GENREG DID THAT?

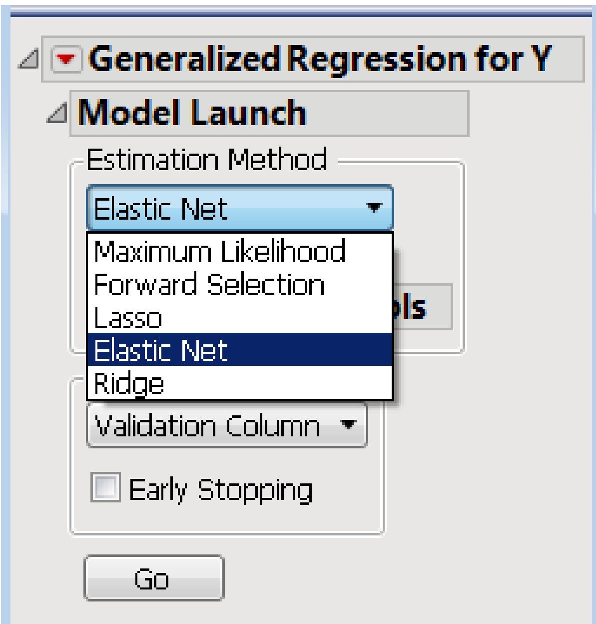
- Clay Barker
- Research Statistician Developer
- JMP Division, SAS Institute

GENREG | WHAT IS IT?

- The Generalized Regression platform was introduced in JMP Pro 11 and got much better in version 12 (more interactive and faster).
- The goal of the platform is to help build better models using variable selection techniques (including penalized regression).
- This platform can be your go-to platform for building regression models.
- From here on, we'll just call it "Genreg".

GENREG | VARIABLE SELECTION TECHNIQUES

- We provide penalized regression techniques (lasso, elastic net, ridge).

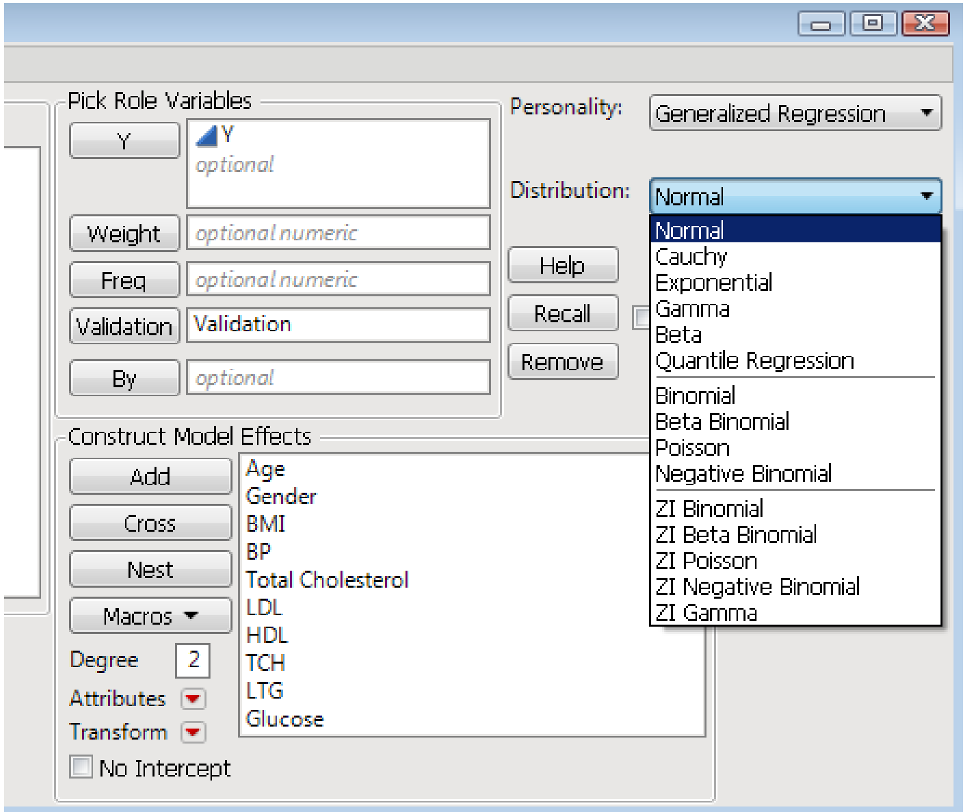


- Lasso does estimation and variable selection simultaneously.
- Elastic Net is similar to Lasso, but also naturally handles highly correlated predictors.
- Forward Selection is a classic: add the most significant predictor, then the next. Keep going until we add everything or run out of degrees of freedom.

- How do lasso and elastic net differ?
  If x_1 and x_2 are highly correlated, lasso will probably pick x_1 **or** x_2.
  Elastic net will probably pick x_1 **and** x_2.

