How to interpret the z-scores in survey reports (theory and practical examples)

Panagiotis Tsiamyrtzis [1] Piet Meijer [2]

[1]Dept. of Mech. Eng., Politecnico di Milano, Italy &
Dept. of Statistics, Athens University of Economics and Business, Greece

pt@aueb.gr

[2]Managing Director ECAT

p.meijer@ecat.nl

Leiden, September 16, 2022

Contents

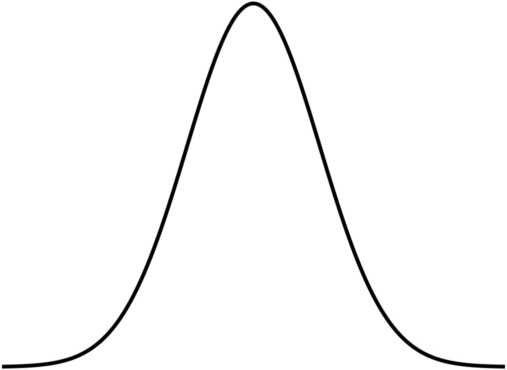
Z-score definition

Z-score definition

- According to **ISO 17043** standard, the External Quality Assessment (EQA) aims to evaluate the deviation of the measured result from the assigned value.
- ECAT's EQA program performs this task with the statistical monitoring of the Z-scores.
- But what is really a z-score?
- Z-score is a standardized measure that can provide valuable information for the state of the laboratory, within a population of similar laboratories.
- On a pure statistical point of view, there is an evaluation of each lab in regard of the analytical performance in its own peer group.
- Let's start by explaining the statistical underline theory.

Basics on the normal distribution

Do you recognize the following statistical continuous distribution?



Yes! This is the **Normal** (aka **Gaussian**) distribution.

Basics on the normal distribution

- The Normal distribution is fully described by two quantities:
  - The mean μ which indicates where the Normal is centered at.
  - The standard deviation σ which reflects how "spread" is the distribution around its center (mean).
- If the random variable X is Normally distributed then we denote it by:

$$X \sim N\left(\mu, \sigma^2\right)$$

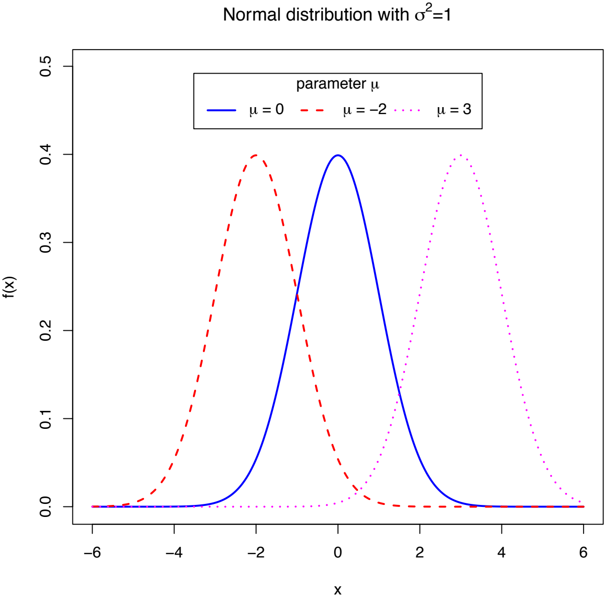
- The Normal random variable for which we have $\mu = 0$ and $\sigma = 1$ is called the **standard** Normal distribution and is denoted by Z, i.e.

$$Z \sim N(0,1)$$

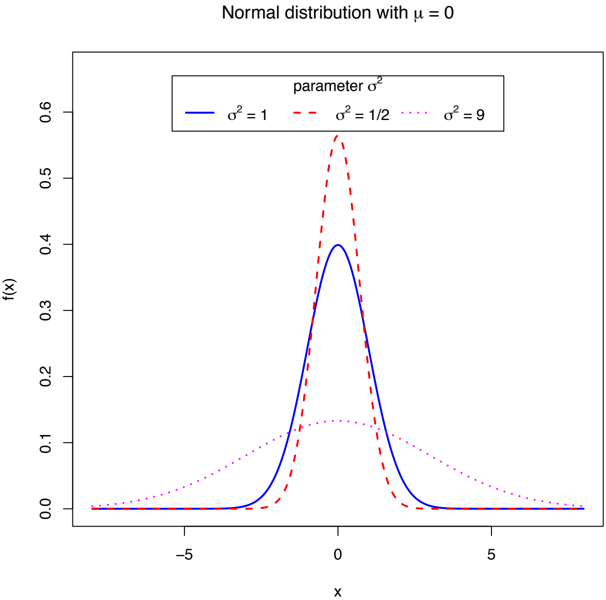
- We can move from $X \sim N\left(\mu, \sigma^2\right)$ to $N(0,1)$ and back:

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \qquad \text{and} \qquad X = \mu + \sigma Z$$

Basics on the normal distribution



Basics on the normal distribution



Z-score definition

- The z-scores are based on the standard Normal distribution.
- Let's assume that we have a homogeneous population of laboratories and we provide the same sample to all. Each one uses it own equipment and reports back a single measurement.
- Thus for each laboratory i we get a value X_i, for which we assume:

$$X_i \sim N(\mu, \sigma^2)$$

  with μ and σ being **unknown** to the laboratories.
- For known (or estimated) μ and σ we can perform the standardization of lab's i result by calculating the z-score, where:

$$z_i = \frac{X_i - \mu}{\sigma}$$

Univariate z-score evaluation

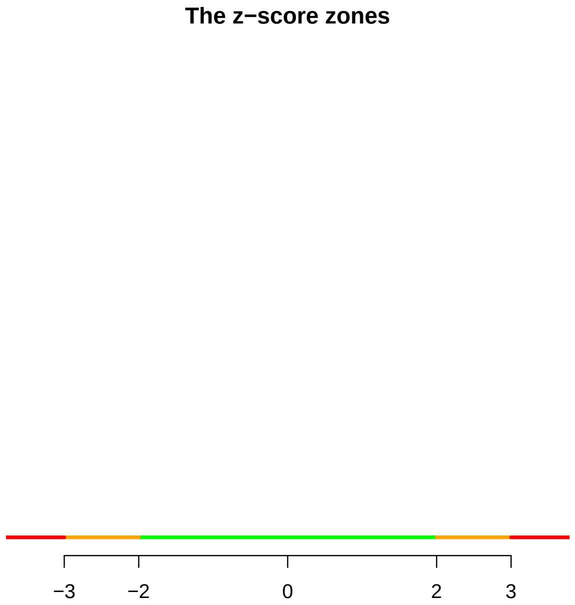
Univariate z-score evaluation

- The Normal is well known for its $68 - 95 - 99.7$ rule, i.e.:

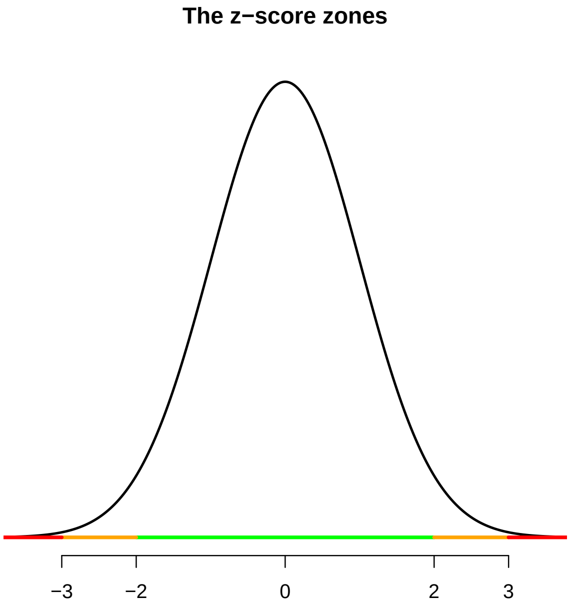
$$\begin{aligned} P(|X-\mu| \leq \sigma) &= P(|Z| \leq 1) &= 0.6826 \\ P(|X-\mu| \leq 2\sigma) &= P(|Z| \leq 2) &= 0.9544 \\ P(|X-\mu| \leq 3\sigma) &= P(|Z| \leq 3) &= 0.9973 \end{aligned}$$

- The above property establishes the well known alarming zones of z-scores (**ISO 17043** recommended by the **ISO 15189** norm). Precisely:
  - If $-2 \leq$ z-score ≤ 2,
    then we give a **Green** card and we interpret it as “Satisfactory”
  - If $-3 <$ z-score ≤ -2 or $2 \leq$ z-score < 3,
    then we give a **Orange** card and we interpret it as “Need Attention”
  - If z-score ≤ -3 or z-score ≥ 3,
    then we give a **Red** card and we interpret it as “Unsatisfactory”

