

بنام خدا

Hossein Vatani

April 24, 2016

Transformation and Polynomia Regression

پسرفت ترادیس شده و چند جمله ای

یکی از اولین مراحل تشکیل نمونه پسرفت (Regression Model)، حدس شکل تابع پسرفت است. ما می توانیم دامنه نمونه (سازی) را با در نظر گرفتن متغیرهای توضیحی خاصی، بطور چشمگیری افزایش دهیم. این شامل متغیرهای شاخص، ترادیس شده (همان تبدیل شده که در ادامه با مفهوم آن آشنا خواهیم شد) و جملات بالاتر و همینطور اثر متقابل آنها خواهد بود. در این بحث ما تنها در مورد ترادیسی و مراتب بالاتر متغیرهای توضیح صحبت خواهیم کرد.

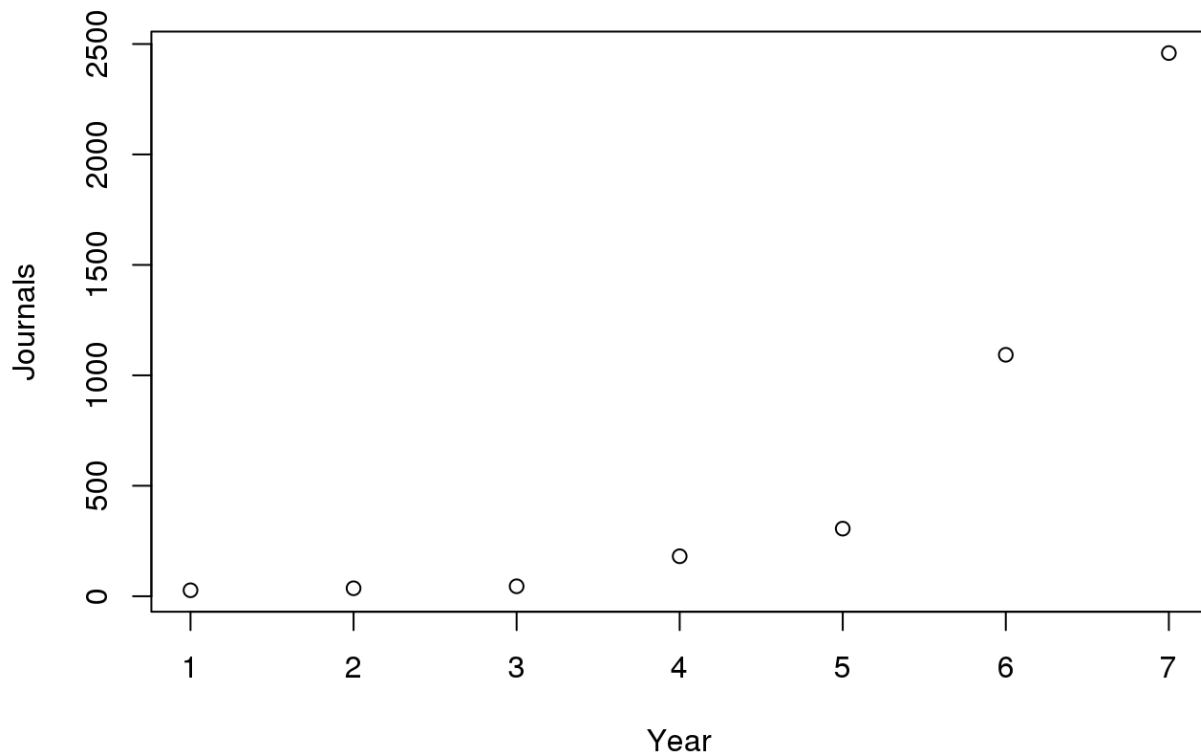
الف-ترادیسی داده (Transforming Data)

زمانیکه یک نمونه پسرفت خطی را ایجاد می کنیم، در واقع با فرض وجود رابطه خطی بین متغیرهای تشریح و پاسخ اینکار را انجام داده ایم. بهر حال، در خیلی از موارد رابطهای غیر خطی بین این دو نوع از متغیر وجود دارد. بعضی اوقات برای درمان این درد! بکار گیری تابع دگرگون ساز (ترادیس) مناسب برای یک یا چند ویا همه متغیرهای تشریح چاره ساز است؛ مانند تابع توان و یا لگاریتم. علاوه بر آن، ترادیسی می تواند در تصحیح تخلف از پیش فرض های نمونه سازی مانند پراکنش خطای ثابت (constant error variance) ویا نرمال بودن، جلوگیری کند.

