



Technische
Universität
Braunschweig



Gruppe 7: Extramarital Affairs

Dania Arshad, Michael Biermann, Nicole Funke, Tim Matschulla & Agatha Stark

Vorstellung Aufgabe

About Dataset

Enclosed with this notebook is the dataset `affairs.csv`. Collected in 1969 by means of a survey by the magazine *Psychology Today*.

The significance of the data collected is shown in the following table:

Variable	Description
affairs	How often engaged in extramarital sexual intercourse during the past year?
gender	Gender
age	Age in years. Coding: 17.5 = under 20, 22 = 20–24, 27 = 25–29, 32 = 30–34, 37 = 35–39, 42 = 40–44, 47 = 45–49, 52 = 50–54, 57 = 55 or over.
yearsmarried	Number of years married. Coding: 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years.
children	Are there children in the marriage?
religiousness	Religiousness level. Coding: 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very.
education	Level of education. Coding: 9 = grade school, 12 = high school graduate, 14 = some college, 16 = college graduate, 17 = some graduate work, 18 = master's degree, 20 = Ph.D., M.D., or other advanced degree.
occupation	Occupation according to Hollingshead classification (reverse numbering)
rating	Self-rating of marriage. Coding: 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, 5 = very happy.

Your Task

Prove / Disprove the following 5 Hypothesis:

1. **Women cheat more often than men.**
2. **People who are not religious have more affairs.**
3. **As long as men have a happy marriage, they have fewer affairs.**
4. **Educated persons rate their marriage higher.**
5. **People with a high income have fewer children, are not religious and have more affairs.**

Vorstellung Datenset

```
# Display the first 20 rows  
print(df.head(20))
```

	affairs	gender	age	yearsmarried	children	religiousness	education	occupation	rating
0	0	male	37.0	10.00	no	3	18	7	4
1	0	female	27.0	4.00	no	4	14	6	4
2	0	female	32.0	15.00	yes	1	12	1	4
3	0	male	57.0	15.00	yes	5	18	6	5
4	0	male	22.0	0.75	no	2	17	6	3
5	0	female	32.0	1.50	no	2	17	5	5
6	0	female	22.0	0.75	no	2	12	1	3
7	0	male	57.0	15.00	yes	2	14	4	4
8	0	female	32.0	15.00	yes	4	16	1	2
9	0	male	22.0	1.50	no	4	14	4	5
10	0	male	37.0	15.00	yes	2	20	7	2
11	0	male	27.0	4.00	yes	4	18	6	4
12	0	male	47.0	15.00	yes	5	17	6	4
13	0	female	22.0	1.50	no	2	17	5	4
14	0	female	27.0	4.00	no	4	14	5	4
15	0	female	37.0	15.00	yes	1	17	5	5
16	0	female	37.0	15.00	yes	2	18	4	3
17	0	female	22.0	0.75	no	3	16	5	4
18	0	female	22.0	1.50	no	2	16	5	5
19	0	female	27.0	10.00	yes	2	14	1	5

Importe

```
# Here you can Import everthing you need  
# YOUR CODE HERE  
import pandas as pd  
import numpy as np  
from scipy import stats  
import matplotlib.pyplot as plt  
import seaborn as sb
```

Hypothese 1

Frauen betrügen häufiger als Männer.

Hypothese 1

Frauen betrügen häufiger als Männer.

- Durchschnittliche Anzahl der Affären je Gruppe (Frauen/ Männer) berechnen

Hypothese 1

Frauen betrügen häufiger als Männer.

- Durchschnittliche Anzahl der Affären je Gruppe (Frauen/ Männer) berechnen

Durchschnittliche Anzahl Affären:

gender

female 1.419048

male 1.496503

Hypothese 1

Frauen betrügen häufiger als Männer.

- Durchschnittliche Anzahl der Affären je Gruppe (Frauen/ Männer) berechnen

Durchschnittliche Anzahl Affären:

gender	
female	1.419048
male	1.496503

Hypothese 1

Frauen betrügen häufiger als Männer.

- Durchschnittliche Anzahl der Affären je Gruppe (Frauen/ Männer) berechnen

Durchschnittliche Anzahl Affären:

gender	
female	1.419048
male	1.496503

→ erste Hinweise gegen die Hypothese

Hypothese 1

- Erstellung von zwei separaten Datenreihen (Zwei-Stichproben-t-test)
- Einseitigen Test berechnen

```
# t-test
t_stat, p_value_two_sided = stats.ttest_ind(women, men, equal_var=False)

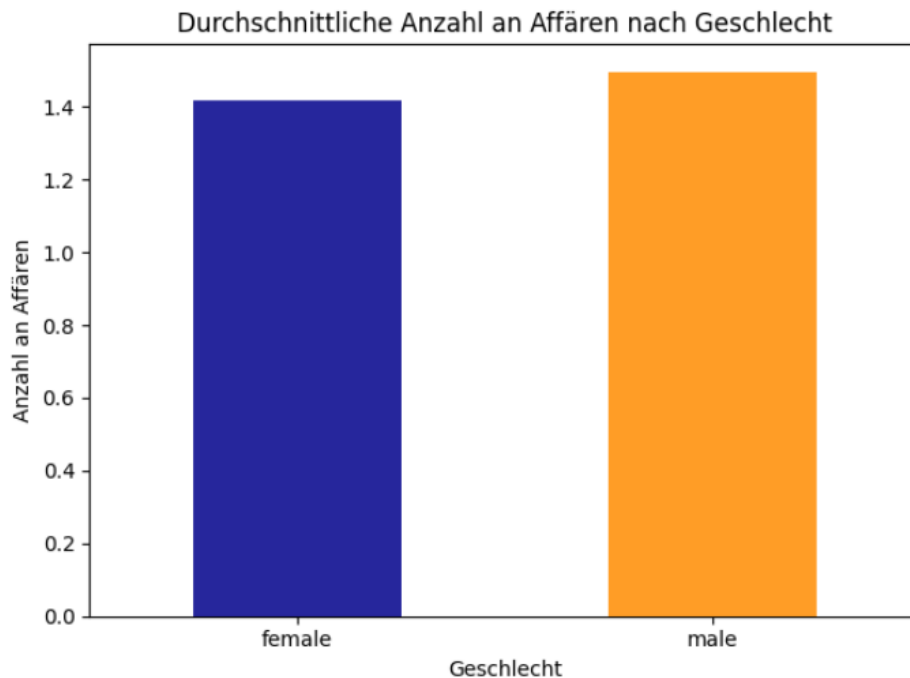
# One-sided p-value
p_value_one_sided = p_value_two_sided / 2 if t_stat > 0 else 1 - (p_value_two_sided / 2)

print("\nT-Statistik:", t_stat)
print("Einseitiger p-Wert:", p_value_one_sided)
```

T-Statistik: -0.28732904440368817
Einseitiger p-Wert: 0.6130197039823697

Hypothese 1

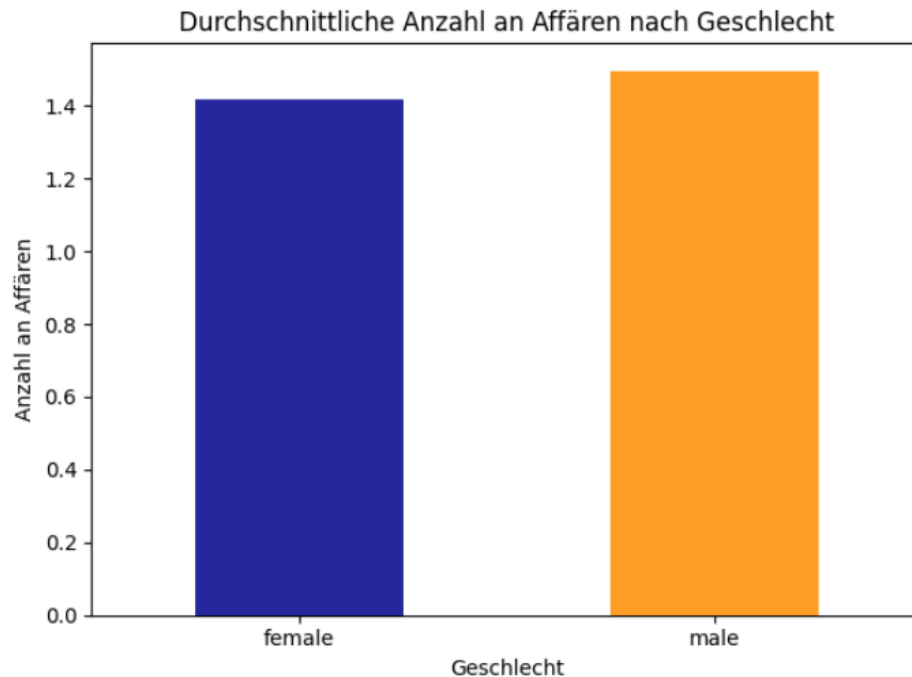
```
# Interpretation
if p_value_one_sided < 0.05:
    print("\nErgebnis: Hypothese 1 wird verifiziert. Frauen betrügen (signifikant) häufiger als Männer")
else:
    print("\nErgebnis: Hypothese 1 wird falsifiziert. Kein signifikanter Beweis dafür, dass Frauen mehr betrügen.")
```



T-Statistik: -0.28732904440368817
Einseitiger p-Wert: 0.6130197039823697

Hypothese 1

```
# Interpretation
if p_value_one_sided < 0.05:
    print("\nErgebnis: Hypothese 1 wird verifiziert. Frauen betrügen (signifikant) häufiger als Männer")
else:
    print("\nErgebnis: Hypothese 1 wird falsifiziert. Kein signifikanter Beweis dafür, dass Frauen mehr betrügen.")
```