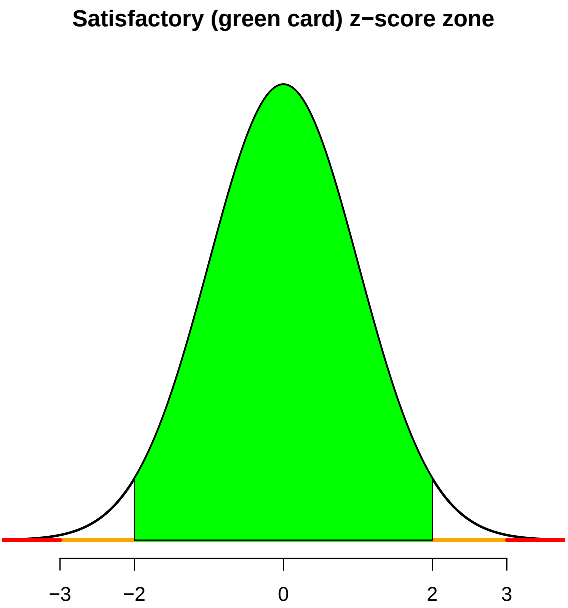
Univariate z-score evaluation



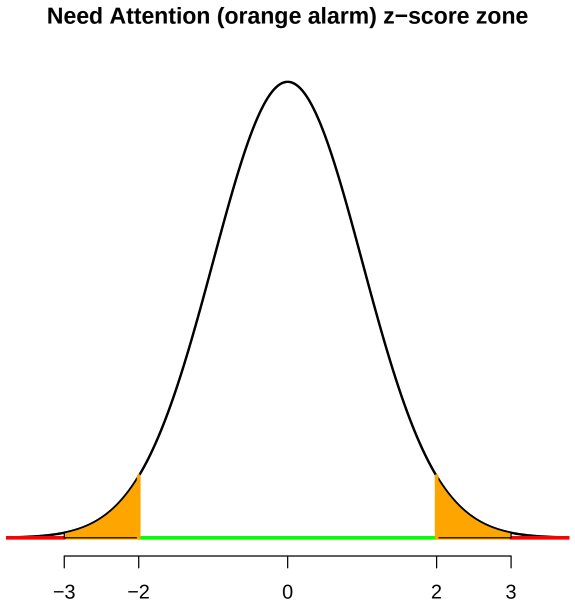
Univariate z-score evaluation



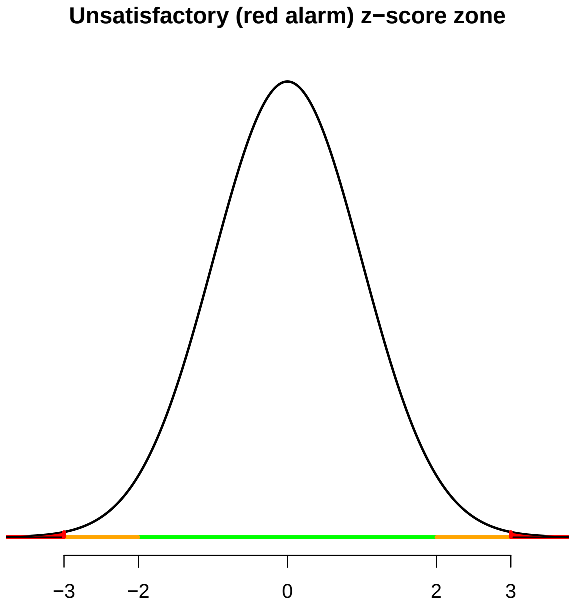
Univariate z-score evaluation



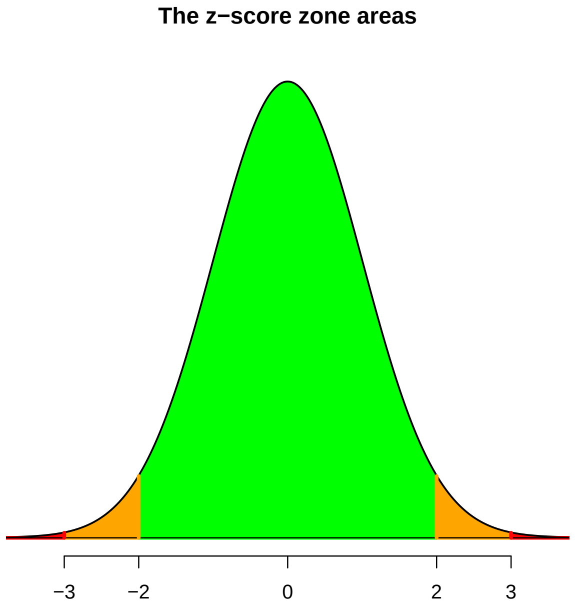
Univariate z-score evaluation



Univariate z-score evaluation



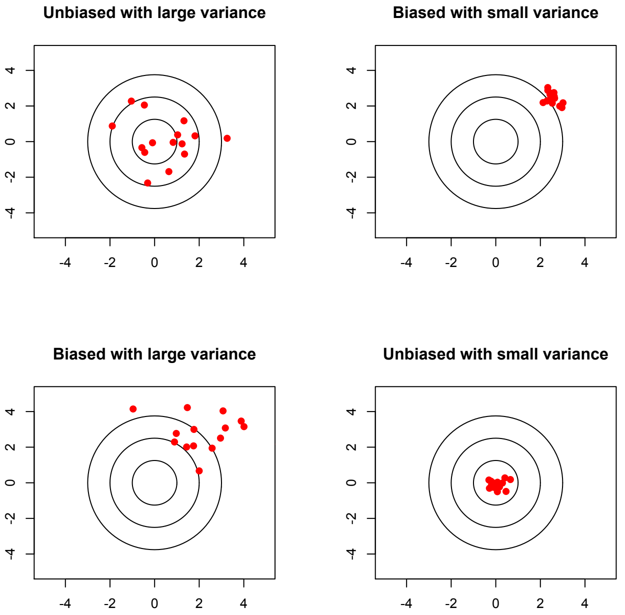
Univariate z-score evaluation



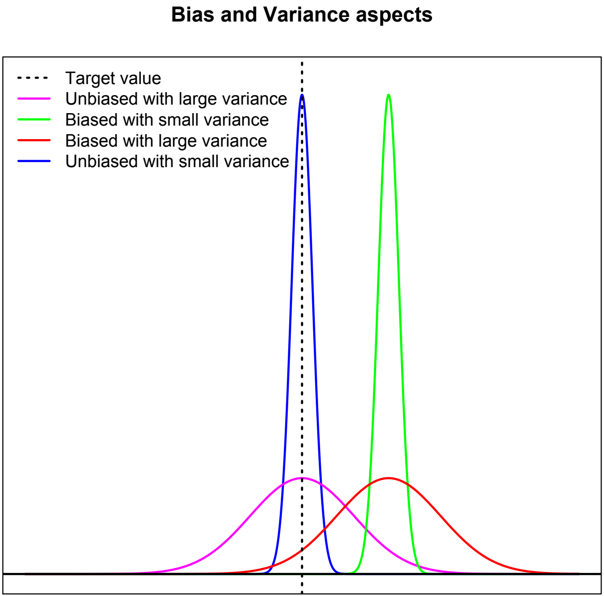
Univariate z-score evaluation

- Assuming the lab performs under the In Control (IC) status there is a risk to get an alarm (0.27% for a red alarm and 4.28% for an orange alarm).
- What happens though if the lab is not "well aligned" with the IC distribution established by the EQA organization?
- A lab performing under Out Of Control (OOC) conditions will tend to have higher (in absolute value) z-scores, leading to an elevated alarm rate.
- The two major OOC issues are related to:
  - **Bias**: how do we perform on average? (biased or unbiased?)
  - **Random Error**: how variable (uncertain) are we?

