Applying six sigma style tools and other lean practices to real world laboratories to reduce error rates and improve efficiencies

Dr Eric Wilkes

Global Manager Analytical Services

Foster's

May 2009

or How to make the lies,
sorry statistics,
work for you!

Eric Wilkes

Six Sigma Beltless

Important introductory note

"*Laboratories are not car manufacturing plants!!!!*"

- The analytical values in labs are expected to vary from sample to sample!
- But good production practices, lean, and statistical controls do have their place and can add value.

What is the lab really there for??

"The main role of a production laboratory is to provide accurate and timely analytical results for their customers in a safe manner."

- To achieve this a good understanding of the nature of the tests and the limitations of the methods is essential.
- The tools and concepts which we can borrow from systems such as Six Sigma can be essential in improving this understanding.

Case Study One

How bad can it get?

Nice Spot!

How bad can it get?

- Large wine production and packaging facility in the Sonoma Valley, California, USA.
- Labs main QA check was cross checking the previous result against the current result for the same batch.
- Other than occasional calibrations no other QA/QC control.
- The result, 30% of the workload was rechecks.

What was happening?

- No investigations we carried out on any analytical issue.
- There wasn’t time!!!!
- This site was a classic example of an out of control system.
- Mistakes tended to feed back on themselves.
- The lab was not meeting any of the roles of a production lab.
- And the really sad bit was that both site management and the lab staff felt they were doing a good job.

Case Study 2

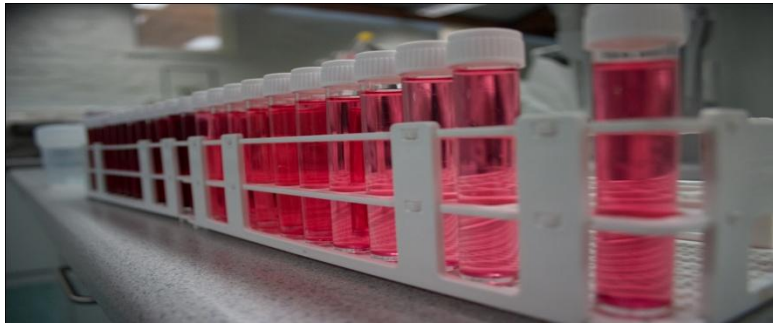
Introducing a run chart culture

Classic FMCG labs

- Many FMCG labs traditionally measured performance against the number of ‘in specification’ production samples.
- They relied on calibrations as QA process for both accuracy and precision.
- A self defeating process in many ways.

Foster's Wine Group

- The Foster's wine group was no different at the beginning of the decade.
- Faith was placed in time proven methods and the experience of technicians.
- Unfortunately this did not provide "evidence of the validity of results".
- The outcome was a lack of confidence by the laboratories customers and assumption that everything was working by the lab staff.
- To address this a change in culture was necessary.



The *"but"* statement!

- Any system introduced must lead to either
  - a required increase in accuracy.
  - an increase in efficiency

"You should never add accuracy or bureaucracy without reason."

- If it works better to do it on a hand drawn bit of paper, do it that way!
- In lean parlance we should do a value stream mapping exercise.

The Solution

- It was decided to add dedicated QA to all quantitative tests.
- This was to be done at the rate of one check sample every 10 tests (10%).
- This data was to be used in run charts with control measures to prompt action.
- However there were a number of core issues to be addressed.
  - What sample?
  - What values for the control limits?
  - What physical system?
  - What action rules?
  - How do we change the culture?

Changing the culture

"Without the buy in from the staff no control system will work!"

- Open discussions where people are helped to recognise that there are issues (and potential disasters) worked well.
- Promote the telling anecdotes. It helps people understand the reasons and the ramifications.
- Don't be afraid to try out various systems no matter how daft they may first seem.
- Don't force fit something from another location just because it worked there. Understand the local problem.
- Don't become bedazzled by technology (or statistics). Simple answers are often the best.

