How can I run a piecewise regression in SPSS?

Authored by

stats writer

June 30, 2024

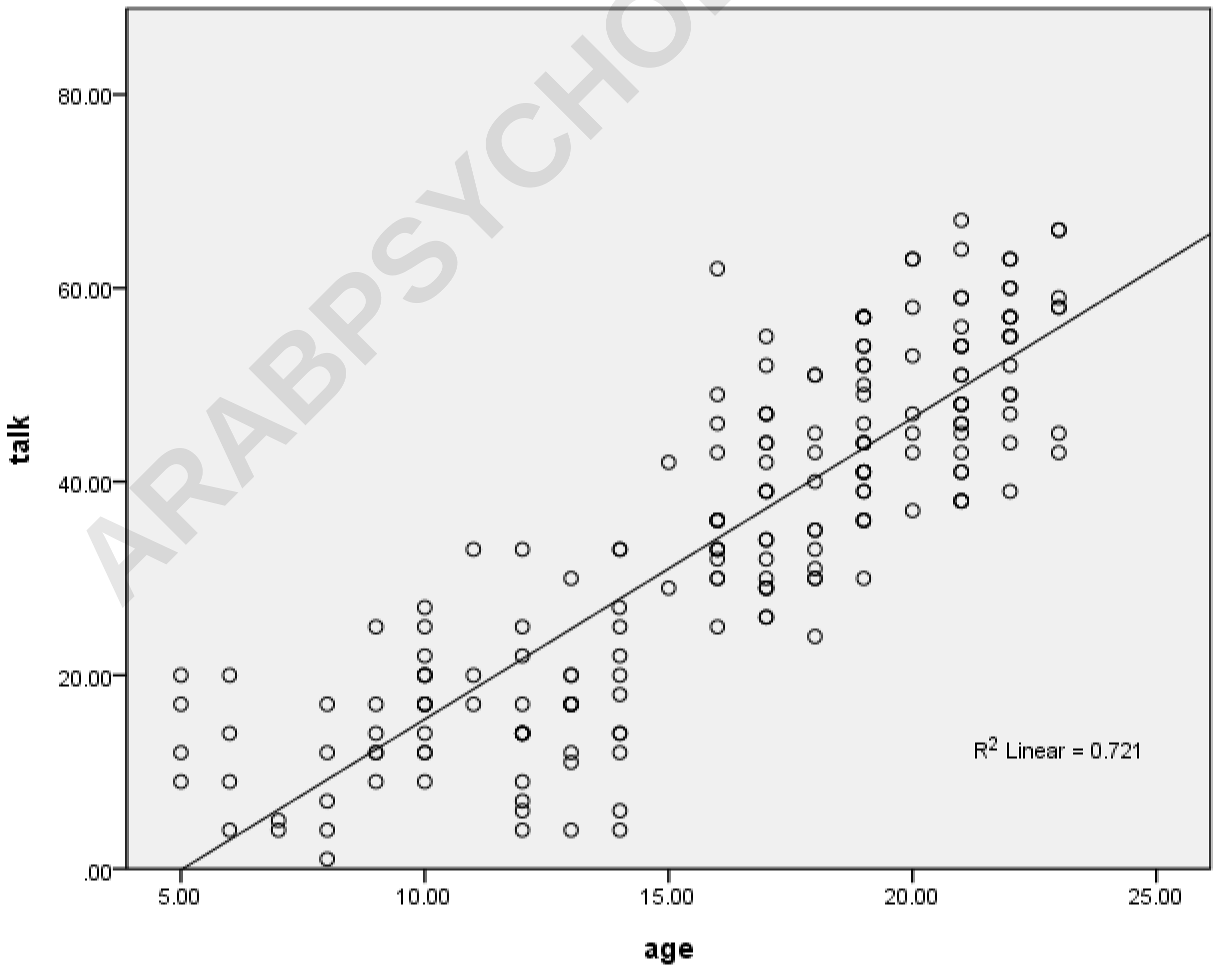
To perform a piecewise regression in SPSS, begin by opening the "Analyze" menu and selecting "Regression" followed by "Linear." Then, select the "Plots" option and choose "Piecewise." Next, specify the variables to be included in the regression and set the breakpoints for the piecewise segments. Finally, click "Run" to generate the results and interpret the output. Piecewise regression in SPSS allows for the analysis of data with non-linear relationships between variables, providing a more accurate and comprehensive understanding of the data.

