Session G3, Tuesday, 3/3, 1:30-2:30 pm, Level: Moderate

Friend or Foe? How to Use Graphical Diagnostics for Scoping Out Discrepant Data

Presented by Mark J. Anderson, PE, CQE, MBA
(Email: mark@statease.com)
Co-authored by Patrick J. Whitcomb

"Errors, like straws, upon the surface flow; He who would search for pearls must dive below."
- John Dryden (1678), All for Love
(from Brian Joiner's "Lurking Variables: Some Examples," *The American Statistician*, Nov 1981)

Disclosure of Biases

- Focus will be on results from designed experiments, not happenstance regression.
- How likely is it that an experimenter will hit every level on every run and collect every response precisely? Our view: *"A pessimist is only an optimist with experience."*
- Sometimes it's in the eye of the beholder as to whether an outlier is viewed as good (you learn something) or bad (not what was expected). It's like the proverbial glass being either half full or half empty – depending on whether one is an optimist or a pessimist. Of course a lean six sigma practitioner, being a "why's" guy, would complain about half the glass being a waste of space – the container is twice as big as it needs to be!
- "Outliers are also known by other names: maverick, flier, straggler or aberrant value." * Hmm, maverick flier – whom does that bring to mind? Let's keep political biases secret!

*(Thomas Murphy & Alex T. Lau, "Dealing with Outliers," *ASTM Standardization News*, Nov/Dec 2008, pp 22-23.)

Learning Outcomes

- Understanding that an outlier is simply a response from an experiment that does not fit the proposed model.
- Becoming wary of immediately deleting discrepant data without first diagnosing the normality of model residuals.
- Seeing how simple transformations such as the log can remedy abnormal residuals
- Achieving a statistically sound view of discrepant data via graphical tools.

Thorny Issues!

How to maintain reasonable balance between two types of errors:

1. Focusing on data that vary only due to common causes, thus introducing bias.
   Anecdotal examples:
   + Positive bias: Cold fusion scientists see what they want,
   – Negative bias: Pig's Eye engineer ignores the unwanted.
2. Overlooking true outlier(s) (special cause),
   - Obscuring real effects (*case later*) or
   - Drawing false conclusions and/or
   - Losing chance in future to:
     - ✓ Prevent failure (*ex. ozone hole*) or
     - ✓ Reproduce breakthrough improvements (discoveries made by accident).

"Stuff" Happens!

- Typographical erors (oops!). *Suggestion: enter data top down, proof-read bottom up*.
- Breakdowns in equipment
- Mistakes by operators
- Non-representative samples
- Bad measurements
- Unknown lurking variables that appear only intermittently

Has any stuff like this ever happened to you?

Two Types of Error

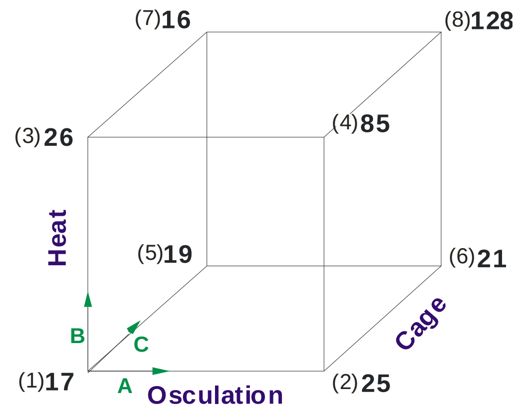
Stuff does not always happen!

Outliers(s)?		What you say: Yes	What you say: No
The truth:	Yes	Correct	False Negative
The truth:	No	False Positive	Correct

Let's look at case studies illustrating both errors.

The Secret To Long Life!

Box's protégé makes bearing breakthrough, but
is it too good to be true?