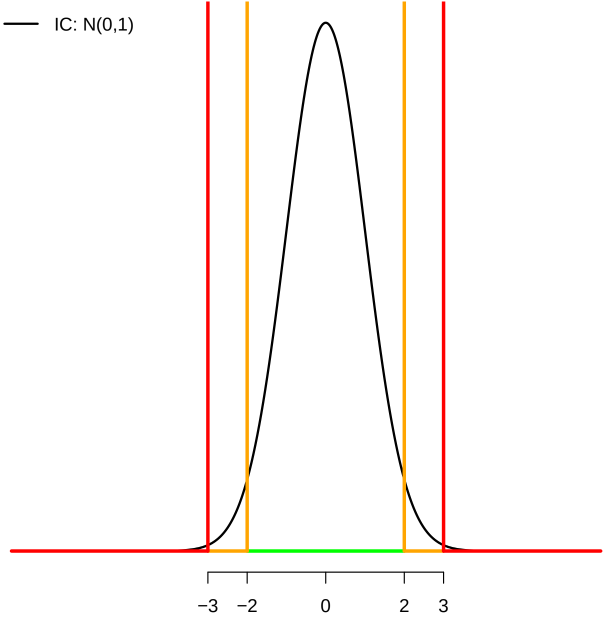
Bias and Uncertainty



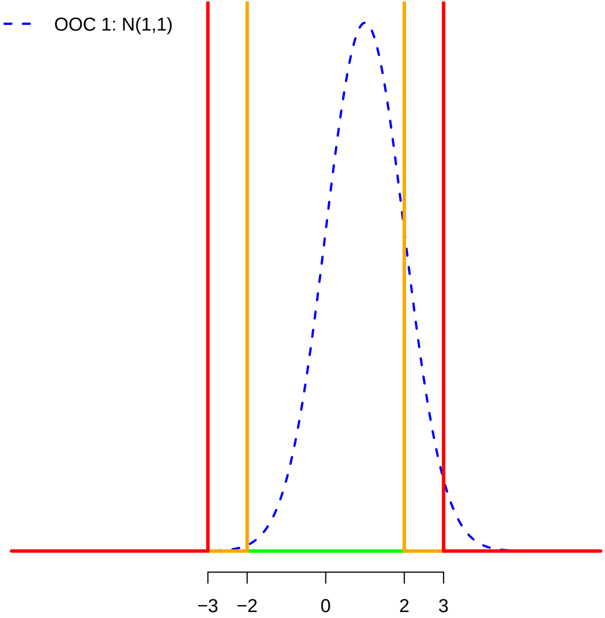
Bias and Uncertainty



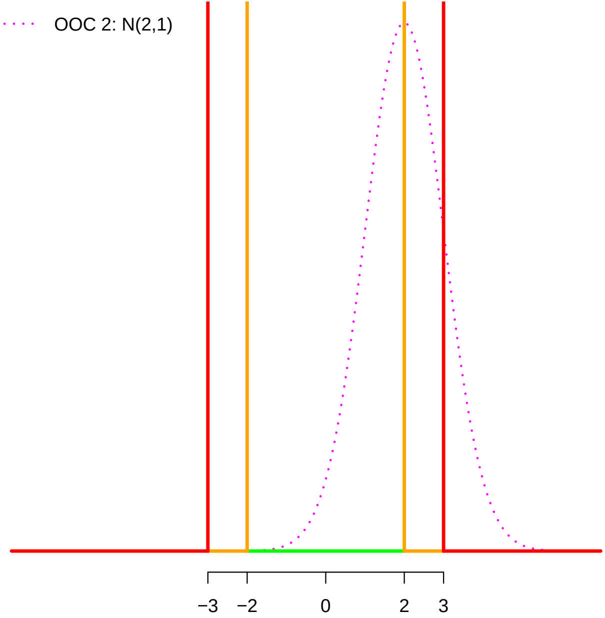
Alarms for various lab's OOC μ cases



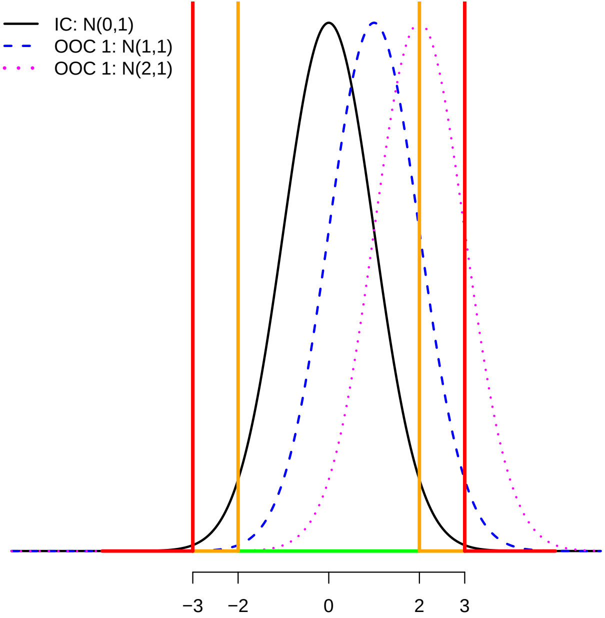
Alarms for various lab's OOC μ cases



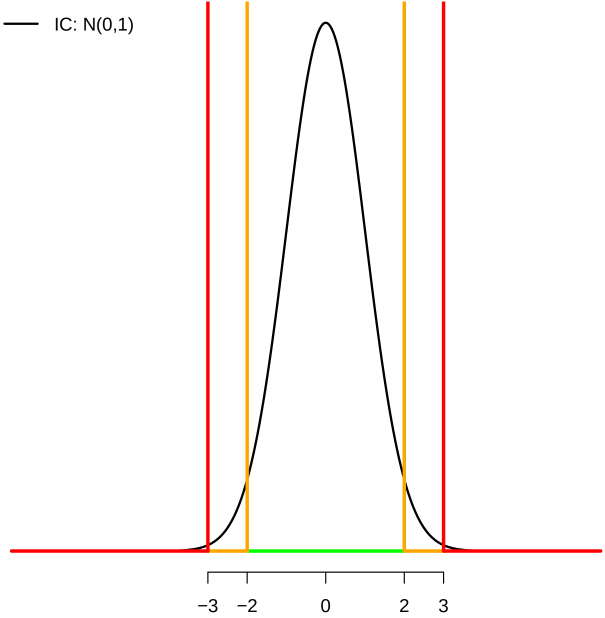
Alarms for various lab's OOC μ cases



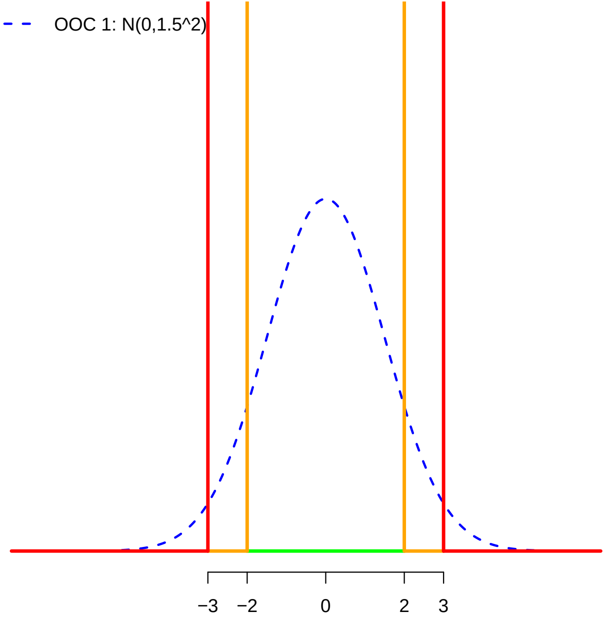
Alarms for various lab's OOC μ cases



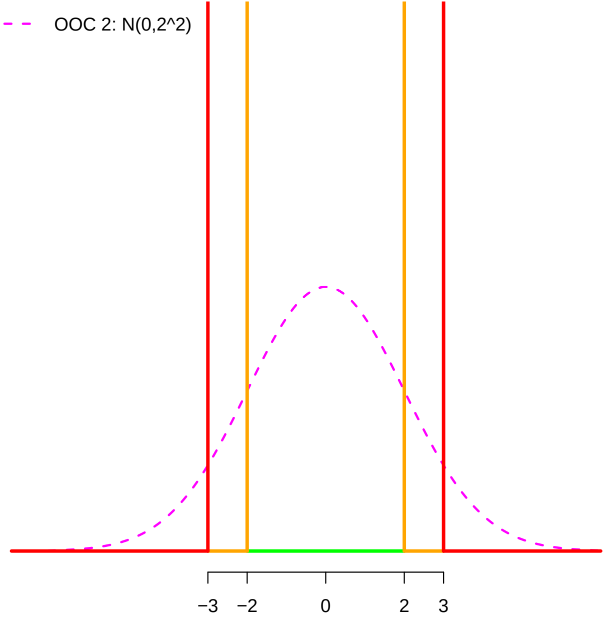
Alarms for various lab's OOC σ cases



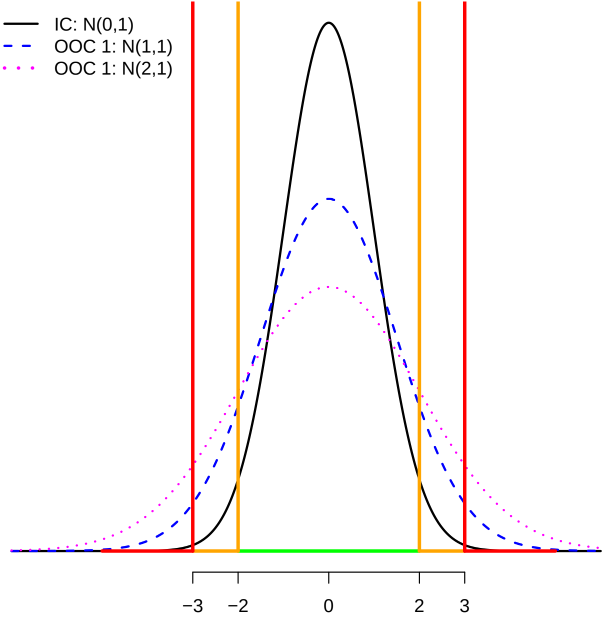
Alarms for various lab's OOC σ cases



Alarms for various lab's OOC σ cases



Alarms for various lab's OOC μ cases



Alarms for various lab's OOC μ or/and σ cases

Mean (μ)	St. Dev (σ)	Green Card	Orange Alarm	Red Alarm
0	1	95.45%	4.28%	0.27%
1	1	84.00%	13.72%	2.28%
2	1	50.00%	34.14%	15.86%
0	1.5	81.76%	13.69%	4.55%
0	2	68.27%	18.37%	13.36%
1	1.5	72.48%	18.02%	9.50%
1	2	62.47%	19.39%	18.14%
2	1.5	49.62%	25.09%	25.29%
2	2	47.72%	20.80%	31.48%

