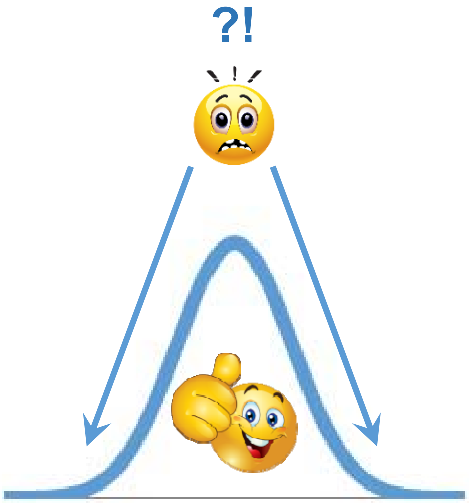
Outing the Outliers
or
"Tails of the Unexpected"

ICEAA 2016 International Training Symposium
Bristol, 17th to 20th October 2016

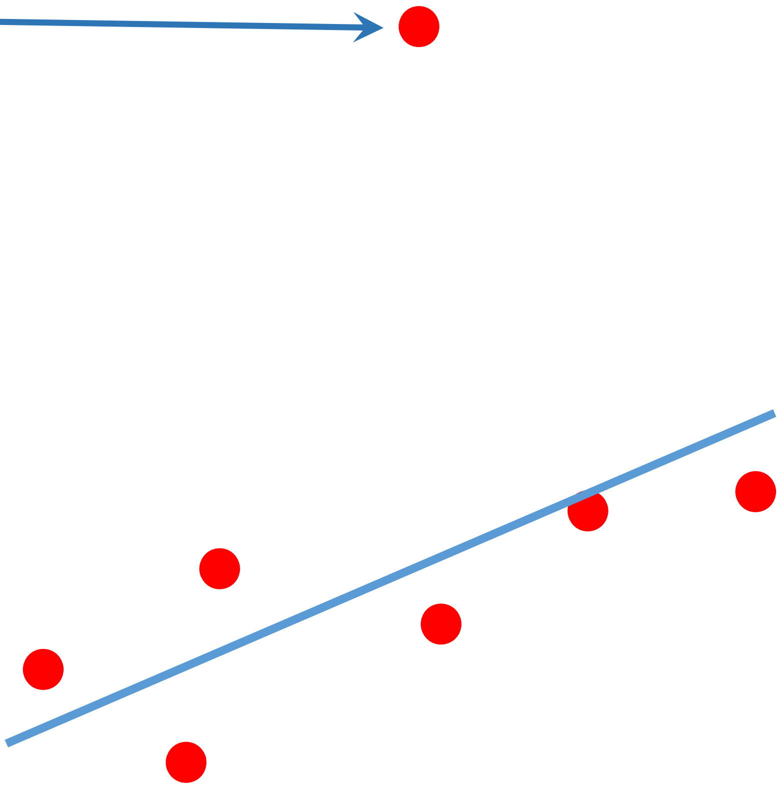
Alan R Jones
Estimata Limited
Promoting TRACEability in Estimating



Outing the Outliers: Agenda

- Outliers: Are we bothered?
  - What do we mean by an “Outlier”?
  - How can we spot one?
- Tukey Fences
  - Traditional and Slimline Fences
- Chauvenet’s Criterion
  - Traditional and Revised SSS Criterion
- Iglewicz and Hoaglin M-Score
  - A MAD Method
- Grubbs’ Test (The definitive test?)
- Doing the J-B Swing
- Summary of Other Tests
  - Not covered in detail
- Summary

You probably won’t recognise those in Red which are offered up for your consideration



Outliers: Are We Bothered?

What do we mean by an “Outlier”?

How can we spot one?

Why should we be bothered by Outliers?

Is it good practice or bad practice to eliminate outliers?

- Some say “Good”; some say “Bad”
- Personally, I’m in favour, … so long as we make an informed decision
- The estimate is the precursor to successful project cost control

  … and a poor estimate leads to challenging cost control

  … or simply no contract award, and therefore no costs to control
- An Outlier can create a bias in the analysis of our data

  … Inappropriate inclusion of data that should be regarded as an outlier may leave a project as unachievable or uncompetitive

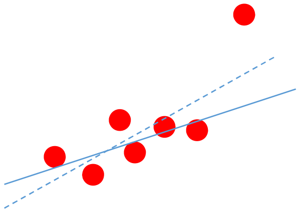
  Similarly, inappropriate exclusion of data as an outlier (when it is not) may also have consequential effects on our ability to deliver projects successfully
- Let’s try to get it right from the beginning

  … history shows that we’ll have plenty more chances to get it wrong again afterwards!
- So, let’s “out” the outliers … but the question is “How?”

Why should we be bothered by Outliers?

Do we think we might have an Outlier or two?

- Assuming we have done the sensible things first i.e. made sure that we:
  - Don't have any cuckoos in the nest so that we are comparing "like with like"
  - Have normalised our data to take account of transient variables like time, scalar difference etc
- The first question to ask is: "*Does it matter if we have an outlier amongst our data?*"

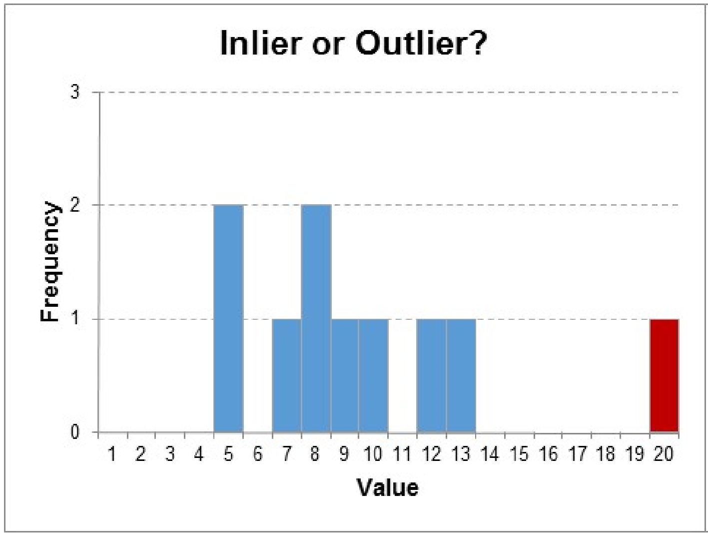


  - If our estimate driver value points towards the "central" values in our data (around the Mean or Median), then we can often get away with just a sensitivity analysis
  - Run the analysis with and without the potential outlier to test the difference it makes
- If it makes a big difference, or we are looking at an estimate to the left or right of our data pot rather than in the central area, then the question we must next resolve is:

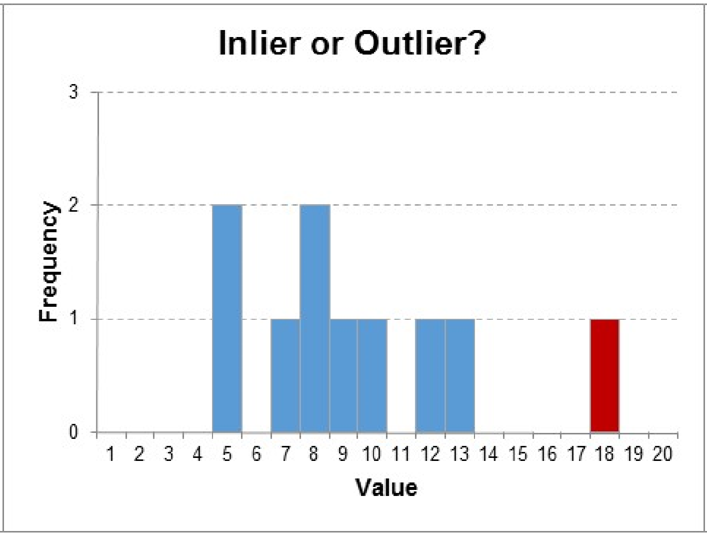
"*Just how far out from the crowd must a data point be, to be classed as an outlier?*"

Spot the Outlier (Audience Participation Required)

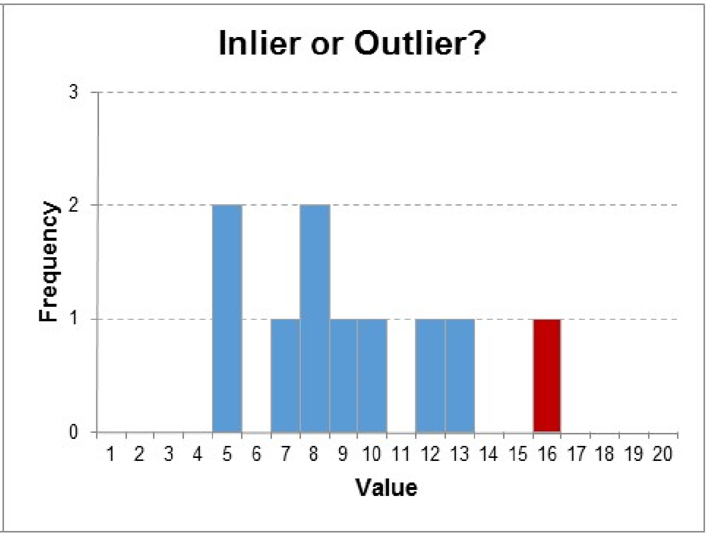
Based on "Gut Feel" you have to decide, do we keep it, exclude it, or toss a coin?



