

## *SPSS*

→ .csv = Comma Separated Values

→ .sav = SPSS Data File

→ .xlsx → Excel Spreadsheet

Exam question:

1. Frequency distribution of age
2. Bar chart
- 3.

Student List List

/ID Name Age

BEGIN DATA

101 Habib 25

202 Sabbir 28

303 Himo 25

302 Rafi 25

→ spss এর তিনটা window আছে

1. Data View window
2. Variable View window
3. Output window

Data file → .sav (data.sav)

Output file → .spo (data.spo)

Syntax file → .sps (data.sps)

## To open an SPSS data file

**Manually:** File → Open → Data → Browse file from source

**Syntax:** GET FILE ="path"

### Example:

```
GET FILE='C:\SPSS Training\DayI\Obesity.sav'.
```

```
GET SAS FILE="C:\SPSS Training\Day I\ Obesity.sas7bdat".
```

```
GET STATA FILE="C:\SPSS Training\Day I\ Obesity.dta".
```

## Open Excel files in SPSS:

**Manually:**

File → Open → Data → Browse Excel file → OK

**Syntax:**

```
GET DATA /TYPE=XLSX
```

```
/FILE='C:\SPSS Training\DayI\Obesity_sample2.xlsx'
```

```
/SHEET=name Obesity_sample2'
```

```
/CELLRANGE=full /READNAMES=on
```

```
/ASSUMEDSTRWIDTH=32767.
```

```
EXECUTE.
```

## COMPUTE Function:

Suppose we want to calculate standardized value of LIFESPAN (X). The formula is,

$$Z = \frac{X - \text{mean}(X)}{\text{Standard deviation}(X)}$$

Suppose mean and standard deviation of LIFESPAN are 835 and 275.

**Manually**

Transform Compute Variable > Define target variable

(STD\_LIFESPAN) > ((LIFESPAN-835)/275)" > OK

**Syntax:**

COMPUTE STD\_LIFESPAN=(LIFESPAN-835)/275.

EXECUTE.

**Recode into different variable:**

**Syntax:**

RECODE LIFESPAN (Lowest thru 299=1) (300 thru 599=2) (600 thru 899=3) (900 thru 1199=4) (1200 thru Highest=5) INTO CAT\_LIFESPAN.

VARIABLE LABELS CAT\_LIFESPAN 'Distribution of Lifespan'.

VALUE LABELS CAT\_LIFESPAN 1 "<300" 2 "300-600" 3 "600-900" 4 "900-1200"  
5 " $\geq$ 1200".

EXECUTE.

**Merge files: Add Cases**

**Syntax:**

Get FILE='G:\Data File\data1.sav'.

ADDFILES /FILE=\*

/FILE='G:\Data File\data2.sav'.

SAVE OUTFILE 'G:\Data File\data3.sav'.

EXECUTE.

**Merge files: Add Variables**

**Syntax:**

GET FILE= "D:\spss\dads.sav".

SORT CASES BY famid.

SAVE OUTFILE="D:\spss\dads2.sav".

GET FILE="D:\spss\faminc.sav".

SORT CASES BY famid.

SAVE OUTFILE="D:\spss\faminc2.sav".

MATCH FILES FILE="D:\spss\dads2.sav"  
/FILE="D:\spss\faminc2.sav"  
/BY famid. SAVE OUTFILE="D:\spss\OnetoOneMerge.sav"

### **Merge files: One to Many Merge**

GET FILE="D:\spss\dads.sav".  
SORT CASES BY famid.  
SAVE OUTFILE="D:\spss\dads2.sav".  
GET FILE="D:\spss\kids.sav".  
SORT CASES BY famid.  
SAVE OUTFILE="D:\spss\kids2.sav".

MATCH FILES FILE="D:\spss\kids2.sav"  
/Table="D:\spss\dads2.sav"  
/BY famid.

SAVE OUTFILE="D:\spss\OnetoManyMerge.sav".

### **Aggregate Statistics:**

Find mean age and weight of children for each family.

#### **Syntax:**

AGGREGATE  
/OUTFILE=\* MODE=ADDVARIABLES  
/BREAK=famid  
/age\_mean\_1=MEAN(age)  
/wt\_mean\_1=MEAN(wt)

→ Dialog box এর last step এ “OK” এর পরিবর্তে “Paste” চাপলে সেই operation এর syntax code syntax file এ লেখা হয়ে যায়। এরপর syntax file থেকে Run করলে output file এ output টা দেখতে পারা যাবে।

এর ফলে কোন operation এর syntax code প্রয়োজন হলে সেখান থেকে নেওয়া যাবে।

→ 1 inch = 0.0254 m

→ 1 kg = 2.20462 pound

### **Correlation Analysis:**

Karl Pearson's correlation coefficient ( $r$ )

Spearman's rank correlation coefficient ( $\rho$ )

## **STATA**

→ command এ br লিখলে import করা data file দেখাবে

→ gen: For creating new variable

gen loss = STARTWEIGHT- ENDWEIGHT

→ Recode Command:

Syntax: old values = new value

Example: recode loss (0/17=0) (18/25=1) , gen(loss\_cat)

The gen option tells recode to create a new variable (loss\_cat) to store the results. If you don't include a gen option, recode will change the original variable.

→ Variable labels:      label variable loss\_cat “loss of weight”

→ Value labels:

label define los 0 “less than 17” 1 “more than 17” label values loss\_cat los

- Rename command: loss\_cat loss\_category
- describe: dataset এর সবগুলো variable এর একটা সাধারণ বিবরণী পাওয়ার জন্য ব্যবহার হয়
- Codebook: ভেরিয়েবলগুলোর নাম, লেবেল এবং ডেটা পরীক্ষা করে ডেটাসেটের একটি বর্ণনামূলক কোডবুক তৈরি করে।
- summarize: minimum, maximum, std, mean এই সব দিবে
- by: প্রতিটা group এর জন্য আলাদা ভাবে কোন operation run করার জন্য

Syntax: by varname: command

Example: by gender: summarize income  
 bysort mpg: summarize price (একই সাথে sort ও করবে)

→ sort: data কে sort করার জন্য

sort mpg

→ if = do this only for observations that satisfy this condition  
 যেসব observation এই condition পূরণ করে, শুধু সেগুলোর ক্ষেত্রেই এটি প্রয়োগ করো

Summarize price if foreign == 1

→ in = do this only for these observation numbers

Syntax: command varlist in range

Example:

sort price

summarize mpg in 1/10

→ tab = “show a table of counts (and sometimes percentages)”

Example: tab foreign

→ help command : লিখলে এই command কি কাজ করে সেটা দেখায়

help tab

→ table = “create a summary table of statistics for groups of data”

Exaple:           table loss\_category, contents( freq mean loss)

→ drop if var1==1

Example: drop gender =1

→ dataset থেকে n সংখ্যক data sample হিসাবে নেওয়ার জন্য

set seed 200304038

sample n, count

n সংখ্যক data রেখে বাকি সবগুলো data file থেকে delete করে দেবে

→if you want 120 observations stratified by gender

set seed 200304038

bysort gender: sample 120, count

60