

```
print("Critical Region: Z < -{:.2f} | Z > {:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2 or z > z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.5f}".format(pvalue))
```

```
Critical Region: Z < -1.96 | Z > 1.96
Z-value: 4.9388
Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak
P value : 0.00000
```

In [ ]:
```
#H0: PShopee = PTokopedia
#H1: PShopee < PTokopedia

alpha = 0.05

XShopee = df_ecommerce[df_ecommerce["ecommerce"] == "Shopee"]["Frekuensi Pengguna"].sum()
XTokopedia = df_ecommerce[df_ecommerce["ecommerce"] == "Tokopedia"]["Frekuensi Pengguna"].sum()

pHat = (XShopee + XTokopedia) / (total_ecommerce + total_ecommerce)
z_alphaper2 = st.norm.ppf(1-alpha/2)
z = (shopee_prop - tokopedia_prop) / np.sqrt(pHat*(1-pHat)*(1/total_ecommerce + 1/total_ecommerce))

print("Critical Region: Z < -{:.2f}".format(z_alphaper2,z_alphaper2))
print("Z-value: {:.4f}".format(z))

if not(z < -z_alphaper2):
    print("Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak")
else:
    print("Karena Z-value berada di dalam Critical Region, maka hipotesis berhasil ditolak")

pvalue = 2*(1-st.norm.cdf(abs(z)))
print ("P value : {:.5f}".format(pvalue))
```

```
Critical Region: Z < -1.96
Z-value: 2.0882
Karena Z-value berada di luar Critical Region, maka hipotesis gagal ditolak
P value : 0.03678
```

Chi-squared test for independence dengan alpha = 5%

In [51]:
```
# H0: Variabel Pekerjaan independen dengan penghasilan per bulan
# H1: Variabel Pekerjaan tidak independen dengan penghasilan per bulan

contigency = pd.crosstab(df['Penghasilan per Bulan'], df['Pekerjaan'])
contigency
```

Out[51]:

Pekerjaan	BUMN	Ibu Rumah Tangga	Karyawan BUMN	Karyawan Swasta	Notaris	Pegawai BUMD	Pegawai Negeri	Pekerja lepas	Pelajar / Mahasiswa	Pengusaha	Pensiunan	Profesional (Dokter, Dosen, Pengacara, dll)	Site Engineer	Swasta	Wiraswasta	Wiraswasta
Penghasilan per Bulan																
< Rp 2 juta	0	4	0	1	0	0	0	1	121	0	1	0	0	0	0	0
> Rp 10 juta	1	4	2	28	0	1	3	0	2	5	0	4	0	1	1	0
Rp 2 juta – Rp 5 juta	0	7	0	14	0	0	1	0	31	2	0	1	0	0	2	1
Rp 5 juta – Rp 10 juta	0	3	0	20	1	0	10	0	1	4	0	6	1	0	0	0

In [91]:
```
Ei = contigency.copy()

for s in contigency.index:
    for o in contigency.columns:
        Ei.loc[s, o] = contigency.loc[s].sum() * contigency.loc[:, o].sum() / contigency.values.sum()

Ei
```

Out[91]:

Pekerjaan	BUMN	Ibu Rumah Tangga	Karyawan BUMN	Karyawan Swasta	Notaris	Pegawai BUMD	Pegawai Negeri	Pekerja lepas	Pelajar / Mahasiswa	Pengusaha	Pensiunan	Profesional (Dokter, Dosen, Pengacara, dll)	Site Engineer	Swasta	Wiraswasta	Wiraswasta
Penghasilan per Bulan																
< Rp 2 juta	0.449123	8.084211	0.898246	28.294737	0.449123	0.449123	6.287719	0.449123	69.614035	4.940351	0.449123	4.940351	0.449123	0.449123	1.347368	0.449123
> Rp 10 juta	0.182456	3.284211	0.364912	11.494737	0.182456	0.182456	2.554386	0.182456	28.280702	2.007018	0.182456	2.007018	0.182456	0.182456	0.547368	0.182456
Rp 2 juta – Rp 5 juta	0.207018	3.726316	0.414035	13.042105	0.207018	0.207018	2.898246	0.207018	32.087719	2.277193	0.207018	2.277193	0.207018	0.207018	0.621053	0.207018
Rp 5 juta – Rp 10 juta	0.161404	2.905263	0.322807	10.168421	0.161404	0.161404	2.259649	0.161404	25.017544	1.775439	0.161404	1.775439	0.161404	0.161404	0.484211	0.161404

In [92]:
```
# H0: Variabel Pekerjaan independen dengan penghasilan per bulan
# H1: Variabel Pekerjaan tidak independen dengan penghasilan per bulan

chi2 = ((contigency - Ei) ** 2 / Ei).sum().sum()

alpha = 0.05
row = 4
col = 16
df = (row-1)*(col-1)

chi2_alpha = st.chi2.ppf(1-alpha, df)

pval = 1 - st.chi2.cdf(chi2, df)

if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0,'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0, \nVariabel Pekerjaan dan penghasilan per bulan memiliki keterkaitan atau hubungan yang signifikan satu sama lain'

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

