FOUNDATION LESSON

WEEK 5 · LESSON 11

Descriptive statistics and explicit conventions

A statistic summarises a dataset under an explicit counting and interpolation convention.

THE MENTAL MODEL

MENTAL MODEL

The mental model

A statistic summarises a dataset under an explicit counting and interpolation convention.

MEAN, MEDIAN AND TWO SD CONVENTIONS

CODE WALKTHROUGH W05-L11-S03

```
01  import pandas as pd
02  s=pd.Series([10.0,20.0,30.0])
03  print("Mean:",s.mean(),"Median:",s.median())
04  print("Population SD:",round(s.std(ddof=0),3))
05  print("Sample SD:",s.std(ddof=1))
```

WHAT TO TRACK

Mean, median and two SD conventions

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

MEAN, MEDIAN AND TWO SD CONVENTIONS · RESULT

CODE WALKTHROUGH W05-L11-S04

```
01  import pandas as pd
02  s=pd.Series([10.0,20.0,30.0])
03  print("Mean:",s.mean(),"Median:",s.median())
04  print("Population SD:",round(s.std(ddof=0),3))
05  print("Sample SD:",s.std(ddof=1))
```

WHAT TO TRACK

Mean, median and two SD conventions

VALIDATE THE RESULT

```
Mean: 20.0 Median: 20.0
Population SD: 8.165
Sample SD: 10.0
```

QUANTILES WITH LINEAR INTERPOLATION

CODE WALKTHROUGH W05-L11-S05

```
01  import pandas as pd
02  s=pd.Series([0,10,20,30])
03  print(s.quantile([.25,.5,.75],interpolation="linear"))
04  print("Observed:",s.count())
```

WHAT TO TRACK

Quantiles with linear interpolation

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

QUANTILES WITH LINEAR INTERPOLATION · RESULT

CODE WALKTHROUGH W05-L11-S06

```
01  import pandas as pd
02  s=pd.Series([0,10,20,30])
03  print(s.quantile([.25,.5,.75],interpolation="linear"))
04  print("Observed:",s.count())
```

WHAT TO TRACK

Quantiles with linear interpolation

VALIDATE THE RESULT

```
0.25     7.5
0.50    15.0
0.75    22.5
dtype: float64
Observed: 4
```

MISSING OBSERVATIONS CHANGE THE DENOMINATOR

CODE WALKTHROUGH W05-L11-S07

```
01  import pandas as pd
02  s=pd.Series([10.0,None,30.0])
03  print("Rows:",len(s),"Observed:",s.count())
04  print("Observed mean:",s.mean())
05  print("Sample variance:",s.var(ddof=1))
```

WHAT TO TRACK

Missing observations change the denominator

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

MISSING OBSERVATIONS CHANGE THE DENOMINATOR · RESULT

CODE WALKTHROUGH W05-L11-S08

```
01  import pandas as pd
02  s=pd.Series([10.0,None,30.0])
03  print("Rows:",len(s),"Observed:",s.count())
04  print("Observed mean:",s.mean())
05  print("Sample variance:",s.var(ddof=1))
```

WHAT TO TRACK

Missing observations change the denominator

VALIDATE THE RESULT

```
Rows: 3 Observed: 2
Observed mean: 20.0
Sample variance: 200.0
```

GUIDED PRACTICE

LEARNER PRACTICE

Guided practice

For [2,4,6], calculate mean, sample variance and sample SD with ddof=1.

PREDICT · MODIFY · RUN

01 Restate the supplied rules.

02 Predict the result.

03 Modify and run the code.

04 Compare actual output.

05 Explain any difference.

INDEPENDENT PRACTICE

LEARNER PRACTICE

Independent practice

For [0,20], compute the 25th percentile using linear and nearest interpolation and explain their difference.

PREDICT · MODIFY · RUN

01 Restate the supplied rules.

02 Predict the result.

03 Modify and run the code.

04 Compare actual output.

05 Explain any difference.

CHECK YOUR UNDERSTANDING

MENTAL MODEL

Check your understanding

Diagnostic: Must sample SD equal population SD on the same values? No. Their denominators differ when ddof differs.

RECAP AND NEXT STEP

Recap and next step

Report counts, define the statistic, state ddof and quantile interpolation, and retain measurement units.

UNDERSTAND APPLY VALIDATE EXPLAIN