Motive

- Applying following SPSS Actions
  - Data Preparation
  - Merging Categories
  - Creating new summary variable
  - Descriptive Statistics – Frequencies (Bar charts , Histograms)
  - Descriptive Statistics – Scatter plots
  - Normality test
- Using two datasets
  - Open-source dataset
    - dataset contains survey results from 435 students enrolled at a university in the United States.
      - ID number , Date of birth , Date of college , Expected date of college graduation , Class rank ,Gender ,Athlete , Height Height ,Weight , Smoking , sprint , MileMinDur , English Score , Reading Score, Math Score , Writing Score , State , LiveOnCampus , HowCommute , CommuteTime , SleepTime , StudyTime
  - Lecture dataset
    - ID , Gender , Age , Marital , Employment , QOL_total , Distress_total , Esteem_Q[1-10]

Open-Source Dataset

- Data Preparation

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
ids	Numeric	5	0		None	None	8	Right	Scale	Input
bday	Date	20	0		None	None	20	Right	Scale	Input
Rank	Numeric	1	0		{1, Freshma...	999	8	Right	Ordinal	Input
Gender	Numeric	1	0		{0, Male}...	999	8	Right	Nominal	Input
Athlete	Numeric	1	0		{0, Non-Athl...	None	8	Right	Nominal	Input
Height	Numeric	5	2		None	None	8	Right	Scale	Input
Weight	Numeric	6	2		None	None	8	Right	Scale	Input
Smoking	Numeric	1	0		{0, Non-smo...	None	8	Right	Nominal	Input
Sprint	Numeric	5	3		None	None	8	Right	Scale	Input
MileMinDur	Date	11	0		None	None	11	Right	Scale	Input
English	Numeric	6	2		None	None	8	Right	Scale	Input
Reading	Numeric	6	2		None	None	8	Right	Scale	Input
Math	Numeric	5	2		None	None	8	Right	Scale	Input
Writing	Numeric	5	2		None	None	8	Right	Scale	Input
State	String	12	0		None	None	12	Left	Nominal	Input
LiveOnCam...	Numeric	1	0		{0, Off-camp...	None	8	Right	Nominal	Input
HowCommute	Numeric	1	0		{0, Walk}...	None	8	Right	Nominal	Input
CommuteTi...	Numeric	2	0		None	None	8	Right	Scale	Input
SleepTime	Numeric	2	0		None	None	8	Right	Scale	Input
StudyTime	Numeric	2	0		None	None	8	Right	Scale	Input
enrolldate	String	20	0		None	None	22	Left	Nominal	Input
expgradate	String	20	0		None	None	22	Left	Nominal	Input
Major	String	58	0		None	None	50	Left	Nominal	Input

- Merging Categories

    0 = "Non-smoker"
    1 = "Past-smoker"
    2 = "Current-smoker"

    Numeric Variable -> Output Variable:

    Smoking --> Smoking_new

    Old --> New:

    0 --> 0
    1 --> 0
    2 --> 1

    .00 = "No_Smoking"
    1.00 = "Yes_Smoking"

- Creating new summary variable

    Target Variable:

    BMI

    =

    Numeric Expression:

    Weight/(Height*Height)

    Type & Label...

- Descriptive Statistics – Frequencies (Bar charts , Histograms)

Statistics

		Gender	Athlete
N	Valid	427	435
	Missing	8	0

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	204	46.9	47.8	47.8
	Female	223	51.3	52.2	100.0
	Total	427	98.2	100.0	
Missing	System	8	1.8		
Total		435	100.0		

Athlete

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-Athlete	251	57.7	57.7	57.7
	Athlete	184	42.3	42.3	100.0
	Total	435	100.0	100.0	