Outlier?
Exclude it



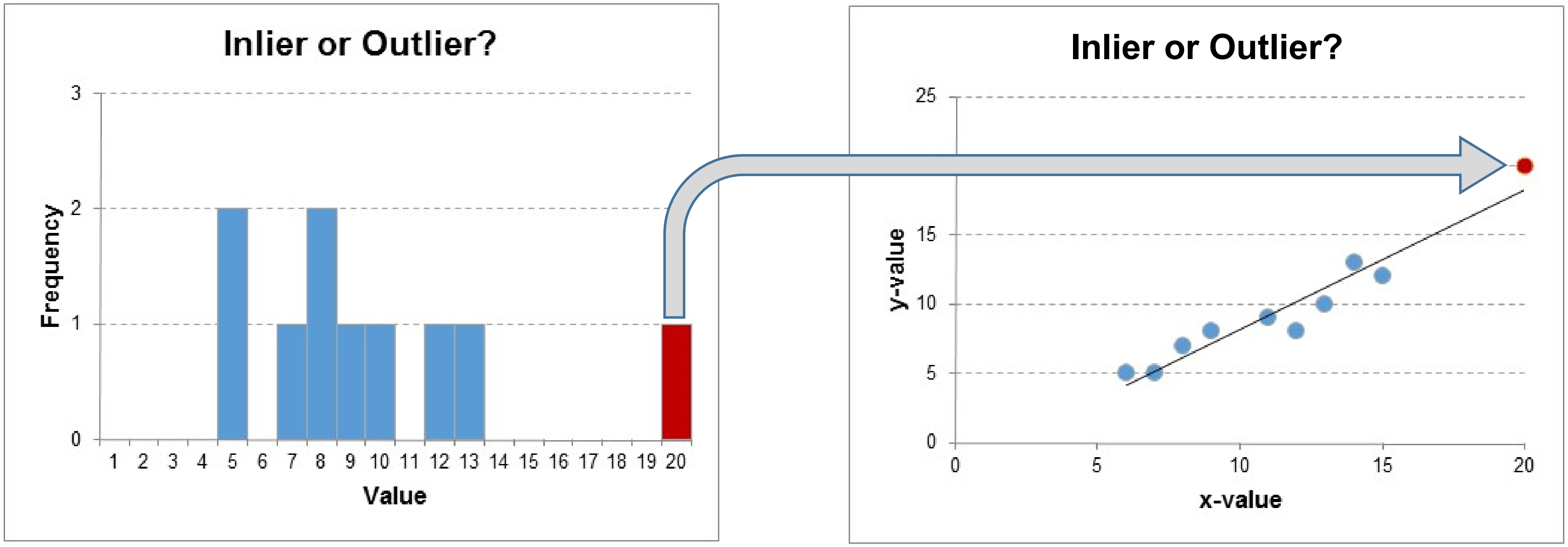
Hmm, probably an Outlier!
… but shall we toss a coin?



Inlier?
Keep it

Spot the Outlier … but now with some Context

**Looking at the same data but including a driver rather than just the value …
Again, based on "Gut Feel" you have to decide, do we keep it or exclude it?**

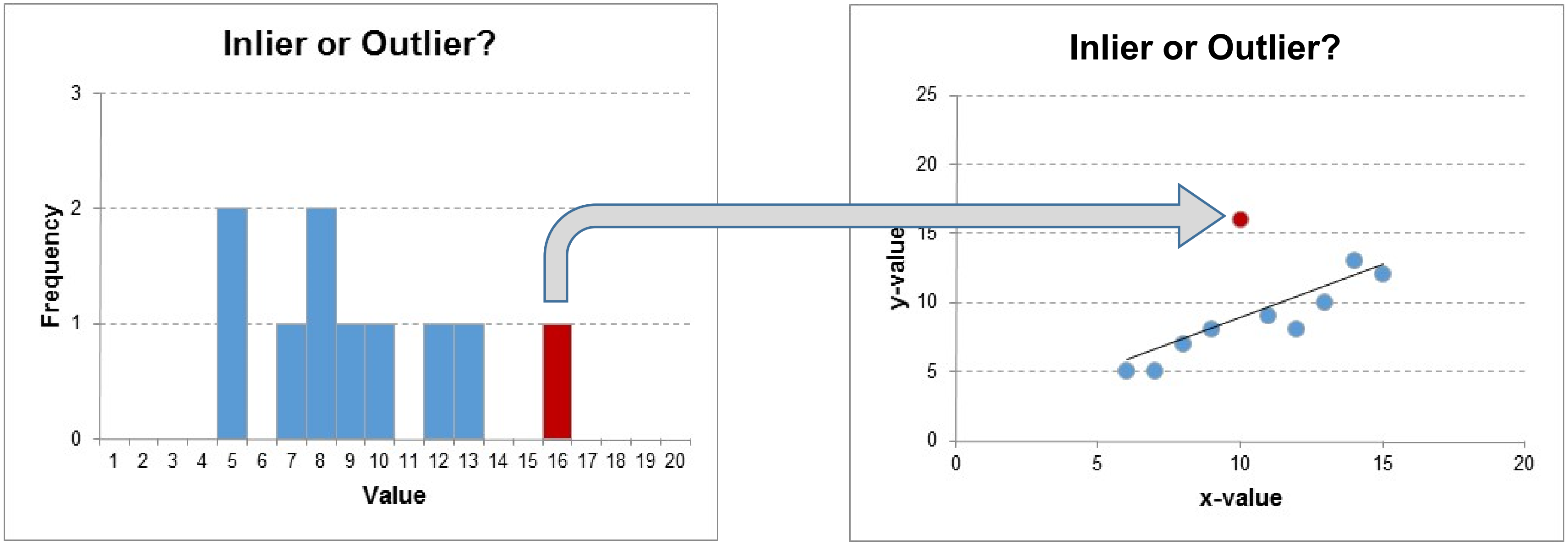


**Outlier?
Exclude it**

**No, Inlier!
Keep it**

Spot the Outlier … but now with some Context

Based on "Gut Feel" you have to decide, do we keep it or exclude it?

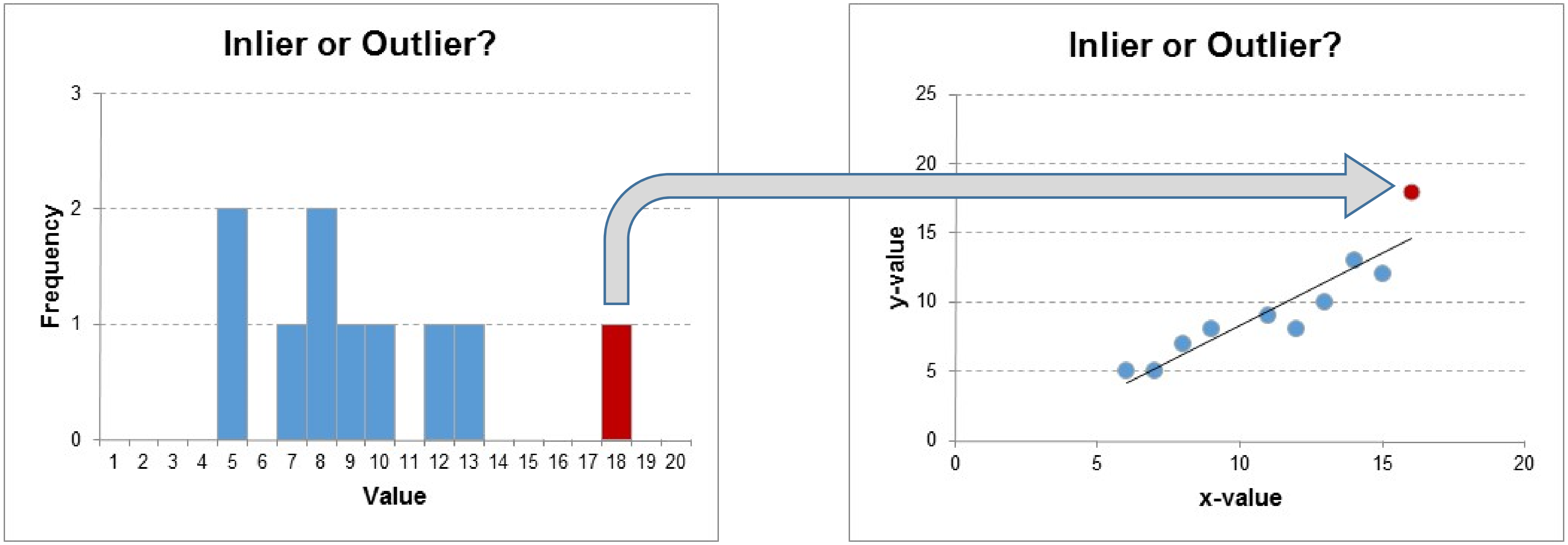


Inlier?
Keep it

No, Outlier!
Exclude it!

Spot the Outlier … but now with some Context

Based on "Gut Feel" you have to decide, do we keep it or exclude it?



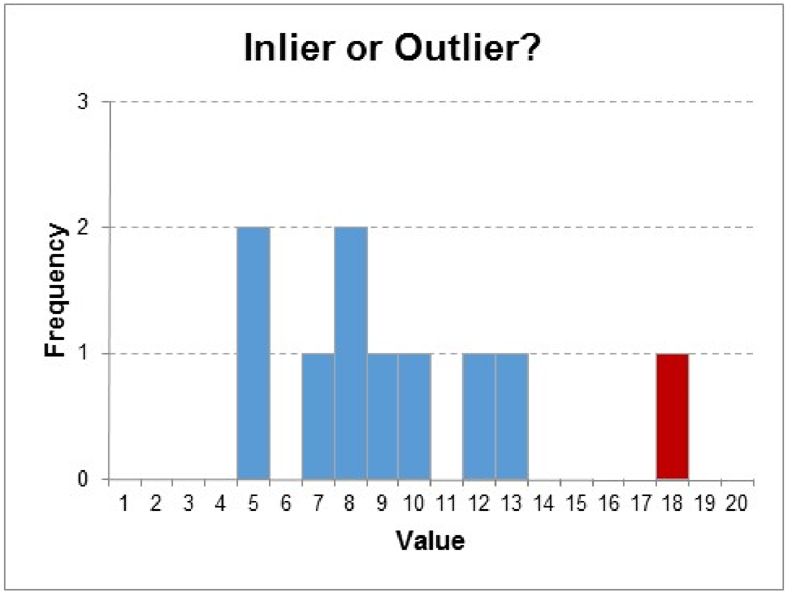
Hmm, probably an Outlier!
… but shall we toss a coin?

Possibly an Inlier
but still not so sure?
Shall we toss a coin?