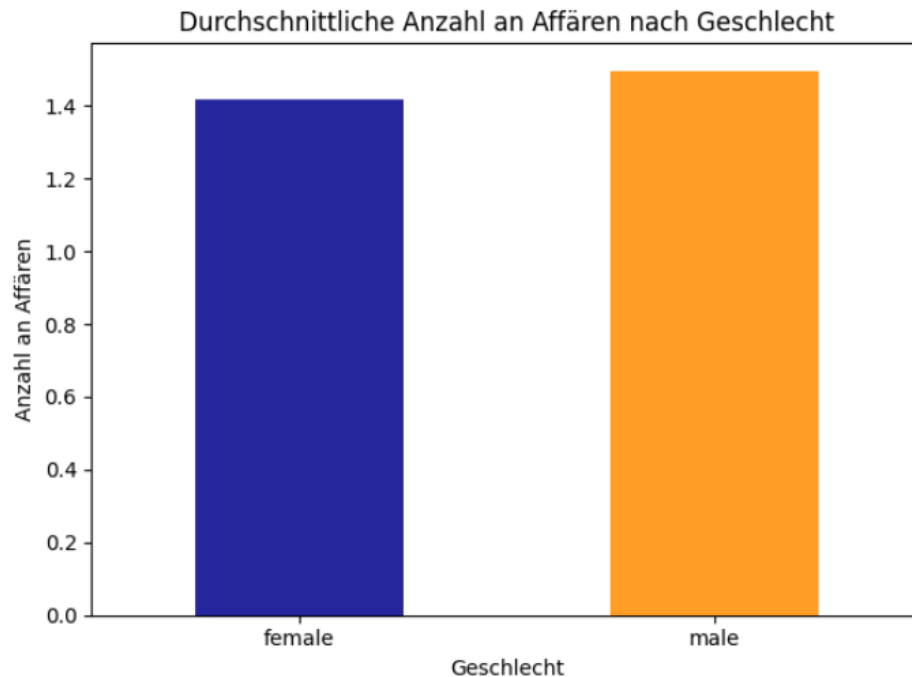


T-Statistik: -0.28732904440368817

Einseitiger p-Wert: 0.6130197039823697

Hypothese 1

```
# Interpretation
if p_value_one_sided < 0.05:
    print("\nErgebnis: Hypothese 1 wird verifiziert. Frauen betrügen (signifikant) häufiger als Männer")
else:
    print("\nErgebnis: Hypothese 1 wird falsifiziert. Kein signifikanter Beweis dafür, dass Frauen mehr betrügen.")
```



T-Statistik: -0.28732904440368817

Einseitiger p-Wert: 0.6130197039823697

Ergebnis: $p\text{-Wert} \geq 0,05$: Kein signifikanter Beweis dafür, dass Frauen mehr betrügen.

Hypothese 2

Personen, die nicht religiös sind, haben mehr Affären.

Hypothese 2

Personen, die nicht religiös sind, haben mehr Affären.

- Aufteilung der Religiosität (Level von 1-5)
- Anzahl der Affären je Religiosität

Hypothese 2

Personen, die nicht religiös sind, haben mehr Affären.

- Aufteilung der Religiosität (Level von 1-5)
- Anzahl der Affären je Religiosität

```
# Filter for affairs and religiousness
both : pd.DataFrame = df[["religiousness","affairs"]]

# Split religiousness (religious levels: 2 and below = not religious, 3 and above = religious)
religious : pd.DataFrame = both[both["religiousness"] >=3]
not_religious : pd.DataFrame = both[both["religiousness"] <=2]
```

Hypothese 2

Personen, die nicht religiös sind, haben mehr Affären.

- Aufteilung der Religiosität (Level von 1-5)
- Anzahl der Affären je Religiosität

```
# Filter for affairs and religiousness
both : pd.DataFrame = df[["religiousness","affairs"]]

# Split religiousness (religious levels: 2 and below = not religious, 3 and above = religious)
religious : pd.DataFrame = both[both["religiousness"] >=3]
not_religious : pd.DataFrame = both[both["religiousness"] <=2]
```

```
# Count the number of people with affairs
count_r : pd.DataFrame = religious["affairs"].value_counts().sort_index()
count_not : pd.DataFrame = not_religious["affairs"].value_counts().sort_index()

# Plot
fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True, sharey=True)
fig.supxlabel("Anzahl der Affären")
fig.supylabel("Anzahl der Personen")

ax1.bar(count_r.keys(), count_r, color="darkblue")
ax1.set_title("Religiös")

ax2.bar(count_not.keys(), count_not, color="darkorange")
ax2.set_title("Nicht religiös")
plt.show()
```

Hypothese 2

Personen, die nicht religiös sind, haben mehr Affären.