Bearing Case – Background

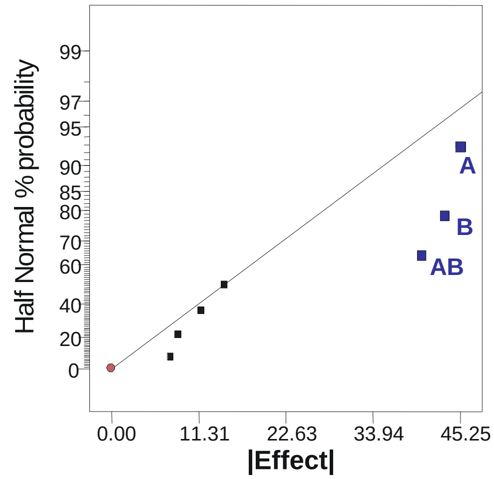
Swedish manufacturer SKF invented rolling bearings in 1919, but by the '70s the Japanese had achieved comparable quality at competitive prices. These pressures inspired SKF engineers (led by Christer Hellstrand, a student of DOE guru George Box*) to quit doing experiments only one factor at a time and try a two-level factorial design.

You will be amazed by the results!

* "George's Column," *Quality Engineering*, Vol. 2, No. 3, p365.

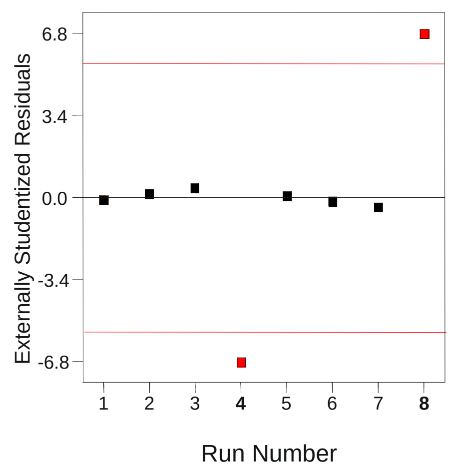
Bearing Case: Half-Normal Plot of Effects

What's wrong with this picture?



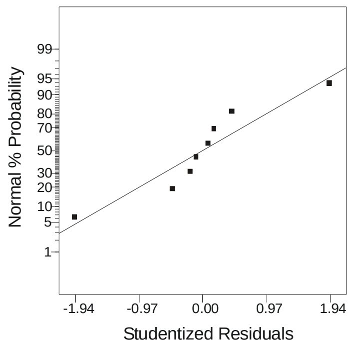
Bearing Case: Apparent Outliers

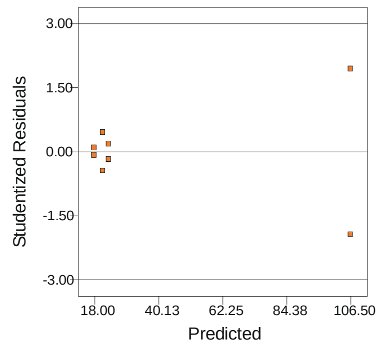
Toss the outliers? Better check first to ID them (go back to cube plot for run #'s).



Bearing Case:
Further Diagnostics

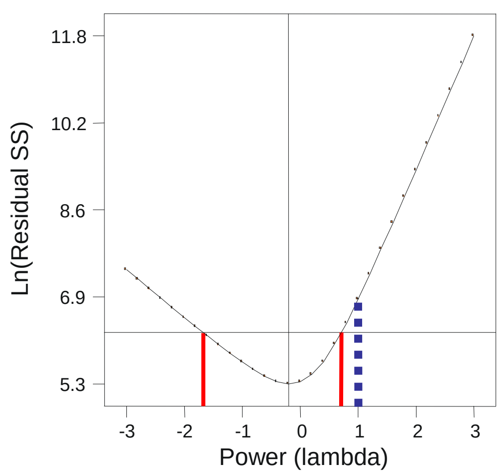
Is assumption correct that residuals are normally distributed with constant variance?





