The Design of JMP

"A few of my favorite things" - John Sall

Why was JMP created?

Why is JMP the way it is?

What are some milestones and stories?

What were we thinking?

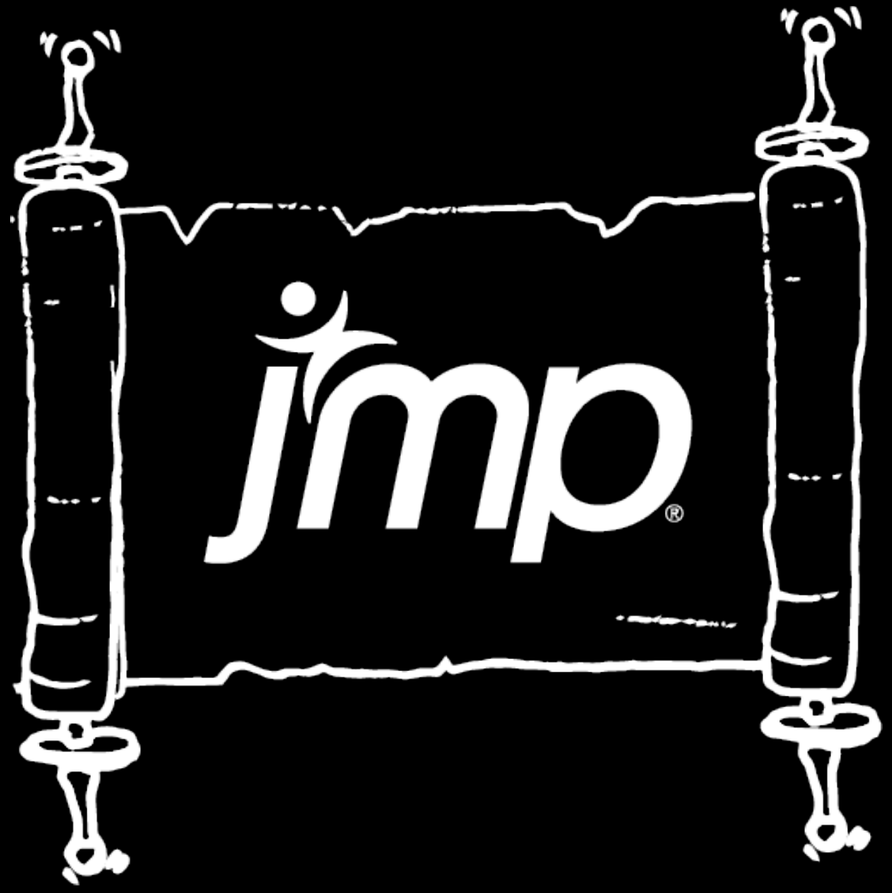
Rule #1

Always do live demos, rather than PowerPoint or Keynote.

Rule #1

… Almost always do live demos, rather than PowerPoint or Keynote.

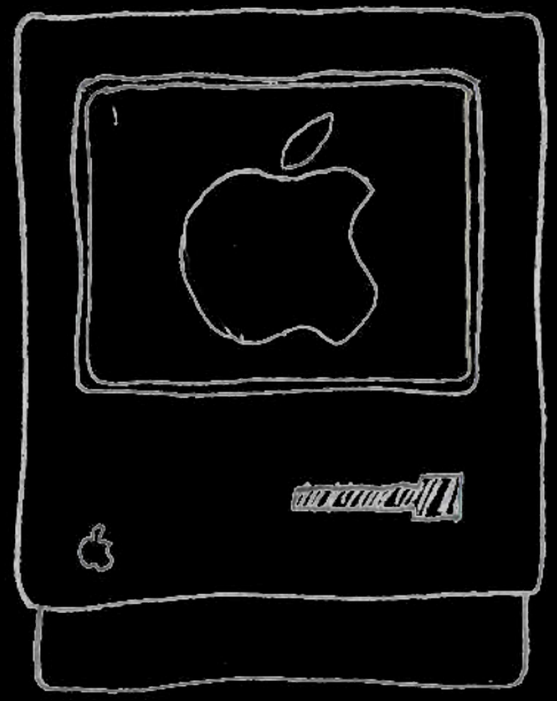
History



Startup: Why Start JMP?

The Creation Story

- In the mid-1980s – complete rewrite of SAS to run on PCs.
- But then the Mac appeared.
  It was a toy at first, but then…
  an awakening.



Startup: The ’80s GUI Revolution

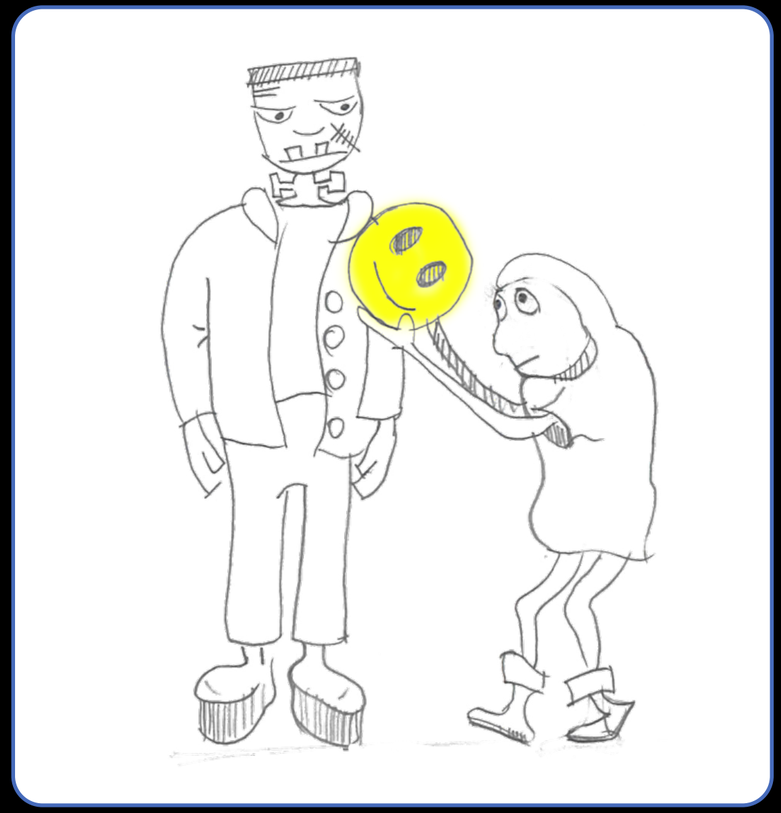
- The Mac arrives in 1984, *and gets good by 1988.*
- Point-and-click beats look-up-in-a-manual-then-type-in-commands-and-submit.
- Graphics beats tables.
- A good UI enables a much, much greater number of researchers to do computing, and not delegate to the programmer class.
- Data analysis can be DIY rather than hire-out-to-experts.

Startup: Advice for the Revolution

Steve Jobs: *Start over from scratch rather than try to evolve an old system to the new UI.*

(Don't just make the UI as a code generator for the old system.)

Adding a UI face.



"Igor, just one more touch and we are done."

Startup: Advice for the Revolution

- Steve Jobs: *Start over from scratch rather than try to evolve an old system to the new UI.* (Don't just make the UI into a code generator for the old system.)
- Design for a less technical user who does not have the patience to look up things in a manual.
- Use graphics everywhere you can.
- Keep the focus on the work. Avoid dialogs and modes.
- Interactively and organically build, rather than plan ahead.

Startup: Reason 2 – Statistical Graphics

A graph should always accompany statistics; it is not just a request-only option.

Response Y1

Summary of Fit

RSquare	0.666542
RSquare Adj	0.629492
Root Mean Square Error	1.236603
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.510001	27.5100	17.9899
Error	9	13.762690	1.5292	Prob > F
C. Total	10	41.272691		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0000909	1.124747	2.67	0.0257*
X1	0.5000909	0.117906	4.24	0.0022*

Response Y2

Summary of Fit

RSquare	0.666242
RSquare Adj	0.629158
Root Mean Square Error	1.237214
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.500000	27.5000	17.9656
Error	9	13.776291	1.5307	Prob > F
C. Total	10	41.276291		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0009091	1.125302	2.67	0.0258*
X2	0.5	0.117964	4.24	0.0022*

Response Y3

Summary of Fit

RSquare	0.666324
RSquare Adj	0.629249
Root Mean Square Error	1.236311
Mean of Response	7.5
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.470008	27.4700	17.9723
Error	9	13.756192	1.5285	Prob > F
C. Total	10	41.226200		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0024545	1.124481	2.67	0.0256*
X3	0.4997273	0.117878	4.24	0.0022*

Response Y4

Summary of Fit

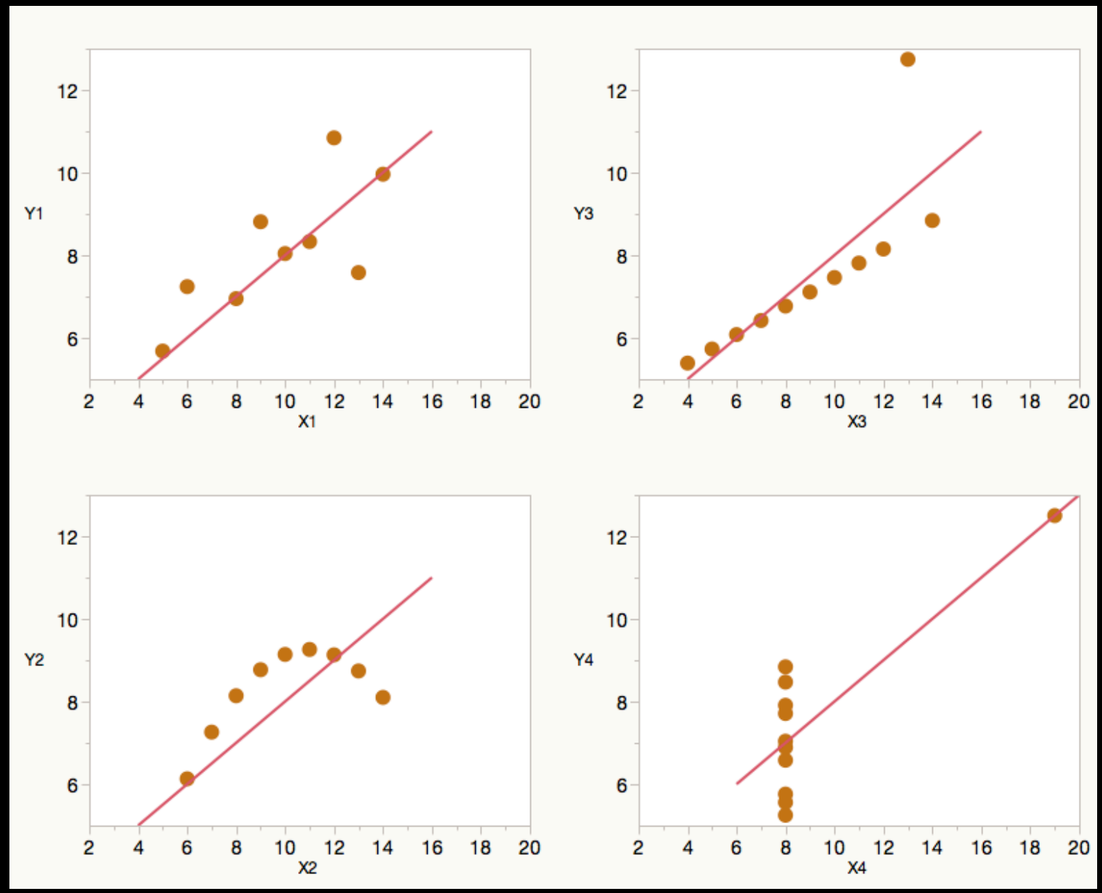
RSquare	0.666707
RSquare Adj	0.629675
Root Mean Square Error	1.235695
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.490001	27.4900	18.0033
Error	9	13.742490	1.5269	Prob > F
C. Total	10	41.232491		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0017273	1.123921	2.67	0.0256*
X4	0.4999091	0.117819	4.24	0.0022*



https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Startup: Reason 3 – SAS Growth

- SAS was growing into an enterprise-class product.
- SAS was a programming language that took some investment to learn.
- We wanted to invest in something that was easier, smaller and more personal than SAS.
- If we didn't, others would take the market.

Startup Team

So we started the *Statistical Instruments* project, with a team of four.

Michael Hecht, Ann Lehman, John Sall, Chung-Wei Ng

The Name: Why Name It JMP?

- We wanted to make a new brand, rather than a SAS sub-brand.
- Of course, lots of the names we liked were already taken.

SAS Institute Inc.
SAS Circle ☐ Box 8000
Cary, NC 27512-8000
Phone (919) 467-8000 ☐ Fax (919) 469-3737

M E M O R A N D U M

TO: John Sall

FROM: Pat Brown PLB

DATE: May 31, 1988

RE: Trademark Search - "Prospector"

We have completed a search on PROSPECTOR to determine its availability for use as an Institute trademark. It does not appear that either PROSPECTOR alone or as SAS/PROSPECTOR are available as there is already a registered trademark, "PROSPECTOR", for computer programs. (A copy of this registration is enclosed.) This registration would bar our use of even SAS/PROSPECTOR for the same type of goods.

In light of the prior registration, it will be necessary to select another name for this new product.

If we can be of further assistance, please contact me.

PLB/SDK

Enclosure
prospect(6)

copa Stat
velox
Statistactory
Statis Faction
empirica
Gamut
EXPERTISE
XPERTIZE
JMP
experior
angler
circinus
Illuma
metridyne
GIST
Aleph
METRICIAN
DATAVISION
Statrix
PRACTITIONER
PARADIGM
NOW YOU CAN JMP
BLITZ
Moebius
metriQue
TO THE RIGHT
inclusion
Iris
ARGOS
SIGMA
FACTOTUM
STAT-O-MATIC

What Stuck

- JMP is a machine instruction to branch to a new place.
- Jump to a new kind of product.
- Jump to a new level of ease-of-use.

What Stuck

- Bill Gjertsen became the marketing person for the product.
- When he wrote a memo about it, he used the phrase “John’s Mac product” and then abbreviated it to J.M.P. later in the memo.

Scientists and Engineers

- SAS was fine for IT and statisticians, for production applications worth a significant investment.

- Engineers and scientists wanted something easy to learn and interactive, on a limited budget.

The UI

little red triangles

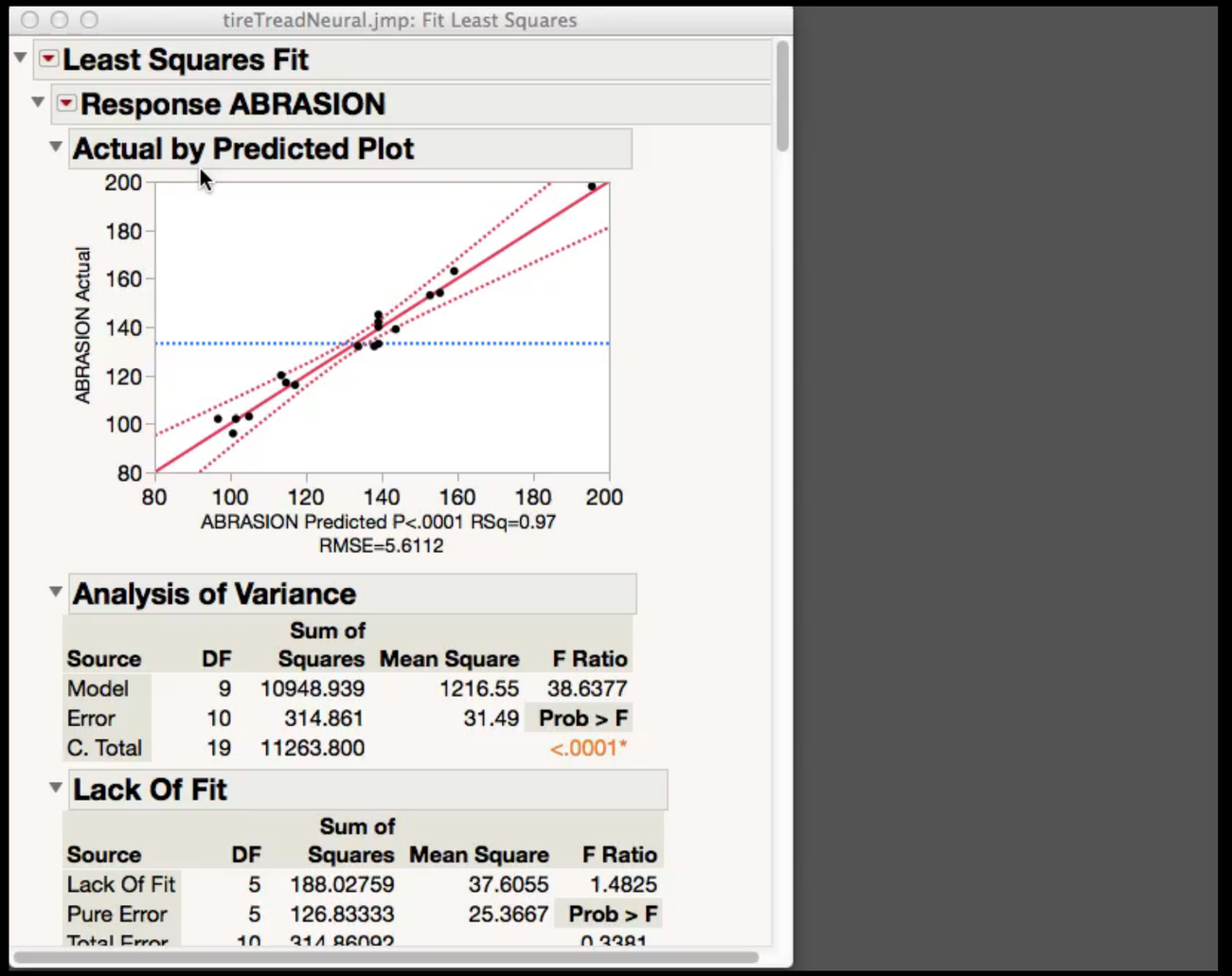
Keep the work surface central, don't divert into modes and dialogs. Always have a view of the document.

- Steve Jobs

little red triangles

- We needed an icon that meant "Click here for more commands."
- We needed to make it in a color that made it easy to find.
- We needed it small, so as not to detract from the document's central focus.

little red triangles



little red triangles

Warning to Novice Users:

If you miss the role of the little red triangle hot spots, you miss most of the features in JMP.

The Document Surface

Documents must contain both text and graphics and be very interactive.

Document must adapt to its content automatically.

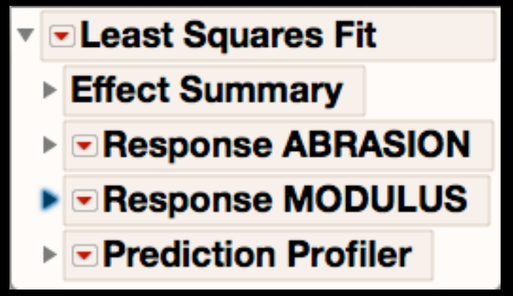
Document must be expandable, customizable, copyable, printable, cursor-active.

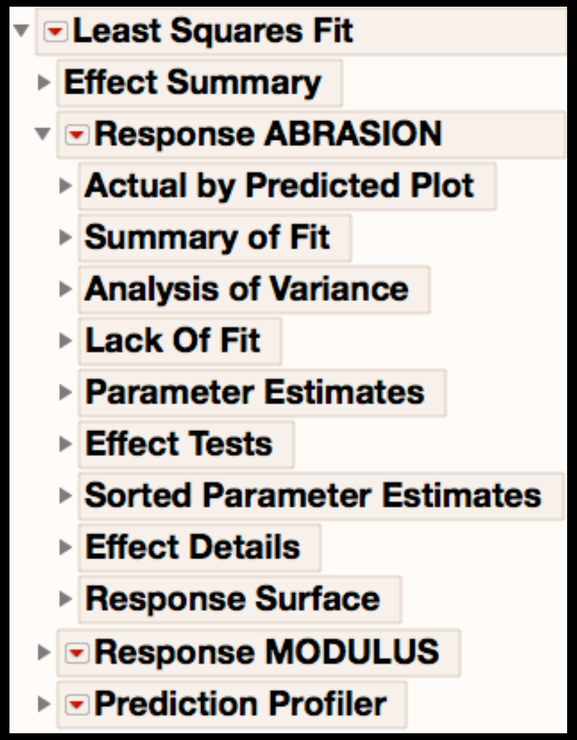
We found four key ideas…

Hypertext

Engelbart: *Hypertext is text displayed on a computer display or other electronic devices with references (hyperlinks) to other text which the reader can immediately access, or where text can be revealed progressively at multiple levels of detail (also called StretchText).* - Wikipedia

- Sections can open and close.
- Sections can nest.
- Serves as an organizer that is a table of contents when closed and becomes the document when open.
- The Outline Node.

