‘(Re)introduction to statistics: dusting off the cobwebs’

Vicki Barwick
LGC

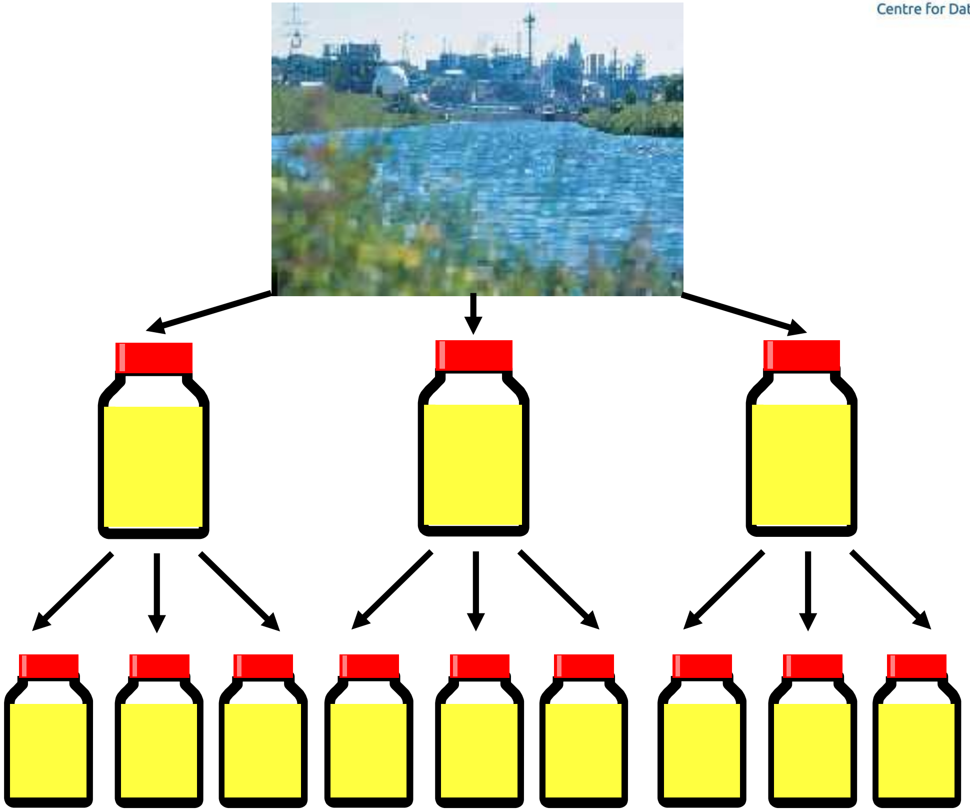
Aoife Morrin
Insight Centre for Data Analysis
DCU

Data – Quality, analysis and integrity workshop
Dublin Castle
14-15 May 2018

Overview

- Sample vs population statistics
- Properties of the normal distribution
- Basic summary statistics
  - mean, standard deviation, relative standard deviation, standard deviation of the mean
- Significance testing
  - procedure
  - different types of test (t-test, F-test, ANOVA)
- Applications of statistics
  - setting limits on control charts
  - interpreting PT scores (z-scores)

Sample vs population (1)



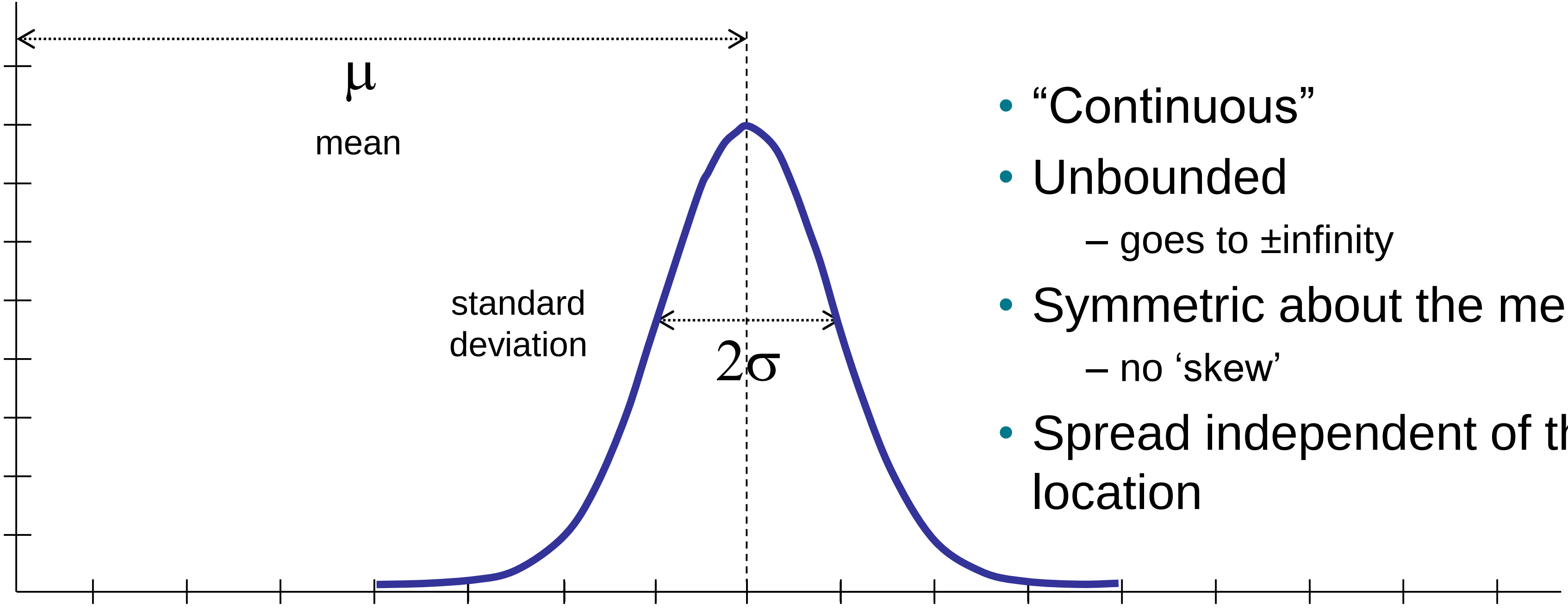
Laboratory samples

Test samples

Sample vs population (2)