-

-
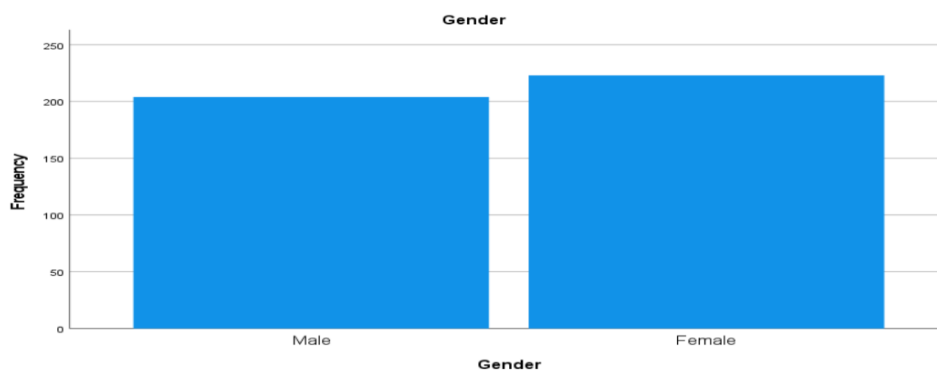
Statistics

		Sprint	English
N	Valid	374	409
	Missing	61	26
Mean		6.58193	82.7876
Minimum		4.503	59.83
Maximum		9.597	101.95

Frequency Table

Sprint

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4.503	1	.2	.3	.3
	4.524	1	.2	.3	.5
	4.556	1	.2	.3	.8
	4.566	1	.2	.3	1.1
	4.574	1	.2	.3	1.3
	4.593	1	.2	.3	1.6
	4.620	1	.2	.3	1.9
	4.621	1	.2	.3	2.1
	4.631	1	.2	.3	2.4
	4.634	1	.2	.3	2.7
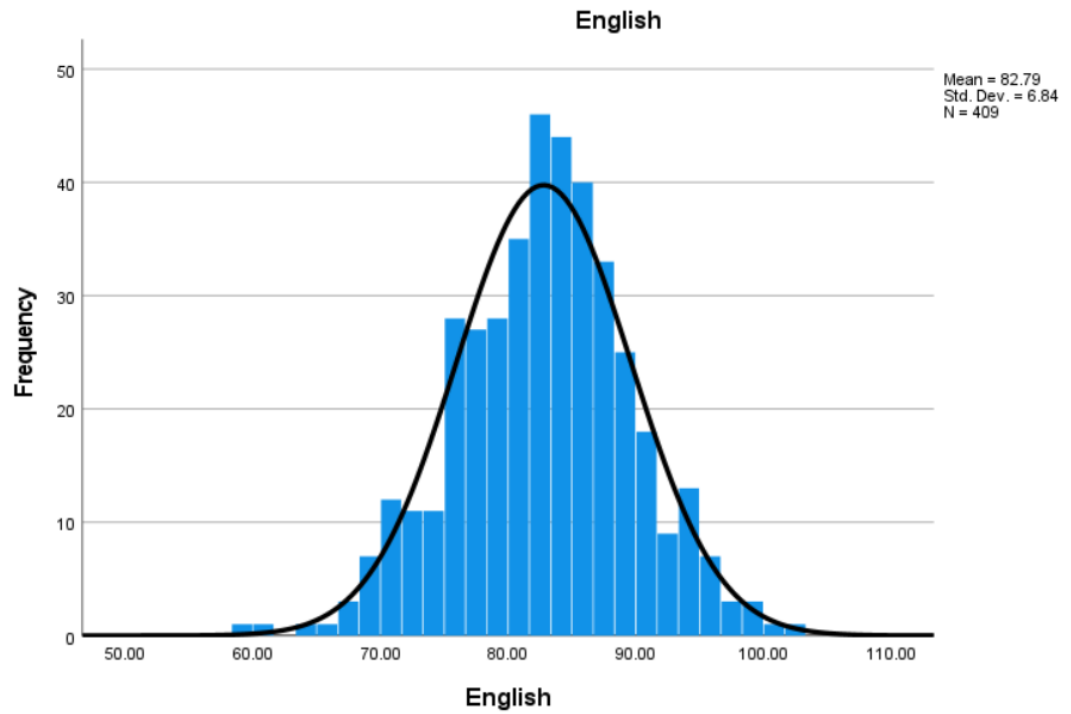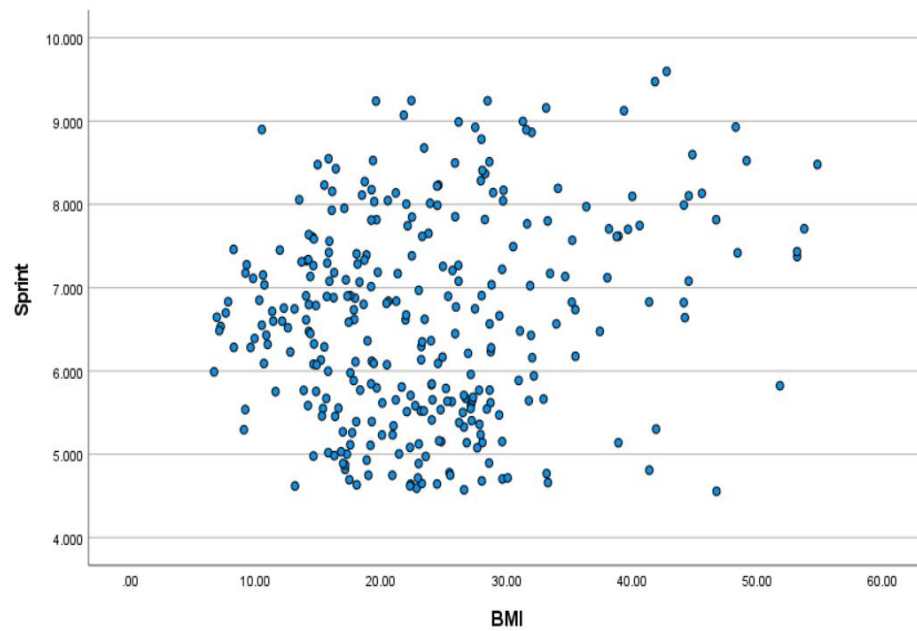
English

