File Edit Selection View Go Run Terminal Help

Statistical Methods and ML models

EXPLORER

OPEN EDITORS

Sprint2_Test_Concept2.1.ipynb

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
        - ML374_52_Height_W...
        - Sprint2_Test_Concept...
      - Sprint2_Test_Concept1.py
    - practice
  - day3
- .markdownlint.json
- notes.md
- prob.py
- syllabus.png
- syllabus.txt

Sprint2_Test_Concept2.1.ipynb

portal code > day2 > concept > Sprint2_Test_Concept2 > Sprint2_Test_Concept2.1.ipynb > # PART 1 - Comare population with sample

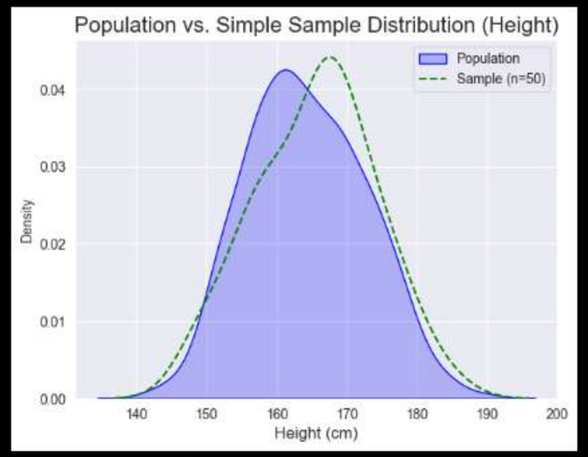
Generate | + Code | + Markdown | Run All | Restart | Clear All Outputs | Jupyter Variables | Outline | ...

.venv (Python 3.12.10)

```python
# PART 1 - Comare population with sample
if len(population_df) < SIMPLE_SAMPLE_SIZE:
    print(f"Not enough rows to sample {SIMPLE_SAMPLE_SIZE} rows. Total rows: {len(population_df)}")
else:
    simple_sample_df = population_df.sample(n=SIMPLE_SAMPLE_SIZE, random_state=RANDOM_SEED)
    print("\nPreview of the sample dataset:")
    simple_sample_df.head()

    sns.kdeplot(population_df['Height_cm'],
                color="blue",
                label='Population',
                fill=True,  # Fill the population curve
                common_norm=False)

    # Overlay the sample distribution (Line only)
    sns.kdeplot(simple_sample_df['Height_cm'],
                color="green",
                label=f'Sample (n={SIMPLE_SAMPLE_SIZE})',
                fill=False, # Do not fill the sample curve
                linestyle='--', # Add dashed line to match sample
                common_norm=False)

    plt.title('Population vs. Simple Sample Distribution (Height)', fontsize=16)
    plt.xlabel('Height (cm)', fontsize=12)
    plt.ylabel('Density') # Kdeplot shows density, not count
    plt.legend()
    plt.show()
```

[55] ✓ 0.1s Python

Preview of the sample dataset:



Trending videos
Puppy Discovers...

Search

ENG
IN

11:57
20-11-2025

File Edit Selection View Go Run Terminal Help

Statistical Methods and ML models

Sprint2_Test_Concept2.1.ipynb U

portal code > day2 > concept > Sprint2_Test_Concept2 > Sprint2_Test_Concept2.1.ipynb > # PART 1 - Comare population with sample

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


```
# PART 2 - Increasing sample size to see effect on distribution
sample_sizes = [5, 15, 25, 35]
fig, axes = plt.subplots(2, 2, figsize=(12, 10))
axes = axes.flatten()
sns.set_style("darkgrid") # Match the style for this specific chart
fig.suptitle('Distribution Shape by Increasing Sample Size', fontsize=18)
for i, size in enumerate(sample_sizes):
    ax = axes[i]

    if len(population_df) < size:
        ax.set_title(f"Sample Size = {size} (Not enough data)")
        continue
    sample_df = population_df.sample(n=size, random_state=RANDOM_SEED)

    # This plot (histplot with KDE) already matches your screenshot
    sns.histplot(sample_df['Height_cm'],
        ax=ax,
        kde=True,
        stat='density',
        common_norm=False)

    ax.set_title(f'Sample Size = {size}')
    ax.set_xlabel('Height (cm)')
    ax.set_ylabel('Density')

plt.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()
```

