File Edit Selection View Go Run Terminal Help

Statistical Methods and ML models

EXPLORER

Sprint2_Test_Concept2.1.py 1, U | Sprint2_Test_Concept3.ipynb U | ML374_S2_Height_Weight_Data_Concept.xlsx U | Extension: Black Formatter

portal code > day2 > concept > Sprint2_Test_Concept3 > Sprint2_Test_Concept3.ipynb > # ----------

Generate + Code + Markdown | Run All | Restart | Clear All Outputs | Jupyter Variables | Outline | .venv (Python 3.12.10)

OPEN EDITORS
- Sprint2_Test_Concept2.1.py port... 1, U
- Sprint2_Test_Concept3.ipynb port... U
- ML374_S2_Height_Weight_Data_C... U
- Extension: Black Formatter

STATISTICAL METHODS AND ML MODELS
- .venv
- notes
  - day1
  - day2
    - resources
    - session1.md
    - session2.md
    - session3.md
    - session4.md
  - day3
  - day4
  - resources
  - day1.md
- portal code
  - day1
  - day2
    - concept
      - Sprint2_Test_Concept2
      - Sprint2_Test_Concept3
        - ML374_S2_Height_W... U
        - Sprint2_Test_Concept... U
      - Sprint2_Test_Concept1.py U
    - practice
  - day3
- .markdownlint.json
- notes.md
- prob.py
- syllabus.png
- syllabus.txt

Generate + Code + Markdown

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

[1] ✓ 9.2s Python

```
sns.set(style="whitegrid")
```

[2] ✓ 0.0s Python

```
# ----------------------------
# Step 1: Load population data from Excel
# ----------------------------
df = pd.read_excel('ML374_S2_Height_Weight_Data_Concept.xlsx')  # replace with your Excel file
population_heights = df['Height_cm']
print("First few rows of population heights:")
print(population_heights.head())
```

[ ] Python

```
# ----------------------------
# Step 2: CLT parameters
# ----------------------------
sample_size = 50          # size of each sample
num_samples = 450         # number of samples to draw
random_seed = 42          # fixed seed for reproducibility
```

[4] ✓ 0.0s Python

```
# ----------------------------
# Step 3: Draw samples and compute means (with fixed seed)
# ----------------------------
sample_means = []
```

Rainy days ahead 31°C | Search | ENG IN | 15:54 20-11-2025

File Edit Selection View Go Run Terminal Help

Statistical Methods and ML models

EXPLORER

OPEN EDITORS

GROUP 1

Sprint2_Test_Concept2.1.py port... 1, U

Sprint2_Test_Concept3.ipynb port... U

ML374_S2_Height_Weight_Data_C... U

Extension: Black Formatter

GROUP 2

Sprint2_Test_Concept3.ipynb port... U

STATISTICAL METHODS AND ML MODELS

.venv

notes

day1

day2

resources

session1.md

session2.md

session3.md

session4.md

day3

day4

resources

day1.md

portal code

day1

day2

concept

Sprint2_Test_Concept2

Sprint2_Test_Concept3

ML374_S2_Height_W... U

Sprint2_Test_Concept... U

Sprint2_Test_Concept1.py U

practice

day3

.markdownlint.json

notes.md

prob.py

syllabus.png

syllabus.txt

Sprint2_Test_Concept2.1.py 1, U | Sprint2_Test_Concept3.ipynb U | ML374_S2_Height_Weight_Data_Concept.xlsx U | Extensior

portal code > day2 > concept > Sprint2_Test_Concept3 > Sprint2_Test_Concept3.ipynb > # ----------

Generate + Code + Markdown | Run All | Restart | Clear All Outputs | Jupyter Variables | .venv (Python 3.12.10)

```
# ----------------------------
# Step 2: CLT parameters
# ----------------------------
sample_size = 50          # size of each sample
num_samples = 450         # number of samples to draw
random_seed = 42          # fixed seed for reproducibility
```

