# Lab Exercise - Descriptive Statistics

Shaik.Hameeda Parvin~G34 PYTHON

1. Open the [Teaching Ratings data](https://courses.cognitiveclass.ai/assets/courseware/v1/8b5e886922d6d1b5cf057fd0ff3b974a/asset-v1:IBM+ST0101EN+v1+type@asset+block/teachingratings.sav" \t "https://courses.cognitiveclass.ai/xblock/[object Object]) in SPSS and switch to variable view. Consider the following variables:  ****age, gender, beauty, eval, tenure, students****. Thought questions: Can you identify which variables are ****continuous variables**** and which ones are ****categorical variables****? Are any of the variables in the above list of ****ordinal**** type?

**age** – Numeric, Scale → **Continuous**

**gender** – String, Nominal → **Categorical**

**beauty** – Numeric, Scale → **Continuous**

**eval** – Numeric, Scale → **Continuous**

**tenure** – String, Nominal → **Categorical**

**students** – Numeric, Scale →**Continuous**

**CATEGORICAL MEANS QUALITATIVE**

**CONTINUOUS MEANS QUANTITIVE**

**No there are no variables of ordinal type**

**2.** Can you identify whether the Teaching Ratings data is a time series, cross-sectional, and/or multivariate data set?

· **Cross-sectional**: Data is collected from multiple professors at one point in time.

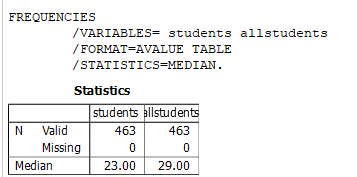
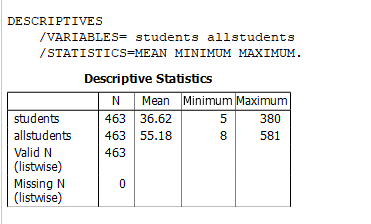
· **Multivariate**: It includes multiple variables for each observation (e.g., age, eval, beauty, etc.).

But not time series.

3. Does the Teaching Ratings data set represents information on an entire population or just a sample?

This dataset represents a **sample** of professors and students, not the entire population of all university instructors.

1. Using SPSS Statistics, find the ****mean****, ****median****, ****minimum****, and ****maximum**** values for ****students**** and ****allstudents**** in the Teaching Ratings data.



**Yes**, the average beauty score **does differ by gender**

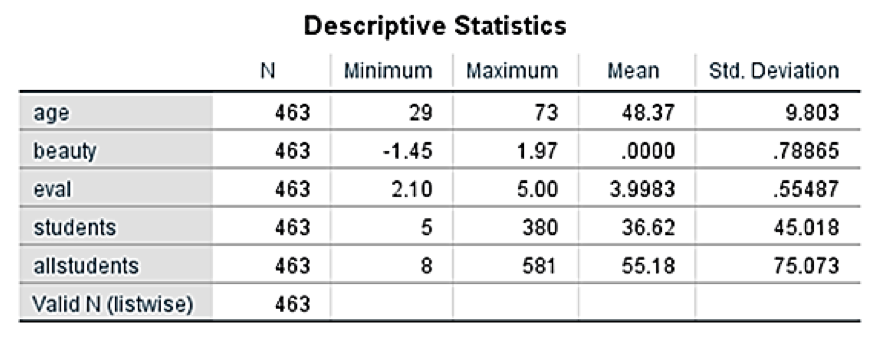
1. What is the mathematical relationship between ****variance**** and ****standard deviation****?

**Standard Deviation (SD)** is the square root of the variance:

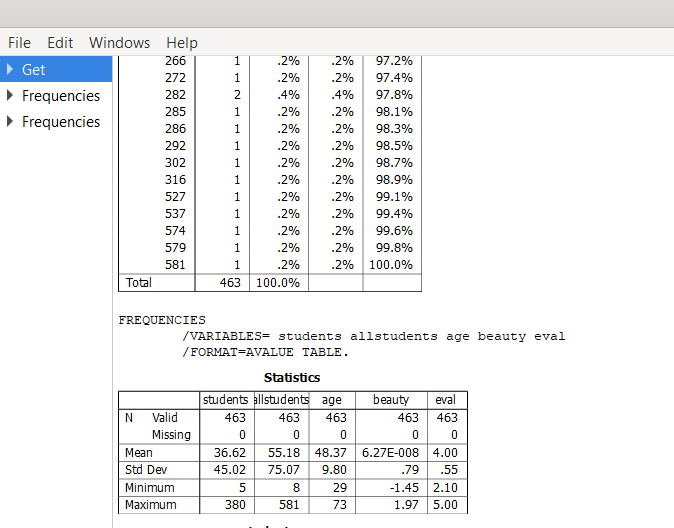
**Variance** is the square of the standard deviation:

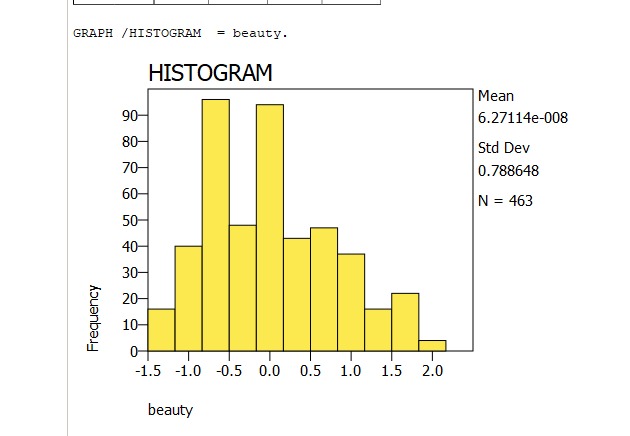
* Variance=(Standard Deviation)2
* Standard Deviation=square root of variance

****6.**** Using SPSS Statistics, try to reproduce the following ****descriptive statistics table**** on the Teaching Ratings data:



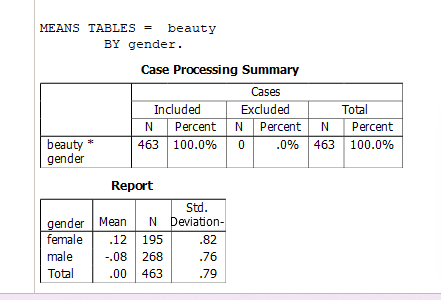
OUTPUT::::::::

****7.**** Create a *histogram* of ****beauty**** and briefly comment on the distribution of data (as a thought question).

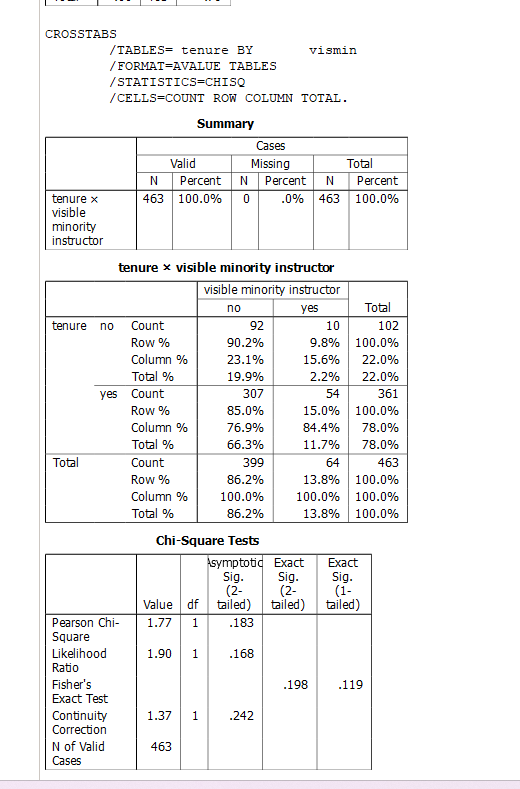


The histogram of the **beauty** variable shows a roughly normal distribution with a slight left skew. Most values are concentrated between **-1.0 and 0.5**, indicating that the majority of beauty scores are around average. The standard deviation is **0.79**, reflecting moderate variability in ratings across **463 observations**.

1. Does average ****beauty**** score differ by ****gender**** in the Teaching Ratings data? Produce averages and standard deviations for each of the two variables



1. Does ****tenure**** status differ by ****vismin**** (visible minority) status? Produce *cross tabulations* explaining what percentage of visible minorities are tenured.



Visible minority instructors have a higher tenure rate (84.4%) than non-minorities (76.9%). However, the chi-square test shows **no significant difference** (p = 0.183). So, tenure status does **not significantly differ** by visible minority status.

1. What is the probability of rolling two dice and getting 9 or less?

When we roll two dice , the total number of possible outcomes is **36** (6 sides × 6 sides).  
To find the probability of getting a sum **less than or equal to 9**, we count all favorable outcomes:

Sums ≤ 9 include: 2, 3, 4, 5, 6, 7, 8, and 9

Number of combinations for each sum:

2 → 1

3 → 2

4 → 3

5 → 4

6 → 5

7 → 6

8 → 5

9 → 4

Total favorable outcomes = 1+2+3+4+5+6+5+4 = **30**

**Probability = 30 / 36 = 0.8333**

1. What is the probability of rolling two dice and getting 7 or less?

We again consider all outcomes where the sum is **7 or less**:

Sum’s : 2, 3, 4, 5, 6, 7

Number of combinations:

2 → 1

3 → 2

4 → 3

5 → 4

6 → 5

7 → 6

Total favorable outcomes = 1+2+3+4+5+6 = **21**

**Probability = 21 / 36 = 0.5833**

1. With an average teaching evaluation score of 4 and standard deviation of 0.55, what is the probability of getting a teaching evaluation of greater than 4.75?

Mean (μ) = 4

Standard deviation (σ) = 0.55

Value (X) = 4.75

**Step 1: Calculate the Z-score**

Z=X−μ/σ= 4.75−40.55 =0.750.55≈1.36

**Step 2: Find the probability from Z-table**

P(Z < 1.36) ≈ **0.9131**

So, P(Z > 1.36) = 1 - 0.9131 = **0.0869**

**Answer: The probability of getting a score greater than 4.75 is approximately 0.087**