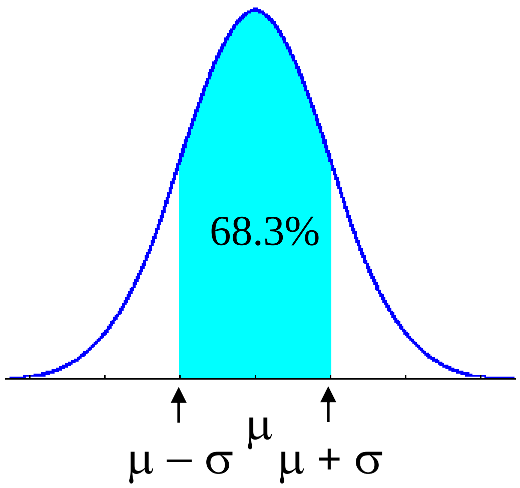
- Laboratories are limited in the number of measurements they can make
- Assume that observations obtained in the laboratory are a random sample from a potentially infinite population
- Population parameters (population mean, population standard deviation)
  - unknown true values of interest
  - represented by Greek alphabet (μ, σ)
- Laboratories use and report ‘sample statistics’
  - provide an estimate of the population parameters
  - represented by Latin alphabet ($\bar{x}$, s)

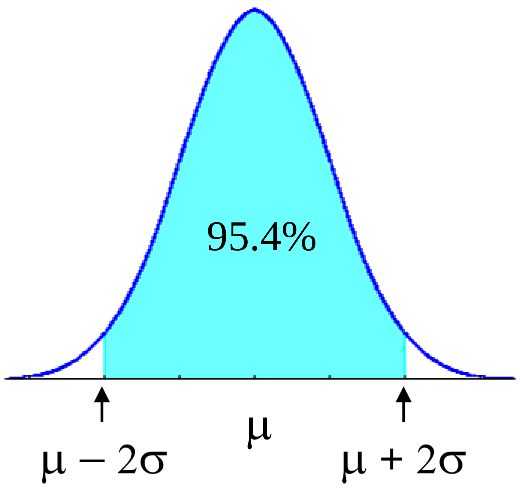
The normal distribution



- "Continuous"
- Unbounded
  - goes to ±infinity
- Symmetric about the mean
  - no 'skew'
- Spread independent of the location

Areas under the normal curve

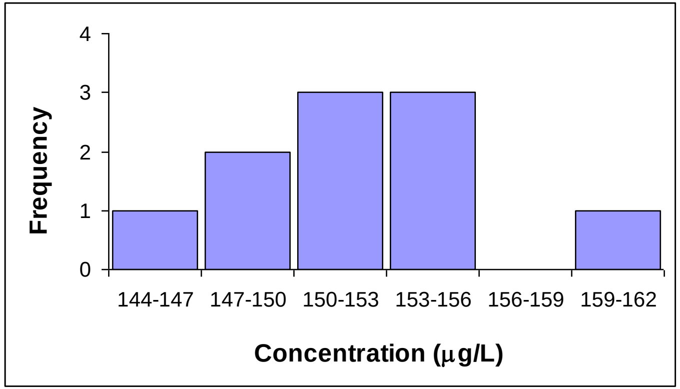




± σ	% population
1.00	68.3
1.64	90.0
1.96	95.0
2.00	95.4
2.57	99.0
3.00	99.7

Summary statistics

Lead (µg/L)	
152	151
155	145
161	155
151	149
156	150



Sample mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 152.5\ \mu g/L$$

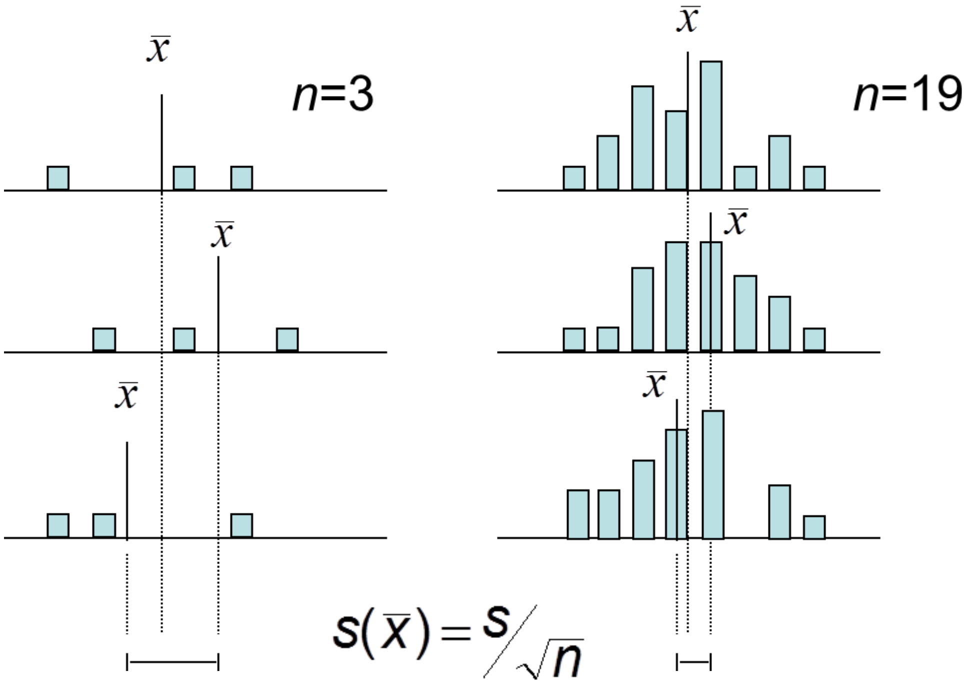
Sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = 4.4\ \mu g/L$$

%relative standard deviation (coefficient of variation)