- Aufteilung der Religiosität (Level von 1-5)
- Anzahl der Affären je Religiosität

```
# Filter for affairs and religiousness
both : pd.DataFrame = df[["religiousness","affairs"]]

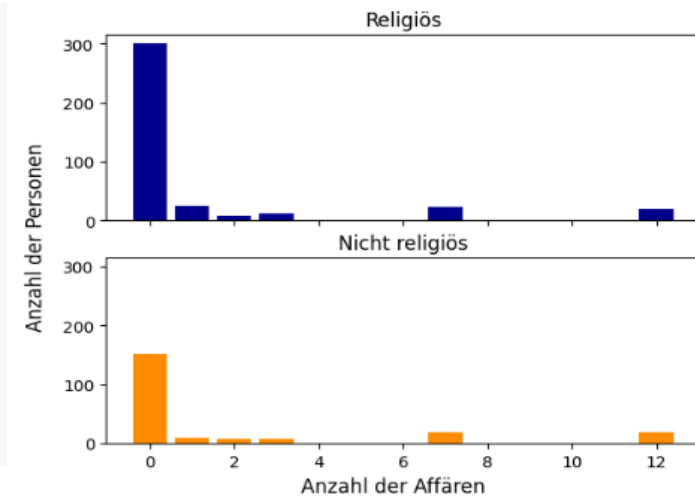
# Split religiousness (religious levels: 2 and below = not religious, 3 and above = religious)
religious : pd.DataFrame = both[both["religiousness"] >=3]
not_religious : pd.DataFrame = both[both["religiousness"] <=2]
```

```
# Count the number of people with affairs
count_r : pd.DataFrame = religious["affairs"].value_counts().sort_index()
count_not : pd.DataFrame = not_religious["affairs"].value_counts().sort_index()

# Plot
fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True, sharey=True)
fig.supxlabel("Anzahl der Affären")
fig.supylabel("Anzahl der Personen")

ax1.bar(count_r.keys(), count_r, color="darkblue")
ax1.set_title("Religiös")

ax2.bar(count_not.keys(), count_not, color="darkorange")
ax2.set_title("Nicht religiös")
plt.show()
```



Hypothese 2

- Prozentzahl der Personen, die mindestens eine Affäre hatten
- Durchschnittliche Affären je Religiosität
- **Ergebnis:** Sehr geringer Unterschied zwischen religiösen und nicht religiösen Personen

```
# Split religious and not religious people for affairs
affairs_r : pd.DataFrame = religious[religious["affairs"] >=1]
affairs_not : pd.DataFrame = not_religious[not_religious["affairs"]>=1]

# Percentage of affairs
pct_affairs_religious : float = round(len(affairs_r)/len(religious)*100, 2)
pct_affairs_not_religious : float = round(len(affairs_not)/len(not_religious)*100, 2)

print(f"{pct_affairs_religious} % der {len(religious)} religiösen Personen hatten im Jahr vor der Befragung mindestens eine Affäre.")
print(f"{pct_affairs_not_religious} % der {len(not_religious)} nicht religiösen Personen hatten im Jahr vor der Befragung mindestens eine Affäre.")
print()

# Filter only affairs for .mean
affairs_r : pd.DataFrame = religious[religious["affairs"]>0]
affairs_not : pd.DataFrame = not_religious[not_religious["affairs"]>0]

# Average affairs
avg_affairs_r : float = round(affairs_r.mean(), 2)
print(f"Die insgesamt {len(affairs_r)} religiösen Personen haben im Durchschnitt {avg_affairs_r} Affären.")

avg_affairs_not : float = round(affairs_not.mean(), 2)
print(f"Die insgesamt {len(affairs_not)} nicht religiösen Personen haben im Durchschnitt {avg_affairs_not} Affären.")
```

22.88 % der 389 religiösen Personen hatten im Jahr vor der Befragung mindestens eine Affäre.
28.77 % der 212 nicht religiösen Personen hatten im Jahr vor der Befragung mindestens eine Affäre.

Die insgesamt 389 religiösen Personen haben im Durchschnitt 1.23 Affären.
Die insgesamt 212 nicht religiösen Personen haben im Durchschnitt 1.86 Affären.

Hypothese 2

- Korrelationstest: Spearman, da nicht normalverteilt
- Plot der Korrelation
- **Ergebnis:** -0.076 (sehr schwach bis keine Korrelation)

