

# On Song Popularity: An Analysis on Song Metadata Using the Million Song Dataset

Stephen Yan\*

(Dated: c. December 2015)

We were tasked with retrieving a unique dataset on which to perform statistical analyses of which there were multiple predictors and at least twenty observations. Using *SQL*, we chose a random sample of songs and their metadata from a public music database on which to perform our analyses and used *SAS* to visualize and manipulate and model our data. We use six predictors: artist hotness, familiarity, song duration, loudness, tempo, and release year to predict song hotness. We run *SAS* implementations for visualizing basic statistics of our data and running multilinear regressions and transformations on our data to determine an appropriate model for song hotness.

## I. THE DATASET

The dataset was obtained from *The Million Song Dataset*<sup>1</sup>, a public dataset of over a million songs used for music analysis and research purposes. Each song in the collection contains over 50 properties ranging from *danceability* and *energy* indices. For our purposes, we will look at *song hotness* as the response variable and use 6 other variables as predictors with 30 observations. A full list of the metadata attributes is also available online.<sup>2</sup> The variables are as follows:

hotIdx	Song hotness index, according to The Echo Nest, when downloaded (December 2010)
artistHotIdx	Artist hotness index, according to The Echo Nest, when downloaded (December 2010)
artistFamIdx	Artist familiar index, according to The Echo Nest, when downloaded (December 2010)
duration	Duration of the track in seconds
loudness	General loudness of the track
tempo	Tempo in beats per minute according to The Echo Nest
year	Year when this song was released, according to musicbrainz.org

## A. Visualization

We begin by running procedures to visualize our data. We produce 30 song observations randomly sampled from the data subset. The full raw data is available in FIG 1.

---

```
data songs;
infile '\\Client\H$\Dropbox\UNC\5\STOR
455\Songs.DAT' dlm='09'x;
input hotIdx artistHotIdx artistFamIdx duration
loudness tempo year;
run;
proc print data=songs;
run;
```

---

Next we wish to obtain some basic statistics on the data and investigate plots against the explanatory variables.

---

```
proc means data=songs;
var hotIdx artistHotIdx artistFamIdx duration
loudness tempo year;
run;
proc gplot data=songs;
plot hotIdx*(artistHotIdx artistFamIdx duration
loudness tempo year);
run;
proc univariate data=songs alpha=0.5;
var hotIdx;
histogram / endpoints=0 to 1 by 0.1;
run;
proc corr data=songs;
var hotIdx artistHotIdx artistFamIdx duration
loudness tempo year;
run;
```

---

On inspection, we observe that *hotIdx* is a roughly bimodal distribution of indices varying generally between 0.2 and 0.9. There are no major outliers and data points for other variables are rather evenly distributed.

We also observed the correlation matrix between variables and there are very few outstanding collinearities except for the relationship between *artistHotIdx* and *artistFamIdx*. We will explore this in depth later.

---

\* sjyan@cs.unc.edu

<sup>1</sup> Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011.

<sup>2</sup> <http://labrosa.ee.columbia.edu/millionsong/sites/default/files/poster.pdf>

## II. MODELING TRANSFORMATIONS

Next we attempt to model the multilinear regression with all variables and no linear transformations.

---

```
proc reg data=songs;
model hotIdx = artistHotIdx artistFamIdx duration
      loudness tempo year / ss1 ss2 VIF TOL r
      influence;
run;
```

---

We obtain  $[Rsqr, adjRsqr]$  as  $[0.3825, 0.2214]$ . We also observe that the variance inflation is moderately high for variables `artistHotIdx` and `artistFamIdx`, indicating that we may have to account for this in future transformations and selections.

Immediately, we attempt to transform the variables to improve the deficiencies in our plain model by log-transforming `hotIdx`.

---

```
data logSongs;
set songs;
logIndex = log(hotIdx);
run;
proc reg data=logSongs;
model logIndex = artistHotIdx artistFamIdx duration
      loudness tempo year / ss1 ss2 VIF TOL;
run;
```

---

The results yield an improvement in correlation. This configuration gives us  $[Rsqr, adjRsqr]$  as  $[0.4097, 0.2557]$ . Our variance inflation across the board looks roughly the same as before, implying little or no change in multicollinearity. We take these results into consideration and think of other ways to transform the data.

We attempt to identify an appropriate exponent  $\lambda$  for data  $Y^k$  by a *Box-Cox power transformation*. We use built-in SAS tools to accomplish this.

---

```
proc transreg data=songs;
model BoxCox(hotIdx / convenient lambda=-2 to 2 by
      0.05)
      = qqpoint(artistHotIdx artistFamIdx duration
      loudness tempo year);
run;
```

---

From these results, we obtain  $\lambda$  as  $-2$  and transform our data accordingly.

---

```
*Select lambda as -3;
data songsK;
set songs;
hIndexRT = hotIdx**(-3);
run;
proc reg data=songsK;
model hIndexRT = artistHotIdx artistFamIdx duration
      loudness tempo year / ss1 ss2 VIF TOL;
run;
```

---

Yet again, this transformation yields us a slight improvement in correlation to  $[Rsqr, adjRsqr]$  as

$[0.4329, 0.2850]$  with no improvement in multicollinearity. We accept this as a final model of all variables and proceed to selection.

## III. MODEL SELECTION

Before we proceed, we visualize a correlation matrix (FIG. 7) to observe any outstanding collinearities to consider when we perform model selection.

---

```
proc sgscatter data=songsK;
matrix hIndexRT artistHotIdx artistFamIdx duration
      loudness tempo year;
run;
```

---

Our beliefs that `artistHotIdx` and `artistFamIdx` are strongly correlated are more confirmed with this visualization as evidenced by plots (2, 3) or (3, 2). Thus we have a few options - we considered the brute option of completely omitting these two variables, however it is likely that they influence the response in some similar way. Formally, we could perform factor analysis to determine the true linear combination of these two variables, but I opt to simply take the difference (`artistFamIdx - artistHotIdx`) and use that as variable `aIndex` when we perform model selection.

---

```
data songsAdj;
set songsK;
aIndex = artistFamIdx - artistHotIdx;
run;
```

---

We are now ready to select a proper model. We will perform *Forward Selection*, *Backward Elimination*, *Stepwise Selection* on the revised dataset (with `aIndex`) and *Stepwise Selection* on the original dataset with respective variables `artistHotIdx` and `artistFamIdx`. We also perform *Maximum Rsqr Improvement Selection*, *Adjusted Rsqr Selection*, and *Mallows'  $C_p$  Selection*.

---

```
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
      duration loudness tempo year /
      selection=FORWARD VIF TOL slentry=0.1;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
      duration loudness tempo year /
      selection=B VIF TOL slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
      duration loudness tempo year /
      selection=STEPWISE VIF TOL slentry=0.15
      slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = artistHotIdx artistFamIdx duration
      loudness tempo year /
      selection=STEPWISE VIF TOL slentry=0.15
```

---

```

slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
duration loudness tempo year /
selection=MAXR VIF TOL;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
duration loudness tempo year /
selection=adjrsq VIF TOL;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx
duration loudness tempo year /
selection=cp VIF TOL;
run;
run;

```

We obtain the following results for each selection algorithm in the format `selection algorithm: (variables), [Rsqr, adjRsqr]`.

```

forward:
(aIndex, duration), [0.2940, 0.2417]

backward:
(aIndex, duration, loudness, year),
[0.4091, 0.3146]

stepwise (aIndex):
(aIndex, duration, loudness, year),
[0.4091, 0.3146]

stepwise (w/o aIndex):
(artistFamIdx), [0.1472, 0.1167]

maxR:
(aIndex, artistFamIdx, duration, loudness,
tempo, year), [0.4329, 0.2850]

adjRsqr:
(aIndex, duration, loudness, year),
[0.4091, 0.3146]

CP:
(aIndex, duration, loudness, year),
[0.4091, 0.3146]

```

The majority of selections yielded the variable configuration `aIndex, duration, loudness, year` with variance inflations corrected to  $< 1.6$ . Although our `Rsqr` coefficient decreased nominally, we can take solace in the fact that *Maximum Rsqr Improvement Selection* only yielded a slightly higher statistic, however with a lower `adjRsqr` statistic in addition to wild variance inflation.

We can now apply a final multilinear regression onto this revised data and plot the residual v. response and normal quantile plots with studentized residuals (FIG. 15) for a final check on linearity.

---

```

proc reg data=songsAdj;
model hIndexRT = aIndex duration loudness year /
ss1 ss2 VIF TOL collin;
plot student.*nqq.;
run;

```

---

#### IV. CONCLUSIVE REMARKS & IMPROVEMENTS

The final model we obtain for the data is:

$$\text{songHotness} = -1465.7 - 226.7 * \text{aggregateArtistIndex} + 0.1622 * \text{duration} - 0.9318 * \text{loudness} + 0.475 * \text{year}$$

Because of the relatively poor correlation coefficients observed throughout our analyses, we suspect that the relationship between song hotness v. artist popularity, song duration, loudness, and release year is not exactly linear, but our residual and normal quantile plots suggest that it is an appropriate fit. Thus it might be more prudent to suspect that there was not enough evidence or enough data to illustrate a good enough model (subset of 30 as opposed to a subset of 100,00+?). The variables aggregate artist index, song duration, loudness, and release year explain 40.91% of the variation in song hotness according to our model. Prior to our analyses, intuition held that song hotness would be most favorably predicted by tempo and loudness of the music - but even then in hindsight, that may very well not be a linear correlation seeing as that both extremely loud and soft songs are unfavorable in comparison to songs with a *sweetspot* timbre and tempo.

For future reference, it is vital to wisely choose the predictors for the response especially when choosing a larger amount of predictors in order to avoid multicollinearity. For our dataset, there were variables such as *danceability* and *key* that could have been explored as well, but we chose the more seemingly stable variables.

Because the song hotness, artist hotness, and artist familiarity indices were collected and assessed by one entity, there may be reason enough to suspect that bias and user anonymity play a role in the discord of our regression.

There were also other ways to transform the data - in applying the Box-Cox method, it is possible that not choosing a sufficiently large enough range resulted in a skewed representation of our model. For our purposes, we chose a power transformation over the log transformation even though both were improvements from the original data.

---

## Appendix

---

FIG. 1. Raw song dataset

Observation	Song Title	Artist
1	Sea_ Subsurface	Aisha Duo
2	Nothing (Album Version)	unloco
3	Rowena	Loudon Wainwright III
4	The Forgotten Lands	David Arkenstone
5	Mais La Vie	Maurane / Lara Fabian
6	Between You And Me	Joy Enriquez
7	Standing On My Own	Agnostic Front
8	Nothing To You	Milk Inc.
9	Hope (Edited Album Version)	Twista FeaturingCee-Lo
10	True Nature	Jane's Addiction
11	Lord of Shadows and Golden Wood	Beherit
12	Looking In Your Eyes	Suicidal Tendencies
13	High Horse	Workhorse
14	Cumberland Gap	Danny Barnes
15	Ripped Pants	Spongebob Squarepants
16	Becoming (LP Version)	Lennie Tristano
17	Zap Zap	Dynamite Deluxe
18	Fortress	Tim Exile
19	Whiskey on the Fire	Aaron Watson
20	The Blinding Sun	Gustavo Santaolalla
21	Gossippo Perpetuo	Perrey And Kingsley
22	Erick Sermon	Erick Sermon
23	Sag Ihnen Bescheid	Headliners
24	This Is Not A Love Song	Public Image Ltd
25	Vitamina	Sonora Carruseles
26	Speechless	Mish Mash
27	Think About Me	Sister Hazel
28	Turn The Heat Up	Bobaflex
29	You Gotta Go	Fountains Of Wayne
30	All About Our Love	Sade