$$\%rsd = \%CV = \frac{s}{\bar{x}} \times 100 = 2.9\%$$

Standard deviation of the mean



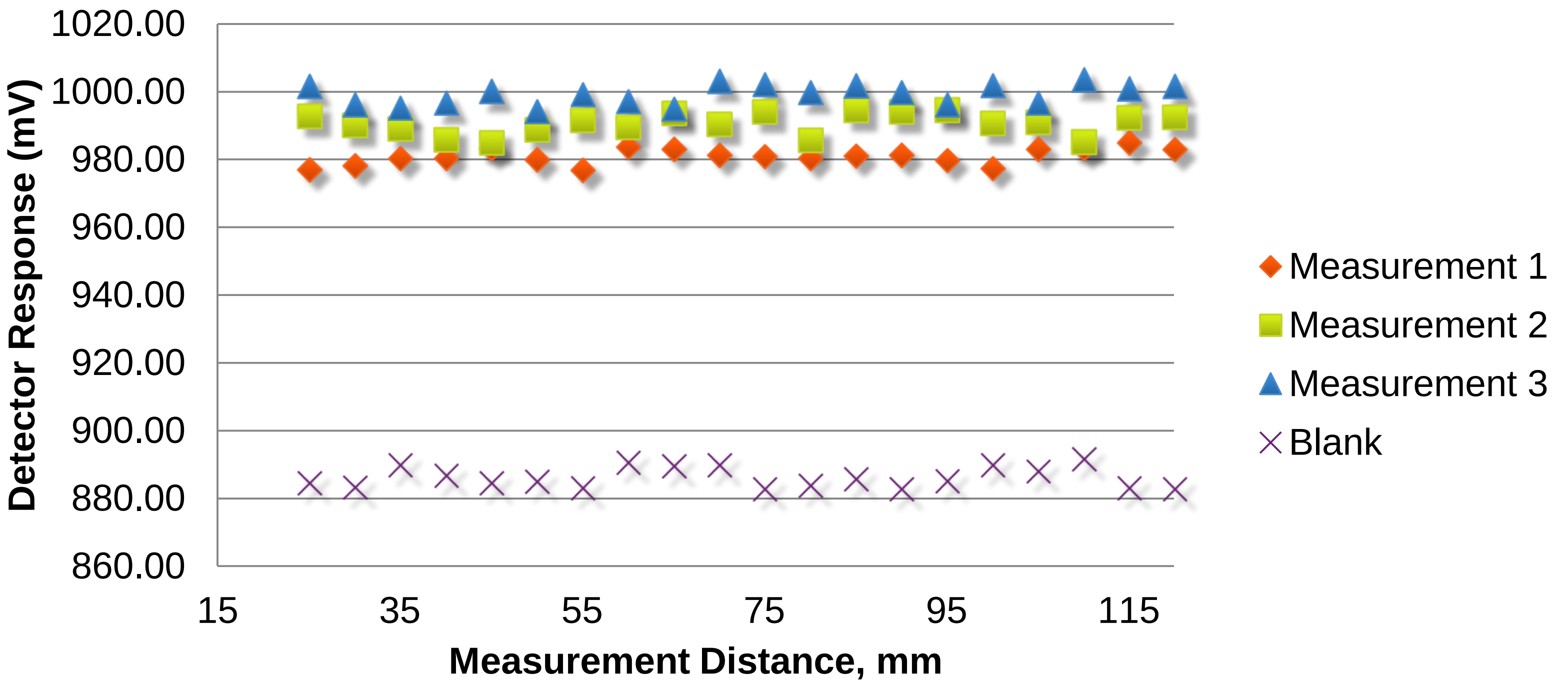
$$s(\bar{x}) = s/\sqrt{n}$$

where *s* is the sample standard deviation

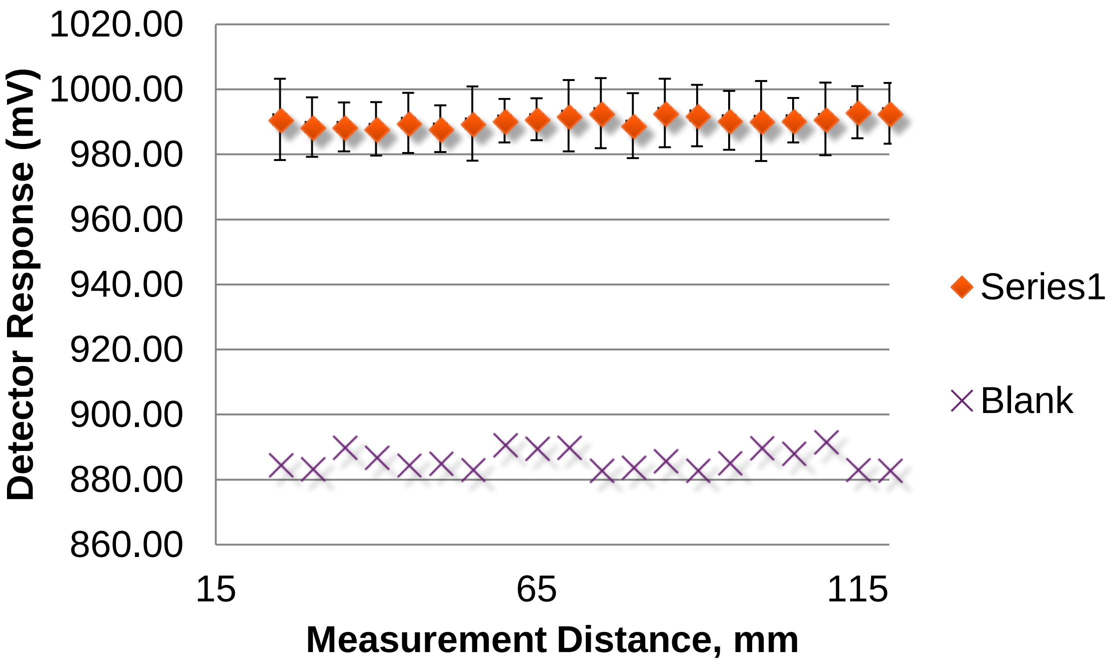
Types of errors

- **Random errors:** cause replicate results to differ from one another, so that the individual results fall on both sides of the average value
    - – affect precision
- **Systematic errors:** cause all the results to be in error in the same sense (e.g. too high)
    - – bias in a method
- **Gross errors:** major errors where the experiment/measurement should be abandoned
    - – should be easily identifiable – clear outliers *etc.*

Processing experimental data – systematic vs random error



Processing experimental data – systematic vs random error

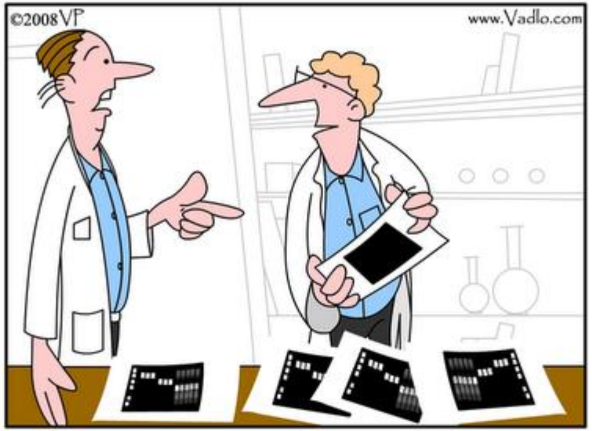


- Error bars quantify the variability
- In this case, the standard deviation is represented by the error bars
- Error bars are representing systematic and random error here!!!