Bearing Case:
Box-Cox Plot

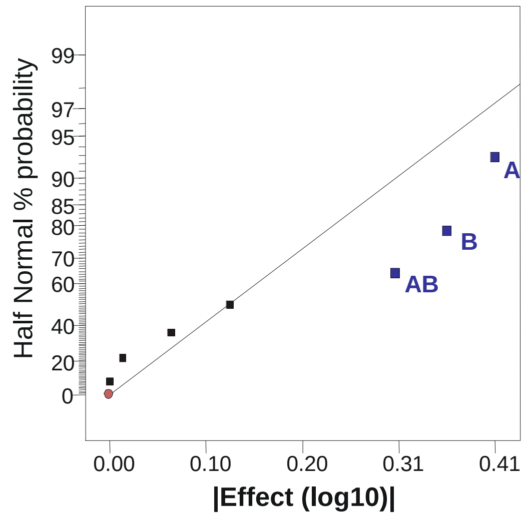
Blue line is outside of red zone



☞ *Transform will help - try log.*

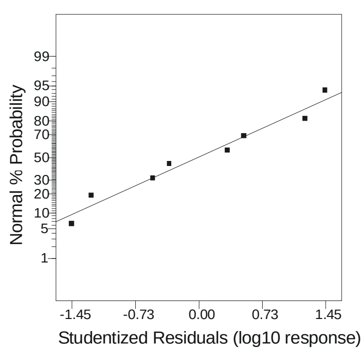
Bearing Case (*after* *log* *transformation*): Plot of Effects

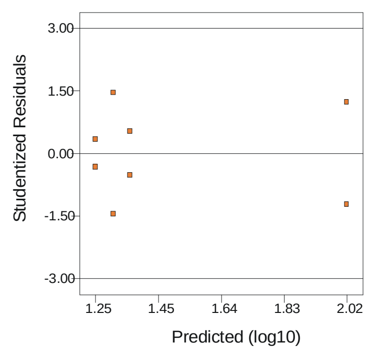
Nice line-up!



Bearing Case (*after* *log* *transformation*): Residual Diagnostic Plots

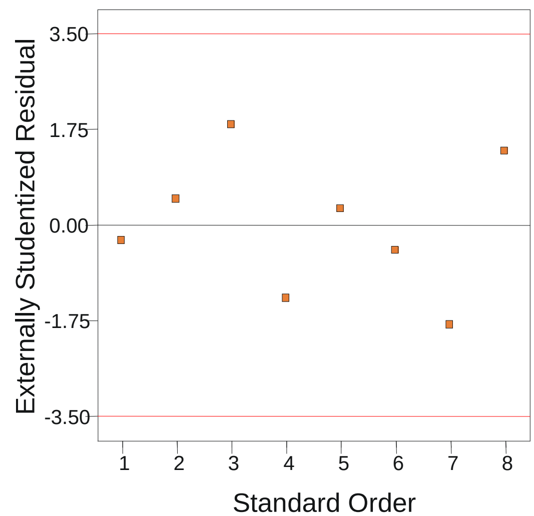
Normally distributed with constant variance!



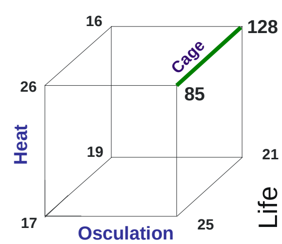


Bearing Case (*after log transformation*): Plot for Detecting Outliers

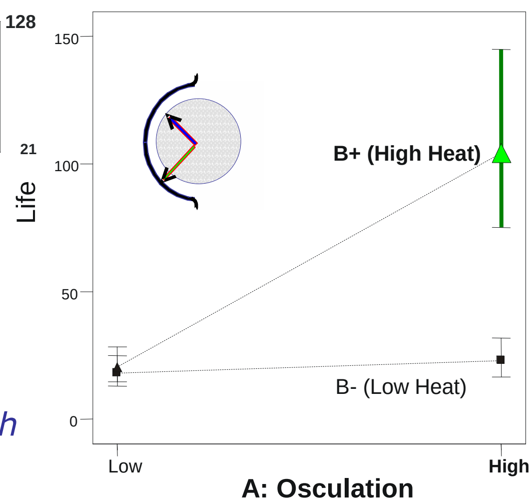
Can include the fab results! None are outliers.



Bearing Case (*after log transformation*): Interaction Plot *(the happy ending!)*



Notice how confidence (LSD) bar widens at breakthrough conditions.