GENREG | GENERALIZED LINEAR MODELS

- And the response doesn’t have to be normally distributed, we can do variable selection for generalized linear models (GLMs) all in the same place.



- Gamma for skewed responses.
- Poisson for count data.
- Negative Binomial for overdispersed count data.
- Binomial for logistic regression.
- Beta for rates.
- Cauchy for robust regression.
- … and quantile regression?

GENREG | EXAMPLE: PREDICTING DIABETES PROGRESSION

Model Comparison

Measures of Fit for Y

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	LS model	Fit Least Squares		0.6121	46.592	36.555	266
Training	Lasso model	Fit Generalized Adaptive Lasso		0.5131	52.199	41.953	266
Test	LS model	Fit Least Squares		0.2697	71.180	56.187	64
Test	Lasso model	Fit Generalized Adaptive Lasso		0.5118	58.195	46.148	64

Adaptive Lasso with Validation Column Validation

Model Summary

Response	Y
Distribution	Normal
Estimation Method	Adaptive Lasso
Validation Method	Validation Column
Mean Model Link	Identity
Scale Model Link	Identity

Measure	Training	Validation	Test
Number of rows	266	112	64
Sum of Frequencies	266	112	64
-LogLikelihood	1428.9322	603.39041	349.43694
BIC	2997.4518	1324.7433	802.84596
AICc	2913.2811	1271.8971	783.08441
Generalized RSquare	0.5151151	0.5472521	0.5524837

Estimation Details

Number of Grid Points	500
Minimum Penalty Fraction	0
Grid Scale	Linear

Solution Path

Parameter Estimates for Centered and Scaled Predictors

Parameter Estimates for Original Predictors

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Intercept	-320.0803	63.468565	25.433169	<.0001*	-444.4764	-195.6842
Age	0	0	0	1.0000	0	0
Gender[1]	14.57521	7.7238566	3.5609122	0.0592	-0.56327	29.713691
BMI	5.3627266	0.9922228	29.211438	<.0001*	3.4180057	7.3074475
BP	1.1024422	0.2639133	17.449748	<.0001*	0.5851817	1.6197027
Total Cholesterol	-0.122971	0.1350058	0.829664	0.3624	-0.387578	0.1416352
LDL	0	0	0	1.0000	0	0
HDL	-0.614639	0.4843082	1.6106335	0.2044	-1.563866	0.3345875
TCH	0	0	0	1.0000	0	0
LTG	58.496377	12.653757	21.370709	<.0001*	33.695468	83.297285
Glucose	0	0	0	1.0000	0	0
(Age-48.5181)*Gender[1]	-0.577619	0.4151188	1.9361469	0.1641	-1.391237	0.2359988

- Traditional least squares model overfits badly. Lasso does a better job predicting new observations.
- The lasso sets some of the coefficients to zero (variable selection).

MORE TO OFFER

- Quickly and easily building models that perform well on new observations is Genreg's bread and butter…
- But with a little creativity, we can use the platform to do some things that you might not realize it can do:

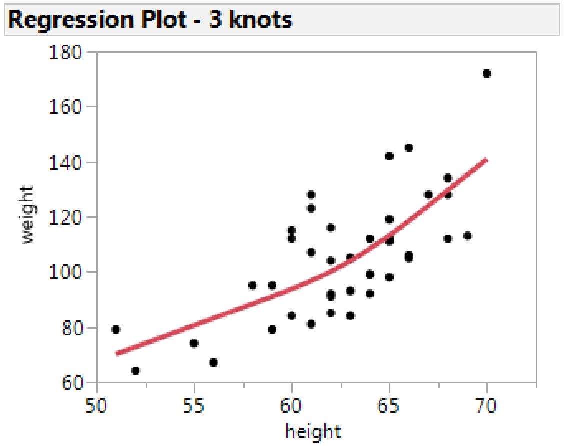
1. Knot selection for splines
2. The Bradley-Terry model
3. Changepoint detection

- We will take a look at how to do these things in Genreg.

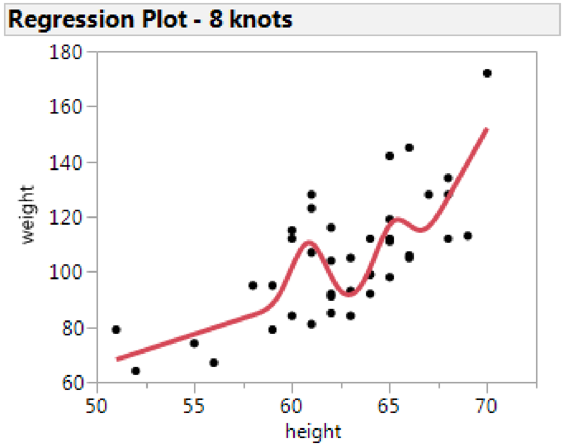
SPLINE KNOT SELECTION

- Splines are piecewise polynomial smoothing functions.
- The polynomials connect at points called knots.
  More knots means more flexibility!
- Splines are very useful but be careful!
  We only want as much flexibility as necessary.
- Too many knots will surely lead to overfitting.