The Sample

- This proved to be surprisingly difficult.
- Please, ***please***, resist using a calibration standard for the first sample of the day!
- In the end for the majority of tests the easiest answer was found to be a 15 litre cask.
  - Cheap
  - Was stable for the one to two weeks it would last.
  - Tended to have analytical values that fell in the middle of the validated range.
- A days testing overlap between casks was made mandatory.
- For analytes where the cask did not meet these parameters a large number of a standard was frozen in small tubes.
  - However, this took a year to validate
  - Very strict guidelines had to be followed around sample prep

Bite the bullet, use the opposition product

Control limits

- The statistics for developing control lines on Shewart based control chart is well defined for production systems (lots of statistics).
- However the practical considerations in a production lab mean that it not always as simple to define as three SD.
- It is not always possible to do a big enough sample set to get a valid indication of “common cause variation”
- Also need to consider possible drift of the QA sample and how to deal with possible changes to matrix on changes to QA sample.

Control Limits

- Initially control lines were generated using accepted analytical values for each test and results from external proficiency programs.
- Later long term SD's were assessed for the control samples, converted to coefficients of variation and applied to typical values to generate the control limits.
- It was important that the later process be undertaken so that the real variation for the actual sites method and equipment be the source of the control.
- Accuracy between sites was addressed by proficiency programs.