The SAS System

11:21 Friday, December 4, 2015 87

Obs	hotIdx	artist HotIdx	artist FamIdx	duration	loudness	tempo	year
1	0.50132	0.41797	0.63807	350.641	21.339	132.888	2005
2	0.69029	0.40360	0.58828	163.866	-5.653	126.113	2001
3	0.38703	0.44277	0.66108	152.137	-12.464	151.294	2009
4	0.36037	0.43094	0.63940	314.592	14.881	141.975	1996
5	0.22577	0.41261	0.58048	198.870	-13.637	107.087	2003
6	0.26586	0.33739	0.59417	260.963	-3.941	96.087	2001
7	0.61018	0.50431	0.70144	105.848	-5.152	112.292	2001
8	0.46054	0.44310	0.65280	202.109	-7.597	140.017	2003
9	0.32774	0.59381	0.88961	272.509	-5.126	106.967	2004
10	0.63019	0.53516	0.71189	229.616	-3.782	92.099	2003
11	0.54155	0.42748	0.59618	202.919	-18.650	130.955	1993
12	0.55065	0.52618	0.73668	167.784	-9.966	166.090	1987
13	0.26586	0.32222	0.47902	251.402	-7.200	85.209	2007
14	0.23963	0.43525	0.47498	160.470	-13.579	189.782	2005
15	0.61561	0.45894	0.62300	76.826	-8.892	122.887	2005
16	0.35223	0.31785	0.51443	271.255	-15.324	200.876	1994
17	0.57785	0.41483	0.63340	86.674	-10.182	93.298	2000
18	0.53003	0.30185	0.62050	369.580	-6.555	110.000	2009
19	0.39201	0.47250	0.65802	171.102	-8.810	130.985	2008
20	0.63925	0.40663	0.64398	117.185	-13.505	90.463	2006
21	0.44440	0.35033	0.53061	127.947	-12.717	135.780	2001
22	0.40512	0.36834	0.57797	198.635	-9.384	97.812	1993
23	0.63823	0.32624	0.53063	216.894	-5.893	160.158	2005
24	0.87362	0.41552	0.68499	251.506	-10.347	127.369	1983
25	0.37598	0.36503	0.56035	230.008	-6.564	108.292	1999
26	0.33471	0.35121	0.50564	211.905	-7.963	124.043	2006
27	0.70717	0.50988	0.73107	188.630	-5.150	138.581	1997
28	0.66068	0.41541	0.58878	147.435	-3.124	93.033	2005
29	0.38527	0.53227	0.71776	126.876	-3.594	85.047	2007
30	0.64069	0.51734	0.77149	160.626	-8.211	161.991	2000

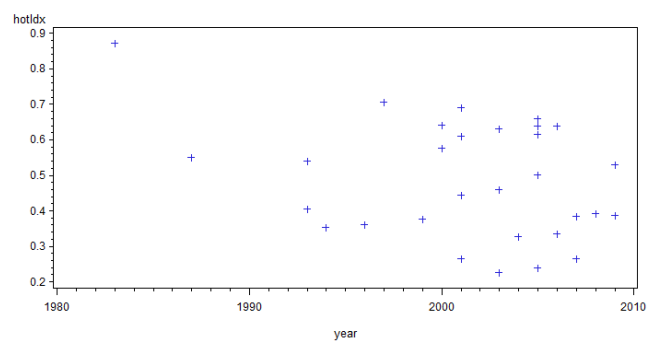
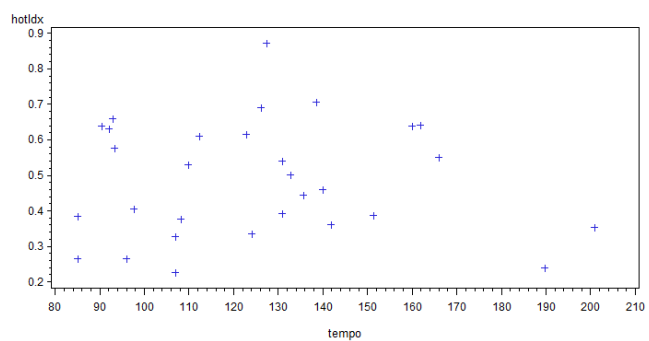
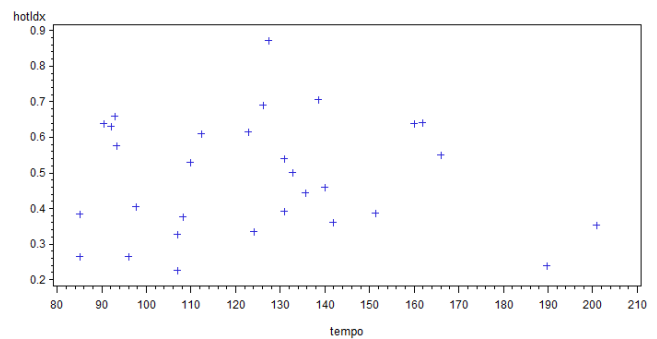
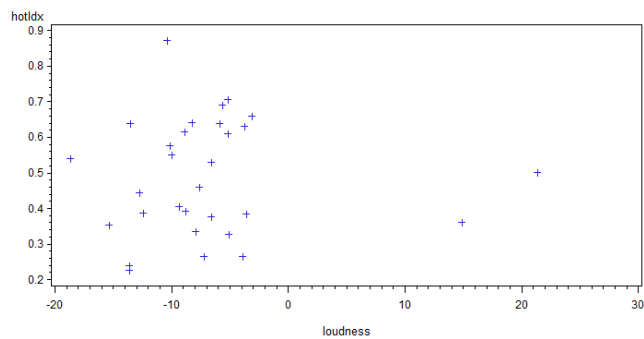
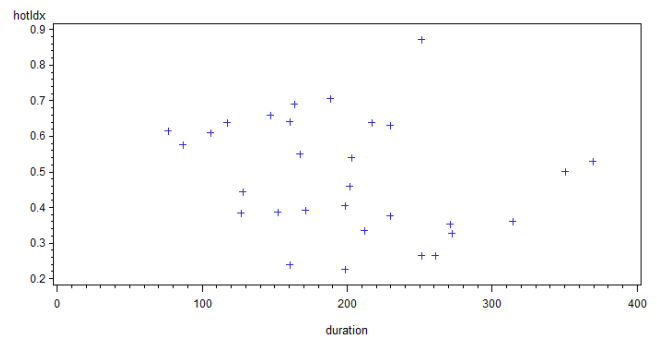
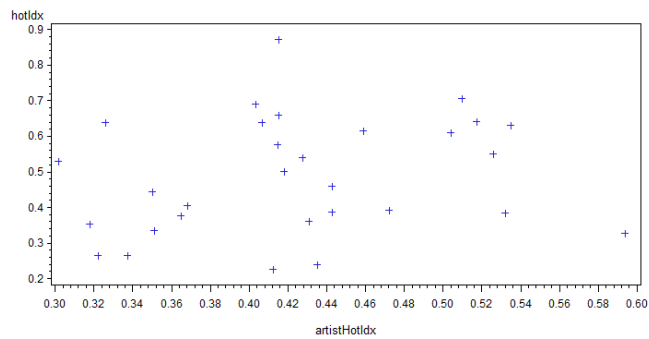
The SAS System

11:21 Friday, December 4, 2015 88

## The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
hotIdx	30	0.4876606	0.1629004	0.2257681	0.8736188
artistHotIdx	30	0.4252308	0.0748211	0.3018471	0.5938084
artistFamIdx	30	0.6278911	0.0920014	0.4749835	0.8896136
duration	30	199.5602540	72.3721651	76.8256700	369.5799500

loudness	30	-6.8914000	7.8439325	-18.6500000	21.3390000
tempo	30	125.3156667	30.0939508	85.0470000	200.8760000
year	30	2001.20	6.2774913	1983.00	2009.00



The SAS System 11:21 Friday, December 4, 2015 92

The UNIVARIATE Procedure  
Variable: hotIdx

Moments

N	30	Sum Weights	30
Mean	0.48766057	Sum Observations	14.6298172
Std Deviation	0.16290038	Variance	0.02653653
Skewness	0.23773367	Kurtosis	-0.6075412
Uncorrected SS	7.90394455	Corrected SS	0.76955947
Coeff Variation	33.4044594	Std Error Mean	0.0297414

Basic Statistical Measures

Location		Variability	
Mean	0.487661	Std Deviation	0.16290
Median	0.480928	Variance	0.02654
Mode	0.265861	Range	0.64785
		Interquartile Range	0.26982

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 16.39669	Pr >  t  <.0001
Sign	M 15	Pr >=  M  <.0001
Signed Rank	S 232.5	Pr >=  S  <.0001

Quantiles (Definition 5)

Level	Quantile
100% Max	0.873619
99%	0.873619
95%	0.707172
90%	0.675486
75% Q3	0.630188
50% Median	0.480928
25% Q1	0.360371
10%	0.265861
5%	0.239629
1%	0.225768
0% Min	0.225768

The SAS System 11:21 Friday, December 4, 2015 93

The UNIVARIATE Procedure  
Variable: hotIdx

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
0.225768	5	0.640688	30
0.239629	14	0.660681	28
0.265861	13	0.690291	2
0.265861	6	0.707172	27
0.327737	9	0.873619	24

The SAS System 11:21 Friday, December 4, 2015 94

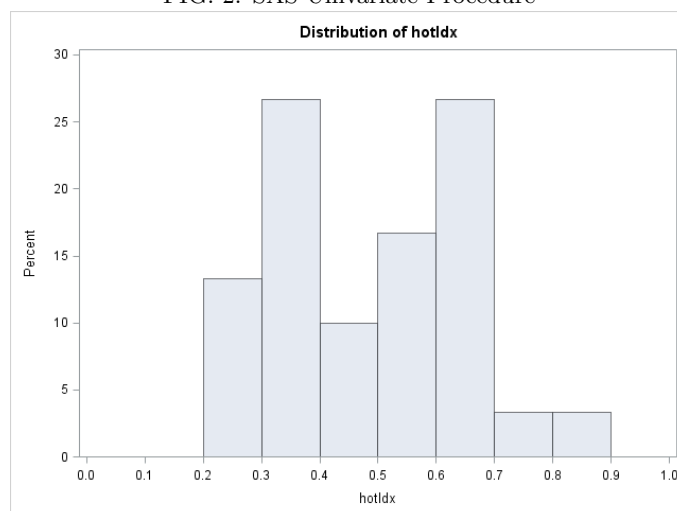
The CORR Procedure

7 Variables: hotIdx artistHotIdx artistFamIdx duration loudness tempo  
year

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
hotIdx	30	0.48766	0.16290	14.62982	0.22577	0.87362
artistHotIdx	30	0.42523	0.07482	12.75692	0.30185	0.59381

FIG. 2. SAS Univariate Procedure



artistFamIdx	30	0.62789	0.09200	18.83673	0.47498	0.88961
duration	30	199.56025	72.37217	5987	76.82567	369.57995
loudness	30	-6.89140	7.84393	-206.74200	-18.65000	21.33900
tempo	30	125.31567	30.09395	3759	85.04700	200.87600
year	30	2001	6.27749	60036	1983	2009

Pearson Correlation Coefficients, N = 30  
 Prob > |r| under H0: Rho=0

	hotIdx	artist HotIdx	artist FamIdx	duration	loudness	tempo	year
hotIdx	1.00000	0.20728 0.2717	0.33787 0.0678	-0.22984 0.2218	0.01381 0.9423	-0.02899 0.8791	-0.33400 0.0712
artistHotIdx	0.20728 0.2717	1.00000	0.83957 <.0001	-0.31435 0.0907	0.10520 0.5801	-0.02217 0.9074	-0.06507 0.7326
artistFamIdx	0.33787 0.0678	0.83957 <.0001	1.00000	-0.04108 0.8294	0.18594 0.3252	-0.13965 0.4617	-0.13632 0.4726
duration	-0.22984 0.2218	-0.31435 0.0907	-0.04108 0.8294	1.00000	0.48934 0.0061	0.07897 0.6783	-0.07933 0.6769
loudness	0.01381 0.9423	0.10520 0.5801	0.18594 0.3252	0.48934 0.0061	1.00000	-0.12721 0.5029	0.12916 0.4963
tempo	-0.02899 0.8791	-0.02217 0.9074	-0.13965 0.4617	0.07897 0.6783	-0.12721 0.5029	1.00000	-0.26404 0.1586
year	-0.33400 0.0712	-0.06507 0.7326	-0.13632 0.4726	-0.07933 0.6769	0.12916 0.4963	-0.26404 0.1586	1.00000

