PSPP Assignment

Robert

2025-06-21

PSPP Import Syntax f

```
GET DATA
  /TYPE=TXT
  /FILE="/home/dragon/Downloads/week4_pspp_data.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /QUALIFIER='"'
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=2
  /VARIABLES=
    ID                F8.0
    Gender            A10
    Age               F8.0
    Education_Level   A20
    Satisfaction_Score F8.1
    Monthly_Income    F8.2.
```

Notes:

- `F8.0` means numeric field (max width 8 digits, 0 decimals).
- `F8.1` or `F8.2` means float with 1 or 2 decimal places.
- `A10`, `A20` are string fields with up to 10 or 20 characters.

A. Understanding the Dataset What are the types of variables in the dataset (categorical vs. continuous)?

Variable Type Breakdown

Variable	Type	Reasoning
ID	*Categorical (nominal)* or *Identifier*	It's just a unique identifier for each row, not used in analysis
Gender	**Categorical (nominal)**	Values: "Male", "Female"

Variable	Type	Reasoning
Age	**Continuous (ratio)**	Numeric with meaningful zero, e.g., age 0 is meaningful
Education_Level	**Categorical (ordinal)**	Values have a ranked order: Primary < Secondary < Tertiary
Satisfaction_Score	**Categorical (ordinal)**	Score from 1 to 5 suggests ordered levels of satisfaction
Monthly_Income	**Continuous (ratio)**	Numeric, measured income values with true zero

---

Summary

- **Categorical (Nominal)**: `Gender`, `ID`
- **Categorical (Ordinal)**: `Education_Level`, `Satisfaction_Score`
- **Continuous (Ratio)**: `Age`, `Monthly_Income`

Which variable(s) are best suited for calculating frequencies?

Variable	Frequency Suitability	Why
Gender	Excellent	Only a few categories (Male, Female)
Education_Level	Excellent	Ordered categories (Primary, Secondary, Tertiary)
Satisfaction_Score	Good	Discrete numeric scores (1–5); behaves like ordinal categories
Age	*Sometimes*	Only if grouped into ranges (e.g., 20–29, 30–39…)
Monthly_Income	Poor	Too many unique values (continuous)
ID	Useless	Every row is unique; not meant for analysis

Which variable(s) are best suited for calculating means and standard deviations?

Variable	Mean/SD Suitability	Reason
Age	Excellent	Continuous numeric; makes sense to calculate average and spread
Monthly_Income	Excellent	Ratio-scale variable; meaningful zero and spread
Satisfaction_Score	*Sometimes*	Technically ordinal, but often treated as numeric in practice
Gender	No	Categorical; mean doesn't make sense
Education_Level	No	Ordinal; ranking exists, but not equal intervals
ID	No	Just a unique identifier — no statistical meaning

B. Running Descriptive Statistics in PSPP

Run frequency tables for the following variables:

```
- # Gender
```

```pspp
FREQUENCIES VARIABLES=Gender Education_Level Satisfaction_Score.
```

Statistics

Satisfaction_Score

N Valid	1000
Missing	0
Mean	3.03
Std Dev	1.44
Minimum	1.0
Maximum	5.0

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	510	51.0%	51.0%	51.0%
	Male	490	49.0%	49.0%	100.0%
Total		1000	100.0%		

- # Education_Level

Education_Level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Primary	301	30.1%	30.1%	30.1%
	Secondary	400	40.0%	40.0%	70.1%
	Tertiary	299	29.9%	29.9%	100.0%
Total		1000	100.0%		

- #Satisfaction_Score

Satisfaction_Score

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.0	212	21.2%	21.2%	21.2%
	2.0	171	17.1%	17.1%	38.3%
	3.0	204	20.4%	20.4%	58.7%

```
      4.0       201   20.1%         20.1%              78.8%
      5.0       212   21.2%         21.2%             100.0%

Total          1000  100.0%
```

Question: What is the most common education level in the dataset?

Education Level	Frequency
Primary	301
Secondary	**400** ← Highest
Tertiary	299

Compute descriptive statistics (mean, median, standard deviation, min, max) for these variables:

- Age

- Monthly_Income

```
                    Descriptive Statistics

                      N    Mean   Std Dev Minimum  Maximum

Age                1000    39.08   12.31      18       60
Monthly_Income     1000 30233.42 7979.69 7252.00 51474.00
Valid N (listwise) 1000
Missing N (listwise)  0
```

Question: What is the average monthly income of participants?

Question: What is the age range of participants?

Metric	Value
Average Monthly Income	$30,233.42
Age Range	42 years
Min Age	18
Max Age	60

C. Interpretation & Reporting

Are there any patterns between Education_Level and Satisfaction_Score?

Optional Task: Create a cross-tabulation (CROSSTABS) of Education_Level vs Satisfaction_Score.

```
Education_Level × Satisfaction_Score
```

		Satisfaction_Score					
		1.0	2.0	3.0	4.0	5.0	Total
Education_Level	Primary	Count	78	46	54	59	64	301
		Row %	25.9%	15.3%	17.9%	19.6%	21.3%	100.0%
		Column %	36.8%	26.9%	26.5%	29.4%	30.2%	30.1%
	Secondary	Count	74	82	83	83	78	400
		Row %	18.5%	20.5%	20.8%	20.8%	19.5%	100.0%
		Column %	34.9%	48.0%	40.7%	41.3%	36.8%	40.0%
	Tertiary	Count	60	43	67	59	70	299
		Row %	20.1%	14.4%	22.4%	19.7%	23.4%	100.0%
		Column %	28.3%	25.1%	32.8%	29.4%	33.0%	29.9%
	Total	Count	212	171	204	201	212	1000
		Row %	21.2%	17.1%	20.4%	20.1%	21.2%	100.0%
		Column %	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Do males and females report similar average incomes?

Optional Task: Split the dataset by Gender and compute mean Monthly_Income.

What percentage of participants fall in each Satisfaction_Score level?

Satisfaction_Score	Count	Percentage
1.0	212	**21.2%**
2.0	171	**17.1%**
3.0	204	**20.4%**
4.0	201	**20.1%**
5.0	212	**21.2%**

Task: Interpret the frequency distribution of Satisfaction_Score.

1. What percentage of participants fall in each Satisfaction_Score level?

You can get this directly from the **"Total Row %"** line at the bottom of the table:

Satisfaction_Score	Count	Percentage
1.0	212	**21.2%**
2.0	171	**17.1%**
3.0	204	**20.4%**
4.0	201	**20.1%**
5.0	212	**21.2%**

Answer:

- The **most common scores** are **1.0 and 5.0**, each with **21.2%** of participants.
- The **least common score** is **2.0** (17.1%).

---

2. Task: Interpret the frequency distribution of Satisfaction_Score

Interpretation:

- The distribution of Satisfaction Scores is **roughly balanced** across all levels.
- However, there's a slight **U-shape**:
  - **More participants selected the extremes (1 and 5)**
  - **Fewer selected mid-level satisfaction (especially 2)**

This suggests:

- Participants tended to feel **either very dissatisfied or very satisfied**.
- Fewer reported "moderate dissatisfaction" (2.0), possibly indicating a **polarized experience** with the service/product being evaluated.

Summary:

Insight	Explanation
Highest % Satisfaction Scores	1.0 and 5.0 (both 21.2%)
Lowest % Satisfaction Score	2.0 (17.1%)
Interpretation of Distribution	Polarized responses; fewer moderate opinions
Implication	Mixed satisfaction experience among users