Ball Bearings
Grand Finale

A small design can work
– if you are both bold <u>and</u> lucky!
(and do not throw out the breakthrough results)

Ultimately SKF improved their actual bearing life
from 41 million revolutions on average
(already better than any competitors),
to 400 million revs* –
nearly a ten-fold improvement!

*("Breaking the Boundaries," *Design Engineering*, Feb 2000, pp 37-38.)

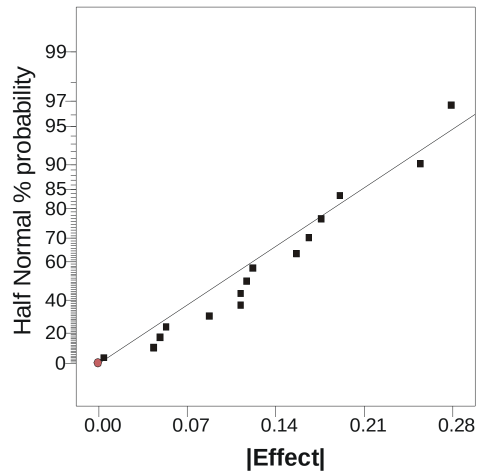
A Case to Test Your Metal

Defect rates on die-cast aluminum disk-drive housing drop dramatically, but is it real?

Run (Std)	A Oil Temp	B Trip	C Metal Temp	D Shot	E Dwell	Fraction Defect
1	350	390	1260	1.60	5.50	0.14
2	450	390	1260	1.60	3.50	0.98
3	350	410	1260	1.60	3.50	0.36
4	450	410	1260	1.60	5.50	0.42
5	350	390	1300	1.60	3.50	1.00
6	450	390	1300	1.60	5.50	0.90
7	350	410	1300	1.60	5.50	0.28
8	450	410	1300	1.60	3.50	0.14
9	350	390	1260	2.20	3.50	0.22
10	450	390	1260	2.20	5.50	0.26
11	350	410	1260	2.20	5.50	0.38
12	450	410	1260	2.20	3.50	0.12
13	350	390	1300	2.20	5.50	0.30
14	450	390	1300	2.20	3.50	0.06
15	350	410	1300	2.20	3.50	0.22
16	450	410	1300	2.20	5.50	0.38

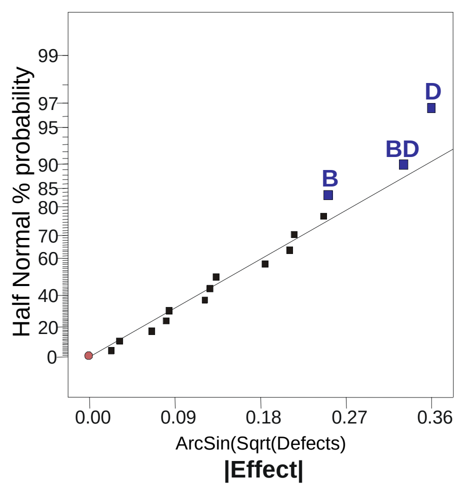
A Case to Test Your Metal
Normal Plot of Effects

- All effects line up so evidently nothing is significant. ☹
- Don't give up yet – try transform: arc-sin square root good for binomial data like this.



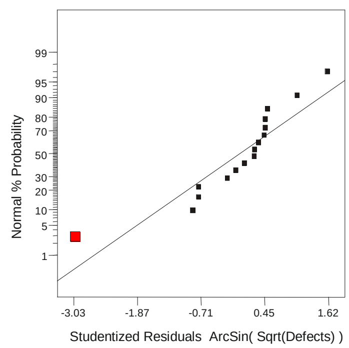
A Case to Test Your Metal
Normal Plot of Effects (*transformed*)

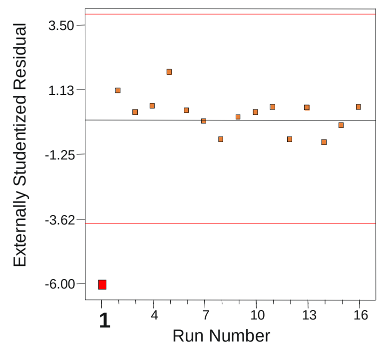
- All effects still line up so arc-sin square root no help. ☹ ☹
- Don't give up yet – pick several of the biggest effects and look for outliers.



