C.5 Curvilinear Regression

The aim of the study is to find an adequate relationship between birth-weight (dependent variable) and gestational age (GA) on the basis of observations on 800 births in a hospital in a developing country.

The following command fits several types of regressions, namely, linear, logarithmic, inverse, quadratic, power, growth and exponential:

```
CURVEFIT /VARIABLES=BirthWeight WITH GA
  /CONSTANT
  /MODEL=LINEAR LOGARITHMIC INVERSE QUADRATIC POWER GROWTH EXPONENTIAL
  /PLOT FIT.
```

The following table provides the model summary and parameter estimates of various types of relationships. The one with maximum R-square is statistically most adequate. There is a marginal difference among linear, logarithmic, inverse and quadratic regressions. In such a situation, you can choose any that is biologically more plausible.

Model Summary and Parameter Estimates

Dependent Variable: BirthWeight

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	.461	681.313	1	798	.000	-4.102	.179	
Logarithmic	.462	686.184	1	798	.000	-20.283	6.319	
Inverse	.461	682.550	1	798	.000	8.497	-220.330	
Quadratic	.463	343.713	2	797	.000	-8.840	.448	-.004
Power	.514	843.637	1	798	.000	3.73E-005	3.075	
Growth	.506	816.918	1	798	.000	-2.304	.087	
Exponential	.506	816.918	1	798	.000	.100	.087	

The independent variable is GA.

The following figure depicts the scatter plot and various types of relationship between the birth-weight and gestational age (Figure C.5(a)).

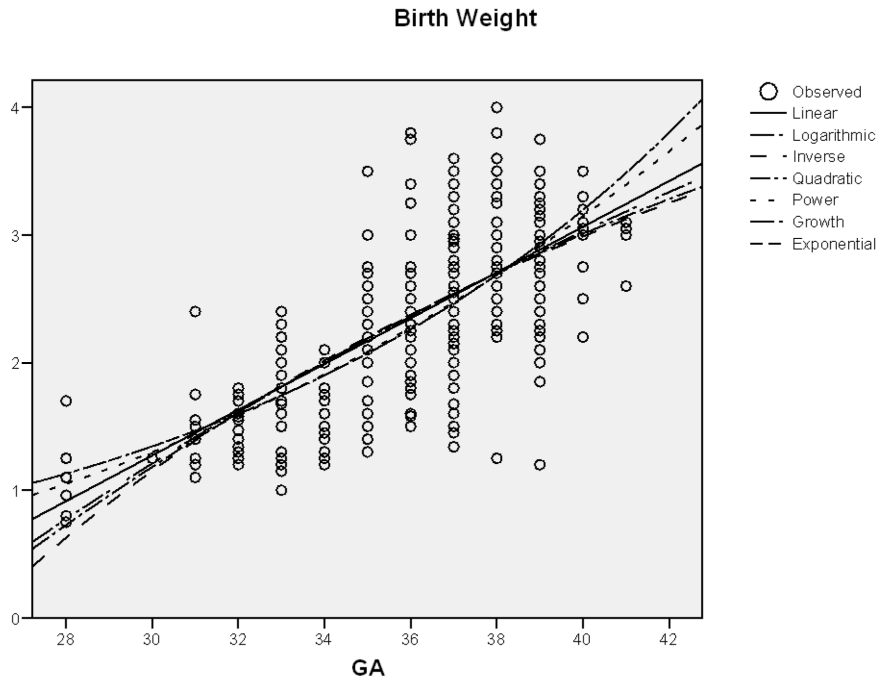


FIGURE C.5(a) Scatter plot and various types of regression of birth weight on GA

This figure does not provide much of a clue in this case. However, the preceding table shows that power relationship has maximum R-square although not particularly high to inspire confidence. But that is best one can obtain with these data. The general form of power equation is

$$Birth\ weight = \beta_0 * GA^{\beta_1}$$

This has two parameters, namely, β_0 and β_1. This is curvilinear as logarithm converts this to a linear form:

$$Log_{10}(Birth\ weight) = Log_{10}(\beta_0) + \beta_1 Log_{10}(GA) .$$

The parameter estimates shown in the preceding table for power regression gives the fitted regression equation:

$$Birth\ weight = 0.0000373 \times GA^{3.075}$$

To get more information on this model such as CI, now fit this model directly to the data.

Command to fit linearized form of the power equation:

```
COMPUTE Log10GA = LG10(GA).
COMPUTE Log10Birtweight = LG10(BirthWeight).
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Log10Birtweight
  /METHOD=ENTER Log10GA
  /SCATTERPLOT=(*ZRESID ,Log10BIRTHWEIGHT)
  /RESIDUALS NORM(ZRESID)
```

```
/SAVE ZRESID.
```

Model summary now is as follows, which is the same as obtained earlier.

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.717(a)	.514	.513	.08048

a Predictors: (Constant), Log10GA
b Dependent Variable: Log10Birtweight

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.464	1	5.464	843.637	.000(a)
	Residual	5.169	798	.006		
	Total	10.633	799			

a Predictors: (Constant), Log10GA
b Dependent Variable: Log10Birtweight

The following table displays the unstandardized and standardized coefficients, and their 95% CI. The coefficients are statistically highly significant (P = 0.000 < 0.001).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.428	.165		-26.788	.000	-4.753	-4.104
	Log10GA	3.075	.106	.717	29.045	.000	2.867	3.283

