WEEK 2 REVISION SHEET: Confidence, Significance & Sampling Distributions

🔷 KEY CONCEPTS TO REMEMBER

1. Population vs. Sample

- **Population**: The full group you want to study.
- **Sample**: A smaller group drawn from the population.
- We estimate the population **parameter** (e.g. μ, σ) using a **sample statistic** (e.g. $\bar{x}$, s).

2. Sampling Error

- **Random Error**: Unpredictable variation between samples. Affects precision but not bia
- **Systematic Error (Bias)**: Consistent over- or underestimation. Affects **accuracy**.
- Always assume random error in well-designed studies (i.e. no bias).

3. Sampling Distribution

- The distribution of a sample statistic (like the sample mean $\bar{x}$) across many samples.
- **Central Limit Theorem**: Regardless of population distribution, the sampling distribution of the mean becomes **normal** if the sample size is large enough.
- Properties:
  - Mean of sampling distribution = **population mean (μ)**
  - SD of sampling distribution = **Standard Error (SE)** = $\sigma / \sqrt{n}$

4. Standard Error (SE)

- Measures how much sample means vary.
- Formula:

```ini
SE = s / √n
```

- Larger samples → smaller SE → more precision

5. Confidence Intervals (CI)

- Range in which we believe the true population parameter lies.
- Formula:

```ini
CI = x̄ ± Z * SE
```

- Common Z-values:
  - 90% → 1.65
  - 95% → 1.96
  - 99% → 2.58

6. Confidence Interval Interpretation

- 95% CI means: "If we repeated the sampling many times, 95% of the intervals would contain the true population mean."
- Wider CI → more confidence, but less precision.

💻 SPSS COMMANDS TO REMEMBER FOR THE EXAM

✅ To Get Descriptive Statistics & Histogram

- Go to: `Analyze > Descriptive Statistics > Frequencies`
- Add your numerical variable (e.g. mobtime)
- **Untick** "Display frequency tables"
- Click **Statistics**, tick: mean, median, mode, SD, min, max
- Click **Charts**, choose: Histogram + Show normal curve

✅ To Calculate Confidence Interval in SPSS

- Go to: `Analyze > Descriptive Statistics > Explore`
- Add variable to **Dependent List**
- Click **Statistics** → Change **Confidence Interval** (default is 95%, you can change it to 90% or 99%)

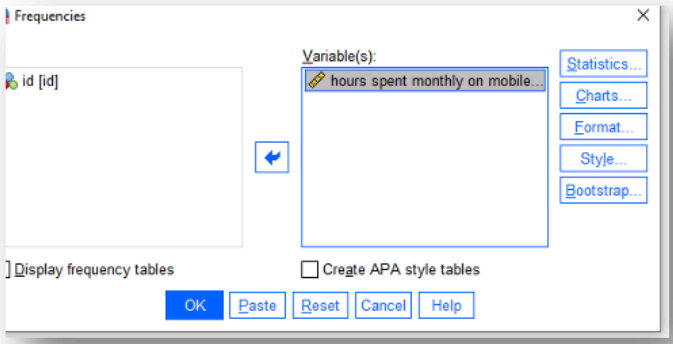
✅ To Get Descriptive Statistics & Histogram:

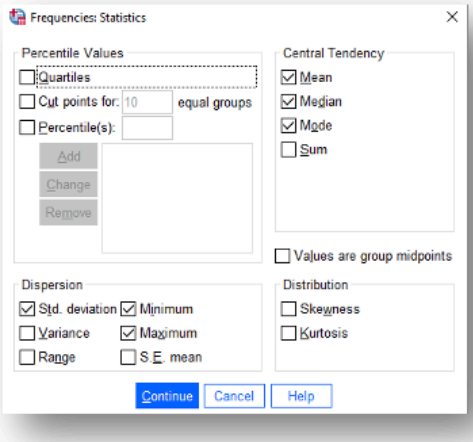
That step is used to:Explore the distribution of a numerical (continuous) variable

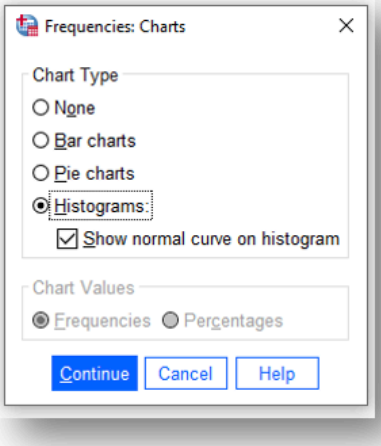
specifically, to summarise it with key statistics and to visualise it with a histogram.