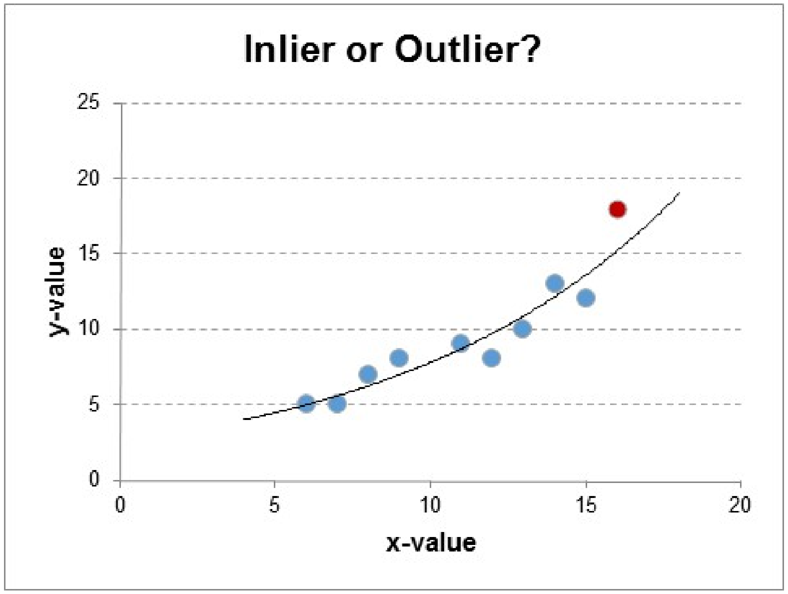
Spot the Outlier … but now with some Context

Based on "Gut Feel" you have to decide, do we keep it or exclude it?

What if it is a non-linear relationship?



Hmm, probably an Outlier!
… but shall we toss a coin?



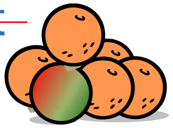
Now it looks more like an Inlier
Let's keep it

What do we mean by the term "Outlier"?

Definition of an Outlier

The Oxford English Dictionary provides four possible definitions of the term Outlier:

1. A person or thing situated away or detached from the main body or system
2. ~~A person or thing differing from all other members of a particular group or set~~
3. ~~*In Geology:* A younger rock formation isolated among older rocks~~
4. *In Statistics:* A data point on a graph or in a set of results that is very much bigger or smaller than the next nearest data point.



Source: OED 2011

Again, what do we mean by the term "Outlier"?

- The OED Definition implies a one dimensional view of an Outlier, but Estimating is rarely one dimensional
  - The concept of an Outlier being an "extreme value", implying a very low or high value relative to all others is quite one-dimensional.
  - As we have just seen, it may be just displaced from the pattern formed by the rest of our data
- Consider this as an alternative definition for Estimators:

  "An outlier is a value that falls substantially outside the pattern of other data. The outlier may be representative of unintended atypical factors, which cannot be easily normalised, or may simply be a value which has a very low probability of occurrence"
- If we accept this definition, how can we use it to identify outliers correctly?

Identifying Outliers: How can we spot one?

Surely there's a better way of spotting an Outlier than "personal opinion"?

- The good news is there is almost a plethora of Numerical Techniques
    - Chauvenet's Criterion
    - Peirce's Criterion
    - Tukey Fences
    - Grubbs' Test
    - Dixon's Q Test
    - Generalised Extreme Studentised Deviate
    - Tietjen-Moore Test
    - Iglewicz and Hoaglin Modified Z Score

We'll have a look at some of these shortly

… and consider some variations on them too

- The bad news is that they don't always point us to the same conclusion
- Perhaps we should follow good estimating practice and use more than one technique?
    - But what if two of them conflict, do we then try a "best of three"?

Do we have the time to do this?

Key Features of these Numerical Techniques?

There are tests that can be used to detect a single outlier … and no more!

- Grubbs' Test
- Dixon's Q Test

Require Tables of Critical Values to be Created or Downloaded

There are tests that can be used iteratively to detect multiple outliers … but one at a time

- Chauvenet's Criterion
- Tukey Fences
- Generalised Extreme Studentised Deviate
- Iglewicz and Hoaglin

One Outlier may mask another

There are tests that can be used to detect multiple outliers … so long as we know how many!

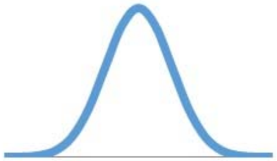
- Peirce's Criterion
- Tietjen-Moor Test

Require Tables of Critical Values to be Created or Downloaded

Key Features of these Numerical Techniques

Whilst these tests are all different, they tend to have one thing in common

- They tend to assume a Normal Distribution or a Distribution that is "Normalesque" such as a Student t Distribution



- Whereas most cost and schedule data is inherently positively skewed … but again that is one-dimensional thinking

  If it is a one dimensional problem, then we are probably better trying to fit a skewed probability distribution such as a Beta Distribution to our data to include all points

We should think of the problem in relation to the pattern in the data

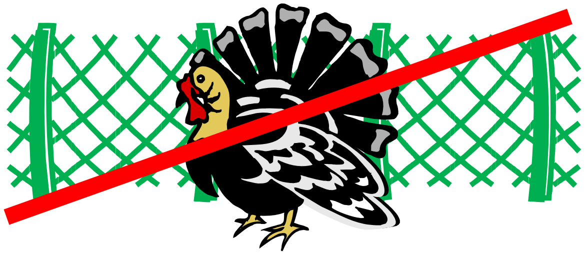
- Does the relationship appear to be linear (apart from the Suspected Outlier)?
- Does the relationship appear to be an Exponential, Power or Logarithmic function?

  i.e. one that we can transform into a Linear Relationship?

- If the answer to either of these is "yes", then we can expect the scatter of the data (or its transformation) to be "Normalesque" around the linear relationship

Tukey Fences

Drawing the Boundary Lines



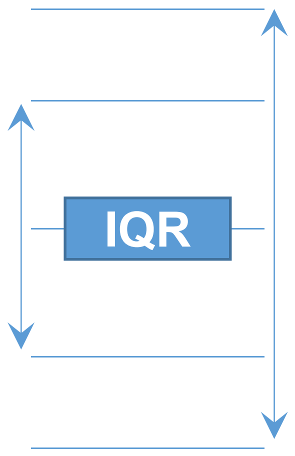
This is no "R" in Tukey !!!

Tukey Fences

How do Tukey Fences work?

The technique identifies two sets of boundaries (fences) for potential outliers based on the Interquartile Range (IQR) of the data sample …

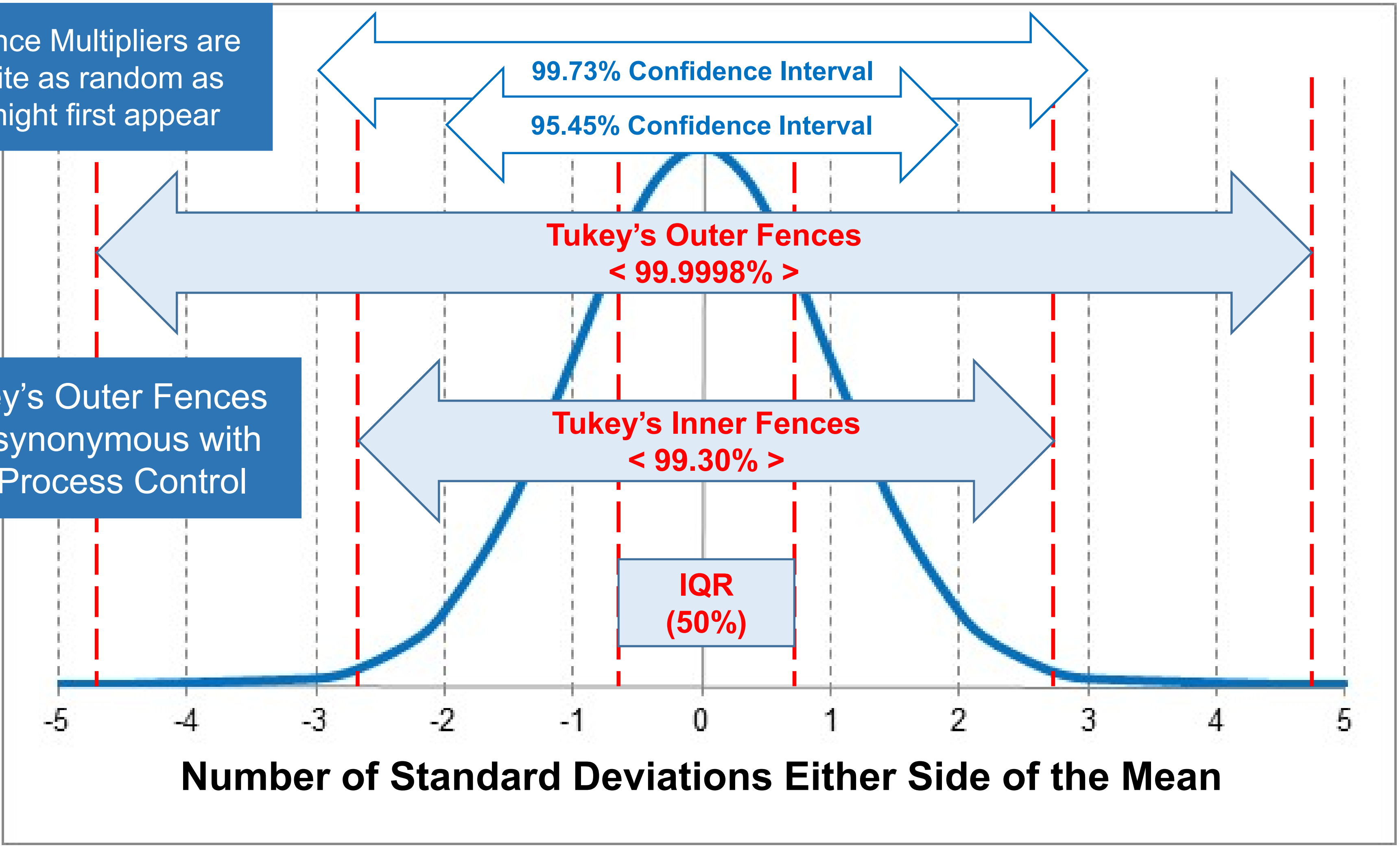
1. Calculate the IQR for the data (the middle 50% Confidence Interval between the 25% and 75% Confidence Levels)
2. Establish Inner Fences for Potential Outliers
   - 25% Confidence Level minus one and a half times the IQR
   - 75% Confidence Level plus one and a half times the IQR
3. Establish Outer Fences for Extreme Outliers
   - 25% Confidence Level minus three times the IQR
   - 75% Confidence Level plus three times the IQR