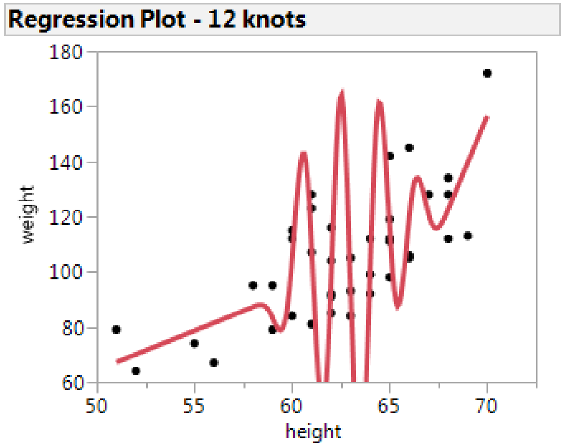
Good



Bad



Ugly



- There are lots of ways to splines, but JMP uses cubic splines in the Fit Model platform.
- A cubic spline for x with k knots looks something like

$$f(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{k-2} \gamma_j g_j(x)$$

  where the $g_j(x)$ are relatively simple cubic functions.
- We can turn the $g_j(x)$ into design columns and use familiar tools for least squares estimation.

  Design matrix: $[\,\mathbf{1} \quad \boldsymbol{x} \quad \boldsymbol{g_1(x)} \quad \boldsymbol{g_2(x)} \quad \ldots \quad \boldsymbol{g_{k-2}(x)}\,]$

- If we could set some of the $\hat{\gamma}_j$ to zero when appropriate, then we could control the flexibility of our spline…

KNOT SELECTION AND THE LASSO

- The lasso does this for us! Using the lasso for estimation/selection, we can be more confident that we will not overfit our data.

Lasso with AICc Validation

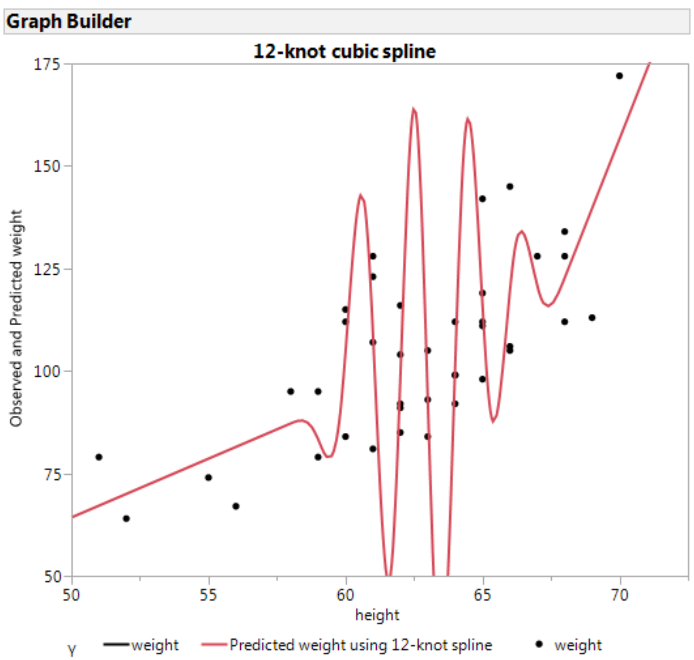
Model Summary

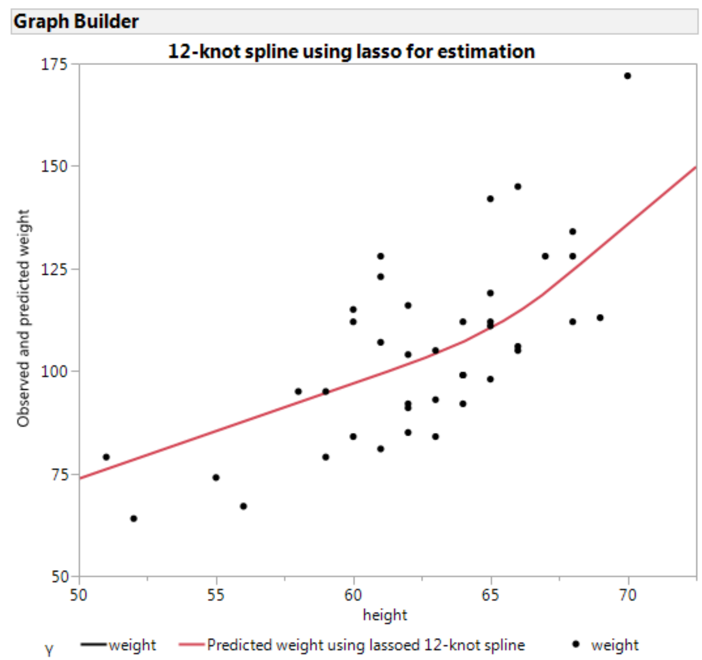
Response	weight
Distribution	Normal
Estimation Method	Lasso
Validation Method	AICc
Mean Model Link	Identity
Scale Model Link	Identity