> We used **Analyse → Descriptive Statistics → Frequencies** and put the variable *mobtime* in the **Variable(s)** tab. We untick **Display** frequencies tables as our variable is numerical continuous. In **Statistics**, we tick **mean**, **median**, **mode**, **min**, **max** and **standard deviation**. We also choose **histogram** in **Charts** and ticked **Show normal curve** (please find screenshots below).
>
> The variable *mobtime* is a numerical continuous variable. The variable is normally distributed (bell shaped and symmetrically distributed around its mean).

*Please note that you will get slightly different values in the **My Sample** column.



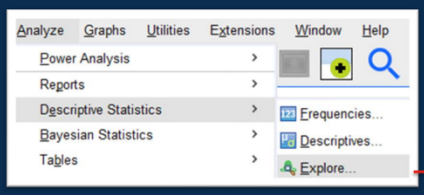


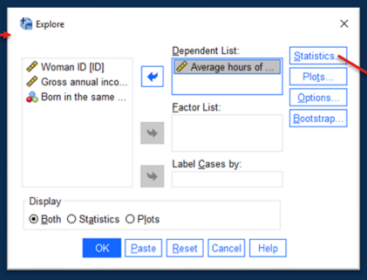


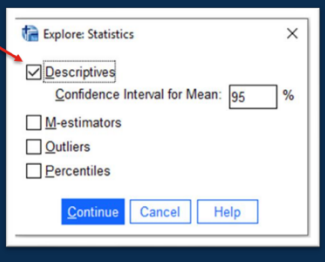
✅ To Calculate Confidence Interval in SPSS

SPSS slide: 'how to'

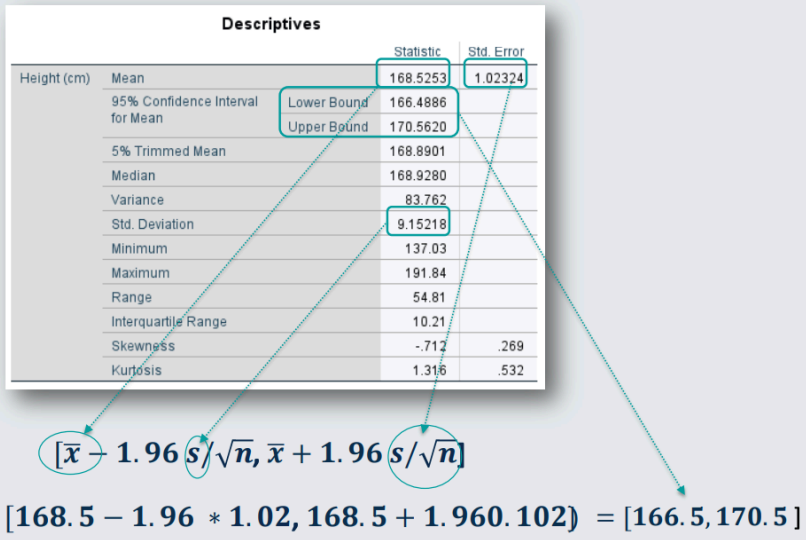
Analyse-> Descriptive Statistics-> Explore -> put the variable in 'Dependent list'-> Statistics-> Change the CI if you want to.







Interpretation slide



Descriptives

			Statistic	Std. Error
Height (cm)	Mean		168.5253	1.02324
	95% Confidence Interval for Mean	Lower Bound	166.4886	
		Upper Bound	170.5620	
	5% Trimmed Mean		168.8901	
	Median		168.9280	
	Variance		83.762	
	Std. Deviation		9.15218	
	Minimum		137.03	
	Maximum		191.84	
	Range		54.81	
	Interquartile Range		10.21	
	Skewness		-.712	.269
	Kurtosis		1.316	.532

$$[\bar{x} - 1.96\, s/\sqrt{n}, \bar{x} + 1.96\, s/\sqrt{n}]$$

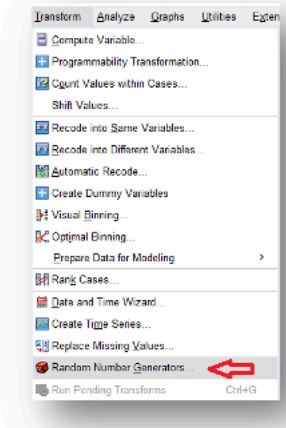
$$[168.5 - 1.96 * 1.02, 168.5 + 1.960.102) = [166.5, 170.5]$$