How can I run a piecewise regression in SPSS? | SPSS FAQ

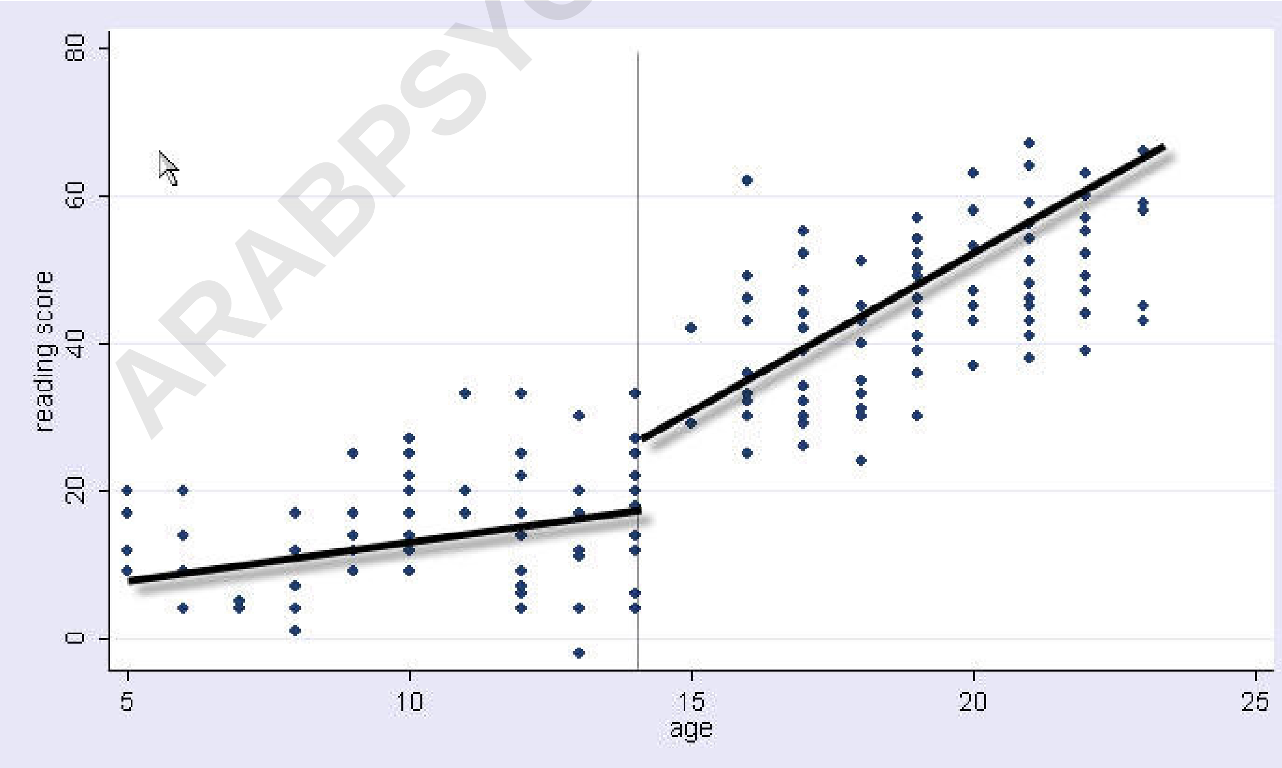
Say that you want to look at the relationship between how much a child talks on the phone and the age of the child. You get a random sample of 200 kids. You ask them how old they are and how many minutes they spend talking on the phone. Finally, you save the data as https://stats.idre.ucla.edu/wp-content/uploads/2016/02/talk.sav. You then make a scatterplot of the data as shown below.

```
GGRAPH
/GRAPHDATASET NAME="iGraphDataset" VARIABLES= talk age
/GRAPHSPEC SOURCE=INLINE INLINETEMPLATE=.
BEGIN GPL
```

```
SOURCE: s=userSource( id( "iGraphDataset" ) )
DATA:talk=col( source(s), name( "talk" ) )
DATA: age=col( source(s), name( "age" ) )
GUIDE: axis( dim( 1 ), label( "age" ) )
GUIDE: axis( dim( 2 ), label( "talk" ) )
SCALE: linear( dim( 1 ), min( 5 ), max( 25 ) )
SCALE: linear( dim( 2 ), min( 0 ), max( 80 ) )
ELEMENT: point( position( ( age * talk ) ) )
END GPL.
```

Thinking about this more, you decide that you think that the amount of time that kids talk on the phone changes dramatically at age 14, and that the slope might change at that age as well. You think that a piecewise regression might make more sense, where before age 14 there is an intercept and linear slope, and after age 14, there is a different intercept and different linear slope, kind of like pictured below with just freehand drawing of what the two regression lines might look like.