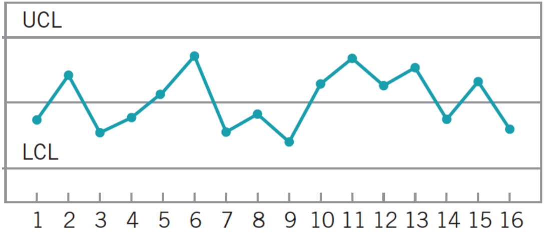
Bivariate z-score evaluation

Bivariate z-score evaluation

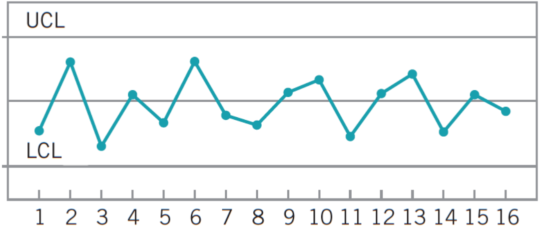
- Medical laboratories typically test two EQA control samples at the same time.
- This will result two z-scores, which we can be studied as two univariate scores (i.e. check their values against the green/orange/red) zone.
- That looks like to be sufficient. Right?
- Well not quite. Studying the z-score as pairs (i.e. bivariate analysis) can provide information that cannot be reached by their univariate evaluation. Let's look in an example:

Why do we need bivariate analysis?

Z-scores of **measurement 1**

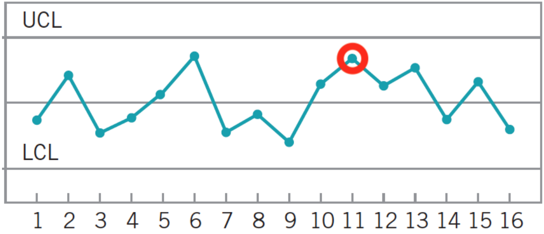


Z-scores of **measurement 2**

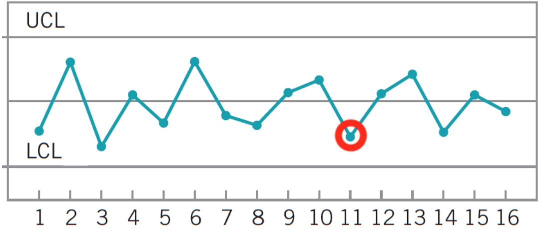


Why do we need bivariate analysis?

Z-scores of **measurement 1** (focus on case 11)

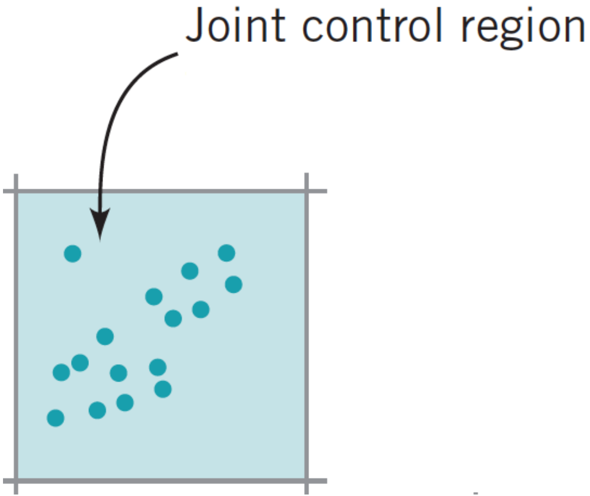


Z-scores of **measurement 2** (focus on case 11)



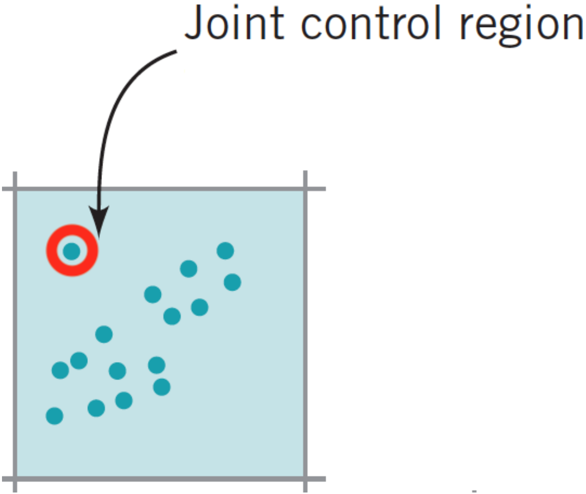
Why do we need bivariate analysis?

Let's visualize the bivariate plot of these data and lets simply combine the univariate control limits to form the square:



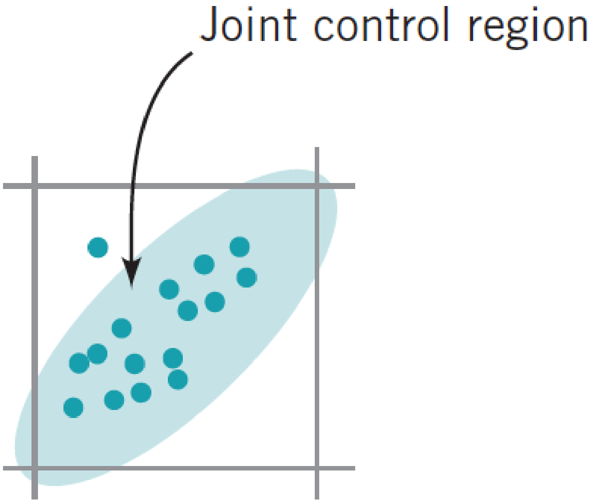
Why do we need bivariate analysis?

Using the square control region we have that case 11 appears as IC. Do you really believe this?



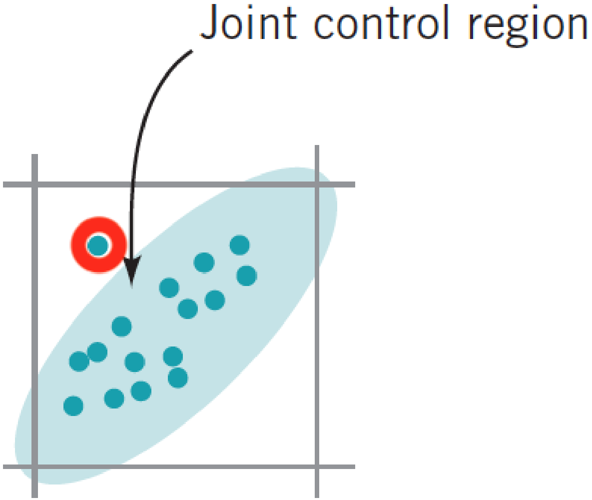
Why do we need bivariate analysis?