Principles of significance testing

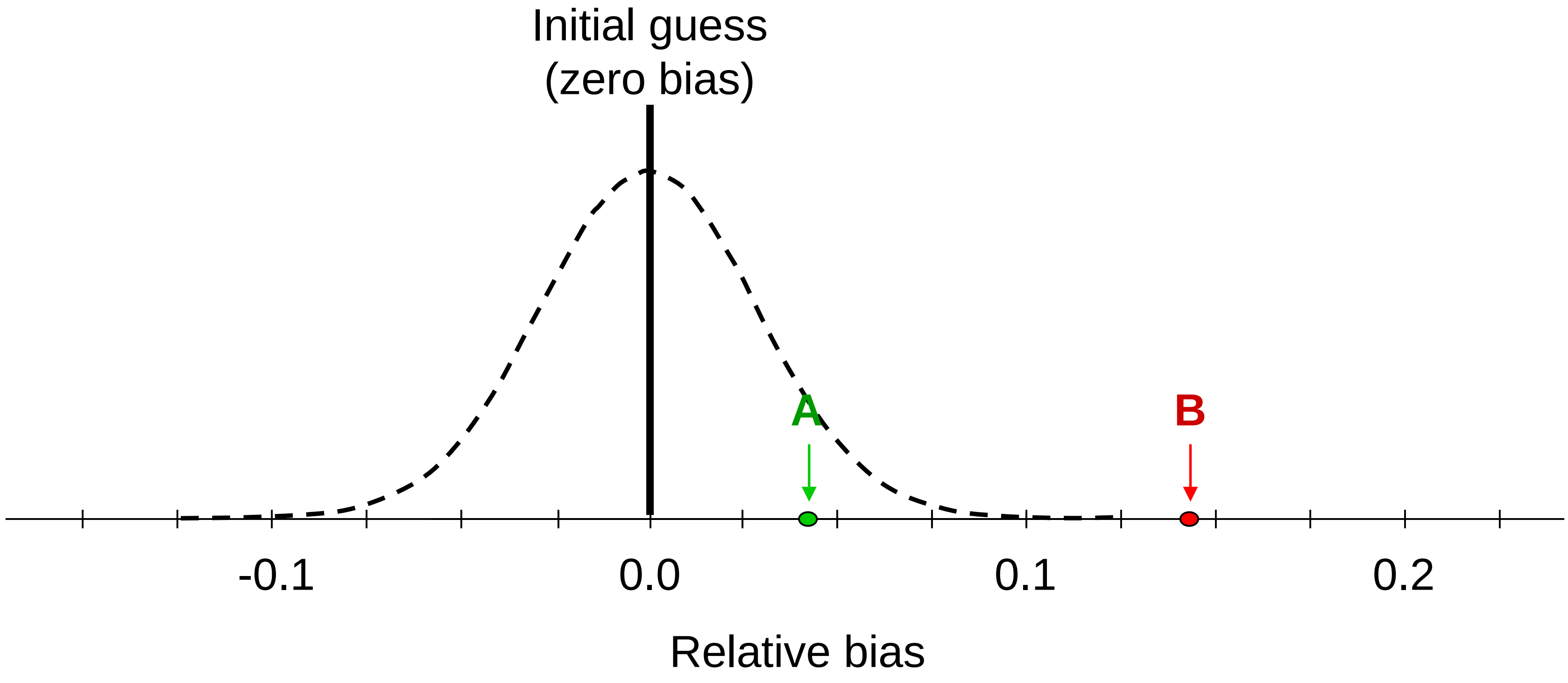
- Make a guess about the true state of affairs (H_0)
  - there is no significant bias/systematic error
  - the precision of two methods is equivalent
  - there are no outliers in a data set
- Ask whether observations are consistent with that guess
  - we calculate the probability that any difference between the observation data and that guess arises solely from random error
- Types of parametric tests
  - *t*-test: Comparing means
  - *F*-test: Comparing variances*
  - analysis of variance (ANOVA): Comparing multiple sets of data

*variance = s^2



Data don't make any sense, we will have to resort to statistics.

Principles of significance testing



Test statistics

- Test statistic

  "A function of a sample of observations which provides a basis for testing a statistical hypothesis"

- Examples:

$$t = \frac{(\bar{x} - x_0)}{s/\sqrt{n}}$$

$$F = \frac{s_1^2}{s_2^2}$$

Significance testing procedure

1. State the question/hypothesis
2. **Select the appropriate test**
3. Choose a level of significance
4. Decide number of tails
5. Calculate degrees of freedom in the data
6. Look up the critical value (tables or software)
7. Calculate the test statistic from the data
8. Compare test statistic with critical value

If test statistic > critical value, result of test is significant → Data not consistent with initial hypothesis

One sample *t*-test

Alternative Hypothesis	t	Tests for
Not equal to x_0 (two-tailed)	$t = \lvert \bar{x} - x_0 \rvert \big/ \left(s/\sqrt{n} \right)$	Any difference?
Greater than x_0 (one-tailed)	$t = (\bar{x} - x_0) \big/ \left(s/\sqrt{n} \right)$	Exceeding reference value/upper limit
Less than x_0 (one-tailed)	$t = (x_0 - \bar{x}) \big/ \left(s/\sqrt{n} \right)$	Below reference value/lower limit

Significance: $t > t_{crit}$

One sample t-test - example of bias evaluation

Data: Bias evaluated through repeat analysis of anhydrous milk fat CRM

- **certified value for cholesterol: 274.9 mg/100 g**
- **mean of results from 11 replicate analyses: 269.3 mg/100 g**
- **standard deviation of results: 1.692 mg/100 g**

- State your question:
  - is there a significant difference between the mean of results from the replicate analysis of a CRM and the certified value?
- Select the test:
  - comparing a mean with a reference value – single sample *t*-test
- Choose level of significance:
  - 5% significance (95% confidence)
- Decide number of tails:
  - two-tailed (interested in a difference in either direction)

Example (continued)

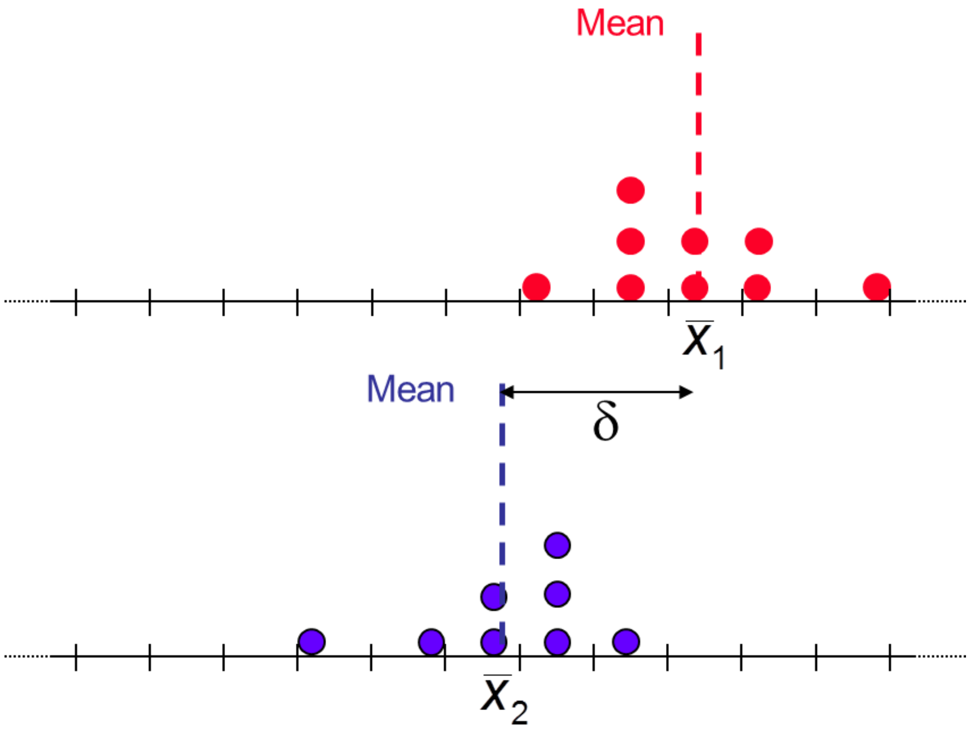
- Calculate degrees of freedom:
  - – degrees of freedom: n-1 = 10
- Look up critical value:
  - – from tables/software, two tailed Student t value for 95% confidence and 10 degrees of freedom: 2.228
- Calculate test statistic from experimental data:

$$t = \frac{|\bar{x} - x_0|}{s/\sqrt{n}} = \frac{|269.33 - 274.7|}{1.692/\sqrt{11}} = 10.53$$