Try 1: Separate regressions

To investigate this, we can run two separate regressions, one for before age 14, and one for after age 14. We can compare the results of these two models.

*** Before age 14.**
compute before14 = (age < 14).
filter by before14.
regression /dep=talk /method=enter age.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.075	3.981		2.029	.047
	age	.682	.382	.225	1.786	.079

a. Dependent Variable: talking on the phone

filter off.

*** At age 14 and after.**
compute after14 = (age >= 14).
filter by after14.
regression /dep=talk /method=enter age.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-24.973	5.709		-4.374	.000
	age	3.629	.301	.718	12.037	.000

a. Dependent Variable: talking on the phone

filter off.

Note how the slopes do seem quite different for the two groups. However, the
intercepts don't make much sense, since they are the predicted time talking on
the phone when one is 0 years old.

Try 2: Separate regression with age centered at 14

Let's rescale (center) age by subtracting 14. Then, when age is 0, that really
refers to being 14 years old.

* age14 subtracts 14 from age, so age is 0 when child is 14.
compute age14 = (age - 14).

* Now, rerun regression when child is below 14.
compute before14 = (age < 14).
filter by before14.

regression /dep=talk /method=enter age14.

filter off.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	17.624	1.752		10.057	.000
	age14	.682	.382	.225	1.786	.079

a. Dependent Variable: talk talking on the phone

*** Now, rerun regression when child is 14 years of age or older.**

compute after14 = (age >= 14).

filter by after14.

regression /dep=talk /method=enter age14.

filter off.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	25.834	1.626		15.884	.000
	age14	3.629	.301	.718	12.037	.000

a. Dependent Variable: talk talking on the phone

Note how the slopes for the two groups stayed the same, but now the intercepts (Constant) are the predicted talking time at

age 14 for the two
groups. We can see that at age 14 there seems to be not only a change
in the slope (from .682 to 3.62) but also a jump in the intercept
(from 17.6 to 25.8). This suggest that at age 14, there is discontinuous
jump in time talking on the phone as well as a change in the slope as well.
However, this is merely suggestive; we should really test this in a combined
model.

Try 3: Combined model, coding for separate slope and intercept

We now combine the two models into a single model. To do this, we need to
create some new variables.

```
compute age1 = (age - 14).
if (age >= 14) age1 = 0 .
compute age2 = (age - 14).
if (age < 14) age2 = 0 .

compute int1 = 1.
```

if (age >= 14) int1 = 0.
compute int2 = 1.
if (age < 14) int2 = 0.
execute.

That might have been confusing, so let us show what these variables look like in a table below. Note that we have a strange person who is 13.99 years old (very, very close to being 14, but
not quite). This person will be helpful for seeing the effect of the jump from going
from being under 14 to being 14.

* Check the coding.

* Save our data file so far.
save outfile = "c:datatalk2.sav".
execute.

* Collapse the data to make the coding easier to see.
aggregate
/outfile=* mode = addvaraibles
/break=age int1 int2 age1 age2
/count=N.

list cases.

age int1 int2 age1 age2 count

5.00 1.00 .00 -9.00 .00 4
6.00 1.00 .00 -8.00 .00 4
7.00 1.00 .00 -7.00 .00 2
8.00 1.00 .00 -6.00 .00 5
9.00 1.00 .00 -5.00 .00 6
10.00 1.00 .00 -4.00 .00 13
11.00 1.00 .00 -3.00 .00 3
12.00 1.00 .00 -2.00 .00 13
13.00 1.00 .00 -1.00 .00 11
13.99 1.00 .00 -.01 .00 1
14.00 .00 1.00 .00 .00 11
15.00 .00 1.00 .00 1.00 2
16.00 .00 1.00 .00 2.00 15
17.00 .00 1.00 .00 3.00 20
18.00 .00 1.00 .00 4.00 12
19.00 .00 1.00 .00 5.00 25
20.00 .00 1.00 .00 6.00 8
21.00 .00 1.00 .00 7.00 22
22.00 .00 1.00 .00 8.00 16
23.00 .00 1.00 .00 9.00 7

**Now we can go back to the talk2.sav data file before we did this
collapsing.**

get file = "c:datatalk2.sav".