Measure	Training
Number of rows	40
Sum of Frequencies	40
-LogLikelihood	165.95791
BIC	346.67133
AICc	341.05867
Generalized RSquare	0.5108163

Parameter Estimates for Original Predictors

Term	Estimate
Intercept	-42.47561
height&Knotted	2.323847
height&Knotted@58	0
height&Knotted@58.90909	0
height&Knotted@59.81818	0
height&Knotted@60.72727	0.023727
height&Knotted@61.63636	0
height&Knotted@62.54545	0
height&Knotted@63.45455	0
height&Knotted@64.36364	0
height&Knotted@65.27273	0
height&Knotted@66.18182	0
Scale	15.333027





…AND THE LASSO

- You can't really know in advance how many knots to use.
- Using the lasso for estimation with cubic splines gives us a safety net.
- We can specify a larger number of knots and the lasso will figure out which ones are really necessary.
- This helps us avoid overfitting our data and saves us time!

Corn data DEMO

THE BRADLEY-TERRY MODEL

BRADLEY-TERRY | WHAT IS IT?

- The Bradley-Terry model is a popular way of modeling the outcome of a comparison or competition.
- Suppose that we have some teams that play each other. When team i plays team j, the Bradley-Terry model gives us

$$\Pr(\text{team } i \text{ beats team } j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}$$

- We can interpret the $\hat{\beta}_k$ as strength measures.

  If $\hat{\beta}_i > \hat{\beta}_j$, then team i is a stronger competitor than team j.
- More than just a tool for ranking sports teams…
  - Relevance of results from a search engine
  - Comparing product desirability (market research)
  - Ranking the quality of academic journals

BRADLEY-TERRY …AND LOGISTIC REGRESSION

- At first glance, the Bradley-Terry model does not look like something we can fit in JMP (except for maybe in the Nonlinear platform)…
- But let's rewrite it,

$$\Pr(\text{team } i \text{ beats team } j) = \frac{e^{\beta_i}}{e^{\beta_i}+e^{\beta_j}}$$

$$= \frac{1}{1+e^{\beta_j-\beta_i}}$$

$$= \frac{1}{1+e^{-(\beta_i-\beta_j)}}$$

- This looks just like a no-intercept logistic regression model!
- Recall the logistic regression probability function

$$\Pr(y = 1 | x) = \frac{1}{1+e^{-x\beta}}$$

BRADLEY-TERRY …AND LOGISTIC REGRESSION

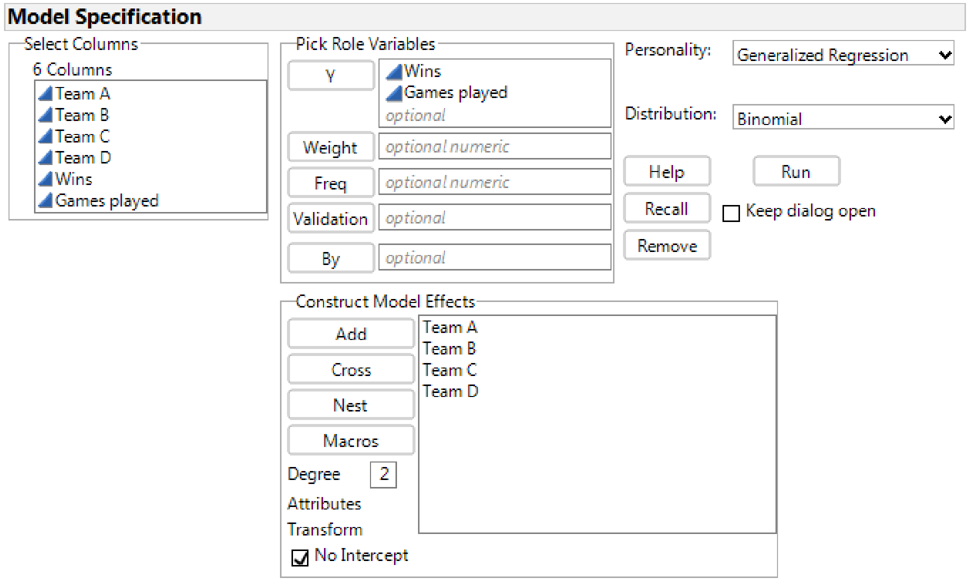
- So we can turn the Bradley-Terry model into a no-intercept logistic regression problem.
- The key is in carefully constructing our data set (next slide).

