Lure your Black Belts into Learning Statistical Tools by using Practical Examples

Don Lifke

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Overview

- Teaching statistical tools to reluctant Black Belt students can be a challenge
- One successful technique is to use examples that they can relate to, and to show that these tools can be applied outside their work lives as well
- Two such examples will be presented
  - Use Fit Model to determine market value of a house
  - Use Design of Experiments to optimize the golf drive

Practical Application of Statistical Modeling: Determining Market Value of a House

You want to sell your house. It has the following attributes:

- **2000 square feet**
- **0.2 acre lot**
- **2 years old**
- **3 bedrooms**
- **3 full bathrooms**

What should your asking price be?

Exercise

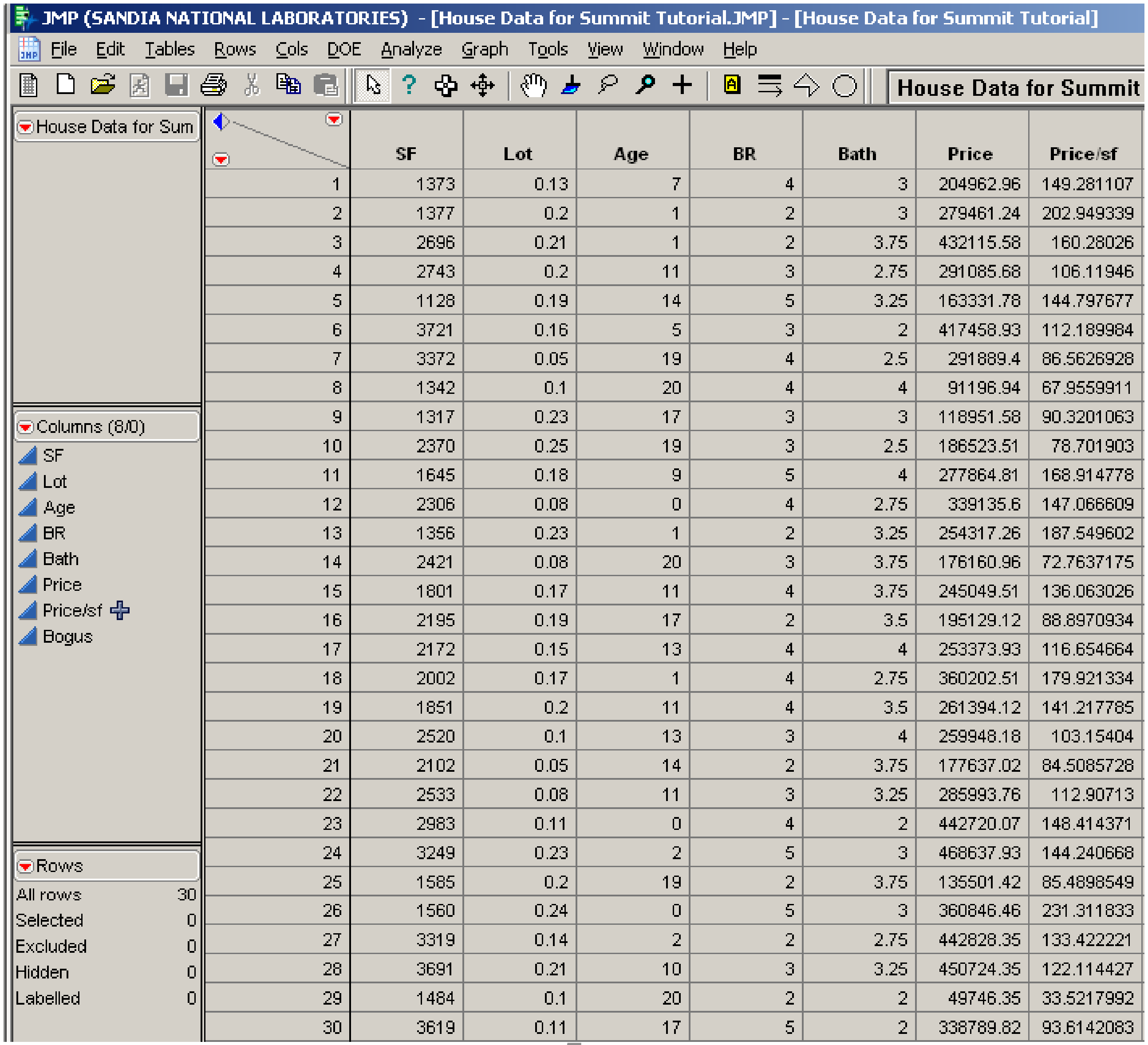
- Students are given the Excel file below with data. They are given 5 minutes to explore the data in Excel.

- Students are asked to provide listing prices based on their analyses. They typically use average $\$/ft^2$.

- Listing Prices:
  - $
  - $
  - $

House Data

Students Provided Same Data in JMP



	SF	Lot	Age	BR	Bath	Price	Price/sf
1	1373	0.13	7	4	3	204962.96	149.281107
2	1377	0.2	1	2	3	279461.24	202.949339
3	2696	0.21	1	2	3.75	432115.58	160.28026
4	2743	0.2	11	3	2.75	291085.68	106.11946
5	1128	0.19	14	5	3.25	163331.78	144.797677
6	3721	0.16	5	3	2	417458.93	112.189984
7	3372	0.05	19	4	2.5	291889.4	86.5626928
8	1342	0.1	20	4	4	91196.94	67.9559911
9	1317	0.23	17	3	3	118951.58	90.3201063
10	2370	0.25	19	3	2.5	186523.51	78.701903
11	1645	0.18	9	5	4	277864.81	168.914778
12	2306	0.08	0	4	2.75	339135.6	147.066609
13	1356	0.23	1	2	3.25	254317.26	187.549602
14	2421	0.08	20	3	3.75	176160.96	72.7637175
15	1801	0.17	11	4	3.75	245049.51	136.063026
16	2195	0.19	17	2	3.5	195129.12	88.8970934
17	2172	0.15	13	4	4	253373.93	116.654664
18	2002	0.17	1	4	2.75	360202.51	179.921334
19	1851	0.2	11	4	3.5	261394.12	141.217785
20	2520	0.1	13	3	4	259948.18	103.15404
21	2102	0.05	14	2	3.75	177637.02	84.5085728
22	2533	0.08	11	3	3.25	285993.76	112.90713
23	2983	0.11	0	4	2	442720.07	148.414371
24	3249	0.23	2	5	3	468637.93	144.240668
25	1585	0.2	19	2	3.75	135501.42	85.4898549
26	1560	0.24	0	5	3	360846.46	231.311833
27	3319	0.14	2	2	2.75	442828.35	133.422221
28	3691	0.21	10	3	3.25	450724.35	122.114427
29	1484	0.1	20	2	2	49746.35	33.5217992
30	3619	0.11	17	5	2	338789.82	93.6142083

JMP Data Table

Sandia National Laboratories

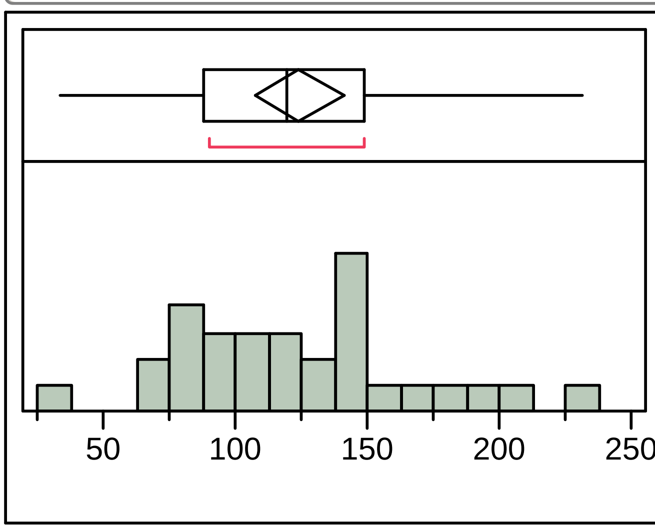
Exercise:
What Will Your Listing Price Be?

Based on Analysis of the distribution of Price/ft^2 :

Average = \$124.36/ft^2
Therefore, \$124.36/ft^2 x 2,000 ft^2 = \$248,720

Distributions

Price/sf



Quantiles

100.0%	maximum	231.31
99.5%		231.31
97.5%		231.31
90.0%		186.79
75.0%	quartile	148.63
50.0%	median	119.38
25.0%	quartile	88.31
10.0%		73.36
2.5%		33.52
0.5%		33.52
0.0%	minimum	33.52

Moments

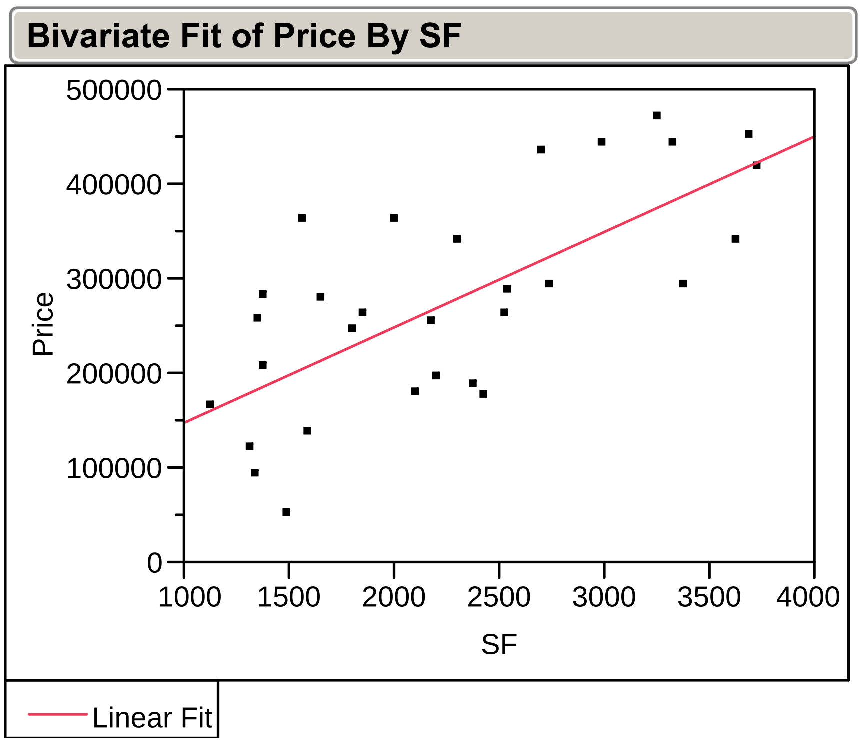
Mean	124.36354
Std Dev	44.020787
Std Err Mean	8.0370594
upper 95% Mean	140.80117
lower 95% Mean	107.92591
N	30

Sandia National Laboratories

Students Analyze JMP Data

- More observant students might say there's a fixed price as well as a cost per square foot
- Perform a Fit Y by X for Price vs. ft^2
- Add a Line Fit

Price = \$45,962 + \$101.34*ft^2 = \$248,642



Analysis of Variance

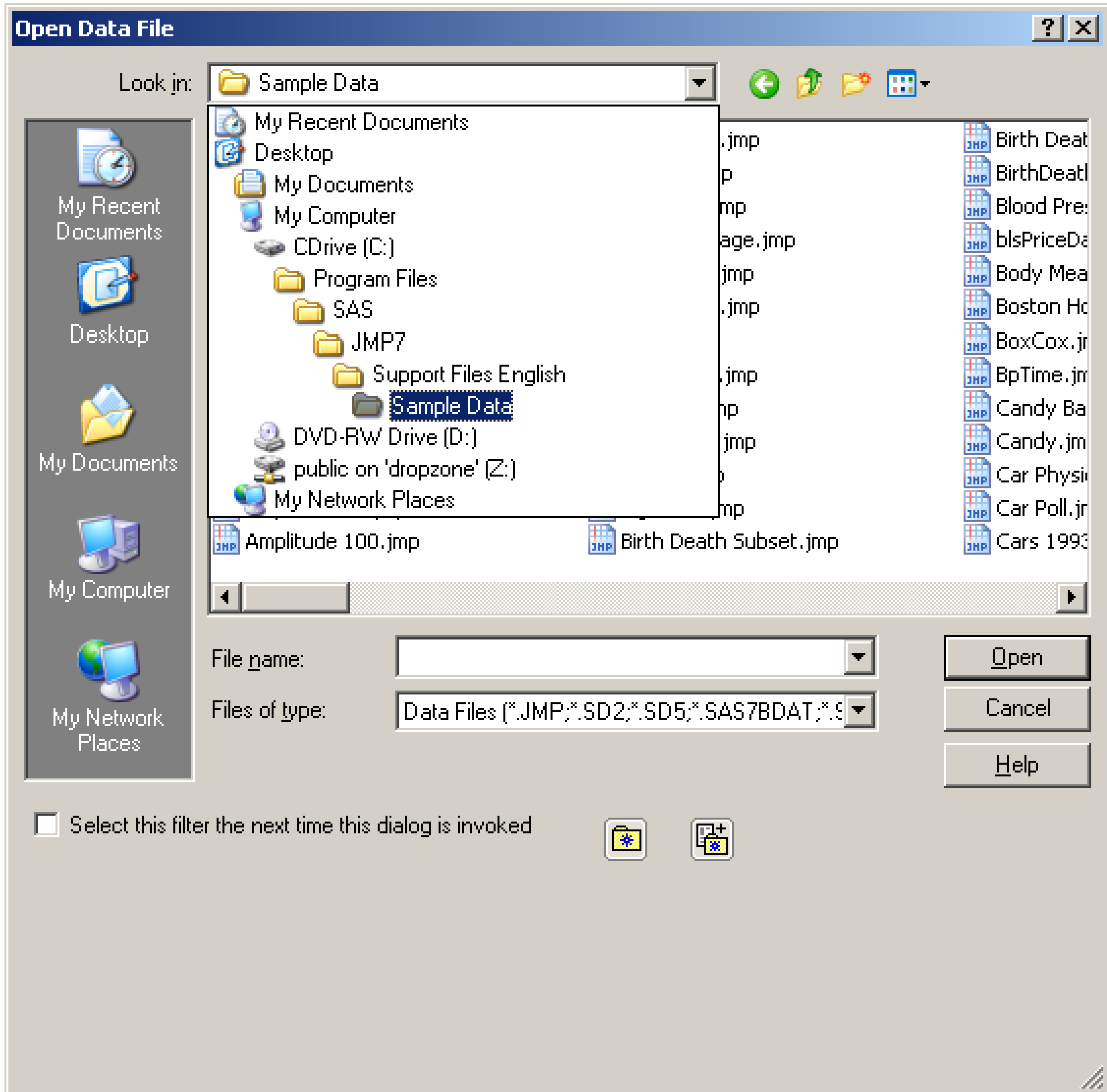
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.8791e+11	1.879e+11	28.1954
Error	28	1.8661e+11	6.6647e+9	**Prob > F**
C. Total	29	3.7452e+11		<.0001*

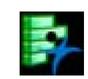
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	45962.927	45654	1.01	0.3227
SF	101.33845	19.0847	5.31	<.0001*