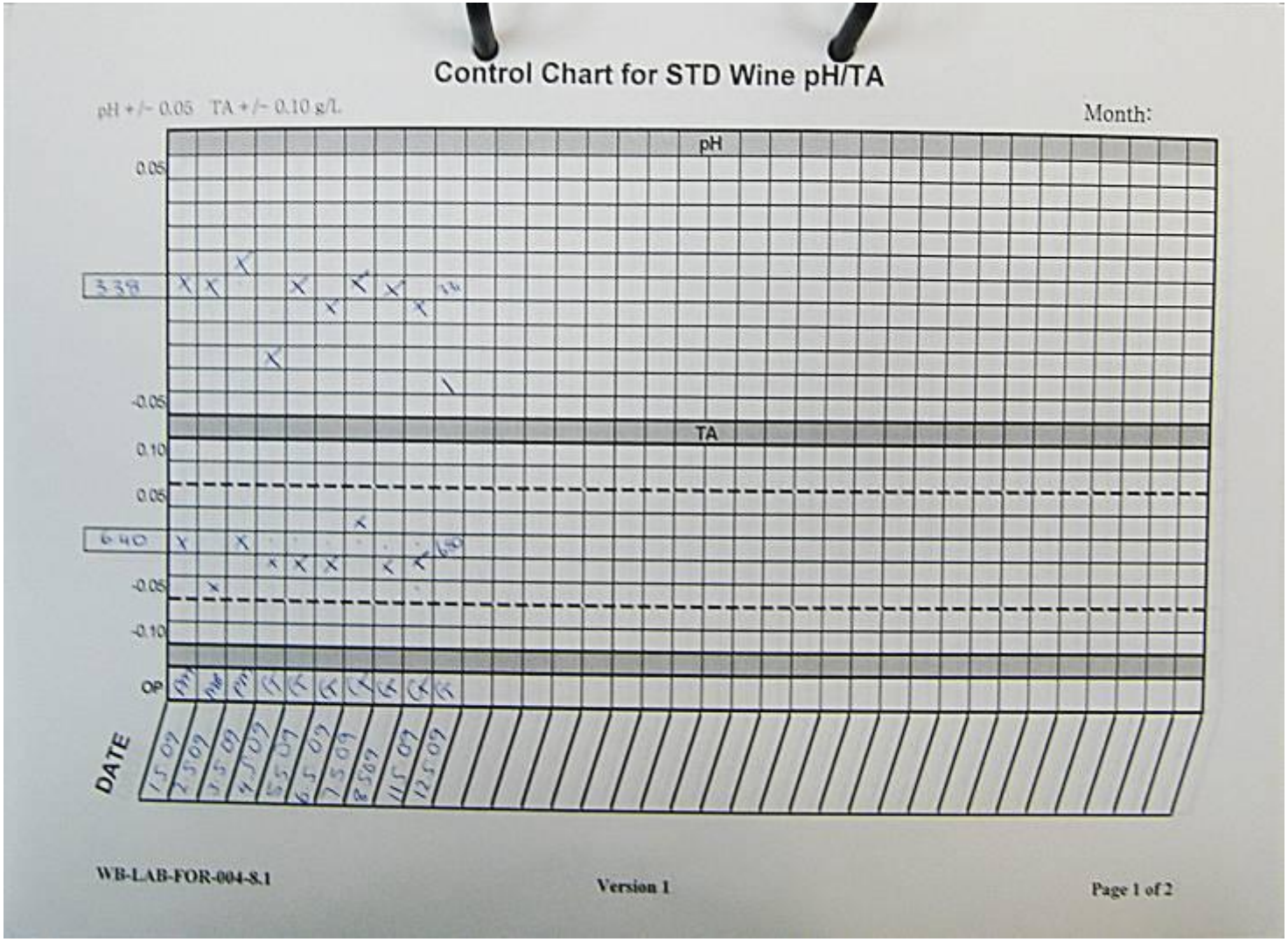
Action Rules

- While the Western Electric and similar systems provide good guidelines for determining if a system is in control, it is important that you carefully assess the particular situation.
- In wine production any one mistake could drive a costly production issue. This was the driving force in limiting time between QA to 10 samples.
- Working with staff it was decided that any out of limit result would spark an assessment of the previous 10 results.
- Should always be horses for courses!
- Once a rule has been made however it must be adhered to religiously.

The physical system

- *Do not add unneeded bureaucracy* should be a mantra in labs.
- It is very tempting to use computer based systems.
- However in many cases after trials of such systems labs have reverted to paper based charts.
- There are a number of reason but primary amongst them are
  - ease of use
  - always accessible
  - are a constant visual aid
- It may be better for the lab manager to be able to access the results remotely, but it is usually the technician doing the work and the technician acting on it.

Typical Control Chart



Did it Work

- One major metric was simply the reduction in time I spent on the phone!!
- It was also evident when investigations for out of spec product began to be started in production rather than in the lab (customer confidence).
- The amount of questioning of methods and issues by lab staff also went up dramatically as their confidence increased.
- In the case of one facility the improvements in efficiency as result confidence went up meant that they managed to absorb a 25% increase in workload over the next two years with improved staff morale.

Case Study 3

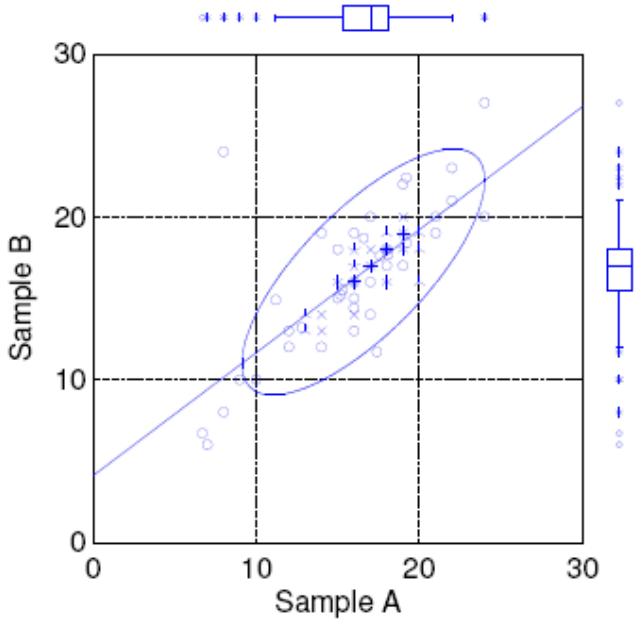
Using proficiency results

Interwinery Analysis Group (IAG)

- An Australian based laboratory proficiency testing group in the wine industry.
- Around 200 laboratory members.
- Every 2 months they provide 2 samples for labs to analyse a range of up 18 analytes.
- The group return the analysed results as Youden plots and summary statistics including individual laboratory Z scores.