The SAS System 11:21 Friday, December 4, 2015 95

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hotIdx

Number of Observations Read 30  
 Number of Observations Used 30

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.29437	0.04906	2.37	0.0622
Error	23	0.47519	0.02066		
Corrected Total	29	0.76956			

Root MSE 0.14374 R-Square 0.3825



Dependent Mean      0.48766      Adj R-Sq      0.2214  
 Coeff Var            29.47480

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS	Type II SS
Intercept	1	17.71117	9.30394	1.90	0.0695	7.13439	0.07487
artistHotIdx	1	-1.61522	0.82746	-1.95	0.0632	0.03306	0.07872
artistFamIdx	1	1.50343	0.63500	2.37	0.0267	0.07000	0.11581
duration	1	-0.00134	0.00053491	-2.50	0.0198	0.08733	0.12961
loudness	1	0.00575	0.00430	1.34	0.1946	0.01476	0.03688
tempo	1	0.00036694	0.00099115	0.37	0.7146	0.01700	0.00283
year	1	-0.00860	0.00460	-1.87	0.0743	0.07222	0.07222

## Parameter Estimates

Variable	DF	Tolerance	Variance Inflation
Intercept	1	.	0
artistHotIdx	1	0.18586	5.38030
artistFamIdx	1	0.20874	4.79069
duration	1	0.47537	2.10364
loudness	1	0.62601	1.59741
tempo	1	0.80076	1.24881
year	1	0.85355	1.17158
The SAS System			11:21 Friday, December 4, 2015 96

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hotIdx

## Output Statistics

Obs	Variable	Dependent Value	Predicted Mean	Std Error Predict	Residual	Std Error Residual	Student Residual	-2	-1	0	1	2	Cook's D
1	0.5013	0.4446	0.1026	0.0567	0.101	0.564		*					0.047
2	0.6903	0.5200	0.0371	0.1703	0.139	1.226		**					0.015
3	0.3870	0.4832	0.0669	-0.0961	0.127	-0.756		*					0.023
4	0.3604	0.5176	0.0878	-0.1572	0.114	-1.382		**					0.162
5	0.2258	0.3768	0.0526	-0.1510	0.134	-1.129		**					0.028
6	0.2659	0.5045	0.0555	-0.2387	0.133	-1.800		***					0.081
7	0.6102	0.6030	0.0532	0.007182	0.134	0.0538							0.000
8	0.4605	0.4787	0.0341	-0.0181	0.140	-0.130							0.000
9	0.3277	0.4904	0.0976	-0.1627	0.105	-1.542		***					0.291
10	0.6302	0.3863	0.0760	0.2439	0.122	1.999		***					0.222
11	0.5416	0.4369	0.0701	0.1046	0.125	0.834		*					0.031
12	0.5506	0.6502	0.0771	-0.0996	0.121	-0.821		*					0.039
13	0.2659	0.2944	0.0742	-0.0285	0.123	-0.232							0.003
14	0.2396	0.2465	0.1043	-0.006859	0.0989	-0.0693							0.001
15	0.6156	0.5452	0.0569	0.0704	0.132	0.533		*					0.008
16	0.3522	0.4357	0.0866	-0.0835	0.115	-0.728		*					0.043
17	0.5778	0.6436	0.0670	-0.0658	0.127	-0.517		*					0.011
18	0.5300	0.3773	0.1017	0.1528	0.102	1.504		***					0.323
19	0.3920	0.4273	0.0471	-0.0353	0.136	-0.260							0.001
20	0.6392	0.5602	0.0630	0.0791	0.129	0.612		*					0.013
21	0.4444	0.5304	0.0551	-0.0860	0.133	-0.648		*					0.010
22	0.4051	0.5519	0.0620	-0.1468	0.130	-1.132		**					0.042
23	0.6382	0.4639	0.0636	0.1743	0.129	1.352		**					0.064
24	0.8736	0.6571	0.0874	0.2165	0.114	1.897		***					0.302
25	0.3760	0.4571	0.0430	-0.0812	0.137	-0.592		*					0.005
26	0.3347	0.3590	0.0482	-0.0243	0.135	-0.179							0.001
27	0.7072	0.5717	0.0451	0.1355	0.136	0.993		*					0.015
28	0.6607	0.4917	0.0517	0.1690	0.134	1.260		**					0.034
29	0.3853	0.5015	0.0635	-0.1163	0.129	-0.902		*					0.028
30	0.6407	0.6231	0.0701	0.0176	0.125	0.140							0.001

## Output Statistics

Obs	RStudent	Hat	Diag H	Cov Ratio	DFFITS
1	0.5551	0.5097		2.5259	0.5660
2	1.2405	0.0665		0.9109	0.3311
3	-0.7485	0.2167		1.4613	-0.3936
4	-1.4110	0.3732		1.1876	-1.0889
5	-1.1357	0.1338		1.0576	-0.4464

6    -1.8997            0.1493            0.5541            -0.7958  
                          The SAS System            11:21 Friday, December 4, 2015    97

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hotIdx

## Output Statistics

Obs	RStudent	Hat H	Diag H	Cov Ratio	DFFITS
7	0.0526	0.1369		1.5802	0.0210
8	-0.1271	0.0562		1.4389	-0.0310
9	-1.5932	0.4615		1.1803	-1.4749
10	2.1509	0.2796		0.4982	1.3400
11	0.8283	0.2381		1.4450	0.4630
12	-0.8145	0.2874		1.5559	-0.5173
13	-0.2268	0.2662		1.8300	-0.1366
14	-0.0678	0.5262		2.8765	-0.0714
15	0.5249	0.1569		1.4840	0.2265
16	-0.7201	0.3634		1.8215	-0.5440
17	-0.5089	0.2170		1.6061	-0.2679
18	1.5487	0.5003		1.3240	1.5497
19	-0.2544	0.1076		1.4984	-0.0883
20	0.6035	0.1920		1.5059	0.2941
21	-0.6395	0.1470		1.4067	-0.2655
22	-1.1390	0.1859		1.1227	-0.5443
23	1.3785	0.1960		0.9506	0.6805
24	2.0206	0.3698		0.6578	1.5478
25	-0.5831	0.0893		1.3463	-0.1826
26	-0.1753	0.1125		1.5231	-0.0624
27	0.9922	0.0984		1.1144	0.3278
28	1.2772	0.1292		0.9501	0.4920
29	-0.8980	0.1954		1.3187	-0.4426
30	0.1369	0.2379		1.7806	0.0765

## Output Statistics

-----DFBETAS-----							
Obs	Intercept	artist HotIdx	artist FamIdx	duration	loudness	tempo	year
1	-0.0388	-0.0034	-0.0243	0.0267	0.4188	0.0942	0.0405
2	0.0615	-0.0748	-0.0049	-0.1943	0.1645	0.0301	-0.0561
3	0.2809	0.1375	-0.1892	0.0949	0.0735	-0.2372	-0.2789
4	-0.3647	-0.0781	0.1508	-0.0180	-0.8070	-0.1476	0.3635
5	0.0294	-0.2383	0.2426	-0.2424	0.2883	0.2366	-0.0311
6	-0.0913	0.4205	-0.2759	-0.0086	-0.0782	0.2872	0.0830
7	0.0028	-0.0008	0.0029	-0.0134	0.0087	-0.0014	-0.0027
8	0.0134	0.0025	-0.0077	-0.0015	0.0043	-0.0152	-0.0132
9	0.4592	-0.2883	-0.3671	-0.7283	0.5759	0.1539	-0.4299
10	-0.0152	1.0507	-0.7300	0.8199	-0.3659	-0.7508	0.0076
11	0.2043	0.2317	-0.2182	0.2368	-0.3094	-0.1412	-0.2072
12	-0.3155	-0.0665	-0.0248	0.0492	-0.0039	-0.1500	0.3233

The SAS System            11:21 Friday, December 4, 2015    98

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hotIdx

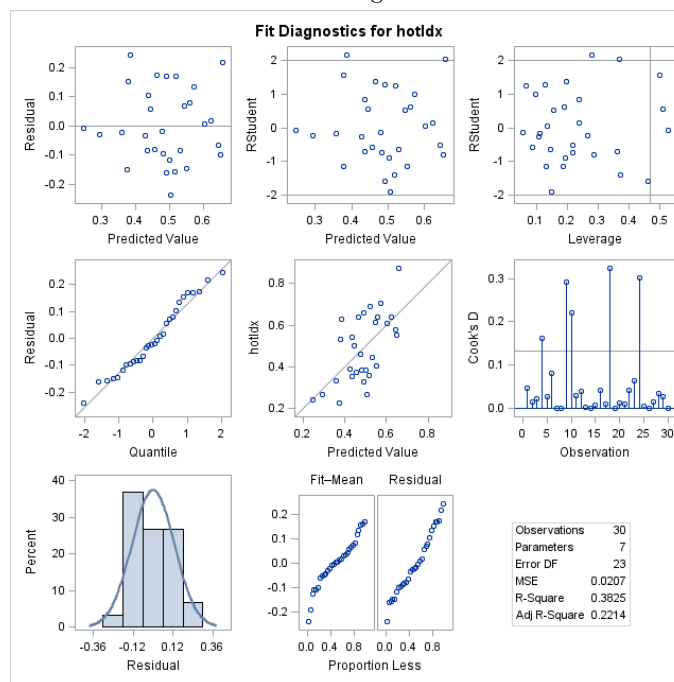
## Output Statistics

-----DFBETAS-----							
Obs	Intercept	artist HotIdx	artist FamIdx	duration	loudness	tempo	year
13	0.0135	-0.0437	0.0751	-0.0609	0.0311	0.0839	-0.0156
14	0.0132	-0.0461	0.0503	-0.0188	0.0145	-0.0228	-0.0132
15	-0.0364	-0.0220	0.0097	-0.1645	0.0712	0.0311	0.0392
16	-0.0539	0.0939	-0.0282	-0.1100	0.1515	-0.3321	0.0579
17	-0.0651	0.1289	-0.0954	0.2045	-0.0729	0.0657	0.0612
18	-0.7232	-0.6811	0.7126	0.6183	-0.5079	0.0254	0.7177
19	0.0623	-0.0212	0.0032	-0.0089	0.0223	-0.0203	-0.0620
20	-0.0772	-0.1559	0.1495	-0.1421	-0.0377	-0.0658	0.0802
21	-0.0190	0.1280	-0.0466	0.1708	-0.0391	-0.0662	0.0148
22	-0.4015	0.0450	0.0677	0.0330	0.0011	0.3103	0.3958
23	-0.2361	-0.3847	0.2277	-0.2211	0.1778	0.4503	0.2412
24	1.2452	-0.2369	0.2640	0.1500	-0.1798	-0.2727	-1.2576
25	-0.0753	-0.0118	0.0588	-0.0384	0.0039	0.0961	0.0728
26	0.0172	-0.0131	0.0308	-0.0129	0.0064	0.0067	-0.0182
27	0.0789	0.0367	0.0620	-0.0246	0.0393	0.0819	-0.0839

28	0.0112	0.0179	-0.1100	-0.2036	0.2181	-0.2211	-0.0005
29	0.0895	-0.1591	0.0742	0.0598	-0.0562	0.1979	-0.0916
30	-0.0161	-0.0271	0.0483	-0.0253	0.0041	0.0507	0.0149

Sum of Residuals	0
Sum of Squared Residuals	0.47519
Predicted Residual SS (PRESS)	0.91407

FIG. 3. SAS Reg Procedure



The SAS System

11:21 Friday, December 4, 2015 99

The REG Procedure

Model: MODEL1

Dependent Variable: logIndex

Number of Observations Read	30
Number of Observations Used	30

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1.49325	0.24887	2.66	0.0414
Error	23	2.15125	0.09353		
Corrected Total	29	3.64449			

Root MSE	0.30583	R-Square	0.4097
Dependent Mean	-0.77605	Adj R-Sq	0.2557
Coeff Var	-39.40881		