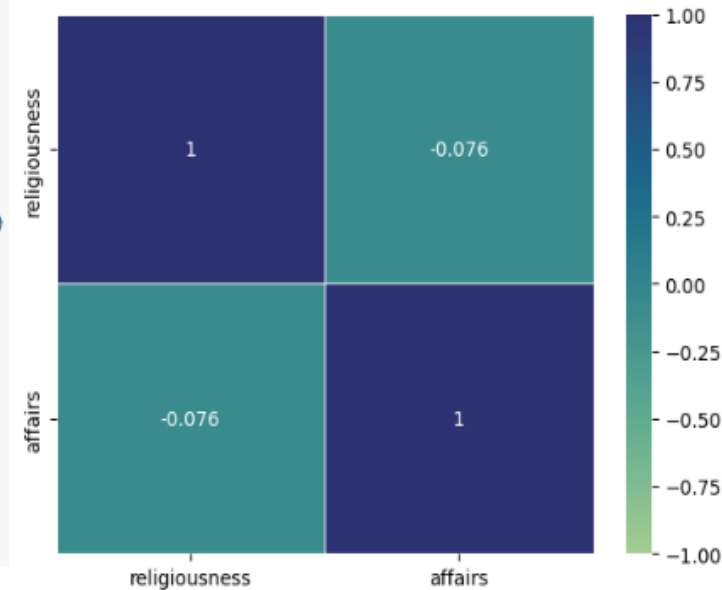
```
# Correlation test
def fkt(x):
    if x <= 2:
        return 0
    else:
        return 1

bi_religiousness : pd.DataFrame = both.copy()
bi_religiousness["religiousness"] : pd.DataFrame = bi_religiousness["religiousness"].map(fkt)

spearmancorr : pd.DataFrame = bi_religiousness.corr(method='spearman', numeric_only=True)

# Plot
sns.heatmap(spearmancorr,
            vmin=-1., vmax=1.,
            xticklabels=spearmancorr.columns,
            yticklabels=spearmancorr.columns,
            cmap='crest',
            annot=True,
            linewidth=0.5)

plt.show()
```



Hypothese 3

Solange Männer eine glückliche Ehe führen, haben sie weniger Affären.

Hypothese 3

Solange Männer eine glückliche Ehe führen, haben sie weniger Affären.

```
# Filter dataset to include only men
df_men = df[df["gender"] == "male"].copy()

# Create binary variable for happy marriage
# 1 = happy (rating >= 4), 0 = less happy (rating < 4)
df_men["happy_marriage"] = np.where(df_men["rating"] >= 4, 1, 0)

# Split into two independent groups
happy_group = df_men[df_men["happy_marriage"] == 1]["affairs"]
unhappy_group = df_men[df_men["happy_marriage"] == 0]["affairs"]
```

Hypothese 3

Solange Männer eine glückliche Ehe führen, haben sie weniger Affären.

```
# Filter dataset to include only men
df_men = df[df["gender"] == "male"].copy()

# Create binary variable for happy marriage
# 1 = happy (rating >= 4), 0 = less happy (rating < 4)
df_men["happy_marriage"] = np.where(df_men["rating"] >= 4, 1, 0)

# Split into two independent groups
happy_group = df_men[df_men["happy_marriage"] == 1]["affairs"]
unhappy_group = df_men[df_men["happy_marriage"] == 0]["affairs"]
```

- Datensatz auf Männer gefiltert
 - Variable „rating“ als Maß: $\geq 4 \rightarrow$ glückliche Ehe | $< 4 \rightarrow$ weniger glückliche Ehe
- Bildung von zwei unabhängigen Gruppen