- Calculated t > critical value (t_{crit}):
  - – → Mean value of the experimental results is significantly different from certified value

Significance testing between sets of data

Two-sample *t*-test



$$t = \frac{\bar{x}_2 - \bar{x}_1}{s_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{pool} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$\nu = n_1 + n_2 - 2$$

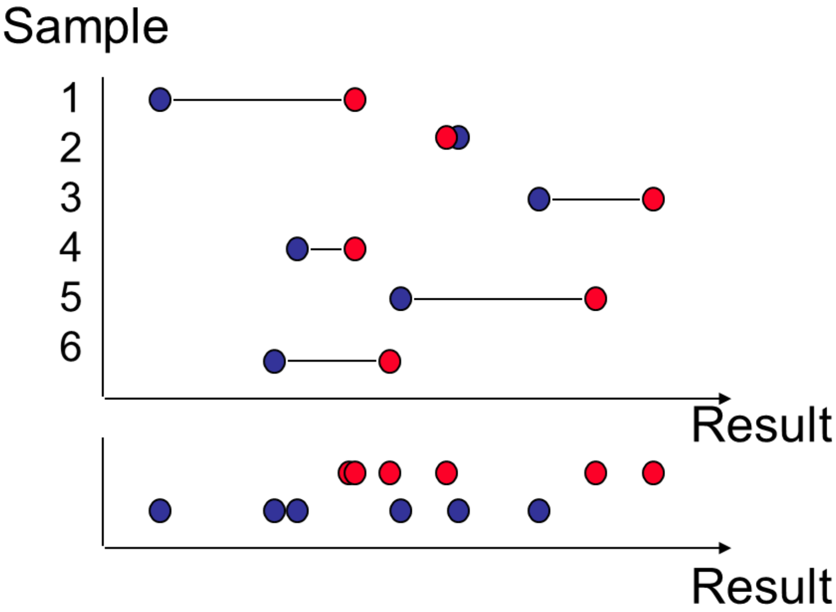
(Assumes equal variance)

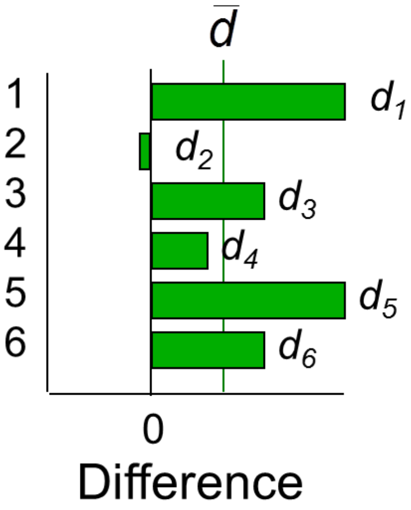
Two sample t-test - example

- BET surface area analysis was carried out on CNT samples that were untreated and treated by argon plasma (m^2/g)
- (Assuming variances to be the same,) does the argon plasma treatment significantly improve surface area?

***Untreated* (*m^2/g*)**	***Argon plasma treated* (*m^2/g*)**
184	281
192	406
194	362
192	327
185	327
191	376
207	

Significance testing between paired samples
Paired sample *t*-test





$$t = \frac{\bar{d}}{s(d)/\sqrt{n}}$$

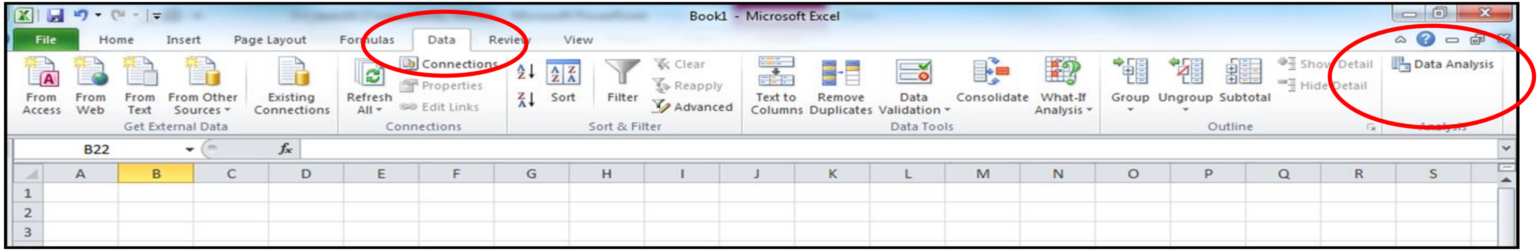
Need *Natural Pairing* of the data

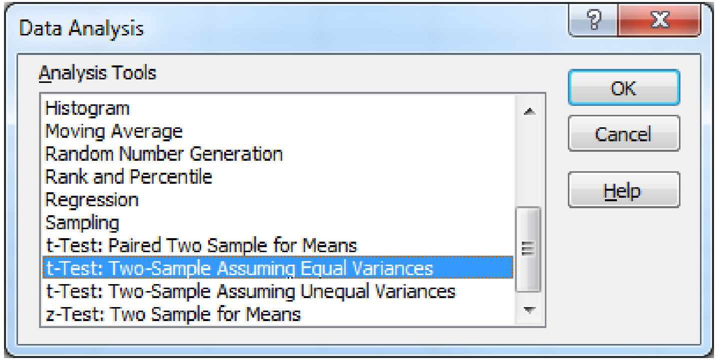
Paired sample t-test - example

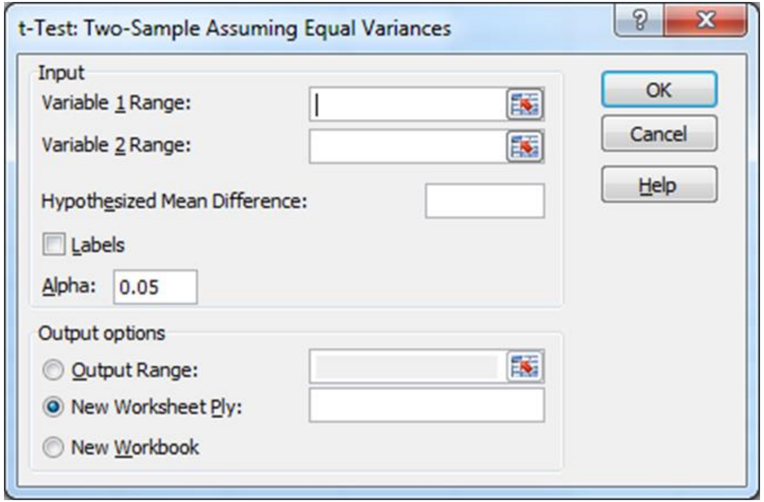
- Where two **methods of analysis** are compared by applying both methods to analyse the SAME set of test materials
- The paracetamol concentration (mg/g) was determined in tablet batches by two different methods – UV and IR – do the methods give the same results?