```
Hasil chi2 test:
chi2: 258.53
crit region: chi2 > 61.66
kesimpulan: chi2 di dalam crit region, reject H0,
Variabel Pekerjaan dan penghasilan per bulan memiliki keterkaitan atau hubungan yang signifikan satu sama lain
p-value: 0.00
```

CleanUp Data

In [95]:
```
df['Durasi Penggunaan Internet per Hari (dalam Jam)'] = df['Durasi Penggunaan Internet per Hari (dalam Jam)'].str.rstrip('%').astype(float) / 100
df
```

Out[95]:

	Jenis Kelamin	Usia	Pendidikan Terakhir	Pekerjaan	Penghasilan per Bulan	Domisili	Durasi Penggunaan Internet per Hari (dalam Jam)	Aktivitas Online Meningkat	Aktivitas yang Meningkat dalam 3 Bulan Terakhir	layanan_aktif_1	...	cara_pembayaran_belanja_online_5	cara_pembayaran_belanja_online_6	cara_pembayaran_belanja_online_7	cara_pembayaran_belanja_online_8	keluhan_belanja_online_1	keluhan_belanja_online_2	keluhan_belanja_online_3	keluhan_belanja_online_4	keluhan_belanja_online_5	keluhan_belanja_online_6
0	Pria	19	S1	Pelajar / Mahasiswa	< Rp 2 juta	Depok	NaN	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	Jumlah barang yang diterima kurang	NaN	NaN
1	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	8.0	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	Barang rusak/ salah tetapi tidak dapat dikemba...	Pembayaran sudah dilakukan; barang tidak tersedia	Jumlah barang yang diterima kurang	NaN	NaN
2	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	10.0	Sama saja	NaN	Mobile Banking	...	Transfer via ATM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Jakarta	7.0	Ya	Akses media sosial	Mobile Banking	...	Transfer via ATM	Kartu Kredit / Debit Online	Melalui minimarket	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	Pembayaran sudah dilakukan; barang tidak tersedia	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	Saldo eMoney/ eWallet berkurang tanpa melakuka...
4	Wanita	20	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Medan	7.0	Ya	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
280	Pria	21	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	8.0	Ya	Streaming video/ film	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
281	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Bogor	14.0	Sama saja	NaN	Mobile Banking	...	NaN	Kartu Kredit / Debit Online	NaN	NaN	Barang rusak/ salah tetapi tidak dapat dikemba...	NaN	NaN	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	NaN
282	Pria	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	padang	15.0	Sama saja	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
283	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Bandung	NaN	Tidak	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	Jumlah barang yang diterima kurang	NaN	NaN
284	Pria	51	S1	Pengusaha	Rp 5 juta – Rp 10 juta	Semarang	8.0	Sama saja	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

285 rows × 150 columns

In [96]:
```
# Menghapus baris dengan nilai NaN pada kolom tertentu
df.dropna(subset=['Durasi Penggunaan Internet per Hari (dalam Jam)'], inplace=True)
df
```

Out[96]:

	Jenis Kelamin	Usia	Pendidikan Terakhir	Pekerjaan	Penghasilan per Bulan	Domisili	Durasi Penggunaan Internet per Hari (dalam Jam)	Aktivitas Online Meningkat	Aktivitas yang Meningkat dalam 3 Bulan Terakhir	layanan_aktif_1	...	cara_pembayaran_belanja_online_5	cara_pembayaran_belanja_online_6	cara_pembayaran_belanja_online_7	cara_pembayaran_belanja_online_8	keluhan_belanja_online_1	keluhan_belanja_online_2	keluhan_belanja_online_3	keluhan_belanja_online_4	keluhan_belanja_online_5	keluhan_belanja_online_6
1	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	8.0	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	Barang rusak/ salah tetapi tidak dapat dikemba...	Pembayaran sudah dilakukan; barang tidak tersedia	Jumlah barang yang diterima kurang	NaN	NaN
2	Wanita	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	10.0	Sama saja	NaN	Mobile Banking	...	Transfer via ATM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Jakarta	7.0	Ya	Akses media sosial	Mobile Banking	...	Transfer via ATM	Kartu Kredit / Debit Online	Melalui minimarket	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	Pembayaran sudah dilakukan; barang tidak tersedia	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	Saldo eMoney/ eWallet berkurang tanpa melakuka...
4	Wanita	20	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Medan	7.0	Ya	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	NaN	NaN	NaN
5	Pria	42	S1	Karyawan Swasta	Rp 2 juta – Rp 5 juta	Surakarta	6.0	Ya	Akses media sosial	NaN	...	NaN	Kartu Kredit / Debit Online	NaN	menggunakan fitur Paylater	Barang yang diperoleh tidak sesuai dengan spes...	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
279	Pria	41	S1	Karyawan Swasta	> Rp 10 juta	Jakarta	10.0	Sama saja	NaN	Mobile Banking	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
280	Pria	21	SMA	Pelajar / Mahasiswa	< Rp 2 juta	Jakarta	8.0	Ya	Streaming video/ film	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
281	Pria	19	SMA	Pelajar / Mahasiswa	Rp 2 juta – Rp 5 juta	Bogor	14.0	Sama saja	NaN	Mobile Banking	...	NaN	Kartu Kredit / Debit Online	NaN	NaN	Barang rusak/ salah tetapi tidak dapat dikemba...	NaN	NaN	NaN	Pembayaran telah dilakukan tetapi tidak terdet...	NaN
282	Pria	19	SMA	Pelajar / Mahasiswa	< Rp 2 juta	padang	15.0	Sama saja	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
284	Pria	51	S1	Pengusaha	Rp 5 juta – Rp 10 juta	Semarang	8.0	Sama saja	Akses media sosial	Mobile Banking	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

214 rows × 150 columns

In [97]:
```
# Contigency Table
contigency2 = pd.crosstab(df['Pendidikan Terakhir'], df['Durasi Penggunaan Internet per Hari (dalam Jam)'])
contigency2
```

Out[97]:

Durasi Penggunaan Internet per Hari (dalam Jam)	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	24.0
Pendidikan Terakhir																		
D3	0	1	0	0	1	1	1	0	4	1	0	0	1	0	0	0	0	0
D4	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0
S1	0	1	11	3	7	6	1	6	2	6	14	1	0	5	2	1	2	1
S2	1	0	1	1	1	2	3	4	0	3	1	1	0	1	0	0	0	1
SMA	0	4	2	6	8	13	9	19	4	13	11	1	4	10	1	1	3	0
SMK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SMP	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

In [98]:
```
# Ei Table

Ei = contigency2.copy()

for s in contigency2.index:
    for o in contigency2.columns:
        Ei.loc[s, o] = contigency2.loc[s].sum() * contigency2.loc[:, o].sum() / contigency2.values.sum()