ما ترادیس کردن را قبل از نمونه سازی انجام میدهم که اینکار برای ایجاد نمونه مناسب و برقراری رابطه خطی بین متغیرها کافیست.

مثال-تعداد نشریه های علمی چاپ شده در سطح جهانی از سال ۱۹۹۱ تا ۱۹۹۷ به شرح زیر می باشد.

Year	Journals
1	27.00
2	36.00
3	45.00
4	181.00
5	306.00
6	1093.00
7	2459.00



باتوجه به نمودار بوضوح مشخص است که رابطه ایندو متغیر غیر خطی است وبصورت نمایی می باشد. می توان اظهار کرد که لگاریتم تعداد نشریات چاپ شده، رابطه ای خطی با سال را ایجاد می کند.

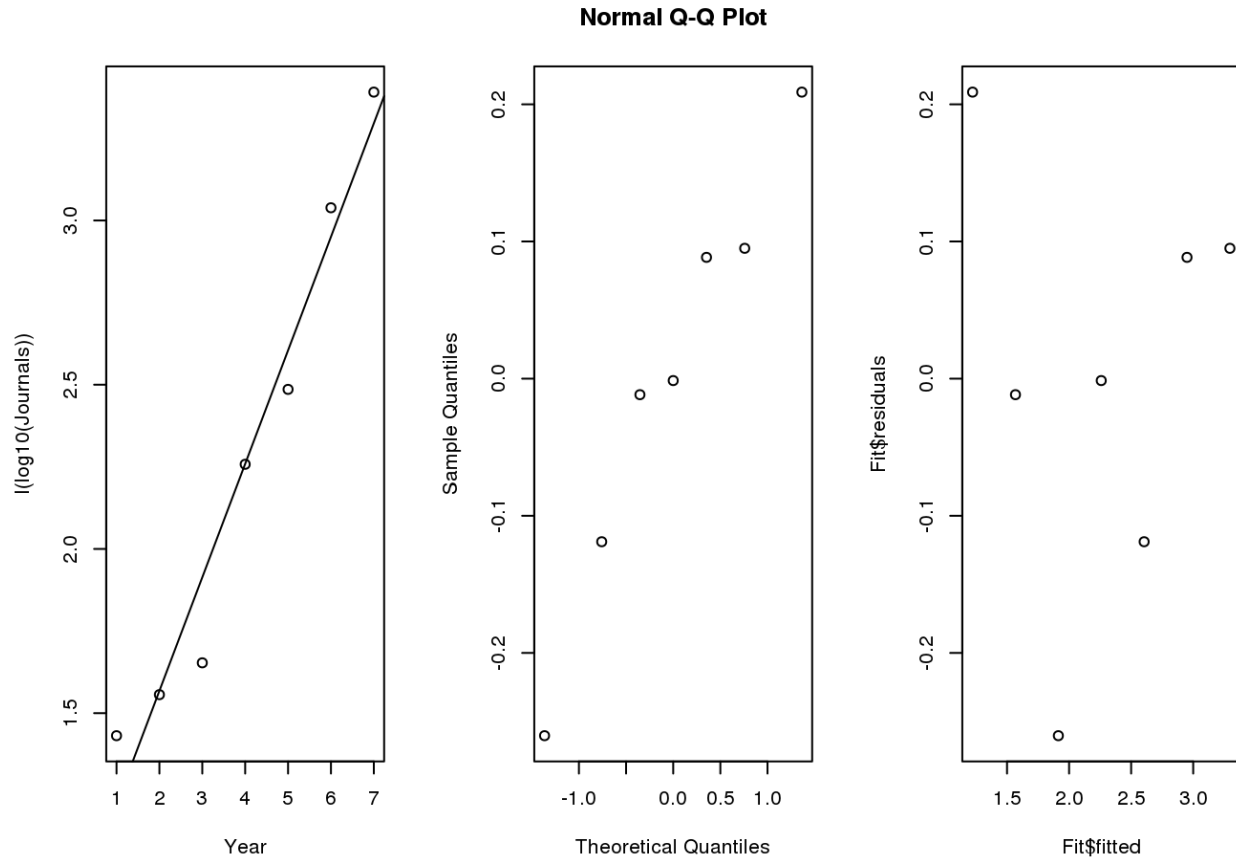
نکته: در زبان R می توان تمام آنچه را که وجود دارد به اصطلاح برنامه نویسان بازنویسی (overwrite) کرد. لذا در مواقعی که می خواهیم مطمئن باشیم هنگام استفاده از دستورات پایه مانند «+»، «-»، «*»، «/» ... و یا در اینجا «log10» دقیقا همان عملگرهای پایه می باشند، از تابعی بنام `as.is()` که تابع "As is" می باشد استفاده می کنیم.

```
Fit=lm(I(log10(Journals))~Year,data=Internet)
summary(Fit)
```

```
##
## Call:
## lm(formula = I(log10(Journals)) ~ Year, data = Internet)
##
## Residuals:
##      1      2      3      4      5      6      7
## 0.208912 -0.011697 -0.260334 -0.001415 -0.118920  0.088431  0.095022
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.87690     0.14270   6.145 0.001659 **
## Year         0.34555     0.03191  10.829 0.000117 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1688 on 5 degrees of freedom
## Multiple R-squared:  0.9591, Adjusted R-squared:  0.9509
## F-statistic: 117.3 on 1 and 5 DF,  p-value: 0.0001165
```