- JMP can do logistic regression in several places, including Genreg.
- The advantage to using Genreg for fitting is that it enables us to do variable selection.
- By doing variable selection, we can quickly distinguish:
  - The strong competitors $(\hat{\beta}_j > 0)$
  - The weak competitors $(\hat{\beta}_j < 0)$
  - The mediocre competitors clumped in the middle $(\hat{\beta}_j = 0)$

- Suppose a little-league (4 teams) basketball season plays out as follows

Team A beats Team B 3 times out of 4	Team B beats Team C 1 time out of 3	Team C beats Team D 3 times out of 4
Team A beats Team C 2 times out of 3	Team B beats Team D 3 times out of 4	
Team A beats Team D 4 times out of 4		

	Team A	Team B	Team C	Team D	Wins	Games played
1	1	-1	0	0	3	4
2	1	0	-1	0	2	3
3	1	0	0	-1	4	4
4	0	1	-1	0	1	3
5	0	1	0	-1	3	4
6	0	0	1	-1	3	4



BRADLEY-TERRY | LITTLE LEAGUE RESULTS

Generalized Regression for Wins

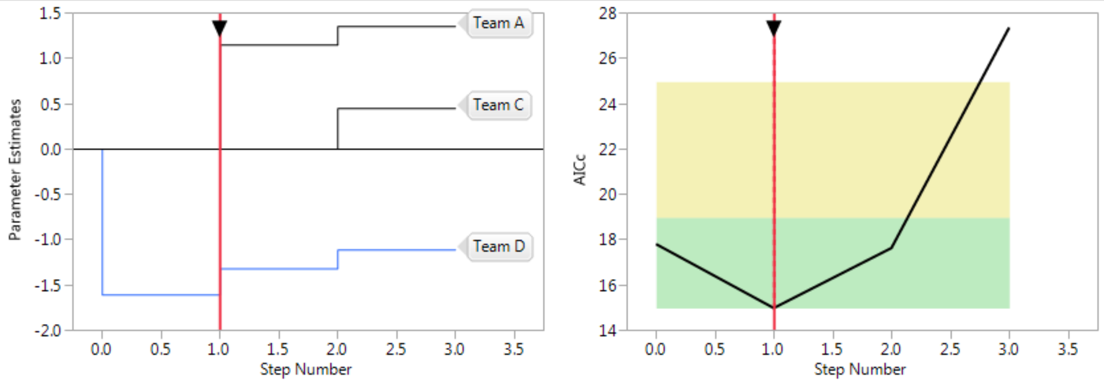
Forward Selection with AICc Validation

Model Summary

Response	Wins
Distribution	Binomial
Estimation Method	Forward Selection
Validation Method	AICc
Probability Model Link	Logit

Measure	Training
Number of rows	6
Sum of Frequencies	6
-LogLikelihood	5.9820987
BIC	13.755957
AICc	14.964197
Generalized RSquare	0.6548318

Solution Path



Parameter Estimates for Original Predictors

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%	Singularity Details
Team A	0	0	0	1.0000	0	0	
Team B	0	0	0	1.0000	0	0	
Team C	0	0	0	1.0000	0	0	
Team D	-1.609438	0.4898979	10.792877	0.0010*	-2.56962	-0.649256	=-Team A-Team B-Team C

- The results are a bit depressing!
- Teams A, B, and C are similar…
  But Team D stinks.

- The Bradley-Terry model is a cool way to model competitions and we can do it in Genreg.

- If we're not interested in the individual competitors but instead want to use their features/attributes to predict outcomes, we can do that too. …but creating the data set gets trickier. We'll look at an example.