Ei
```

Out[98]:

Durasi Penggunaan Internet per Hari (dalam Jam)	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	24.0
Pendidikan Terakhir																		
D3	0.051402	0.308411	0.719626	0.514019	0.873832	1.130841	0.771028	1.542056	0.308411	1.387850	1.439252	0.154206	0.205607	0.925234	0.205607	0.102804	0.257009	0.102804
D4	0.014019	0.084112	0.196262	0.140187	0.238318	0.308411	0.210280	0.420561	0.084112	0.378505	0.392523	0.042056	0.056075	0.252336	0.056075	0.028037	0.070093	0.028037
S1	0.322430	1.934579	4.514019	3.224299	5.481308	7.093458	4.836449	9.672897	1.934579	8.705607	9.028037	0.967290	1.289720	5.803738	1.289720	0.644860	1.612150	0.644860
S2	0.093458	0.560748	1.308411	0.934579	1.588785	2.056075	1.401869	2.803738	0.560748	2.523364	2.616822	0.280374	0.373832	1.682243	0.373832	0.186916	0.467290	0.186916
SMA	0.509346	3.056075	7.130841	5.093458	8.658879	11.205607	7.640187	15.280374	3.056075	13.752336	14.261682	1.528037	2.037383	9.168224	2.037383	1.018692	2.546729	1.018692
SMK	0.004673	0.028037	0.065421	0.046729	0.079439	0.102804	0.070093	0.140187	0.028037	0.126168	0.130841	0.014019	0.018692	0.084112	0.018692	0.009346	0.023364	0.009346
SMP	0.004673	0.028037	0.065421	0.046729	0.079439	0.102804	0.070093	0.140187	0.028037	0.126168	0.130841	0.014019	0.018692	0.084112	0.018692	0.009346	0.023364	0.009346

In [99]:
```
# H0: Variabel Pendidikan Terakhir independen dengan Durasi Penggunaan Internet per Hari (dalam Jam)
# H1: Variabel Pendidikan Terakhir tidak independen dengan Durasi Penggunaan Internet per Hari (dalam Jam)

chi2 = ((contigency2 - Ei) ** 2 / Ei).sum().sum()

alpha = 0.05
row = 7
col = 18
df = (row-1)*(col-1)

chi2_alpha = st.chi2.ppf(1-alpha, df)

pval = 1 - st.chi2.cdf(chi2, df)

if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0,\nVariabel Pendidikan Terakhir tidak independen dengan Durasi Penggunaan Internet per Hari'

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

```
Hasil chi2 test:
chi2: 129.71
crit region: chi2 > 126.57
kesimpulan: chi2 di dalam crit region, reject H0,
Variabel Pendidikan Terakhir tidak independen dengan Durasi Penggunaan Internet per Hari
p-value: 0.03
```

Chi-squared test for Homogenity dengan alpha = 5%

In [102…
```
#H0: Distribusi usia tiap Jenis Kelamin sama/homogen
#H1: Distribusi Usia tiap jenis Kelamin tidak sama

alpha = 0.05

contigency3 = pd.crosstab(df['Jenis Kelamin'], df['Usia'])
contigency3
```

Out[102…

Usia	16	17	18	19	20	21	22	23	24	25	…	51	52	53	54	55	56	57	58	59	61
Jenis Kelamin																					
Pria	0	1	4	41	18	3	1	3	3	1	…	3	0	2	2	1	2	1	1	0	1
Wanita	1	4	11	46	14	8	3	4	0	2	…	1	5	4	3	1	0	0	0	1	0

2 rows × 44 columns

In [103…
```
# Ei Table

Ei = contigency3.copy()

for s in contigency3.index:
    for o in contigency3.columns:
        Ei.loc[s, o] = contigency3.loc[s].sum() * contigency3.loc[:, o].sum() / contigency3.values.sum()

Ei
```

Out[103…

Usia	16	17	18	19	20	21	22	23	24	25	…	51	52	53	54	55	56	57	58	59	61
Jenis Kelamin																					
Pria	0.477193	2.385965	7.157895	41.515789	15.270175	5.249123	1.908772	3.340351	1.431579	1.431579	…	1.908772	2.385965	2.863158	2.385965	0.954386	0.954386	0.477193	0.477193	0.477193	0.477193
Wanita	0.522807	2.614035	7.842105	45.484211	16.729825	5.750877	2.091228	3.659649	1.568421	1.568421	…	2.091228	2.614035	3.136842	2.614035	1.045614	1.045614	0.522807	0.522807	0.522807	0.522807

2 rows × 44 columns

In [104…
```
#H0: Distribusi usia tiap Jenis Kelamin sama/homogen
#H1: Distribusi Usia tiap jenis Kelamin tidak sama

chi2 = ((contigency3 - Ei) ** 2 / Ei).sum().sum()

alpha = 0.05
row = 2
col = 41
df = (row-1)*(col-1)

chi2_alpha = st.chi2.ppf(1-alpha, df)

pval = 1 - st.chi2.cdf(chi2, df)

if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0,\nDistribusi usia tiap jenis kelamin homogen atau sama'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0'

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

```
Hasil chi2 test:
chi2: 53.27
crit region: chi2 > 55.76
kesimpulan: chi2 di luar crit region, fail to reject H0,
Distribusi usia tiap jenis kelamin homogen atau sama
p-value: 0.08
```

CleanUp Data

In [107…
```
df.dropna(subset=['18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking ]'], inplace=True)

null_values = df['18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking ]'].isnull().sum()
print("Jumlah null values dalam kolom '18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking ]':", null_values)
```

