MATH 101
Lia Vas

Polynomial Regressions. Finding the Best Model

Quadratic model. Choosing option **QuadReg** in **STAT** menu, you can make a quadratic model for your data. This model fits your points on the parabola $ax^2 + bx + c$

Cubic model. Choosing option **CubicReg** in **STAT**, you can make a cubic model for your data. This model fits your points on the curve $ax^3 + bx^2 + cx + d$.

Quartic model. Choosing option **QuartReg** in **STAT**, you can make a quartic model for your data. This model fits your points on the curve $ax^4 + bx^3 + cx^2 + dx + e$.

Power model. Choosing option **PwrReg** in **STAT**, you can make a power model for your data. This model fits your points on the curve ax^b.

To graph two regression curves on the same plot, enter your data in a list and then place one curve in Y_1 and the other in Y_2. For example, to graph both the linear and quadratic regression curve, use **LinReg(ax+b)** $\mathbf{Y_1}$ and **QuadReg** $\mathbf{Y_2}$. This will give you both linear (in Y_1) and quadratic (in Y_2) model of your data.

Finding the best model: The **coefficient of determination** R^2 tells you how well your model describes the data. It takes values between 0 and 1 and the closer it is to 1, the better your data fit the regression curve. It is important to keep in mind that the value of R^2 alone is not the only factor that you should consider when choosing the best model. Other factors include:

- The validity of the model.
- The effectiveness of the model.

The examples below illustrate these issues.

Testing the validity of a model

Problem. A projectile is fired upwards from the ground. The height of the projectile above the ground is shown in the following table:

Time (seconds)	0	0.5	1	1.5	2	2.5
Height (feet)	0	20.5	31.36	36.25	30.41	28.23

a) Find a good model to fit this data. b) Find the time at which the projectile hit the ground.

Discussion. The linear, exponential and logarithmic models are not good fits as they are always increasing or decreasing and the data is first increasing and then decreasing. So, you can start by finding the quadratic, cubic and quartic models. Look at the graphs. Note that the cubic and quartic polynomial start increasing instead of decreasing after the maximum height is reached. Because of this, these models are not very appropriate. The quadratic model seems to fit the data and the reality of the situation the best. For this model $R^2 = 0.9738$ which is pretty close to 1 so this supports the decision to use this model.

Solution. a) Quadratic model is $y = -12.87x^2 + 42.22x + 1.177$ and $R^2 = 0.9738$. b) The object falls to the ground 3.31 second after it is thrown up.

Testing the effectiveness of a model

Problem. Healthcare costs have been increasing over the years. The following data shows the average cost of healthcare per person from 1976 to 1998:

Year	1976	1980	1987	1993	1998
Cost (per person)	618	860	1324	1865	2256

a) Find a model that fits the data well. b) Find the time the average healthcare cost will reach \$2800 per person.

Discussion. Let $x = 0$ denotes the year 1970. Find the quadratic, cubic and quartic model. Look at the graphs. Compare R^2 for all models. Note that cubic model increases complexity of the equation without changing the value of R^2 significantly (also it starts to decrease at some point and the data does not indicate that). The quartic model, although accurate because $R^2 = 1$ also starts decreasing which does not fit the data. The exponential model has $R^2 = .988$ so a good case could be made for choosing it. The logarithmic model is not appropriate as the scatterplot indicates concave up data, not concave down. Thus, the quadratic model is the best.

Solution. 3. a) Quadratic Model: $y = .7686x^2 + 49.204x + 290.8067$, $R^2 = .9988$ b) $y = 2800$ when $x = 33.48$. So, by year 2004, the healthcare cost will be over \$2800 per person.

Practice Problems.

1. The population present in a bacteria culture over 5 days is given in the table below:

time (days)	0	1	2	3	4	5
population	30	133	214	337	527	819

   a) Find a good model for the data. b) Estimate the population after 7 days.

2. A company decides to develop a cost equation based on the quantity of the product produced in a day. They collected the following data:

quantity produced	20	35	50	65	80	95	110
cost	642.35	766.48	858.82	928.83	1005.32	1078.82	1140.79

   a) Find a good model for this data. b) According to the model, how much will producing 195 units cost the company? c) How many units could be produced for \$800?

3. The table below shows the yield (in mg) of a chemical reaction in the first 6 minutes.

time (minutes)	1	2	3	4	5	6
yield (mg)	1.2	6.9	9.3	12.7	14.1	15.7

   a) Use the scatterplot to find the best model to fit this data. b) Using that model, determine in how many minutes will the yield be 20 mg.

4. The table below shows the concentration of a drug in a patients bloodstream t hours after it was administered.

time (hours)	.5	1	1.5	2	2.5	3
concentration (mg/cc)	.16	.19	.2	.19	.18	.17

Find a cubic model to fit this data. When was the concentration .1825 mg/cc?

5. After the winter break, 3 students came to school sick with the flu. The following table shows the number of students infected with the flu depending on the number of days after the winter break.

time (days)	0	5	10	15	20	25	30
number of infected students	3	6	14	23	23	21	9

Find the quadratic and quartic model that fit this data. Which model appears to better fit the data? Using the better model, find the day at which the number of infected students will reach the maximum. When will the number of infected students drop to zero?

Solutions

1. a) The cubic model is a slightly better fit than quadratic. For cubic $R^2 = .9999$. The quartic model has almost the same R^2 as cubic so it increases complexity without adding much accuracy. Logarithmic is concave down instead of up and exponential does not have very high R^2. b) 1886 bacteria.

2. a) The leading coefficient of cubic model is $3.25 \cdot 10^{-4}$. Since this number is relatively small, this means that the cubic curve is almost a quadratic curve. R^2 of quadratic is .997 and the cubic has $R^2 = .9993$ which suggests that they are almost equally efficient. Considering effectiveness, we are leaning towards quadratic. However, the quadratic curve starts decreasing after about $x = 200$. Thus, if it is to predict the cost if more than 200 items are produced, the cubic model is better. If the production is not expected to exceed 200 items, the quadratic is better since it is more efficient.

   The leading coefficient of quartic model is $-9.7 \cdot 10^{-6}$. As this number is very small, this means that this curve is almost a cubic. Thus, when deciding between these two models, cubic seems to be a better choice. Also, the quartic model starts decreasing and the data does not indicate that should happen.

   b) Using quadratic: 195 units cost \$1287. c) Using quadratic: about 42 units. Using cubic: about 41 units.

3. The quartic and logarithmic model have almost the same R^2. The other models are not very accurate. As logarithmic is simpler than quartic (especially considering that taking $\ln x$-values instead of x-values would give you a linear model), we can choose logarithmic. b) 10.34 min.

4. In .8 hours and 2.4 hours.

5. Quartic is better. Maximum 19.5 days after. No infected students 33 days after.