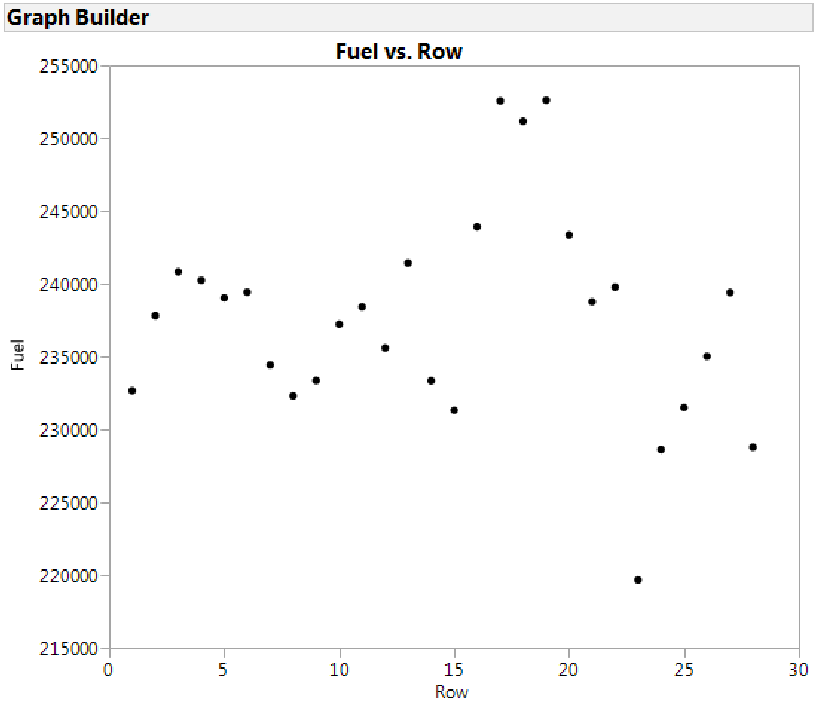
Basketball data DEMO

CHANGEPOINT DETECTION

CHANGEPOINT DETECTION

WHAT IS IT?

- Changepoint detection is a very important problem in data analysis.
- When our data have a natural ordering, a changepoint occurs when the mean of our response shifts up or down.

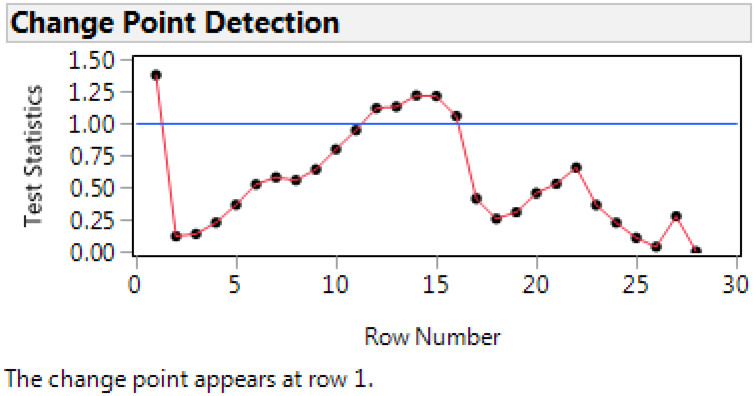


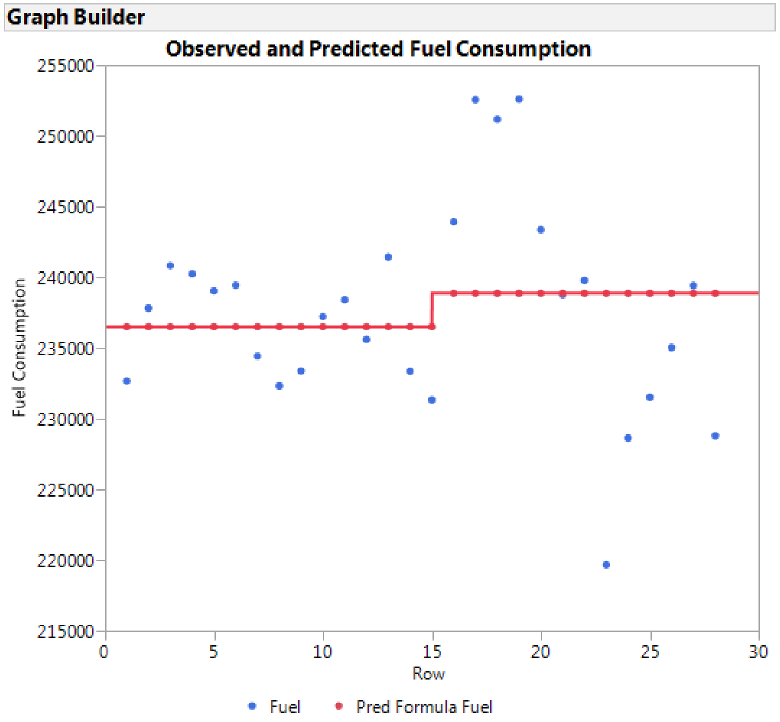
- Steam Turbine Historical.jmp from sample data.
- Monitoring daily fuel consumption.
- It looks like the mean shifts up around row 15 and back down around row 20.
- But we don’t want to eyeball it, we want to let the data decide if/when shifts happen.

CHANGEPOINT DETECTION

IN MULTIVARIATE CONTROL CHART PLATFORM

- The Multivariate Control Chart platform provides a changepoint detection tool, but it only makes it easy to find a single shift.
- And it doesn't work too well for this example.





FUSED LASSO

- The fused lasso is a promising technique for finding changepoints.

$$\hat{\boldsymbol{\mu}} = \arg\min_{\mu} \sum_{i=1}^{n} (y_i - \mu_i)^2 + \lambda \sum_{j=2}^{n} |\mu_j - \mu_{j-1}|$$

- So each point gets its own mean, but we try to set many of the means equal to their neighbors. They get “fused” together.