```
Jumlah null values dalam kolom '18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking ]': 0
```

In [108…
```
#H0: Distribusi Frekuensi penggunaan mobile banking tiap domisili sama/homogen
#H1:Distribusi Frekuensi penggunaan mobile banking tiap domisili tidak sama/homogen

alpha = 0.05

contigency4 = pd.crosstab(df['Domisili'], df['18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking ]'])
contigency4
```

Out[108…

18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking]	2-5 kali per bulan	6-9 kali per bulan	> 10 kali per bulan	Kurang dari/ setidaknya 1 kali per bulan	Tidak Pernah
Domisili					
Aceh	0	1	0	0	0
BandarLampung	0	1	0	0	0
Bandung	16	15	34	0	0
Bekasi	2	2	2	1	0
Bogor	0	4	4	0	0
Bogor , Jawa Barat	1	0	0	0	0
Cilacap	0	0	1	0	0
Cilegon	0	0	1	0	0
Cimahi	0	1	0	0	0
Depok	2	3	6	0	0
Garut	0	0	1	0	0
Gorontalo	0	0	1	0	0
Jakarta	16	10	44	2	0
Kabupaten Bogor	0	0	1	0	0
Kisaran	0	0	1	0	0
Klaten	1	0	0	0	0
Lancaster	0	1	0	0	0
Madiun	0	1	0	0	0
Makassar	2	1	1	0	0
Malang	1	1	3	0	0
Medan	2	0	5	0	0
München, Germany	0	0	1	0	0
Padang	1	0	1	0	0
Palembang	1	0	2	0	0
Palu	0	0	2	0	0
Pematangsiantar	0	1	0	0	0
Pontianak	1	0	0	0	0
Purwokerto	1	0	0	1	0
Semarang	2	2	2	0	0
Sosok	1	0	0	0	0
Sukabumi	0	1	0	0	0
Surabaya	4	1	5	0	0
Surakarta	0	0	2	0	0
Surakarta	0	1	0	0	0
Tangerang	0	5	11	1	0
Tanjungpinang	0	0	1	0	0
Tebing Tinggi	0	0	0	0	1
Yogyakarta	2	0	3	0	0
bandar lampung	0	0	1	0	0
bogor	0	1	1	0	0
medan	0	0	1	0	0

In [109…

```
jumlah_baris, jumlah_kolom = contigency4.shape
print("Jumlah baris:", jumlah_baris)
print("Jumlah kolom:", jumlah_kolom)
```

```
Jumlah baris: 41
Jumlah kolom: 5
```

In [110…

```
# Ei Table

Ei = contigency4.copy()

for s in contigency4.index:
    for o in contigency4.columns:
        Ei.loc[s, o] = contigency4.loc[s].sum() * contigency4.loc[:, o].sum() / contigency4.values.sum()

Ei
```

Out[110…

18. Bagaimana frekuensi penggunaan Channel Bank berikut? [Mobile Banking]	2-5 kali per bulan	6-9 kali per bulan	> 10 kali per bulan	Kurang dari/ setidaknya 1 kali per bulan	Tidak Pernah
Domisili					
Aceh	0.221344	0.209486	0.545455	0.019763	0.003953
BandarLampung	0.221344	0.209486	0.545455	0.019763	0.003953
Bandung	14.387352	13.616601	35.454545	1.284585	0.256917
Bekasi	1.549407	1.466403	3.818182	0.138340	0.027668
Bogor	1.770751	1.675889	4.363636	0.158103	0.031621
Bogor , Jawa Barat	0.221344	0.209486	0.545455	0.019763	0.003953
Cilacap	0.221344	0.209486	0.545455	0.019763	0.003953
Cilegon	0.221344	0.209486	0.545455	0.019763	0.003953
Cimahi	0.221344	0.209486	0.545455	0.019763	0.003953
Depok	2.434783	2.304348	6.000000	0.217391	0.043478
Garut	0.221344	0.209486	0.545455	0.019763	0.003953
Gorontalo	0.221344	0.209486	0.545455	0.019763	0.003953
Jakarta	15.936759	15.083004	39.272727	1.422925	0.284585
Kabupaten Bogor	0.221344	0.209486	0.545455	0.019763	0.003953
Kisaran	0.221344	0.209486	0.545455	0.019763	0.003953
Klaten	0.221344	0.209486	0.545455	0.019763	0.003953
Lancaster	0.221344	0.209486	0.545455	0.019763	0.003953
Madiun	0.221344	0.209486	0.545455	0.019763	0.003953
Makassar	0.885375	0.837945	2.181818	0.079051	0.015810
Malang	1.106719	1.047431	2.727273	0.098814	0.019763
Medan	1.549407	1.466403	3.818182	0.138340	0.027668
München, Germany	0.221344	0.209486	0.545455	0.019763	0.003953
Padang	0.442688	0.418972	1.090909	0.039526	0.007905
Palembang	0.664032	0.628458	1.636364	0.059289	0.011858
Palu	0.442688	0.418972	1.090909	0.039526	0.007905
Pematangsiantar	0.221344	0.209486	0.545455	0.019763	0.003953
Pontianak	0.221344	0.209486	0.545455	0.019763	0.003953
Purwokerto	0.442688	0.418972	1.090909	0.039526	0.007905
Semarang	1.328063	1.256917	3.272727	0.118577	0.023715
Sosok	0.221344	0.209486	0.545455	0.019763	0.003953
Sukabumi	0.221344	0.209486	0.545455	0.019763	0.003953
Surabaya	2.213439	2.094862	5.454545	0.197628	0.039526
Surakarta	0.442688	0.418972	1.090909	0.039526	0.007905
Surakarta	0.221344	0.209486	0.545455	0.019763	0.003953
Tangerang	3.762846	3.561265	9.272727	0.335968	0.067194
Tanjungpinang	0.221344	0.209486	0.545455	0.019763	0.003953
Tebing Tinggi	0.221344	0.209486	0.545455	0.019763	0.003953
Yogyakarta	1.106719	1.047431	2.727273	0.098814	0.019763
bandar lampung	0.221344	0.209486	0.545455	0.019763	0.003953
bogor	0.442688	0.418972	1.090909	0.039526	0.007905
medan	0.221344	0.209486	0.545455	0.019763	0.003953

In [111…

