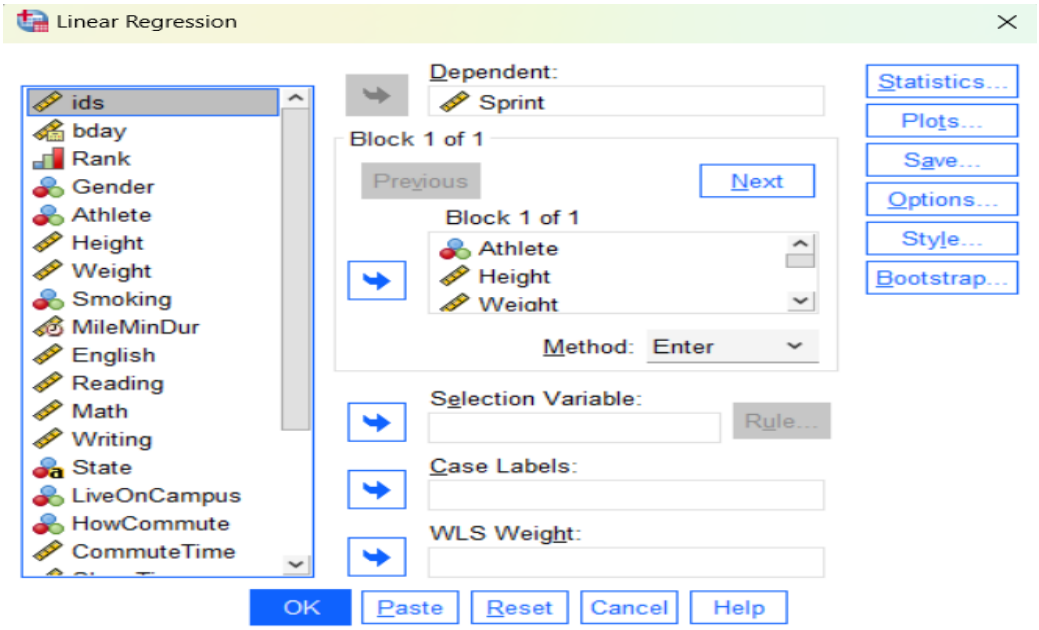
Motive

- Applying following SPSS Actions
  - Simple Ordinary least squares (OLS) linear regression (Single Independent variable)
  - Multiple Ordinary least squares (OLS) linear regression (Multiple Independent variable)
- Assumptions
  - *(1) Independence, (2) Normality, (3) Equal variance, (4) Linearity*
- Using three datasets
  - Open-source dataset1
    - 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
  - Open-source dataset2
    - dataset contains heights for 1078 pairs of father/son
      - father_height/ son_height
  - Lecture dataset
    - ID , Gender , Age , Marital , Employment , QOL_total , Distress_total , Esteem_Q[1-10]
- Steps
  - **• Analyze > Regression > Linear …**

Open-Source Dataset1



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.767	.754		4.994	<.001
	Athlete	-1.757	.105	-.707	-16.790	<.001
	Weight	.058	.009	.265	6.424	<.001
	Height	-.009	.001	-.303	-7.112	<.001
	MileMinDur	.003	.000	.311	7.794	<.001
	English	-.004	.006	-.021	-.665	.507
	SleepTime	-.009	.016	-.018	-.553	.581
	StudyTime	-.004	.009	-.014	-.427	.670

a. Dependent Variable: Sprint

Equation:

Sprint Time = 3.767-1.757*Athelete + 0.058*Weight -0.009*Height + 0.003 * MileDur – 0.004 * English - 0.009 * sleeptime – 0.004 * StudyTime

Sig value

Athlete , Height , Weight , MileDur have significant impact on sprint time sine sig < 0.05

English , study time , sleep time don't have significant impact on sprint time since sig > 0.05

Relation type

Athlete , Height is opposite of sprint time since sign of coefficient is negative

Weight , MileDur having positive relation with sprint time since sign of coefficient is positive