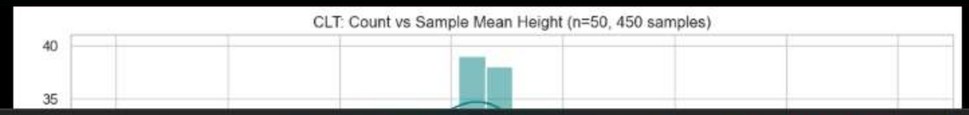
[4] 0.0s Python

```
# ----------------------------
# Step 3: Draw samples and compute means (with fixed seed)
# ----------------------------
sample_means = []

rng = np.random.default_rng(random_seed)  # NumPy random generator for reproducibility

for _ in range(num_samples):
    # sample with fixed random generator
    sample = population_heights.sample(n=sample_size, random_state=rng.integers(0, 100000))
    sample_mean = sample.mean()
    sample_means.append(sample_mean)

sample_means = np.array(sample_means)
```

[5] 0.2s Python

```
# ----------------------------
# Step 4: Plot histogram (Count vs Sample Mean) with KDE
# ----------------------------
plt.figure(figsize=(12,6))
sns.histplot(sample_means, bins=30, kde=True, color='teal')
plt.title(f"CLT: Count vs Sample Mean Height (n={sample_size}, {num_samples} samples)")
plt.xlabel("Sample Mean of Height_cm")
plt.ylabel("Count")
plt.show()
```

[8] 0.2s Python



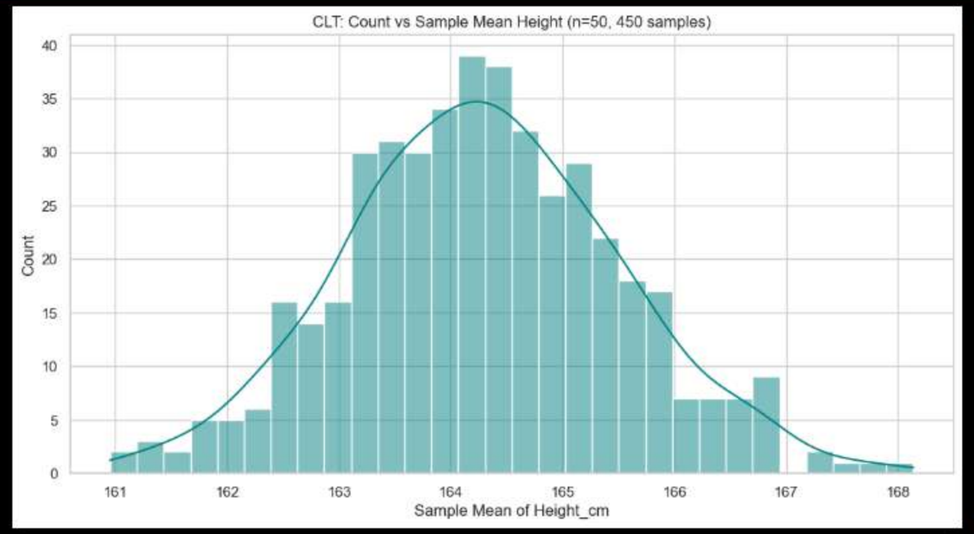
Sprint2_Test_Concept3.ipynb U

portal code > day2 > concept > Sprint2_Test_Concept3 > Sprint2_Test_Concept3.ipynb > # ----------

Generate + Code + Markdown | Run All | Restart | Clear All Outputs | Jupyter Variables | .venv (Python 3.12.10)

```
plt.ylabel("Count")
plt.show()
```

[8] 0.2s Python




```
# ----------------------------
# Step 5: Print summary statistics
# ----------------------------
print(f"Population mean: {population_heights.mean():.2f}")
print(f"Population std deviation: {population_heights.std():.2f}")
print(f"Sampling distribution mean: {sample_means.mean():.2f}")
print(f"Sampling distribution std deviation (Standard Error): {sample_means.std():.2f}")
```

[7] 0.0s Python

```
Population mean: 164.18
Population std deviation: 8.63
Sampling distribution mean: 164.28
Sampling distribution std deviation (Standard Error): 1.22
```

1 cm of rain
Saturday

Search

ENG
IN

15:55
20-11-2025