```
#H0: Distribusi Frekuensi penggunaan mobile banking tiap domisili sama/homogen
#H1 :Distribusi Frekuensi penggunaan mobile banking tiap domisili tidak sama/homogen

chi2 = ((contigency4 - Ei) ** 2 / Ei).sum().sum()

alpha = 0.05
row = jumlah_baris
col = jumlah_kolom
df = (row-1)*(col-1)
```

```
chi2_alpha = st.chi2.ppf(1-alpha, df)

pval = 1 - st.chi2.cdf(chi2, df)

if chi2 < chi2_alpha:
    kesimpulan = 'chi2 di luar crit region, fail to reject H0'
else:
    kesimpulan = 'chi2 di dalam crit region, reject H0,\nDistribusi frekuensi penggunaan mobile banking berbeda secara signifikan di setiap domisili yang diuji'

print(f'''
Hasil chi2 test:
chi2: {chi2:.2f}
crit region: chi2 > {chi2_alpha:.2f}
kesimpulan: {kesimpulan}
p-value: {pval:.2f}
''')
```

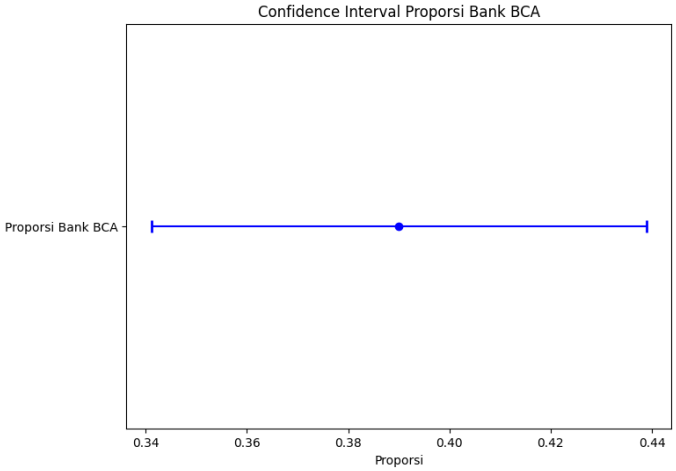
```
Hasil chi2 test:
chi2: 370.11
crit region: chi2 > 190.52
kesimpulan: chi2 di dalam crit region, reject H0,
Distribusi frekuensi penggunaan mobile banking berbeda secara signifikan di setiap domisili yang diuji
p-value: 0.00
```

Aththariq Lisan
18222013 - STI - K01

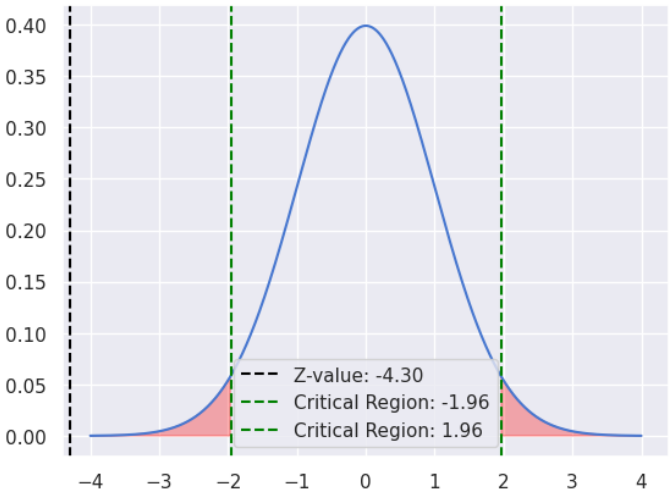
Insight from Data

1. Proporsi bank BCA

Dari hasil analisis yang didapat, kita dapat menafsirkan bahwa dengan tingkat kepercayaan 95%, kita memperkirakan bahwa proporsi penggunaan Bank BCA di populasi secara keseluruhan berada di rentang antara 0.3411 hingga 0.4390. Ini berarti kita memiliki keyakinan sebesar 95% bahwa proporsi penggunaan Bank BCA berada dalam rentang ini berdasarkan sampel yang digunakan.



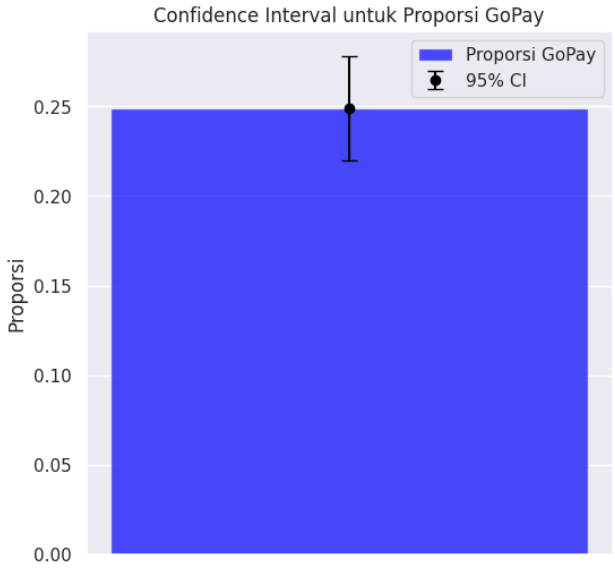
Hasil dari pengujian z-value juga menunjukkan bahwa nilai Z berada di dalam critical region. Artinya, kita menolak hipotesis nol (H0) bahwa proporsi penggunaan Bank BCA sama dengan 0.5. Dengan kata lain, proporsi penggunaan Bank BCA dari sampel yang diuji tidak sama dengan 0.5 secara signifikan pada tingkat signifikansi 5%.



Aththariq Lisan
18222013 - STI - K01

2. **Proporsi Gopay**

