

Palautettava tehtävä 5

June 1, 2025

1 Tehtävä 5: Korrelaatiot ja korrelaatiomatriisi

Tässä notebookissa: 1. Ladataan `Opinnäytetyökysely.xlsx` pandas-DataFrameksi (sheet "Kysely"). 2. **Tehtävä 1:** Lasketaan korrelaatiokertoimet ja selitysasteet kahdelle muuttujaparille: - "Hankin itse aktiivisesti tietoa työni aiheesta" vs. "Tutkimusaiheeni kiinnosti minua" - "Sain riittävästi ohjausta" vs. "Opinnäytetyön tekemiseen kulunut aika (kuukautta)". 3. **Tehtävä 2:** Lasketaan uusi sarake "Kokonaistyytyväisyys ohjaukseen" ohjaustyytyväisyys-kysymyksistä ja tutkitaan korrelaatio "Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan" vs. uusi sarake. 4. **Tehtävä 3:** Luodaan korrelaatiomatriisi seuraaville sarakkeille: - "Olin innostunut opinnäytetyötä tehdessäni" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: aihevalinnassa" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: tutkimuskysymysten rajauksessa" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: menetelmien valinnassa" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: aineiston analyysissä" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: johtopäätösten ja yhteenvedon tekemisessä" - "Ohjaajan tuki opinnäytetyön eri vaiheissa: raportoinnissa"

1.1 1) Kirjastojen tuonti ja Excel-datan lataus

Tuodaan pandas ja luetaan sheet "Kysely" DataFrameen.

Varmista, että `Opinnäytetyökysely.xlsx` on samassa kansiossa kuin tämä notebook (tai päivitä polku muuttujaan `file_path`).

```
[11]: import pandas as pd

# Funktiot Task 1-3
def task1_corr(df: pd.DataFrame) -> None:
    """
    Tehtävä 1: Korrelaatio ja selitysaste
    """
    var1a = "Hankin itse aktiivisesti tietoa työni aiheesta"
    var1b = "Tutkimusaiheeni kiinnosti minua"
    corr1 = df[var1a].corr(df[var1b])
    r2_1 = corr1 ** 2

    var2a = "Sain riittävästi ohjausta"
    var2b = (
        "Opinnäytetyön tekemiseen kulunut aika ensimmäisistä"
        " aihekaavailuista työn valmistumiseen:kuukautta"
    )
```

```

corr2 = df[var2a].corr(df[var2b])
r2_2 = corr2 ** 2

print("Tehtävä 1:")
print(
    f" 1) Corr({var1a}, {var1b}) = {corr1:.2f}, R2 = {r2_1:.2f}"
)
print(
    f" 2) Corr({var2a}, ...aika) = {corr2:.2f}, R2 = {r2_2:.2f}"
)

def task2_corr(df: pd.DataFrame) -> None:
    """
    Tehtävä 2: Uusi sarake ja korrelaatio
    """
    sales_questions = [
        "Ohjaajani panos tuki työtäni",
        "Saamani ohjaus oli asiantuntevaa",
        "Saamani ohjaus oli motivoivaa",
        "Työni ohjaaja vastasi nopeasti tiedusteluihini",
        "Ohjaustilanteet eivät tuntuneet minusta pelottavilta",
        "Ohjaajaani oli helppo lähestyä",
        "Luotin ohjaajani neuvoihin",
    ]
    df["Kokonaistyytyväisyys ohjaukseen"] = df[sales_questions].mean(axis=1)

    var_choice = "Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan"
    var_sat = "Kokonaistyytyväisyys ohjaukseen"
    corr = df[var_choice].corr(df[var_sat])
    r2 = corr ** 2

    print("Tehtävä 2:")
    print(
        f" Corr({var_choice}, {var_sat}) = {corr:.2f}, R2 = {r2:.2f}"
    )

def task3_corr_matrix(df: pd.DataFrame) -> None:
    """
    Tehtävä 3: Korrelaatiomatriisi
    """
    cols = [
        "Olin innostunut opinnäytetyötä tehdessäni",
        "Ohjaajan tuki opinnäytetyön eri vaiheissa:aiheen valinnassa",
        "Ohjaajan tuki opinnäytetyön eri vaiheissa:tutkimuskysymysten tai_
    ➔kehittämistehtävän rajauksessa",
        "Ohjaajan tuki opinnäytetyön eri vaiheissa:menetelmien valinnassa",
        "Ohjaajan tuki opinnäytetyön eri vaiheissa:aineiston analyysissa",
    ]

```

```

        "Ohjaajan tuki opinnäytetyön eri vaiheissa:johtopäätösten ja
        yhteenvedon tekemisessä",
        "Ohjaajan tuki opinnäytetyön eri vaiheissa:raportoinnissa",
    ]
    corr_matrix = df[cols].corr()
    print("Tehtävä 3: Korrelaatiomatriisi")
    display(corr_matrix)

# Ladataan data
file_path = r"D:\GitHub\PythonDataAnalytics\doc\Opinnäytetyökysely.xlsx"

df = pd.read_excel(file_path, sheet_name="Kysely", engine="openpyxl")

# Vahvistetaan, että df on ladattu
df.shape

```

[11]: (242, 42)

1.2 2) Tehtävä 1: Korrelaatiot ja selitysasteet

Kahden muuttujaparin korrelaatiot ja selitysasteet laskettuna: - **Hankin itse aktiivisesti tietoa työni aiheesta** vs. **Tutkimusaiheeni kiinnosti minua**
 - **Sain riittävästi ohjausta** vs. **Opinnäytetyön tekemiseen kulunut aika (kuukautta)**

[12]: task1_corr(df)

Tehtävä 1:

- 1) Corr(Hankin itse aktiivisesti tietoa työni aiheesta, Tutkimusaiheeni kiinnosti minua) = 0.36, $R^2 = 0.13$
- 2) Corr(Sain riittävästi ohjausta, ...aika) = -0.01, $R^2 = 0.00$

1.3 3) Tehtävä 2: “Kokonaistyytyväisyys ohjaukseen” ja korrelaatio

Luodaan sarake “Kokonaistyytyväisyys ohjaukseen” kyselyn ohjaustyytyväisyyttä kuvaavien muuttujien keskiarvoista, ja lasketaan sen korrelaatio muuttujan “**Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan**” kanssa.