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.891[a]	.794	.787	.566455

a. Predictors: (Constant), StudyTime, English, Height, SleepTime, MileMinDur, Weight, Athlete

R square is 0.794 and this is the score of model strength and how model is describing the impact of these variables on sprint time but still 0.206 of the value is impacted by other variables

Adjusted R square is also more indicative (more robust) (more aaccuracy) then R square and usually less than R square

Std error is 0.566 and it's absolute value (not percentage)

ANOVA[a]

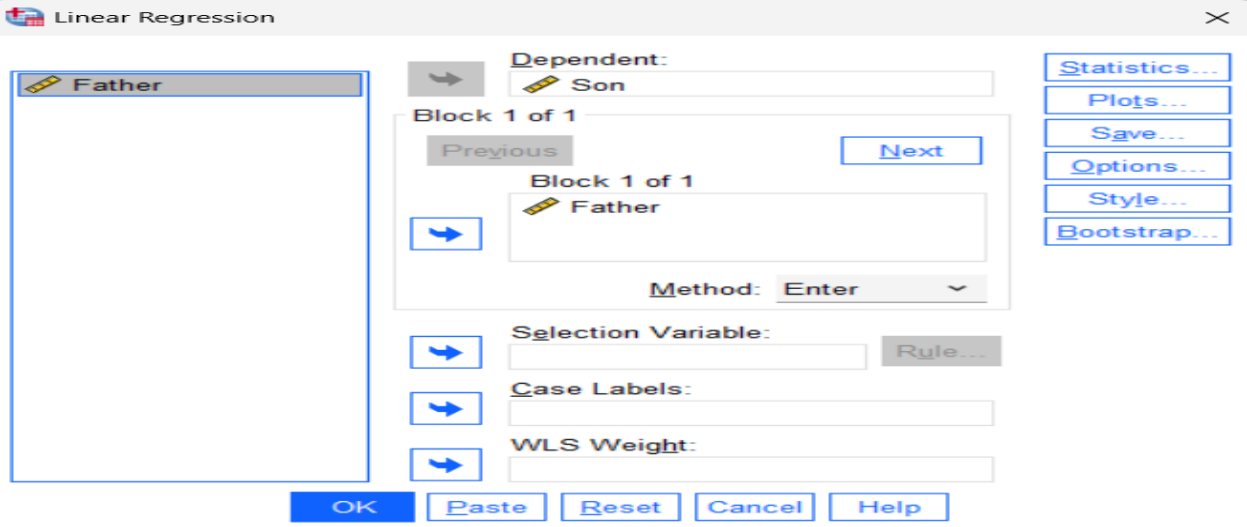
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	259.337	7	37.048	115.461	<.001[b]
	Residual	67.383	210	.321		
	Total	326.720	217			

a. Dependent Variable: Sprint

b. Predictors: (Constant), StudyTime, English, Height, SleepTime, MileMinDur, Weight, Athlete

As sig < 0.05 so the total test is significant and even if there's some variables not significant but the other variables are dominant

Open-Source Dataset2



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	33.893	1.833		18.491	<.001
	Father	.514	.027	.501	18.997	<.001

a. Dependent Variable: Son

Equation:

Son = 33.893+0.524*Father

Sig value

Father have significant impact on Son since sig < 0.05

Relation type

Father having positive relation with Son since sign of coefficient is positive

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.501[a]	.251	.250	2.4381

a. Predictors: (Constant), Father

R square is 0.251 and this is the score of model strength and how model is describing the impact of these variables on Son but still 0.749 of the value is impacted by other variables

Adjusted R square is also more indicative (more robust) (more aaccuracy) then R square and usually less than R square

Std error is 2.4381 and it's absolute value (not percentage)

ANOVA[a]