Tablet Batch No.	UV	Near-IR
1	84.63	83.15
2	84.38	83.72
3	84.08	83.84
4	84.41	84.20
5	83.82	83.92
6	83.55	84.16
7	83.92	84.02
8	83.69	83.60
9	84.06	84.13
10	84.03	84.24

Excel® data analysis tools







Interpreting significance test results in Excel®

- Excel also quotes the results of a significance test in terms of a probability (p-level)
- Probability of obtaining a test statistic at least as extreme as the one that was actually observed assuming that H_0 is true
- If p-level > 0.05 – it is not significant, i.e., your data is likely to agree with the H_0
- If p-level < 0.05 – it is significant, i.e., your data is not likely to agreee with the H_0

Publishing/reporting stats - examples

- *"A t-test was performed to determine if there was a significant difference between film thickness when films were deposited by spin-coating and printing. The mean film thickness for spin-coating ($\bar{x}$=772.57, s = 13.56, n=7) was not significantly different to that for printing ($\bar{x}$=780.86, s=10.42, n=7), test statistic = 1.28, two-tail, p=0.22, providing no evidence that film thickness was influenced by the method of deposition."*
- *"A t-test was performed to determine if there was a greater swelling response achieved in the presence of catalase. The difference in swelling responses was found to be significant after a swelling time of 495 min ($p<0.05$; one-tailed; n=3)."*

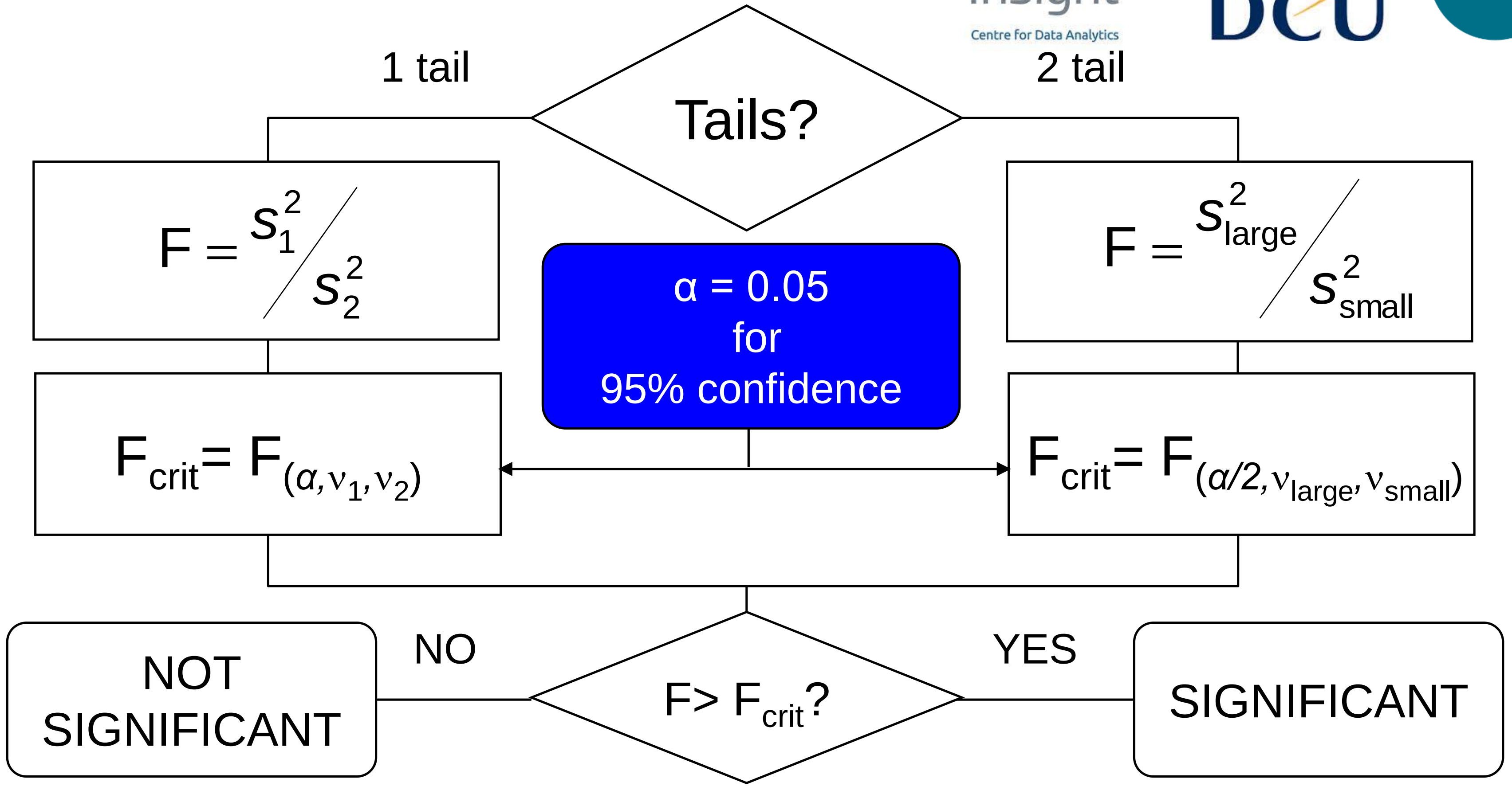
The F-test

- To compare the spread, use the ratio of variances:

$$F = {s_1^2}/{s_2^2}$$

- This ratio, the ‘F-statistic’, can be compared with values in tables (the ‘F-test’)

Rules for the F-test



Finding F_{crit}

- Calculate degrees of freedom (ν)
  $\nu_1 = n_1 - 1$ $\nu_2 = n_2 - 1$
- Use standard table of values
- Or use Excel Data Analysis Tool or F.INV.RT function
- Significance: $F > F_{crit}$