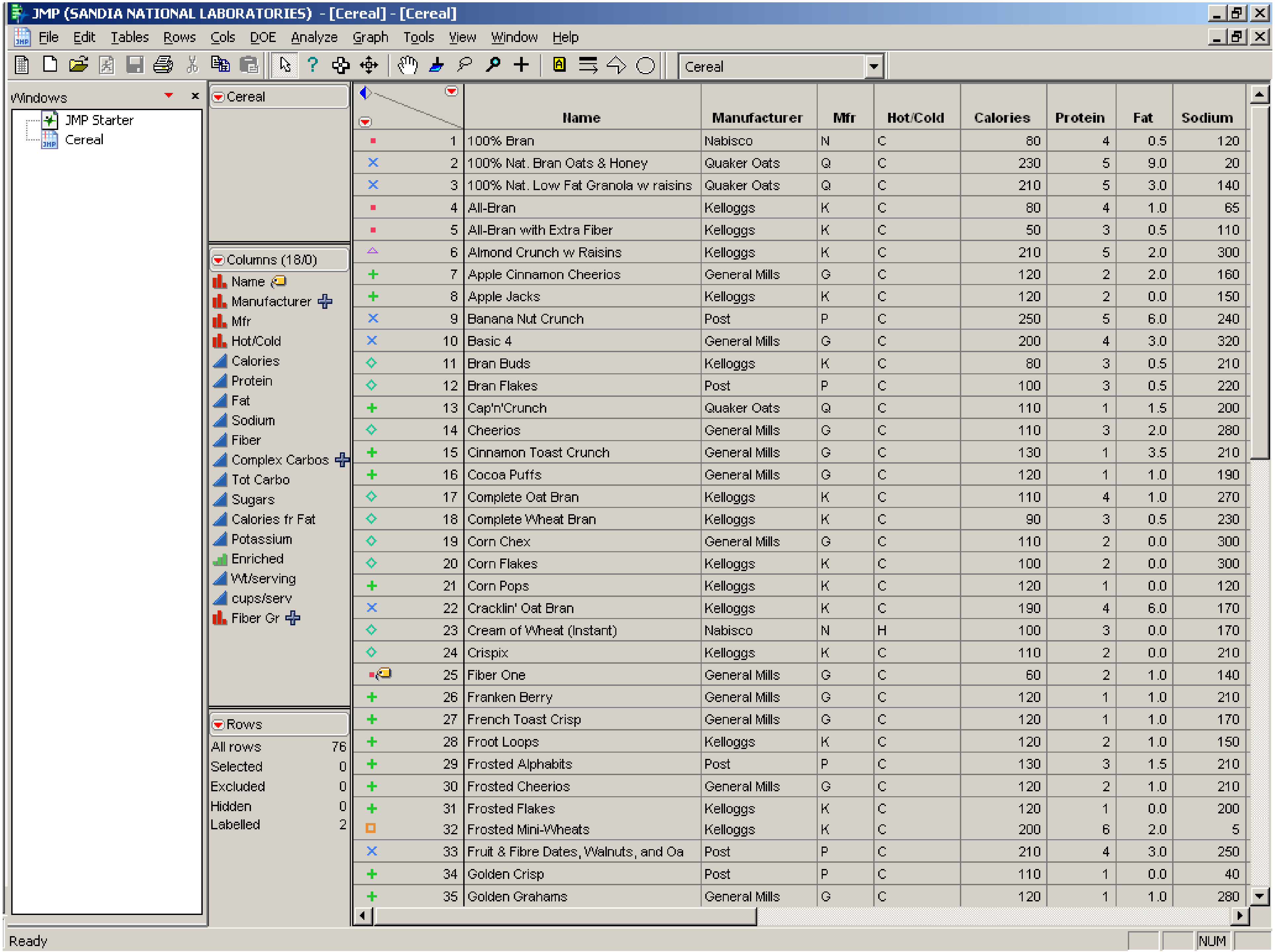
Sandia National Laboratories

Before Proceeding with the House Example, Teach Students Data Exploration using Cereal File




JMP Data Table

Sandia National Laboratories



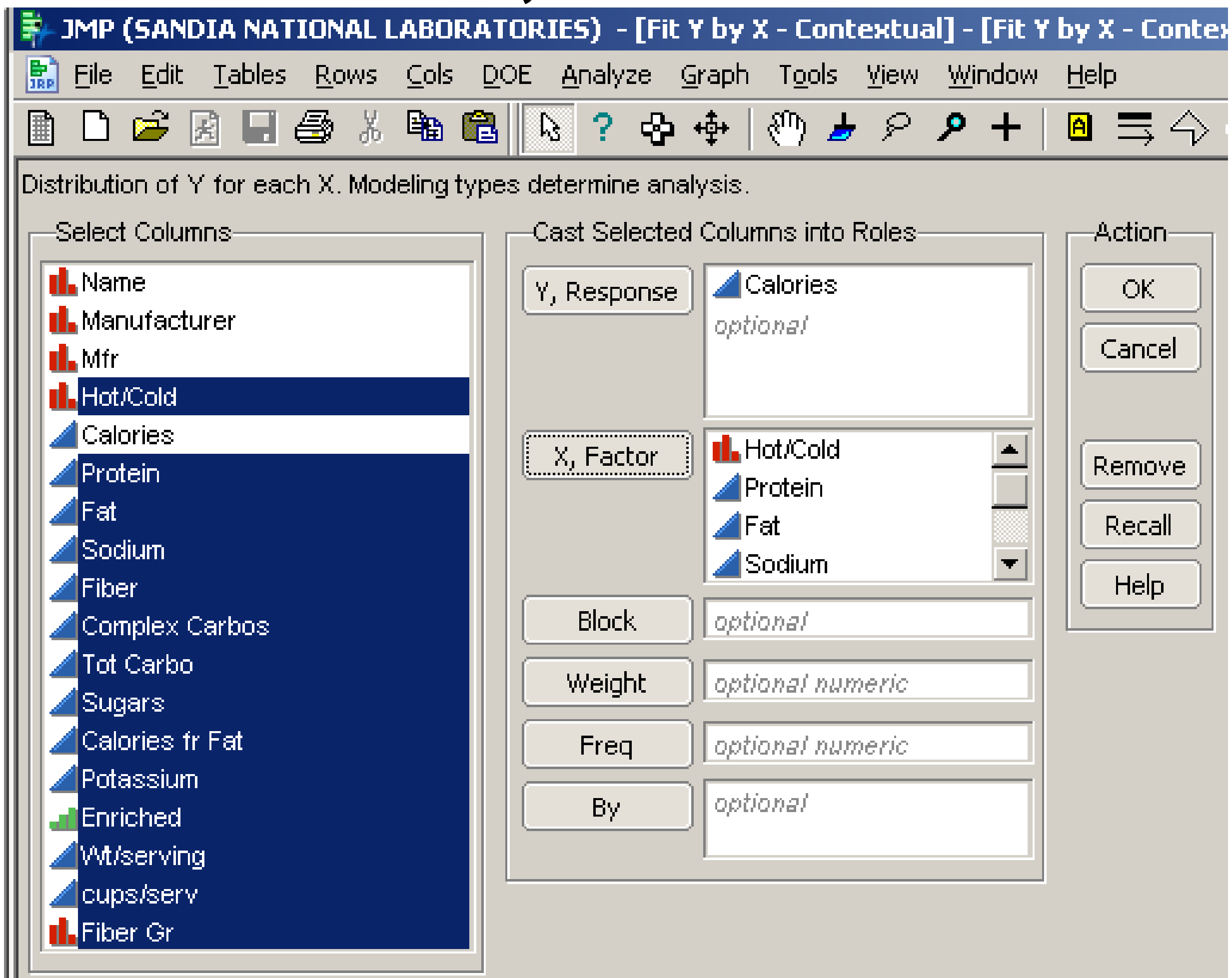
	Name	Manufacturer	Mfr	Hot/Cold	Calories	Protein	Fat	Sodium
1	100% Bran	Nabisco	N	C	80	4	0.5	120
2	100% Nat. Bran Oats & Honey	Quaker Oats	Q	C	230	5	9.0	20
3	100% Nat. Low Fat Granola w raisins	Quaker Oats	Q	C	210	5	3.0	140
4	All-Bran	Kelloggs	K	C	80	4	1.0	65
5	All-Bran with Extra Fiber	Kelloggs	K	C	50	3	0.5	110
6	Almond Crunch w Raisins	Kelloggs	K	C	210	5	2.0	300
7	Apple Cinnamon Cheerios	General Mills	G	C	120	2	2.0	160
8	Apple Jacks	Kelloggs	K	C	120	2	0.0	150
9	Banana Nut Crunch	Post	P	C	250	5	6.0	240
10	Basic 4	General Mills	G	C	200	4	3.0	320
11	Bran Buds	Kelloggs	K	C	80	3	0.5	210
12	Bran Flakes	Post	P	C	100	3	0.5	220
13	Cap'n'Crunch	Quaker Oats	Q	C	110	1	1.5	200
14	Cheerios	General Mills	G	C	110	3	2.0	280
15	Cinnamon Toast Crunch	General Mills	G	C	130	1	3.5	210
16	Cocoa Puffs	General Mills	G	C	120	1	1.0	190
17	Complete Oat Bran	Kelloggs	K	C	110	4	1.0	270
18	Complete Wheat Bran	Kelloggs	K	C	90	3	0.5	230
19	Corn Chex	General Mills	G	C	110	2	0.0	300
20	Corn Flakes	Kelloggs	K	C	100	2	0.0	300
21	Corn Pops	Kelloggs	K	C	120	1	0.0	120
22	Cracklin' Oat Bran	Kelloggs	K	C	190	4	6.0	170
23	Cream of Wheat (Instant)	Nabisco	N	H	100	3	0.0	170
24	Crispix	Kelloggs	K	C	110	2	0.0	210
25	Fiber One	General Mills	G	C	60	2	1.0	140
26	Franken Berry	General Mills	G	C	120	1	1.0	210
27	French Toast Crisp	General Mills	G	C	120	1	1.0	170
28	Froot Loops	Kelloggs	K	C	120	2	1.0	150
29	Frosted Alphabits	Post	P	C	130	3	1.5	210
30	Frosted Cheerios	General Mills	G	C	120	2	1.0	210
31	Frosted Flakes	Kelloggs	K	C	120	1	0.0	200
32	Frosted Mini-Wheats	Kelloggs	K	C	200	6	2.0	5
33	Fruit & Fibre Dates, Walnuts, and Oa	Post	P	C	210	4	3.0	250
34	Golden Crisp	Post	P	C	110	1	0.0	40
35	Golden Grahams	General Mills	G	C	120	1	1.0	280

A Few Setup Changes

- Rows → Clear Row States
- File → Preferences
  - Click Reports
    - Change Graph Marker Size to medium.
  - Click Platforms
    - Select Distribution. Under Options, select Stack.
  - Click OK.

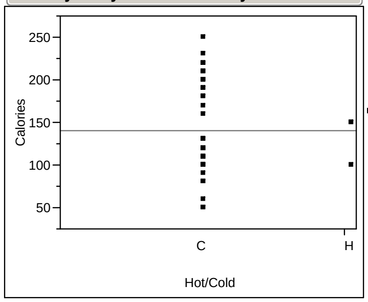
What Affects Calories?

Analyze → Fit Y by X
Select Calories for Y, those below for X

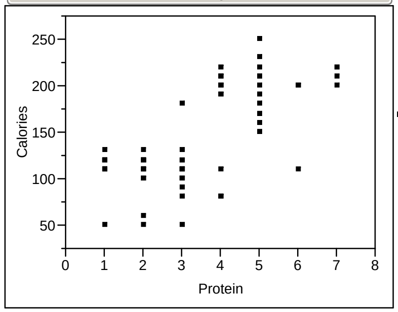


Fit Y by X Plots

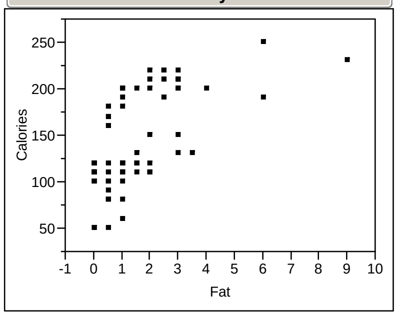
Oneway Analysis of Calories By Hot/Cold



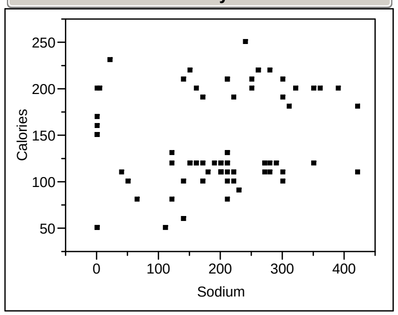
Bivariate Fit of Calories By Protein



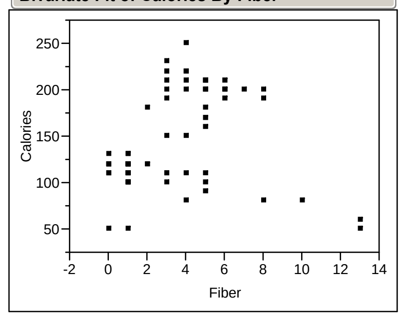
Bivariate Fit of Calories By Fat



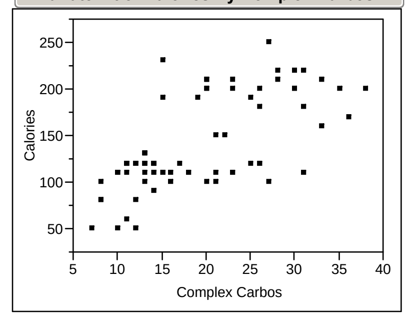
Bivariate Fit of Calories By Sodium



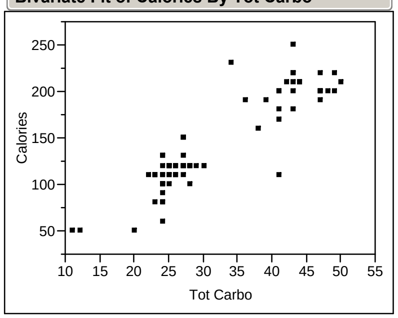
Bivariate Fit of Calories By Fiber



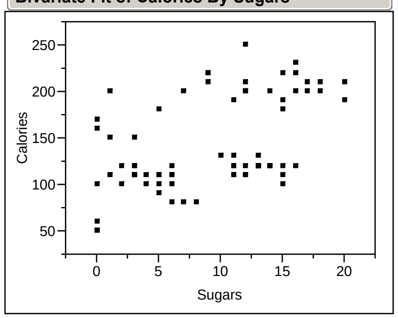
Bivariate Fit of Calories By Complex Carbos



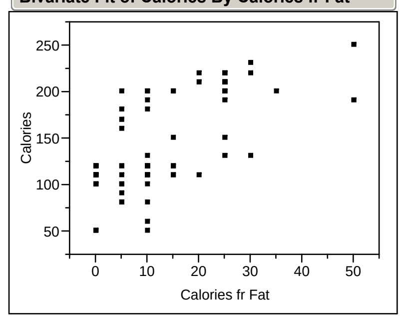
Bivariate Fit of Calories By Tot Carbo



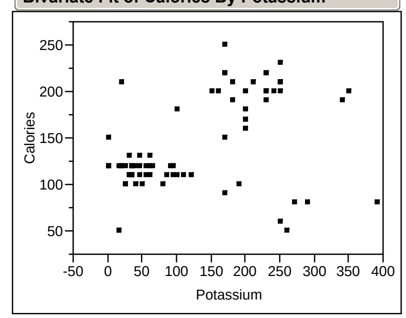
Bivariate Fit of Calories By Sugars



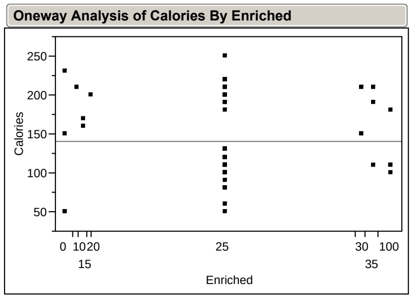
Bivariate Fit of Calories By Calories fr Fat



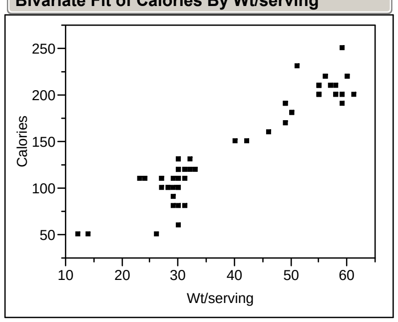
Bivariate Fit of Calories By Potassium



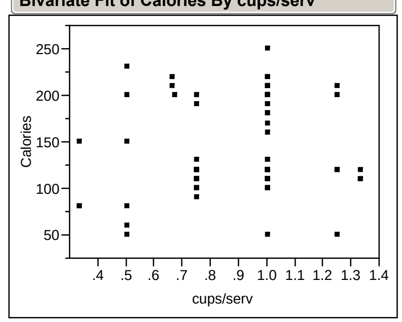
Oneway Analysis of Calories By Enriched



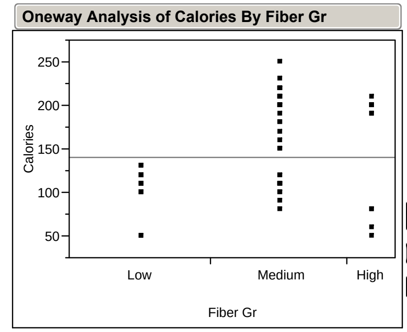
Bivariate Fit of Calories By Wt/serving



Bivariate Fit of Calories By cups/serv

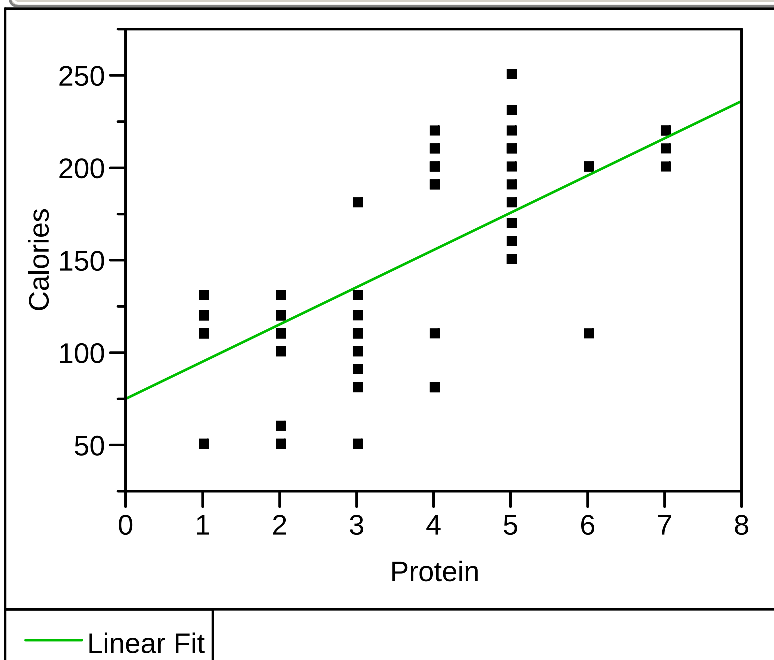


Oneway Analysis of Calories By Fiber Gr



Look Closer at Protein
Right-click title bar – Fit Line

Bivariate Fit of Calories By Protein



Linear Fit

Calories = 74.874142 + 20.200669 Protein

Summary of Fit

RSquare	0.495772
RSquare Adj	0.488958
Root Mean Square Error	35.46409
Mean of Response	140.5263
Observations (or Sum Wgts)	76