www.interwinery.com.au

Free Sulphur Dioxide

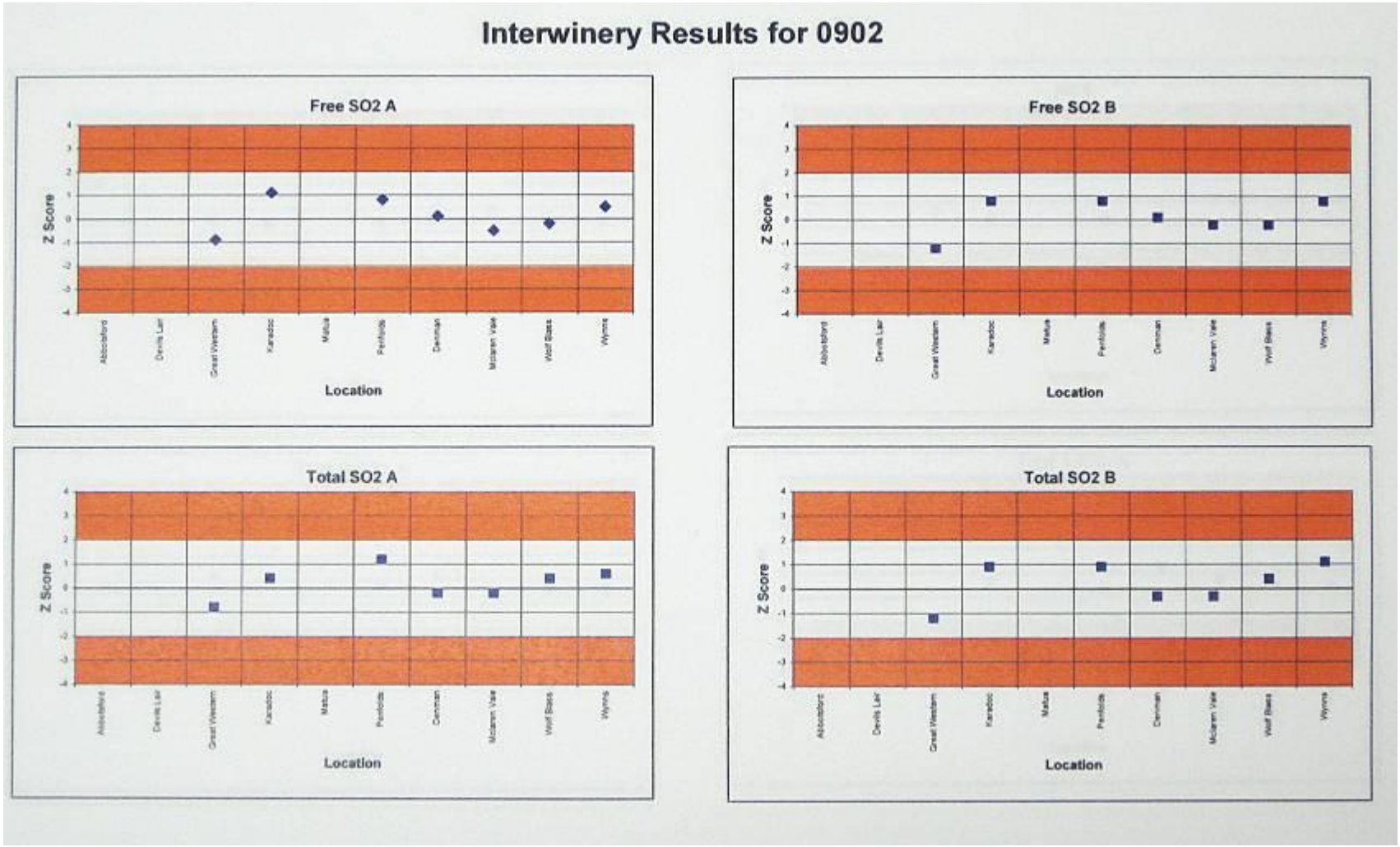


	Free S02 A	Free S02 B
	mg/L	mg/L
N of cases	122	122
Minimum	7	6
Maximum	24	27
Range	17	21
Median	17	17
Mean	17	17
95% CI Upper	17	17
95% CI Lower	16	16
Standard Dev	3	3
C.V. %	17.6	17.6
Lab Result		
Lab Z-Score Result		