✅ To Select a Random Sample

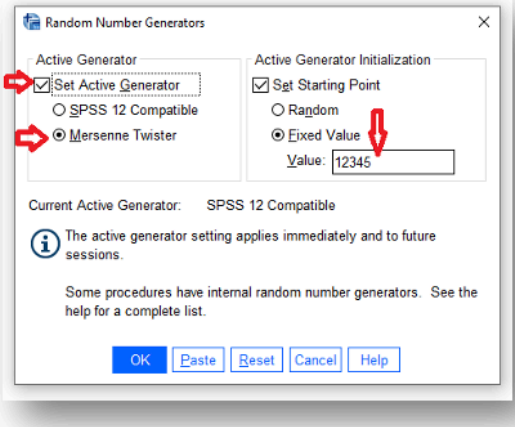
- Set random seed:
  - `Transform > Random Number Generators` → enter a seed (e.g. 24893)
- Select sample:
  - `Data > Select Cases > Random sample of cases`
  - Choose: “Exactly 200 cases from the first 2000”
  - SPSS will create a filter variable

You will now create your very own random sample. Following the guide below, please set a 'seed' (the starting point for the random number generator). Please create your seed as a random number with 5 digits at least, different to that of the picture and your classmates' (as far as you know). Take a note of this number in the box below.

a. Go to **Transform → Random Number Generators**



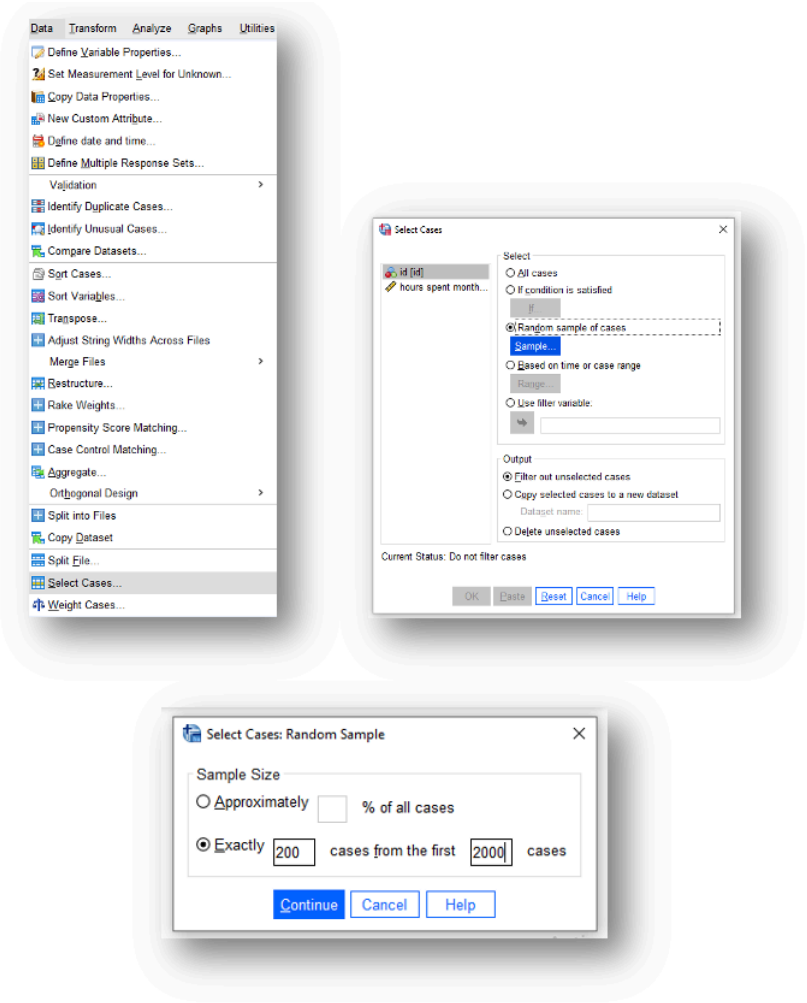
b. Click the boxes as shown in the picture below and put your own random seed in the **Value** box (please do not use 12345, as in the example, and use a number with at least 5 digits).