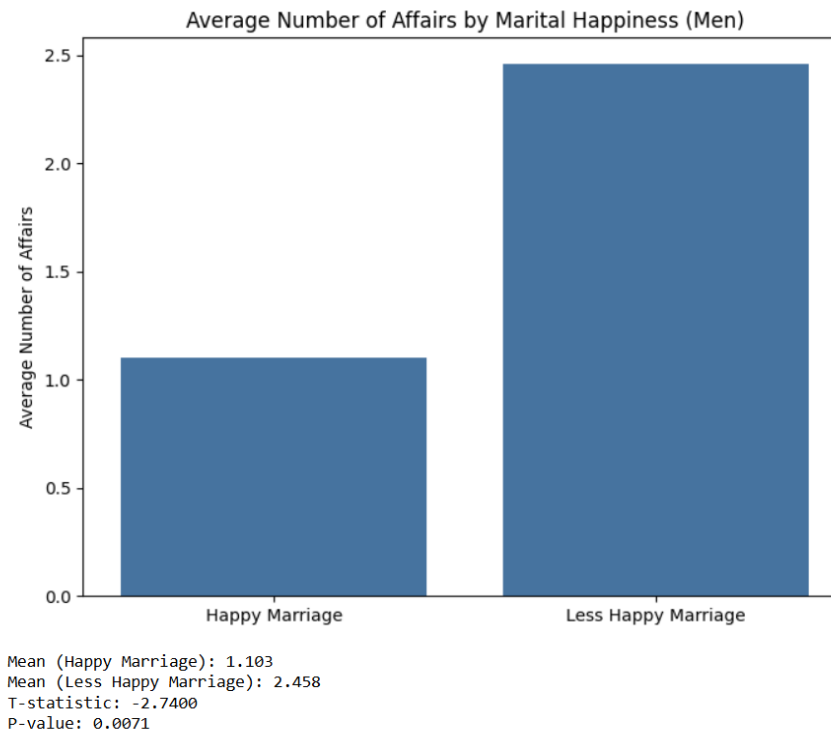
Hypothese 3

```
# Calculate group means
mean_happy = happy_group.mean()
mean_unhappy = unhappy_group.mean()

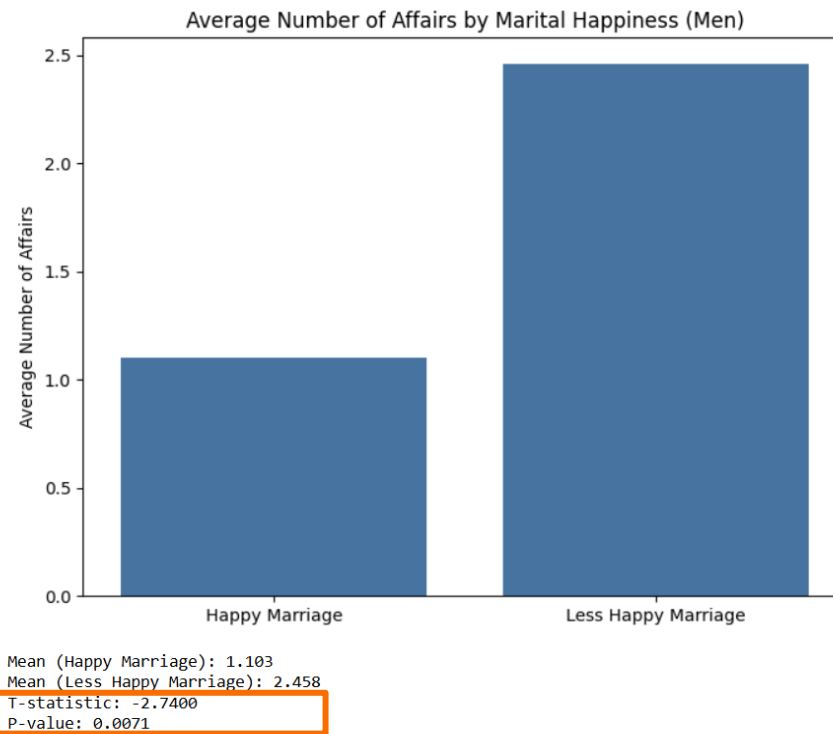
# Perform Welch's t-test (robust against unequal variances)
t_stat, p_value = stats.ttest_ind(
    happy_group,
    unhappy_group,
    equal_var=False
)
```

- Vergleich zweier unabhängiger Mittelwerte
- Zwei-Stichproben-T-Test

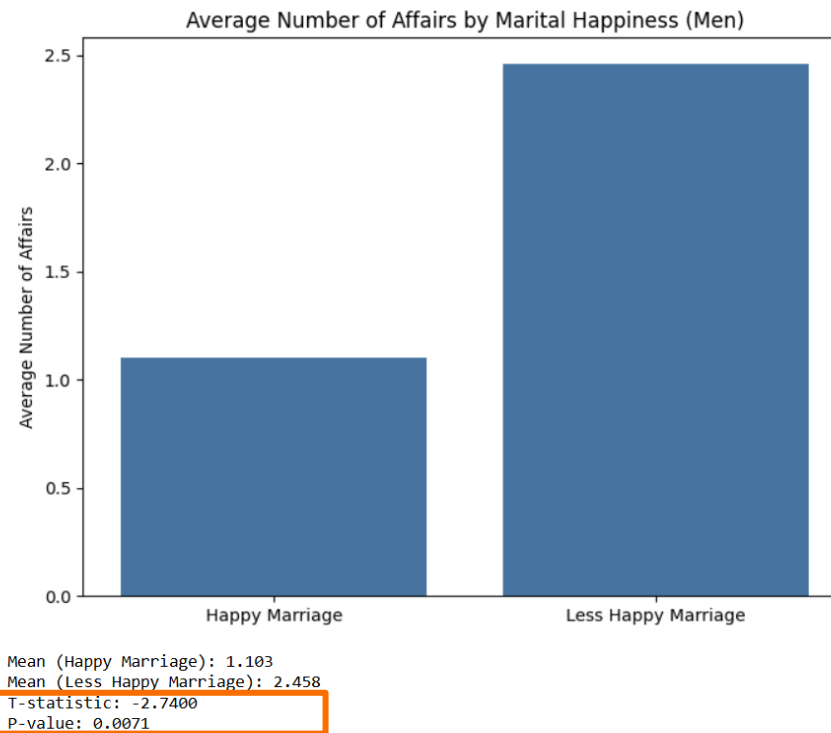
Hypothese 3



Hypothese 3



Hypothese 3



- Durchschnittliche Affären: glückliche Ehe → niedriger | weniger glückliche Ehe → höher
- Nullhypothese verworfen
- **Ergebnis:** p-Wert < 0,05: Negativer Zusammenhang zwischen Eheglück und Affären

Hypothese 4

Höher gebildete Personen bewerten ihre Ehe positiver.

Hypothese 4

Höher gebildete Personen bewerten ihre Ehe positiver.