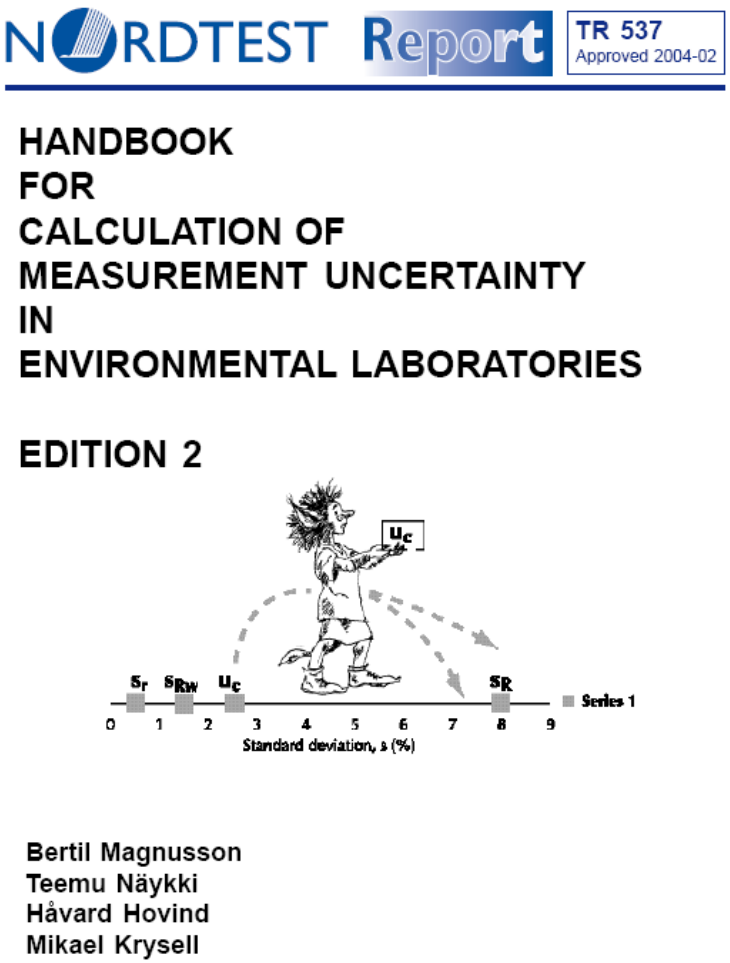
What to do with the results?

- Fosters has 10 participating laboratories in the group.
- Traditionally each laboratory has looked at its results in isolation and treated them on a case by case basis (i.e. reactive process)
- This has done little to improve the either efficiency or accuracy and tends to lead to stressful and not always useful investigations of individual spurious results.
- The question, is how do we get better use out of this information.

Internal Report



Part of the answer.



- The tools developed by the Nordtest group lend themselves well to analysing the outputs from proficiency programs.
- By looking at the results of the contributing labs as a whole v's the Foster's lab group and individual labs it is possible to
  - prioritise areas to improve performance
  - develop valid expectations for the precision of industry standard tests.

So what did we get?

(maths free area!)

Required Accuracy

Analysis	Basic Test	Typical Result	Accuracy Required (+/-)	CV
Acetic	1.000	0.500	0.050	0.100
Alcohol	1.000	12.000	0.100	0.008
Ascorbic	1.000	50.000	10.000	0.200
Ca		50.000	5.000	0.100
Citric		0.500	0.100	0.200
Copper		0.500	0.100	0.200
DCO2		1.000	0.100	0.100
FSO2	1.000	30.000	3.000	0.100
GF	1.000	5.000	0.500	0.100
Iron		1.000	0.200	0.200
K		1000.000	100.000	0.100
Malic	1.000	0.500	0.100	0.200
Na		100.000	10.000	0.100
pH	1.000	3.400	0.050	0.015
RS	1.000	5.000	0.500	0.100
SG		0.994	0.001	0.001
TA	1.000	5.750	0.500	0.087
TSO2	1.000	100.000	5.000	0.050
TURB		0.500	0.100	0.200
VA	1.000	0.500	0.050	0.100

Fosters Wine ANZ - Actual Performance

Analysis	Basic Test	Resultf	SD Fosters	CV
Acetic	1.00	0.41	0.03	0.06
Alcohol	1.00	13.47	0.06	0.00
Ascorbic	1.00	59.58	11.76	0.20
Ca		66.53	10.65	0.16
Citric		0.10	0.01	0.10
Copper		0.23	0.05	0.22
DCO2		0.40	0.03	0.07
FSO2	1.00	22.34	2.46	0.11
GF	1.00	2.99	0.19	0.06
Iron		0.88	0.17	0.20
K		836.47	62.02	0.07
Malic	1.00	0.48	0.09	0.19
Na		44.93	4.77	0.11
pH	1.00	3.41	0.02	0.01
RS	1.00	4.22	0.24	0.06
SG		0.99	0.00	0.00
TA	1.00	5.75	0.34	0.06
TSO2	1.00	81.65	6.89	0.08
TURB		1.05	0.40	0.38
VA	1.00	0.49	0.06	0.12