A Case to Test Your Metal
Diagnostic Plots

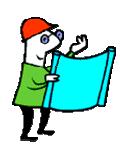
Problem revealed: Obvious outlier! Special cause?





A Case to Test Your Metal
Investigation of Outlier

Diagnostic plots ID'd run number 1 as a discrepant outcome. The foreman, when confronted with this statistical evidence, broke down and confessed that his crew overlooked this particular combination of factors. They then tried to make up for it by coming in early the following week, after shutting down the foundry over the weekend, to sneak the missing run in before the engineer came in to work.

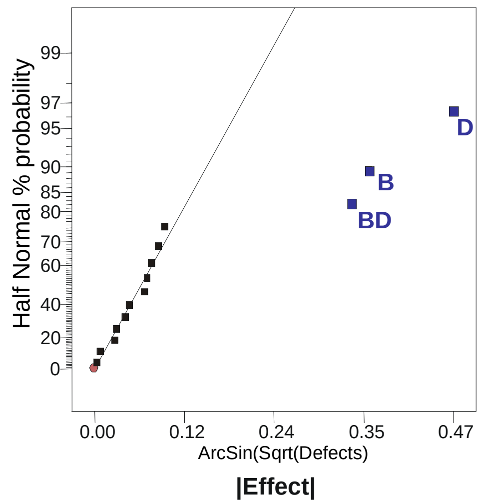


Next step: delete this run and re-analyze.

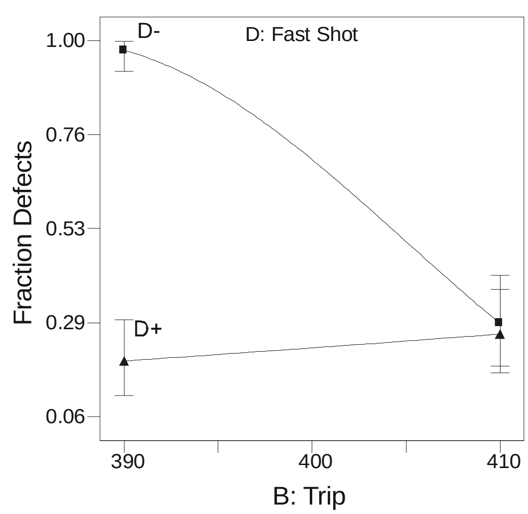
A Case to Test Your Metal *(outlier removed)*
Normal Plot of Effects

Family of effects now clearly significant. ☺

(Warning: missing data causes information loss, but not serious in this case.)



A Case to Test Your Metal *(outlier removed)*
Interaction Plot *(the happy ending!)*



Answer now obvious: Do not run both B & D low.

Conclusions

An outlier is a response from an experiment that does not fit the proposed model. Before jumping to any conclusions, consider these possibilities:

1. The model is faulty, not the data *(Ex. bearing case)*.
   - Watch for bad patterns on diagnostic plots: normal, residuals versus predicted, outlier (externally studentized) and Box-Cox.
   - Try transformation such as a log, square root (for counts), inverse (for rates) or arcsin square root (for fraction defects).

Conclusions

2. The result really is an outlier *(Ex. die-cast aluminum)*.
   - Look for possible errors in data entry, or in response measurement, or in the conduct of that particular experimental run.
   - If the response really differs at that particular combination of the design factors, further study may lead to an important discovery!

Above all – avoid bias:

> *"The first principle is that you must not fool yourself--and you are the easiest person to fool."* -- *Richard Feynman*

References

For more detail and references, see the basis for this talk:

"How to Use Graphs to Diagnose and Deal with Bad Experimental Data," by Mark Anderson & Patrick Whitcomb, *Quality Engineering*, April 2007.

Thanks for listening!
-- *Mark*

Mark J. Anderson
mark@statease.com