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	91509.03	91509.0	72.7589
Error	74	93069.92	1257.7	Prob > F
C. Total	75	184578.95		<.0001*

Parameter Estimates

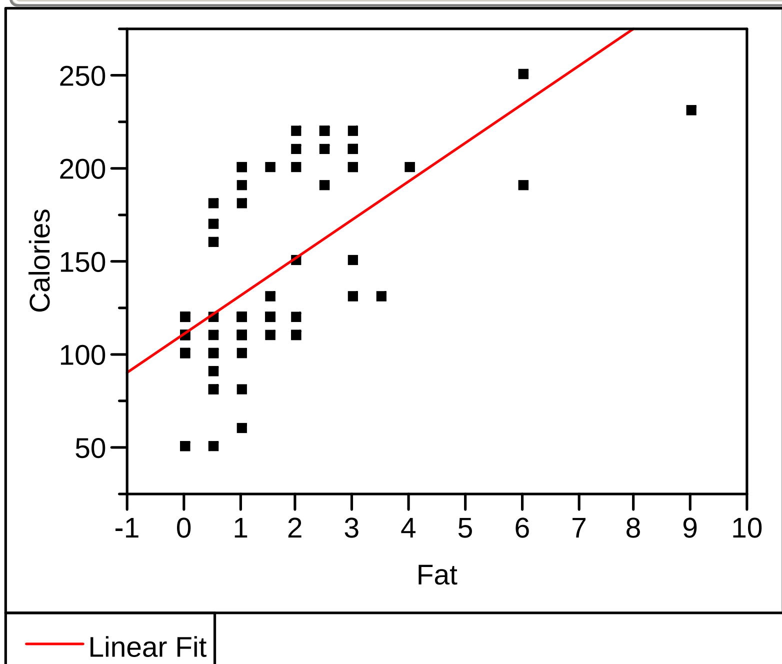
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	74.874142	8.705646	8.60	<.0001*
Protein	20.200669	2.368223	8.53	<.0001*

It appears that each g of protein adds 20.2 calories

Sandia National Laboratories

Look Closer at Fat
Right-click title bar – Fit Line

Bivariate Fit of Calories By Fat



Linear Fit

Calories = 110.77451 + 20.555796 Fat

It appears that each g of fat adds 20.6 calories

Summary of Fit

RSquare	0.4173
RSquare Adj	0.409425
Root Mean Square Error	38.12395
Mean of Response	140.5263
Observations (or Sum Wgts)	76

Analysis of Variance

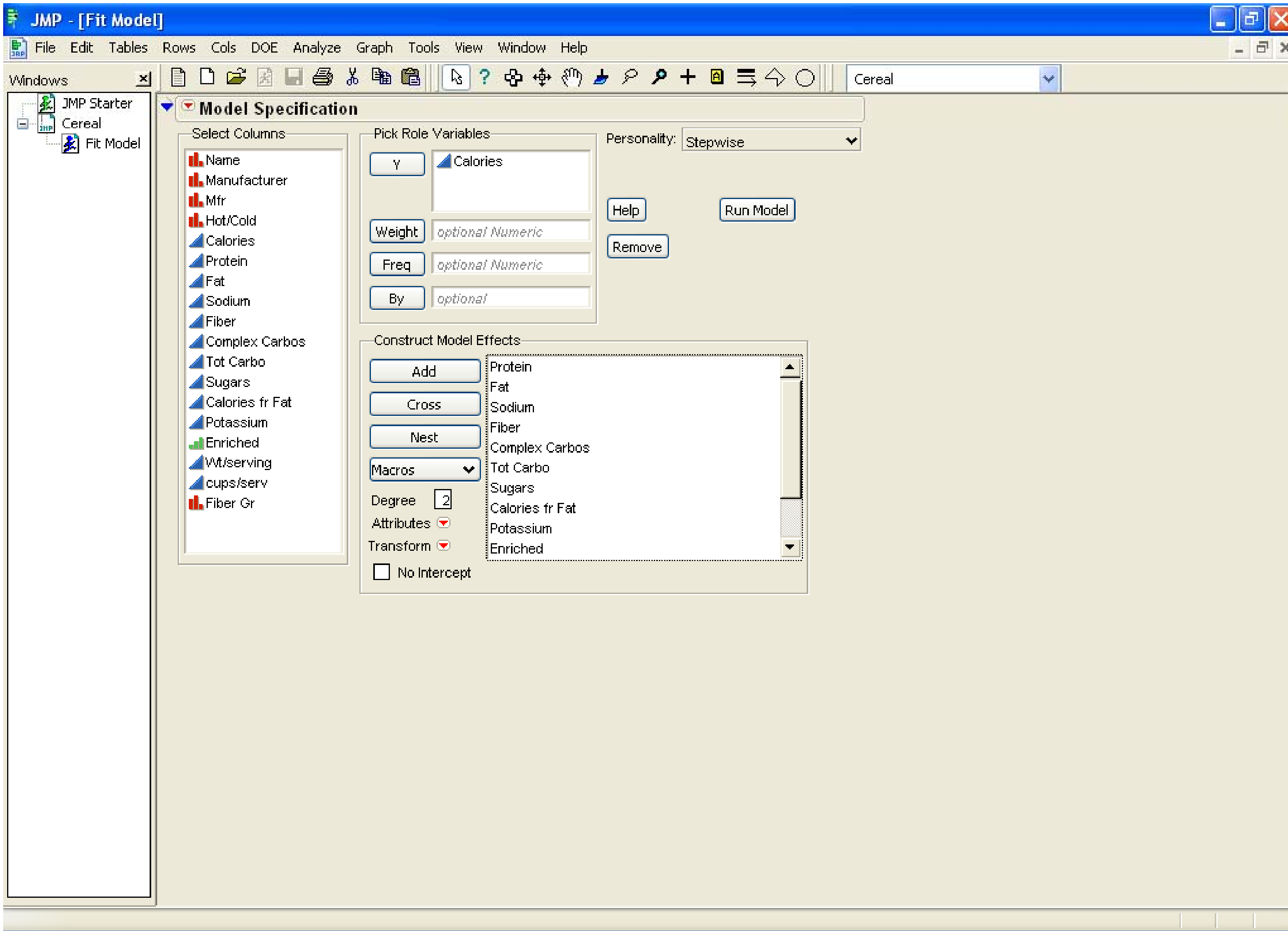
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	77024.73	77024.7	52.9949
Error	74	107554.22	1453.4	Prob > F
C. Total	75	184578.95		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	110.77451	5.985571	18.51	<.0001*
Fat	20.555796	2.82369	7.28	<.0001*

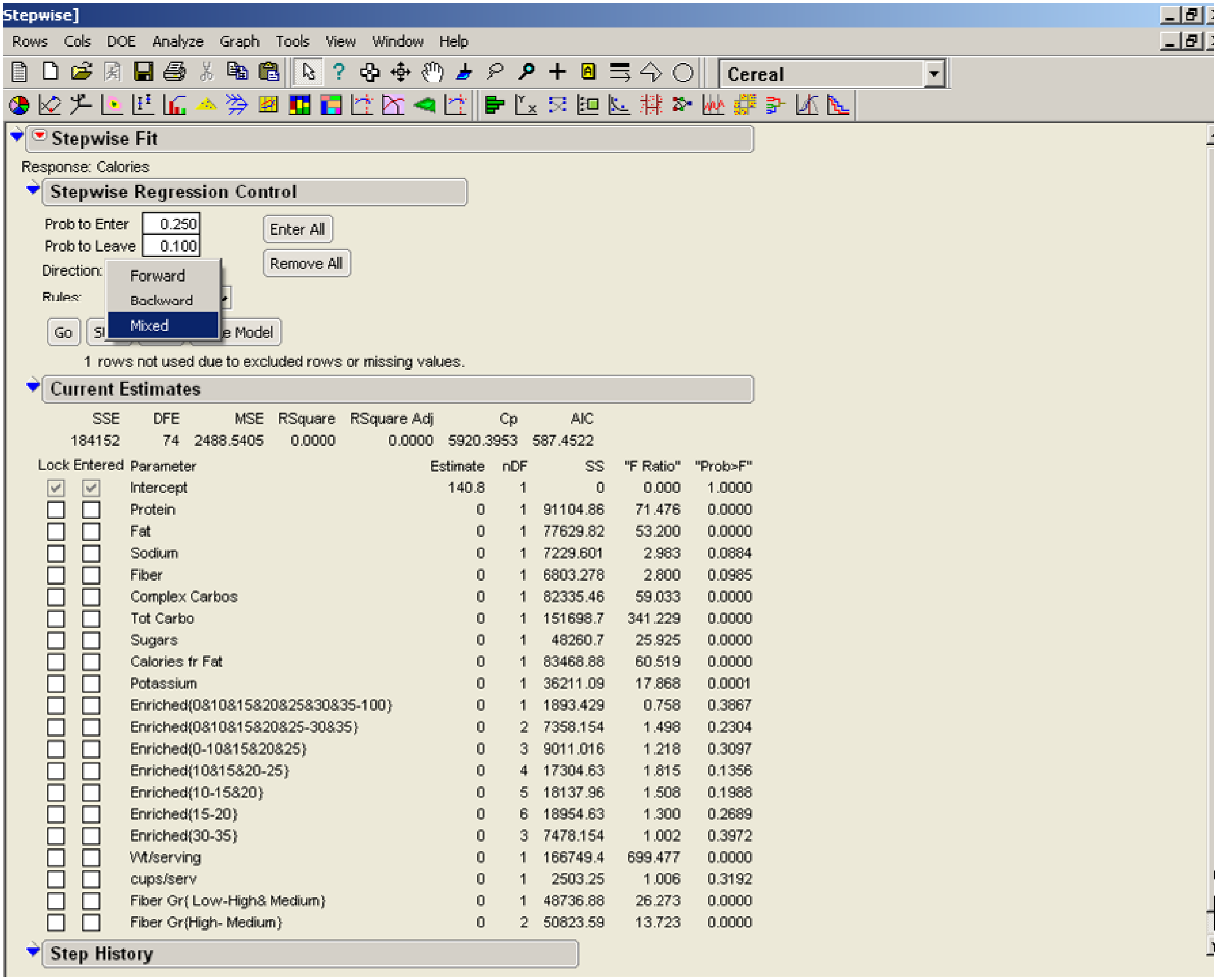
Sandia National Laboratories

Analyze → Fit Model
Calories in Y, all below that in X



andia
ational
boratories

Select Personality → Stepwise, Run Model
Change Direction to Mixed



Stepwise]

Rows Cols DOE Analyze Graph Tools View Window Help

Cereal

Stepwise Fit

Response: Calories

Stepwise Regression Control

Prob to Enter 0.250

Prob to Leave 0.100

Direction: Forward / Backward / Mixed

Rules:

Enter All

Remove All

Go

e Model

1 rows not used due to excluded rows or missing values.

Current Estimates

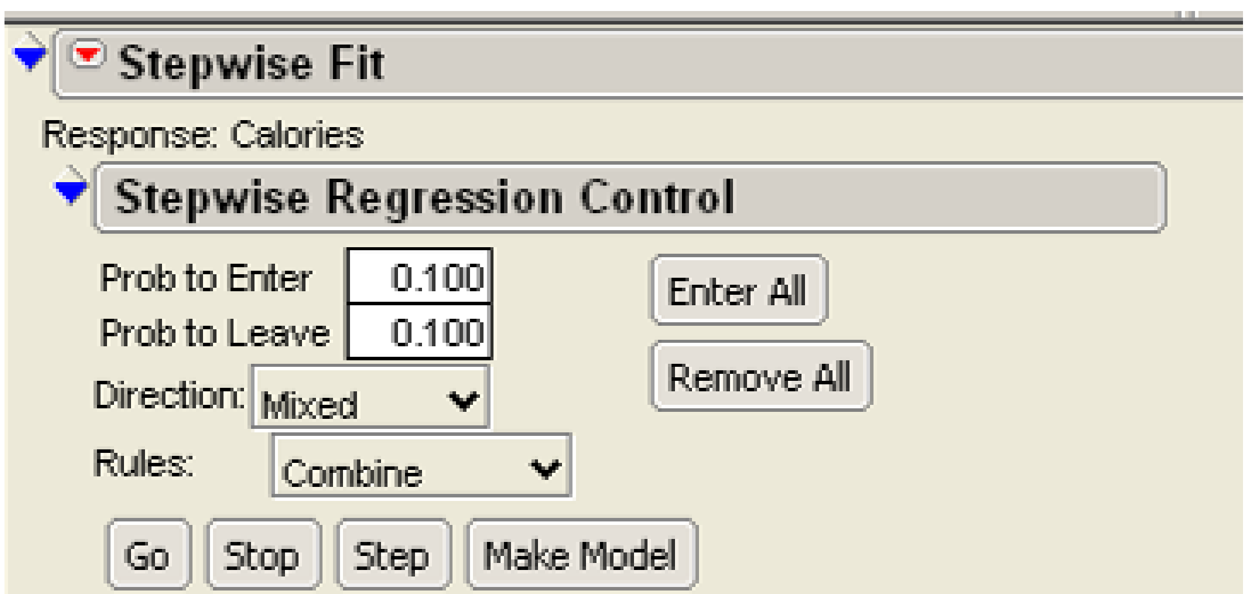
SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
184152	74	2488.5405	0.0000	0.0000	5920.3953	587.4522

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	140.8	1	0	0.000	1.0000
☐	☐	Protein	0	1	91104.86	71.476	0.0000
☐	☐	Fat	0	1	77629.82	53.200	0.0000
☐	☐	Sodium	0	1	7229.601	2.983	0.0884
☐	☐	Fiber	0	1	6803.278	2.800	0.0985
☐	☐	Complex Carbos	0	1	82335.46	59.033	0.0000
☐	☐	Tot Carbo	0	1	151698.7	341.229	0.0000
☐	☐	Sugars	0	1	48260.7	25.925	0.0000
☐	☐	Calories fr Fat	0	1	83468.88	60.519	0.0000
☐	☐	Potassium	0	1	36211.09	17.868	0.0001
☐	☐	Enriched{0&10&15&20&25&30&35-100}	0	1	1893.429	0.758	0.3867
☐	☐	Enriched{0&10&15&20&25-30&35}	0	2	7358.154	1.498	0.2304
☐	☐	Enriched{0-10&15&20&25}	0	3	9011.016	1.218	0.3097
☐	☐	Enriched{10&15&20-25}	0	4	17304.63	1.815	0.1356
☐	☐	Enriched{10-15&20}	0	5	18137.96	1.508	0.1988
☐	☐	Enriched{15-20}	0	6	18954.63	1.300	0.2689
☐	☐	Enriched{30-35}	0	3	7478.154	1.002	0.3972
☐	☐	Wt/serving	0	1	166749.4	699.477	0.0000
☐	☐	cups/serv	0	1	2503.25	1.006	0.3192
☐	☐	Fiber Gr{ Low-High& Medium}	0	1	48736.88	26.273	0.0000
☐	☐	Fiber Gr{High- Medium}	0	2	50823.59	13.723	0.0000