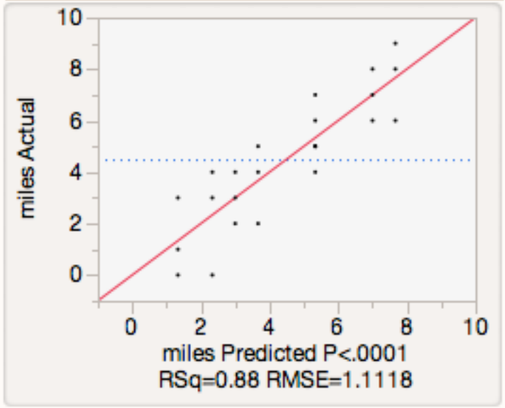




Boxes and Glue

Response miles

Actual by Predicted Plot



Parameter Estimates

Term	Estimate	Std Error
Intercept	4.4583333	0.42287
species[COYOTE]	1.4583333	0.42287
season[fall]	-0.625	0.393083
season[spring]	1.7083333	0.393083
season[summer]	0.875	0.393083
species[COYOTE]*season[fall]	0.0416667	0.393083
species[COYOTE]*season[spring]	-0.625	0.393083
species[COYOTE]*season[summer]	0.875	0.393083

Don Knuth's TeX

OutlineBox(1)
{Y("miles")}
HelpKey=FitLS
Response miles

OutlineBox(2)
HelpKey=Fit YByPred
Actual by Predicted Plot

ListBox(4)

AxisBox(1)
miles Actual

FrameBox(1)
HelpKey=FitLS Leverage
LineSeg
LineSeg
MarkerSeg

TextEditBox(1)
miles Actual

AxisBox(2)
miles Predicted P<.0001
RSq=0.88 RMSE=1.1118

TextEditBox(2)
miles Predicted P<.0001

TextEditBox(3)
RSq=0.88 RMSE=1.1118

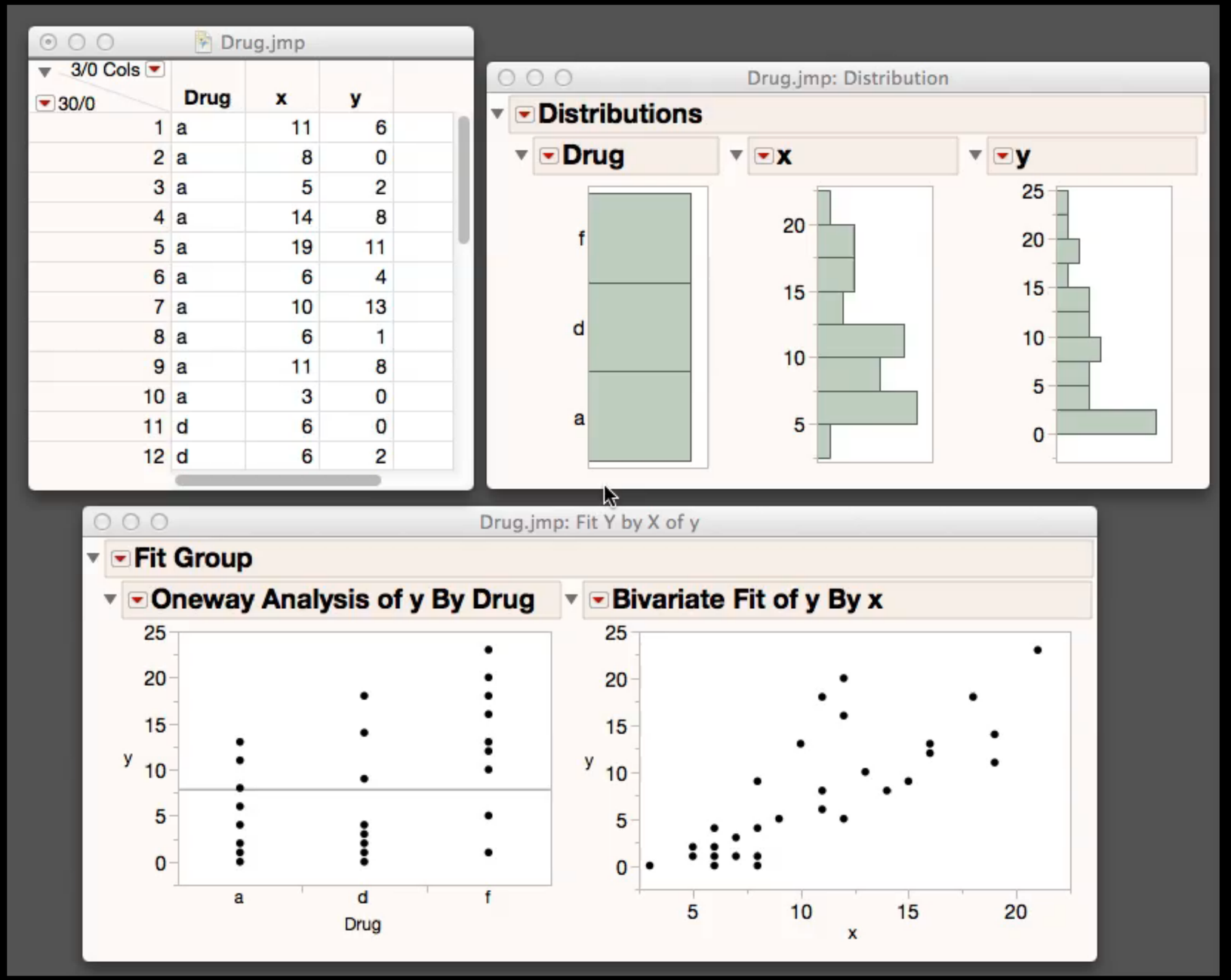
OutlineBox(4)
HelpKey=Fit Estimates
Parameter Estimates

TableBox(2)

StringColBox(2) Term	NumberColBox(2) Estimate	NumberColBox(3) Std Error	NumberColBox(4) DFDen	NumberColBox(5) t Ratio	NumberColBox(6) Prob>|t|
Intercept	4.4583333	0.42287	4	10.54	0.0005*
species[COYOTE]	1.4583333	0.42287	4	3.45	0.0261*
season[fall]	-0.625	0.393083	12	-1.59	0.1378
season[spring]	1.7083333	0.393083	12	4.35	0.0010*
season[summer]	0.875	0.393083	12	2.23	0.0459*
species[COYOTE]*season[fall]	0.0416667	0.393083	12	0.11	0.9173
species[COYOTE]*season[spring]	-0.625	0.393083	12	-1.59	0.1378
species[COYOTE]*season[summer]	0.875	0.393083	12	2.23	0.0459*

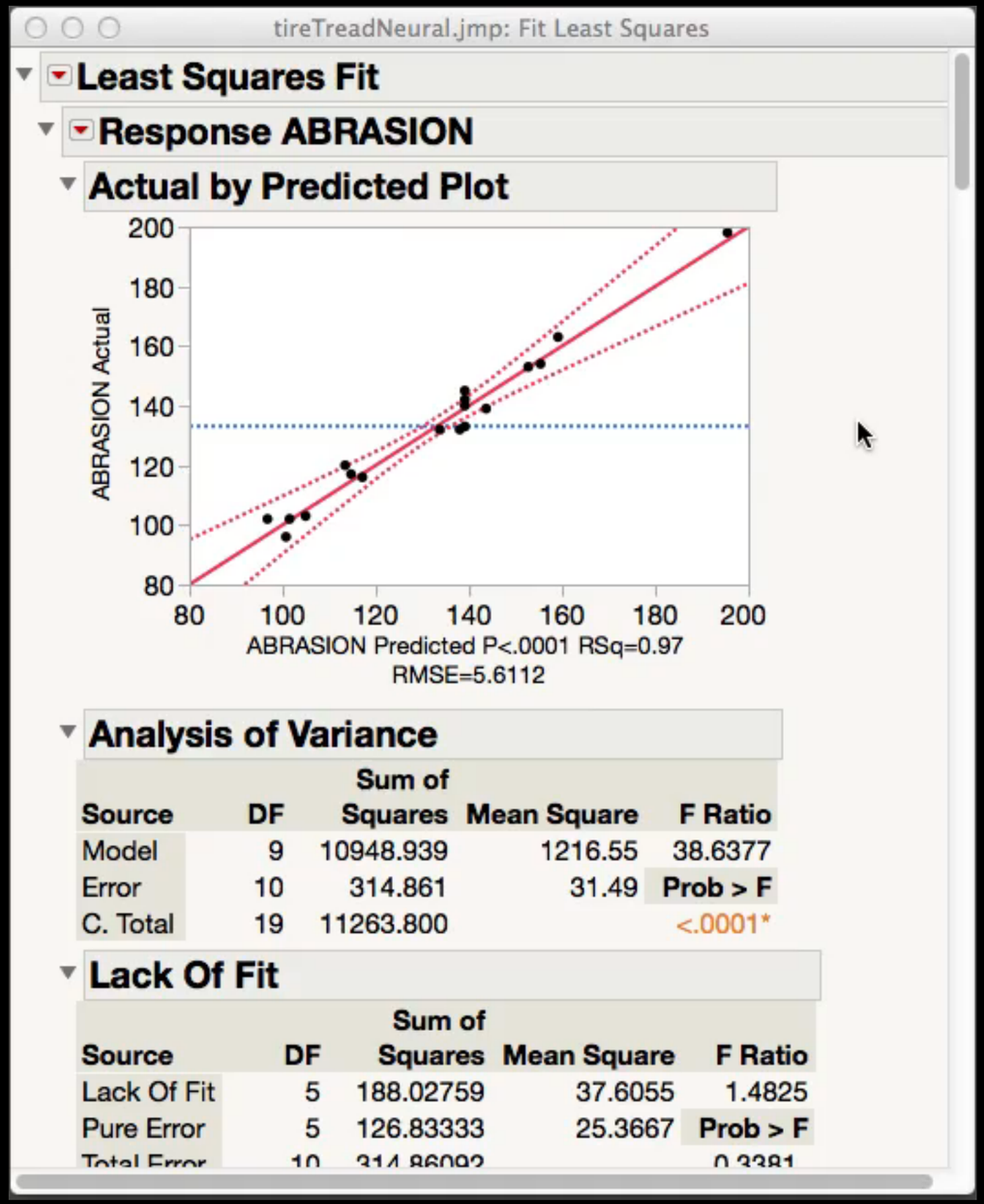
Linked Selection

Identifying
Selecting
Brushing



Dynamic Linking by Rows

- Linked rows have limited, if any, benefit to a large portion of users.
- Which users?
- Those who run in Windows Maximized mode, where you can only see one window at a time.



Smart Scrolling

- Page-oriented or continuous scroll?
- Both
- “Sticky title”

- Steve Jobs was adamant that there should be only one button on a mouse.
- Or even no buttons.
- But sometimes you need to both point and say that you wanted a context menu for that place.

Apple “invented” the two-handed context click.

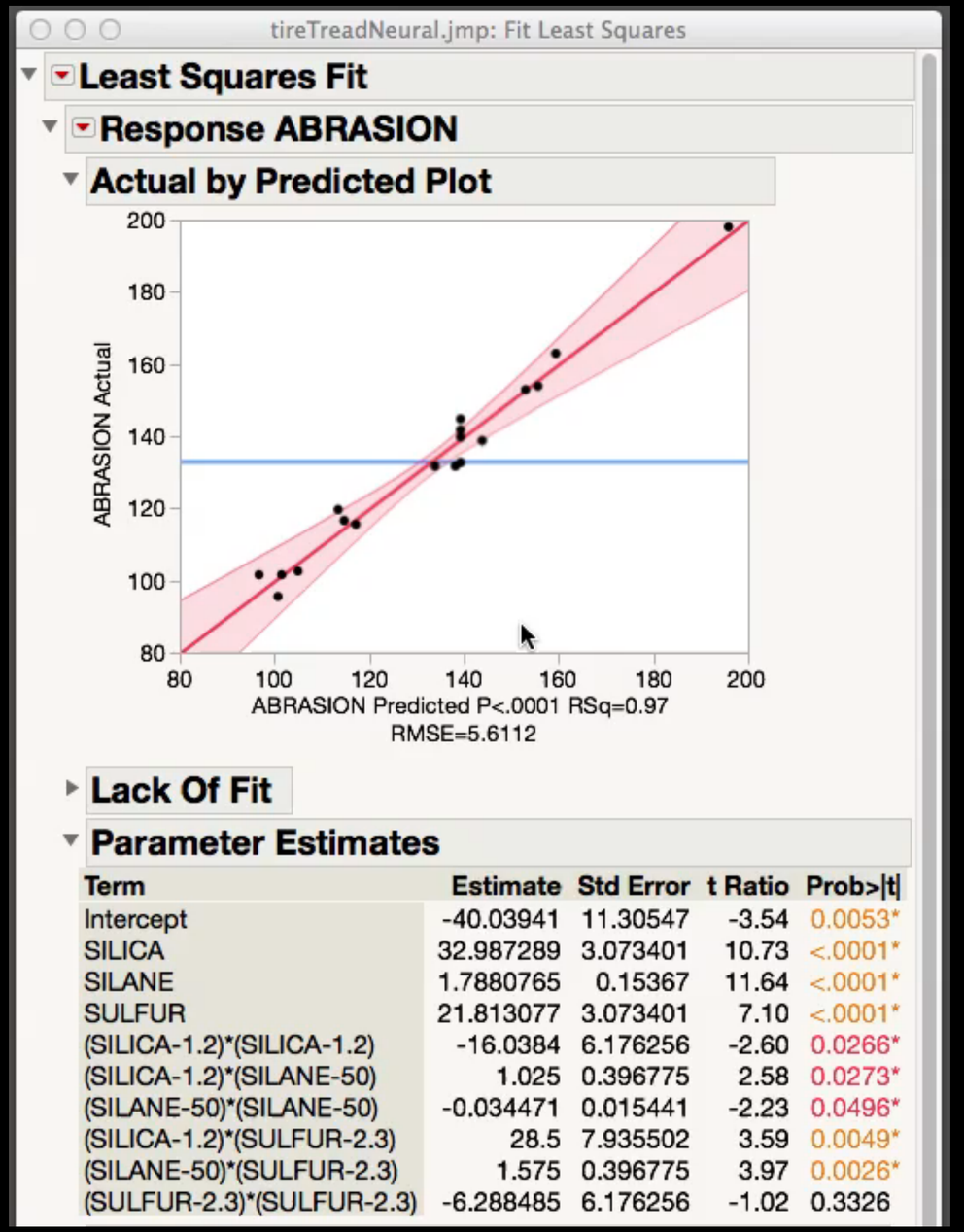
But one company thought a two-button mouse would be far easier.

Microsoft and the rise of the middle finger.

Left-click for selection
Right-click for context menu

Context Clicking Everywhere

For customization



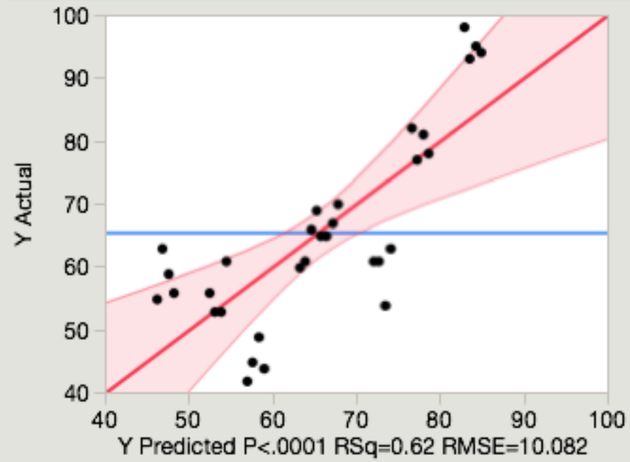
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-40.03941	11.30547	-3.54	0.0053*
SILICA	32.987289	3.073401	10.73	<.0001*
SILANE	1.7880765	0.15367	11.64	<.0001*
SULFUR	21.813077	3.073401	7.10	<.0001*
(SILICA-1.2)*(SILICA-1.2)	-16.0384	6.176256	-2.60	0.0266*
(SILICA-1.2)*(SILANE-50)	1.025	0.396775	2.58	0.0273*
(SILANE-50)*(SILANE-50)	-0.034471	0.015441	-2.23	0.0496*
(SILICA-1.2)*(SULFUR-2.3)	28.5	7.935502	3.59	0.0049*
(SILANE-50)*(SULFUR-2.3)	1.575	0.396775	3.97	0.0026*
(SULFUR-2.3)*(SULFUR-2.3)	-6.288485	6.176256	-1.02	0.3326

Color-Keyed

- Background is light gray or beige.
- Graphics frames are white, and they "pop."
- Red triangles easily found.
- P-values, loadings, correlations colored by value.

Response Y

Actual by Predicted Plot



Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	65.5	1.782241	36.75	<.0001*
F	-0.6875	1.782241	-0.39	0.7028
Ct	9.75	1.782241	5.47	<.0001*
A	-0.3125	1.782241	-0.18	0.8622
T	5.375	1.782241	3.02	0.0057*
Cn	-3.125	1.782241	-1.75	0.0913

Effect Tests

Effect Details

Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4
1500m	2.34334	0.45598	-0.61233	0.42617
Pole Vault	0.98255	0.08337	-0.13498	-0.02076
High Jump	0.58271	0.17023	-0.35803	-0.06839
Long Jump	-0.04758	-0.54366	-0.12770	-0.05579
Discus	0.43876	1.19124	-0.22715	0.43821
400m	-0.39054	0.42294	0.41052	-0.07569
100m	-1.09641	0.23434	0.89945	-0.06694
Javelin	-0.35705	0.09886	0.64548	-0.04826
100m hurdles	-0.92866	-0.10786	0.49114	0.03803
Shot Put	-0.00643	0.23130	-0.03980	1.02655

Correlations

	100m	Long Jump	Shot Put	High Jump	400m	100m hurdles	Discus	Pole Vault	Javelin	1500m
100m	1.0000	-0.3106	0.1501	-0.2380	0.4971	0.4209	0.2620	-0.1889	0.4834	-0.5267
Long Jump	-0.3106	1.0000	-0.0881	-0.0385	-0.0009	0.0932	-0.4295	-0.0793	-0.2427	-0.0012
Shot Put	0.1501	-0.0881	1.0000	-0.1554	0.0306	0.1908	0.4318	-0.1798	0.0305	0.0413
High Jump	-0.2380	-0.0385	-0.1554	1.0000	0.0140	-0.0991	0.0620	-0.1124	-0.3302	0.2599
400m	0.4971	-0.0009	0.0306	0.0140	1.0000	0.1124	0.3439	-0.1314	0.3088	-0.2033
100m hurdles	0.4209	0.0932	0.1908	-0.0991	0.1124	1.0000	0.0168	-0.0182	0.3019	-0.4121
Discus	0.2620	-0.4295	0.4318	0.0620	0.3439	0.0168	1.0000	-0.0564	0.0792	0.1968
Pole Vault	-0.1889	-0.0793	-0.1798	-0.1124	-0.1314	-0.0182	-0.0564	1.0000	-0.1995	0.4009
Javelin	0.4834	-0.2427	0.0305	-0.3302	0.3088	0.3019	0.0792	-0.1995	1.0000	-0.2063
1500m	-0.5267	-0.0012	0.0413	0.2599	-0.2033	-0.4121	0.1968	0.4009	-0.2063	1.0000

Summary: UI Ideas

- Hot spots for optional analysis.
- Hypertext for organizing, unfolding.
- Boxes and glue for layout.
- Dynamic linking for selection, etc.
- Smart scrolling to keep titles sticky.
- Customization to serve user preferences.
- Color-keyed for *pre-attentive* cognition.

The Platforms

Rules for Analyses: 1

Ask less

- Do not ask the user anything the computer can determine by itself.
- Do not force the user to make decisions that have reasonable defaults, especially when the user can change things later.
- Do not force the user to become experts before they analyze. The software should limit itself to good choices.

Rules for Analyses: 2

Remember more

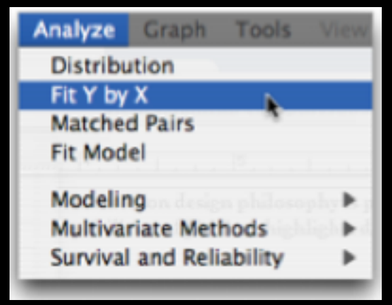
- The modeling type is assigned in the data table; platforms do not have to ask the user to specify it later.
- Make it easy to have the software remember things so that user doesn't have to specify them again. Remember things as column properties.

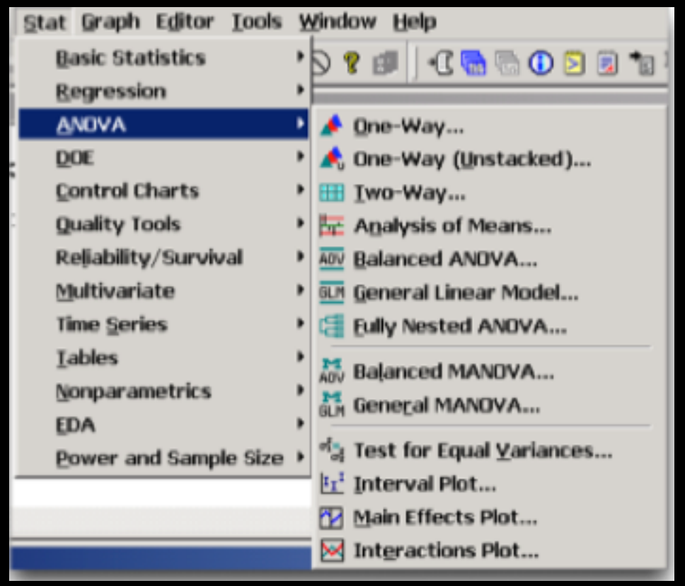
Rules for Analyses: 3

Cover more ground with fewer commands

- Make the platforms generic so they adapt to the modeling type where possible.
- Provide structure to the choices.
- Make the platform general so there are no limitations or requirements. All must support missing values, excluded rows, etc.

Why the Platform?