- Data falling outside the Outer Fences can be safely assumed to be Extreme Outliers.
- Data within the Outer but outside the Inner Fences are potential outliers, depending on the underlying distribution …

Tukey Fences cf. Standard Normal Distribution

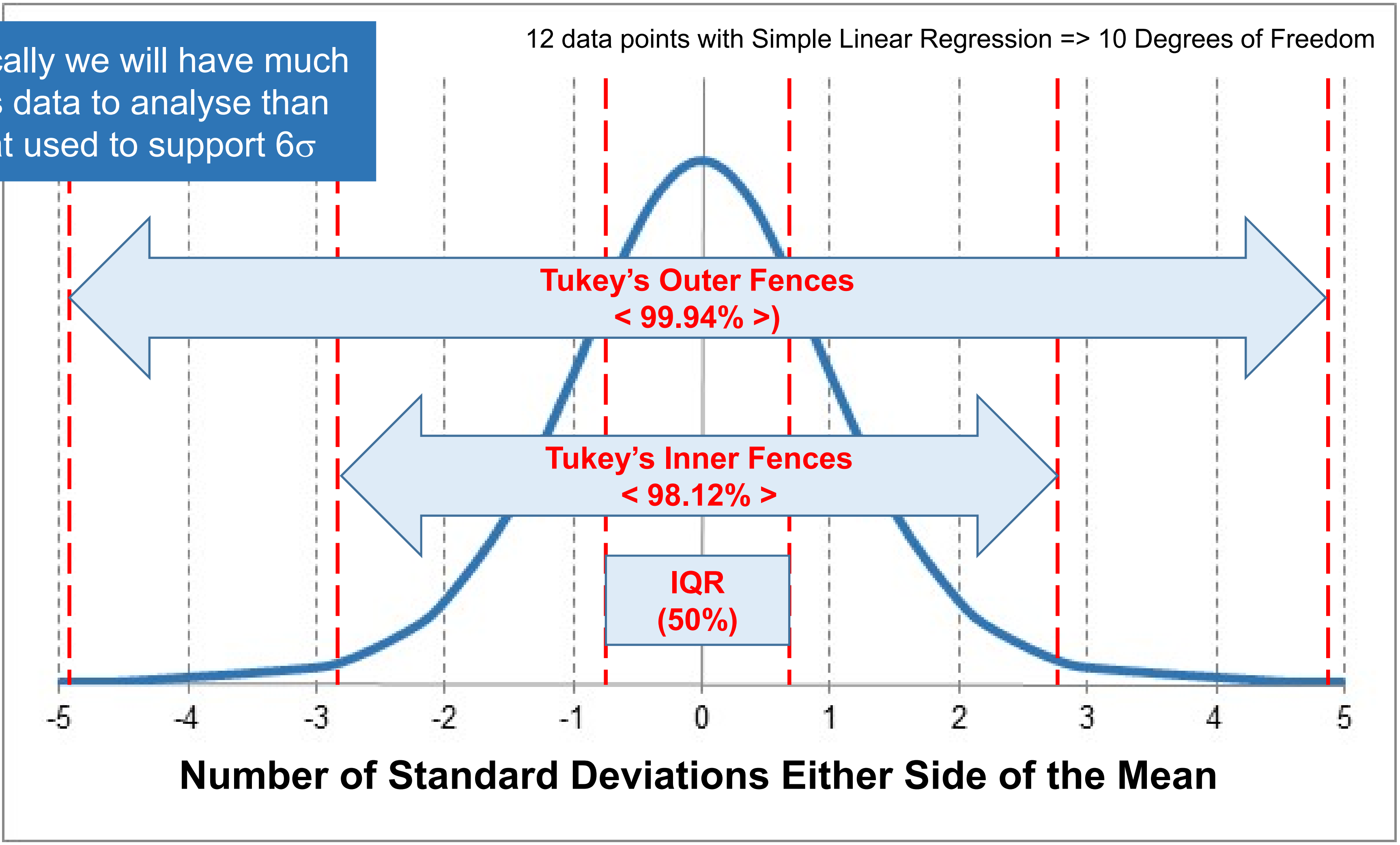
The Fence Multipliers are not quite as random as they might first appear



Tukey's Outer Fences are synonymous with 6σ Process Control

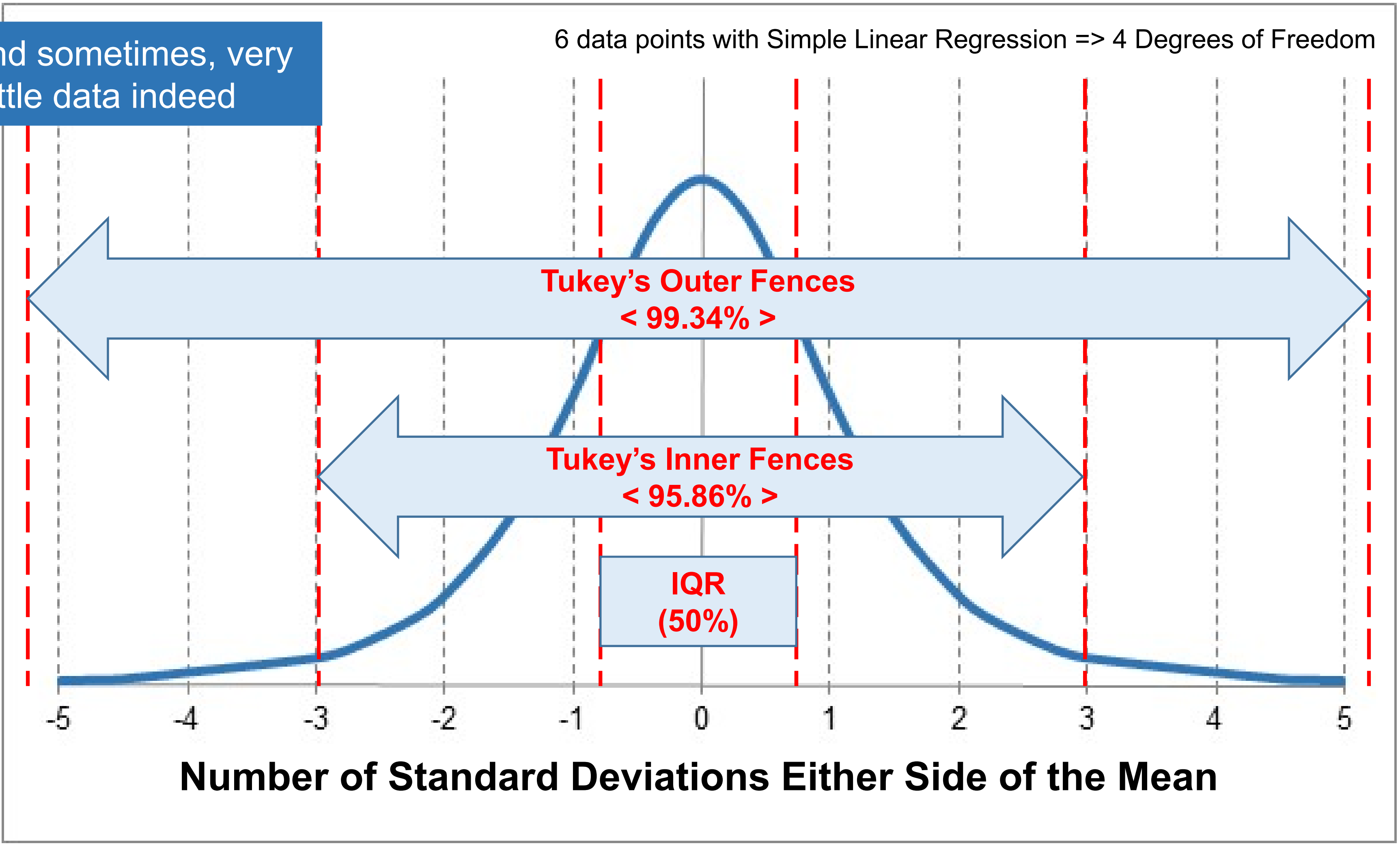
Tukey Fences cf Student t-Distribution + 10 df