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS	Type II SS
Intercept	1	31.31697	19.79609	1.58	0.1273	18.06746	0.23408
artistHotIdx	1	-3.94591	1.76060	-2.24	0.0350	0.17340	0.46982
artistFamIdx	1	3.65513	1.35110	2.71	0.0126	0.38914	0.68453
duration	1	-0.00318	0.00114	-2.79	0.0103	0.45186	0.73028
loudness	1	0.01473	0.00915	1.61	0.1210	0.11989	0.24249
tempo	1	0.00125	0.00211	0.59	0.5602	0.10751	0.03268

	year	1	-0.01606	0.00979	-1.64	0.1147	0.25145	0.25145
--	------	---	----------	---------	-------	--------	---------	---------

## Parameter Estimates

Variable	DF	Tolerance	Variance Inflation
Intercept	1	.	0
artistHotIdx	1	0.18586	5.38030
artistFamIdx	1	0.20874	4.79069
duration	1	0.47537	2.10364
loudness	1	0.62601	1.59741
tempo	1	0.80076	1.24881
year	1	0.85355	1.17158

FIG. 4. SAS Reg Procedure

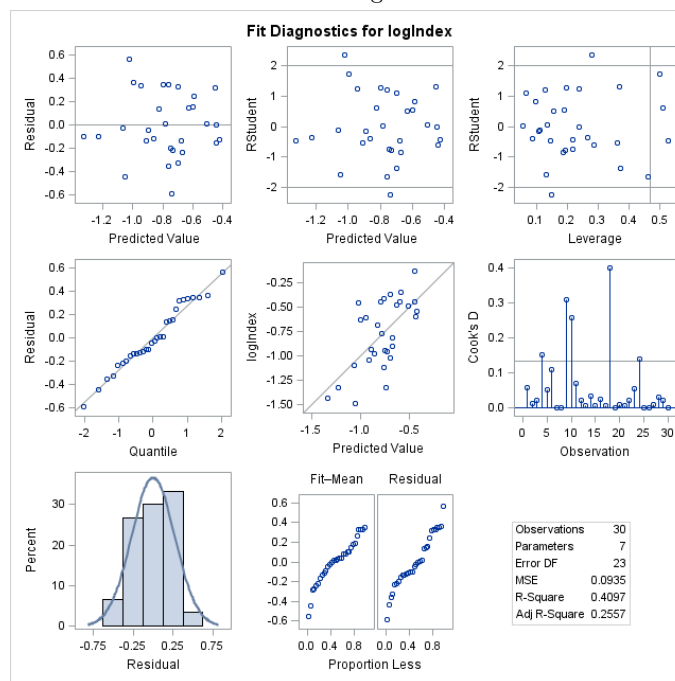
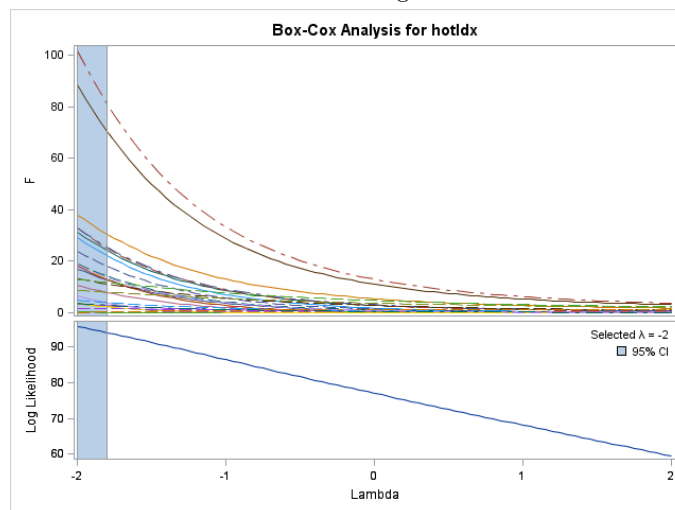


FIG. 5. SAS Transreg Procedure



The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hIndexRT

Number of Observations Read 30  
 Number of Observations Used 30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282.42099	47.07017	2.93	0.0286
Error	23	369.90536	16.08284		
Corrected Total	29	652.32635			

Root MSE 4.01034 R-Square 0.4329  
 Dependent Mean 6.07712 Adj R-Sq 0.2850  
 Coeff Var 65.99088

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS	Type II SS
Intercept	1	-307.66834	259.58501	-1.19	0.2480	1107.93985	22.59275
artistHotIdx	1	61.03969	23.08670	2.64	0.0145	28.84248	112.42523
artistFamIdx	1	-55.99957	17.71686	-3.16	0.0044	94.78809	160.67893
duration	1	0.04383	0.01492	2.94	0.0074	70.42604	138.72475
loudness	1	-0.23531	0.11999	-1.96	0.0621	36.99989	61.85042
tempo	1	-0.02541	0.02765	-0.92	0.3677	27.07877	13.57824
year	1	0.15779	0.12840	1.23	0.2316	24.28572	24.28572

Parameter Estimates

Variable	DF	Tolerance	Variance Inflation
Intercept	1	.	0
artistHotIdx	1	0.18586	5.38030
artistFamIdx	1	0.20874	4.79069
duration	1	0.47537	2.10364
loudness	1	0.62601	1.59741
tempo	1	0.80076	1.24881
year	1	0.85355	1.17158

FIG. 6. SAS Reg Procedure

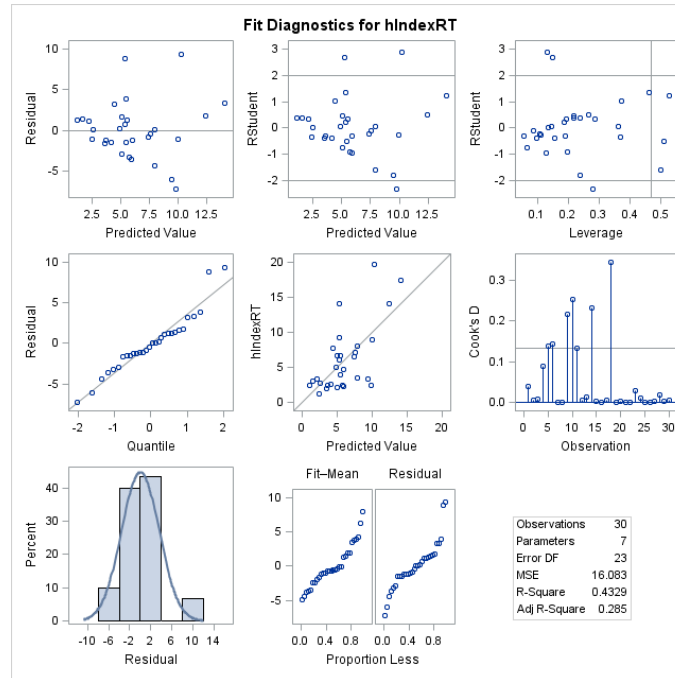
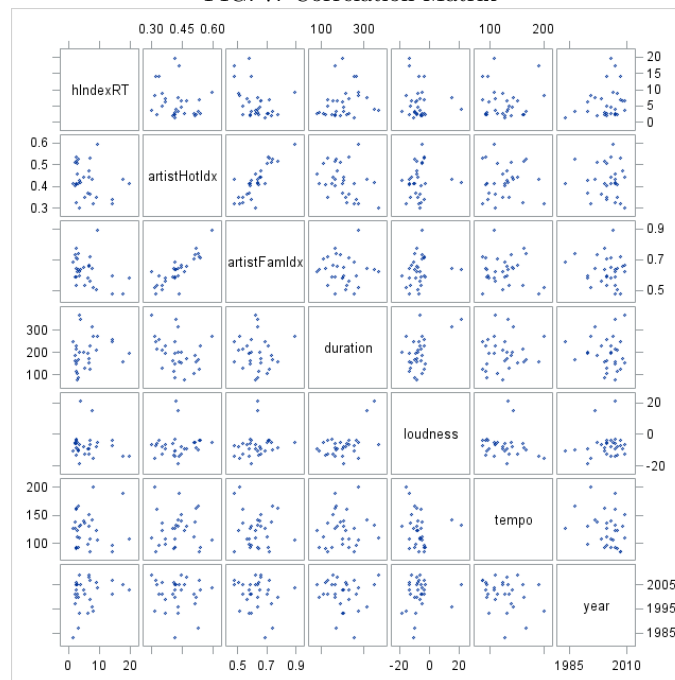


FIG. 7. Correlation Matrix



The SAS System 11:21 Friday, December 4, 2015 103

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

Forward Selection: Step 1

Variable aIndex Entered: R-Square = 0.1528 and C(p) = 8.3621

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	99.68574	99.68574	5.05	0.0327
Error	28	552.64062	19.73716		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	13.58603	3.43825	308.17345	15.61	0.0005
aIndex	-37.05174	16.48673	99.68574	5.05	0.0327

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable duration Entered: R-Square = 0.2940 and C(p) = 4.6362

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	191.77541	95.88770	5.62	0.0091
Error	27	460.55095	17.05744		
Corrected Total	29	652.32635			

The SAS System 11:21 Friday, December 4, 2015 104

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Forward Selection: Step 2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.33712	3.33966	196.56871	11.52	0.0021
aIndex	-52.34084	16.67951	167.96885	9.85	0.0041
duration	0.02680	0.01153	92.08967	5.40	0.0279

Bounds on condition number: 1.1843, 4.7373

No other variable met the 0.1000 significance level for entry into the model.

#### Summary of Forward Selection

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	aIndex	1	0.1528	0.1528	8.3621	5.05	0.0327
2	duration	2	0.1412	0.2940	4.6362	5.40	0.0279

The SAS System 11:21 Friday, December 4, 2015 105

The REG Procedure

Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

#### Analysis of Variance

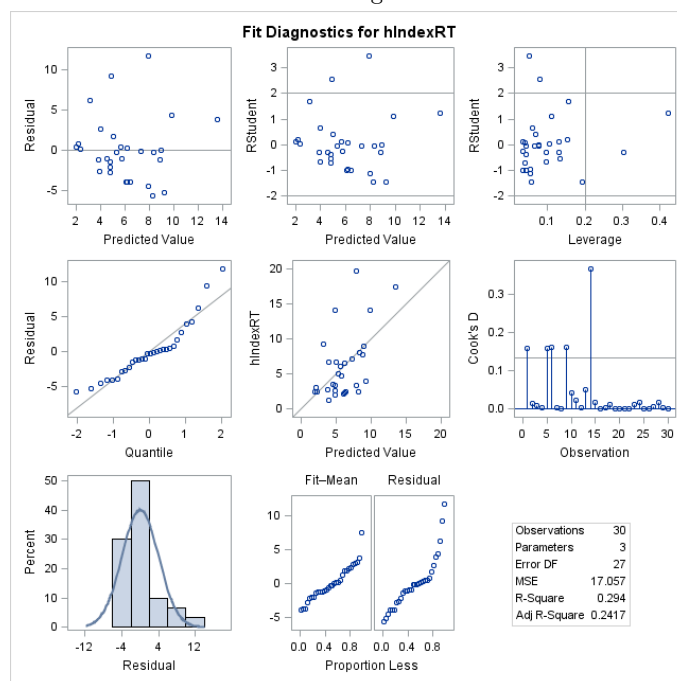
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	191.77541	95.88770	5.62	0.0091
Error	27	460.55095	17.05744		
Corrected Total	29	652.32635			

Root MSE 4.13007 R-Square 0.2940  
Dependent Mean 6.07712 Adj R-Sq 0.2417  
Coeff Var 67.96096

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	11.33712	3.33966	3.39	0.0021	.	0
aIndex	1	-52.34084	16.67951	-3.14	0.0041	0.84437	1.18432
duration	1	0.02680	0.01153	2.32	0.0279	0.84437	1.18432

FIG. 8. SAS Reg Procedure



The SAS System

11:21 Friday, December 4, 2015 106

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.4329 and C(p) = 7.0000  
The model is not of full rank. A subset of the model which is of full rank is chosen.



## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282.42099	47.07017	2.93	0.0286
Error	23	369.90536	16.08284		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-307.66834	259.58501	22.59275	1.40	0.2480
aIndex	-55.99957	17.71686	160.67893	9.99	0.0044
artistHotIdx	5.04011	11.47008	3.10535	0.19	0.6645
duration	0.04383	0.01492	138.72475	8.63	0.0074
loudness	-0.23531	0.11999	61.85042	3.85	0.0621
tempo	-0.02541	0.02765	13.57824	0.84	0.3677
year	0.15779	0.12840	24.28572	1.51	0.2316

Bounds on condition number: 2.1036, 53.2

Backward Elimination: Step 1

Variable artistHotIdx Removed: R-Square = 0.4282 and C(p) = 5.1931

The SAS System 11:21 Friday, December 4, 2015 107

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Backward Elimination: Step 1

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	279.31564	55.86313	3.59	0.0144
Error	24	373.01071	15.54211		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-292.82758	253.01479	20.81815	1.34	0.2585
aIndex	-54.32265	17.00763	158.55667	10.20	0.0039
duration	0.04067	0.01285	155.68248	10.02	0.0042
loudness	-0.21670	0.11036	59.91985	3.86	0.0613
tempo	-0.02419	0.02705	12.42945	0.80	0.3801
year	0.15158	0.12546	22.68623	1.46	0.2388

Bounds on condition number: 1.6135, 33.784

Backward Elimination: Step 2

Variable artistFamIdx Entered: R-Square = 0.4329 and C(p) = 7.0000

NOTE: The variable which previously had small tolerance is now allowed to enter after removal of some variables from the model.

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282.42099	47.07017	2.93	0.0286
Error	23	369.90536	16.08284		
Corrected Total	29	652.32635			

The SAS System 11:21 Friday, December 4, 2015 108

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Backward Elimination: Step 2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-307.66834	259.58501	22.59275	1.40	0.2480
aIndex	-61.03969	23.08670	112.42523	6.99	0.0145
artistFamIdx	5.04011	11.47008	3.10535	0.19	0.6645
duration	0.04383	0.01492	138.72475	8.63	0.0074
loudness	-0.23531	0.11999	61.85042	3.85	0.0621
tempo	-0.02541	0.02765	13.57824	0.84	0.3677
year	0.15779	0.12840	24.28572	1.51	0.2316

Bounds on condition number: 2.4064, 63.215

Backward Elimination: Step 3

Variable artistFamIdx Removed: R-Square = 0.4282 and C(p) = 5.1931

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	279.31564	55.86313	3.59	0.0144
Error	24	373.01071	15.54211		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-292.82758	253.01479	20.81815	1.34	0.2585
aIndex	-54.32265	17.00763	158.55667	10.20	0.0039
duration	0.04067	0.01285	155.68248	10.02	0.0042
loudness	-0.21670	0.11036	59.91985	3.86	0.0613
tempo	-0.02419	0.02705	12.42945	0.80	0.3801
year	0.15158	0.12546	22.68623	1.46	0.2388

Bounds on condition number: 1.6135, 33.784

Backward Elimination: Step 4

The SAS System 11:21 Friday, December 4, 2015 109

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Backward Elimination: Step 4

Variable tempo Removed: R-Square = 0.4091 and C(p) = 3.9659

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-359.86178	240.68605	34.46566	2.24	0.1474
aIndex	-49.33357	16.00210	146.53726	9.50	0.0049
duration	0.03794	0.01243	143.58644	9.31	0.0053
loudness	-0.20173	0.10865	53.14988	3.45	0.0752
year	0.18338	0.11983	36.10455	2.34	0.1385

Bounds on condition number: 1.5224, 20.636

All variables left in the model are significant at the 0.1500 level.

#### Summary of Backward Elimination

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1		artistHotIdx	5	0.0048	0.4282	5.1931	0.19	0.6645
2	artistFamIdx		6	0.0048	0.4329	7.0000	0.19	0.6645
3		artistFamIdx	5	0.0048	0.4282	5.1931	0.19	0.6645
4		tempo	4	0.0191	0.4091	3.9659	0.80	0.3801
The SAS System					11:21	Friday, December 4, 2015 110		

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

#### Analysis of Variance

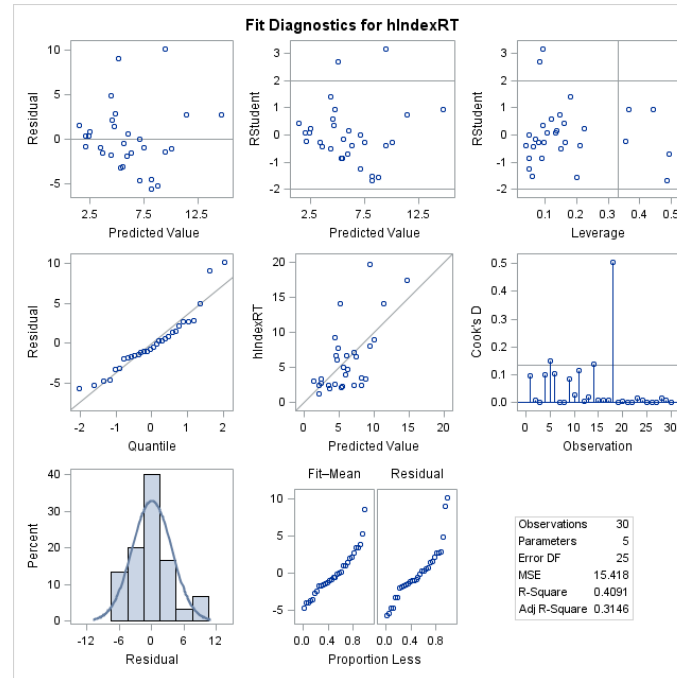
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Root MSE 3.92653 R-Square 0.4091  
Dependent Mean 6.07712 Adj R-Sq 0.3146  
Coeff Var 64.61168

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	-359.86178	240.68605	-1.50	0.1474	.	0
aIndex	1	-49.33357	16.00210	-3.08	0.0049	0.82918	1.20602
duration	1	0.03794	0.01243	3.05	0.0053	0.65684	1.52244
loudness	1	-0.20173	0.10865	-1.86	0.0752	0.73197	1.36617
year	1	0.18338	0.11983	1.53	0.1385	0.93950	1.06439

FIG. 9. SAS Reg Procedure



The SAS System

11:21 Friday, December 4, 2015 111

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Number of Observations Read 30  
 Number of Observations Used 30

Stepwise Selection: Step 1

Variable aIndex Entered: R-Square = 0.1528 and C(p) = 8.3621

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	99.68574	99.68574	5.05	0.0327
Error	28	552.64062	19.73716		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	13.58603	3.43825	308.17345	15.61	0.0005
aIndex	-37.05174	16.48673	99.68574	5.05	0.0327

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable duration Entered: R-Square = 0.2940 and C(p) = 4.6362

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	191.77541	95.88770	5.62	0.0091

Error	27	460.55095	17.05744
Corrected Total	29	652.32635	

The SAS System 11:21 Friday, December 4, 2015 112

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Stepwise Selection: Step 2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.33712	3.33966	196.56871	11.52	0.0021
aIndex	-52.34084	16.67951	167.96885	9.85	0.0041
duration	0.02680	0.01153	92.08967	5.40	0.0279

Bounds on condition number: 1.1843, 4.7373

Stepwise Selection: Step 3

Variable loudness Entered: R-Square = 0.3538 and C(p) = 4.2108

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	230.78165	76.92722	4.74	0.0091
Error	26	421.54471	16.21326		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	8.41379	3.76212	81.09422	5.00	0.0341
aIndex	-52.60802	16.26244	169.66905	10.46	0.0033
duration	0.03586	0.01267	129.85152	8.01	0.0089
loudness	-0.16955	0.10931	39.00624	2.41	0.1330

Bounds on condition number: 1.5043, 12.011

Stepwise Selection: Step 4

Variable year Entered: R-Square = 0.4091 and C(p) = 3.9659

The SAS System 11:21 Friday, December 4, 2015 113

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Stepwise Selection: Step 4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-359.86178	240.68605	34.46566	2.24	0.1474
aIndex	-49.33357	16.00210	146.53726	9.50	0.0049
duration	0.03794	0.01243	143.58644	9.31	0.0053
loudness	-0.20173	0.10865	53.14988	3.45	0.0752
year	0.18338	0.11983	36.10455	2.34	0.1385

Bounds on condition number: 1.5224, 20.636

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

#### Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	aIndex		1	0.1528	0.1528	8.3621	5.05	0.0327
2	duration		2	0.1412	0.2940	4.6362	5.40	0.0279
3	loudness		3	0.0598	0.3538	4.2108	2.41	0.1330
4	year		4	0.0553	0.4091	3.9659	2.34	0.1385
The SAS System					11:21 Friday, December 4, 2015 114			

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

#### Analysis of Variance

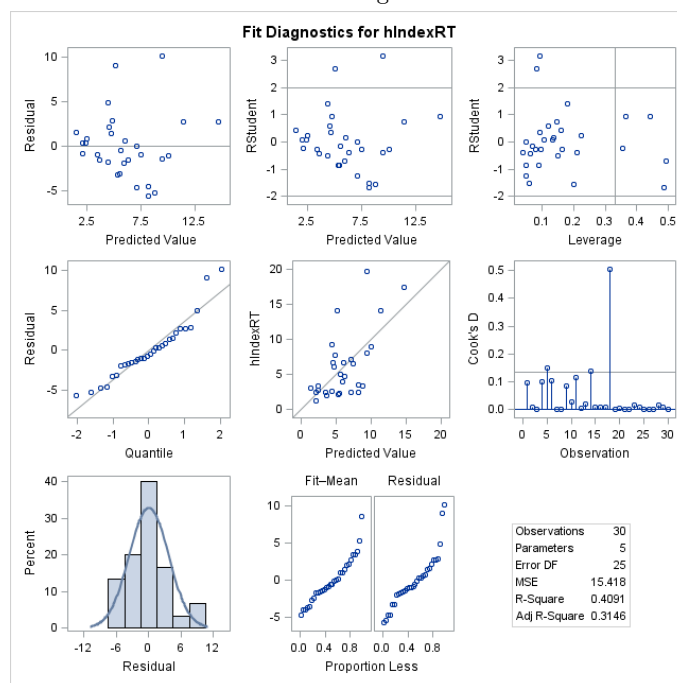
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Root MSE 3.92653 R-Square 0.4091  
Dependent Mean 6.07712 Adj R-Sq 0.3146  
Coeff Var 64.61168

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	-359.86178	240.68605	-1.50	0.1474	.	0
aIndex	1	-49.33357	16.00210	-3.08	0.0049	0.82918	1.20602
duration	1	0.03794	0.01243	3.05	0.0053	0.65684	1.52244
loudness	1	-0.20173	0.10865	-1.86	0.0752	0.73197	1.36617
year	1	0.18338	0.11983	1.53	0.1385	0.93950	1.06439

FIG. 10. SAS Reg Procedure



The SAS System

11:21 Friday, December 4, 2015 115

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Number of Observations Read 30  
 Number of Observations Used 30

Stepwise Selection: Step 1

Variable artistFamIdx Entered: R-Square = 0.1472 and C(p) = 8.5912

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	96.00114	96.00114	4.83	0.0364
Error	28	556.32521	19.86876		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	18.49447	5.70738	208.63207	10.50	0.0031
artistFamIdx	-19.77628	8.99688	96.00114	4.83	0.0364

Bounds on condition number: 1, 1

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

## Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
------	------------------	------------------	----------------	------------------	----------------	------	---------	--------

```

1  artistFamIdx              1      0.1472      0.1472      8.5912      4.83      0.0364
                                The SAS System      11:21 Friday, December 4, 2015 116