باتوجه به خروجی نمونه پیش بینی ساخته شده بصورت $\log(\hat{y}) = 0.8769 + 0.3455x$ می باشد. و در ادامه جهت بررسی درستی این نمونه داریم

```
par(mfrow=c(1,3))
plot(Year,I(log10(Journals)),data=Internet)
abline(Fit)
qqnorm(Fit$residuals)
plot(Fit$fitted,Fit$residuals)
```



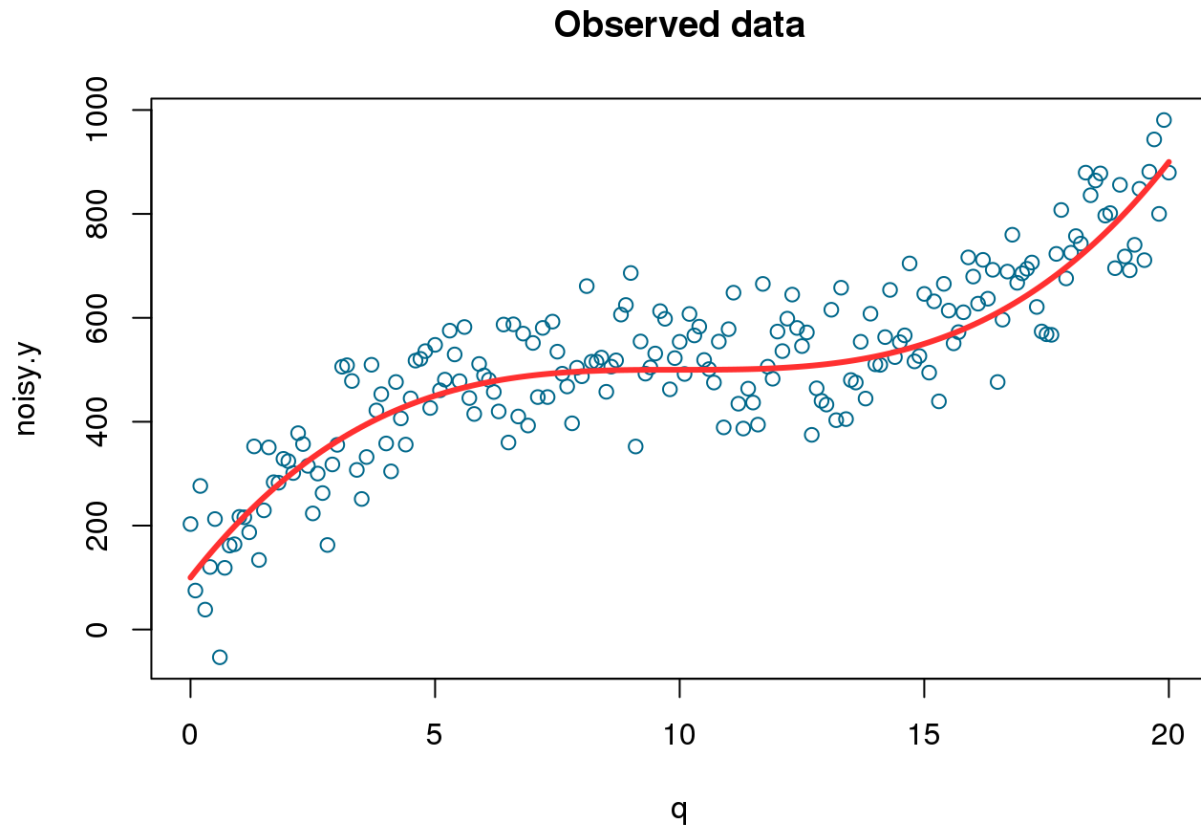
توجه شود که اگر بخواهیم از رابطه بدست آمده برای پیش بینی استفاده کنیم، پس از حصول جواب باید عدد ۱۰ را بتوان جواب برسانیم.