Typically we will have much less data to analyse than that used to support 6σ

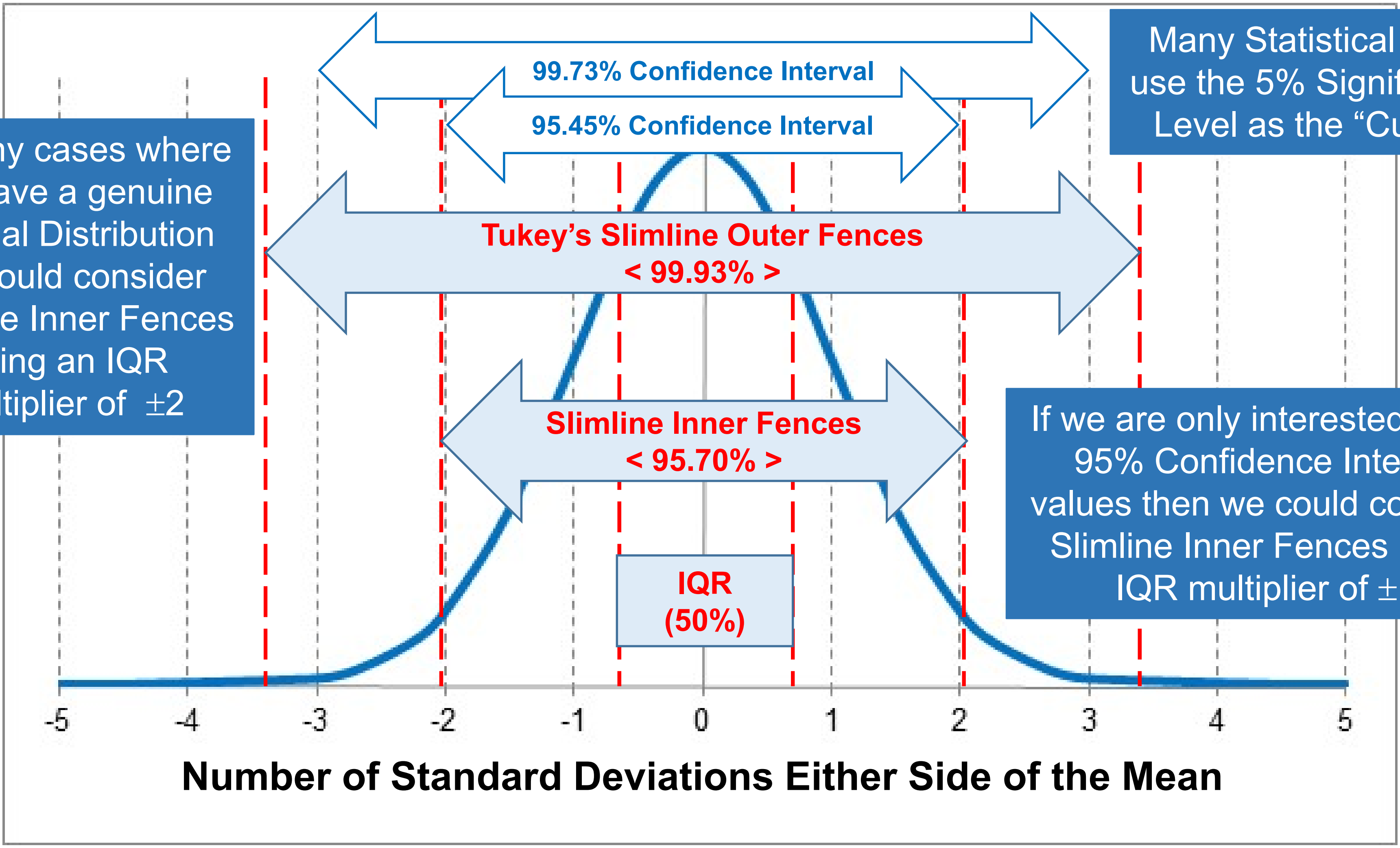


Tukey Fences cf Student t-Distribution + 4 df

… and sometimes, very little data indeed



Consider Slimline Tukey Fences for 'Normal' Data



In many cases where we have a genuine Normal Distribution we could consider Slimline Inner Fences using an IQR multiplier of ±2

Many Statistical Tests use the 5% Significance Level as the "Cut-off"

If we are only interested in the 95% Confidence Interval values then we could consider Slimline Inner Fences using IQR multiplier of ±1

Tukey Fences Example (Traditional)

x	y	Line of Best Fit	Difference to LoBF
6	5	4.18	0.82
7	5	5.23	-0.23
8	7	6.27	0.73
9	8	7.31	0.69
11	9	9.40	-0.40
12	8	10.44	-2.44
13	10	11.48	-1.48
14	13	12.52	0.48
15	12	13.56	-1.56
16	18	14.61	3.39

Count	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

Applied to Regression Error Value

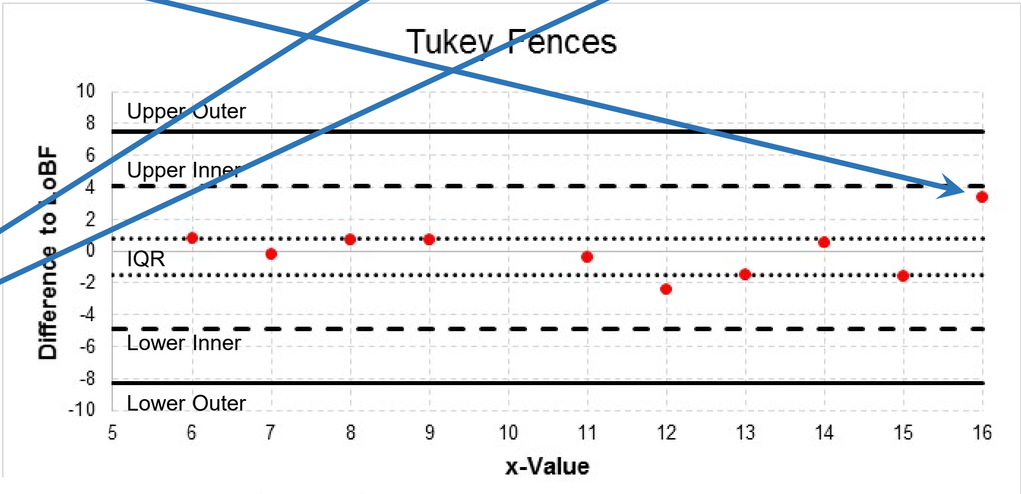
Tukey

Less than the Upper Inner Fence

= Q3 + 3 x IQR

Calculated using Excel's SLOPE and INTERCEPT functions (no test of significance performed)

Using QUARTILE.EXC function in Microsoft Excel



Note: 4 Quartiles are defined by 5 boundary parameters 0 for start and end of 1st, 2nd, 3rd and 4th Quartiles

Tukey Fences Example (Slimline)

x	y	Line of Best Fit	Difference to LoBF
6	5	4.18	0.82
7	5	5.23	-0.23
8	7	6.27	0.73
9	8	7.31	0.69
11	9	9.40	-0.40
12	8	10.44	-2.44
13	10	11.48	-1.48
14	13	12.52	0.48
15	12	13.56	-1.56
16	18	14.61	3.39

Count	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

IQR	Quartile 1	-1.50	2.25
	Quartile 2	0.13	
	Quartile 3	0.75	

Using QUARTILE.EXC function in Microsoft Excel

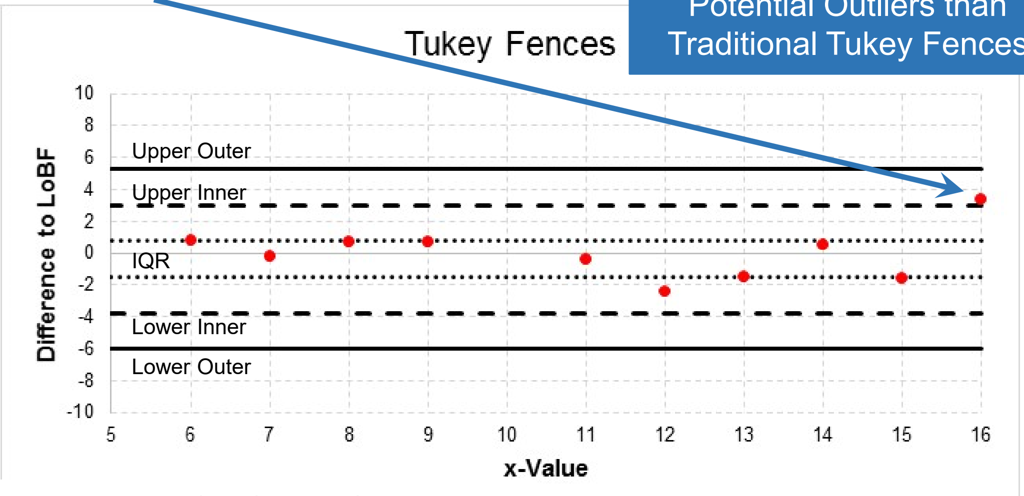
Applied to Regression Error Value

	Fence Multiplier	Fence Position
Lower Outer	-2	-6.01
Lower Inner	-1	-3.75
IQR		2.25
Upper Inner	1	3.01
Upper Outer	2	5.26

Slimline Tukey Fences

Greater than the Upper Inner Fence

Slimline Tukey Fences are much less tolerant of Potential Outliers than Traditional Tukey Fences



Note: 4 Quartiles are defined by 5 boundary parameters 0 for start and end of 1st, 2nd, 3rd and 4th Quartiles

Chauvenet's Criterion

Counting the Unexpected Arisings

Chauvenet's Criterion

How does Chauvenet's Criterion work?

It asks "*How many data points can I expect this far from the mean given the sample size I have?*"

1. Assumes a Normal Distribution (*we'll come back to that*)
2. Calculates a "Standardised Difference Z-Score" for the suspect data point:
   - Suppose we have n data points x_1 to x_n in our sample
   - If the Mean of the sample is $\bar{x}$ and the Standard Deviation is s,

   then the Absolute value of the Z-Score for the i^{th} data point would be: $Z_i = \frac{|x_i - \bar{x}|}{s}$
3. Calculates the probability of having an Absolute Z-Score of this value or greater, (i.e. includes the mirror image of having a negative Z-Score of this value or less)
4. Calculates how many data points we might reasonably expect from a Standard Normal Distribution based on our sample of data points (rounded to the nearest integer) by applying the calculate probability to the sample size
5. If our "suspect" data point has a zero expectation, it is a potential outlier

Chauvenet's Criterion Example (Traditional)

x	y	Line of Best Fit	Difference to LoBF
6	5	4.18	0.82
7	5	5.23	-0.23
8	7	6.27	0.73
9	8	7.31	0.69
11	9	9.40	-0.40
12	8	10.44	-2.44
13	10	11.48	-1.48
14	13	12.52	0.48
15	12	13.56	-1.56
16	18	14.61	3.39