Mean = 82.79
Std. Dev. = 6.84
N = 409

-
  -
- Descriptive Statistics – Scatter plots

-
  -
- Normality test

English

Mean = 82.79
Std. Dev. = 6.84
N = 409

○
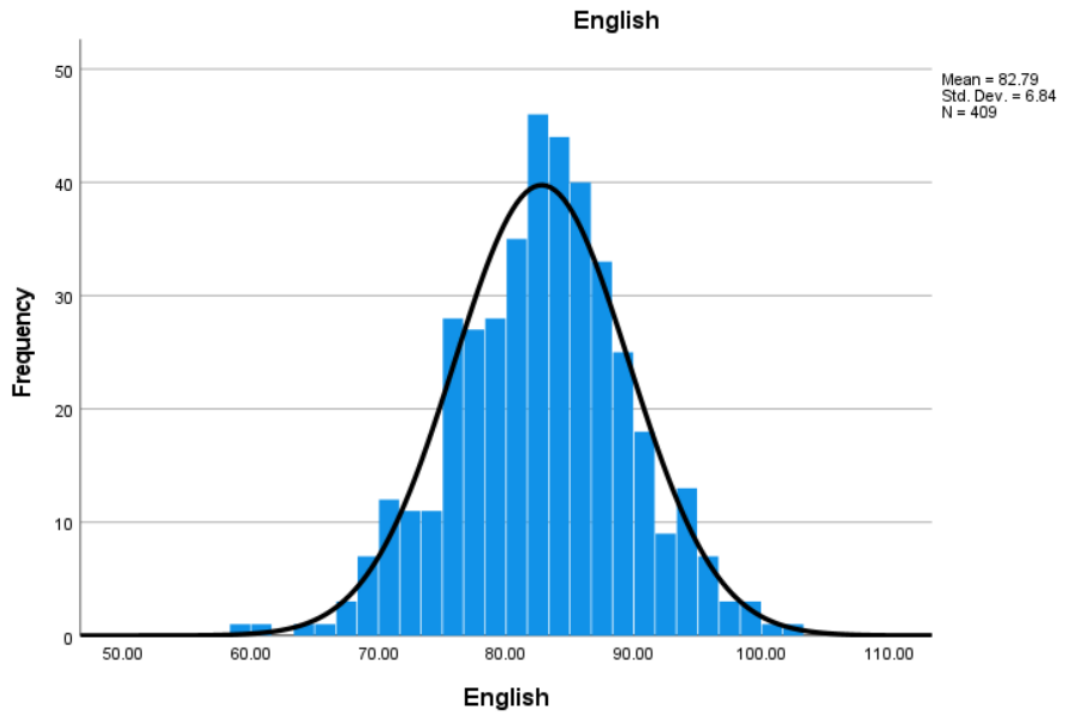
Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
English	409	94.0%	26	6.0%	435	100.0%