   Confidence interval (CI) menunjukkan rentang perkiraan proporsi GoPay di populasi dengan tingkat kepercayaan tertentu. Rentang ini adalah 0.2196 hingga 0.2780 dengan tingkat kepercayaan 95%. Ini berarti ada keyakinan sebesar 95% bahwa proporsi penggunaan GoPay di populasi sesungguhnya berada dalam rentang ini berdasarkan sampel yang diuji.

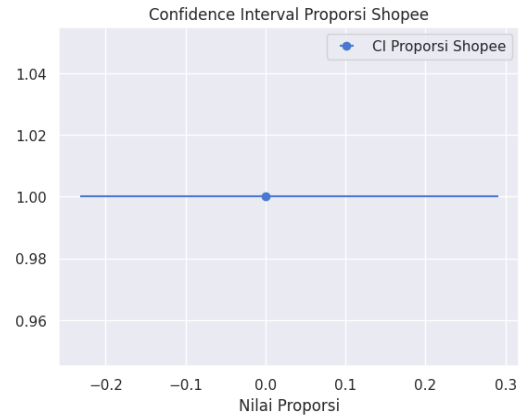


   Hasil uji hipotesis juga menunjukkan bahwa terdapat perbedaan signifikan antara proporsi penggunaan GoPay yang diamati dalam sampel dengan nilai yang diasumsikan (0.3) dengan tingkat kepercayaan 95%. Dikarenakan nilai Z berada di dalam critical region ($-1.96 < Z < 1.96$), kita menolak hipotesis nol bahwa proporsi GoPay adalah 0.3.

Aththariq Lisan
18222013 - STI - K01

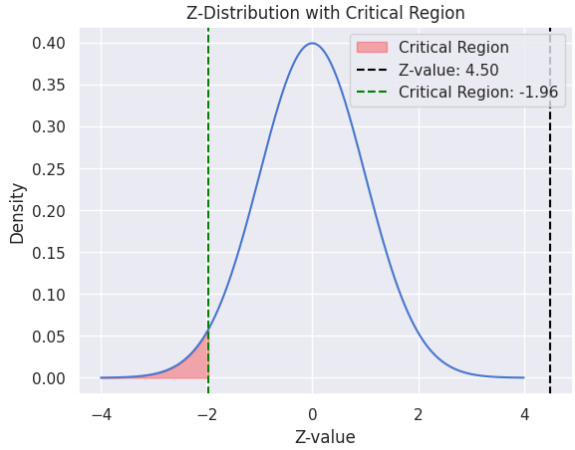
3. **Proporsi Shopee**

Untuk Confidence Interval proporsi Shopee (0.2320 < Proporsi Shopee < 0.2907), ini berarti kita memiliki keyakinan 95% bahwa proporsi penggunaan Shopee di populasi umum berada di rentang antara 0.2320 hingga 0.2907, berdasarkan sampel yang diambil. kita memiliki keyakinan sebesar 95% bahwa proporsi penggunaan Shopee umumnya ada dalam rentang tersebut. Ini memberikan gambaran tentang seberapa akurat perkiraan kita tentang penggunaan Shopee di antara sampel yang diuji.



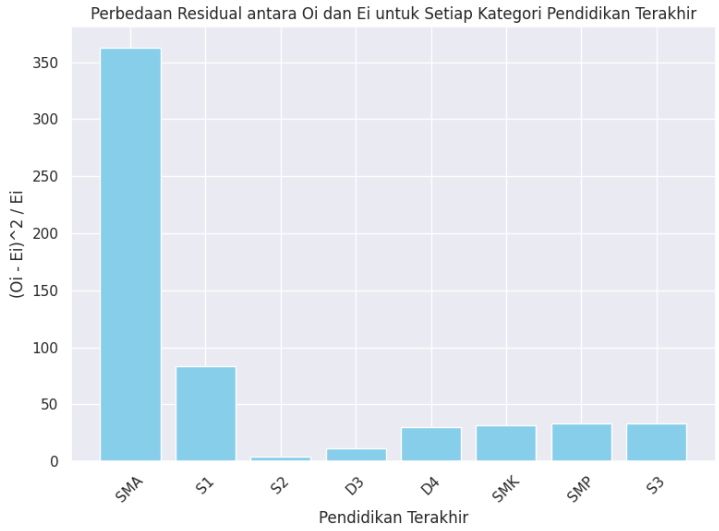
Analisis z-tets yang menguji hipotesis terhadap proporsi penggunaan Shopee yang diduga kurang dari 0.2 juga menemukan bahwa Z-value (4.4985) berada di luar Critical Region (-1.96). Hal ini menunjukkan bahwa kita dapat menolak H0 (hipotesis nol) karena Z-value yang diperoleh jauh lebih besar dari batas kritis yang ditentukan. P-value yang sangat kecil (0.00000684) juga menunjukkan bahwa kita memiliki bukti yang kuat untuk menolak H0.

Dalam konteks ini, kita memiliki cukup bukti untuk menyimpulkan bahwa proporsi penggunaan Shopee jauh lebih tinggi daripada 0.2.

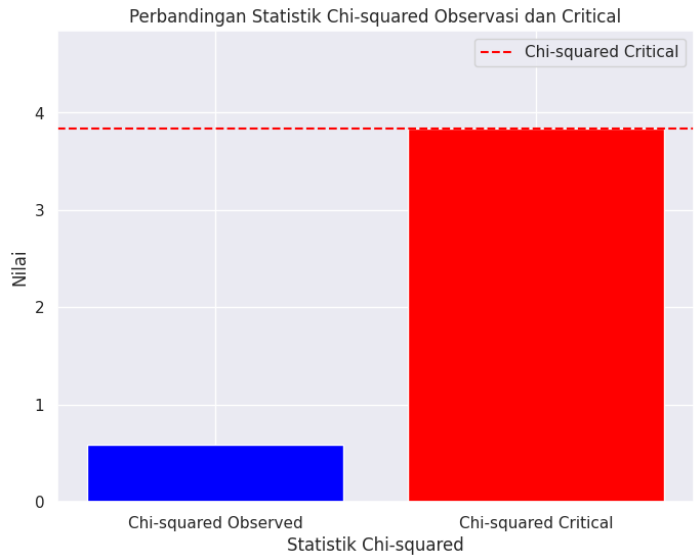


Aththariq Lisan
18222013 - STI - K01

4. **Distribusi Pendidikan Terakhir**