ب-پسرفت چندجمله ای (Polynomial Regression)

نمونه پسرفت چندجمله ای (Polynomial Regression) زمانی کاربرد دارد که بتوانیم بگوییم رابطه بین متغیرها بصورت منحنی می باشد. نمونه پسرفت مرتبه p به نمونه ای گفته می شود که رابطه متغیر تشریح با متغیر پاسخ چند جمله ای از مرتبه p باشد.

مثال- (فایل موجود نیست) برای این مثال داده سازی زیر را انجام می دهیم

```
set.seed(20)
q=seq(0, 20, 0.1)
y=500+ 0.4* (q-10)^3
noise <- rnorm(length(q), mean=10, sd=80)
noisy.y <- y + noise
plot(q,noisy.y,col='deepskyblue4',xlab='q',main='Observed data')
lines(q,y,col='firebrick1',lwd=3)
```



حال نمونه مورد نظر را ایجاد می کنیم. به تابع poly توجه شود.

```
Fit=lm(noisy.y~q + I(q^2) + I(q^3))  
summary(Fit)
```

```
##
## Call:
## lm(formula = noisy.y ~ q + I(q^2) + I(q^3))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -212.326  -51.186    4.276   61.485  165.960
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   87.98080   21.99621    4.00 8.96e-05 ***
## q             125.38245    9.54842   13.13 < 2e-16 ***
## I(q^2)        -12.16859    1.11097  -10.95 < 2e-16 ***
## I(q^3)         0.39719    0.03651   10.88 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.42 on 197 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.8001
## F-statistic: 267.8 on 3 and 197 DF,  p-value: < 2.2e-16
```

اگر فقط ضرایب و عدد ثابت رابطه پیش بینی را بخواهیم می توان از فرمان زیر استفاده کرد

Fit

```
##
## Call:
## lm(formula = noisy.y ~ q + I(q^2) + I(q^3))
##
## Coefficients:
## (Intercept)          q          I(q^2)          I(q^3)
##      87.9808      125.3824      -12.1686       0.3972
```

با توجه به خروجی رابطه مورد نظر جهت پیش بینی را می توان اینگونه بیان نمود:
 $\hat{y} = 513.6 + 2075.9q - 108.0q^2 + 864.0q^3$ و در نهایت نمودار را با خط برازش شده نمونه پسرفت نمایش می دهیم.

```
plot(q,noisy.y,col='deepskyblue4',xlab='q',main='Observed data')
lines(q,Fit$coefficients[1]+Fit$coefficients[2]*q+Fit$coefficients[3]*q^2+Fit$coefficients[4]*q^3)
```