- For our example, we might end up with something like

$$\hat{\mu}_1 = \hat{\mu}_2 = \cdots = \hat{\mu}_{15}$$
$$\hat{\mu}_{16} = \hat{\mu}_{17} = \cdots = \hat{\mu}_{20}$$
$$\hat{\mu}_{21} = \hat{\mu}_{22} = \cdots = \hat{\mu}_{28}$$

- That would imply mean shifts at rows 16 and 21.
- Bummer we can’t do this in JMP…

FUSED LASSO

- …Or maybe we can do it!
- Instead of the original form, we could rewrite the model

$$E(Y_1) = \mu$$
$$E(Y_2) = E(Y_1) + \beta_2 = \mu + \beta_2$$
$$E(Y_3) = E(Y_2) + \beta_3 = \mu + \beta_2 + \beta_3$$
$$\vdots$$
$$E(Y_j) = E(Y_{j-1}) + \beta_j = \mu + \sum_{k=2}^{j} \beta_j$$

- If $\hat{\beta}_j \neq 0$, then the mean shifts at point j.
- Easy to express as a design matrix: $n \times n$ lower triangular matrix of ones.

$$n = 6: \quad X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

CHANGEPOINT DETECTION

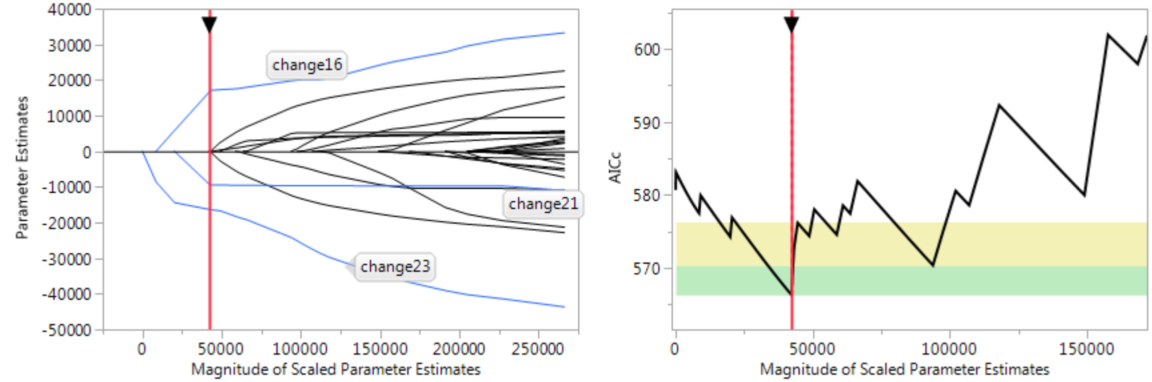
FUSED LASSO

- Using this new parameterization and the Lasso for estimation is equivalent to the fused lasso!
- So yes, we can do the fused lasso in Genreg using a carefully crafted design matrix.