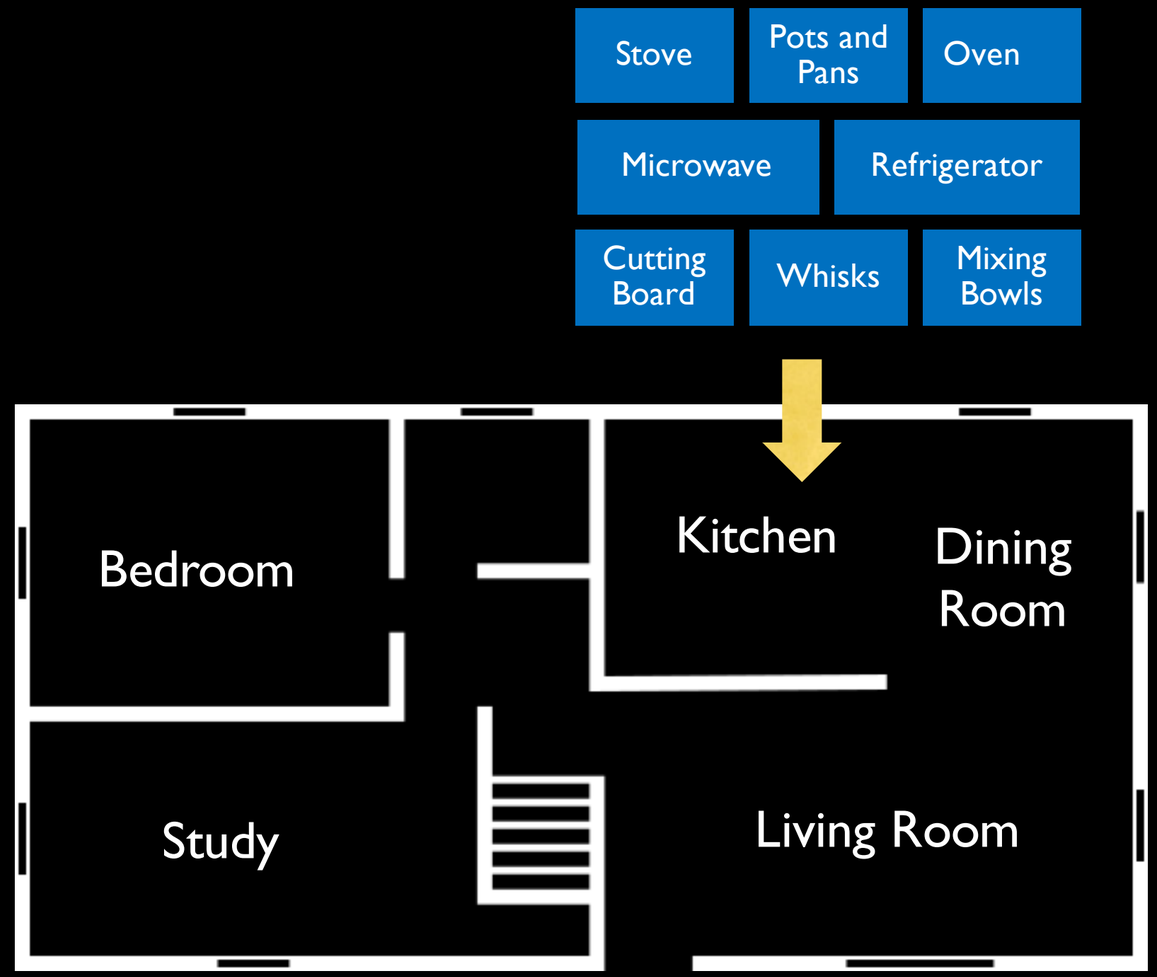


Launch by situation, rather than from a huge enumeration of all the analytical methods.

- Limited choices – so it's easy to find.
- Choices by situation – non-technical.
- Quick to launch – minimal specification.
- Don't have to plan ahead.

Why the Platform?

- Think of rooms in a house, each with specialized roles.
- A kitchen has all the features for preparing meals.



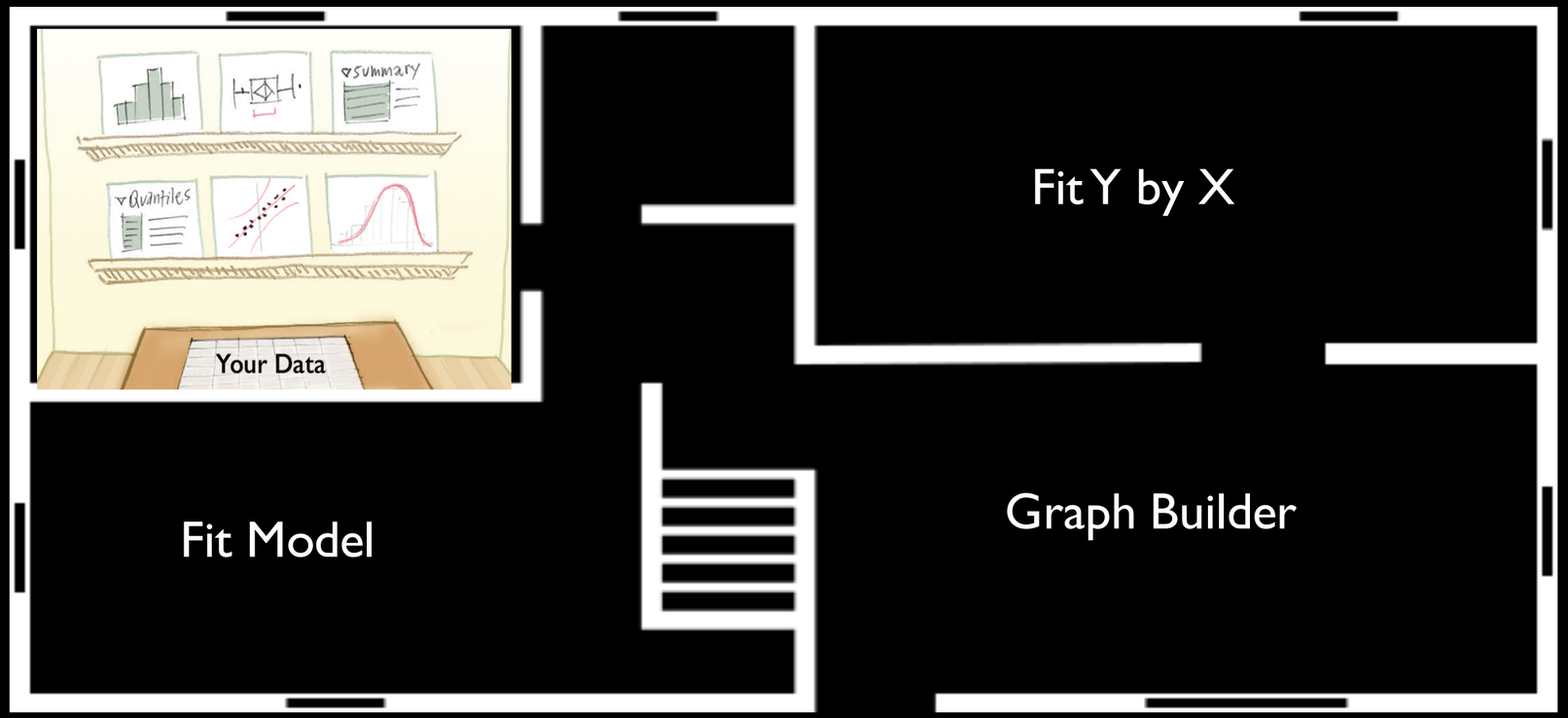
Why the Platform?

- Platforms are places with specialized roles.
- In a room, all the features are useful in certain situations.

The "look at variables individually" room.

The "look at Y variables by X variables" room.

Distribution

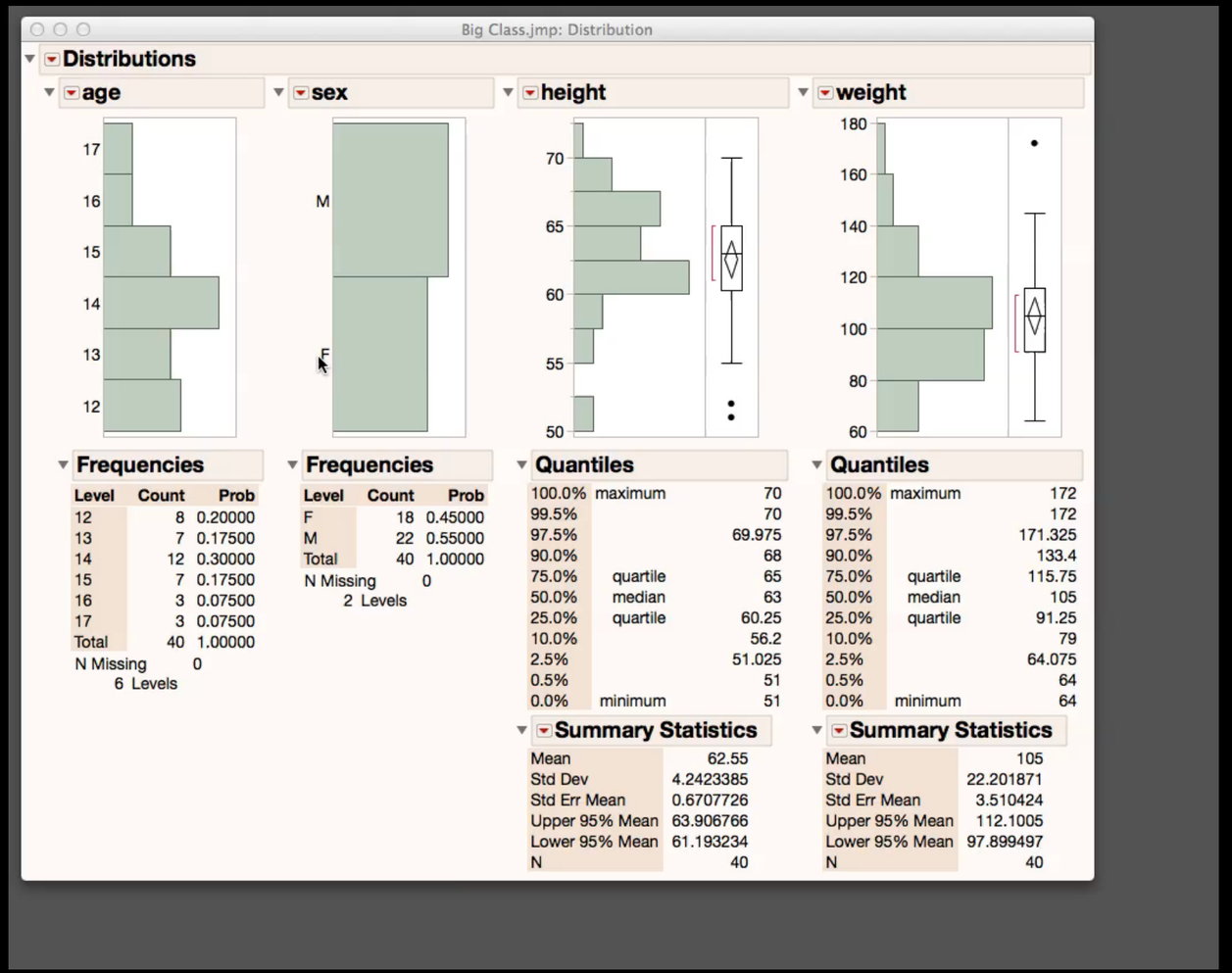


The "fit Y by many X's" room.

The graph-making shop.

The First Platform

- Distribution (univariate) – many variables analyzed individually.
- Generic – continuous variables done differently than categorical (nominal and ordinal).
- Interactive histogram linked to rows in the data.
- Interactive with bin width and position.
- Details on demand.



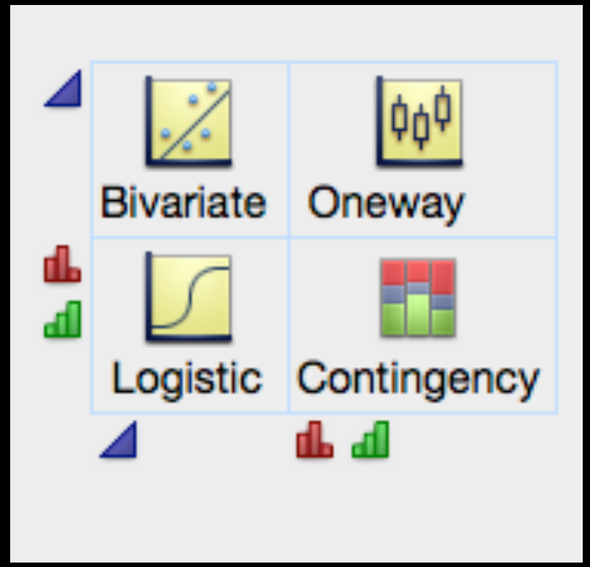
Which Situations Make Platforms?

- One at a time (Distribution).
- Two at a time, where one is response, the other factor (Fit Y by X).
- Matched pairs (before and after).
- Fit Model.
- Modeling, Multivariate.
- Subject-matter areas: quality, reliability, consumer research.

The Second Platform

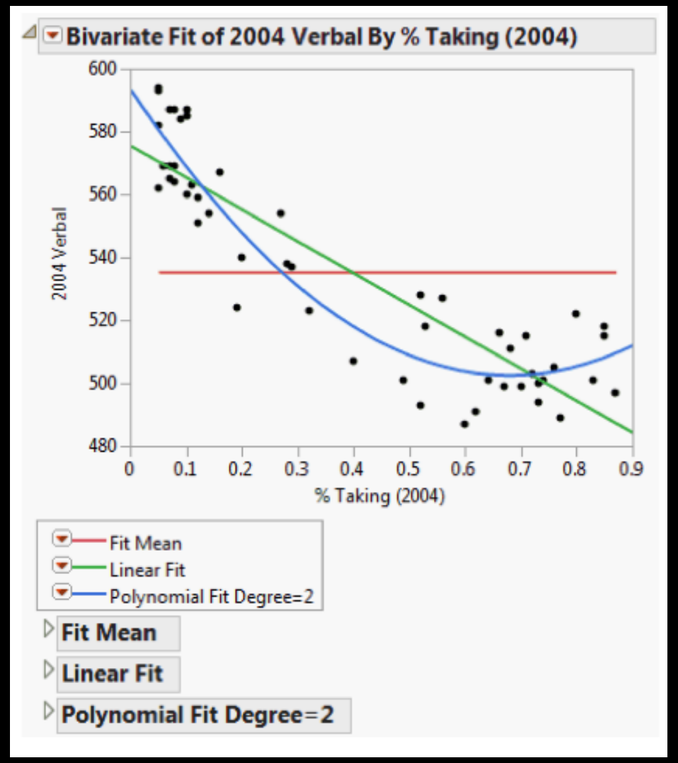
Fit Y by X is really four platforms:

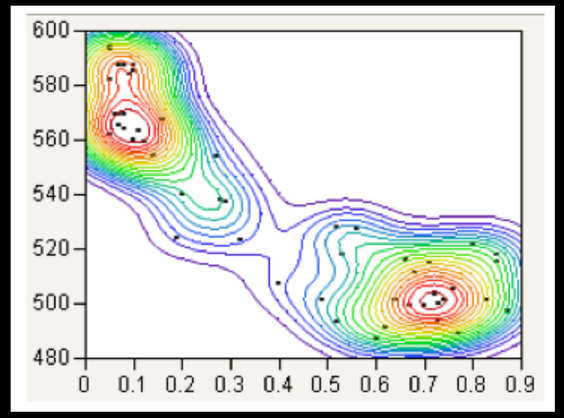
- Continuous-by-continuous
  scatterplot with regression fits.
- Continuous-by-categorical
  one-way analysis of variance with side-by-side points.
- Categorical-by-categorical
  contingency table with mosaic chart.
- Categorical-by-continuous
  logistic regression with logistic plot.

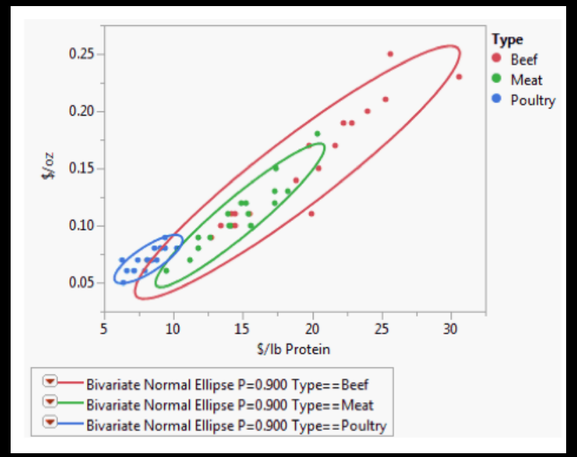


Bivariate – Fit Continuous Y by Continuous X

- Everything you might want to do with two continuous variables.
- Regressions to degree, transformed regressions.
- Correlations with normal density contour ellipses.
- Smoothers – splines.
- Nonparametric density contours.

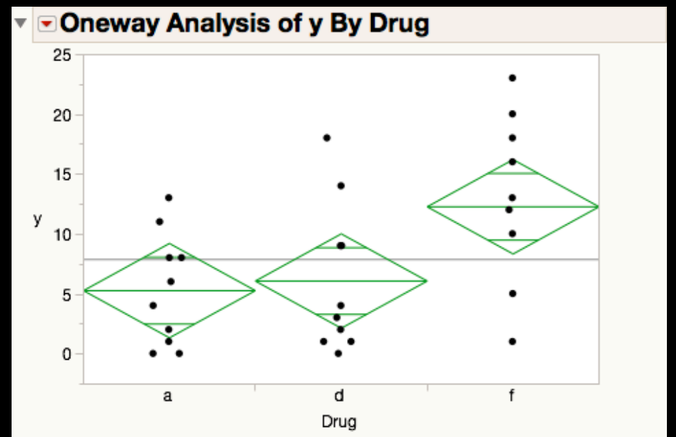






Oneway – Fit Continuous Y by Categorical X

- Everything you might want to do with a continuous Y and a categorical X
- Means, ANOVA, T-Test
- Nonparametric tests
- Multiple comparisons
- Graphics
- Comparison circles



Oneway Analysis of y By Drug

Oneway Anova

Summary of Fit

Rsquare	0.227826
Adj Rsquare	0.170628
Root Mean Square Error	6.070878
Mean of Response	7.9
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Drug	2	293.6000	146.800	3.9831	0.0305*
Error	27	995.1000	36.856		
C. Total	29	1288.7000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
a	10	5.3000	1.9198	1.3609	9.239
d	10	6.1000	1.9198	2.1609	10.039
f	10	12.3000	1.9198	8.3609	16.239

Std Error uses a pooled estimate of error variance

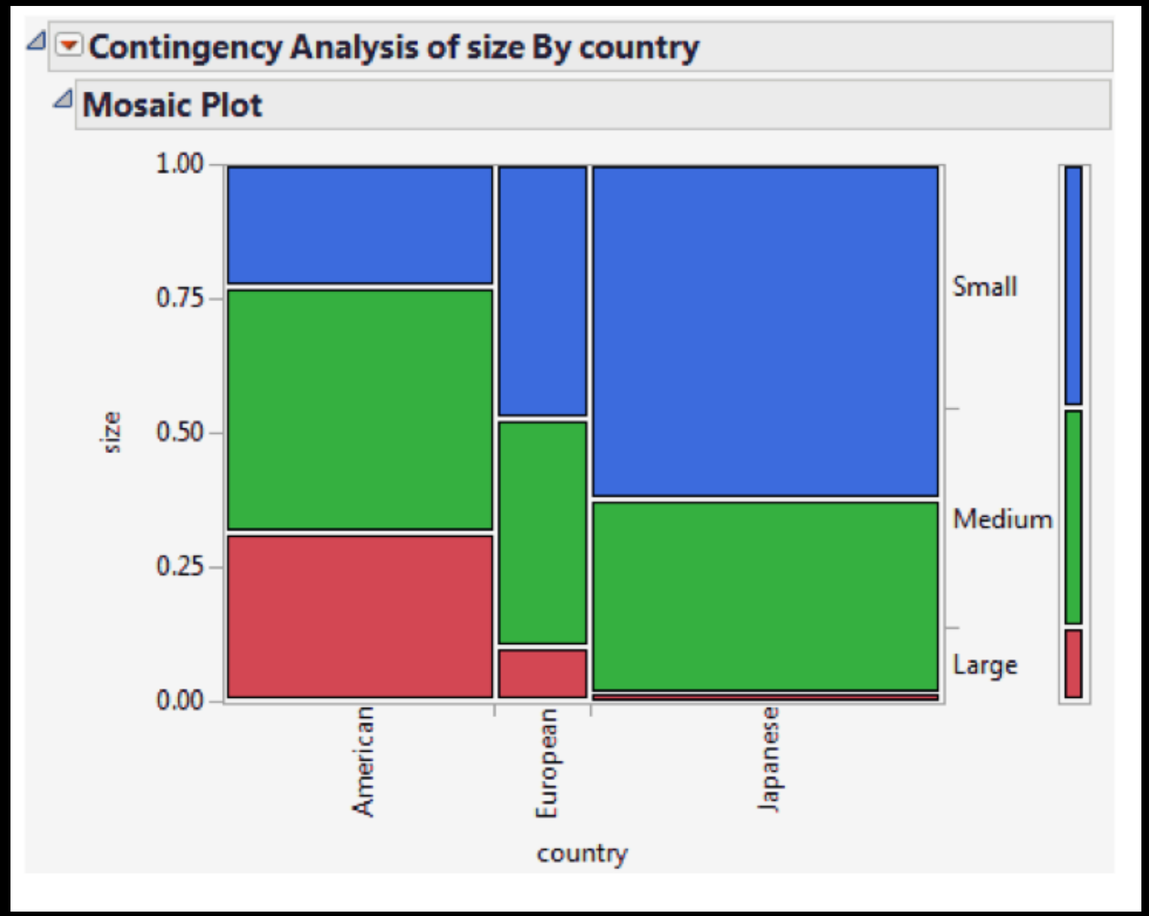
Welch's Test

F Ratio	DFNum	DFDen	Prob > F
3.3942	2	17.406	0.0569

Welch Anova testing Means Equal, allowing Std Devs Not Equal

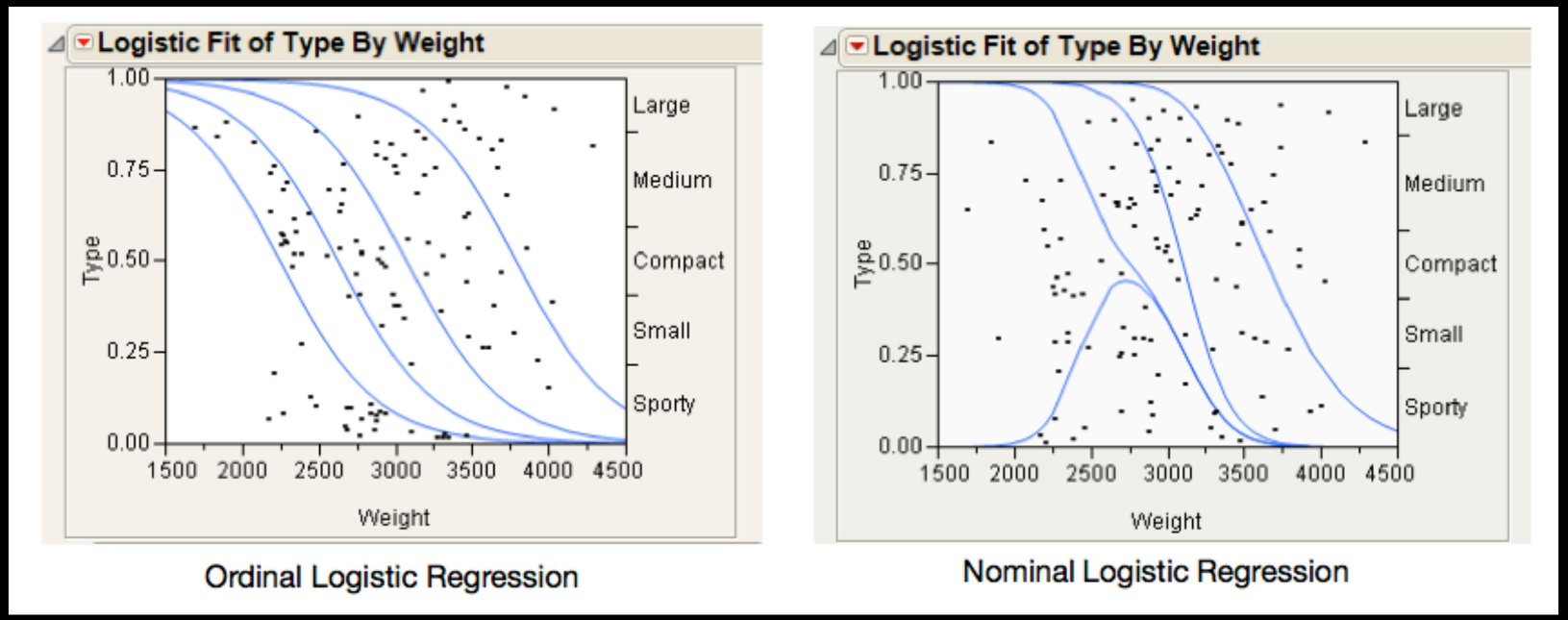
Contingency – Fit Categorical Y by Categorical X