   Distribusi dari variabel "Pendidikan Terakhir" tidak mengikuti distribusi seragam. Dari hasil chi-squared test, nilai chi2 yang diperoleh (622.79) jauh melebihi nilai kritis yang ditetapkan (15.51) pada tingkat signifikansi alpha 0.05. Hal ini mengarah pada penolakan H0 (hipotesis nol), dengan p-value yang sangat rendah (0.00), menunjukkan bukti yang sangat kuat bahwa distribusi pendidikan terakhir tidak bersifat seragam.



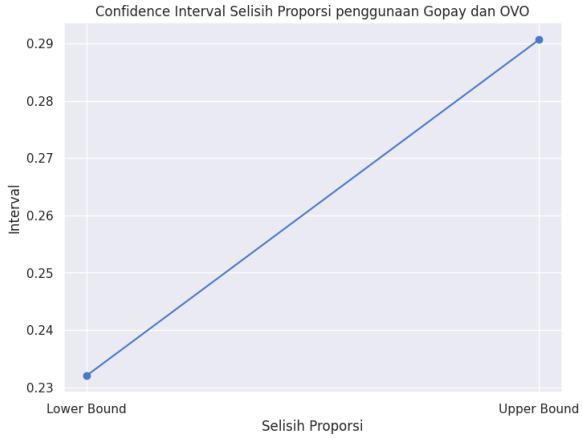
5. **Distribusi Jenis Kelamin**
   Dalam konteks uji chi-squared untuk keseragaman distribusi Jenis Kelamin, nilai chi2 yang diperoleh (0.59) berada di luar daerah kritis (critical region) yang ditetapkan (3.84). Hal ini mengarah pada kesimpulan bahwa tidak ada cukup bukti untuk menolak H0 (H0: Distribusi Jenis Kelamin = uniform distribution), yang berarti distribusi Jenis Kelamin dapat dianggap sebagai distribusi seragam.



Aththariq Lisan
18222013 - STI - K01

6. **Perbedaan Proporis Gopay dan Ovo**

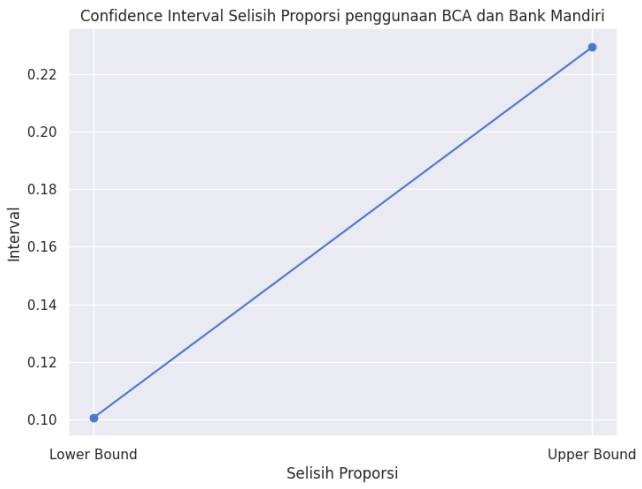
   Hasil Confidence Interval untuk selisih proporsi penggunaan GoPay dan OVO menunjukkan rentang antara 0.02724 hingga 0.10546. Ini mengindikasikan bahwa dengan tingkat kepercayaan 95%, perbedaan proporsi penggunaan GoPay dan OVO pada populasi umum diperkirakan berada dalam rentang ini berdasarkan sampel yang digunakan. Rentang ini menunjukkan bahwa proporsi penggunaan GoPay cenderung lebih tinggi daripada OVO, namun ada fluktuasi dalam perbedaannya yang dapat berada di kisaran tersebut.



7. **Perbedaan Proporsi Bank BCA dan Mandiri**

   Dari Confidence Interval Selisih Proporsi penggunaan BCA dan Bank Mandiri (0.00269 < P BCA - P Bank Mandiri < 0.08326), insight yang dapat diambil adalah bahwa dengan tingkat kepercayaan 95%, kita memperkirakan selisih proporsi penggunaan BCA dan Bank Mandiri pada populasi secara keseluruhan berada di rentang antara 0.00269 hingga 0.08326.

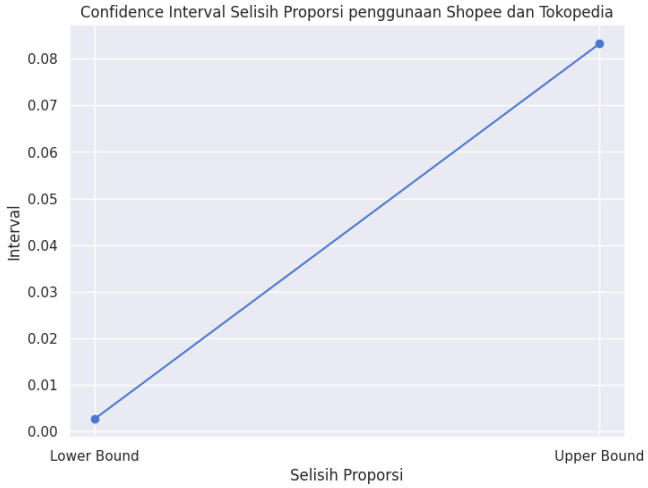