Generalized Regression for Fuel

Lasso with AICc Validation

Solution Path

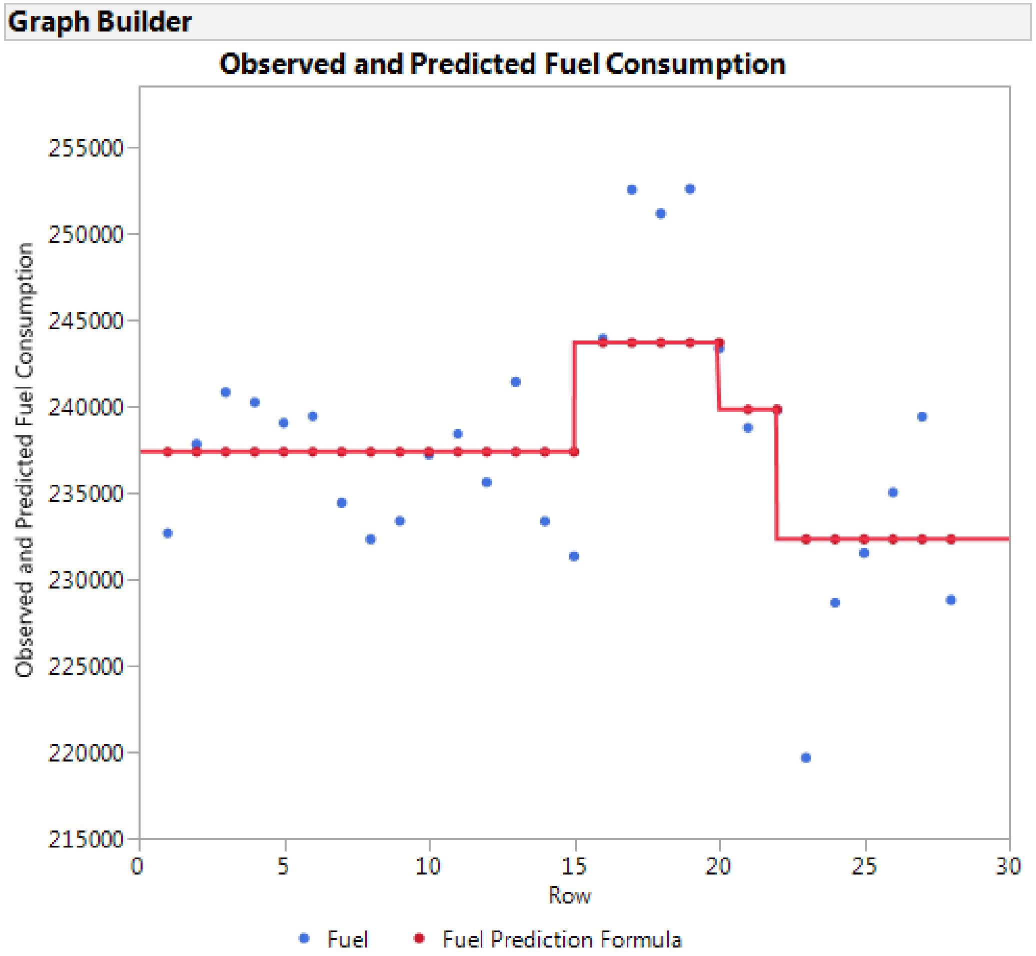


Parameter Estimates for Original Predictors

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Intercept	237378.03	1297.189	33486.87	<.0001*	234835.59	239920.48
change2	0	0	0	1.0000	0	0
change3	0	0	0	1.0000	0	0
change4	0	0	0	1.0000	0	0
change5	0	0	0	1.0000	0	0
change6	0	0	0	1.0000	0	0
change7	0	0	0	1.0000	0	0
change8	0	0	0	1.0000	0	0
change9	0	0	0	1.0000	0	0
change10	0	0	0	1.0000	0	0
change11	0	0	0	1.0000	0	0
change12	0	0	0	1.0000	0	0
change13	0	0	0	1.0000	0	0
change14	0	0	0	1.0000	0	0
change15	0	0	0	1.0000	0	0
change16	6317.4919	8308.7634	0.5781182	0.4471	-9967.385	22602.369
change17	0	0	0	1.0000	0	0
change18	0	0	0	1.0000	0	0
change19	0	0	0	1.0000	0	0
change20	0	0	0	1.0000	0	0
change21	-3873.96	7868.5925	0.2423908	0.6225	-19296.12	11548.198
change22	0	0	0	1.0000	0	0
change23	-7506.438	3438.8293	4.764821	0.0290*	-14246.42	-766.4566
change24	0	0	0	1.0000	0	0
change25	0	0	0	1.0000	0	0
change26	0	0	0	1.0000	0	0
change27	0	0	0	1.0000	0	0
Scale	4752.3567	2318.0881	4.2029882	0.0404*	208.98742	9295.7259

The mean shifts at rows 16, 21, and 23.

STEAM TURBINE EXAMPLE



- The fused lasso results are much more reasonable than what we saw earlier.

FUSED LASSO

- The fused lasso is a very powerful tool for finding changepoints in our data.
- If we set up our data properly, we can fit this model using the lasso in Genreg!

  …and it is easy to set up the data.

  This JSL snippet gets you most of the way there.

```
d = J(n,n,0);          // n is the number of observations
for(i=1, i<=n, i++,
   d[i,1::i] = 1;
);
as table(d);
```

Lebron James data and TV show data Demos

SUMMARY

- Genreg is the go-to place for building generalized linear models.
- The interactive nature of the platform makes it easy to quickly build models that will generalize well to new observations.

- But there is more to it!
- We have looked at examples using the platform for
  1. Knot selection for splines
  2. Modeling competitions using the Bradley-Terry model
  3. Searching for changepoints in our response
  4. Screening for outliers (and adjusting for them)

- The powerful variable selection techniques available in Genreg allow us to take on more problems than we could otherwise.

SOME IMPORTANT RESOURCES

- Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.
- Zou, Hui, and Trevor Hastie. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005): 301-320.
- Hesterberg, Tim, et al. "Least angle and ℓ1 penalized regression: A review."*Statistics Surveys* (2008): 61-93.
- Hoerl, Arthur E., and Robert W. Kennard. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12.1 (1970): 55-67.
- Zou, Hui. "The adaptive lasso and its oracle properties." *Journal of the American statistical association* 101.476 (2006): 1418-1429.
- Tibshirani, Robert, et al. "Sparsity and smoothness via the fused lasso."*Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.1 (2005): 91-108.
- Bradley, Ralph Allan, and Milton E. Terry. "Rank analysis of incomplete block designs: The method of paired comparisons." *Biometrika* 39.3-4 (1952): 324-345.