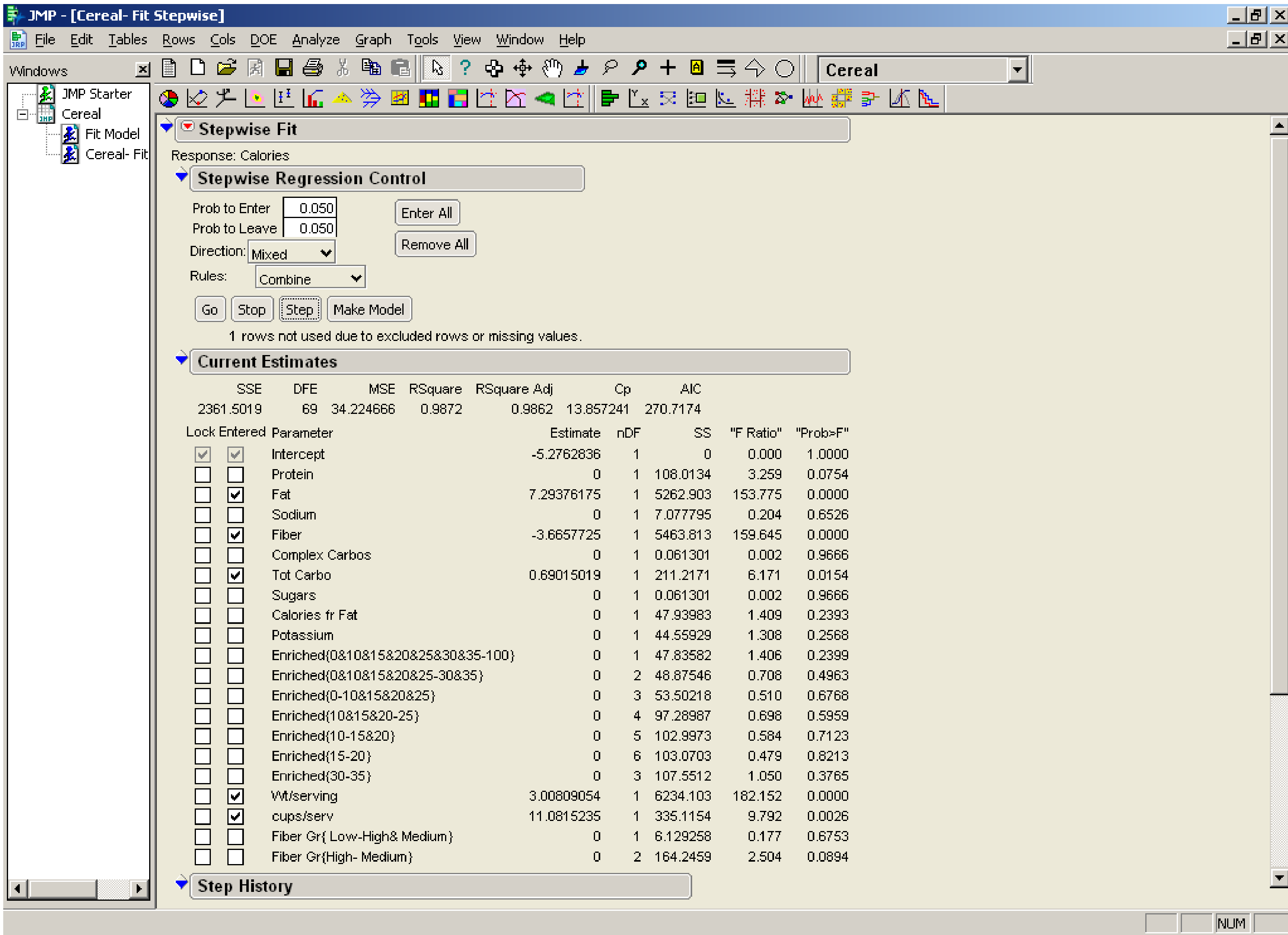
Step History

A Few Changes First

- You already changed direction to Mixed.
- Change "Prob to Enter" and "Prob to Leave" to 0.100



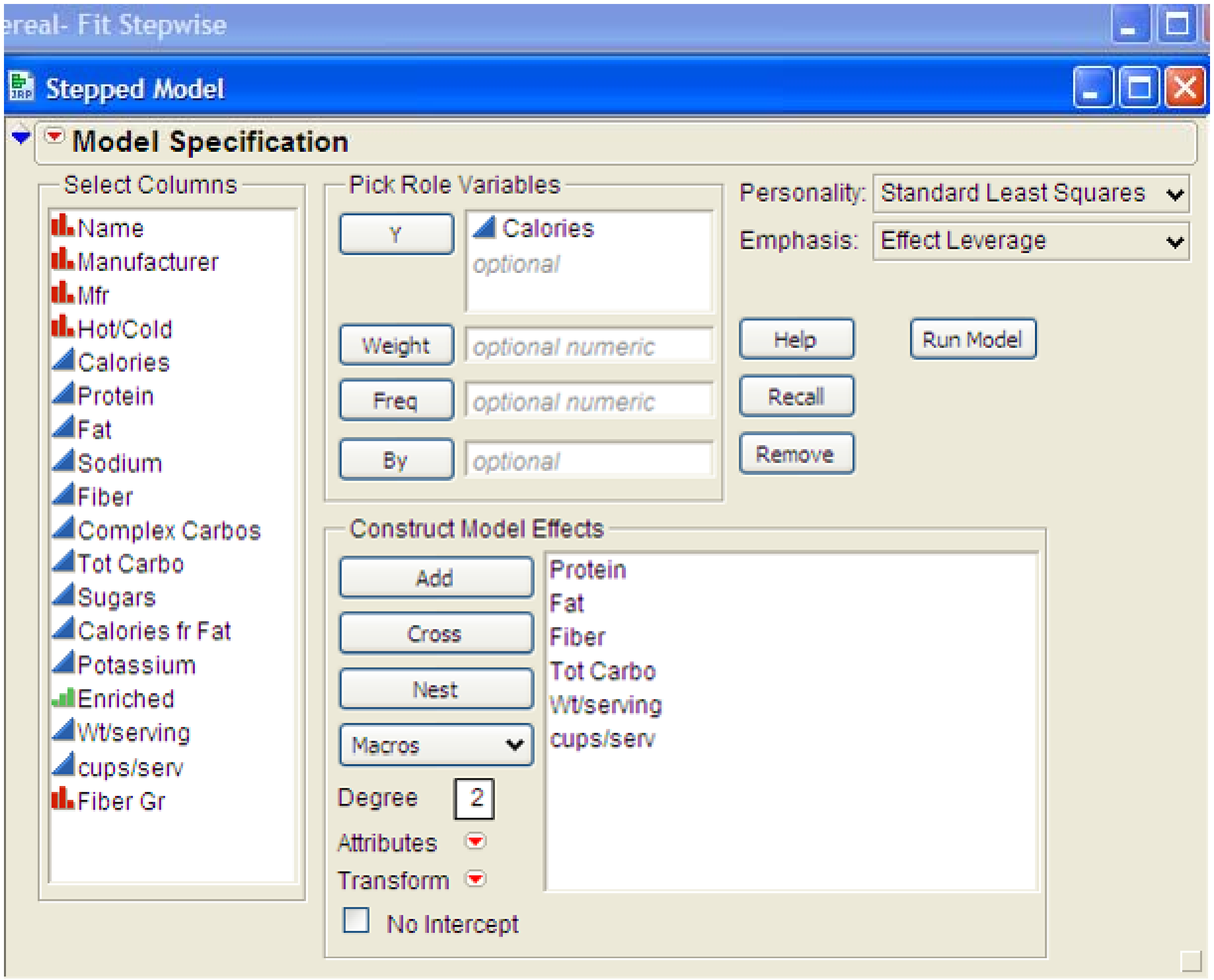
Click Step and Watch Factors Get Added to the Model



SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
2361.5019	69	34.224666	0.9872	0.9862	13.857241	270.7174

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	-5.2762836	1	0	0.000	1.0000
☐	☐	Protein	0	1	108.0134	3.259	0.0754
☐	☑	Fat	7.29376175	1	5262.903	153.775	0.0000
☐	☐	Sodium	0	1	7.077795	0.204	0.6526
☐	☑	Fiber	-3.6657725	1	5463.813	159.645	0.0000
☐	☐	Complex Carbos	0	1	0.061301	0.002	0.9666
☐	☑	Tot Carbo	0.69015019	1	211.2171	6.171	0.0154
☐	☐	Sugars	0	1	0.061301	0.002	0.9666
☐	☐	Calories fr Fat	0	1	47.93983	1.409	0.2393
☐	☐	Potassium	0	1	44.55929	1.308	0.2568
☐	☐	Enriched{0&10&15&20&25&30&35-100}	0	1	47.83582	1.406	0.2399
☐	☐	Enriched{0&10&15&20&25-30&35}	0	2	48.87546	0.708	0.4963
☐	☐	Enriched{0-10&15&20&25}	0	3	53.50218	0.510	0.6768
☐	☐	Enriched{10&15&20-25}	0	4	97.28987	0.698	0.5959
☐	☐	Enriched{10-15&20}	0	5	102.9973	0.584	0.7123
☐	☐	Enriched{15-20}	0	6	103.0703	0.479	0.8213
☐	☐	Enriched{30-35}	0	3	107.5512	1.050	0.3765
☐	☑	Wt/serving	3.00809054	1	6234.103	182.152	0.0000
☐	☑	cups/serv	11.0815235	1	335.1154	9.792	0.0026
☐	☐	Fiber Gr{ Low-High& Medium}	0	1	6.129258	0.177	0.6753
☐	☐	Fiber Gr{High- Medium}	0	2	164.2459	2.504	0.0894

Click Make Model

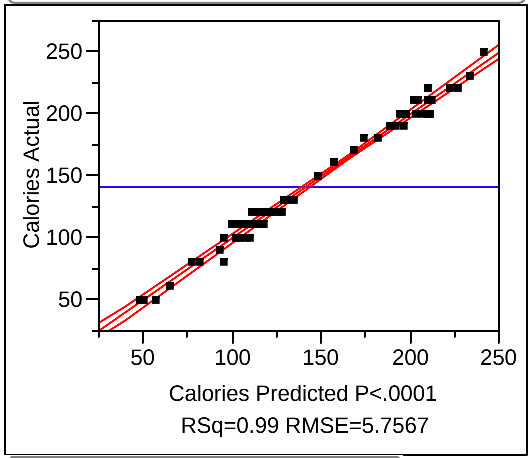


Sandia National Laboratories

Click Run Model

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.987763
RSquare Adj	0.986683
Root Mean Square Error	5.756695
Mean of Response	140.8
Observations (or Sum Wgts)	75

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	181898.51	30316.4	914.8112
Error	68	2253.49	33.1	Prob > F
C. Total	74	184152.00		<.0001*

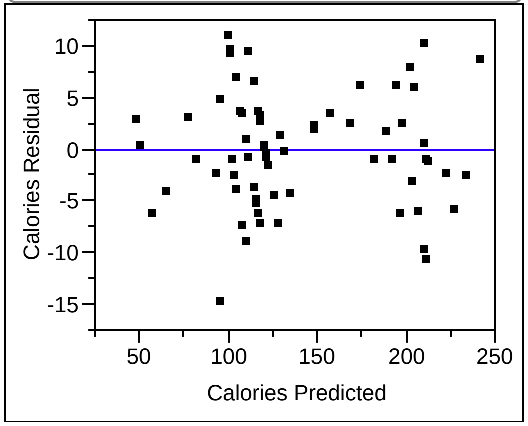
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	65	2253.4886	34.6691	.
Pure Error	3	0.0000	0.0000	Prob > F
Total Error	68	2253.4886		.
				Max RSq
				1.0000

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-4.230658	3.478264	-1.22	0.2281
Protein	1.3019557	0.721158	1.81	0.0754
Fat	7.414865	0.582652	12.73	<.0001*
Fiber	-3.874463	0.308004	-12.58	<.0001*
Tot Carbo	0.8724868	0.291431	2.99	0.0038*
Wt/serving	2.755463	0.260157	10.59	<.0001*
cups/serv	10.097625	3.527139	2.86	0.0056*

Residual by Predicted Plot



andia
ational
iboratories

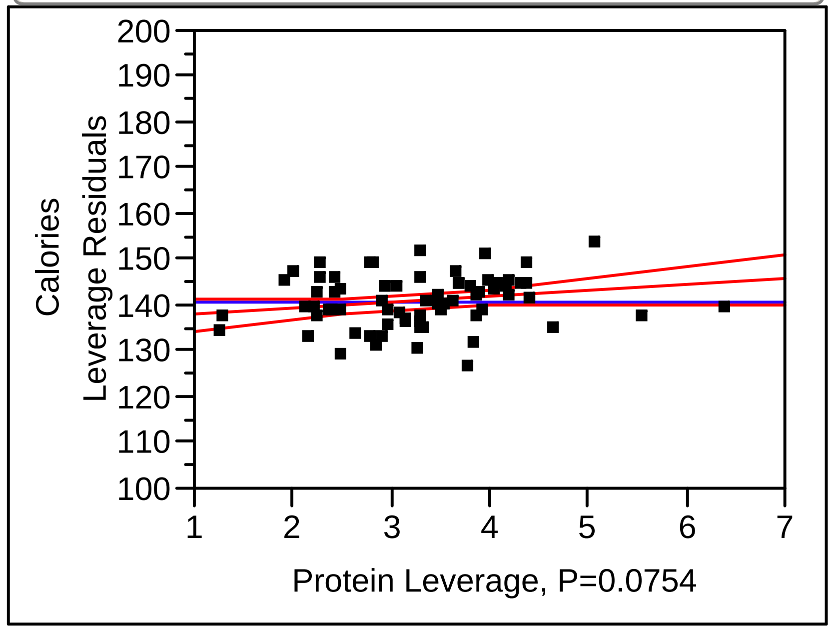
What is the Effect of Protein Now?

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-4.230658	3.478264	-1.22	0.2281
Protein	1.3019557	0.721158	1.81	0.0754
Fat	7.414865	0.582652	12.73	<.0001*
Fiber	-3.874463	0.308004	-12.58	<.0001*
Tot Carbo	0.8724868	0.291431	2.99	0.0038*
Wt/serving	2.755463	0.260157	10.59	<.0001*
cups/serv	10.097625	3.527139	2.86	0.0056*

Protein

Leverage Plot



Recall that it was 20.2 calories per g when we just looked at calories vs. protein. Nutritionists tell us that the real number is 4 calories per g of protein.

Sandia National Laboratories

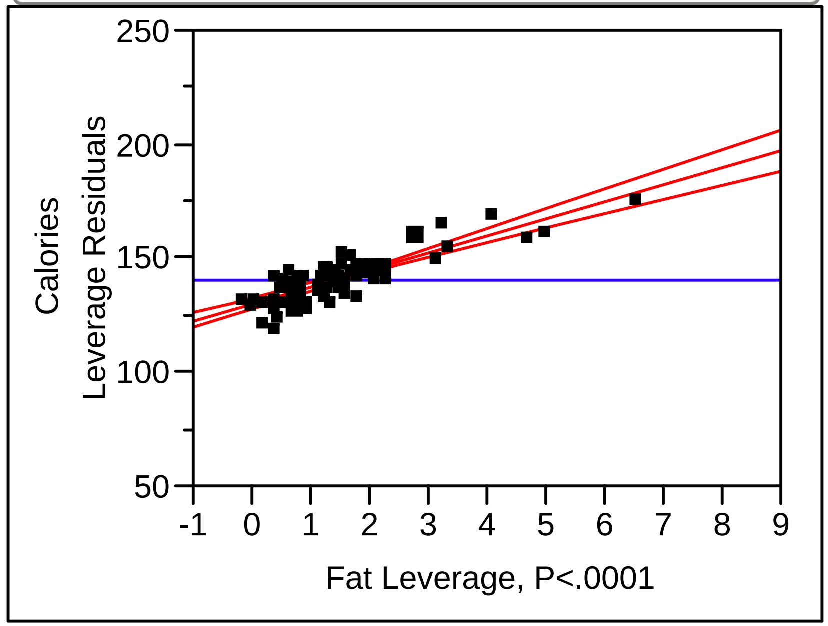
What is the Effect of Fat Now?

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-4.230658	3.478264	-1.22	0.2281
Protein	1.3019557	0.721158	1.81	0.0754
Fat	7.414865	0.582652	12.73	<.0001*
Fiber	-3.874463	0.308004	-12.58	<.0001*
Tot Carbo	0.8724868	0.291431	2.99	0.0038*
Wt/serving	2.755463	0.260157	10.59	<.0001*
cups/serv	10.097625	3.527139	2.86	0.0056*

Fat

Leverage Plot



Recall that it was 20.6 calories per g when we just looked at calories vs. fat. Nutritionists tell us that the real number is 9 calories per g of fat.