Count, n	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

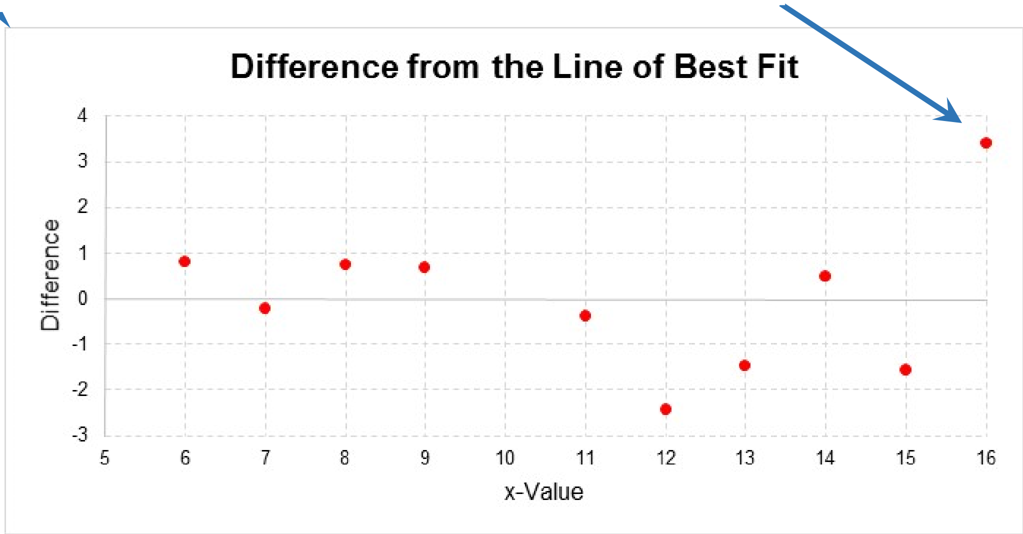
2*(1-NORM.S.DIST(ABS(Z),TRUE))

Multiply Probability by Sample Size

No points expected

Calculated using Excel's SLOPE and INTERCEPT functions (no test of significance performed)

Z-Score based on Difference to Best Fit Line ABS(Error)/Std Dev


Difference from the Line of Best Fit

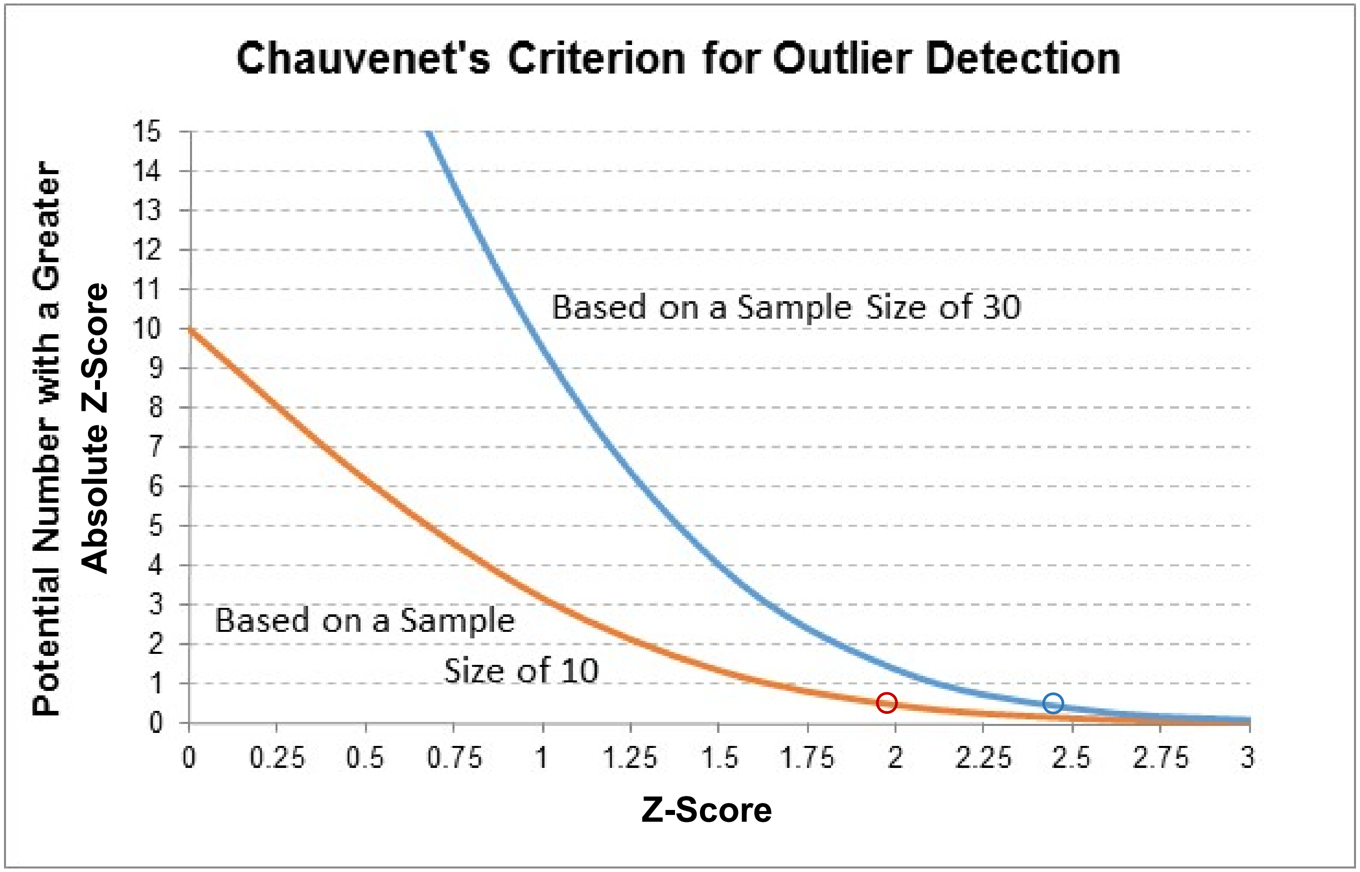
Chauvenet's Criterion – A Potential Flaw in the Logic

- It implies probabilistically that we have a greater chance of getting a remote value with larger sample sizes
- Implies a Critical Value of the Z-Score increases with the number of data points

Data Points: 10
Critical Value: 1.96

Data Points: 30
Critical Value: 2.39

- **Surely, it is either an outlier, or it is not?**



Small Sample Size

A Potential SSS Variation on Chauvenet's Criterion

Countering the downside of Chauvenet's Criterion

Things we can do differently thanks to Guinness!

Other Stouts are available

- Chauvenet's Criterion assumes that data will be distributed Normally around the sample mean
  - That's not unreasonable for a sufficiently large sample size where we are considering fitting a straight line by Least Squares (i.e. Regression)
  - But it is fundamentally flawed for small sample sizes … as is often the domain of Estimators
- Instead we should consider that our Standardised Z-Score has a Student t Distribution
  - This in effect widens the spread of Confidence Interval values for smaller samples
  - It is more "forgiving" or "lenient" when considering points further from the mean
  - However, it still doesn't mean that it is more correct necessarily … the sample mean may not be representative of the whole population … but the assumption is more consistent with any Least Squares Regression we may want to perform

Chauvenet's Criterion Example Revisited

x	y	Line of Best Fit	Difference to LoBF	Absolute Z-Score	Prob > Z ~ t(0,n-1)	Expected # Points	Rounded # Points
6	5	4.18	0.82	0.497	63.1%	6.310	6
7	5	5.23	-0.23	0.138	89.3%	8.930	9
8	7	6.27	0.73	0.446	66.6%	6.663	7
9	8	7.31	0.69	0.420	68.4%	6.844	7
11	9	9.40	-0.40	0.241	81.5%	8.147	8
12	8	10.44	-2.44	1.487	17.1%	1.713	2
13	10	11.48	-1.48	0.903	39.0%	3.903	4
14	13	12.52	0.48	0.291	77.8%	7.775	8
15	12	13.56	-1.56	0.954	36.5%	3.649	4
16	18	14.61	3.39	2.069	6.8%	0.685	1

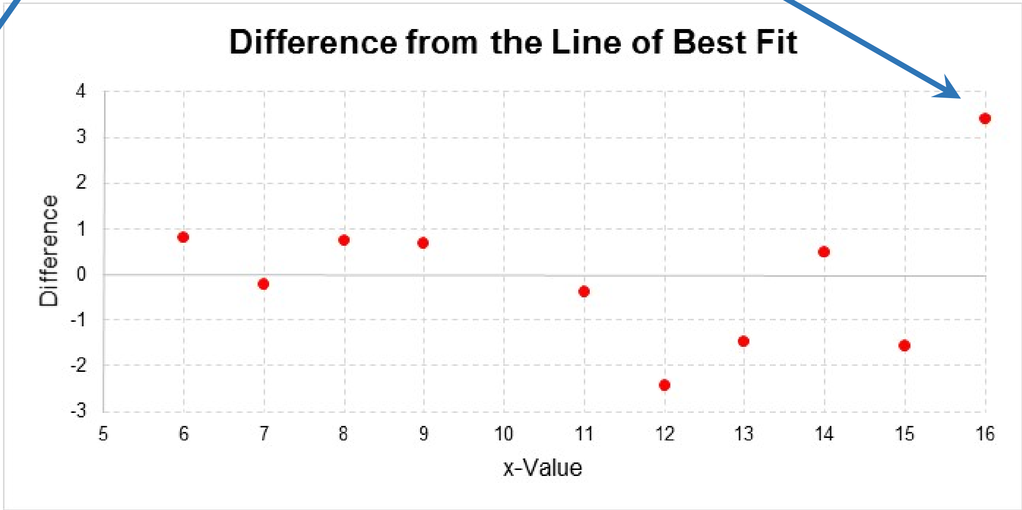
T.DIST.2T(ABS(Z),n-1)

Multiply Probability by Sample Size

One point expected

Count, n	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

Z-Scores are usually associated with Normal Distributions but here we will use it with a t-Distribution



Iglewicz and Hoaglin

“The MAD Method”

There's MADness in this Method

Iglewicz and Hoaglin recommend the use of a Modified Z-Score

Noting that the Mean of the Sample data is adversely affected by Outliers Iglewicz and Hoaglin considered a Median-based test

1. Assumes a Normal Distribution
2. Calculates a Modified Z-Score, M_i (M-Score) for the suspect data point:
    - Suppose we have n data points x_1 to x_n in our sample
    - If the **Median** of the sample is $\tilde{x}$, and the **Median Absolute Deviation** is $\boldsymbol{MAD}$, then the Absolute value of the M-Score for the i^{th} data point would be:

$$M_i = 0.6745 \frac{|x_i - \tilde{x}|}{MAD}$$

3. Any point with a M-Score of 3.5 or greater should be considered a potential outlier

Note: The constant 0.6745 is largely redundant; we could instead just use a Critical Value of 5.1891 (but that is not quite so memorable, is it?)