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
English	.033	409	.200[*]	.997	409	.543

*. This is a lower bound of the true significance.
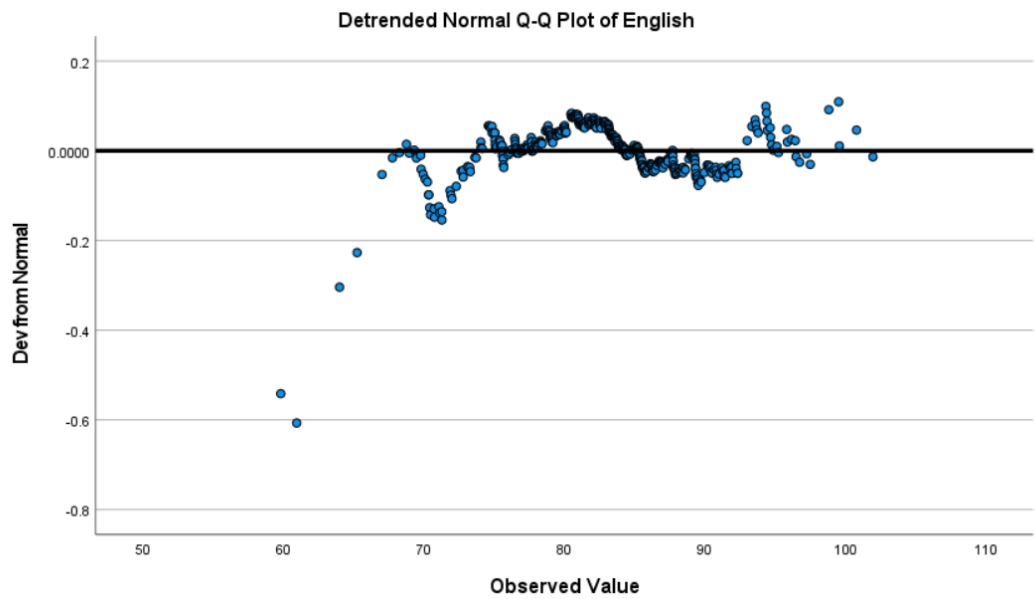
a. Lilliefors Significance Correction

○

- Sig > 0.05 -> English Exam data is normal

Normal Q-Q Plot of English
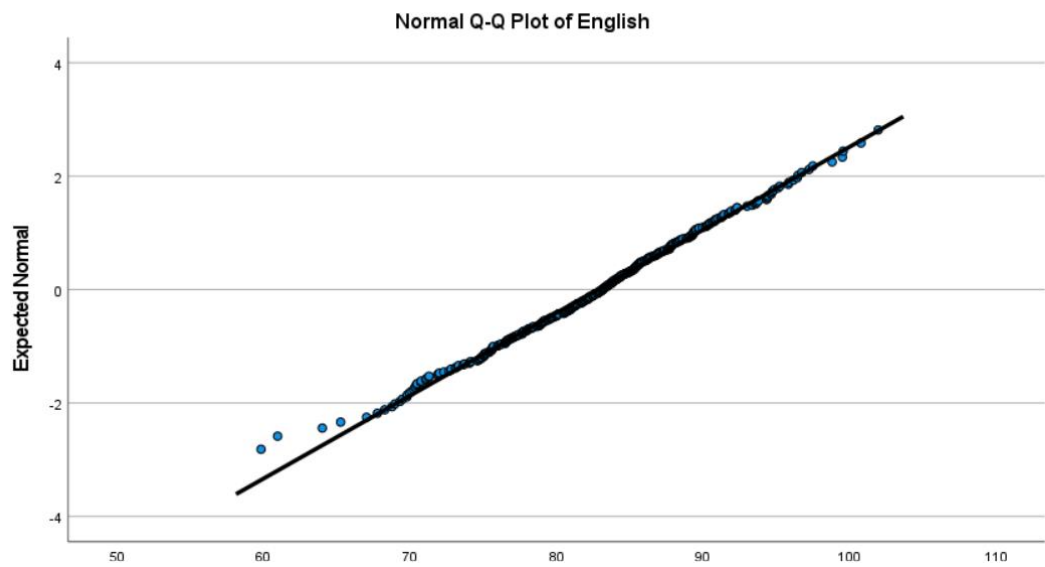
- Most of lines on line so it's normal