Taking into account though that the data are **correlated**, the proper joint statistical control region will form an ellipse:



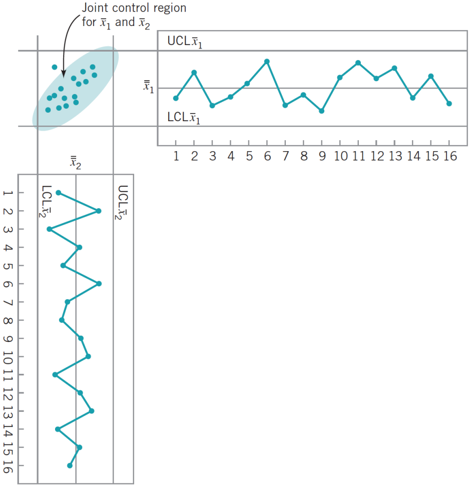
Why do we need bivariate analysis?

Thus using the proper bivariate analysis we get that case 11 is outside of the control region (i.e. we will get an OOC alarm):



Why do we need bivariate analysis?

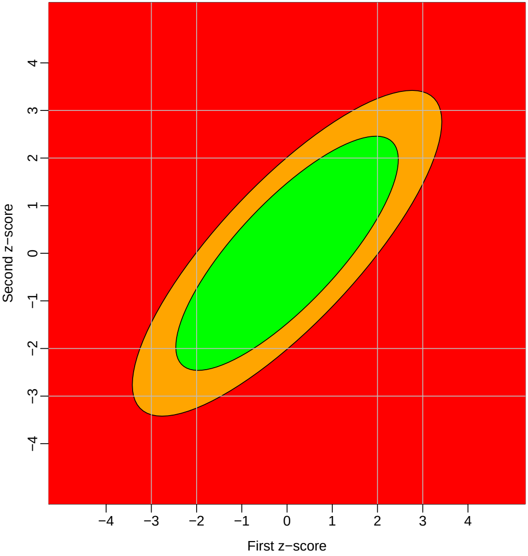
The above example was taken from the classic book *"Introduction to Statistical Quality Control"* by D. Montgomery (a highly recommended textbook regarding the statistical aspects of quality control):



Bivariate z-score evaluation

- When we test at the same time, it is expected to have correlated EQA control results and this correlation will transfer to the respective z-score results within an analytical run.
- The goal of the bivariate analysis is to enrich and **not** replace the existing univariate analysis.
- In the two-dimensional plane (where the z-score pairs are plotted), we will introduce three regions: green/orange/red.
- Assuming bivariate normality, we start by identifying the bivariate outliers (utilizing the Hotelling's T^2 statistic).
- The outlying data are removed and all remaining data provide estimates of the parameters of the underline bivariate distribution.
- Based on these estimates two nested ellipses are drawn with the same center, corresponding to confidence levels 95% and 99.7%.

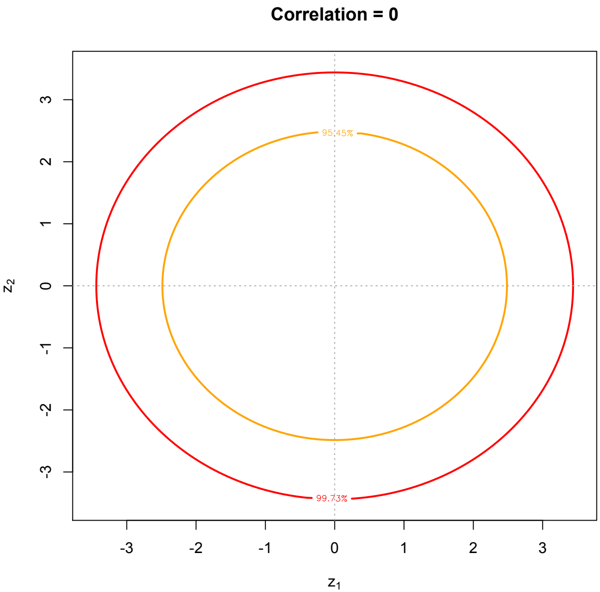
Bivariate z-score evaluation



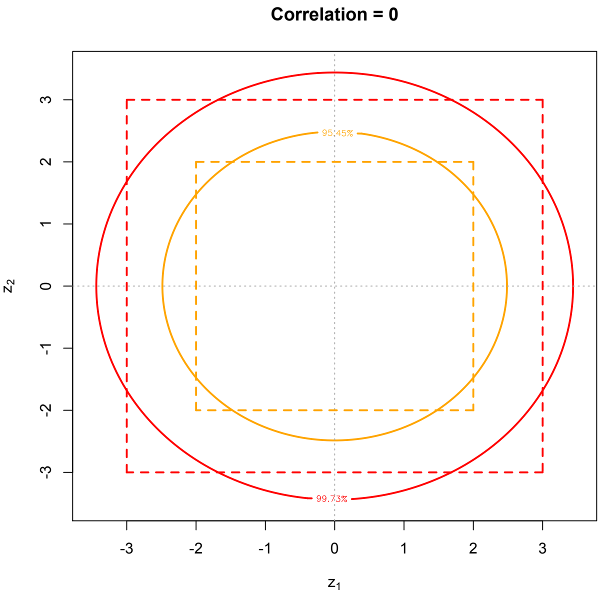
Bivariate z-score evaluation

- These ellipses split the plane in three non-overlapping regions that will indicate the status of the bivariate points:
  - "Acceptable" (green card), if a pair of z-scores lies within the inner ellipse (green region)
  - "Need attention" (orange alarm) if a pair of z-scores lies between the two ellipses (orange region) or
  - "Unsatisfactory" (red alarm) if a pair of z-scores lies outside the outer ellipse (red region).
- The stronger the correlation (i.e. the bigger its absolute value), the more narrow these ellipses will be, while on the other extreme of correlation close to zero (indicating independent z-scores), the ellipses will turn to circles.