```

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

## Analysis of Variance

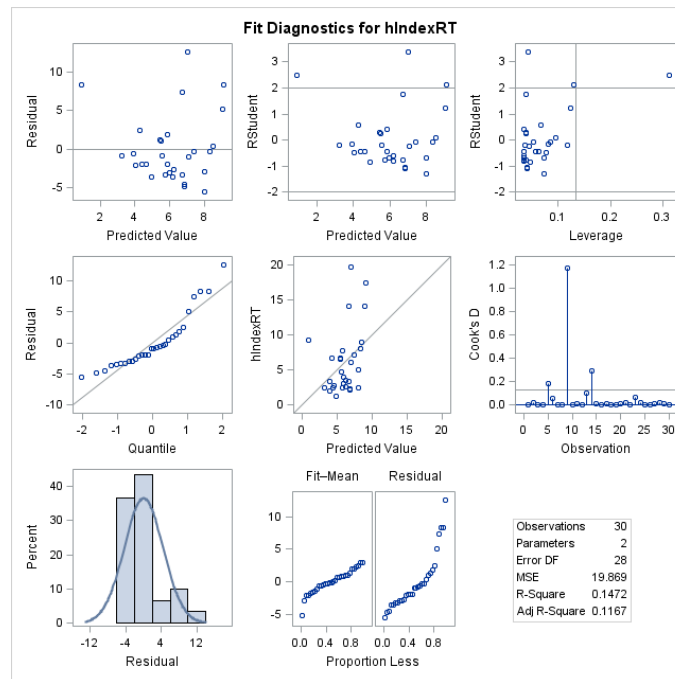
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	96.00114	96.00114	4.83	0.0364
Error	28	556.32521	19.86876		
Corrected Total	29	652.32635			

Root MSE	4.45744	R-Square	0.1472
Dependent Mean	6.07712	Adj R-Sq	0.1167
Coeff Var	73.34794		

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	18.49447	5.70738	3.24	0.0031	.	0
artistFamIdx	1	-19.77628	8.99688	-2.20	0.0364	1.00000	1.00000

FIG. 11. SAS Reg Procedure



The SAS System 11:21 Friday, December 4, 2015 117

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

Maximum R-Square Improvement: Step 1



Variable aIndex Entered: R-Square = 0.1528 and C(p) = 8.3621

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	99.68574	99.68574	5.05	0.0327
Error	28	552.64062	19.73716		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	13.58603	3.43825	308.17345	15.61	0.0005
aIndex	-37.05174	16.48673	99.68574	5.05	0.0327

Bounds on condition number: 1, 1

The above model is the best 1-variable model found.

Maximum R-Square Improvement: Step 2

Variable duration Entered: R-Square = 0.2940 and C(p) = 4.6362

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	191.77541	95.88770	5.62	0.0091
Error	27	460.55095	17.05744		
Corrected Total	29	652.32635			

The SAS System 11:21 Friday, December 4, 2015 118

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Maximum R-Square Improvement: Step 2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.33712	3.33966	196.56871	11.52	0.0021
aIndex	-52.34084	16.67951	167.96885	9.85	0.0041
duration	0.02680	0.01153	92.08967	5.40	0.0279

Bounds on condition number: 1.1843, 4.7373

The above model is the best 2-variable model found.

Maximum R-Square Improvement: Step 3

Variable loudness Entered: R-Square = 0.3538 and C(p) = 4.2108

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	230.78165	76.92722	4.74	0.0091
Error	26	421.54471	16.21326		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	8.41379	3.76212	81.09422	5.00	0.0341
aIndex	-52.60802	16.26244	169.66905	10.46	0.0033
duration	0.03586	0.01267	129.85152	8.01	0.0089

loudness            -0.16955          0.10931          39.00624          2.41   0.1330

Bounds on condition number: 1.5043, 12.011

The above model is the best 3-variable model found.

Maximum R-Square Improvement: Step 4

The SAS System            11:21 Friday, December 4, 2015 119

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Maximum R-Square Improvement: Step 4

Variable year Entered: R-Square = 0.4091 and C(p) = 3.9659

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-359.86178	240.68605	34.46566	2.24	0.1474
aIndex	-49.33357	16.00210	146.53726	9.50	0.0049
duration	0.03794	0.01243	143.58644	9.31	0.0053
loudness	-0.20173	0.10865	53.14988	3.45	0.0752
year	0.18338	0.11983	36.10455	2.34	0.1385

Bounds on condition number: 1.5224, 20.636

The above model is the best 4-variable model found.

Maximum R-Square Improvement: Step 5

Variable tempo Entered: R-Square = 0.4282 and C(p) = 5.1931

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	279.31564	55.86313	3.59	0.0144
Error	24	373.01071	15.54211		
Corrected Total	29	652.32635			

The SAS System            11:21 Friday, December 4, 2015 120

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Maximum R-Square Improvement: Step 5

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-292.82758	253.01479	20.81815	1.34	0.2585
aIndex	-54.32265	17.00763	158.55667	10.20	0.0039
duration	0.04067	0.01285	155.68248	10.02	0.0042
loudness	-0.21670	0.11036	59.91985	3.86	0.0613
tempo	-0.02419	0.02705	12.42945	0.80	0.3801
year	0.15158	0.12546	22.68623	1.46	0.2388

Bounds on condition number: 1.6135, 33.784

The above model is the best 5-variable model found.

Maximum R-Square Improvement: Step 6

Variable artistFamIdx Entered: R-Square = 0.4329 and C(p) = 7.0000

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282.42099	47.07017	2.93	0.0286
Error	23	369.90536	16.08284		
Corrected Total	29	652.32635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-307.66834	259.58501	22.59275	1.40	0.2480
aIndex	-61.03969	23.08670	112.42523	6.99	0.0145
artistFamIdx	5.04011	11.47008	3.10535	0.19	0.6645
duration	0.04383	0.01492	138.72475	8.63	0.0074
loudness	-0.23531	0.11999	61.85042	3.85	0.0621
tempo	-0.02541	0.02765	13.57824	0.84	0.3677
year	0.15779	0.12840	24.28572	1.51	0.2316

Bounds on condition number: 2.4064, 63.215

The SAS System 11:21 Friday, December 4, 2015 121

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Maximum R-Square Improvement: Step 6

The above model is the best 6-variable model found.

No further improvement in R-Square is possible.

The SAS System 11:21 Friday, December 4, 2015 122

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282.42099	47.07017	2.93	0.0286
Error	23	369.90536	16.08284		
Corrected Total	29	652.32635			

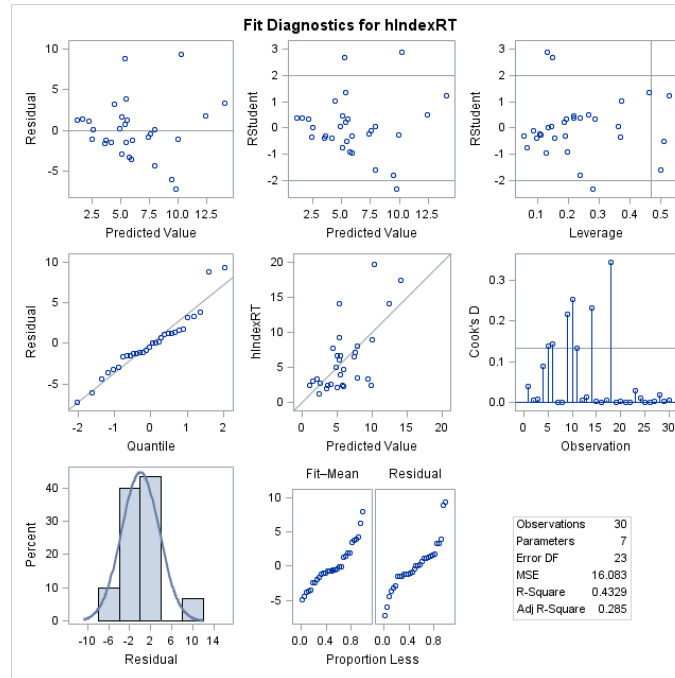
Root MSE 4.01034 R-Square 0.4329  
Dependent Mean 6.07712 Adj R-Sq 0.2850  
Coeff Var 65.99088

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	-307.66834	259.58501	-1.19	0.2480	.	0
aIndex	1	-61.03969	23.08670	-2.64	0.0145	0.41555	2.40645
artistFamIdx	1	5.04011	11.47008	0.44	0.6645	0.49802	2.00797
duration	1	0.04383	0.01492	2.94	0.0074	0.47537	2.10364
loudness	1	-0.23531	0.11999	-1.96	0.0621	0.62601	1.59741
tempo	1	-0.02541	0.02765	-0.92	0.3677	0.80076	1.24881
year	1	0.15779	0.12840	1.23	0.2316	0.85355	1.17158

The SAS System 11:21 Friday, December 4, 2015 123

FIG. 12. SAS Reg Procedure



The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Adjusted R-Square Selection Method

Number of Observations Read 30  
Number of Observations Used 30

Number in Model	Adjusted R-Square	R-Square	Variables in Model
4	0.3146	0.4091	aIndex duration loudness year
5	0.3091	0.4282	aIndex duration loudness tempo year
4	0.2964	0.3934	aIndex duration loudness tempo
5	0.2897	0.4121	aIndex artistFamIdx duration loudness year
5	0.2897	0.4121	aIndex artistHotIdx duration loudness year
5	0.2897	0.4121	artistHotIdx artistFamIdx duration loudness year
6	0.2850	0.4329	aIndex artistFamIdx duration loudness tempo year
6	0.2850	0.4329	aIndex artistHotIdx duration loudness tempo year
6	0.2850	0.4329	artistHotIdx artistFamIdx duration loudness tempo year
3	0.2792	0.3538	aIndex duration loudness
5	0.2698	0.3957	aIndex artistHotIdx duration loudness tempo
5	0.2698	0.3957	aIndex artistFamIdx duration loudness tempo
5	0.2698	0.3957	artistHotIdx artistFamIdx duration loudness tempo
4	0.2509	0.3542	aIndex artistHotIdx duration loudness
4	0.2509	0.3542	aIndex artistFamIdx duration loudness
4	0.2509	0.3542	artistHotIdx artistFamIdx duration loudness
3	0.2501	0.3277	aIndex duration year
2	0.2417	0.2940	aIndex duration
3	0.2362	0.3152	aIndex duration tempo
4	0.2301	0.3363	aIndex duration tempo year
4	0.2226	0.3298	aIndex artistHotIdx duration year
4	0.2226	0.3298	aIndex artistFamIdx duration year
4	0.2226	0.3298	artistHotIdx artistFamIdx duration year
3	0.2164	0.2975	aIndex artistFamIdx duration
3	0.2164	0.2975	aIndex artistHotIdx duration
3	0.2164	0.2975	artistHotIdx artistFamIdx duration
4	0.2086	0.3177	aIndex artistFamIdx duration tempo
4	0.2086	0.3177	aIndex artistHotIdx duration tempo
4	0.2086	0.3177	artistHotIdx artistFamIdx duration tempo
5	0.2002	0.3381	aIndex artistHotIdx duration tempo year
5	0.2002	0.3381	aIndex artistFamIdx duration tempo year
5	0.2002	0.3381	artistHotIdx artistFamIdx duration tempo year

4	0.1420	0.2603	artistFamIdx duration loudness year
2	0.1295	0.1895	aIndex artistFamIdx
2	0.1295	0.1895	aIndex artistHotIdx
2	0.1295	0.1895	artistHotIdx artistFamIdx
3	0.1271	0.2174	artistFamIdx duration year
3	0.1263	0.2167	aIndex artistHotIdx year
3	0.1263	0.2167	aIndex artistFamIdx year