IWAG Actual Performance

Analysis	Basic Test	Result	SD IWAG	CV
Acetic	1.00	0.41	0.05	0.13
Alcohol	1.00	13.47	0.18	0.01
Ascorbic	1.00	59.58	15.50	0.26
Ca		66.53	12.33	0.19
Citric		0.10	0.04	0.39
Copper		0.23	0.06	0.28
DCO2		0.40	0.14	0.35
FSO2	1.00	22.34	3.70	0.17
GF	1.00	2.99	0.50	0.17
Iron		0.88	0.24	0.27
K		836.47	118.97	0.14
Malic	1.00	0.48	0.16	0.34
Na		44.93	8.51	0.19
pH	1.00	3.41	0.05	0.01
RS	1.00	4.22	0.77	0.18
SG		0.99	0.00	0.00
TA	1.00	5.75	0.72	0.13
TSO2	1.00	81.65	8.41	0.10
TURB		1.05	0.56	0.54
VA	1.00	0.49	0.07	0.15

- Based on 2 years of IAG results
- Brenton Porter, Group Chemist Wine ANZ, Foster's

How realistic were our expectations?

- For the first time the group could define if the expected analytical accuracy was in line with industry performance.
- We could identify test where we were under performing as a group and focus our efforts.

Required Accuracy

Analysis	Basic Test	Typical Result	Accuracy Required (+/-)	CV
Acetic	1.000	0.500	0.050	0.100
Alcohol	1.000	12.000	0.100	0.008
Ascorbic	1.000	50.000	10.000	0.200
Ca		50.000	5.000	0.100
Citric		0.500	0.100	0.200
Copper		0.500	0.100	0.200
DCO2		1.000	0.100	0.100
FSO2	1.000	30.000	3.000	0.100
GF	1.000	5.000	0.500	0.100
Iron		1.000	0.200	0.200
K		1000.000	100.000	0.100
Malic	1.000	0.500	0.100	0.200
Na		100.000	10.000	0.100
pH	1.000	3.400	0.050	0.015
RS	1.000	5.000	0.500	0.100
SG		0.994	0.001	0.001
TA	1.000	5.750	0.500	0.087
TSO2	1.000	100.000	5.000	0.050
TURB		0.500	0.100	0.200
VA	1.000	0.500	0.050	0.100

Fosters Wine ANZ - Actual Performance

Analysis	Basic Test	Resultf	SD Fosters	CV
Acetic	1.00	0.41	0.03	0.06
Alcohol	1.00	13.47	0.06	0.00
Ascorbic	1.00	59.58	11.76	0.20
Ca		66.53	10.65	0.16
Citric		0.10	0.01	0.10
Copper		0.23	0.05	0.22
DCO2		0.40	0.03	0.07
FSO2	1.00	22.34	2.46	0.11
GF	1.00	2.99	0.19	0.06
Iron		0.88	0.17	0.20
K		836.47	62.02	0.07
Malic	1.00	0.48	0.09	0.19
Na		44.93	4.77	0.11
pH	1.00	3.41	0.02	0.01
RS	1.00	4.22	0.24	0.06
SG		0.99	0.00	0.00
TA	1.00	5.75	0.34	0.06
TSO2	1.00	81.65	6.89	0.08
TURB		1.05	0.40	0.38
VA	1.00	0.49	0.06	0.12

IWAG Actual Performance

Analysis	Basic Test	Result	SD IWAG	CV
Acetic	1.00	0.41	0.05	0.13
Alcohol	1.00	13.47	0.18	0.01
Ascorbic	1.00	59.58	15.50	0.26
Ca		66.53	12.33	0.19
Citric		0.10	0.04	0.39
Copper		0.23	0.06	0.28
DCO2		0.40	0.14	0.35
FSO2	1.00	22.34	3.70	0.17
GF	1.00	2.99	0.50	0.17
Iron		0.88	0.24	0.27
K		836.47	118.97	0.14
Malic	1.00	0.48	0.16	0.34
Na		44.93	8.51	0.19
pH	1.00	3.41	0.05	0.01
RS	1.00	4.22	0.77	0.18
SG		0.99	0.00	0.00
TA	1.00	5.75	0.72	0.13
TSO2	1.00	81.65	8.41	0.10
TURB		1.05	0.56	0.54
VA	1.00	0.49	0.07	0.15

We can now manage our expectations.

- Control charts can be constructed based on both valid expectations of group and industry performance.
- Warnings can be based around the group performance.
- Control limits around industry performance.
- Product specifications do not outstrip analytical reality.
- This means that investigations tend to focus on real issues and specifications are based on real capability.
- Having these control expectations based around industry history allows us to leverage this information when dealing with third party providers.