- Vergleich der Durchschnittswerte
- Korrelationsanalyse (Pearson-Korrelation)

Hypothese 4

Höher gebildete Personen bewerten ihre Ehe positiver.

- Vergleich der Durchschnittswerte
- Korrelationsanalyse (Pearson-Korrelation)

```
df = pd.read_csv("affairs.csv")  
  
print(df.groupby("education")["rating"].mean())  
  
corr, p_value = stats.pearsonr(df["education"],df["rating"])  
  
print("Correlation:", corr)  
  
print("P-value:", p_value)
```

Hypothese 4

- Niedriges Bildungsniveau (education = 9):
Durchschnittliche Bewertung ≈ 2.57
- Höheres Bildungsniveau (education = 16):
Durchschnittliche Bewertung ≈ 4.24

```
education
9      2.571429
12     3.454545
14     3.896104
16     4.243478
17     3.943820
18     3.866071
20     4.012500
Name: rating, dtype: float64
Correlation: 0.10930347290914201
P-value: 0.007316835380154443
```

Hypothese 4

- Niedriges Bildungsniveau (education = 9):
Durchschnittliche Bewertung ≈ 2.57
- Höheres Bildungsniveau (education = 16):
Durchschnittliche Bewertung ≈ 4.24

```
education
9      2.571429
12     3.454545
14     3.896104
16     4.243478
17     3.943820
18     3.866071
20     4.012500
Name: rating, dtype: float64
Correlation: 0.10930347290914201
P-value: 0.007316835380154443
```

Hypothese 4

- Niedriges Bildungsniveau (education = 9):

Durchschnittliche Bewertung ≈ 2.57

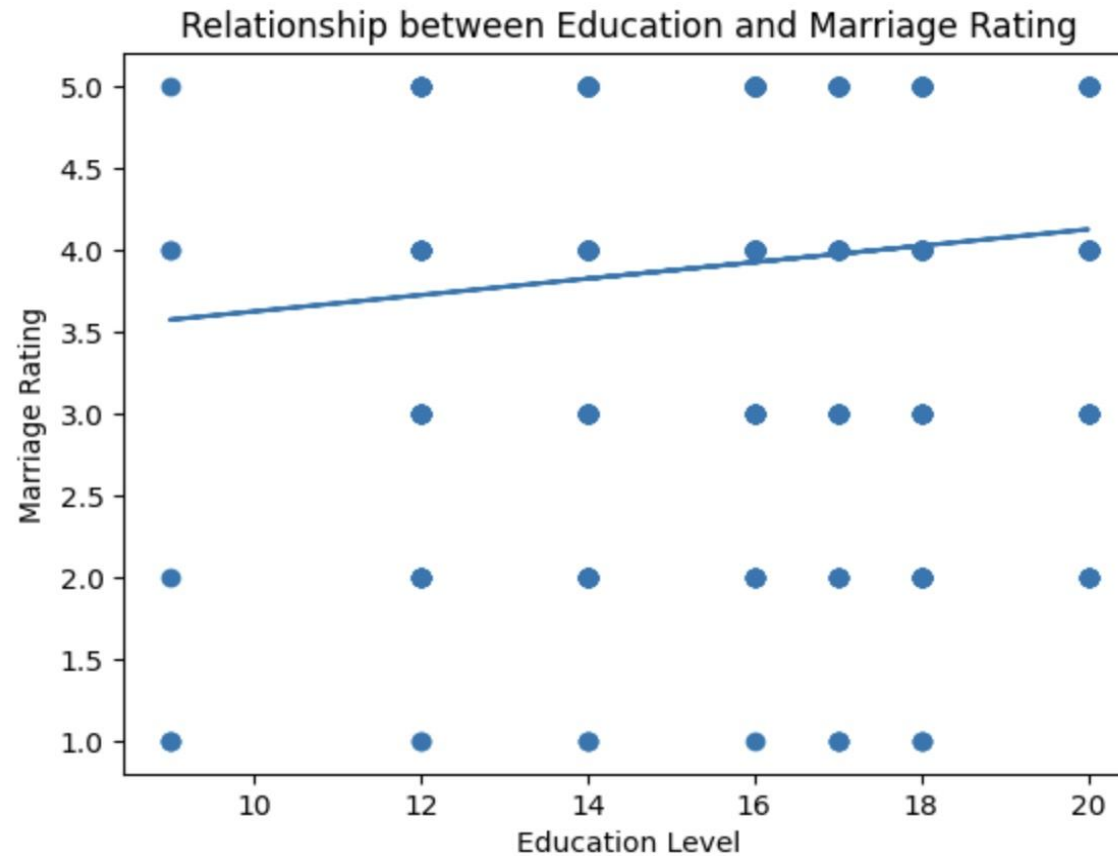
- Höheres Bildungsniveau (education = 16):

Durchschnittliche Bewertung ≈ 4.24

```
education
9      2.571429
12     3.454545
14     3.896104
16     4.243478
17     3.943820
18     3.866071
20     4.012500
Name: rating, dtype: float64
Correlation: 0.10930347290914201
P-value: 0.007316835380154443
```

- Ergebnis:** p-Wert < 0.05 : Signifikanter Beweis dafür, dass Personen mit höherem Bildungsniveau ihre Ehe im Durchschnitt besser bewerten.