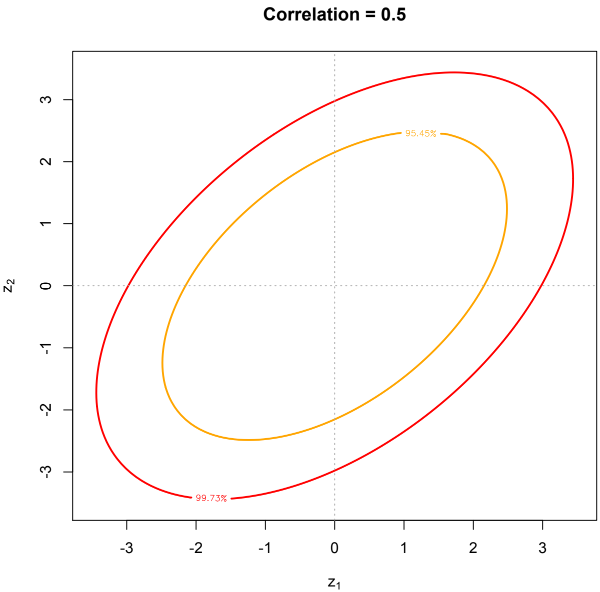
Bivariate z-score evaluation



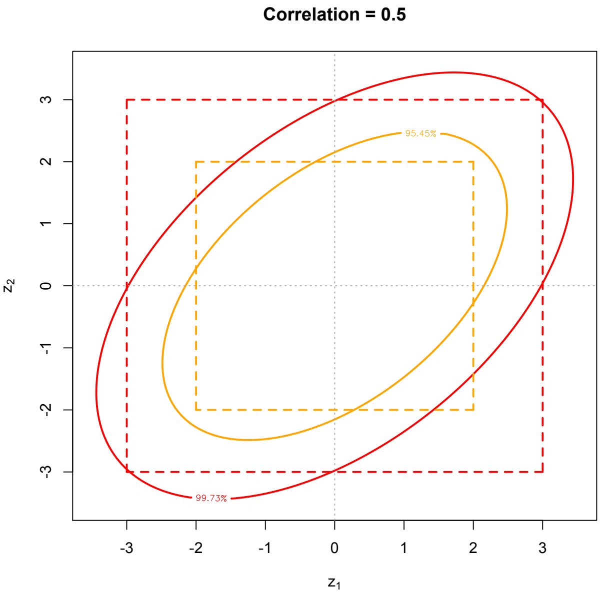
Bivariate z-score evaluation



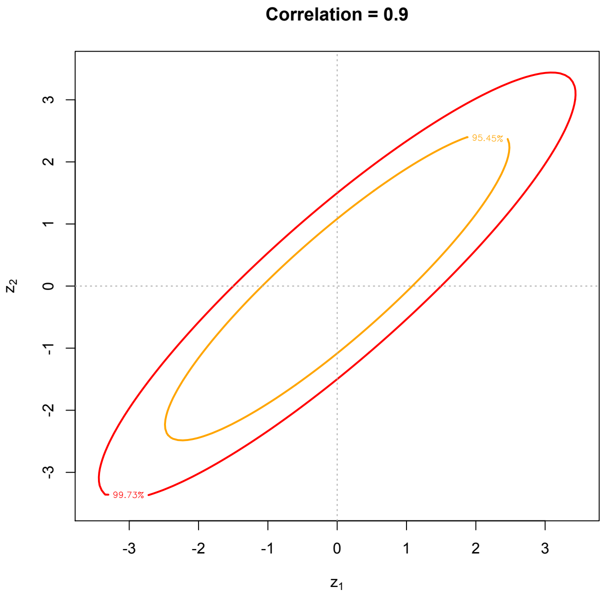
Bivariate z-score evaluation



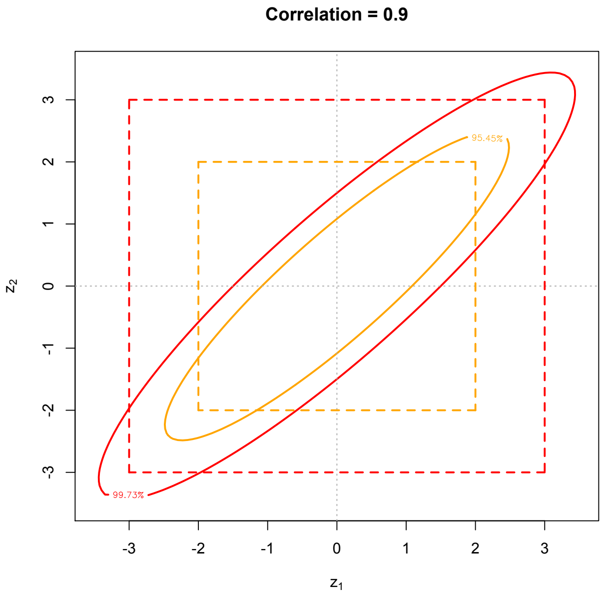
Bivariate z-score evaluation



Bivariate z-score evaluation



Bivariate z-score evaluation



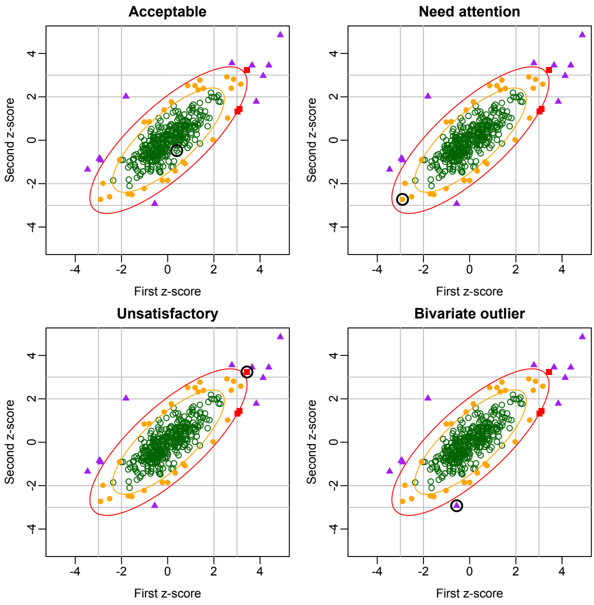
Bivariate z-score evaluation

In the ECAT's survey reports, the data points are plotted using a distinct symbol and color accordingly to the class that they belong:

- Green open circles will indicate the points that are “Acceptable”
- Orange filled discs will indicate the points that “Need attention”
- Red filled squares will indicate the points that are “Unsatisfactory”
- Purple filled triangles will indicate the points that are “Bivariate outliers”

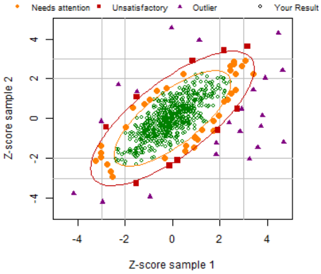
Next we illustrate the various scenarios using a real data from an ECAT survey.