Iglewicz & Hoaglin Example

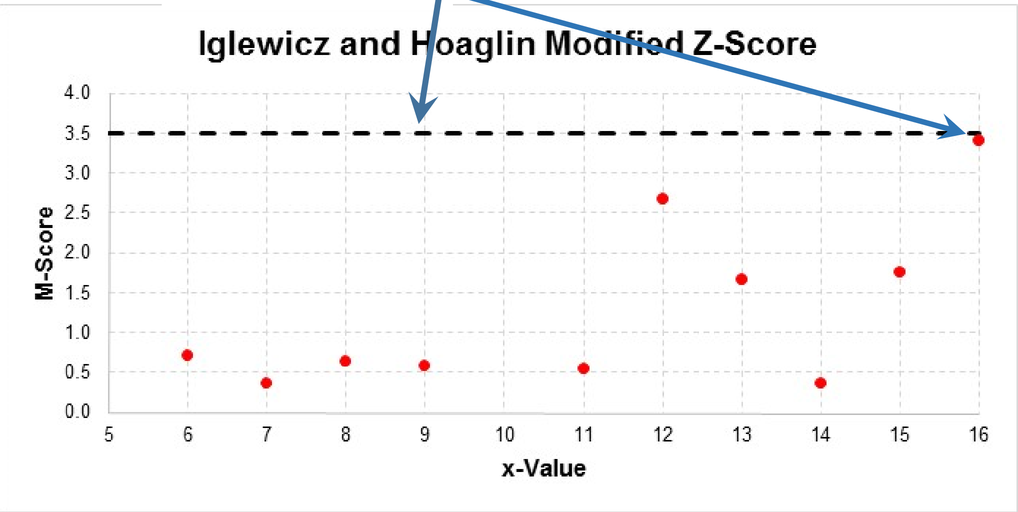
x	y	Line of Best Fit	Difference to LoBF	Median Abs Dev	Absolute M-Score
6	5	4.18	0.82	0.69	0.718
7	5	5.23	-0.23	0.35	0.367
8	7	6.27	0.73	0.61	0.631
9	8	7.31	0.69	0.56	0.587
11	9	9.40	-0.40	0.52	0.543
12	8	10.44	-2.44	2.56	2.669
13	10	11.48	-1.48	1.61	1.672
14	13	12.52	0.48	0.35	0.367
15	12	13.56	-1.56	1.69	1.760
16	18	14.61	3.39	3.27	3.402

Count	10				
Median	11.5	8.5	9.92	0.13	0.65
Provisional Regression Slope			1.04		
Provisional Regression Intercept			-2.07		

$$M_i = 0.6745 \frac{|x_i - \hat{x}|}{MAD}$$

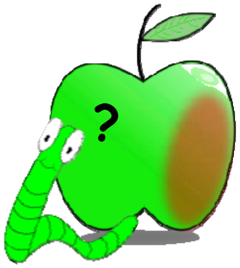
Calculated using Excel's SLOPE and INTERCEPT functions (no test of significance performed)

Critical Value for an Outlier is an M-Score of 3.5



Grubbs' Test for a Single Outlier

Tails of the Unexpected



Grubbs' Test

Maximum Deviation Compared with the Standard Deviation

Considered by many to be the most robust Outlier Test

Compares the Maximum Deviation from the Mean with the Standard Deviation:

1. Assumes a Normal Distribution
2. Calculates a G-Statistic as the Maximum Deviation divided by the Standard Deviation:
   - Suppose we have *n* data points x_1 to x_n in our sample
   - If the Mean of the sample is $\bar{x}$ and the Standard Deviation is s,

   then the G-Statistic for the sample would be: $G = \frac{\max\limits_{n} |x_i - \bar{x}|}{s}$

This is really just the Maximum Z-Score

3. If G > Critical Value then the associated Data point is an Outlier

The downside is that the Critical Value is calculated based on the number of Data Points, ***n*** in the sample, and the Confidence Level Cut-off, α we want to apply:

The good news is that we can download a table of Critical Values them from the internet

$$G > \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\frac{\alpha}{2n},n-2}}{n-2+t^2_{\frac{\alpha}{2n},n-2}}}$$

Yes, seriously!
I just wish I had a camera to take a picture of your faces right now!

Grubbs' Test Example

x	y	Line of Best Fit	Difference to LoBF	Abs Deviation from the Mean	G: |Abs Dev| (Std Dev)
6	5	4.18	0.82	0.82	0.497
7	5	5.23	-0.23	0.23	0.138
8	7	6.27	0.73	0.73	0.446
9	8	7.31	0.69	0.69	0.420
11	9	9.40	-0.40	0.40	0.241
12	8	10.44	-2.44	2.44	1.487
13	10	11.48	-1.48	1.48	0.903
14	13	12.52	0.48	0.48	0.291
15	12	13.56	-1.56	1.56	0.954
16	18	14.61	3.39	3.39	2.069

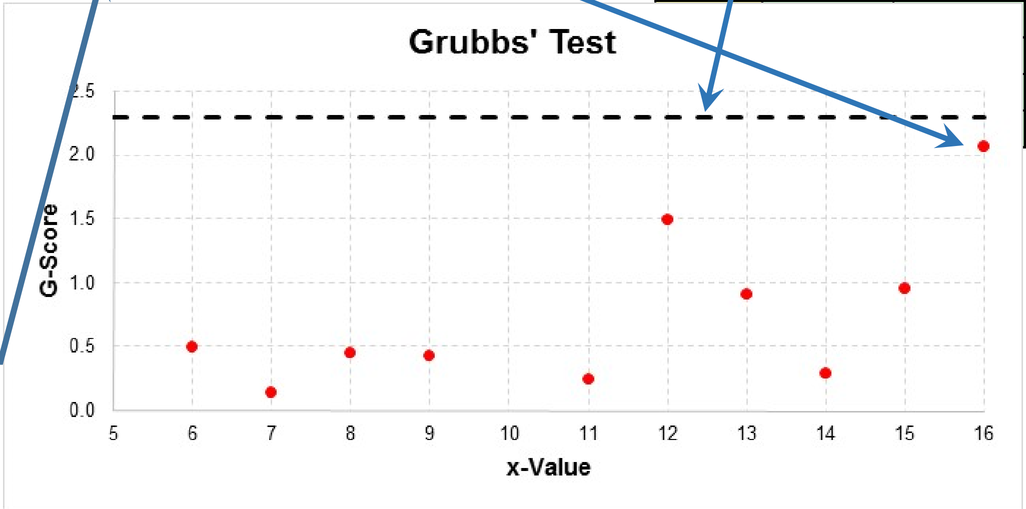
Count	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

Sample Size, n	Critical Values Grubbs' Test @ 5% Level	 Grubbs' Test @ 10% Level
4	1.481	1.463
5	1.715	1.671
6	1.887	1.822
7	2.020	1.938
8	2.127	2.032
9	2.215	2.110
10	2.290	2.176
11	2.355	2.234
12	2.412	2.285
13	2.462	2.331
14	2.507	2.372

Critical Value for an Outlier at the 5% Significance and a sample size of 10 is a Max G-Score of 2.290

Calculated using Excel's SLOPE and INTERCEPT functions (no test of significance performed)

G-Score is identical to Z-Score used for Chauvenet's Criterion



Doing the J-B Swing!

A Skewness and Excess Kurtosis Perspective

An embryonic idea for further investigation

Doing The J-B Swing

It might sounds like the name of a Jazz Band but …

- The Jarque-Bera Statistic is used as a Test for Normality
- For a sample size of ***n***, it combines measures of Skewness, ***g*** and Excess Kurtosis, ***k***:

$$JB = \frac{n}{6}\left(g^2 + \frac{k^2}{4}\right)$$

- A Normal Distribution has a JB Statistic of Zero
- A t-Distribution with ten data points has a JB Statistic of $^5/_{12}$
- A Chi-Squared Test with 2 degrees of freedom can be used to test significance of the sample's JB Statistic
- The proposition behind the "**J-B Swing**" as a "Rule of Thumb" Test for an Outlier, is a "Before and After" event:

Does the residual data sample become significantly "more Normal" with the removal of the potential outlier?

Doing The J-B Swing

	x	y	Line of Best Fit	Difference to LoBF
	6	5	4.18	0.82
	7	5	5.23	-0.23
	8	7	6.27	0.73
	9	8	7.31	0.69
	11	9	9.40	-0.40
	12	8	10.44	-2.44
	13	10	11.48	-1.48
	14	13	12.52	0.48
	15	12	13.56	-1.56
	16	18	14.61	3.39
Count, n	10			
Mean	11.1	9.5	9.50	0.00
Std Dev	3.48	3.98	3.63	1.64
Provisional Regression Slope			1.04	
Provisional Regression Intercept			-2.07	

< Suspected Outlier

Skewness	With Potential Outlier	0.59	*g*
	Without Potential Outlier	-0.65	
Excess Kurtosis	With Potential Outlier	1.08	*k*
	Without Potential Outlier	-1.04	

- There is a valid argument that the J-B Statistic should be recalculated for the “after” event based on the difference to the revised Line of Best Fit (LoBF)
- As a short-cut, to avoid having to do this we should be looking for a large positive swing of say 50%+

Using Excel Function
CHISQ.DIST.RT(ABS(JB),2)

$$JB = \frac{n}{6}\left(g^2 + \frac{k^2}{4}\right)$$

Jarque-Bera Statistic	With Potential Outlier	1.07
	Without Potential Outlier	1.15
Jarque-Bera Significance	With Potential Outlier	59%
	Without Potential Outlier	56%
	JB Significance Swing	-2%

Other Outlier Tests

For reference only – not discussed here

Some Other Tests (For Reference)

- **Peirce's Criterion** – another Z-Score Test (pre-dating Chauvenet's Criterion), which considers the maximum allowable deviation form the Mean.