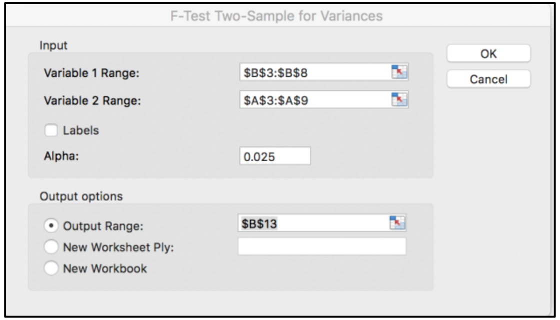
$$F = {s_1^2}/{s_2^2}$$

	ν_1			
ν_2	**3**	**5**	**9**	**∞**
3	15.4	14.9	14.5	13.9
5	7.8	7.1	6.7	6.0
9	5.1	4.5	4.0	3.3
∞	3.1	2.6	2.1	1.0

97.5% (α=0.025) 1-tailed F table
(used for 95% (α=0.05) 2-tailed test)

Excel output – F-test

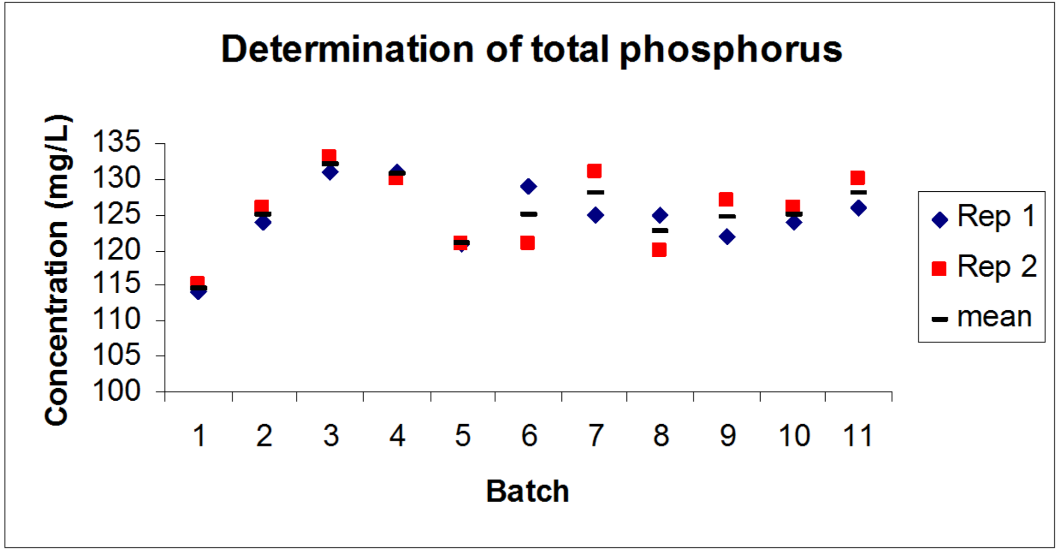
- BET analysis question from earlier – we want to verify if the assumption is true – that the variances are the same?
- Note: need to use an Alpha value of 0.025 for a 95% confidence level



F-Test Two-Sample for Variances		
	Variable 1	Variable 2
Mean	346.5	192.1429
Variance	1940.3	57.14286
Observations	6	7
df	5	6
F	33.95525	
P(F<=f) one-tail	0.000251	
F Critical one-tail	5.987565	

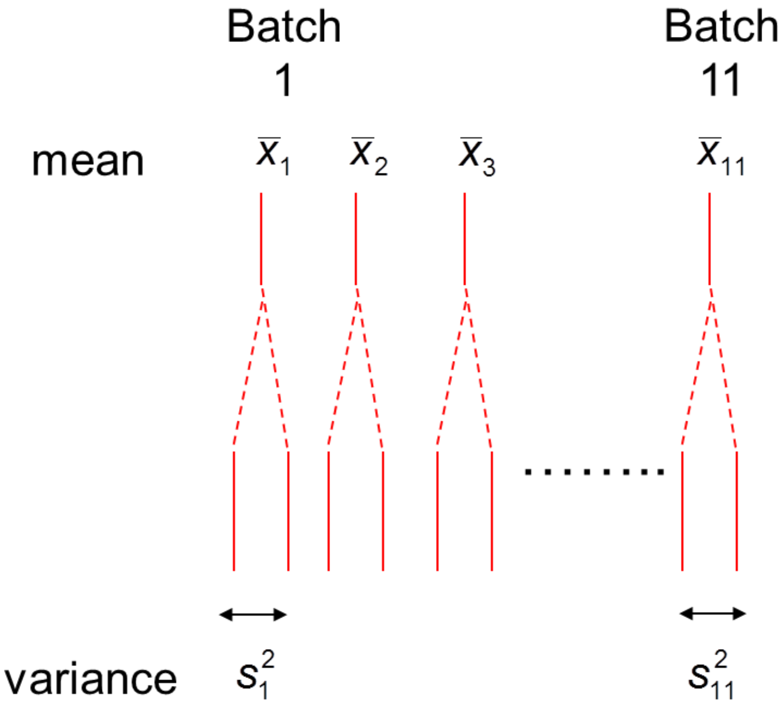
Comparing multiple groups of data

- Variation between duplicates (within-batch)
- Variation between batches – measurements made on different days



- Does the variation increase significantly when measurements are made on different days?

Within- and between-group effects



Total variance has contributions from

- Random variation between duplicates (within-batch)
- Variation between results obtained in different batches (between-batch)

Analysis of variance (ANOVA)

- ANOVA separates different sources of variation
  - e.g. the within- and between-batch variation in results
- Different sources of variation can be compared to determine whether they are significantly different
  - e.g. is the between-batch variability in results significantly greater than the within-batch variability?
- H_0 is that all samples are drawn from same population
- Method validation precision study
  - can be useful to know where variation in results is coming from
    - within-batch vs. between-batch

ANOVA: single factor - example

- 4 different batches of disposable, screen-printed electrodes are used to fabricate a lactate biosensor. The electrodes are modified with enzyme and their amperometric responses to lactate are measured (μA) (n=3). Before combining all of the data, one-way ANOVA is used to determine if the different batches of electrodes are giving statistically different results.

