Regression and Transformed Data

Section 4.1 – Exponential Models

Consider the following data of average movie ticket prices.

(source: http://natoonline.org/data/ticket-price/)

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Cost of Movie Ticket** | **Log(ticketcost)** | **Ln(ticketcost)** |
| 1948 | $0.36 | **-.4437** | **-1.022** |
| 1958 | $0.68 | **-.1675** | **-.3857** |
| 1968 | $1.22 | **0.08636** | **.19885** |
| 1978 | $2.37 | **.37475** | **.86289** |
| 1988 | $4.11 | **.61384** | **1.4134** |

2. Draw a rough scatterplot.

1. Note, the data **do not** appear to be linear. Calculate r2 and make a residual plot in your calculator. What do these two items tell you about the linearity of our original data?

r2 is 0.91, which does not demonstrate that my data is nonlinear. However, the residual plot reveals a clear curved pattern. Paired with the fact that the original scatterplot appears curved, I can say that this data is not linear.

1. I believe that this data is growing exponentially.

|  |
| --- |
| **Ratios** |
| **1.89** |
| **1.79** |
| **1.94** |
| **1.73** |

Test this theory by checking the ratios $\frac{y\_{n}}{y\_{n-1}}$.

Since all of these ratios are close to the same number, I am reasonably happy that it may be best modeled by an exponential equation.

1. Since we believe the data might be exponential, take the log of only the y variable. Fill in the chart at the top of the page.

Done. (I also took the natural logs … you don’t need to do both, I am just showing both so that you can check your answers no matter what you did!)

5. Draw a rough scatterplot of this transformed data.

Woohoo! Look how linear that transformed data is…NICE =)

1. Find the LSRL of the transformed data. Is this a good idea? Reference your transformed scatterplot and r2?

$$\hat{y}=-52.203267+0.0265732x$$

r2 =.9993

The line on the transformed data does a good job according to r2 (which is higher or my transformed data than the original data). Also, my transformed scatterplot appears linear and the ratios calculated earlier indicate exponential growth

7. Draw a rough residual plot. What do you think about this plot?



My residual plot provides further evidence that the exponential model is a good fit for this data. The scattered points show no pattern, indicating that my exponential transformation was successful in linearizing the data.

1. Use your regression line to find a prediction of the cost of movie tickets today.
2. $\hat{y}=-52.203267+0.0265732x$

**ln**$\hat{y}$ $=-120.2024654$$+.0611870099x$ **log**$\hat{y}$ **=** $-52.203267$$+0.0265732x$

**ln**$\hat{y}$ **=** $=-120.2024654$$+.0611870099(2014)$ **log**$\hat{y}$ **=** $-52.203267$$+0.0265732(2014)$

**ln**$\hat{y}$ **= 3.028172539 log**$\hat{y}$ **= 1.3151186**

$e^{ln\hat{y}}$**=**$e^{3.028172539}$$10^{log\hat{y}}$**=**$10^{1.3151578}$

$\hat{y}=$**$20.66** $\hat{y}=\$$**20.66**

\*\*Rounding makes a big difference when we are working with numbers like this…that’s why you show your work ☺\*\*

9. Look up the average cost of a movie ticket today. Why is our prediction so far from the actual cost?

10. "Untransform" your LSRL to find a regression equation in terms of x and y instead of x and log y.

**log**$\hat{y}$ **=** $-52.203267$$+0.0265732x$

$10^{log\hat{y} }$**=**$10^{-52.203267 +0.0265732x}$

$$\hat{y}= 6.26×10^{-53}\*1.063098^{x}$$

**ln**$\hat{y}$ **=** $-120.2024654$$+.0611870099x$

$e^{ln\hat{y} }$**=**$e^{-120.2024654 +.0611870099x}$

$$\hat{y}= 6.26×10^{-53}\*1.063098^{x}$$