Your seed value is: 12345

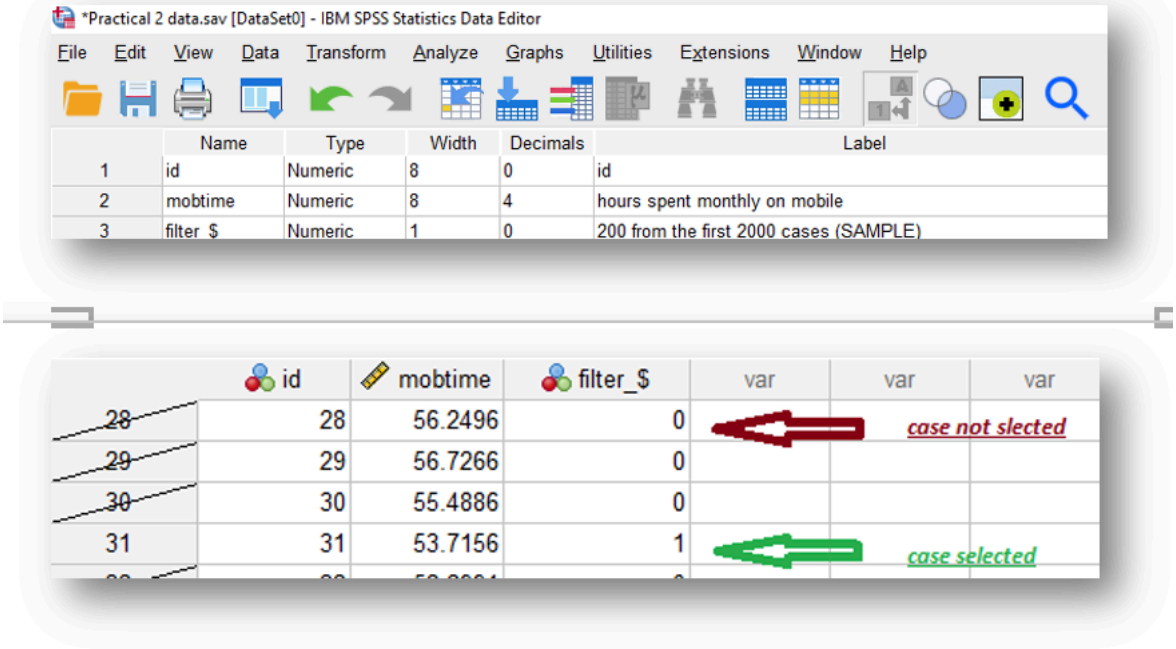
You are now ready to create your very own random sample, as if you were sampling from the population for a research project. Use the guidelines below to randomly select 200 cases from the (pretend) population of 2000.

c. Go to **Data → Select cases → Random sample of cases** and click **Sample.** Put **Exactly** 200 **cases from the first** 2000**.** Press **continue** and then **ok.**



SPSS has now created a **filter** variable to indicate which of the cases were selected for your random sample.

SPSS has now created a **filter** variable to indicate which of the cases were selected for your random sample.

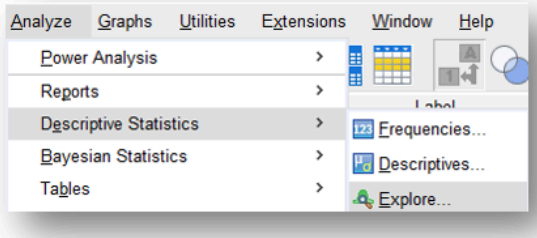


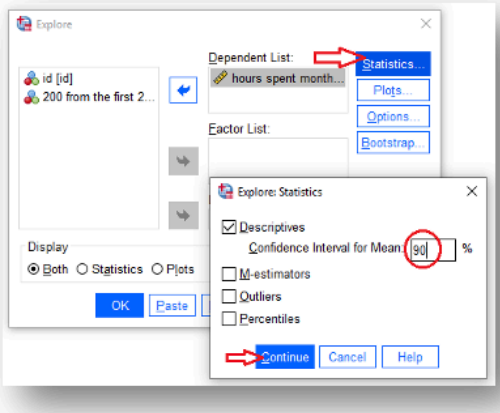
*Please note that in your sample different cases will be selected than this example.

✅**Compute analytically the 90% Confidence Interval**

The 90% CI is given by the formula [$\bar{x}$-1.65*SE,$\bar{x}$+1.65*SE]. Substituting my sample's values into the formula, the 90% CI is:
[50.86-1.65*0.68,50.86+1.65*0.68] = [49.74,51.98]

This is the same value produced by SPSS using the Explore command. We used **Analyse → Descriptive Statistics → Explore** and put the variable *mobtime* in the **Dependent list** tab. We click **Statistics** and changed the **Confidence interval** to 90%. Please find screenshots below.





Descriptives

			Statistic	Std. Error
hours spent monthly on mobile	Mean		50.860752	.6797075
	90% Confidence Interval for Mean	Lower Bound	49.737503	
		Upper Bound	51.984000	