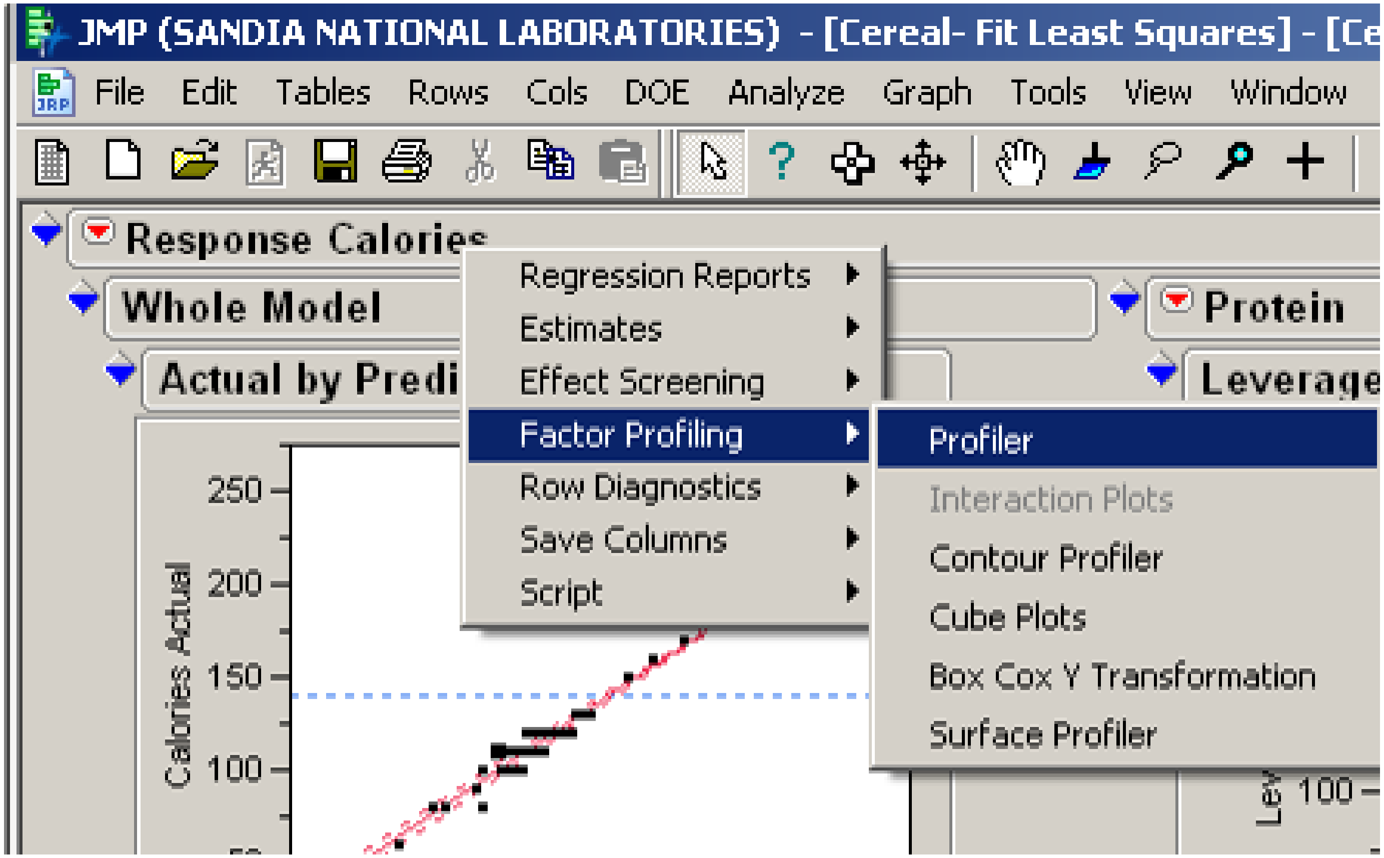
The Model is Revealing

Calories per gram	Individual Fit Y by X	Fit Model	Nutritionists[1]
Protein	20.2	1.3	4
Fat	20.6	7.4	9

[1] http://www.nutristrategy.com/nutrition/calories.htm

Sandia National Laboratories

Right-click Response Calories title bar
Select Factor Profiling → Profiler

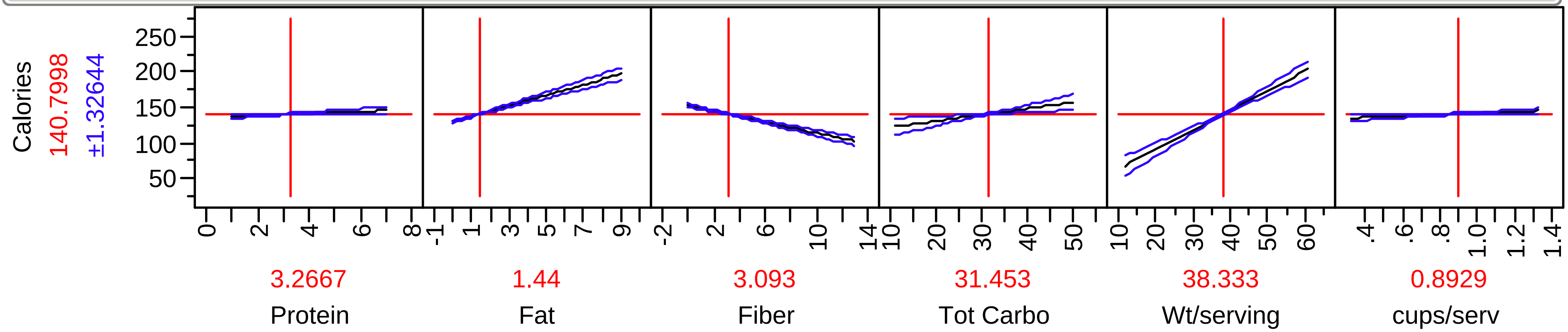


Scroll Down to Prediction Profiler

Try moving the vertical lines.

Which factors could you change in order to reduce calories?

Prediction Profiler



Sandia National Laboratories

Back to the House Exercise

You want to sell your house. It has the following features:

- **2000 square feet**
- **0.2 acre lot**
- **2 years old**
- **3 bedrooms**
- **3 full bathrooms**

You Want to Sell Your House

- Your real estate agent pulls up the set of data for recent home sales in your zip code, and tells you the average selling price was $124.36 per square foot.
- Your real estate agent breaks out the calculator and tells you your home is worth $\$124.36/\text{ft}^2 \times 2{,}000\ \text{ft}^2 = \$248{,}720$.
- Your real estate agent tells you to list your house for $260,000. "That leaves a little room for negotiating," they explain.
- You're just about to sign the listing paperwork, but you remember the modeling you just learned using cereal data

Should You Listen to Your real estate agent? Exercise

- Create a model for home price, including only significant factors.
- Determine the value of your home based on the model.
- Capture the students' listing prices on the board.
- Are these much different than what your real estate agent recommended?

Solutions

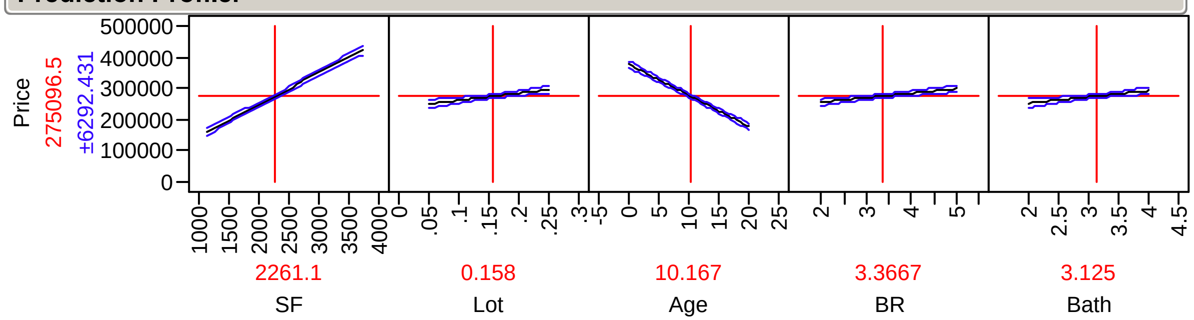
Create a model for home price, including only significant factors.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3242.1849	28037.38	0.12	0.9089
SF	100.26837	4.377378	22.91	<.0001*
Lot	228519.2	55577.05	4.11	0.0004*
Age	-9954.605	456.565	-21.80	<.0001*
BR	14362.019	2925.965	4.91	<.0001*
Bath	19803.935	5364.676	3.69	0.0011*

Prediction Expression

Prediction Profiler



What Does the Model Tell You?

- Which factors are statistically significant?
- What are the coefficients for these factors?
- In particular, what is the coefficient for $/square foot?

Solutions

Determine the value of your home based on the model.

$$3242.18 + 100.27*(2000) + 228519.20*(0.2) - 9954.60*(2) + 14362.02*(3) + 19803.93*(3) =$$

$332,075

Should you listen to your real estate agent and list your house for $260,000?

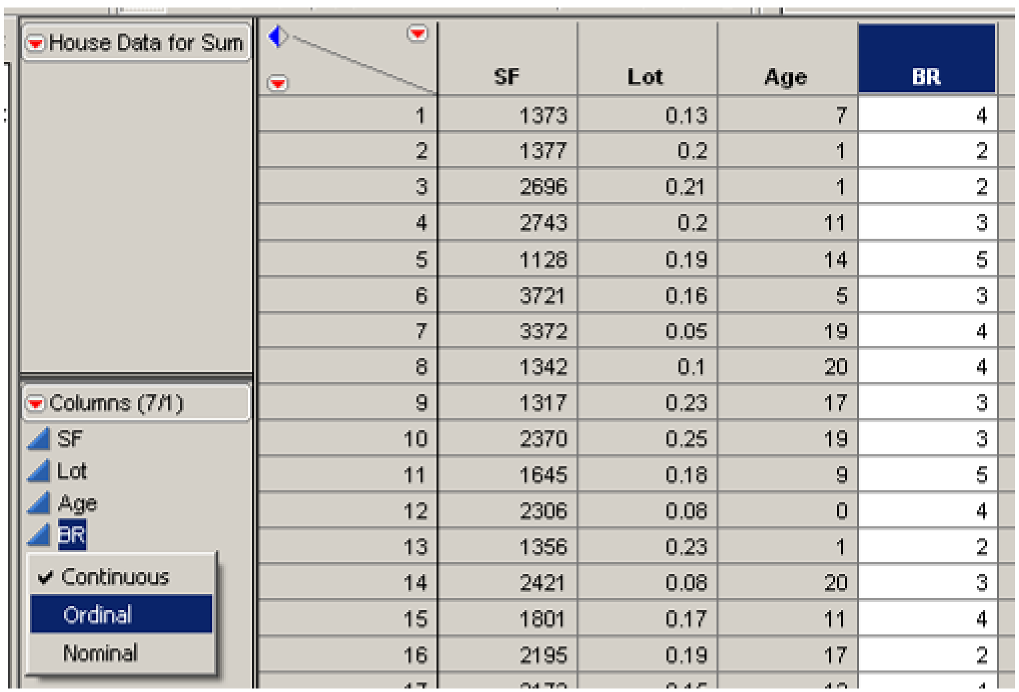
Sandia National Laboratories

A Different Approach

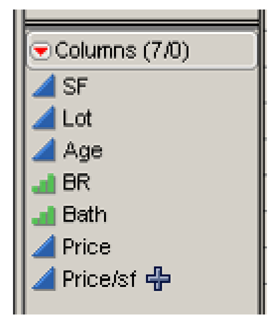
- Three types of variables
  - Continuous
    - » Time
    - » Distance
  - Ordinal
    - » Character data with an order (poor, fair, good, better, best)
    - » Numerical data with unequal spacing (4 = strongly agree, 3 = agree, 2 = disagree, 1 = strongly disagree)
  - Nominal
    - » Character data with no specific order (green, blue, yellow)
    - » Numerical data with no specific order (NASCAR car #)
- Should BR and Bath be treated as continuous variables?
- What if we had treated them as Ordinal Variables?

Treating BR and Bath as Ordinal

- **If Time Permits, change BR and Bath to Ordinal and redo the analysis**

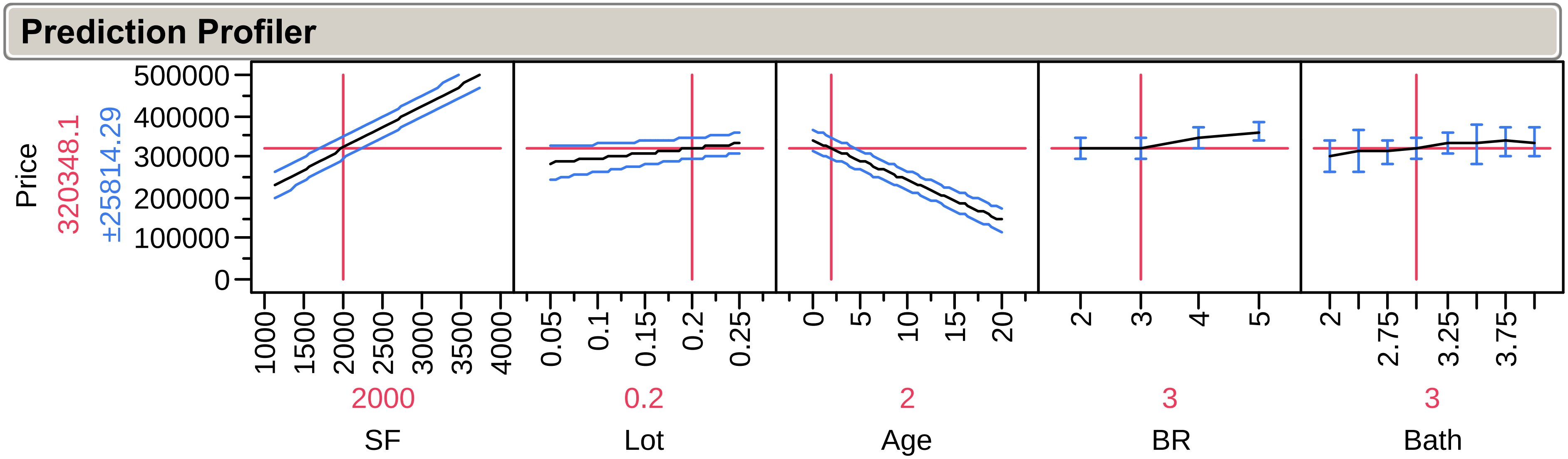


	SF	Lot	Age	BR
1	1373	0.13	7	4
2	1377	0.2	1	2
3	2696	0.21	1	2
4	2743	0.2	11	3
5	1128	0.19	14	5
6	3721	0.16	5	3
7	3372	0.05	19	4
8	1342	0.1	20	4
9	1317	0.23	17	3
10	2370	0.25	19	3
11	1645	0.18	9	5
12	2306	0.08	0	4
13	1356	0.23	1	2
14	2421	0.08	20	3
15	1801	0.17	11	4
16	2195	0.19	17	2



Sandia National Laboratories

Use Prediction Profiler

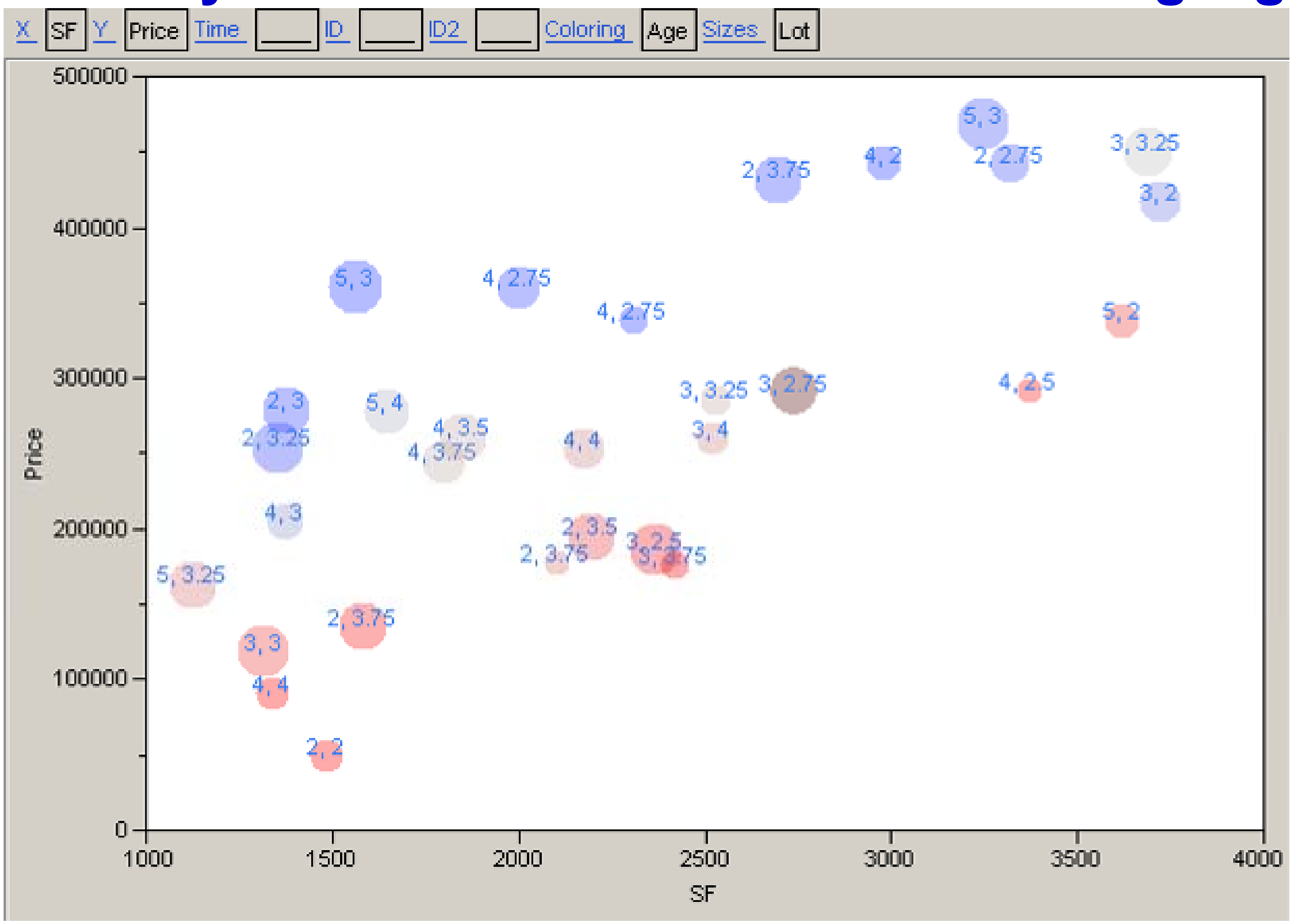


What is the predicted price now?