- Everything you might want to do with two categorical variables.
- Crosstabs.
- Mosaic plot. First subdivide by X category proportions. Then subdivide by Y categories.
- If the rates are the same in each group (marginal homogeneity) the sections line up well.



Logistic – Fit Categorical Y by Continuous X

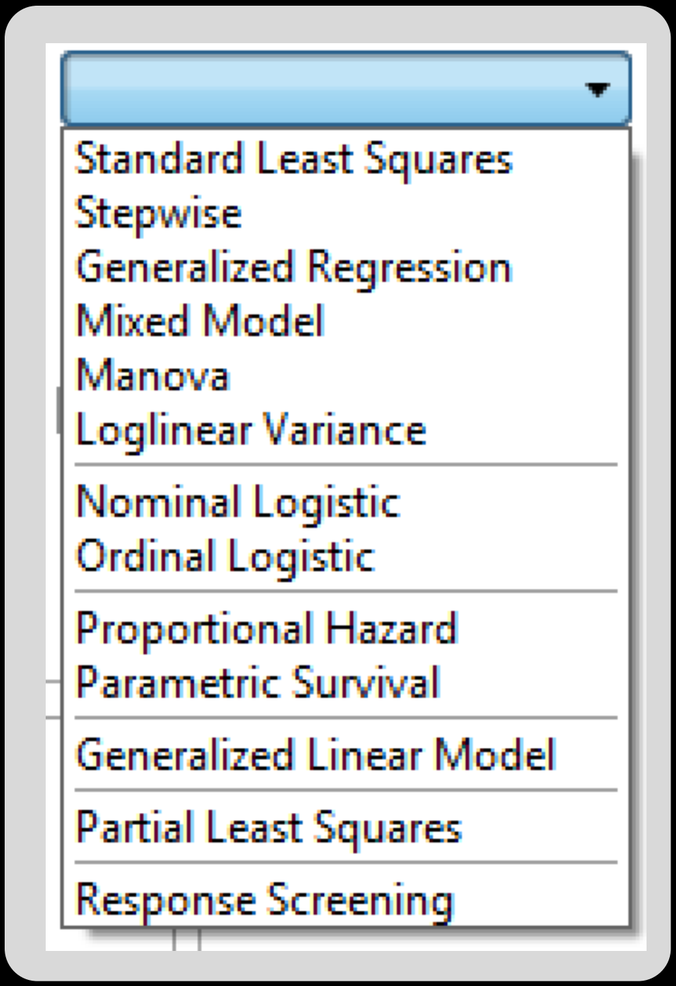
The response is the probability across the Y categories.
That probability is the distance between curves.



Fit Model

Model Dialog

- Standard Least Squares
- Stepwise
- Generalized Regression
- Mixed Model
- MANOVA
- Log-linear Variance
- Nominal Logistic
- Ordinal Logistic
- Proportional Hazard
- Parametric Survival
- Generalized Linear Model
- PLS
- Response Screening



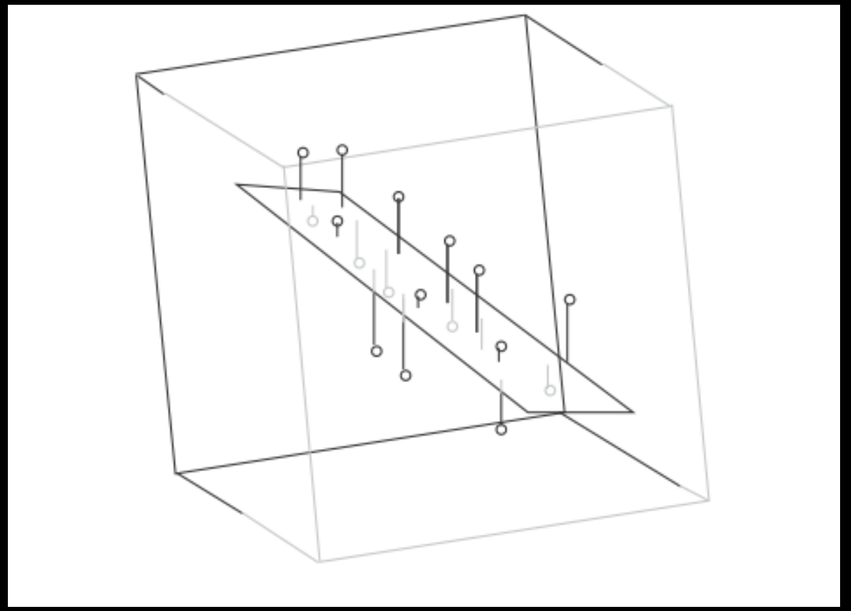
JMP PRO

General effects: polynomial effects, interactions, nesting.

Statistical Graphics

Statistical Graphics

- A graph for every statistical test.
- The best graph for each situation.
- Graphs that encourage visualization of forces between data and model.



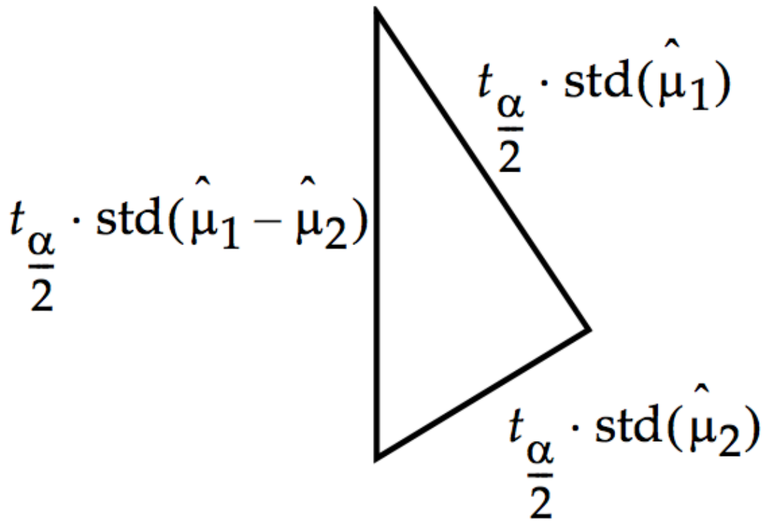
Comparing Two Means with Unequal Sample Sizes

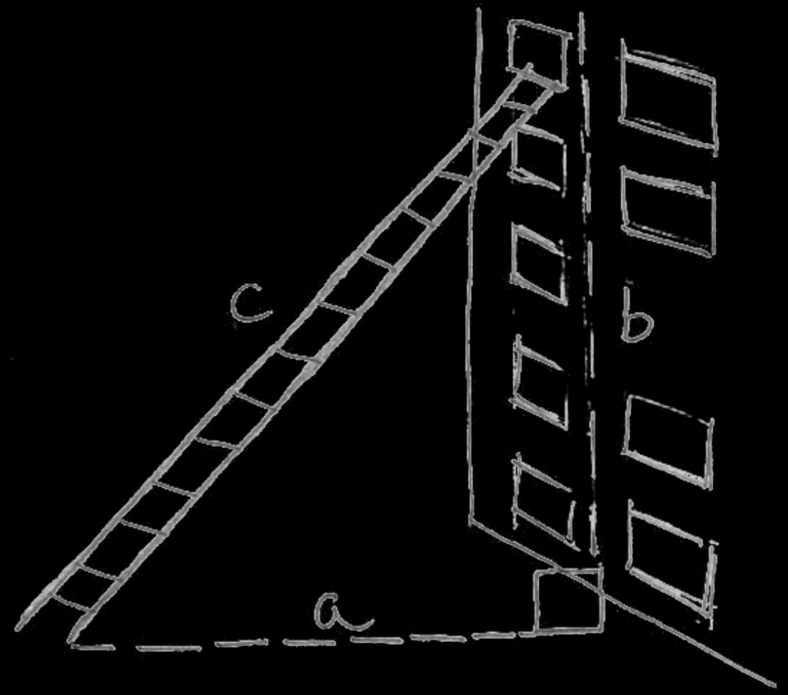
$$\text{LSD} = t_{\alpha/2}\text{std}(\hat{\mu}_1 - \hat{\mu}_2)$$

$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

Pythagorean Theorem
$c^2 = a^2+b^2$

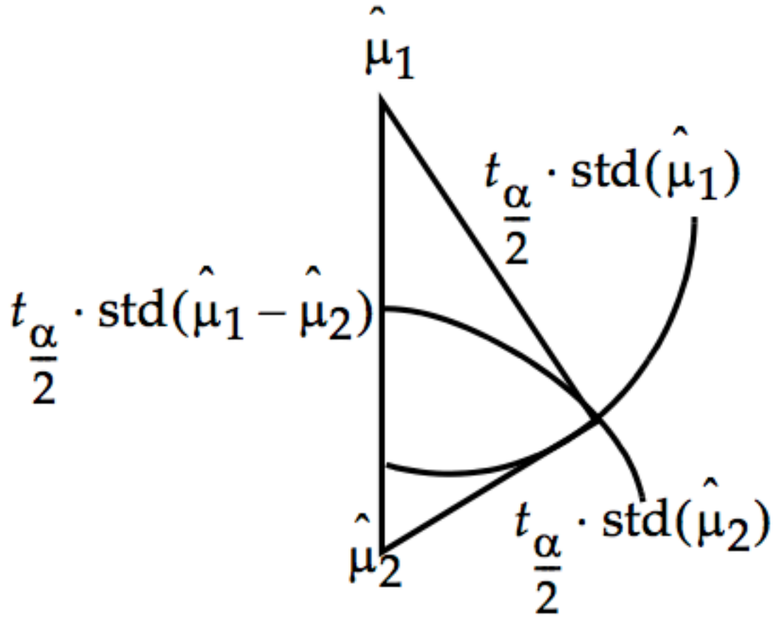
Figure 6.32 Relationship of the Difference between Two Means





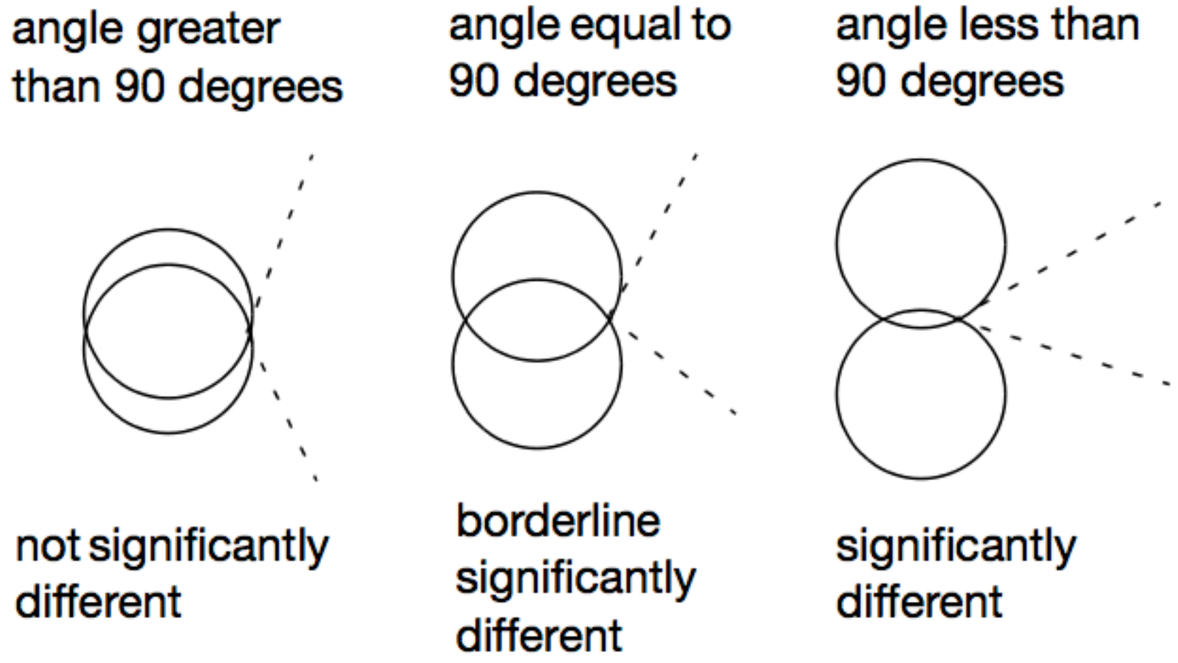
Circles an LSD Apart Intersect at Right Angles

Figure 6.33 Geometric Relationship of *t*-test Statistics

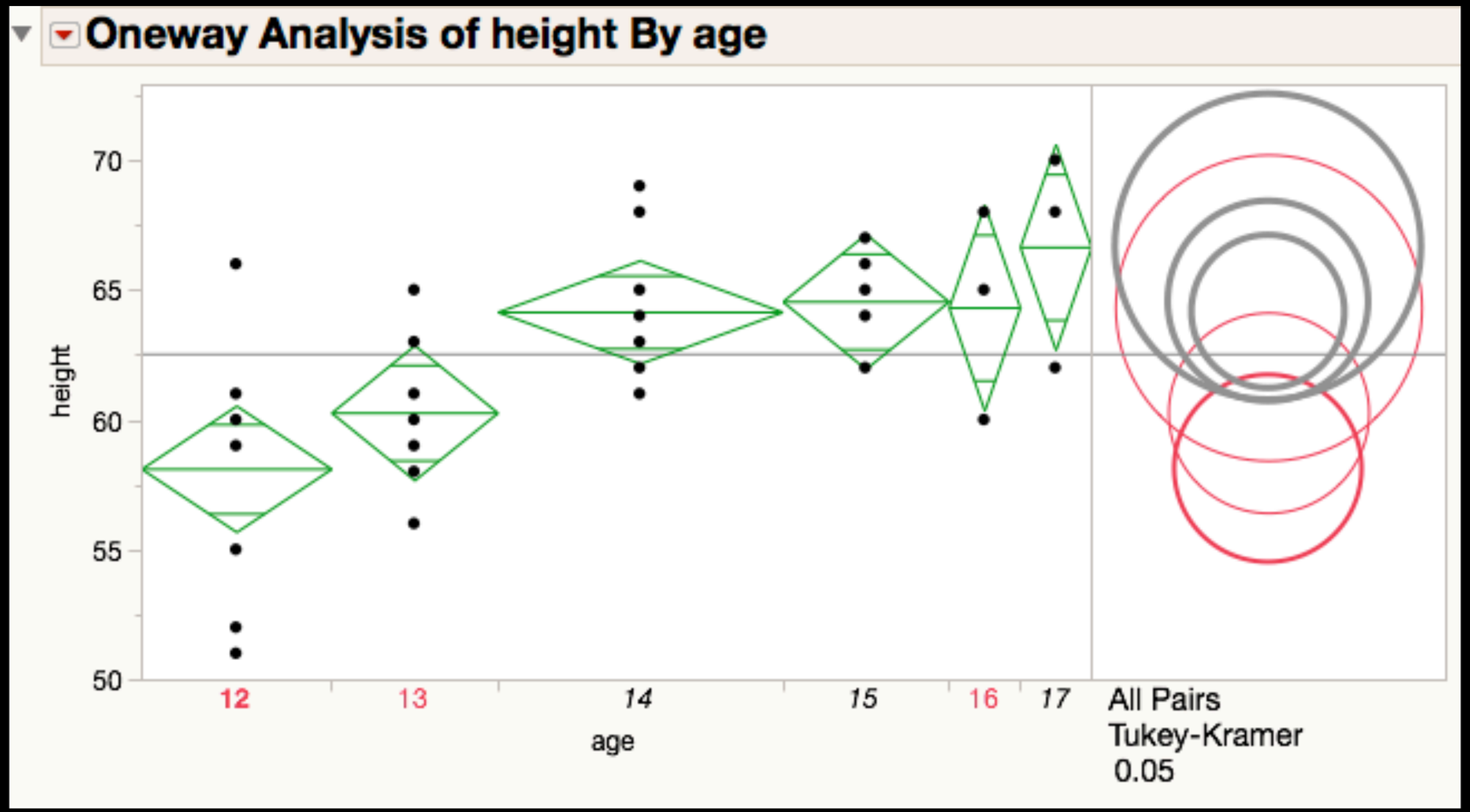


Comparison Circles

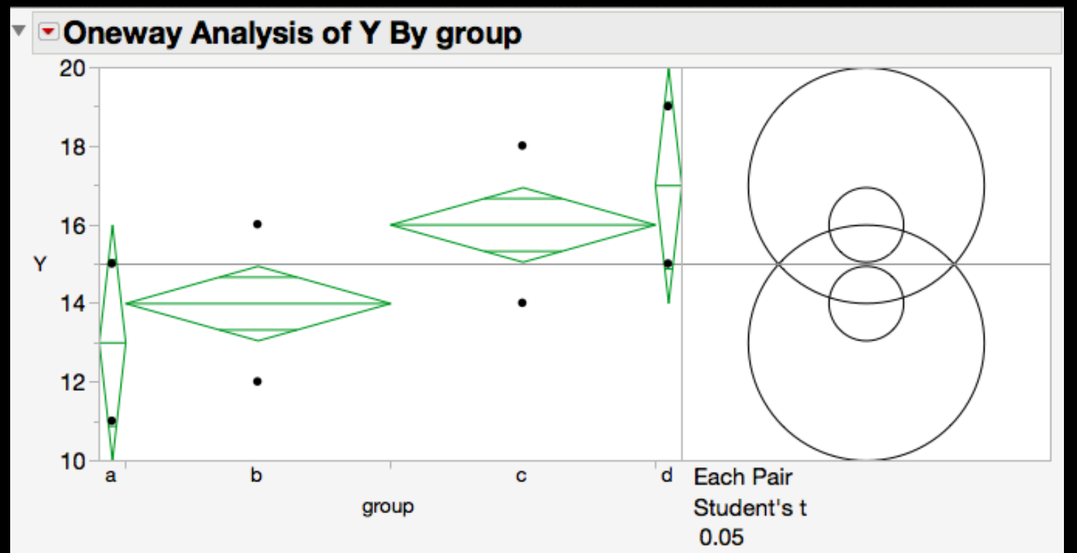
Figure 6.11 Angles of Intersection and Significance