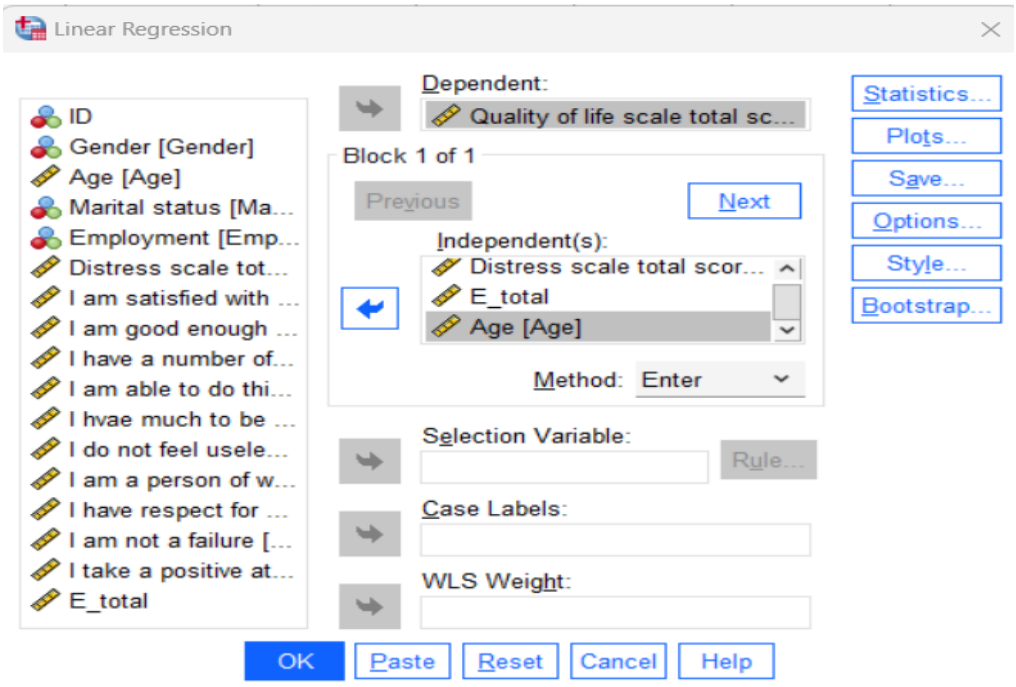
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2145.351	1	2145.351	360.897	<.001[b]
	Residual	6396.281	1076	5.944		
	Total	8541.632	1077			

a. Dependent Variable: Son

b. Predictors: (Constant), Father

As sig < 0.05 so the total test is significant

Lecture Dataset



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	12.558	1.409		8.913	<.001
	Distress scale total score	-.429	.060	-.480	-7.190	<.001
	E_total	.216	.044	.328	4.916	<.001
	Age	-.005	.012	-.022	-.443	.659

a. Dependent Variable: Quality of life scale total score

Equation:

Quality of life = 12.558 – 0.429*Distress+0.216*E_total-0.005*Age

Sig value

Distress , E_total have significant impact on Quality of life sine sig < 0.05
Age doesn't have significant impact on quality of life since sig > 0.05

Relation type

Distress,Age is opposite of quality of life since sign of coefficient is negative
E-total having positive relation with quality of life since sign of coefficient is positive

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.742[a]	.551	.544	2.937

a. Predictors: (Constant), Age, Distress scale total score, E_total

R square is 0.551 and this is the score of model strength and how model is describing the impact of these variables on quality of life but still 0.449 of the value is impacted by other variables

Ifwe removed the age and keep only the other two variables , R square will increase to 0.560 and that's why sig value of age is > 0.05 (not significant)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.748[a]	.560	.556	2.916

a. Predictors: (Constant), E_total, Distress scale total score

Adjusted R square is also more indicative (more robust) (more aaccuracy) then R square and usually less than R square

Std error is 2.916 and it's absolute value (not percentage) and here the error in second model is less than the error in first model

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2012.972	3	670.991	77.779	<.001[b]
	Residual	1639.110	190	8.627		
	Total	3652.082	193			

a. Dependent Variable: Quality of life scale total score

b. Predictors: (Constant), Age, Distress scale total score, E_total

As sig < 0.05 so the total test is significant and even if there's one variable not significant (Age) but the other two variable are dominant