Replicates	**Batch 1**	**Batch 2**	**Batch 3**	**Batch 4**
1	10.2	10.6	10.3	10.5
2	10.2	10.8	10.4	10.7
3	10.0	10.9	10.7	10.4
Mean				

ANOVA: single factor in Excel®

- There are sources of error in all measurements, so its normal for the means to be different. We want to determine if the error is:
  - – just in the measurement (random error) or
  - – between the batches (systematic error)

- We have two potential sources of variance:
  - – run to run errors
  - – the batches may actually be different

ANOVA: single factor in Excel®

Sum of Squares

Mean Square (σ_0^2)

$MS_{between}/MS_{within}$

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.61583333	3	0.20527778	7.94623656	0.00876534	4.06618055
Within Groups	0.20666667	8	0.02583333			
Total	0.8225	11				

- SS – sum of the squares
  - between groups: the difference in the means between batches
  - within groups: the random error within a given batch
- df – degrees of freedom
- MS – mean of the SS values (SS/df)

ANOVA: single factor in Excel®

- H_0: All samples are drawn from same population. Specifically, there is no major difference between means of batches
- $F > F_{crit}$, H_0 is rejected

OR

- P-value < 0.05 H_0 is significant
- Therefore, samples are not drawn from same population. Specifically there is a major difference between the means of the batches

NOTE: ANOVA does NOT indicate WHICH batch is different from others – Need to look at a post-hoc analysis

ANOVA: Single Factor - Total Phosphorus

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	459.8182	10	45.98182	5.620	0.004312	2.854
Within Groups	90.00	11	8.181818			
Total	549.8182	21				

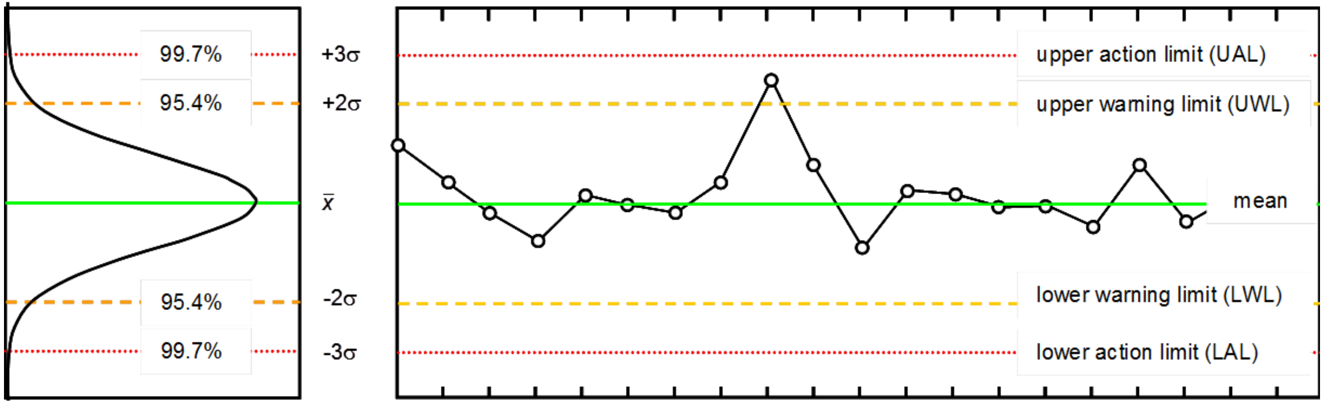
$F>F_{crit}$, $P<0.05 \Rightarrow$ Significant difference between results obtained in different batches

Applications of statistics in QC & QA

- Interpretation of quality control results
  - control charts
- Proficiency testing scores

Shewhart chart (x-chart)

- Used to monitor bias and precision
- Individual control values plotted in time ordered sequence



- Key features
  - – central line
  - – upper and lower warning limits
  - – upper and lower action limits

Also known as an 'individuals chart'

Scoring PT results

- PT results commonly reported as a performance score
  - calculated by the scheme organiser
- Z-score (most common score in analytical chemistry) is calculated as

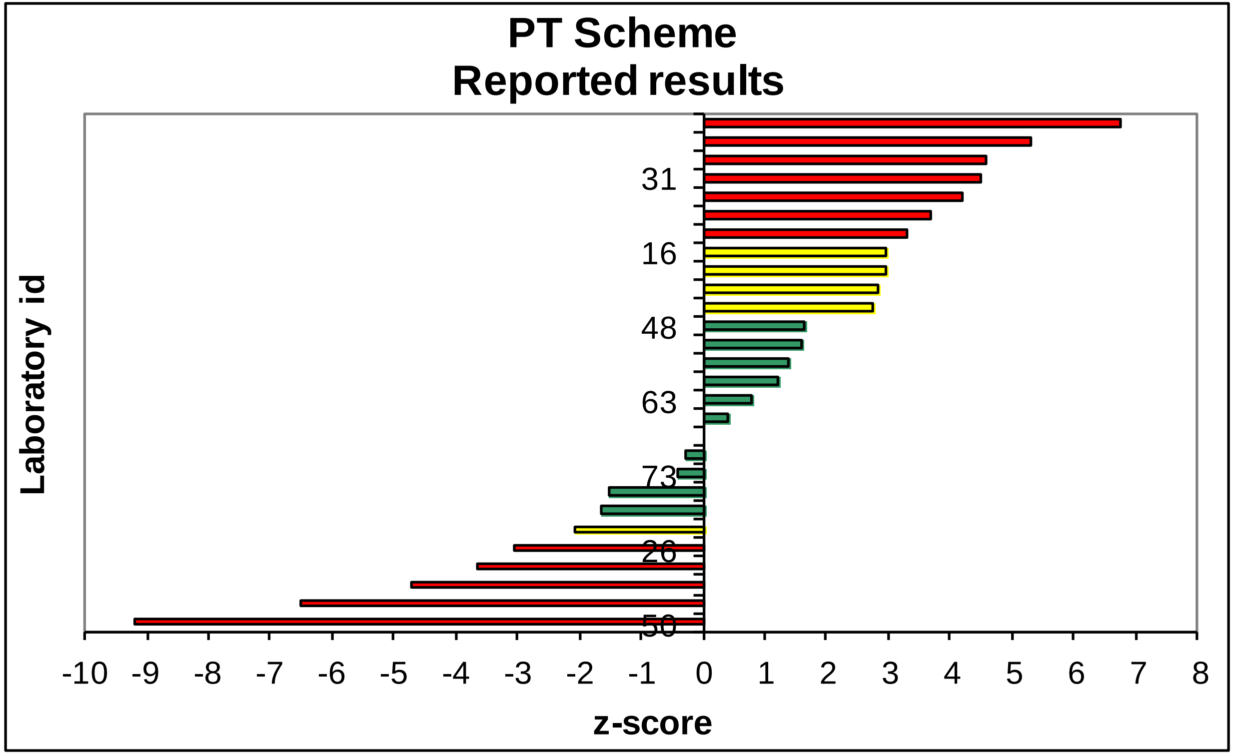
$$z_i = \frac{(x_i - x_{pt})}{\sigma_{pt}}$$

x_i is the result submitted by the participant

x_{pt} is the assigned value determined by the co-ordinator

σ_{pt} is the is the standard deviation for proficiency assessment

Interpreting PT results



$\lvert z \rvert < 2$	Satisfactory
$2 < \lvert z \rvert < 3$	Questionable
$\lvert z \rvert > 3$	Unsatisfactory

Thank you for listening

Enjoy the rest of the workshop