Comparison Circles



Comparison Circles



Freq Freq

Oneway Anova

Means Comparisons

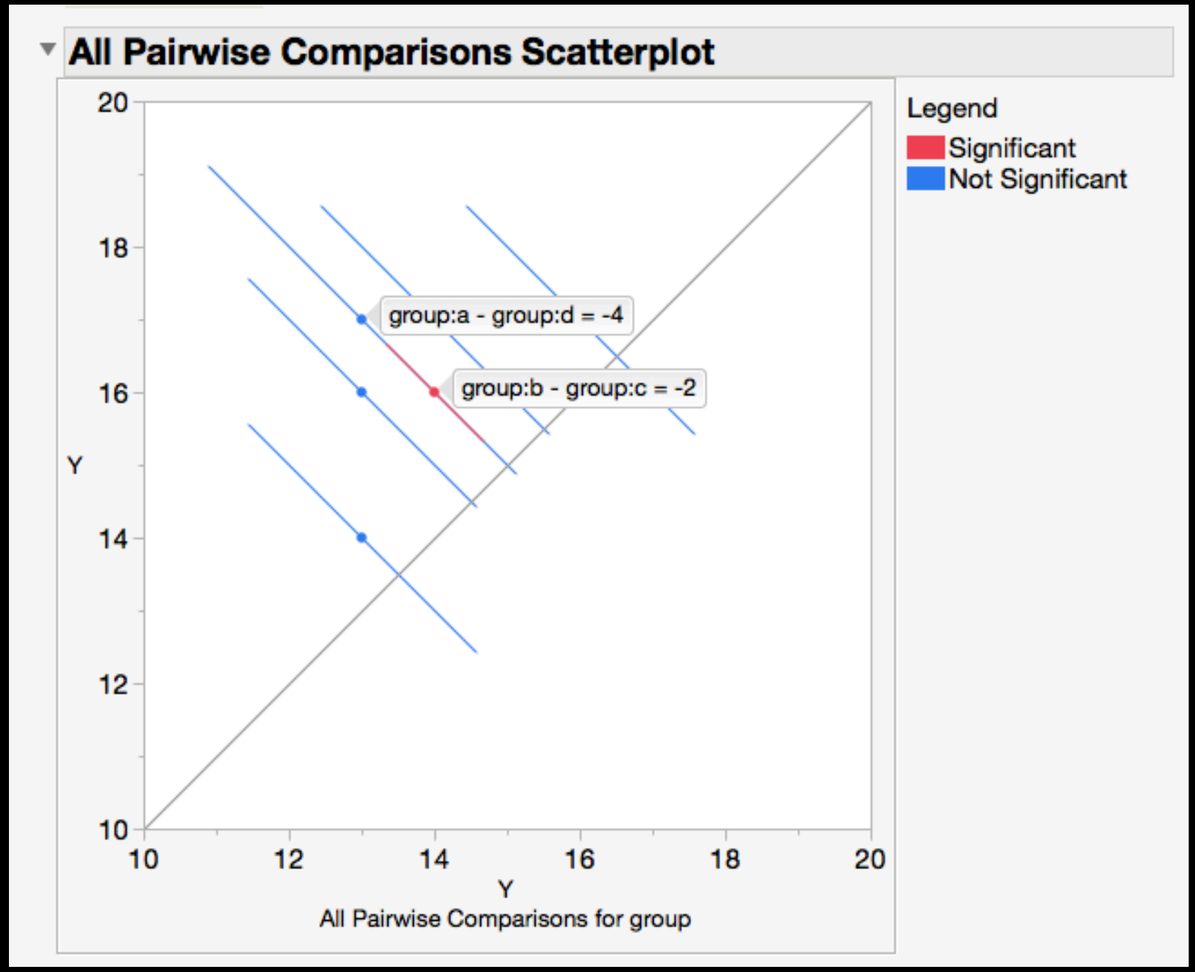
Comparisons for each pair using Student's t

Connecting Letters Report

Level		Mean
d	A B	17.000000
c	A	16.000000
b	B	14.000000
a	A B	13.000000

Levels not connected by same letter are significantly different.

Dif-o-gram

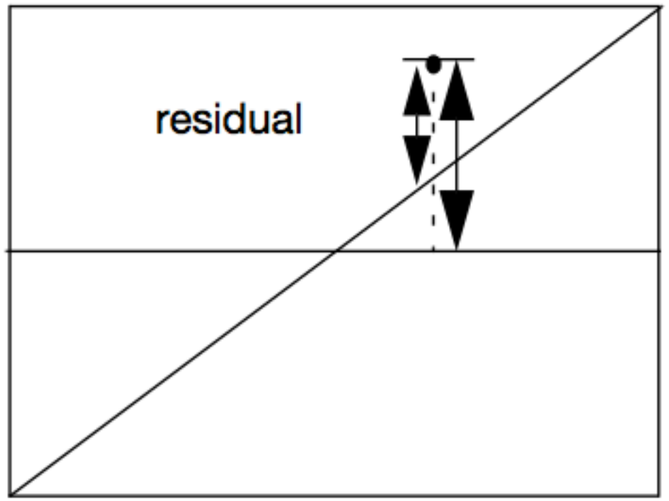


An element for each pair, rather than for each mean.

Fit Model
Standard Least Squares

- What graph tells the story of the significance of each effect?
- “Added variable plots” (Cook Weisberg) and “residual leverage plots” (Belsley, Kuh, Welch) only applied to continuous regressors, but they were on the right track.
- We needed a generalization that applied to any effect, any hypothesis.

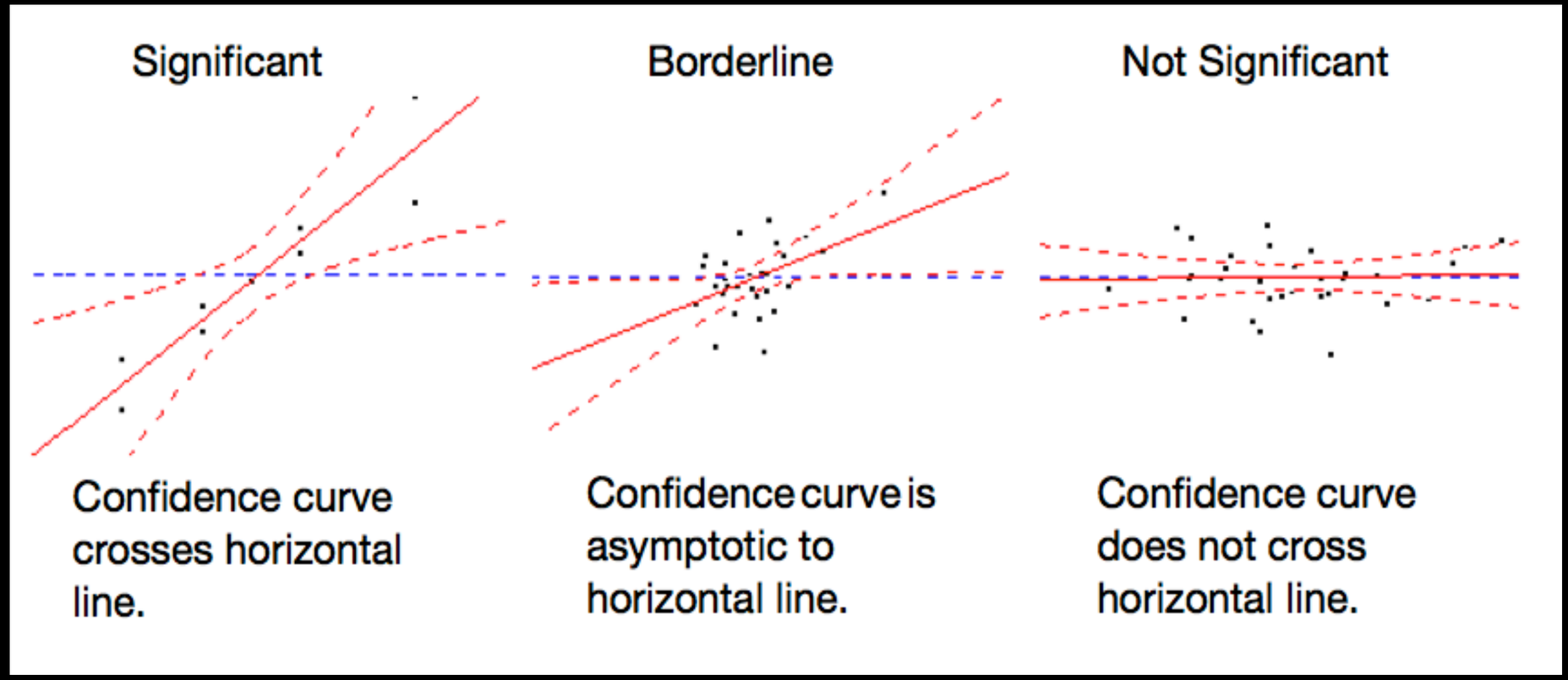
Leverage Plot



residual constrained by hypothesis

Points at the extremes exert greater leverage than points near the middle exert.

Leverage Plot

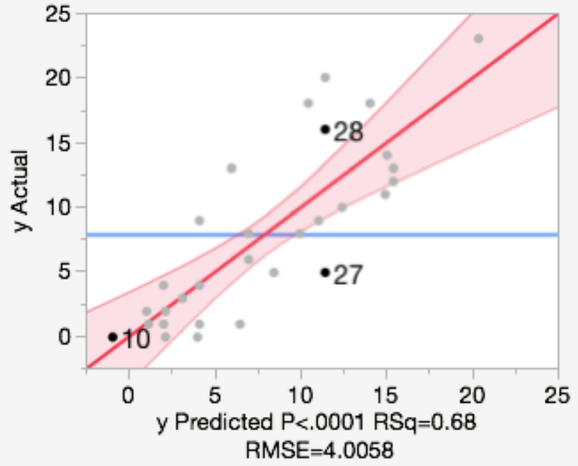


Each Test Explained Point-by-Point

Response y

Whole Model

Actual by Predicted Plot



Effect Summary

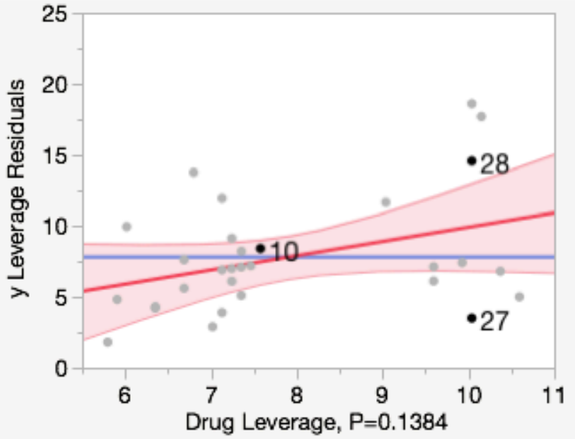
Lack Of Fit

Analysis of Variance

Parameter Estimates

Drug

Leverage Plot

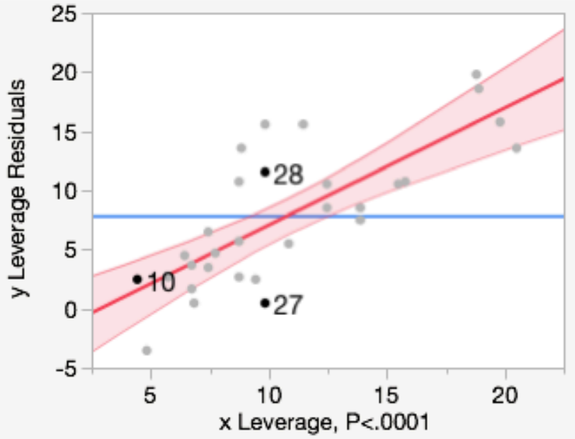


Least Squares Means Table

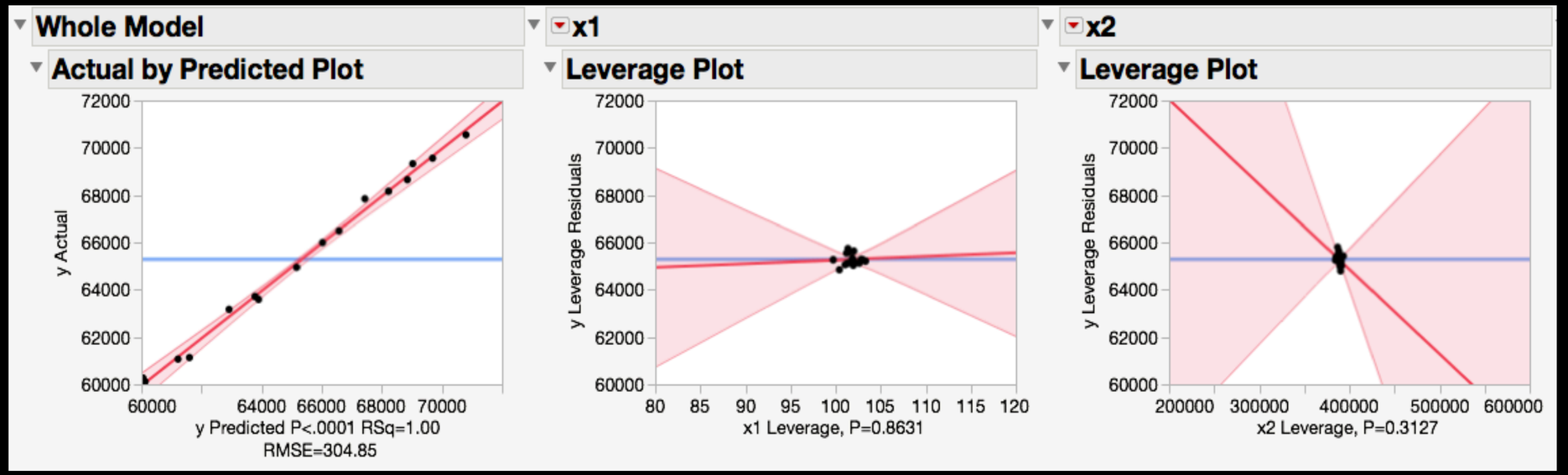
Level	Least Sq Mean	Std Error	Mean
a	6.714963	1.2884943	5.3000
d	6.823935	1.2724690	6.1000
f	10.161102	1.3159234	12.3000

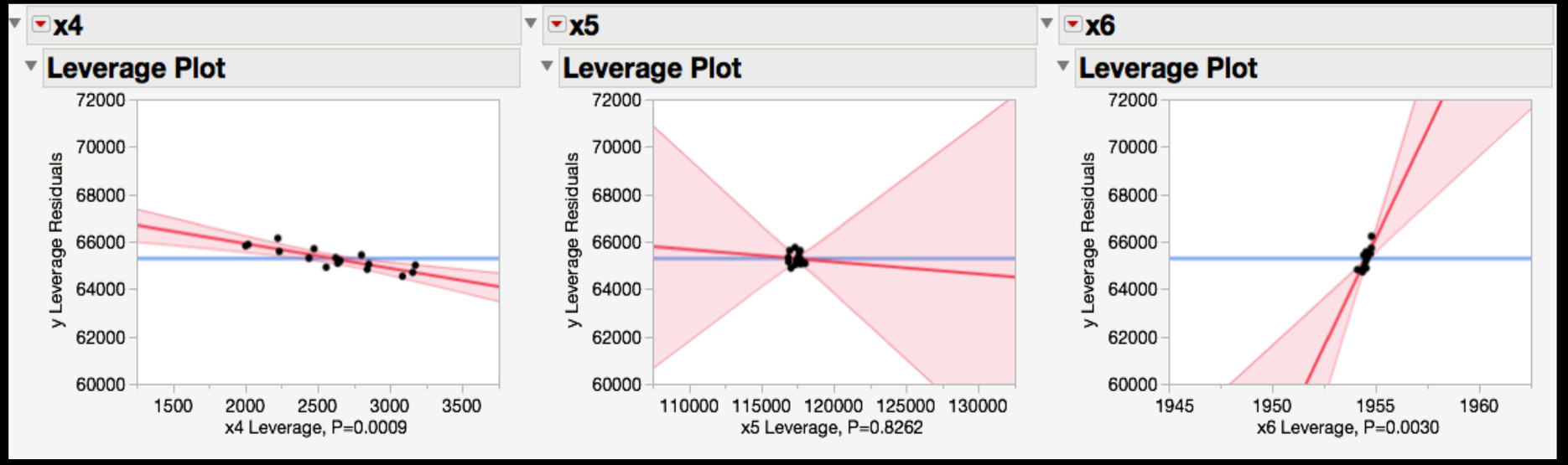
x

Leverage Plot

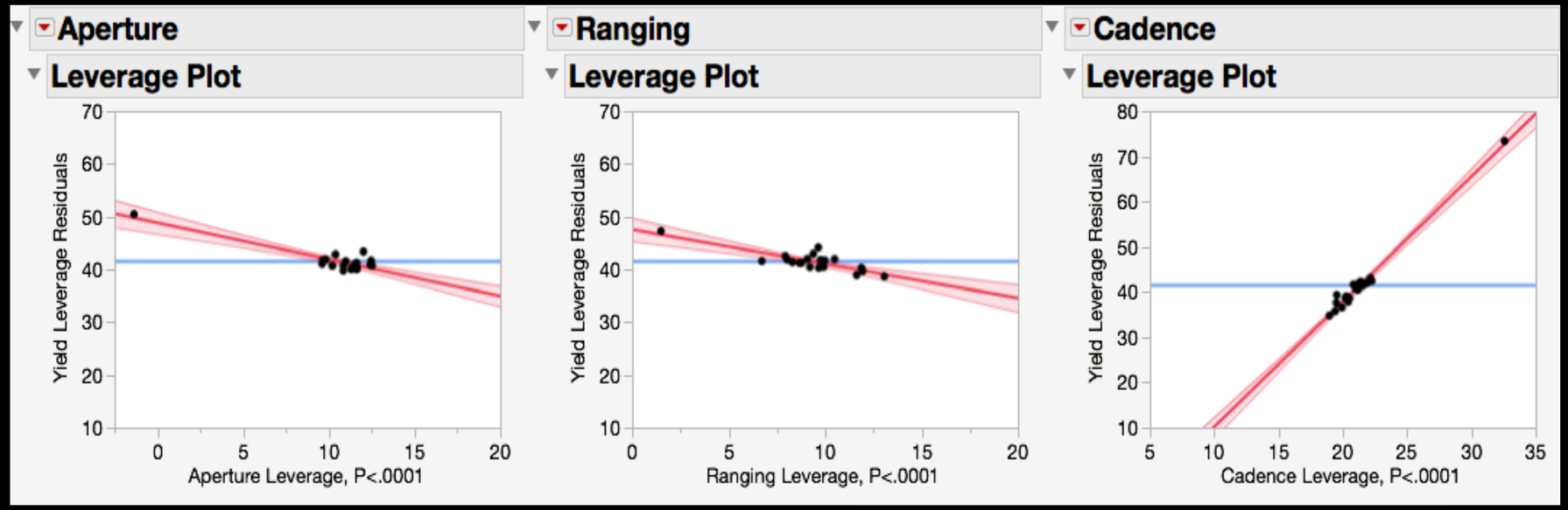


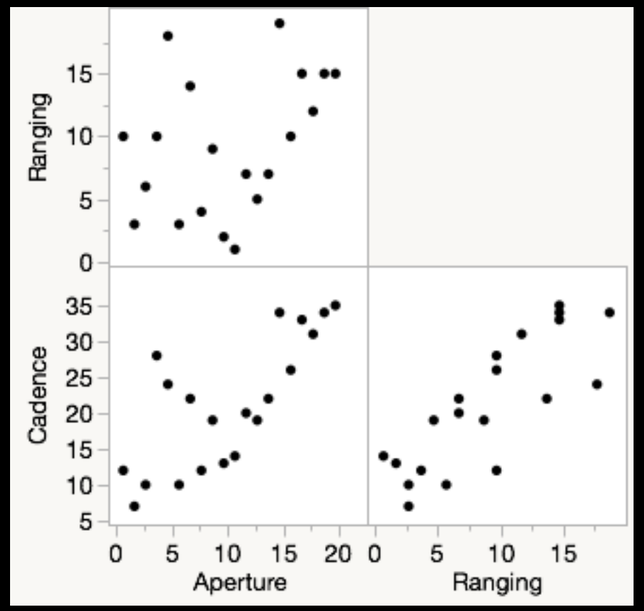
Leverage in Observational Data

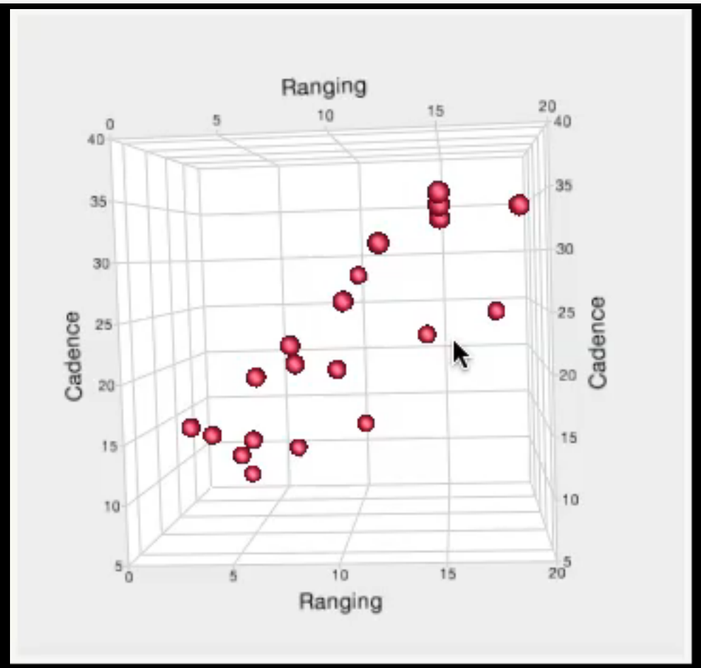




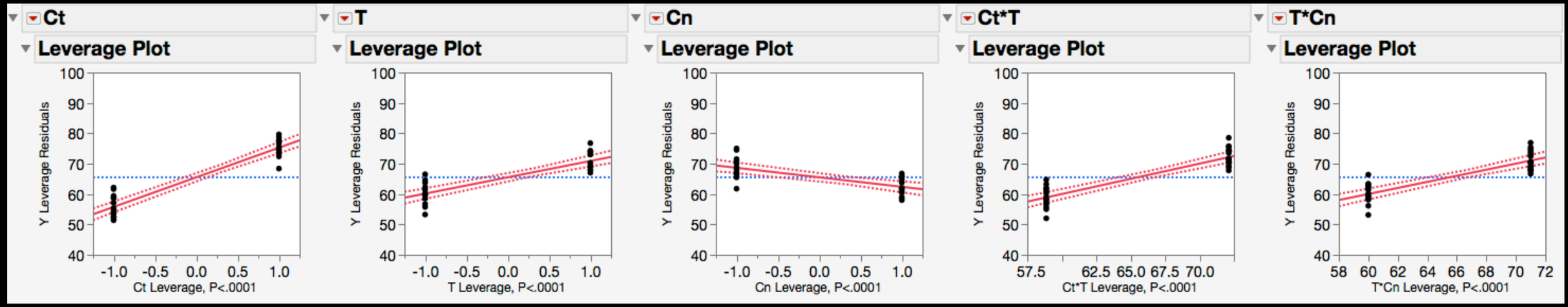
Hidden Leverage







Leverage in DOE



Boring, all points have the same maximum X-leverage.