**Now we are ready to run our combined regression. We will
put in the intercept for both groups, so we don't need an intercept from SPSS so we use the origin option to put the regression
through the origin (i.e., no intercept). This is necessary because our model has
an implied constant, int1 plus int2
adds up to 1. Note that the r-square is not valid for this model and should not be
reported.**

*** Run the regression, compare to try 2.
regression /origin /dependent=talk
/method=enter int1 int2 age1 age2
/save=pred(yhat).**

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	int1	17.624	2.009	.255	8.771	.000
	int2	25.834	1.547	.557	16.701	.000
	age1	.682	.438	.045	1.558	.121
	age2	3.629	.287	.422	12.656	.000

a. Dependent Variable: talk talking on the phone

b. Linear Regression through the Origin

Now let's obtain the predicted values (shown in the table below) and relate those to the meaning of the coefficients above.

means tables=yhat by age.

yhat Unstandardized Predicted Value * age

yhat Unstandardized Predicted Value

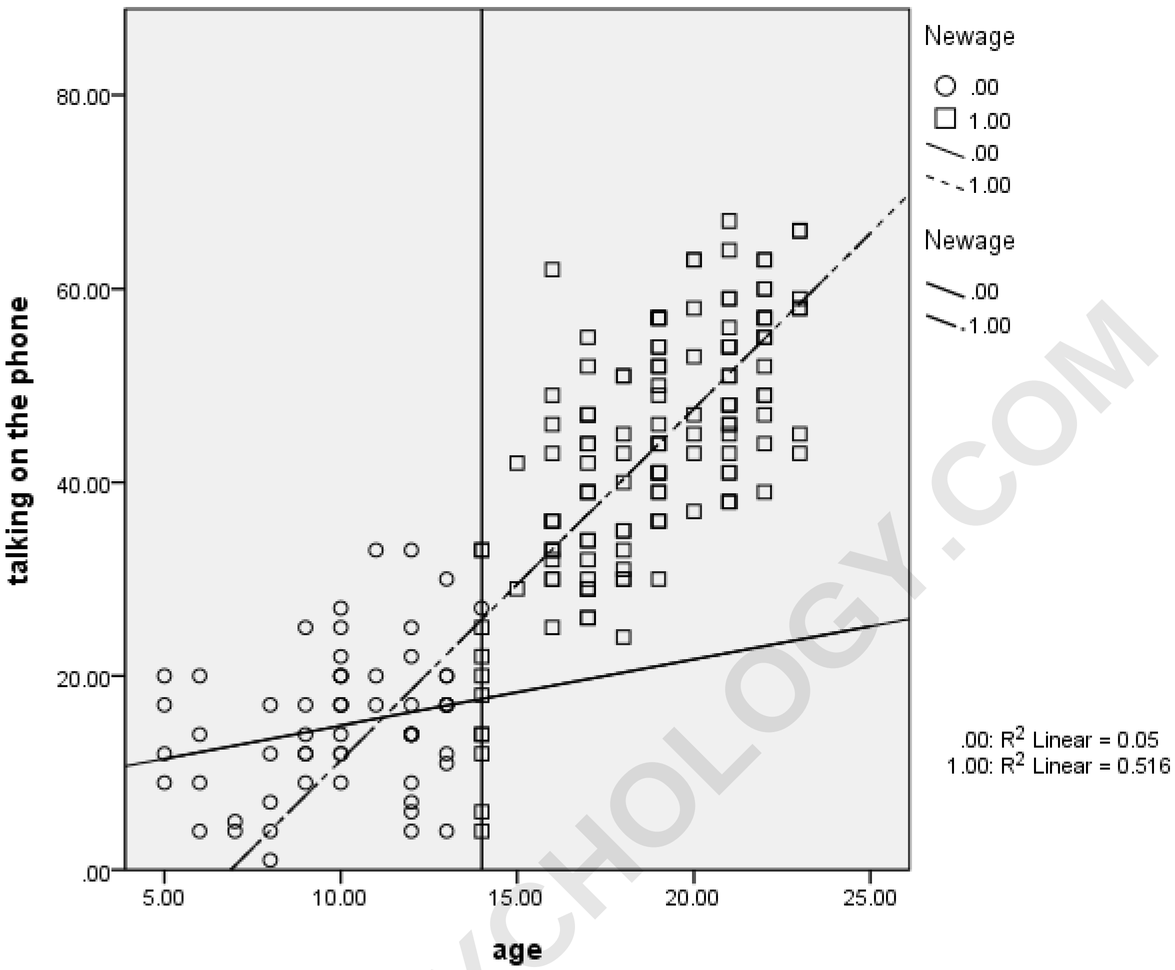
age	Mean	N	Std. Deviation
5.00	11.48639	4	.00000000
6.00	12.16831	4	.00000000
7.00	12.85023	2	.00000000
8.00	13.53215	5	.00000000
9.00	14.21407	6	.00000000
10.00	14.89599	13	.00000000
11.00	15.57791	3	.00000000
12.00	16.25983	13	.00000000
13.00	16.94175	11	.00000000
13.99	17.61685	1	.
14.00	25.83397	11	.00000000
15.00	29.46302	2	.00000000
16.00	33.09206	15	.00000000
17.00	36.72111	20	.00000000
18.00	40.35015	12	.00000000
19.00	43.97920	25	.00000000
20.00	47.60824	8	.00000000
21.00	51.23729	22	.00000000
22.00	54.86634	16	.00000000
23.00	58.49538	7	.00000000
Total	34.41000	200	15.14670989