Detrended Normal Q-Q Plot of English

- No pattern an random so it's normal

Lecture Dataset

- Data Preparation

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
ID	Numeric	8	0		None	None	10	Right	Nominal	Input
Gender	Numeric	8	0	Gender	{0, male}...	999	8	Right	Nominal	Input
Age	Numeric	8	0	Age	None	999	8	Right	Scale	Input
Marital	Numeric	8	0	Marital status	{1, married, ...	999	8	Right	Nominal	Input
Employment	Numeric	8	0	Employment	{1, no job}...	999	11	Right	Nominal	Input
QOL_total	Numeric	8	0	Quality of life s...	None	999	10	Right	Scale	Input
Distress_total	Numeric	8	0	Distress scale t...	None	999	8	Right	Scale	Input
Esteem_Q1	Numeric	8	0	I am satisfied w...	None	None	8	Right	Scale	Input
Esteem_Q2	Numeric	8	0	I am good enough	None	None	8	Right	Scale	Input
Esteem_Q3	Numeric	8	0	I have a number...	None	None	8	Right	Scale	Input
Esteem_Q4	Numeric	8	0	I am able to do ...	None	None	8	Right	Scale	Input
Esteem_Q5	Numeric	8	0	I hvae much to ...	None	None	8	Right	Scale	Input
Esteem_Q6	Numeric	8	0	I do not feel us...	None	None	8	Right	Scale	Input
Esteem_Q7	Numeric	8	0	I am a person o...	None	None	8	Right	Scale	Input
Esteem_Q8	Numeric	8	0	I have respect f...	None	None	8	Right	Scale	Input
Esteem_Q9	Numeric	8	0	I am not a failure	None	None	8	Right	Scale	Input
Esteem_Q10	Numeric	8	0	I take a positive...	None	None	8	Right	Scale	Input

- Merging Categories

Numeric Variable -> Output Variable:

Employment --> Employment_new

Old --> New:

1 --> 1
2 --> 2
3 --> 2

Add

Change

1.00 = "no_job"
2.00 = "having_job"

Employment	Employment_new
2	2.00
3	2.00
3	2.00
2	2.00
1	1.00

- Creating new summary variable

Compute Variable   ✕

Target Variable:

Esteem_total

Type & Label...

Numeric Expression:

SUM(Esteem_Q1,Esteem_Q2,Esteem_Q3,Esteem_Q4,Esteem_Q5,Esteem_Q6,Esteem_Q7Esteem_Q8,Esteem_Q9,Esteem_Q10)

=

Gender [Gender]
Age [Age]

Function group:

- Descriptive Statistics – Frequencies (Bar charts , Histograms)

Statistics

		Gender	Marital status	Employment
N	Valid	200	200	200
	Missing	0	0	0
Minimum		0	1	1
Maximum		1	3	3

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	91	45.5	45.5	45.5
	female	109	54.5	54.5	100.0
	Total	200	100.0	100.0	

Marital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	married, common law	63	31.5	31.5	31.5
	widow, divorce, separate	65	32.5	32.5	64.0
	single, never married	72	36.0	36.0	100.0
	Total	200	100.0	100.0	