Product specs and method realities

- Many labs quote method capabilities based very much their expectation or “what we have been told”.
- Production specs often don’t take into account the reality of testing.
- A production facility was routinely adding 1mg/l SO2 additions based on the lab results.
- The very best 1 SD that the lab ever achieved on control samples was 3 mg/L
- Do you see the problem here?

We can focus our (limited) resources more intelligently to get real improvements in efficiency.

Laboratory	Performance%		
	0806	0901	0902
Laboratory A	92	92	92
Laboratory B	81	76	86
Laboratory C	78	73	76
Laboratory D	83	82	83
Laboratory E	87	8	
Laboratory F	93	8	
Laboratory G	98	10	
Laboratory H	98	9	
Laboratory I	85	8	

It is important to base individual test analysis on a rolling average for the year to ensure that resources are allocated against real trends.

FSO2

Laboratory	*Count*	*Typical Result (Units)*	*Required Accuracy (+/-)*	*Actual Accuracy (+/-)*	*Performance %*
Devils Lair	2	30.00	3.00	2.08	100
Penfolds	12	30.00	3.00	2.43	100
Rosemount Mclaren Vale	12	30.00	3.00	2.44	100
Denman	12	30.00	3.00	2.85	100
Wynns	12	30.00	3.00	3.14	96
Matua	10	30.00	3.00	3.26	92
Wolf Blass	12	30.00	3.00	3.55	85
Great Western	12	30.00	3.00	4.17	72
Karadoc	12	30.00	3.00	4.86	62

Acknowledgements

- **Brenton Porter** developed the use of the Nordtest methods for the analysis of IAG results.
- **Timra Trethewey** is currently managing the wine QA program.
- **Melinda Cristophersen** who has been developing much of the statistics we now use in our lab systems.
- The lab staff and management from the Foster's group (all 140 of them) who make it all happen.

Case Study 4

Introduction of a global QA program

(no rocket science in this one!)

The Situation

- In 2002 Foster's had 8 Australian and 7 US wine labs.
- In Australia there were also 8 key 3rd party suppliers of product who had integral analytical labs.
- Product was routinely shipped between these facilities both nationally and internationally.
- There was no program in place to ensure that results from different sites were aligned.
- Frequent disputes and recriminations regarding product quality led to production delays.
- In the end it was ***"always the labs fault"***

The Solution (at least part of it)

- Introduce a global program of enforced QA that could be used to
  - Show laboratories were performing to expectations
  - Identify issues quickly before they could impact on production
  - Drive standardisation across the laboratory space.
  - Focus capital and human resources where they would have the biggest impact.

The Restrictions

- No investment in IT infrastructure allowed! Excel at best.
- Overall cost must be limited both for samples and support.
- Added workload to laboratories must be small.
- The program had to have a rapid response rate.
- It must reflect day to day operations of the lab, not special case samples.

What to test

- To manage lab workload the testing was restricted to 5 core analytes.
- However they were to be tested every day as part of lab standard practice.
- They were to be introduced as much as possible as standard samples, not special.
- A range of non-core analytes would be tested monthly.

The samples

- To keep costs down standard 5 litre casks were used.
- Initially these were purchased from the local bottle shop.
- They were despatched to labs once a month to arrive in time for testing at the start of each month.

Reporting and Analysis

- All results were reported by email on a standard Excel result sheet on a weekly basis.
- The consolidated results were return within a week to all labs in the form of another spreadsheet.
- Control values were defined by the 2 SD from the cask mean.
- Later Z charts and charts showing individual lab result ranges were also added.
- Results for 3rd party laboratories were reported separately, but showed the Foster's performance in the form of control lines based on internal lab SD.

What worked!

- For the first time a number of the laboratories had a a QA process!
- Coefficients of Variation (CV) dropped by a whopping 40% on some analytes in the first year.
- Analysis of individual lab variations led to targeted placement of capital and the acceptance of standard methods.
- Disputes over product quality performance shifted from the laboratory to investigation into production and shipping issues (mostly).

What didn’t work.

- Sourcing consistent cask samples and ensuring storage conditions were well managed was problematic leading to noise in results.
- SD measures based on the current test set tended to be a weak tool as
  - n was still relatively small
  - Obviously 5% of results were always going to fall outside the controls.
- Addressed the first issue with samples purchased directly from producer and clear guidelines on handling for the labs.
- The later is being address by using long term historical results to develop warning and control lines.