Published

The American Statistician, November 1990, Vol. 44, No. 4

Leverage Plots for General Linear Hypotheses

JOHN SALL*

Leverage plots are a generalization of partial-regression leverage plots, extending the idea to apply to general linear hypothesis tests. Leverage plots can show the point-by-point composition of the sum of squares for a hypothesis test. They are valuable in revealing the degree of fit, the parameter estimates, the residuals, a measure of the variance of the fit, influential points, nonfitting points, nonlinearities, and even collinearity.

KEY WORDS: Added variable plots; Influence; Partial-regression leverage plots; Statistical graphics.

1. INTRODUCTION

Plotting the raw data for your multiple regression gives you a fairly useful picture but one that might mislead. For an analogy from medicine, consider the case of a doctor who can learn a lot by examining the patient from the outside, but a series of X-ray pictures that show how things look inside is often essential for the best diagnosis.

What the linear-model doctor needs is an inside picture of the data—one that shows how each observation functions

name. Belsley, Kuh, and Welsch (1980), who cited Mosteller and Tukey, matured the idea, calling it the "partial-regression leverage plot." Cook and Weisberg (1982) also developed the concept further and called it an "added variable plot." Chambers, Cleveland, Kleiner, and Tukey (1983) termed it "adjusted variable plots" and credited it to Gnanadesikan. Since then the idea has appeared in the literature extensively under all these and other names.

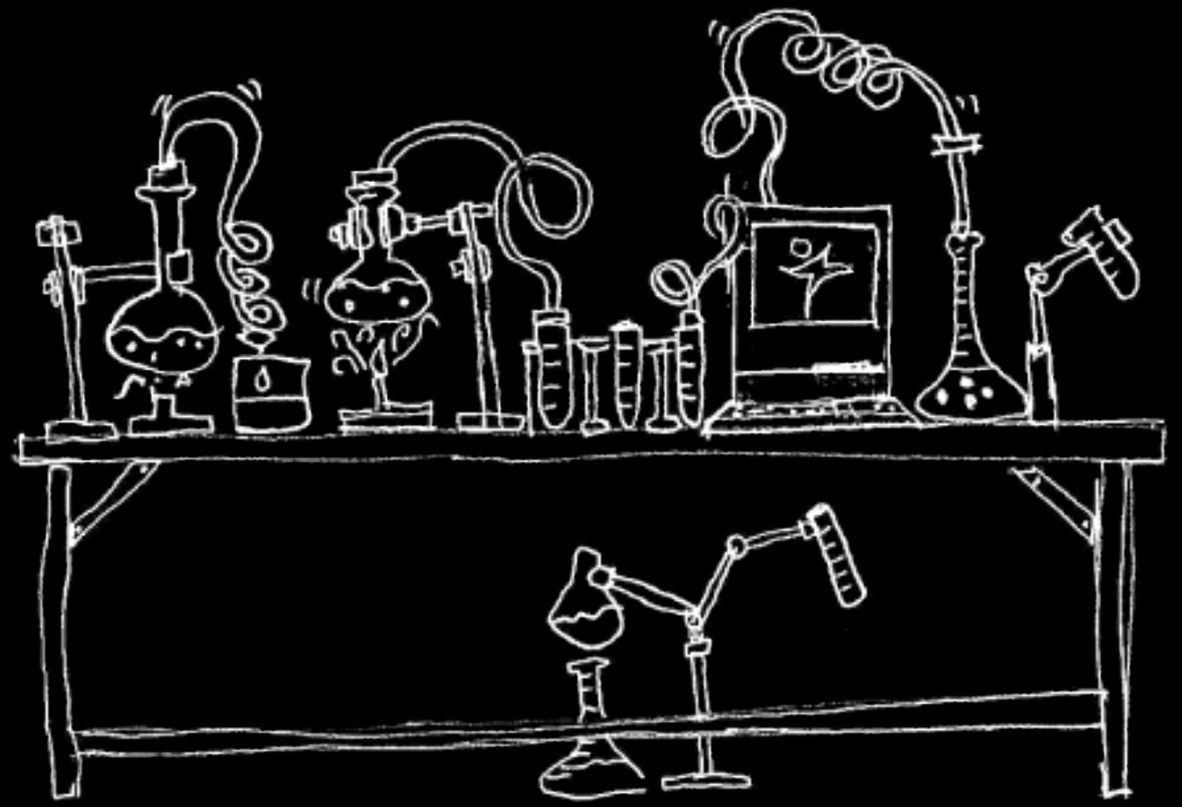
Here the idea is generalized to apply to any linear hypothesis, and I propose the simple term "leverage plot." This term seems to be a suitable shortened form of "partial-regression leverage plot." The term "added variable plot" is a less suitable term to adapt for the extension to general linear hypotheses.

2. LEVERAGE PLOTS FOR GENERAL LINEAR HYPOTHESES

Assume a standard linear model with an intercept term, fixed effects, and iid normal random errors. Suppose that the estimable hypothesis of interest is $L\beta = 0$. We want a plot that characterizes this test by plotting points so that the

http://www.jstor.org/stable/2684358

Design of Experiments

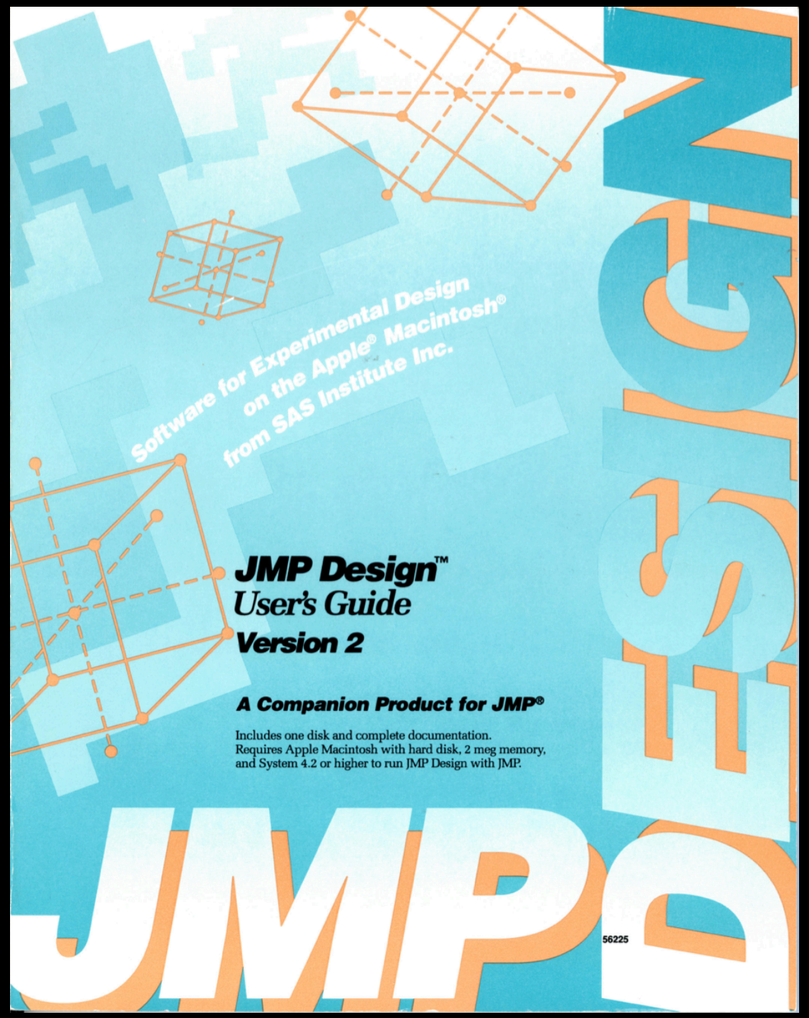


DOE

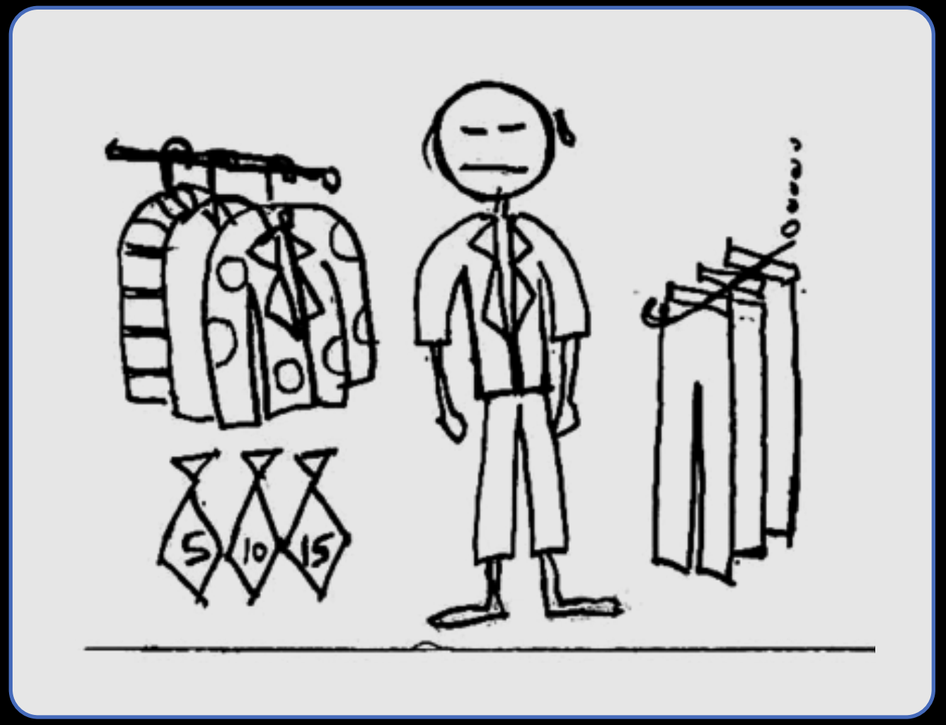
- The key to improving products and processes is trial and error.
- The most efficient trial and error is through designed experiments.
- Learn the most from a given number of experimental runs.

DOE

- We started a separate traditional product to do DOE…
- …but then Brad Jones arrived with fresh ideas.

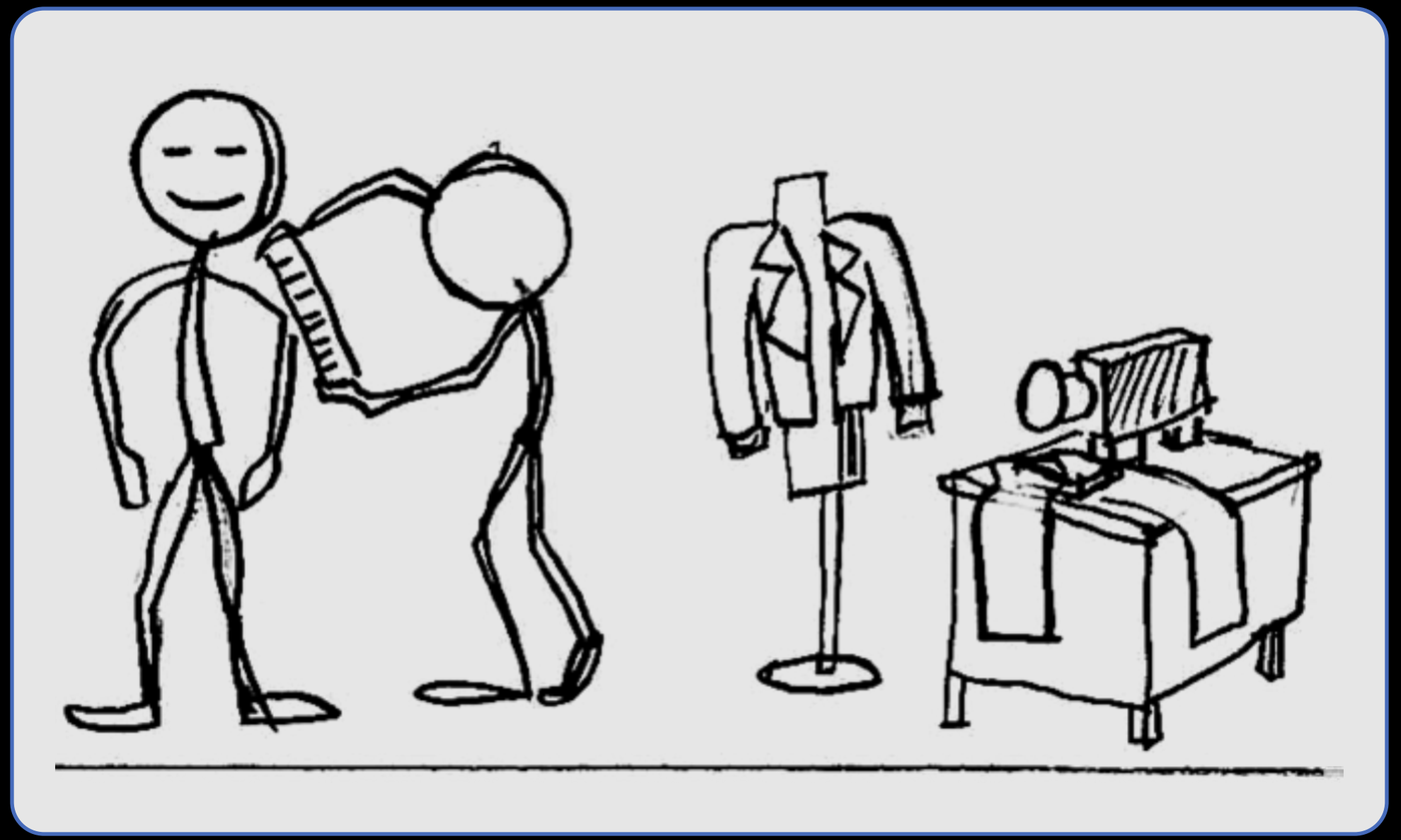


Old DOE



Pick from a design catalog,
name factors, then run.

New DOE



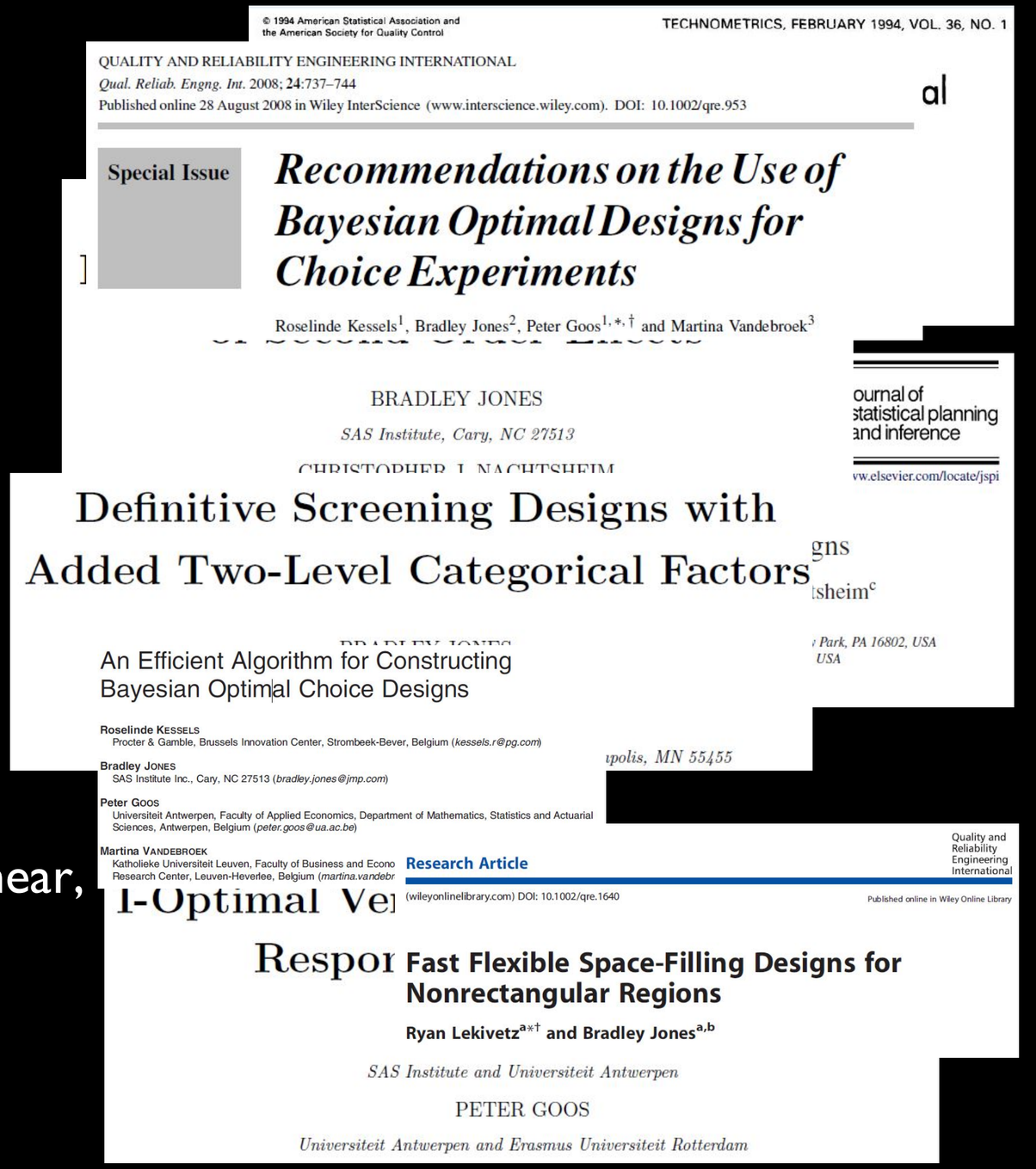
Enter your unique specs,
build a custom design, then run.

DOE

- Focus on the engineer's specific situation and needs.
- Design for the engineer's run budget, not what is in a table.
- Design for what the engineer wants to be estimable.
- Design for any combination of factor types.
- Allow restrictions on the factor space.
- In JMP, Custom Design is the default (optimal design).
- However, not in the standard textbooks by 2000.

DOE

- D-Optimal Design by coordinate exchange
- Bayesian D-Optimal design and supersaturated
- I-Optimal Design for response surfaces
- Split Plot D-Optimal, then Split Plot I-Optimal
- Minimum-aliasing designs
- Definitive Screening Designs
- Specialized Designs (Choice, Spacefilling, Nonlinear, Covering)

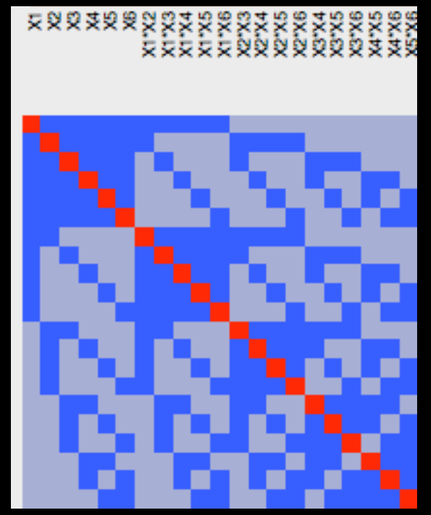


Split Plot

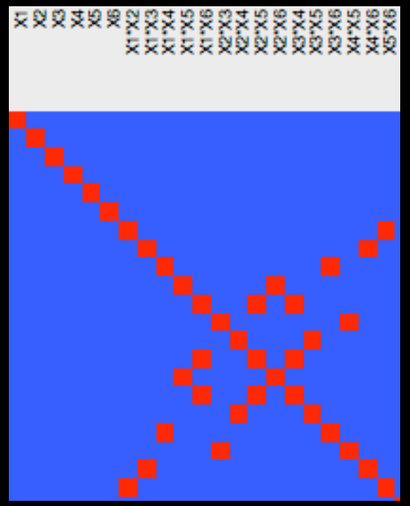
- It is said that most industrial experiments are really split-plot experiments. Some treatments are applied to a group of runs.
- Brad built an optimal DOE designer for Split Plots. Peter Goos had just advanced some of the research in the area.
- But you have to fit the model too. REML was the way to estimate, but how to test hypotheses?
- Kenward-Roger adjustment was invented just in time,1995.
- Chris Gotwalt found the way to calculate it efficiently.
- With small run sizes, components often went negative, but you had to allow this to get the size right. Chris innovated here too.
- The only thing the fitter needs to know is that the whole plot identifier is *random*.
- For experimenters, the situation changed from something impossible to something easy.