Sandia National Laboratories

One More Tip: Visually Display your Data!

How many dimensions are shown in this single graph?



Sandia National Laboratories

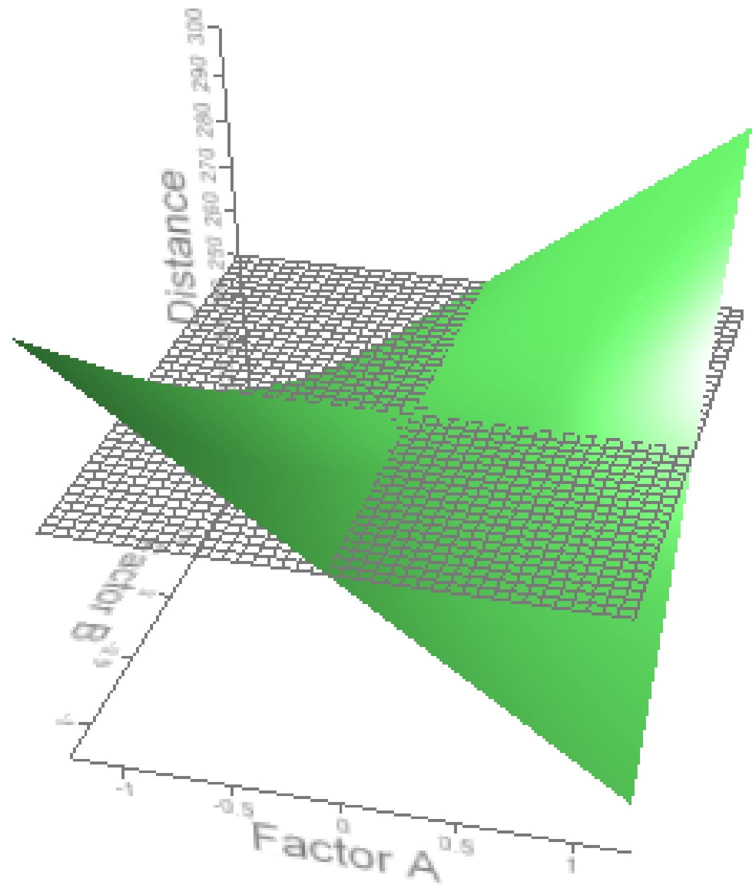
Practical Application of Design of Experiments: Optimize the Golf Drive

- **Simple Golf Example**
  - **We want to increase the distance of our golf drive**
  - **We suspect changing distance from the ball and right hand position may be factors**
  - **How would you approach increasing drive distance by varying these factors?**

Design of Experiments vs. Traditional Approach

- Traditional Approach
  - Vary one factor at a time
  - Look for changes
  - Experimenters call this an OFAT experiment (One Factor At a Time)
- Problems with Traditional Approach
  - Does not catch interactions
  - Requires multiple experiments (one for each factor)

Suppose our Distance vs. Two Factors Looks Like This



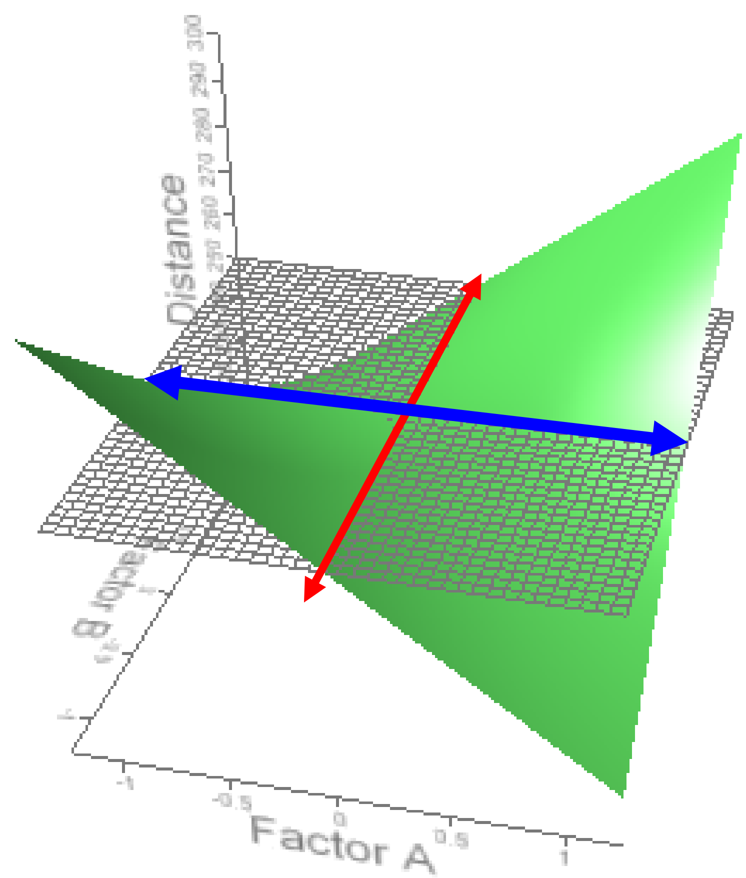
JMP Document

The Traditional Approach

- **Vary "distance from ball" in one experiment (blue line)**

- **Vary "right hand position" in another experiment (red line)**

- **Our conclusion would incorrectly be, "Neither factor affects distance."**



Is There a Better Way?

- **A Designed Experiment would Change Both Factors Simultaneously**
- **Example: 2-factor, 2-level Full Factorial**
  - **There are 2 factors at two levels, or 2^2 combinations**

	Distance from ball	Right hand
1	Close	Weak
2	Close	Strong
3	Far	Weak
4	Far	Strong

Sandia National Laboratories

Using DOE on the Golf Swing

- I've taken various golf lessons throughout the past 10 years

- Once I learned DOE, I quickly realized I was a victim of the OFAT approach to experimental design during these lessons

- Some of the factors instructors typically vary:
  - Right Hand Position (Weak to Strong)
  - Stance Width
  - Distance to Ball (Reach)
  - Ball Forward / Backward in Stance

A Better Way

- As instructors identified the "optimum" for a particular factor, they found that I had to readjust the other factors as well, to compensate for the change in the one factor.
- This told me that I had *interactions* present.
- This was a great opportunity to apply DOE.
- Disclaimer: I am not very good at golf. Sample video:

Designing The Experiment

- Used Custom Design Response Surface Methodology (RSM)
- Three Replicates and Four Center Points
- Results in 80 runs (80 balls)

My Design

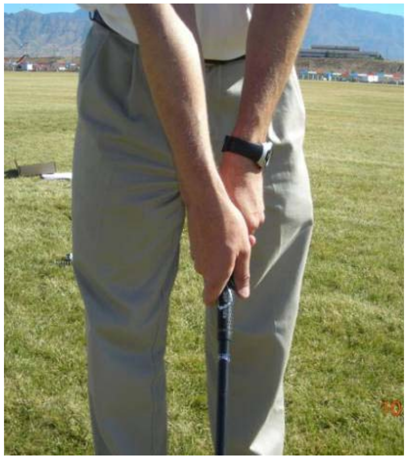
Four Factors at Three Levels Each

- Right Hand
  - » Weak
  - » Neutral
  - » Strong
- Stance Width
  - » Narrow
  - » Normal
  - » Wide
- Distance to Ball
  - » Close
  - » Middle
  - » Far
- Ball in Stance
  - » Back
  - » Middle
  - » Forward