a. Dependent Variable: Log10Birtweight

The regression equation is

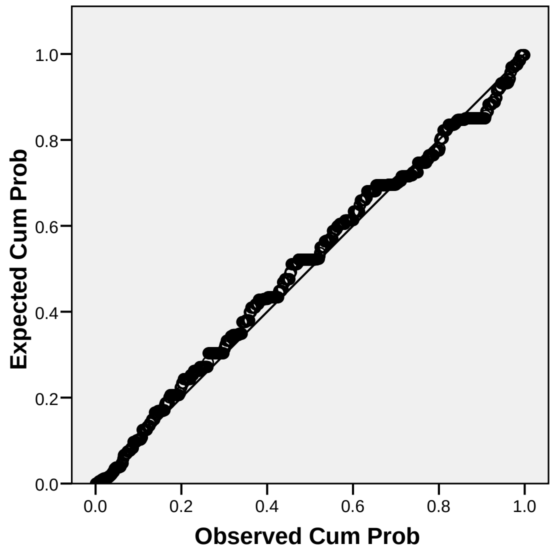
$$Log_{10}(Birth\ weight) = -4.428 + 3.075 \times Log_{10}(GA)$$

The antilog of -4.428 is 0.0000373 and the equation is the same as obtained before.

First panel of Figure C.5(b) displays the Normal P-P plot to study the Gaussianity of the residuals. There is some deviation around 0.9 otherwise all the points lie on the expected line. This confirms that the deviation from Gaussianity is not large, and establishes that the tests of significance and CIs are not invalid.

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Log10Birtweight



Dependent Variable: Log10BIRTHWEIGHT

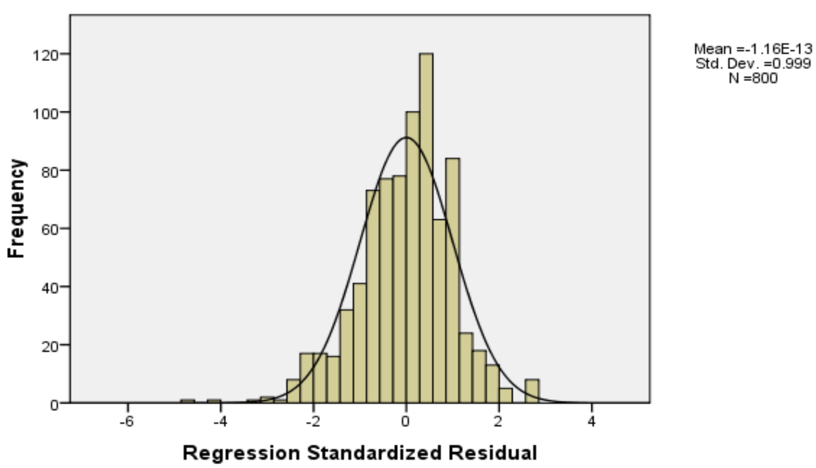


FIGURE C.5(b) Normal P-P plot and histogram of residuals for checking the Gaussian distribution of residuals

Command to calculate the observed and predicted mean birth weight at each gestational age and to draw the graph using the power regression:

```
AGGREGATE
  /OUTFILE=*
  MODE=ADDVARIABLES
  /BREAK=GA
  /BirthWeight_mean_1 = MEAN(BirthWeight).
COMPUTE predicted_BW = 0.0000373*GA ** 3.075.
EXECUTE.
GRAPH
  /SCATTERPLOT(OVERLAY)=GA GA  WITH BirthWeight_mean_1 predicted_BW (PAIR)
  /MISSING=LISTWISE.
```

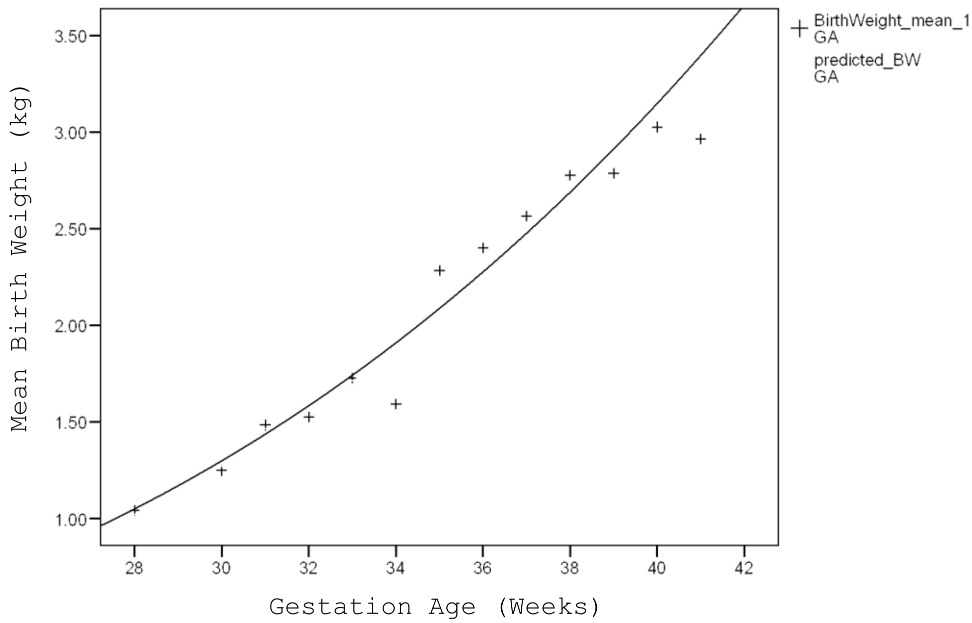


FIGURE C.5(c) Observed and predicted means of birth weight by power regression on gestational age

To avoid complexity of plot of 800 points, this graph shows only the mean birth weight for each gestation age. The fitted regression curve shows faster than linear rise in birth weight as the gestation age increased. Examine biological admissibility of this possibility before accepting the regression curve.