DOE

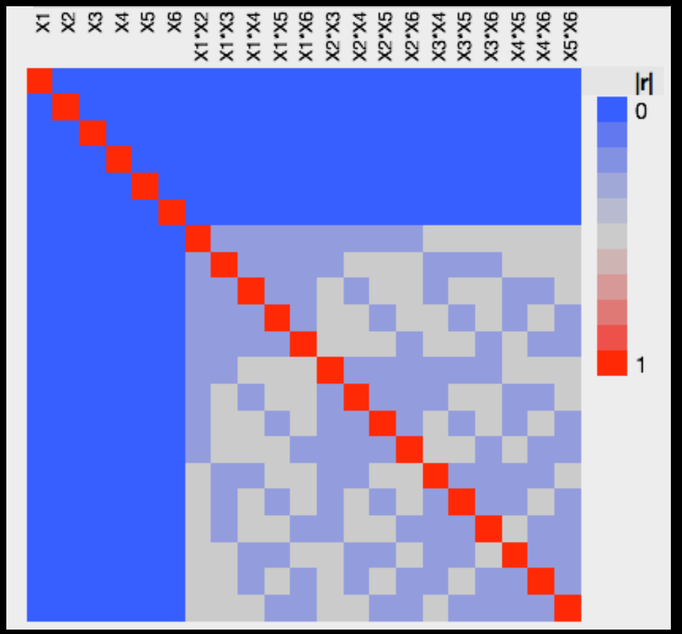
- Historically, small screening designs were limited to two-level main effects. No curvature, no interactions.
- *Definitive Screening Designs* allowed experimenters to fit quadratics and find large interactions for the run budget that you would ordinarily only fit a main-effects screening design with all the interactions confounded.
- And there is more on the way…



12 run PB



16 run FF



13 run DSD

DOE

DOE is in kind of a golden age, where practitioners get the latest technology years ahead of what is in the textbooks.

Listen to Doug Montgomery's plenary talk,
The Flight of the Phoenix

Exploiting the Fit

You have fit a model, then what?

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	37.802855	5.10993	7.40	<.0001*
crim	-0.108011	0.032865	-3.29	0.0011*
zn	0.0464205	0.013727	3.38	0.0008*
indus	0.0205586	0.061496	0.33	0.7383
chas[0]	-1.343367	0.43079	-3.12	0.0019*
nox	-17.76661	3.819744	-4.65	<.0001*
rooms	3.8098652	0.417925	9.12	<.0001*
age	0.0006922	0.01321	0.05	0.9582
distance	-1.475567	0.199455	-7.40	<.0001*
radial	0.3060495	0.066346	4.61	<.0001*
tax	-0.012335	0.003761	-3.28	0.0011*
pt	-0.952747	0.130827	-7.28	<.0001*
b	0.0093117	0.002686	3.47	0.0006*
lstat	-0.524758	0.050715	-10.35	<.0001*

Response Surface Exploration

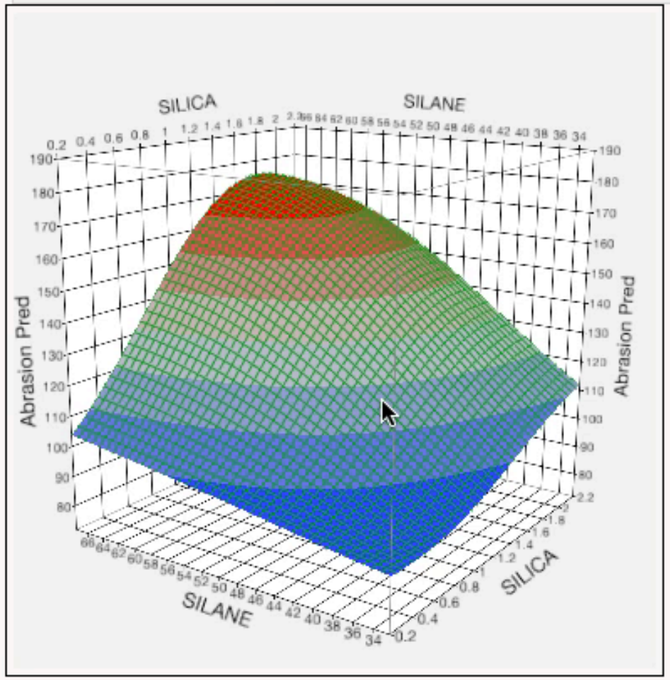
- Understand the slopes, the curvature, the interactions, the sensitivities.
- Optimize and characterize the factor space with respect to the responses.
- Create prediction formulas.
- Sensitivities to factor variation - Simulation

TireTreadGP: Profiler of Abrasion Pred

Profiler

Prediction Profiler

Abrasion Pred



Response Grid Slider

138.96

Independent Variables

X	Y		Value	Grid
◉	○	SILICA	1.155	☐
○	◉	SILANE	51.31	☐
○	○	SULFUR	2.25	

☐ Lock Z Scale

Appearance

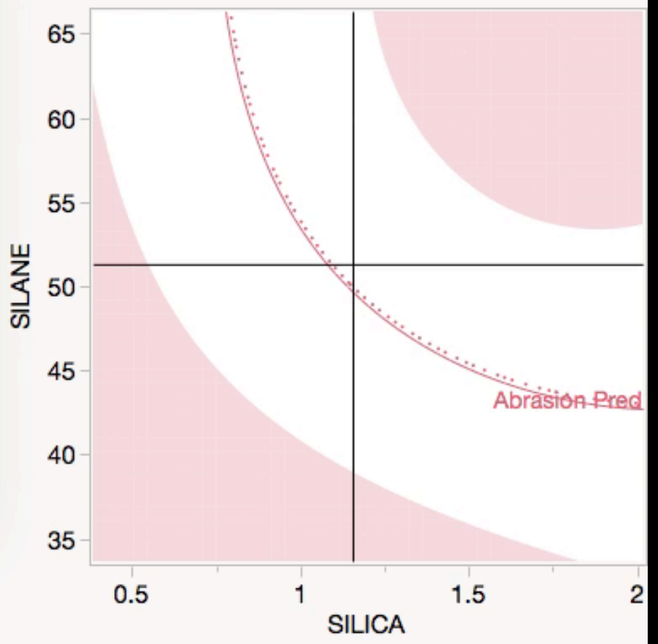
TireTreadGP: Contour Profiler of Abrasion

Profiler

Contour Profiler

Horiz	Vert	Factor	Current X
◉	○	SILICA	1.155
○	◉	SILANE	51.31
○	○	SULFUR	2.25

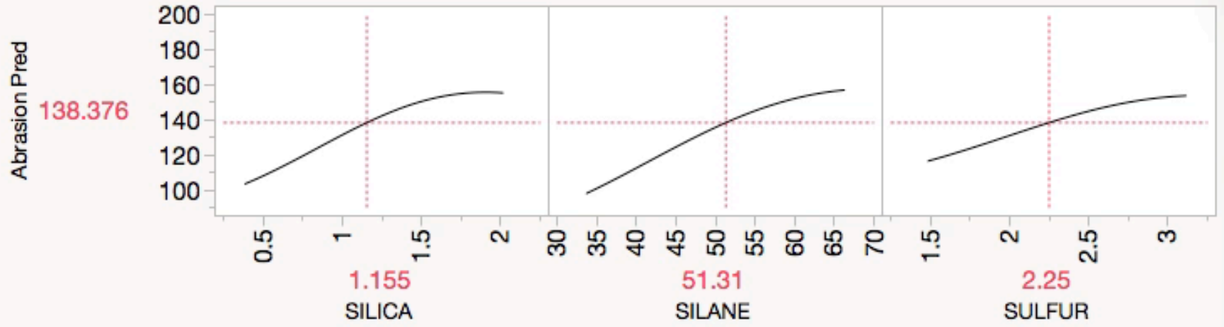
Response	Contour	Current Y
— Abrasion Pred	135	138.37597



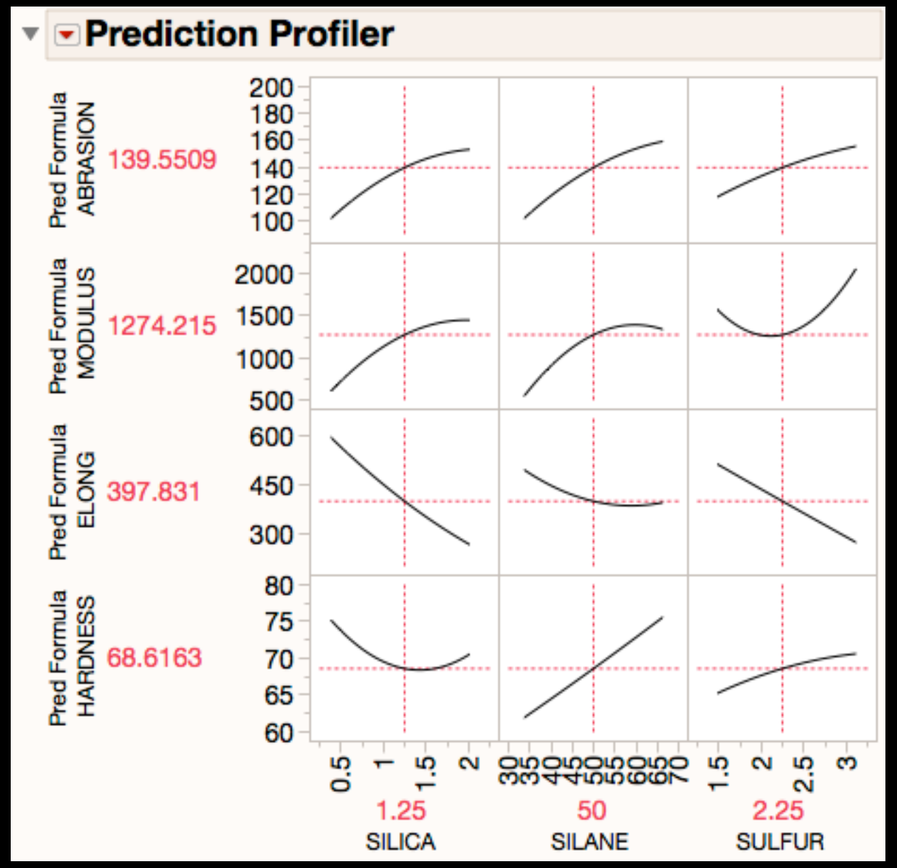
TireTreadGP: Profiler of Abrasion Pred

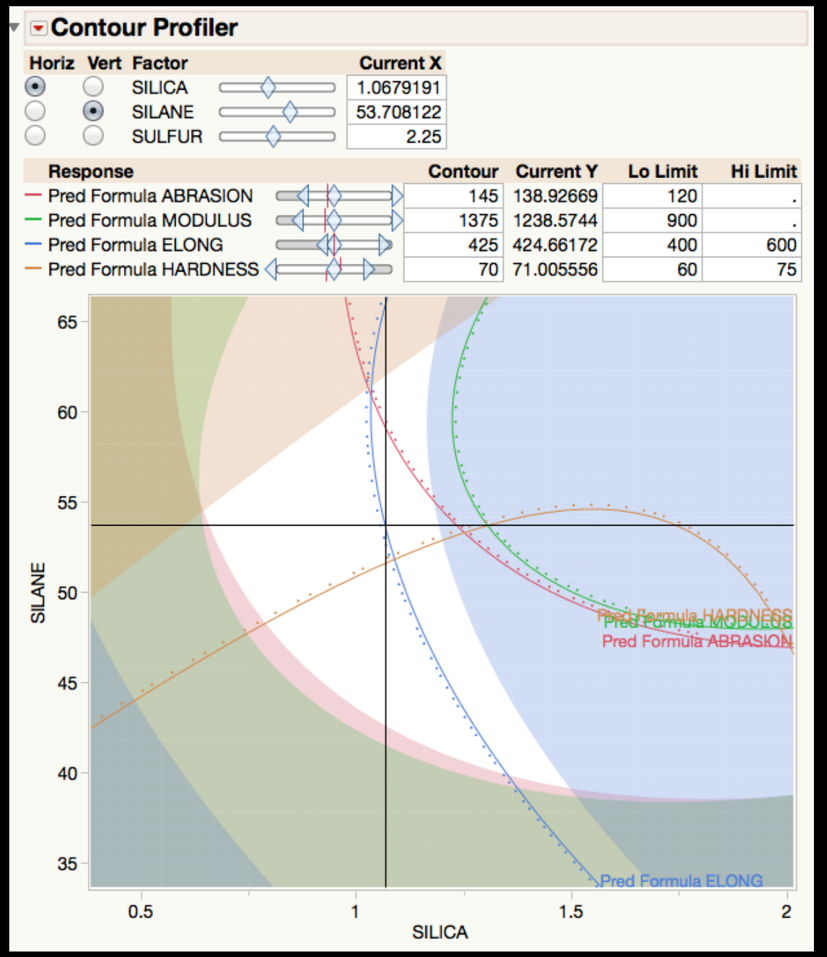
Profiler

Prediction Profiler

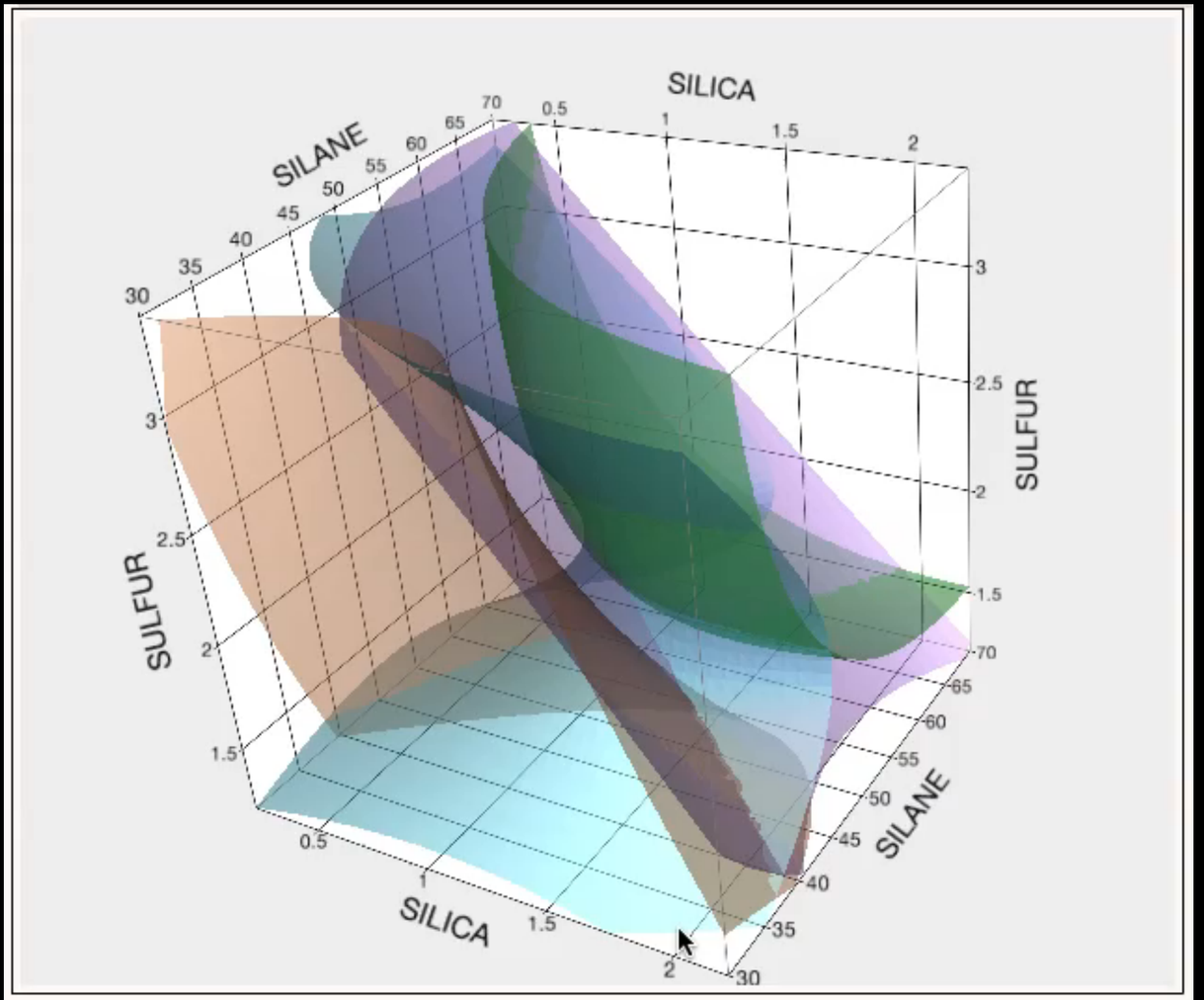


Multiple Cross-Sections





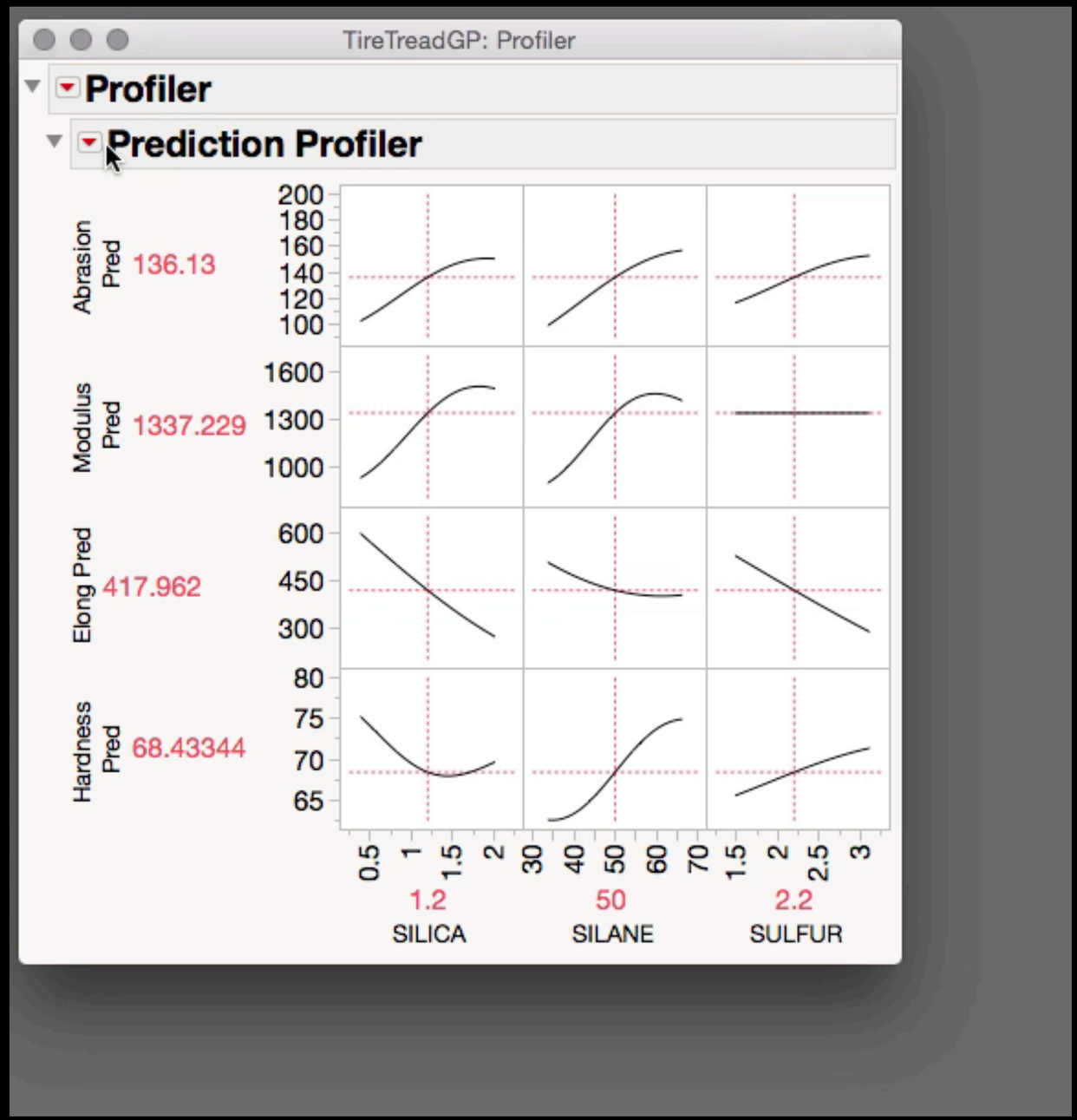
3D Iso-Contours



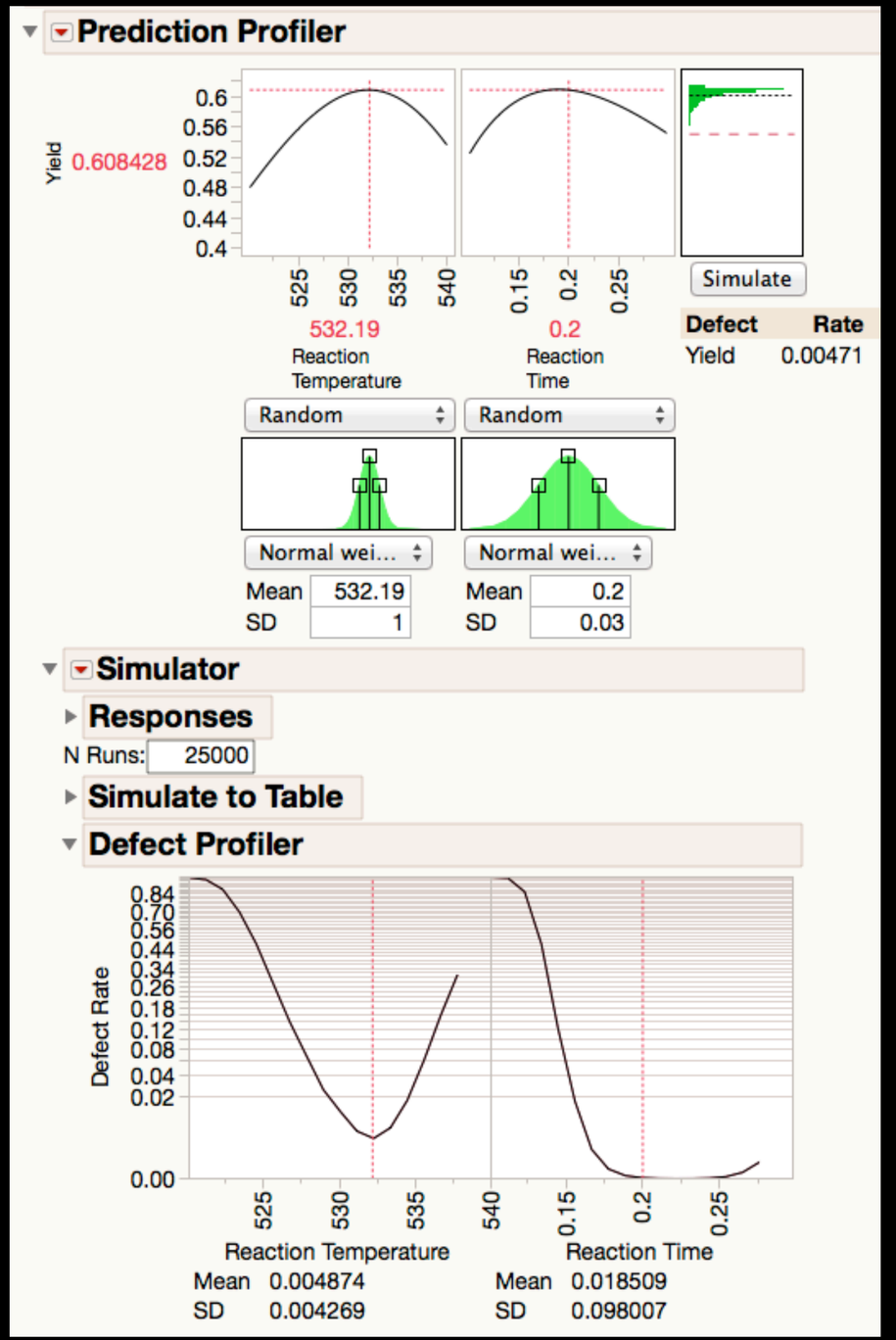
Dependent Variables

Formula		Value		Surface
Pred Formula ABRASION		132		Both sides
Pred Formula MODULUS		1200		Both sides
Pred Formula ELONG		400		Both sides
Pred Formula HARDNESS		65		Both sides