- **Tietjen-Moore Test** – requires us to know in advance how many outliers we think have (*defeats the objective a little, wouldn't you say*) but similar in that sense to Peirce's Criterion

- **Generalized Extreme Studentised Deviate (ESD)** – despite its name sounding like a radicalised student protest movement from the nineteen-sixties or seventies, is a more general purpose version of Grubbs' Test allowing multiple outliers to be detected

- **Dixon's Q-Test** – works on the premise that an outlier by definition is significantly distant from the rest of the data. The Test compares the distance between the potential outlier and its nearest neighbour (i.e. the gap) in comparison to the overall range of the data (including the potential outlier.)

Testing the Tests

How do the tests perform against each other?

Testing the Tests

How do the various Tests Compare with our data?

Test	Basis	Result	Comment
Tukey Fences (Traditional)	IQR ± Multipliers of 1.5 and 3	Not an Outlier	Inner Fences normally adequate to detect Outliers
Tukey Fences (Slimline)	IQR ± Multipliers	Potential Outlier	Geared to 5% Significance Test for larger data samples
Chauvenet's Criterion (Normal Assumption)	Expected Number with observed Z-Score	Potential Outlier	Often conflicts with other tests
Chauvenet's Criterion (Revised for SSS)	Expected Number with observed Z-Score	Not an Outlier	Adapts Chauvenet Criterion to SSS using t-Distribution
Iglewicz and Hoaglin MAD Score	M-Score Critical Value	Not an Outlier	Median Based Test
Grubb's Test	G-Max Critical Value (Z-Score Equivalent)	Not an Outlier	Probably the most robust test for a Single Outlier
J-B Swing	J-B Statistic "Before and After" Swing	Not an Outlier	Use as a Rule of Thumb Indicator only

… but that was just a one-off

Testing the Tests

- **Fifteen samples of ten data points gave the following results**

If Grubbs' Test is the most robust …

Generally Consistent

Promising, but needs some more "fine tuning" to assure robustness

				Sample Example Number														
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Grubbs' Test	Critical Z @	95%	2.290		Outlier?				Outlier?					Outlier?				
	Critical Z @	90%	2.176		Outlier?				Outlier?					Outlier?				
Chauvenet's Criterion	Based on Normal Dist	No Expected with Z-Score		0	0	1	0	1	0	1	1	0	1	0	1	1	0	0
		Traditional Test Result		Outlier?	Outlier?		Outlier?		Outlier?			Outlier?		Outlier?			Outlier?	Outlier?
	Based on Student t	No Expected with Z-Score		1	0	2	1	1	0	1	1	1	2	0	1	1	1	1
		Revised SSS Test Result			Outlier?				Outlier?					Outlier?				
Iglewicz & Hoaglin	Max M-Score >		3.5						Outlier?					Outlier?				
Tukey Fences	Traditional	Inner Fence	1.5 IQR		Outlier?		Outlier?		Outlier?			Outlier?		Outlier?				
		Outer Fence	3 IQR															
	Slimline	Inner Fence	1 IQR	Outlier?	Outlier?		Outlier?		Outlier?			Outlier?		Outlier?	Outlier?		Outlier?	Outlier?
		Outer Fence	2 IQR						Outlier?					Outlier?				
Jarque-Bera Swing Indicator	JB Swing	Significance Swing		-2%	78%	-11%	-4%	-27%	55%	-3%	-1%	43%	-2%	12%	11%	-16%	-27%	10%
		Swing >	50%		Outlier?				Outlier?									

High degree of commonality but very intolerant of Outliers

Very Outlier Friendly!

Summary

Outing the Outliers

Outing the Outliers: Summary

Why is it important for Contract Acquisition and Delivery that we correctly identify Outliers?

The initial estimate is pivotal to success, in terms of competitiveness and achievability

So, let's not set ourselves up for a fall …

If in our data analysis and estimating phase we put in:

- Something that is out of **C**ontext (a Cuckoo)
- Something that is **R**emote (Extreme Value)
- Something that is out of **A**lignment (or Anomalous)
- Something that **P**robably won't occur again

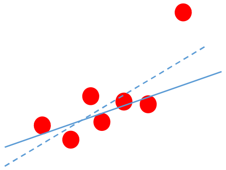
I'm struggling here for a suitable acronym. Can anyone help me out with one?

After all, what you put in is what you get out

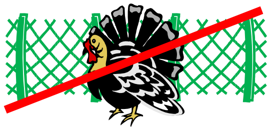
So, out with the …

Outing the Outliers: Summary

- Virtually all Outlier Tests assume a “Normalesque” Distribution
- Check first that we have eliminated any Cuckoos
- Check also that we have normalised our data



- Look for a Linear Pattern in our data. Transform to Linear where we can
- Apply Tests to the scatter around the Linear Pattern (Expect Normalesque)



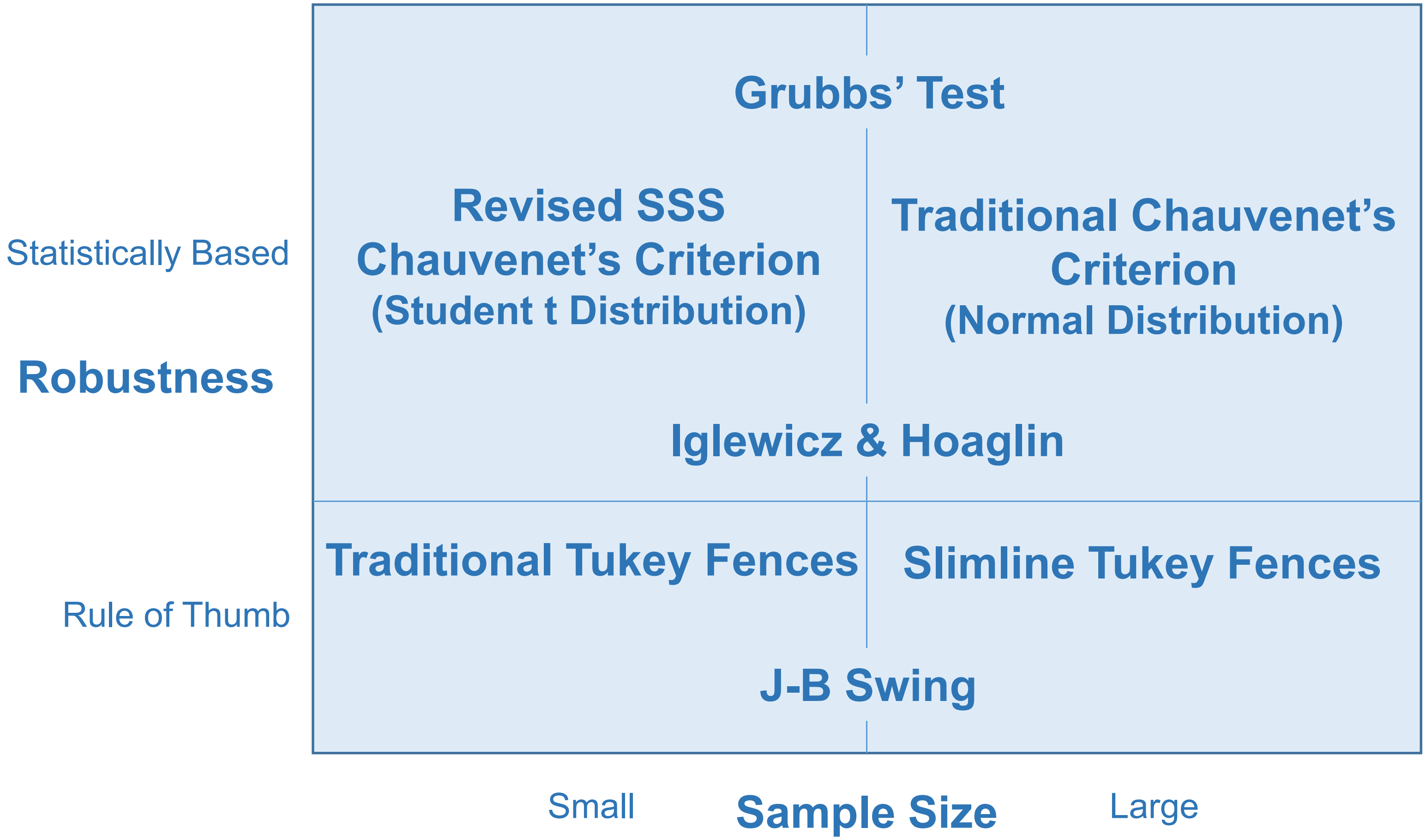
- Try Tukey Fences first – simple and visual
  - Extreme Outliers (beyond the Outer Fences) … remove data as Outliers
  - Potential Outliers (between the Inner and Outer Fences) … verify with another technique (possibly Slimline Outer Fences)

- Check initial conclusions with either:
  - Grubbs’ Test (for a single outlier)
  - Revised SSS Chauvenet’s Criterion (using Student t Distribution)
  - Iglewicz and Hoaglin’s MAD Method

  or by Doing the J-B Swing

Any Questions?

Thank you for your patience and indulgence



Grubbs' Test

Revised SSS Chauvenet's Criterion (Student t Distribution)

Traditional Chauvenet's Criterion (Normal Distribution)

Iglewicz & Hoaglin

Traditional Tukey Fences

Slimline Tukey Fences

J-B Swing

Statistically Based

Robustness

Rule of Thumb

Small **Sample Size** Large

Table position is not intended to be an absolute indicator of hierarchical statistical robustness or reliability