[13]: task2_corr(df)

Tehtävä 2:

Corr(Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan, Kokonaistyytyväisyys ohjaukseen) = 0.28, $R^2 = 0.08$

1.4 4) Tehtävä 3: Korrelaatiomatriisi

Luodaan DataFrame sarakkeista: - “Olin innostunut opinnäytetyötä tehdessäni”
 - “Ohjaajan tuki opinnäytetyön eri vaiheissa: aihevalinnassa”
 - “Ohjaajan tuki opinnäytetyön eri vaiheissa: tutkimuskysymysten tai kehittämistehtävän ratkaisemisessa”

- “Ohjaajan tuki opinnäytetyön eri vaiheissa: menetelmien valinnassa”
- “Ohjaajan tuki opinnäytetyön eri vaiheissa: aineiston analyysissä”
- “Ohjaajan tuki opinnäytetyön eri vaiheissa: johtopäätösten ja yhteenvedon tekemisessä”
- “Ohjaajan tuki opinnäytetyön eri vaiheissa: raportoinnissa”

Ja lasketaan näiden muuttujien korrelaatiomatriisi.

```
[14]: task3_corr_matrix(df)
```

Tehtävä 3: Korrelaatiomatriisi

Olin innostunut

```

    opinnäytetyötä tehdessäni \
Olin innostunut opinnäytetyötä tehdessäni      1.000000
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee... 0.425152
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki... 0.275876
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet... 0.270424
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei... 0.261830
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto... 0.184460
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor... 0.181031
    ↪

```

Ohjaajan tuki opinnäytetyön

```

    eri vaiheissa:aiheen valinnassa \
Olin innostunut opinnäytetyötä tehdessäni      0.425152
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee... 1.000000
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki... 0.403532
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet... 0.331662
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei... 0.382156
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto... 0.374488
    ↪
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor... 0.263227
    ↪

```

Ohjaajan tuki opinnäytetyön

```

    eri vaiheissa:tutkimuskysymysten tai kehittämistehtävän rajauksessa \

```

Olin innostunut opinnäytetyötä tehdessäni	0.275876	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee...	0.403532	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki...	1.000000	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet...	0.709703	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei...	0.661632	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto...	0.585724	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor...	0.386080	□
↪		

Ohjaajan tuki opinnäytetyön □

↪eri vaiheissa:menetelmien valinnassa \		
Olin innostunut opinnäytetyötä tehdessäni	0.270424	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee...	0.331662	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki...	0.709703	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet...	1.000000	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei...	0.709834	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto...	0.677735	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor...	0.430803	□
↪		

Ohjaajan tuki opinnäytetyön □

↪eri vaiheissa:aineiston analyysissa \		
Olin innostunut opinnäytetyötä tehdessäni	0.261830	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee...	0.382156	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki...	0.661632	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet...	0.709834	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei...	1.000000	□
↪		
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto...	0.690570	□
↪		

```

Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor...
↪ 0.515566

Ohjaajan tuki opinnäytetyön
↪ eri vaiheissa:johtopäätösten ja yhteenvedon tekemisessä \
Olin innostunut opinnäytetyötä tehdessäni
↪ 0.184460
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee...
↪ 0.374488
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki...
↪ 0.585724
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet...
↪ 0.677735
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei...
↪ 0.690570
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto...
↪ 1.000000
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor...
↪ 0.619559

Ohjaajan tuki opinnäytetyön
↪ eri vaiheissa:raportoinnissa
Olin innostunut opinnäytetyötä tehdessäni
↪ 0.181031
Ohjaajan tuki opinnäytetyön eri vaiheissa:aihee...
↪ 0.263227
Ohjaajan tuki opinnäytetyön eri vaiheissa:tutki...
↪ 0.386080
Ohjaajan tuki opinnäytetyön eri vaiheissa:menet...
↪ 0.430803
Ohjaajan tuki opinnäytetyön eri vaiheissa:ainei...
↪ 0.515566
Ohjaajan tuki opinnäytetyön eri vaiheissa:johto...
↪ 0.619559
Ohjaajan tuki opinnäytetyön eri vaiheissa:rapor...
↪ 1.000000

```

2 Yhteenveto

Jupyter-notebookissa: - Ladataan `Opinnäytetyökysely.xlsx` pandas-DataFrameksi (sheet: “Kysely”). - Tehtävä 1: Korrelaatiot ja selitysasteet kahdelle muuttujaparille: 1) “Hankin itse aktiivisesti tietoa työni aiheesta” vs. “Tutkimusaiheeni kiinnosti minua”
2) “Sain riittävästi ohjausta” vs. “Opinnäytetyön tekemiseen kulunut aika (kuukautta)” - Tehtävä 2: Luodaan uusi sarake **“Kokonaistyytyväisyys ohjaukseen”** ohjaustyytyväisyys-kysymysten keskiarvosta, ja lasketaan sen korrelaatio **“Pystyin itse vaikuttamaan opinnäytetyöni ohjaajan valintaan”** kanssa. - Tehtävä 3: Muodostetaan korrelaatiomatriisi seitsemälle muuttujalle,

jotka kuvaavat innostusta ja ohjaajan tukea eri vaiheissa.