Desirability Optimization



Robust Process Engineering - Simulation



Reaching Out

- Develop a very technically adept sales group
- New users
    - Predictive modeling
    - Consumer research
    - Bigger data
- Internationalization

Localization and Unicode

- We changed to Unicode in JMP 6.
- Programming strings became much better.
- We had a good system to extract and localize.

	english (us)	sex	ID	age	height	weight	فارسی	日本語	中文 (Trad)	中文 (Simp)	tiếng việt	ภาษาไทย	Русский язык	dansk	deutsch
1	ALICE	F	10	13	61	107	آیسا	さと未	佩佩	吾尔开希	Anh Đào	นภัสร	Аделаида	Anna	Angelika
2	AMY	F	29	15	64	112	افشانه	一愛	姍姍	周润发	Anh Thư	อัญชลี	Анна	Anne	Anja
3	BARBARA	F	9	13	60	112	اندیشه	七海	安娜	奇白石	Bắc	ดุษณี	Валентина	Bente	Bärbel
4	CAROL	F	19	14	63	84	اوزن	佳緒	寶珠	张国荣	Bạch Tuyết	ดารณี	Вера	Else	Catharina
5	ELIZABETH	F	17	14	62	91	بهشته	如春	小蘭	张曼玉	Bảo Châu	วรรณา	Зинаида	Hanne	Elke

	english (us)	sex	español	euskera	français	gailge	ελληνικά	íslenska	magyar	norsk	shqip	slovenščina	svenska	suomi
1	ALICE	F	Agüed	Ahuña	Adélaïde	Áine	Αβηιχα	Anna	Apollónia	Berit	Bilbilesha	Alojzija	Anita	Aino
2	AMY	F	Angélica	Adoniñe	Adèle	Bébhinn	Αβροξενα	Ásta	Barbala	Bjørg	Currana	Amalija	Anna	Anneli
3	BARBARA	F	Bárbara	Bilebañe	Anaïs	Bláthnaid	Βαναυσις	Elín	Catalÿn	Ingebj…	Dallënd…	Ana	Birgitta	Aurora
4	CAROL	F	Belén	Burtzeña	Cécile	Dáirine	Γαλατεια	Erla	Caterina	Johanne	Dëlirime	Angela	Christ…	Elina
5	ELIZABETH	F	Begoña	Deiñe	Éléonore	Éibhleann	Ειρηνη	Guðbjörg	Chrÿstina	Jorunn	Erëzake	Antonija	Emma	Emilia

french

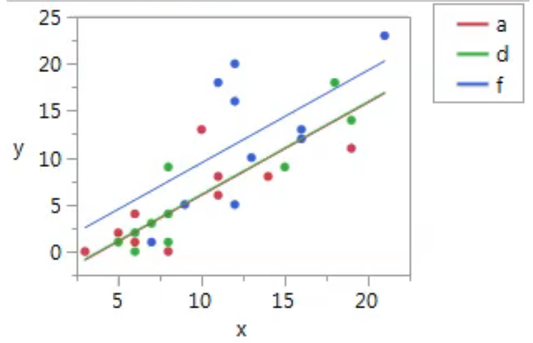
l73787.na.sas.com

Drug - Fit Least Squares - JMP Pro

Response y

Whole Model

Regression Plot



Summary of Fit

RSquare	0.676261
RSquare Adj	0.638906
Root Mean Square Error	4.005778
Mean of Response	7.9
Observations (or Sum Wgts)	30

Analysis of Variance

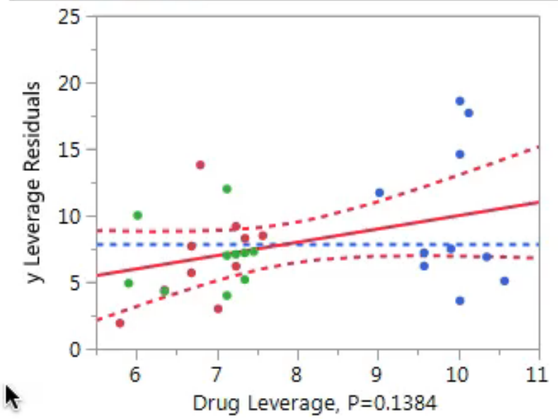
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	871.4974	290.499	18.1039
Error	26	417.2026	16.046	Prob > F
C. Total	29	1288.7000		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	18	254.86926	14.1594	0.6978
Pure Error	8	162.33333	20.2917	Prob > F
Total Error	26	417.20260		0.7507

Drug

Leverage Plot



Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
a	6.714963	1.2884943	5.3000
d	6.823935	1.2724690	6.1000
f	10.161102	1.3159234	12.3000

x

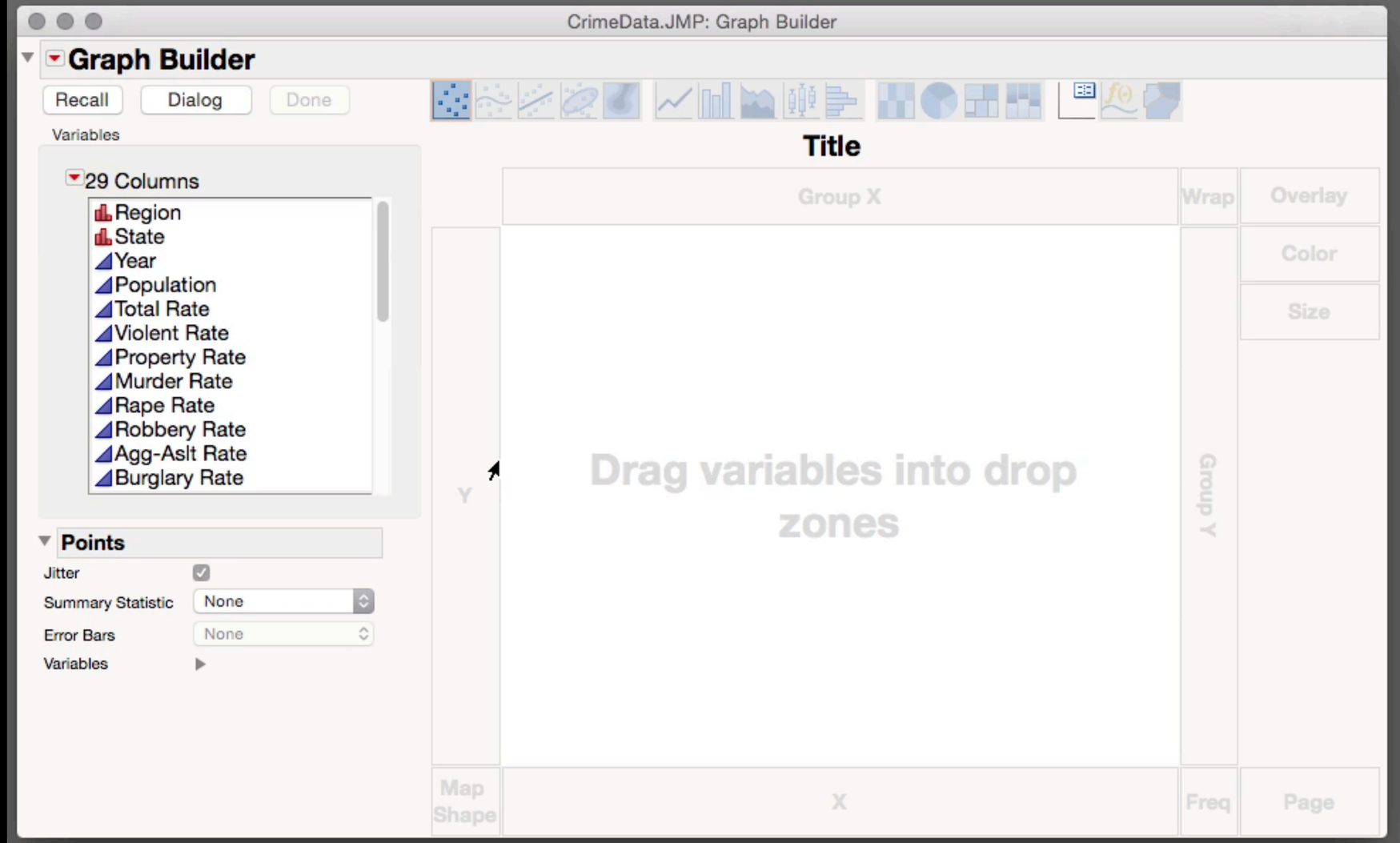
Leve

y Leverage Residuals

Next-Gen Platforms

Drag and Drop

- No launch dialog – just drag a column into a role target.
- Tabulate was the first platform to embrace a drag-and-drop approach.
- Graph Builder did it in an especially rich and immediate way.


CrimeData.JMP: Graph Builder
Graph Builder
Recall
Dialog
Done
Variables
29 Columns
Region
State
Year
Population
Total Rate
Violent Rate
Property Rate
Murder Rate
Rape Rate
Robbery Rate
Agg-Aslt Rate
Burglary Rate
Points
Jitter
Summary Statistic
None
Error Bars
None
Variables
Title
Group X
Wrap
Overlay
Color
Size
Y
Drag variables into drop zones
Group Y
Map Shape
X
Freq
Page

Scripting

JMP Scripting Language

- Originally, we banned programming interface, thinking it would corrupt us from supporting point and click.
- … But we needed a way to save an analysis to redo it another day … and a way to loop … and a way to extend JMP, to do things that were not built into the product.
- So in JMP 4 we introduced JSL.
- High-level data types, object-sending to mimic the interactive features.

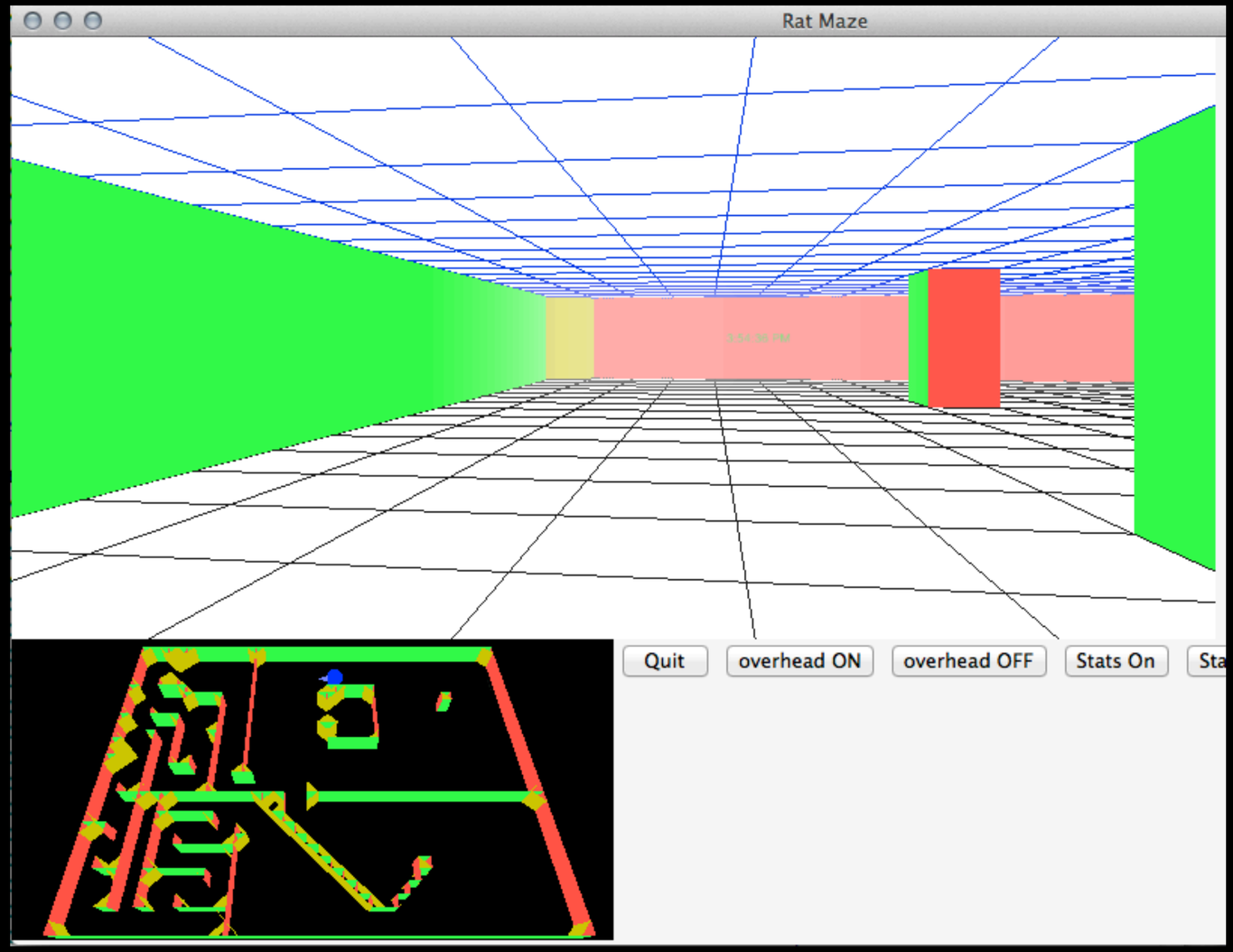
Craige Hales tested patterns by writing a Fortran interpreter to run the Adventure game.

"… a maze of twisty little passages"

wikipedia.org/wiki/Colossal_Cave_Adventure

```
Welcome to Adventure!!  Would you like instructions?

>yes

Somewhere nearby is Colossal Cave, where others have found fortunes in
 treasure and gold, though it is rumored that some who enter are never
 seen again.  Magic is said to work in the cave.  I will be your eyes
 and hands.  Direct me with commands of 1 or 2 words.  I should warn
 you that I look at only the first four letters of each word, so you'll
 have to enter "NORTHEAST" as "NE" to distinguish it from "NORTH".
 (Should you get stuck, type "HELP" for some general hints.  For infor-
 mation on how to end your adventure, etc., type "INFO".)
                        - - -
 This program was originally developed by Willie Crowther.  Most of the
 features of the current program were added by Don Woods (DON @ SU-AI).
 The current version was done by Kent Blackett and Bob Supnik.  It is
 DECUS Program 11-340.
                        - - -
 The rehost to a PC, under Microsoft Fortran, was done by Ken Plotkin.
 See file PCADVENT.DOC for details.
                        - - -
 This is a port of the FORTRAN code to JSL, March, 2009.
                        - - -

 You are standing at the end of a road before a small brick building.
 Around you is a forest.  A small stream flows out of the building and
 down a gully.

>go in

You are inside a building, a well house for a large spring.
 There are some keys on the ground here.
 There is a shiny brass lamp nearby.
 There is food here.
 There is a bottle of water here.

>take food

Ok

>quit

Do you really want to quit now?

>yes

Ok

You scored  27 out of a possible 350, using    3 turns.
You are You are obviously a Rank Amateur.  Better luck next time.
 To achieve the next higher rating, you need  9 more points.
```

Craige tested sockets in JSL by first writing a Web browser, then adapting a 3D Rat Maze game to be multi-user across a network.



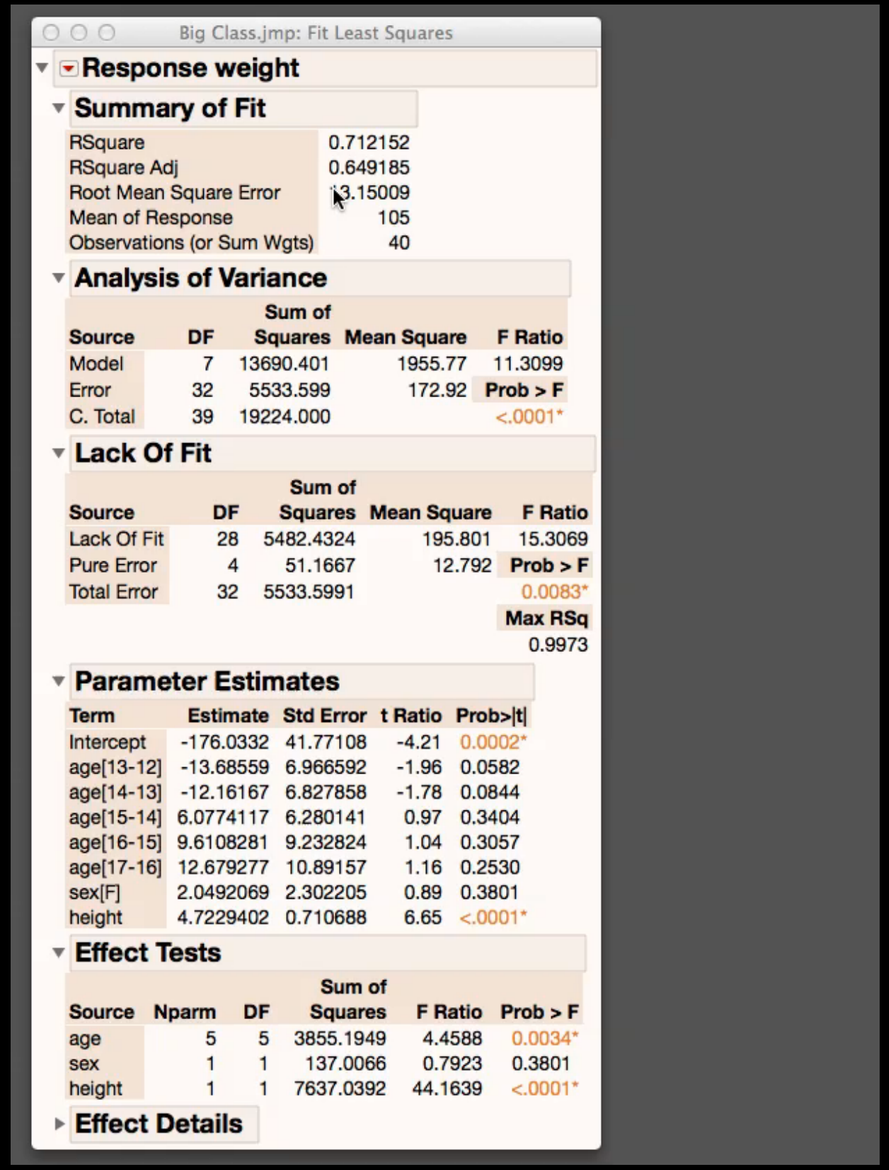
Scripting Advances

- Associative arrays, patterns, sockets, blobs, images
- Script editor
- Name spaces
- JSL debugger
- JSL Performance Profiler
- Add-ins and the JMP File Exchange
- Expression data type in data table columns

Short Topics

Many Helps

- Help menu
- JMP Starter
- Menu item tooltips
- Question cursor/tool
- Help buttons
- Circle hover help
- Books
- Script Index
- Statistics Index
- Tutorial
- community.jmp.com



Redo as an Enabler

Some valuable features are almost free

- Data Filter and Local Data Filter
- Column Switcher
- Bootstrap
- Future features…

Possible Challenges

- “The desktop is dead.” – Client-server/Web/Cloud
- Production batch runs.
- “Software should be free.” – Open source
- Differentiating from SAS in the market.
- “It has to be in the textbooks.”

Concluding Remarks

How to Think

In statistics we used to think like a lawyer: We know the result and we just have to prove it.

In statistics we now need to think like a detective: We don't already know, and we need to discover new things about our data.

Serendipity Discovery

- The easier it is to look at things, the more ways you will look, and the more things you will find.
- If you feel lucky, you will open your eyes to notice things, and you will be lucky. - Richard Wiseman