The SAS System 11:21 Friday, December 4, 2015 124

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

## Adjusted R-Square Selection Method

Number in Model	Adjusted R-Square	R-Square	Variables in Model
3	0.1263	0.2167	artistHotIdx artistFamIdx year
2	0.1232	0.1837	aIndex year
1	0.1226	0.1528	aIndex
2	0.1199	0.1806	artistFamIdx year
2	0.1170	0.1779	artistFamIdx duration
1	0.1167	0.1472	artistFamIdx
3	0.1097	0.2018	artistFamIdx duration loudness
5	0.1066	0.2606	artistFamIdx duration loudness tempo year
3	0.1023	0.1952	aIndex artistHotIdx tempo
3	0.1023	0.1952	aIndex artistFamIdx tempo
3	0.1023	0.1952	artistHotIdx artistFamIdx tempo
3	0.0967	0.1901	aIndex artistFamIdx loudness
3	0.0967	0.1901	aIndex artistHotIdx loudness
3	0.0967	0.1901	artistHotIdx artistFamIdx loudness
2	0.0958	0.1581	aIndex tempo
3	0.0954	0.1890	aIndex loudness year
4	0.0947	0.2195	aIndex artistHotIdx loudness year
4	0.0947	0.2195	aIndex artistFamIdx loudness year
4	0.0947	0.2195	artistHotIdx artistFamIdx loudness year
3	0.0923	0.1862	artistFamIdx loudness year
4	0.0923	0.2175	artistFamIdx duration tempo year
2	0.0921	0.1547	aIndex loudness
4	0.0921	0.2173	aIndex artistHotIdx tempo year
4	0.0921	0.2173	aIndex artistFamIdx tempo year
4	0.0921	0.2173	artistHotIdx artistFamIdx tempo year
3	0.0904	0.1845	duration loudness year
3	0.0899	0.1840	aIndex tempo year
3	0.0864	0.1809	artistFamIdx tempo year
2	0.0861	0.1491	artistFamIdx loudness
3	0.0860	0.1805	artistFamIdx duration tempo
2	0.0855	0.1486	artistFamIdx tempo
4	0.0811	0.2078	artistFamIdx duration loudness tempo
4	0.0676	0.1962	aIndex artistFamIdx loudness tempo
4	0.0676	0.1962	aIndex artistHotIdx loudness tempo
4	0.0676	0.1962	artistHotIdx artistFamIdx loudness tempo
3	0.0638	0.1607	aIndex loudness tempo
4	0.0599	0.1895	aIndex loudness tempo year
5	0.0578	0.2203	aIndex artistHotIdx loudness tempo year
5	0.0578	0.2203	aIndex artistFamIdx loudness tempo year
5	0.0578	0.2203	artistHotIdx artistFamIdx loudness tempo year
4	0.0561	0.1863	artistFamIdx loudness tempo year
4	0.0557	0.1859	artistHotIdx duration loudness year
4	0.0547	0.1851	duration loudness tempo year
3	0.0530	0.1510	artistFamIdx loudness tempo

The SAS System 11:21 Friday, December 4, 2015 125

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

## Adjusted R-Square Selection Method

Number in Model	Adjusted R-Square	R-Square	Variables in Model
2	0.0320	0.0987	duration year
2	0.0265	0.0937	duration loudness
2	0.0255	0.0927	artistHotIdx year
1	0.0208	0.0545	year
5	0.0172	0.1866	artistHotIdx duration loudness tempo year
3	0.0151	0.1170	artistHotIdx duration year
1	0.0101	0.0442	artistHotIdx
2	0.0075	0.0760	loudness year
3	0.0054	0.1083	artistHotIdx loudness year
1	0.0021	0.0365	duration

3	-.0001	0.1034	duration tempo year
3	-.0023	0.1014	artistHotIdx duration loudness
3	-.0060	0.0981	artistHotIdx tempo year
2	-.0080	0.0615	artistHotIdx duration
2	-.0085	0.0611	tempo year
3	-.0086	0.0957	duration loudness tempo
2	-.0172	0.0530	artistHotIdx loudness
4	-.0194	0.1212	artistHotIdx duration tempo year
1	-.0220	0.0132	loudness
3	-.0256	0.0805	loudness tempo year
2	-.0264	0.0444	artistHotIdx tempo
4	-.0299	0.1121	artistHotIdx loudness tempo year
2	-.0349	0.0365	duration tempo
1	-.0354	0.0003	tempo
4	-.0406	0.1030	artistHotIdx duration loudness tempo
3	-.0467	0.0615	artistHotIdx duration tempo
3	-.0563	0.0530	artistHotIdx loudness tempo
2	-.0598	0.0132	loudness tempo

NOTE: Models of not full rank are not included.

The SAS System 11:21 Friday, December 4, 2015 126

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read	30
Number of Observations Used	30

#### Analysis of Variance

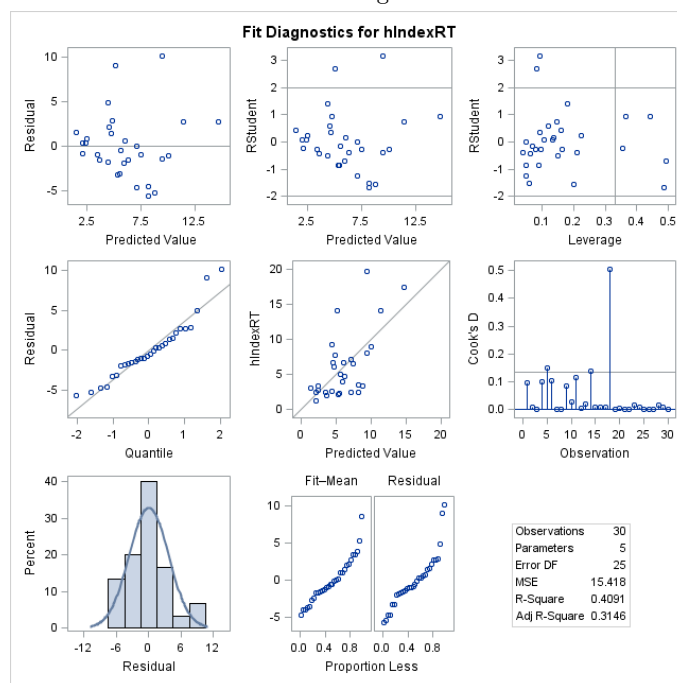
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Root MSE	3.92653	R-Square	0.4091
Dependent Mean	6.07712	Adj R-Sq	0.3146
Coeff Var	64.61168		

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	-359.86178	240.68605	-1.50	0.1474	.	0
aIndex	1	-49.33357	16.00210	-3.08	0.0049	0.82918	1.20602
duration	1	0.03794	0.01243	3.05	0.0053	0.65684	1.52244
loudness	1	-0.20173	0.10865	-1.86	0.0752	0.73197	1.36617
year	1	0.18338	0.11983	1.53	0.1385	0.93950	1.06439

FIG. 13. SAS Reg Procedure



The SAS System

11:21 Friday, December 4, 2015 127

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

C(p) Selection Method

Number of Observations Read	30
Number of Observations Used	30

Number in Model	C(p)	R-Square	Variables in Model
4	3.9659	0.4091	aIndex duration loudness year
3	4.2108	0.3538	aIndex duration loudness
4	4.6037	0.3934	aIndex duration loudness tempo
2	4.6362	0.2940	aIndex duration
5	5.1931	0.4282	aIndex duration loudness tempo year
3	5.2707	0.3277	aIndex duration year
3	5.7757	0.3152	aIndex duration tempo
5	5.8443	0.4121	aIndex artistFamIdx duration loudness year
5	5.8443	0.4121	aIndex artistHotIdx duration loudness year
5	5.8443	0.4121	artistHotIdx artistFamIdx duration loudness year
4	6.1937	0.3542	aIndex artistHotIdx duration loudness
4	6.1937	0.3542	aIndex artistFamIdx duration loudness
4	6.1937	0.3542	artistHotIdx artistFamIdx duration loudness
3	6.4943	0.2975	aIndex artistFamIdx duration
3	6.4943	0.2975	aIndex artistHotIdx duration
3	6.4943	0.2975	artistHotIdx artistFamIdx duration
5	6.5100	0.3957	aIndex artistHotIdx duration loudness tempo
5	6.5100	0.3957	aIndex artistFamIdx duration loudness tempo
5	6.5100	0.3957	artistHotIdx artistFamIdx duration loudness tempo
4	6.9188	0.3363	aIndex duration tempo year
6	7.0000	0.4329	aIndex artistFamIdx duration loudness tempo year
6	7.0000	0.4329	aIndex artistHotIdx duration loudness tempo year
6	7.0000	0.4329	artistHotIdx artistFamIdx duration loudness tempo year
4	7.1839	0.3298	aIndex artistHotIdx duration year
4	7.1839	0.3298	aIndex artistFamIdx duration year
4	7.1839	0.3298	artistHotIdx artistFamIdx duration year
4	7.6731	0.3177	aIndex artistFamIdx duration tempo
4	7.6731	0.3177	aIndex artistHotIdx duration tempo
4	7.6731	0.3177	artistHotIdx artistFamIdx duration tempo
1	8.3621	0.1528	aIndex

1	8.5912	0.1472	artistFamIdx
5	8.8457	0.3381	aIndex artistHotIdx duration tempo year
5	8.8457	0.3381	aIndex artistFamIdx duration tempo year
5	8.8457	0.3381	artistHotIdx artistFamIdx duration tempo year
2	8.8733	0.1895	aIndex artistFamIdx
2	8.8733	0.1895	aIndex artistHotIdx
2	8.8733	0.1895	artistHotIdx artistFamIdx
2	9.1107	0.1837	aIndex year
2	9.2340	0.1806	artistFamIdx year

The SAS System 11:21 Friday, December 4, 2015 128

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

C(p) Selection Method

Number in Model	C(p)	R-Square	Variables in Model
2	9.3430	0.1779	artistFamIdx duration
3	9.7414	0.2174	artistFamIdx duration year
3	9.7725	0.2167	aIndex artistHotIdx year
3	9.7725	0.2167	aIndex artistFamIdx year
3	9.7725	0.2167	artistHotIdx artistFamIdx year
4	10.0006	0.2603	artistFamIdx duration loudness year
2	10.1462	0.1581	aIndex tempo
2	10.2848	0.1547	aIndex loudness
3	10.3736	0.2018	artistFamIdx duration loudness
2	10.5108	0.1491	artistFamIdx loudness
2	10.5338	0.1486	artistFamIdx tempo
3	10.6439	0.1952	aIndex artistHotIdx tempo
3	10.6439	0.1952	aIndex artistFamIdx tempo
3	10.6439	0.1952	artistHotIdx artistFamIdx tempo
3	10.8479	0.1901	aIndex artistFamIdx loudness
3	10.8479	0.1901	aIndex artistHotIdx loudness
3	10.8479	0.1901	artistHotIdx artistFamIdx loudness
3	10.8939	0.1890	aIndex loudness year
3	11.0088	0.1862	artistFamIdx loudness year
3	11.0773	0.1845	duration loudness year
3	11.0953	0.1840	aIndex tempo year
3	11.2238	0.1809	artistFamIdx tempo year
3	11.2385	0.1805	artistFamIdx duration tempo
4	11.6559	0.2195	aIndex artistHotIdx loudness year
4	11.6559	0.2195	aIndex artistFamIdx loudness year
4	11.6559	0.2195	artistHotIdx artistFamIdx loudness year
4	11.7400	0.2175	artistFamIdx duration tempo year
4	11.7467	0.2173	aIndex artistHotIdx tempo year
4	11.7467	0.2173	aIndex artistFamIdx tempo year
4	11.7467	0.2173	artistHotIdx artistFamIdx tempo year
5	11.9904	0.2606	artistFamIdx duration loudness tempo year
3	12.0431	0.1607	aIndex loudness tempo
4	12.1315	0.2078	artistFamIdx duration loudness tempo
1	12.3487	0.0545	year
3	12.4376	0.1510	artistFamIdx loudness tempo
2	12.5562	0.0987	duration year
4	12.6029	0.1962	aIndex artistFamIdx loudness tempo
4	12.6029	0.1962	aIndex artistHotIdx loudness tempo
4	12.6029	0.1962	artistHotIdx artistFamIdx loudness tempo
2	12.7605	0.0937	duration loudness
1	12.7670	0.0442	artistHotIdx
2	12.7985	0.0927	artistHotIdx year
4	12.8731	0.1895	aIndex loudness tempo year
4	13.0037	0.1863	artistFamIdx loudness tempo year

The SAS System 11:21 Friday, December 4, 2015 129

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

C(p) Selection Method

Number in Model	C(p)	R-Square	Variables in Model
4	13.0186	0.1859	artistHotIdx duration loudness year
4	13.0518	0.1851	duration loudness tempo year
1	13.0801	0.0365	duration
2	13.4786	0.0760	loudness year
5	13.6256	0.2203	aIndex artistHotIdx loudness tempo year
5	13.6256	0.2203	aIndex artistFamIdx loudness tempo year
5	13.6256	0.2203	artistHotIdx artistFamIdx loudness tempo year
3	13.8155	0.1170	artistHotIdx duration year



1	14.0233	0.0132	loudness
2	14.0645	0.0615	artistHotIdx duration
2	14.0837	0.0611	tempo year
3	14.1664	0.1083	artistHotIdx loudness year
3	14.3672	0.1034	duration tempo year
2	14.4127	0.0530	artistHotIdx loudness
3	14.4476	0.1014	artistHotIdx duration loudness
1	14.5496	0.0003	tempo
3	14.5830	0.0981	artistHotIdx tempo year
3	14.6775	0.0957	duration loudness tempo
2	14.7615	0.0444	artistHotIdx tempo
5	14.9907	0.1866	artistHotIdx duration loudness tempo year
2	15.0801	0.0365	duration tempo
3	15.2971	0.0805	loudness tempo year
4	15.6431	0.1212	artistHotIdx duration tempo year
4	16.0129	0.1121	artistHotIdx loudness tempo year
2	16.0232	0.0132	loudness tempo
3	16.0644	0.0615	artistHotIdx duration tempo
4	16.3846	0.1030	artistHotIdx duration loudness tempo
3	16.4127	0.0530	artistHotIdx loudness tempo

NOTE: Models of not full rank are not included.