JMP Data Table

Sandia National Laboratories

Right Hand Settings



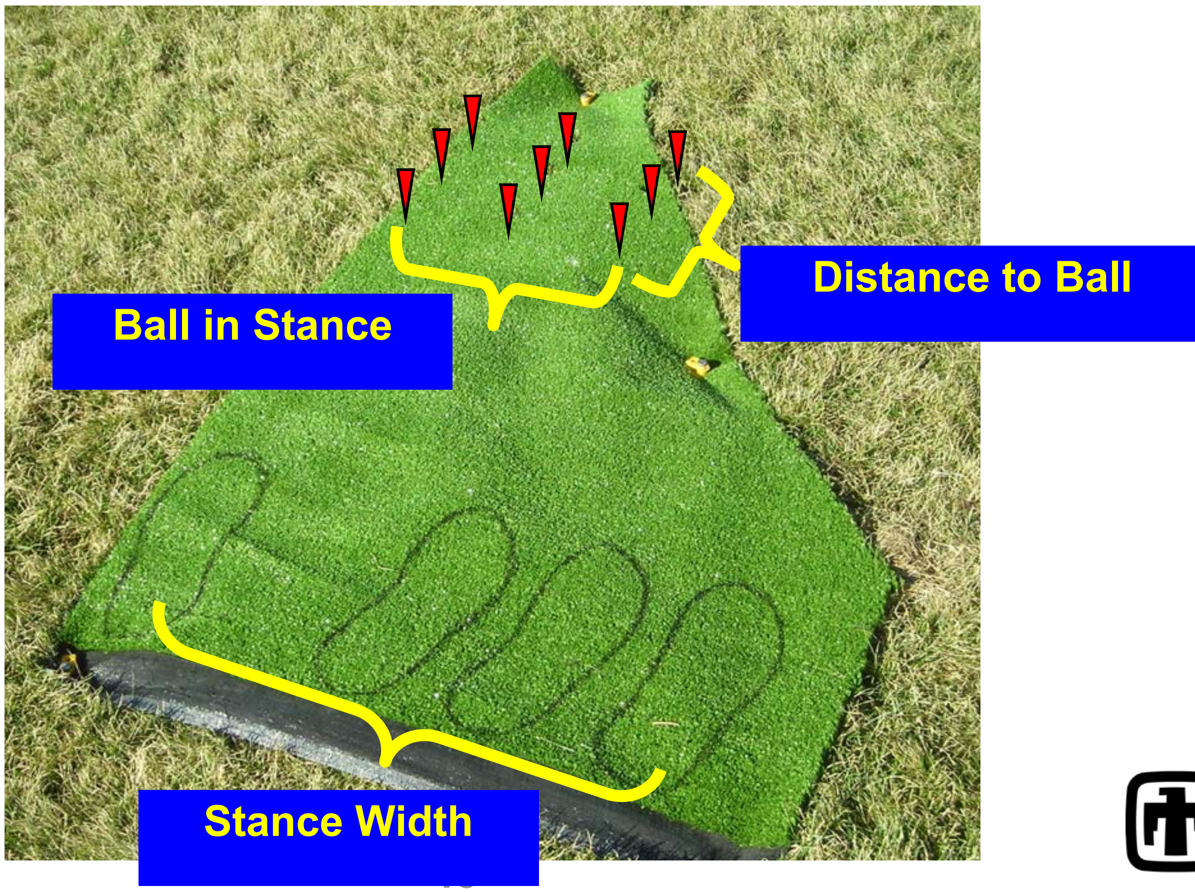
Weak

Neutral

Strong

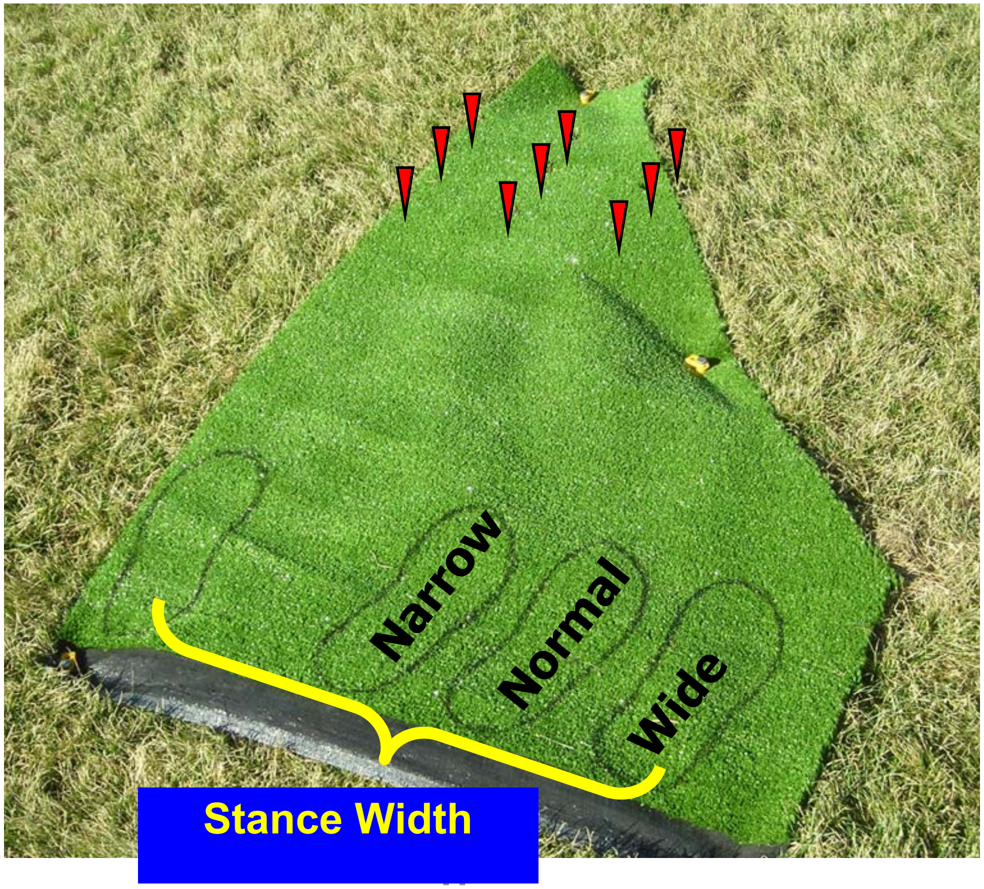
The Other Factors

Used a mat as a template



The Other Factors

Stance Width

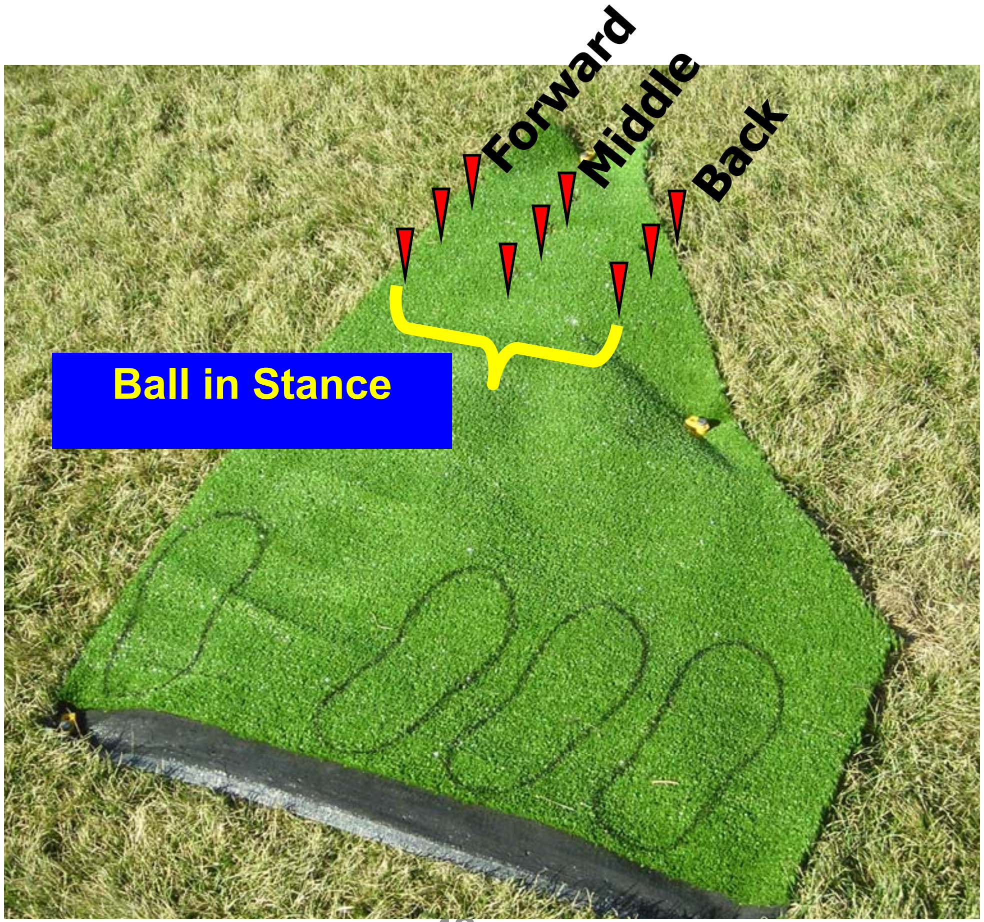


The Other Factors

Distance to Ball

The Other Factors

Ball in Stance

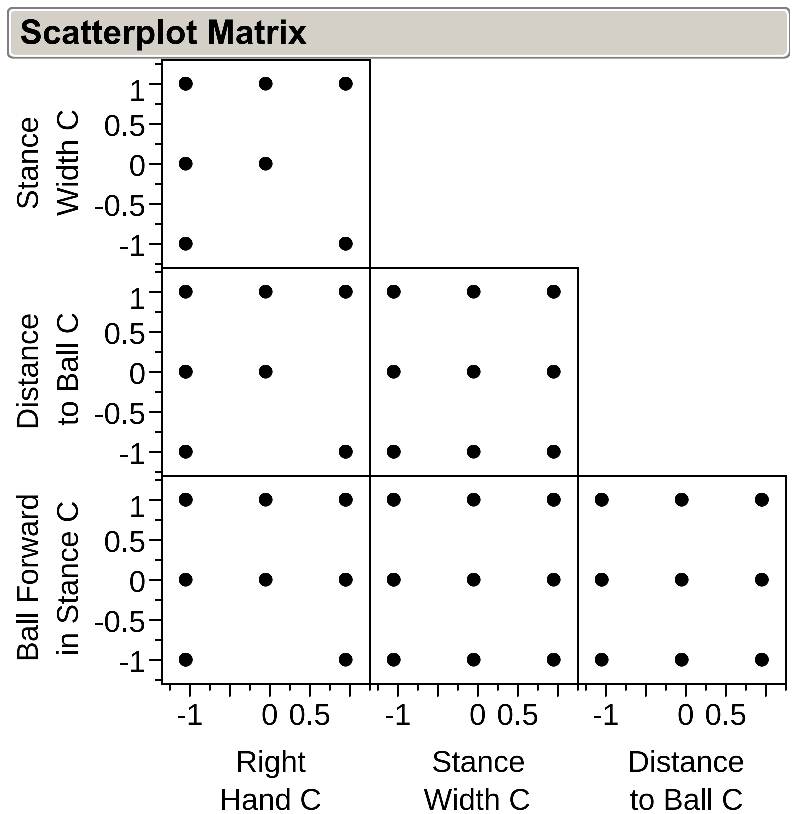


Experiment Details

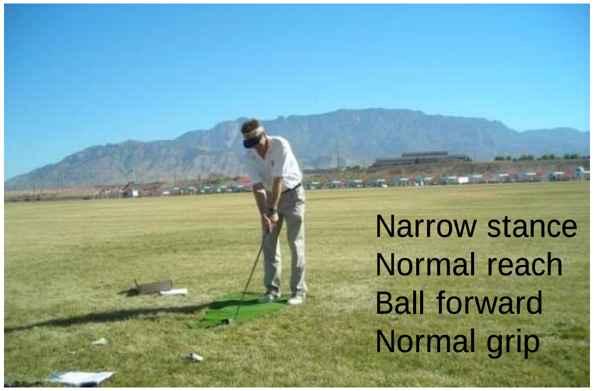
Settings for the First 20 Balls

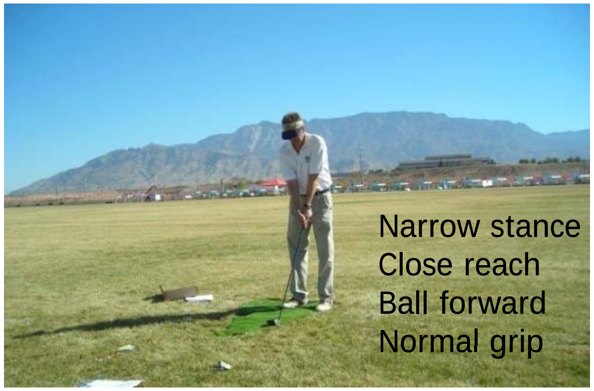
Ball	Right Hand	Stance Width	Distance to Ball	Ball in Stance
1	Neutral	Normal	Middle	Middle
2	Neutral	Normal	Middle	Middle
3	Neutral	Normal	Middle	Middle
4	Neutral	Wide	Far	Forward
5	Neutral	Normal	Middle	Middle
6	Weak	Narrow	Middle	Back
7	Weak	Wide	Far	Back
8	Strong	Wide	Far	Middle
9	Weak	Wide	Close	Middle
10	Strong	Narrow	Far	Forward
11	Weak	Normal	Close	Back
12	Strong	Narrow	Far	Forward
13	Weak	Narrow	Far	Middle
14	Strong	Wide	Far	Middle
15	Neutral	Normal	Middle	Middle
16	Weak	Wide	Middle	Forward
17	Strong	Narrow	Close	Middle
18	Weak	Wide	Far	Back
19	Weak	Normal	Far	Forward
20	Weak	Wide	Middle	Forward

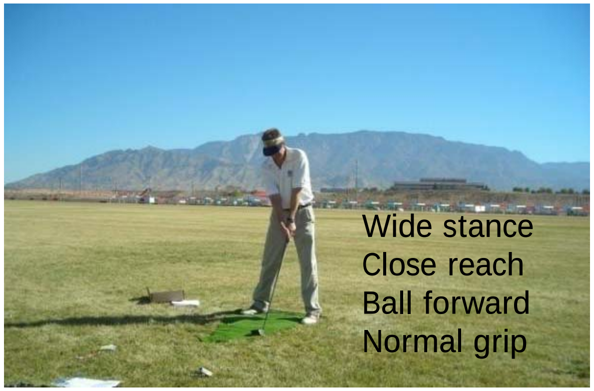
The Design Space

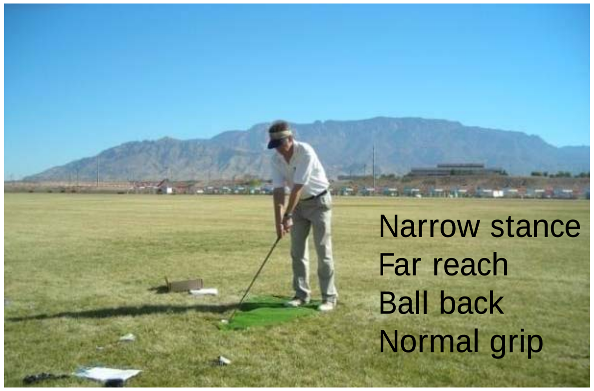


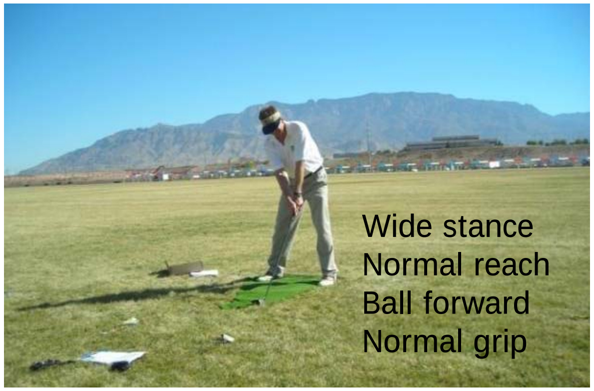
Sample of Some of the "Extreme" Set-ups











Sandia National Laboratories

Experiment Location

Albuquerque International Balloon Fiesta Park

Very heavy rough (little or no roll)

The Methodology

- Number 80 balls
- Hit balls in randomized order
- Track ball location using GPS
  - GPS receiver on laptop
  - Accurate to within 5 feet
- Convert GPS coordinates to distance and angle

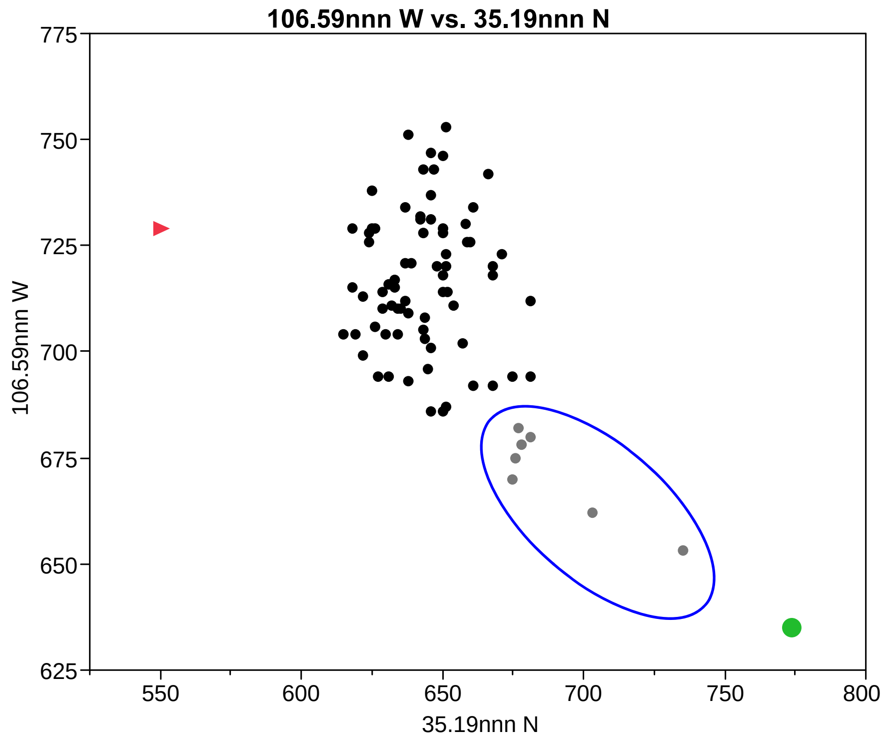
Sandia National Laboratories

What about Outliers?

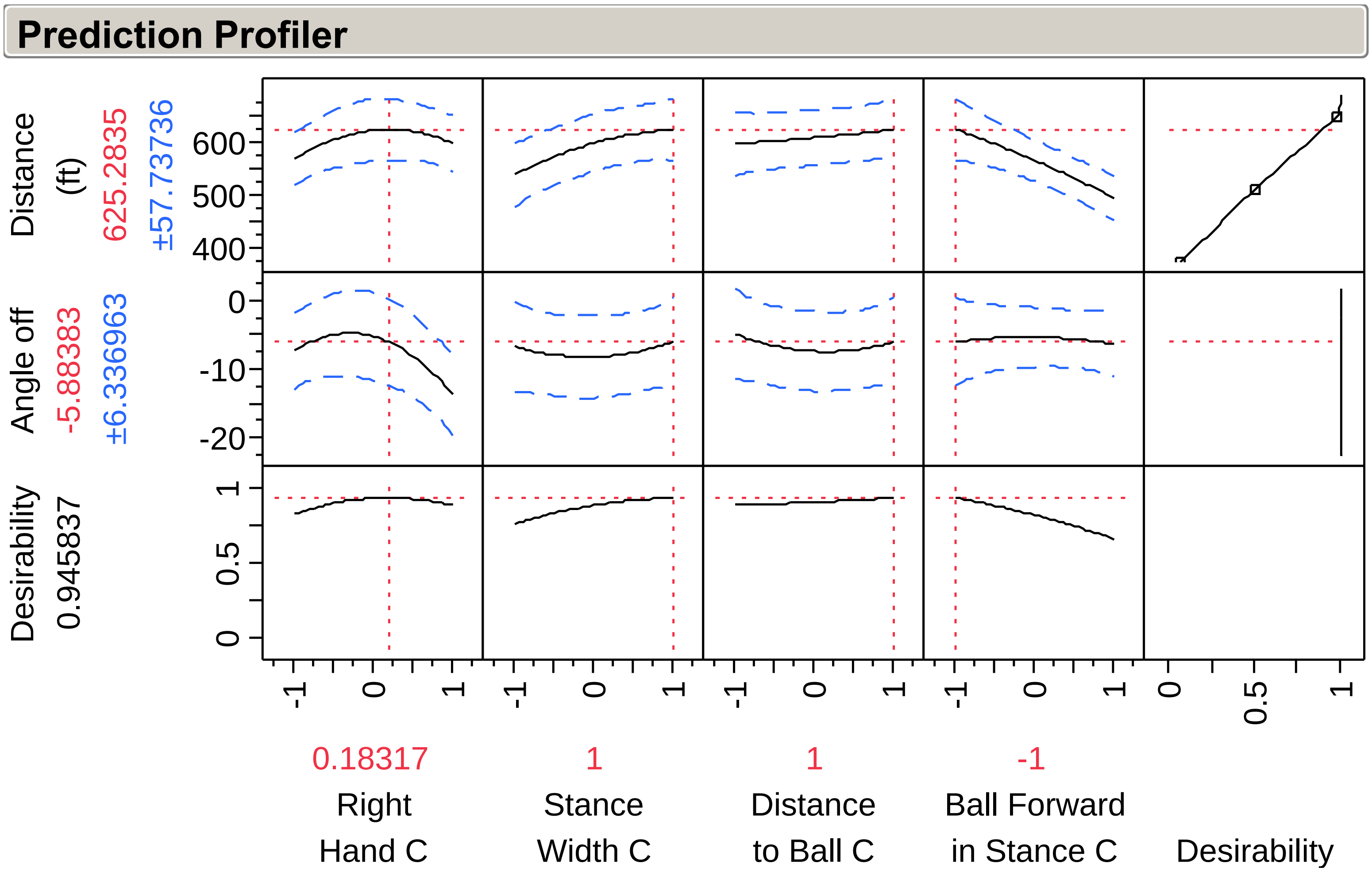
- Complete shanks were rerun at the end (still random)
- These were expected … DOEs should start by pushing variables to the extremes
- For example, hitting a ball with a narrow stance, far reach, and the ball back in your stance is tough to do for an amateur like me

What about Outliers?