○

○
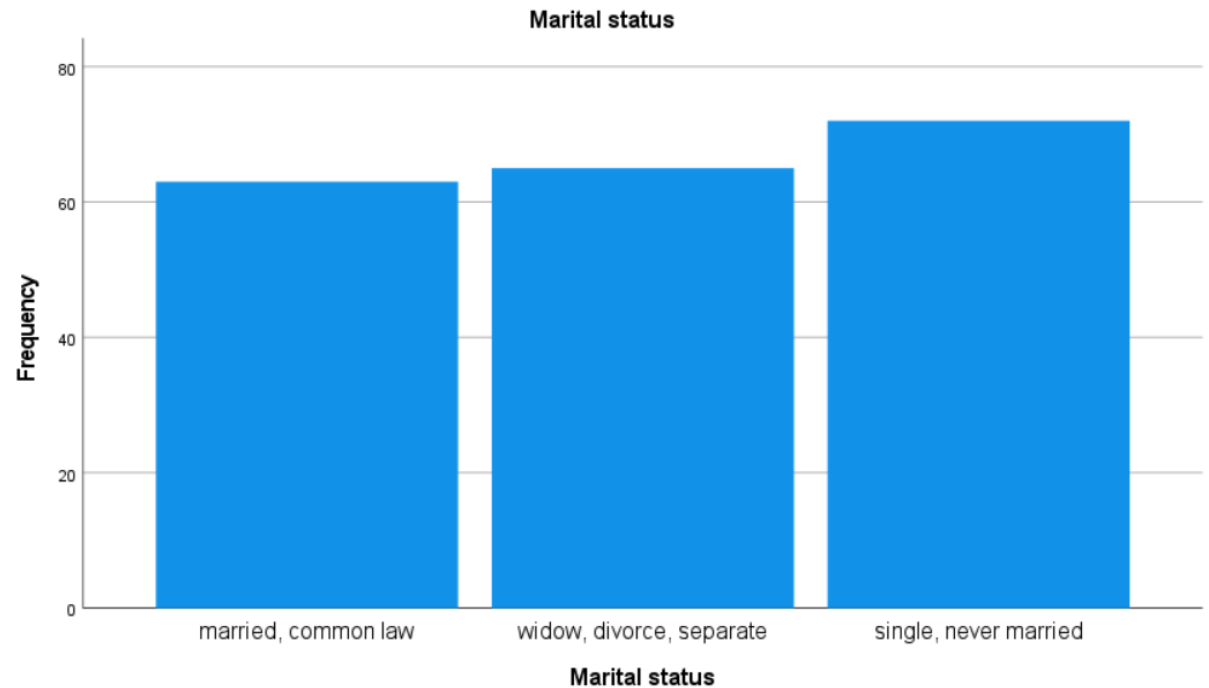
Statistics

		Age	Quality of life scale total score	Distress scale total score
N	Valid	196	200	198
	Missing	4	0	2
Mean		44.20	11.45	10.68
Minimum		13	3	1
Maximum		80	20	20

○

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	13	2	1.0	1.0	1.0
	19	6	3.0	3.1	4.1
	20	6	3.0	3.1	7.1
	21	1	.5	.5	7.7
	22	4	2.0	2.0	9.7
	23	8	4.0	4.1	13.8
	24	6	3.0	3.1	16.8
	25	1	.5	.5	17.3
	26	2	1.0	1.0	18.4