The SAS System 11:21 Friday, December 4, 2015 130

The REG Procedure  
Model: MODEL1  
Dependent Variable: hIndexRT

Number of Observations Read	30
Number of Observations Used	30

#### Analysis of Variance

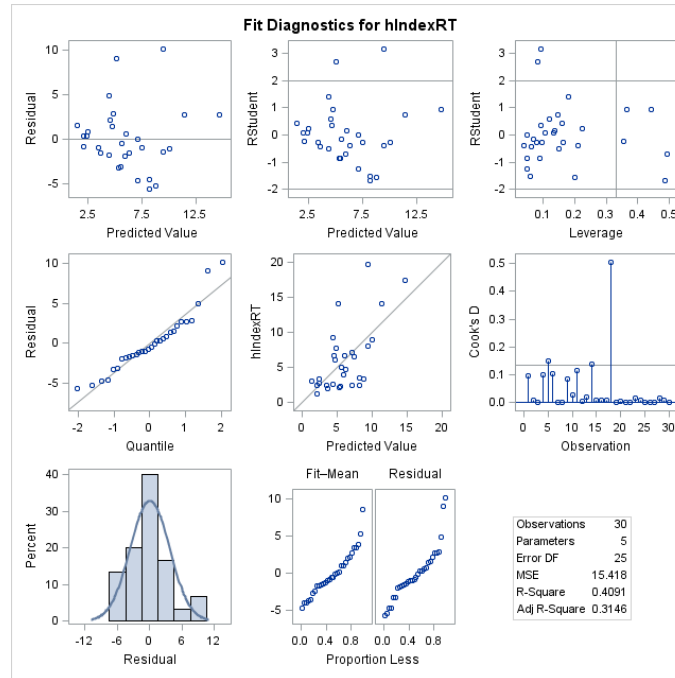
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Root MSE	3.92653	R-Square	0.4091
Dependent Mean	6.07712	Adj R-Sq	0.3146
Coeff Var	64.61168		

#### Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Tolerance	Variance Inflation
Intercept	1	-359.86178	240.68605	-1.50	0.1474	.	0
aIndex	1	-49.33357	16.00210	-3.08	0.0049	0.82918	1.20602
duration	1	0.03794	0.01243	3.05	0.0053	0.65684	1.52244
loudness	1	-0.20173	0.10865	-1.86	0.0752	0.73197	1.36617
year	1	0.18338	0.11983	1.53	0.1385	0.93950	1.06439

FIG. 14. SAS Reg Procedure



The SAS System

14:47 Friday, December 4, 2015 173

The REG Procedure

Model: MODEL1

Dependent Variable: hIndexRT

Number of Observations Read 30  
Number of Observations Used 30

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	266.88619	66.72155	4.33	0.0085
Error	25	385.44016	15.41761		
Corrected Total	29	652.32635			

Root MSE 3.92653 R-Square 0.4091  
Dependent Mean 6.07712 Adj R-Sq 0.3146  
Coeff Var 64.61168

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS	Type II SS	Tolerance
Intercept	1	-359.86178	240.68605	-1.50	0.1474	1107.93985	34.46566	.
aIndex	1	-49.33357	16.00210	-3.08	0.0049	99.68574	146.53726	0.82918
duration	1	0.03794	0.01243	3.05	0.0053	92.08967	143.58644	0.65684
loudness	1	-0.20173	0.10865	-1.86	0.0752	39.00624	53.14988	0.73197
year	1	0.18338	0.11983	1.53	0.1385	36.10455	36.10455	0.93950

## Parameter Estimates

Variable	DF	Variance Inflation
Intercept	1	0
aIndex	1	1.20602
duration	1	1.52244
loudness	1	1.36617
year	1	1.06439

The SAS System

14:47 Friday, December 4, 2015 174

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: hIndexRT

## Collinearity Diagnostics

Number	Eigenvalue	Condition Index
1	4.33882	1.00000
2	0.57452	2.74810
3	0.05186	9.14715
4	0.03480	11.16570
5	0.00000445	987.19164

## Collinearity Diagnostics

Number	-----Proportion of Variation-----				
	Intercept	aIndex	duration	loudness	year
1	4.640779E-7	0.00235	0.00352	0.01119	4.673853E-7
2	1.041306E-7	0.00223	0.01581	0.59225	1.058947E-7
3	0.00002145	0.04168	0.94009	0.33724	0.00002186
4	0.00003546	0.93455	0.02790	0.02039	0.00003627
5	0.99994	0.01919	0.01267	0.03893	0.99994

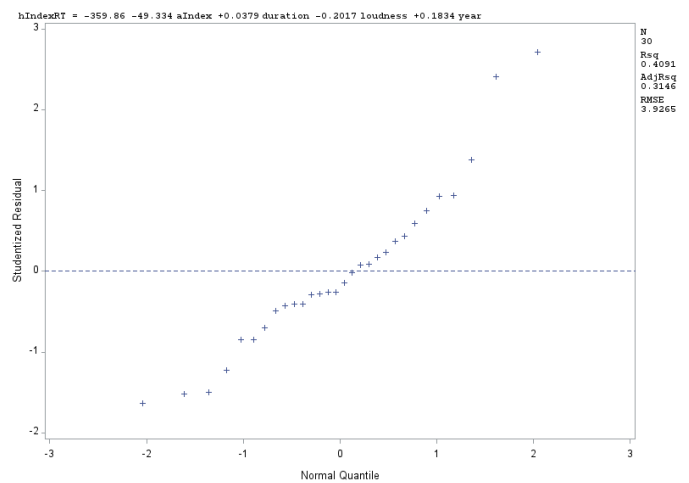
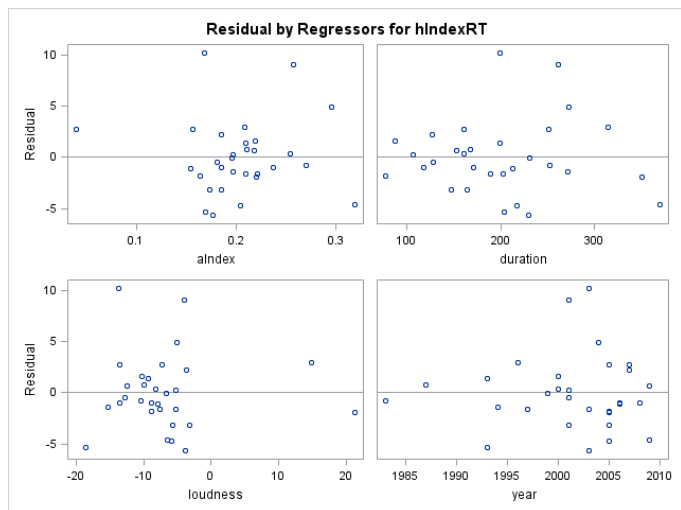
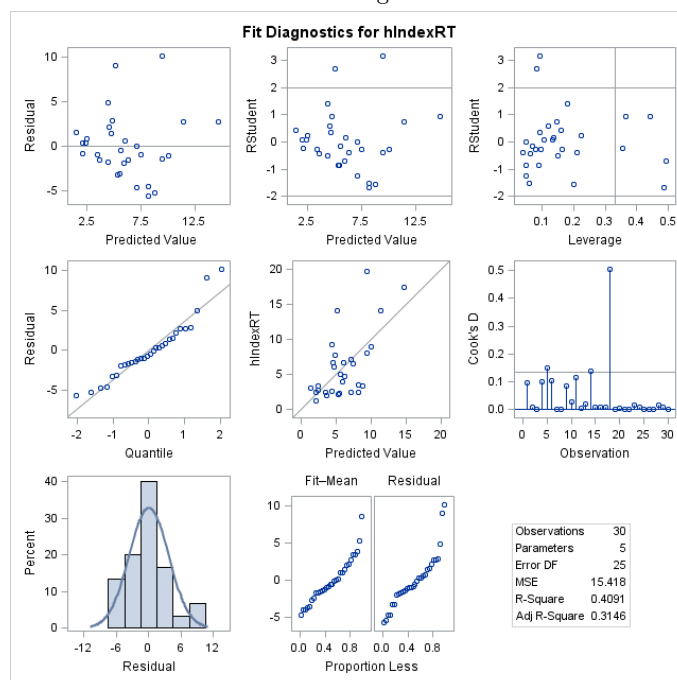


FIG. 15. SAS Reg Procedure



```

data songs;
infile '\\Client\H$\Dropbox\UNC\5\STOR 455\Songs.DAT' dlm='09'x;
input hotIdx artistHotIdx artistFamIdx duration loudness tempo year;
run;
proc print data=songs;
run;
proc means data=songs;
var hotIdx artistHotIdx artistFamIdx duration loudness tempo year;
run;
proc gplot data=songs;
plot hotIdx*(artistHotIdx artistFamIdx duration loudness tempo year);
run;
proc univariate data=songs alpha=0.5;
var hotIdx;
histogram / endpoints=0 to 1 by 0.1;
run;
proc corr data=songs;
var hotIdx artistHotIdx artistFamIdx duration loudness tempo year;
run;
proc reg data=songs;
model hotIdx = artistHotIdx artistFamIdx duration loudness tempo year / ss1 ss2 VIF TOL r influence;
run;
data logSongs;
set songs;
logIndex = log(hotIdx);
run;
proc reg data=logSongs;
model logIndex = artistHotIdx artistFamIdx duration loudness tempo year / ss1 ss2 VIF TOL;
run;
proc transreg data=songs;
model BoxCox(hotIdx / convenient lambda=-2 to 2 by 0.05)
= qqpoint(artistHotIdx artistFamIdx duration loudness tempo year);
run;
*Select lambda as -2;
data songsK;
set songs;
hIndexRT = hotIdx** -2;
run;
proc reg data=songsK;
model hIndexRT = artistHotIdx artistFamIdx duration loudness tempo year / ss1 ss2 VIF TOL;
run;
proc sgscatter data=songsK;
matrix hIndexRT artistHotIdx artistFamIdx duration loudness tempo year;
run;
data songsAdj;
set songsK;
aIndex = artistFamIdx - artistHotIdx;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=FORWARD VIF TOL slentry=0.1;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=B VIF TOL slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=STEPWISE VIF TOL slentry=0.15 slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = artistHotIdx artistFamIdx duration loudness tempo year /
selection=STEPWISE VIF TOL slentry=0.15 slstay=0.15;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=MAXR VIF TOL;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=adjrsq VIF TOL;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex artistHotIdx artistFamIdx duration loudness tempo year /
selection=cp VIF TOL;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex duration loudness year /
ss1 ss2 VIF TOL;
plot student.*nqq.;
run;
proc reg data=songsAdj;
model hIndexRT = aIndex duration loudness year /

```

```
ss1 ss2 VIF TOL collin;  
run;
```