Below we show a graph of the results.

```
compute newage = 0.
if age ge 14 newage = 1.
exe.

GGRAPH
/GRAPHDATASET NAME="iGraphDataset"
VARIABLES= talk age newage
/GRAPHSPEC SOURCE=INLINE EDITABLE=YES
INLINETEMPLATE=.
```

```
BEGIN GPL
SOURCE: s=userSource( id( "iGraphDataset" ) )
DATA: talk=col( source(s), name( "talk" ) )
DATA: age=col( source(s), name( "age" ) )
DATA: newage=col( source(s), name( "newage" ), unit.category() )
GUIDE: axis( dim( 1 ), label( "age" ) )
GUIDE: axis( dim( 2 ), label( "talking on the phone" ) )
GUIDE: legend( aesthetic( aesthetic.shape.interior ), label( "newage" ) )
SCALE: linear( dim( 1 ), min( 5 ), max( 25 ) )
SCALE: linear( dim( 2 ), min( 0 ), max( 80 ) )
GUIDE: form.line(position(14, *))
ELEMENT: point( position( age * talk ), shape.interior(newage) )
ELEMENT: line( position( smooth.linear(age * talk ) ), shape.interior( newage) )
GUIDE: legend(aesthetic(aesthetic.shape), label("Newage") )
END GPL.
```



You might want to test whether the difference in the intercepts
is 0 or whether the change in the slopes is different from 0. The next section
shows how we can do this.

Try 4: Alternate coding, coding to compare intercepts and slopes

This is another way you can code this model. Note that we include age14 and

age2 for the two terms for age, and include the intercept (by not excluding it) and int2 to represent the intercept values.
With this coding, age2 and int2 represent the change in slope
and intercept from being
less than 14 to being 14 and older.

```
regression
/dependent=talk
/method=enter age14 age2 int2
/save=pred(yhat2).
```

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	17.624	2.009		8.771	.000
	age14	.682	.438	.186	1.558	.121
	age2	2.947	.523	.518	5.631	.000
	int2	8.210	2.536	.220	3.238	.001

a. Dependent Variable: talk talking on the phone

Using this coding scheme, here is the meaning of the coefficients.

As you can see, the coefficients for age2 and int2 now focus on the change that results from becoming 14

years old.

Below we compute the predicted values calling them yhat2. Note how the predicted values are the same for this model and the prior model, because the models are essentially the same, they are just parameterized differently.

means tables=yhat2 by age.

Report

yhat2 Unstandardized Predicted Value

age	Mean	N	Std. Deviation
5.00	11.48639	4	.00000000
6.00	12.16831	4	.00000000
7.00	12.85023	2	.00000000
8.00	13.53215	5	.00000000
9.00	14.21407	6	.00000000
10.00	14.89599	13	.00000000
11.00	15.57791	3	.00000000
12.00	16.25983	13	.00000000
13.00	16.94175	11	.00000000
13.99	17.61685	1	.
14.00	25.83397	11	.00000000
15.00	29.46302	2	.00000000
16.00	33.09206	15	.00000000
17.00	36.72111	20	.00000000
18.00	40.35015	12	.00000000
19.00	43.97920	25	.00000000
20.00	47.60824	8	.00000000
21.00	51.23729	22	.00000000
22.00	54.86634	16	.00000000
23.00	58.49538	7	.00000000
Total	34.41000	200	15.14670989

Summary

This brief FAQ compared different ways of creating piecewise regression models. All of these models are equivalent,
just parameterized differently. They all generate the exact predicted values. The differences in parameterization are merely a rescrambling
of the intercepts and slopes for the two segments of the regression model. You
can choose the coding strategy that you like best, although it is often useful
to use both schemes.