○

Age
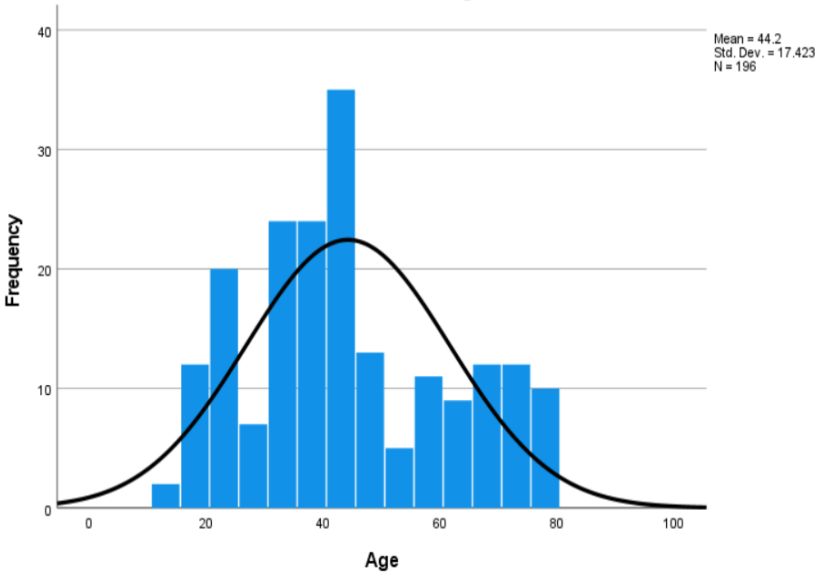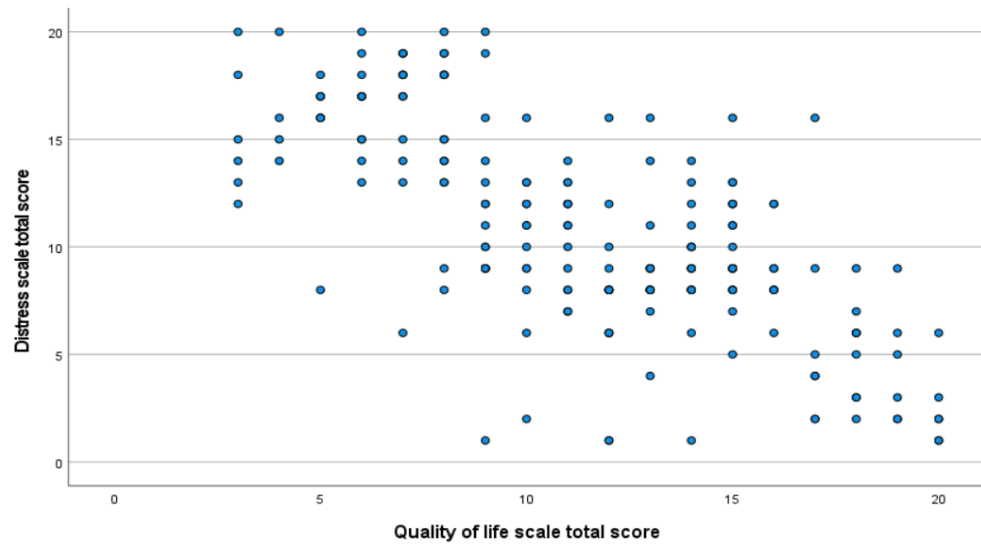
Mean = 44.2
Std. Dev. = 17.423
N = 196

○

- Descriptive Statistics – Scatter plots

  o

- Normality test

  o
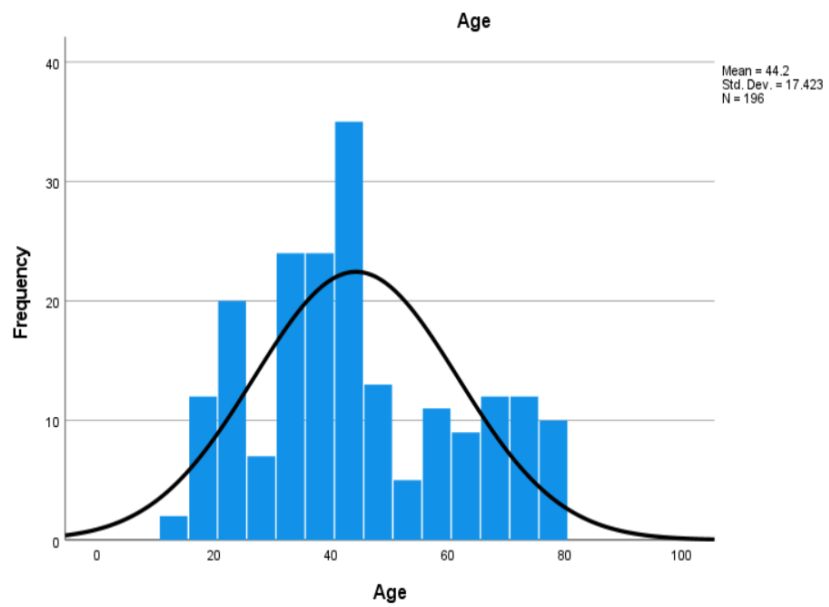
Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Age	196	98.0%	4	2.0%	200	100.0%

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age	.132	196	<.001	.946	196	<.001

a. Lilliefors Significance Correction

- ○
  - ▪ Sig < 0.05 -> Age data is not normal

Normal Q-Q Plot of Age

- ○
  - ▪ Many points not on the line so it's not normal

Detrended Normal Q-Q Plot of Age

- ○
  - ▪ There's psttern not random do it's not normal