Hypothese 4



Hypothese 5

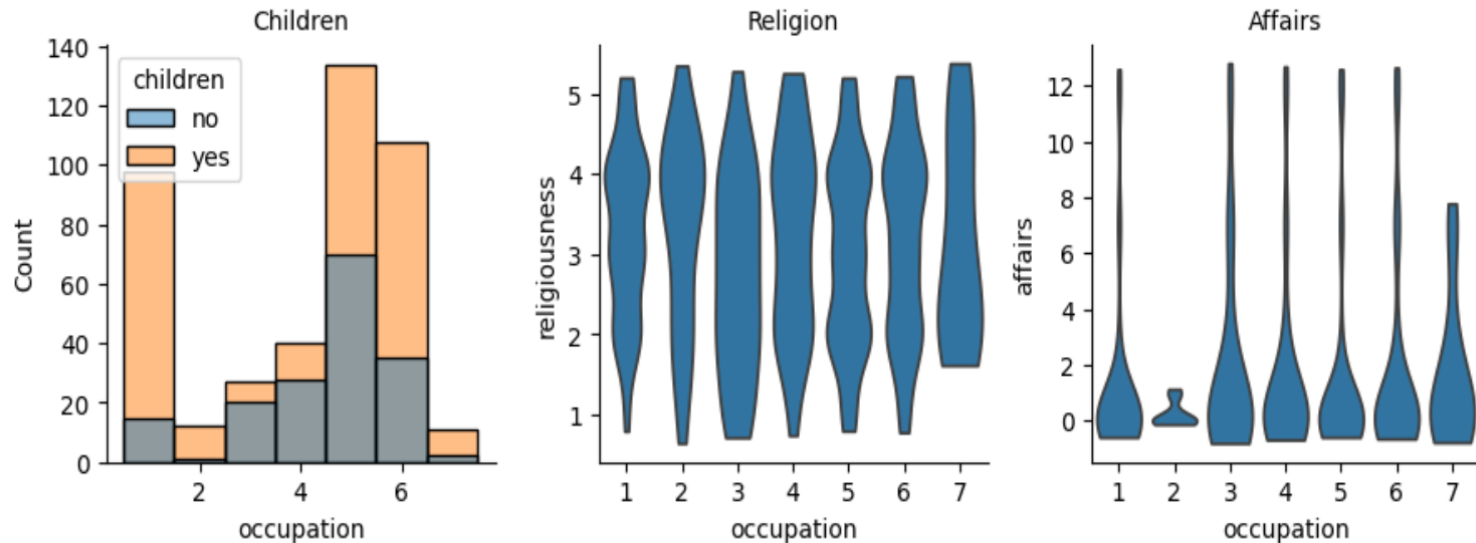
Menschen mit einem höheren Einkommen haben weniger Kinder, sind weniger religiös und haben mehr Affären

Hypothese 5

Menschen mit einem höheren Einkommen haben weniger Kinder, sind weniger religiös und haben mehr Affären

- Keine Informationen über Anzahl der Kinder → Korrelation mit Ja/Nein Antworten bestimmen
- Keine Informationen über Einkommen → Berufsgruppe als äquivalent verwendet

Hypothese 5



- Verteilungen darstellen als Säulen-/Violinendiagramme → Überblick
- Violinplot: stellt Verteilung über y-Achse für Kategorien auf x-Achse dar (vgl.: Bevölkerungsverteilungsdiagramme)
- Ungefähr gleichmäßig verteilt: Korrelationsberechnung für Überprüfung

Hypothese 5

```
# Calculate correlation for non-numerical information about children using Pearson
corr_children: pd.DataFrame = pd.DataFrame(data=affairs_df["children"].str.get_dummies(sep=',').corrwith(affairs_df["occupation"],
                                                                                                     method="spearman"),
                                           columns=["Correlation of Children and Income"])

# Calculate correlations for numerical values
corr_other: pd.DataFrame = affairs_df.corr(method='spearman', numeric_only=True)
```

- Berechnung der Korrelation zwischen den gewünschten Größen: nicht-numerisch und numerisch
- Spearman Korrelation, da Werte nicht normalverteilt sind

Hypothese 5

	Children	Religion	Affairs
Correlation to income	-0.02	-0.02	0.04

- **Ergebnis:** Tendenzen in Richtung der Hypothese
 - Korrelation deutlich zu schwach, um nachgewiesen zu sein
- Hypothese widerlegt

Vielen Dank für Eure und Ihre Aufmerksamkeit!



Technische
Universität
Braunschweig



Gruppe 7: Extramarital Affairs

Dania Arshad, Michael Biermann, Nicole Funke, Tim Matschulla & Agatha Stark