- Even after Re-hitting Shanks, there were Still Some Outliers
- I Excluded These because they were “Semi-shanks”



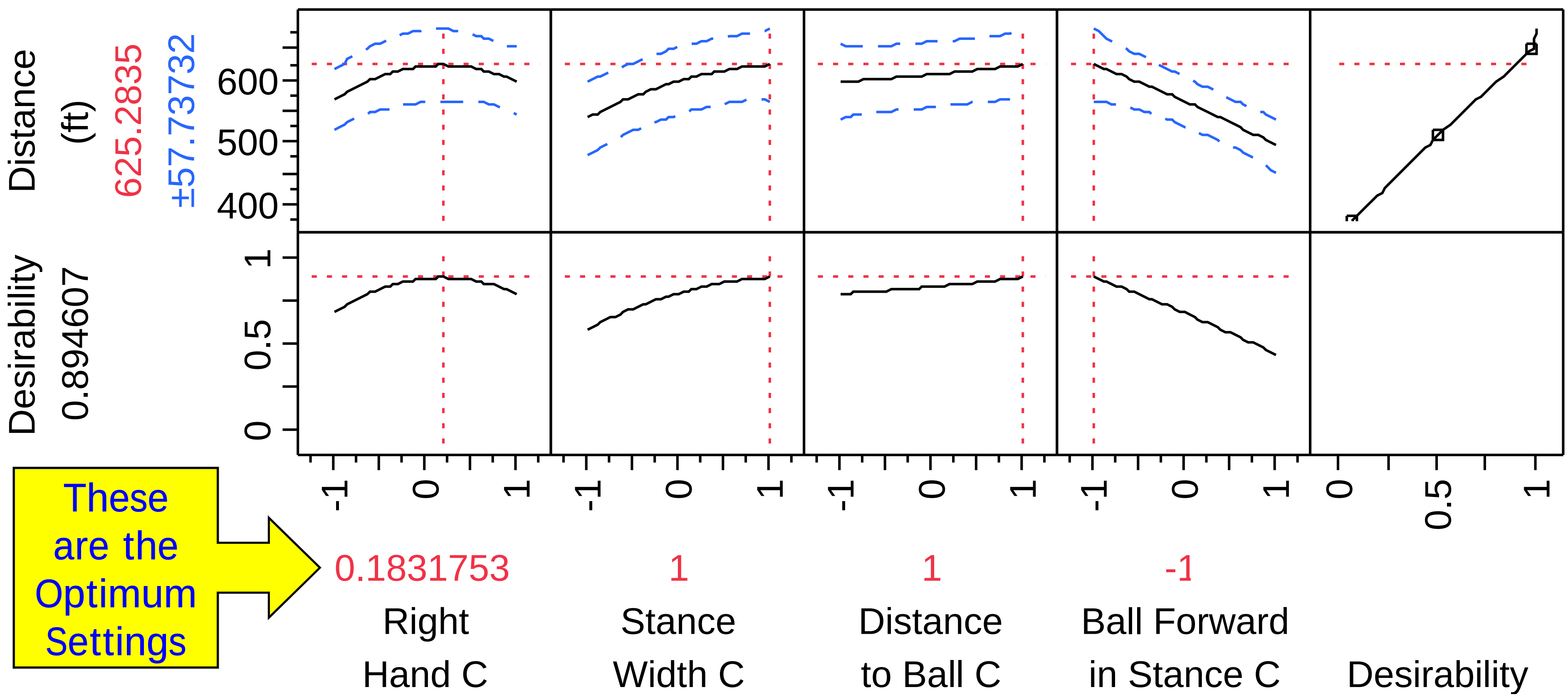
Results (Excluding Seven Outliers)



Sandia National Laboratories

Taking a Closer Look

Prediction Profiler



These are the Optimum Settings

Slightly Strong Right Hand

Wide Stance

Doesn't Matter Much

Ball Back in Stance

Sandia National Laboratories

Optimum Settings for Me

Right Hand = 0.18
- » Weak = -1
- » Neutral = 0
- » Strong = 1

Stance Width = 1
- » Narrow = -1
- » Normal = 0
- » Wide = 1

Distance to Ball = 1
- » Close = -1
- » Middle = 0
- » Far = 1

Ball in Stance = -1
- » Back = -1
- » Middle = 0
- » Forward = 1

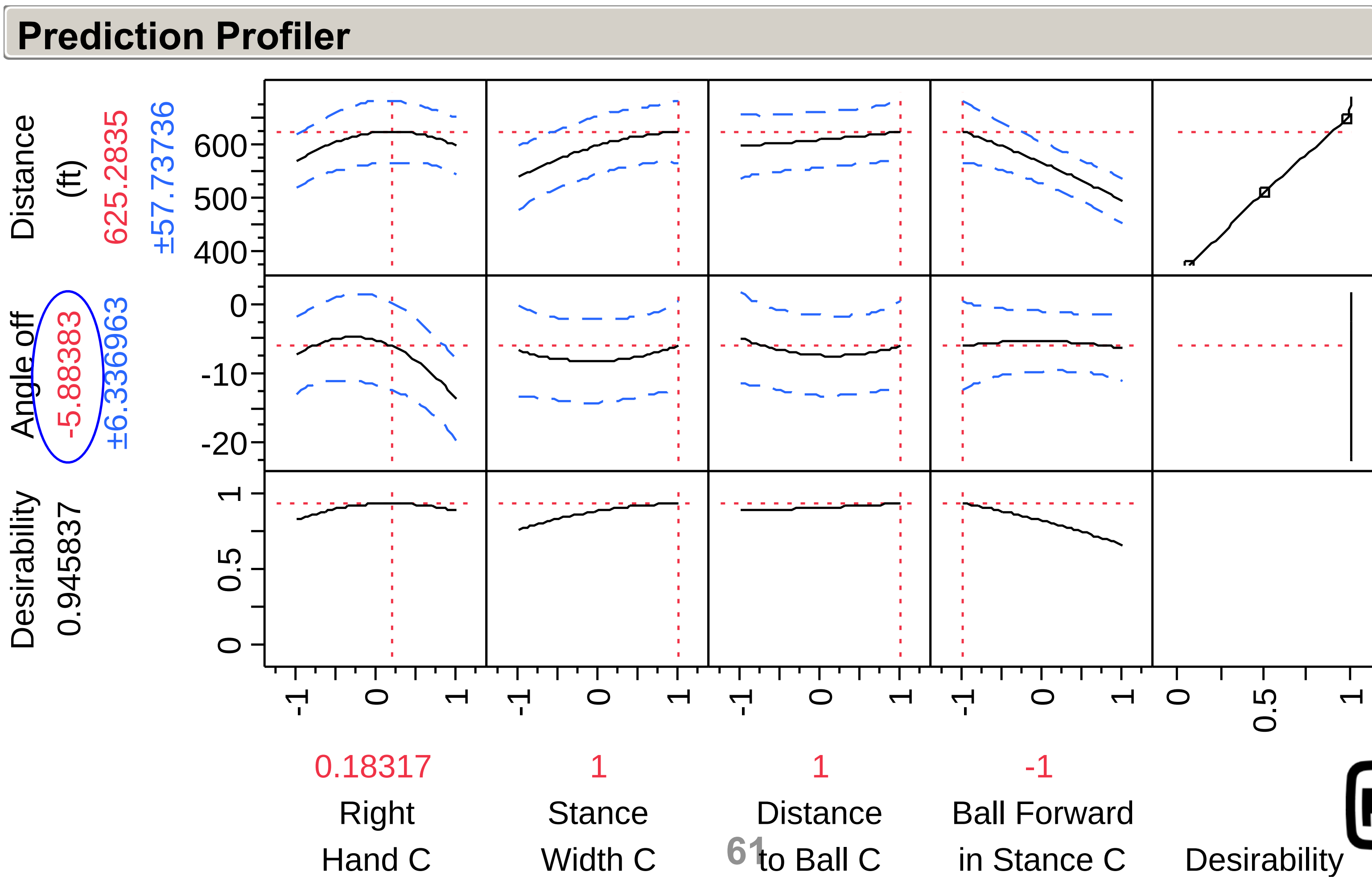
What about Angle?

- In addition to learning what settings are optimum, I also learned what angle I can expect the ball to fly

- My GPS data allowed me to calculate angle data (in addition to distance data)

What about Angle?

- An "Aha!" moment for me: I need to align my feet 6 degrees left of my target



What about Angle?

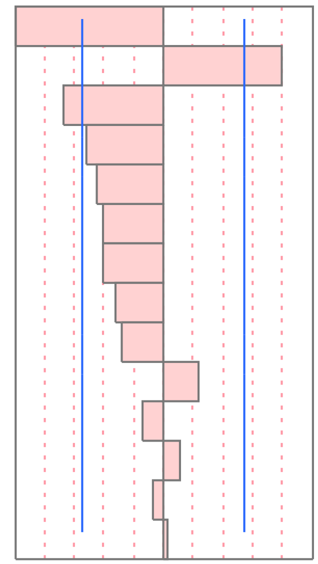
- I need to align my feet 6 degrees left of my target

Sandia National Laboratories

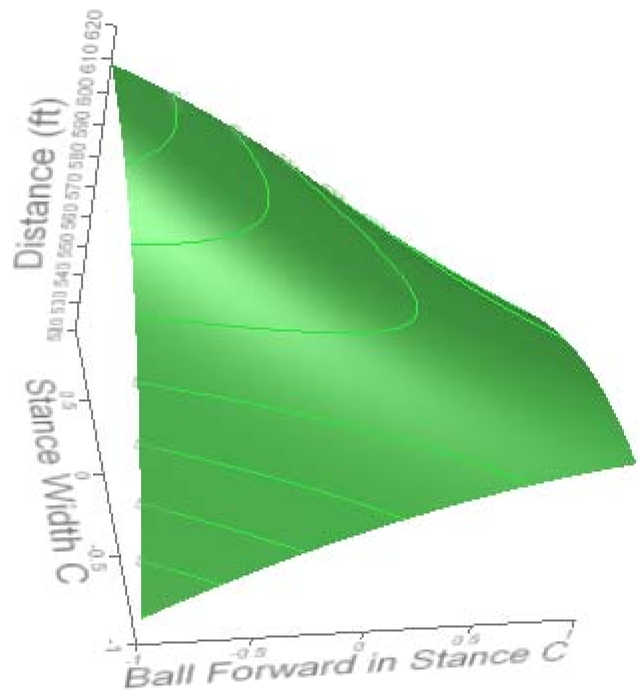
Significance of Factors

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
(Stance Width C-0.06849)*(Ball Forward in Stance C-0.0274)	-32.43008	8.918967	-3.64	0.0006*
Right Hand C	19.589935	6.777577	2.89	0.0054*
(Distance to Ball C-0.06849)*(Ball Forward in Stance C-0.0274)	-21.70966	8.918967	-2.43	0.0180*
(Right Hand C+0.09589)*(Right Hand C+0.09589)	-39.36472	20.65248	-1.91	0.0616
Distance to Ball C	-11.25945	6.898557	-1.63	0.1081
Ball Forward in Stance C	-11.23923	7.502341	-1.50	0.1395
(Right Hand C+0.09589)*(Ball Forward in Stance C-0.0274)	-11.39382	7.689659	-1.48	0.1438
(Right Hand C+0.09589)*(Stance Width C-0.06849)	-8.365868	7.098898	-1.18	0.2434
(Stance Width C-0.06849)*(Stance Width C-0.06849)	-17.24583	16.56152	-1.04	0.3020
Stance Width C	6.2352511	7.008115	0.89	0.3773
(Ball Forward in Stance C-0.0274)*(Ball Forward in Stance C-0.0274)	-6.127163	12.53297	-0.49	0.6268
(Stance Width C-0.06849)*(Distance to Ball C-0.06849)	3.1510016	7.715673	0.41	0.6845
(Right Hand C+0.09589)*(Distance to Ball C-0.06849)	-1.645035	6.956826	-0.24	0.8139
(Distance to Ball C-0.06849)*(Distance to Ball C-0.06849)	1.9091748	17.52639	0.11	0.9136

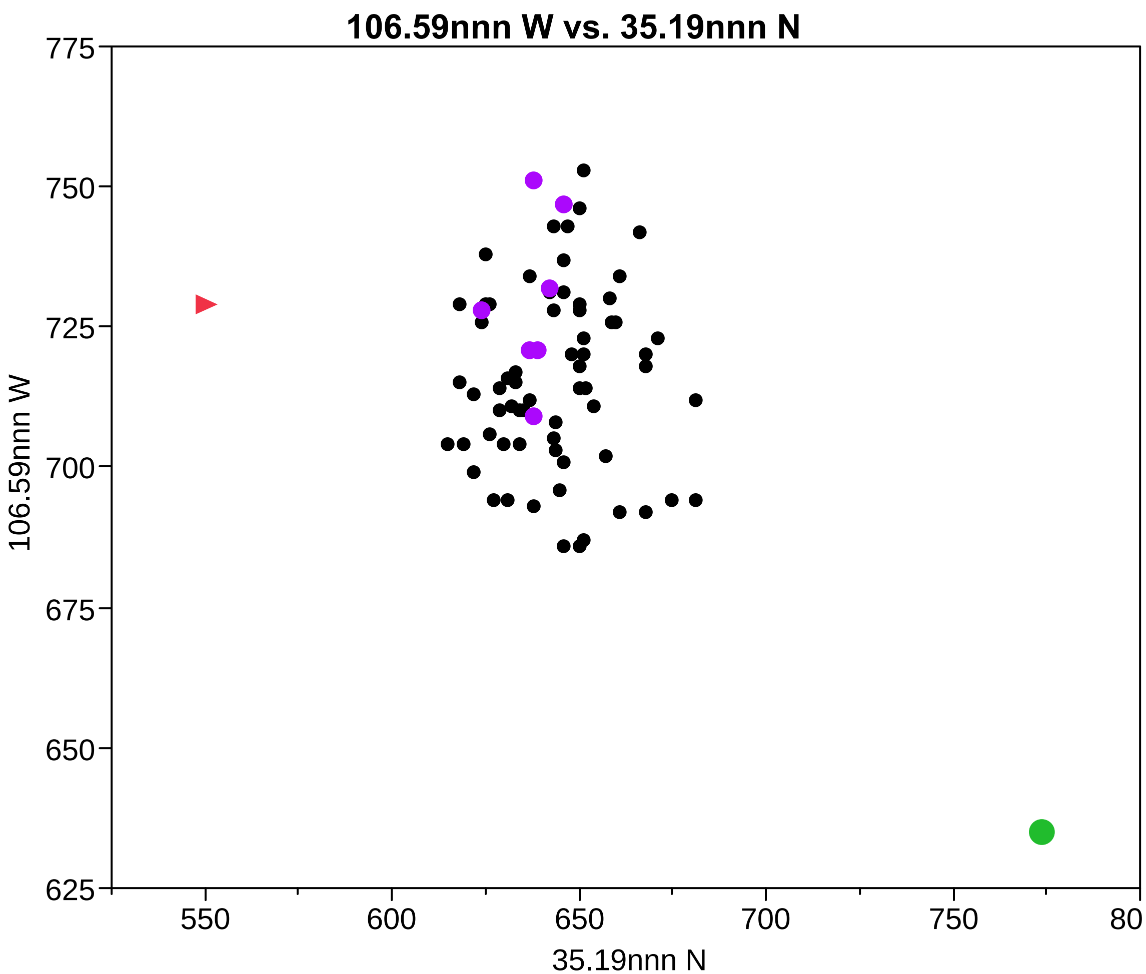


“Ball in Stance” and ”Stance Width” interaction



Use Data Filter to see Various Combinations such as Back & Wide

Graph Builder



Data Filter

8 matching rows

Select Show Include

Right Hand

Neutral Strong Weak

Stance Width

Narrow Normal Wide

Distance to Ball

Close Far Middle

Ball Forward in Stance

??? (2)
Back (20)
Forward (24)
Middle (36)

Sandia National Laboratories

Validating Results

- Unfortunately, access to Balloon Fiesta Park is a paperwork nightmare
- The true test is on the golf course
- I’ve been playing golf with my brother for 15 years and have never beat him
- With this new set-up, I tied him (missed a birdie putt on 18 or I would have beat him)
- I played my very next round in Phoenix. I strive for 6 pars, and only accomplish that about half the time. I had 11 pars that day!

Other Applications of DOE

- **Design of New Equipment, such as a new putter**
  - What Factors are significant with respect to minimizing putt variation?
    - » Moment of Inertia?
    - » Center of Gravity?
    - » Shaft Length?
- **Comparing Existing Equipment**
  - Is a hybrid better than an iron?
  - Which driver loft angle, shaft material, and shaft length are best for me?

Conclusion

- This portion of the presentation wasn't about showing you how I improved my game; it was about showing you how DOE can be used even on the most obscure processes

- Design of Experiments was used to Optimize the Set-up
  - Right Hand Position (Weak to Strong)
  - Stance Width
  - Distance to Ball (Reach)
  - Ball Forward / Backward in Stance

- The Interaction between *Ball in Stance* and *Stance Width* would have never been detected by varying only one of these at a time!

- A well-designed experiment can give us much more information at a fraction of the cost of multiple experiments

Sandia National Laboratories

Student Feedback on Using Practical Examples

- “Great new ways to simply look at existing data”
- “Tools and modeling should be added to BB training”
- “Make it (statistical modeling) mandatory for BB certification”
- “The examples were great. They really held my interest much more than boring technical examples.”

Questions / Discussion

Sandia National Laboratories

"… all models are wrong; the practical question is how wrong do they have to be to not be useful …"

George Box and Norman Draper, Empirical Model Building and Response Surfaces, John Wiley, 1987, pg. 74

Sandia National Laboratories