[53] ✓ 0.4s Python

Distribution Shape by Increasing Sample Size

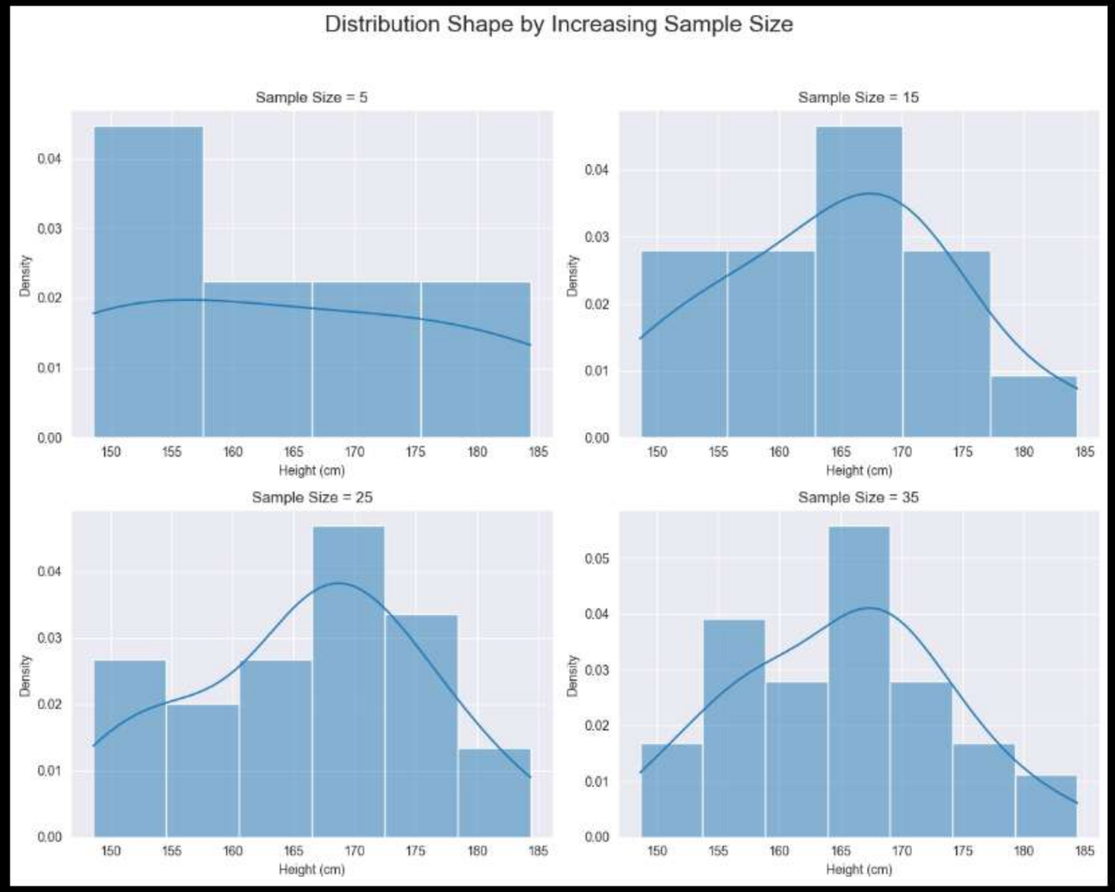
Sprint2_Test_Concept2.1.ipynb U

portal code > day2 > concept > Sprint2_Test_Concept2 > Sprint2_Test_Concept2.1.ipynb > # PART 1 - Comare population with sample

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

```
plt.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()
```

[53] ✓ 0.4s Python



29°C Mostly sunny

Search

ENG IN

11:58 20-11-2025