   Ini menunjukkan bahwa terdapat perbedaan dalam proporsi penggunaan antara BCA dan Bank Mandiri. Namun, rentang interval ini menunjukkan bahwa perbedaan ini bisa sangat kecil (0.00269) hingga cukup signifikan (0.08326) tergantung pada populasi sebenarnya.



Aththariq Lisan
18222013 - STI - K01

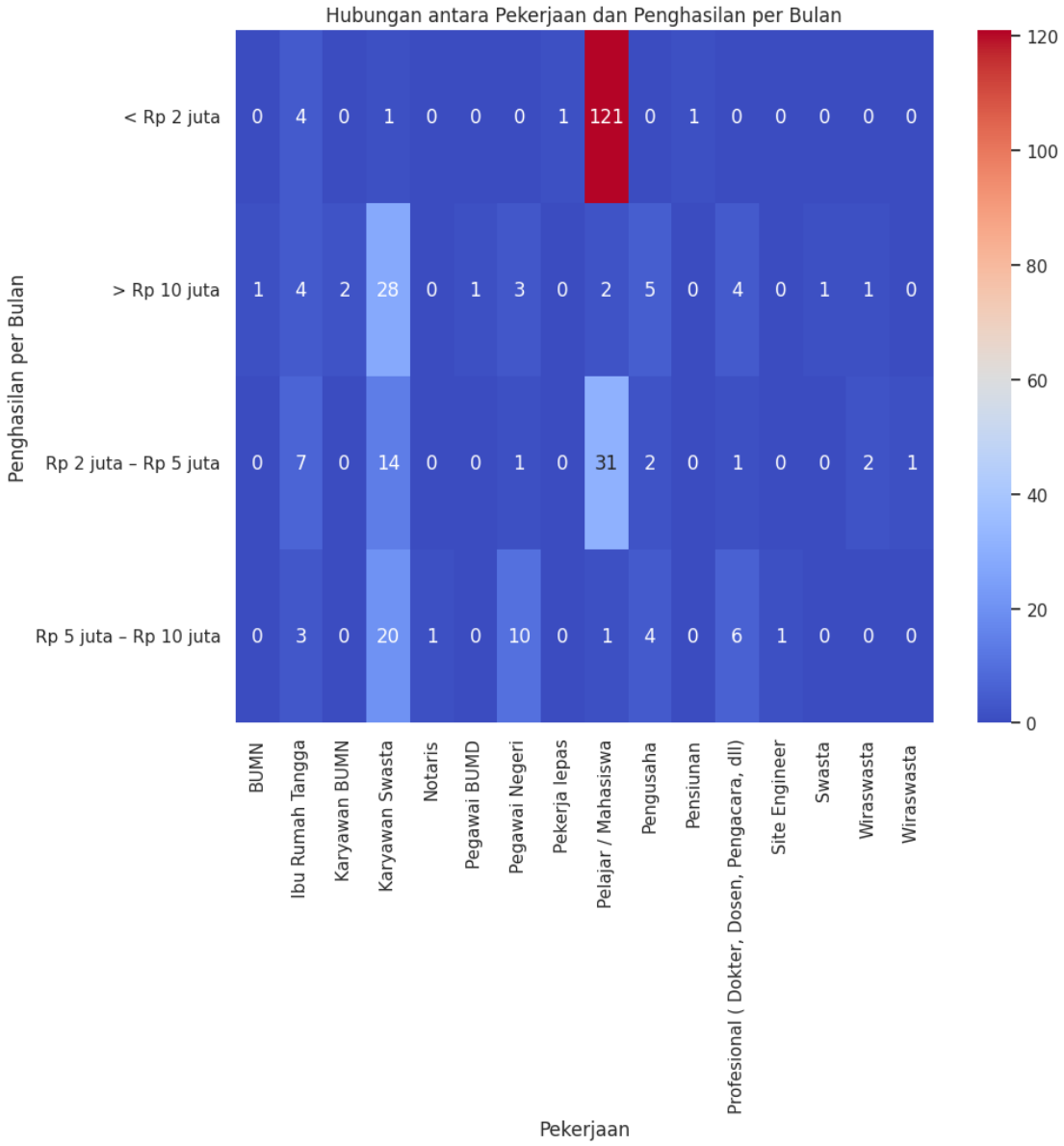
8. **Perbedaan Proporsi Shopee dan Tokopedia**
   Dari hasil Confidence Interval Selisih Proporsi penggunaan Shopee dan Tokopedia (0.00269 < P Shopee - P Tokopedia < 0.08326), insight yang bisa diambil adalah bahwa dengan tingkat kepercayaan 95%, kita memperkirakan bahwa selisih proporsi penggunaan Shopee dan Tokopedia pada populasi secara keseluruhan berada dalam rentang antara 0.00269 hingga 0.08326.

   Hal ini menunjukkan bahwa ada perbedaan dalam proporsi penggunaan antara Shopee dan Tokopedia, tetapi rentang interval ini menunjukkan bahwa perbedaan ini bisa kecil (0.00269) hingga cukup signifikan (0.08326) tergantung pada populasi sebenarnya.



9. **Independensi antara Variabel Pekerjaan dan penghasilan per bulan**
   Dari hasil uji Chi-squared untuk independensi antara Variabel Pekerjaan dan penghasilan per bulan (chi2 = 258.53, df = 45, p-value: 0.00), kita dapat menyimpulkan bahwa terdapat hubungan yang signifikan antara Variabel Pekerjaan dan penghasilan per bulan dalam dataset yang digunakan.

   Insight yang dapat diambil adalah bahwa pekerjaan seseorang dapat memiliki dampak atau keterkaitan yang signifikan terhadap penghasilan per bulan mereka. Ini menunjukkan adanya asosiasi yang kuat antara jenis pekerjaan yang dijalani seseorang dengan tingkat penghasilan yang mereka peroleh.



10. **Distribusi frekuensi penggunaan mobile banking di setiap domisili**

    Hasil uji chi-squared menunjukkan bahwa distribusi frekuensi penggunaan mobile banking berbeda secara signifikan di setiap domisili yang diuji. Hal ini diperkuat oleh nilai chi-squared yang dihitung berada di dalam daerah kritis dengan p-value

yang sangat rendah (< 0.05), sehingga H0 ditolak.