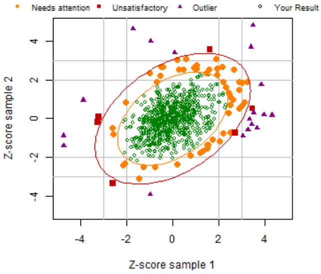
Bivariate z-score evaluation - case study

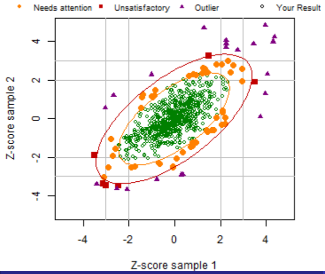


Case Studies

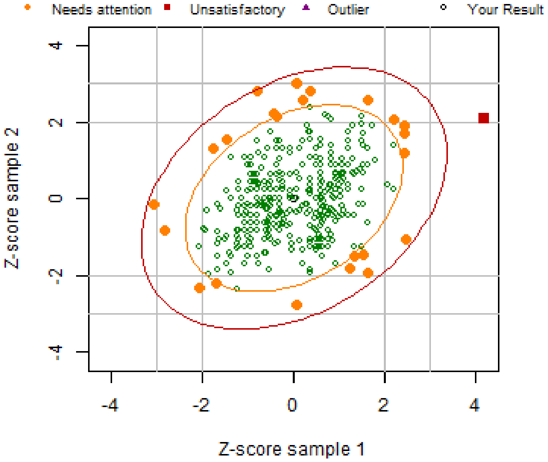
D-Dimer case studies: 2021-D4, 2022-D1, 2022-D2



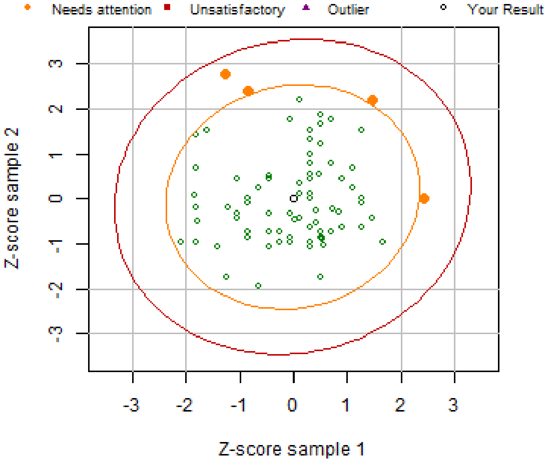




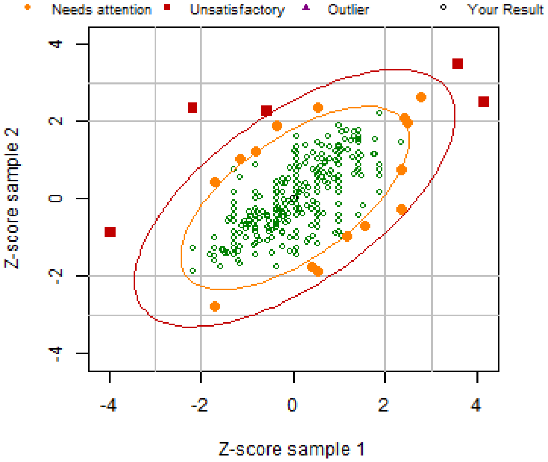
Case study: 2022-M2 Antithrombin



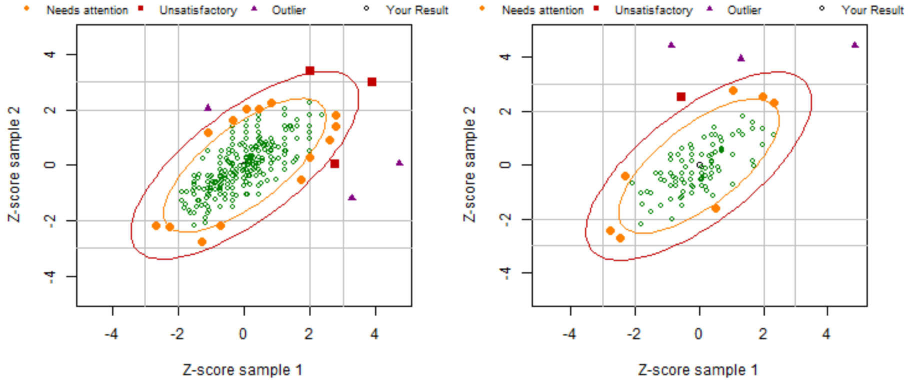
Case study: 2022-M2 Protein C Clot activity



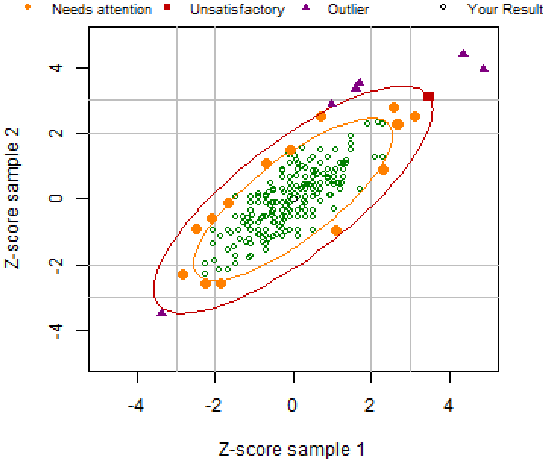
Case study: 2022-M2 Free Protein S antigen



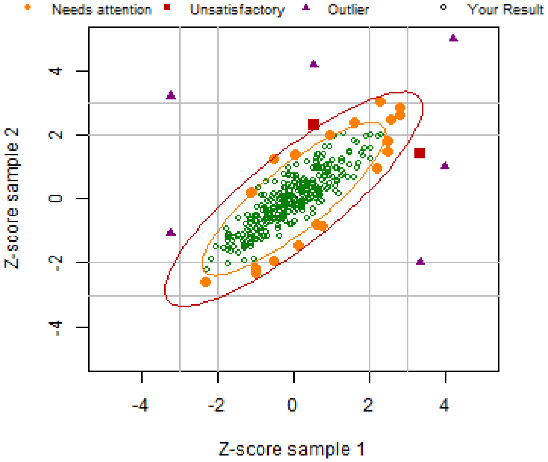
Case study: 2022-M2 Factor VIII clot assay & 2022-M2 Factor VIII chromogenic assay



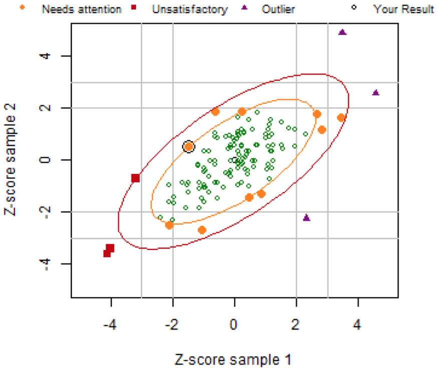
Case study: 2022-M2 Factor XII clot activity

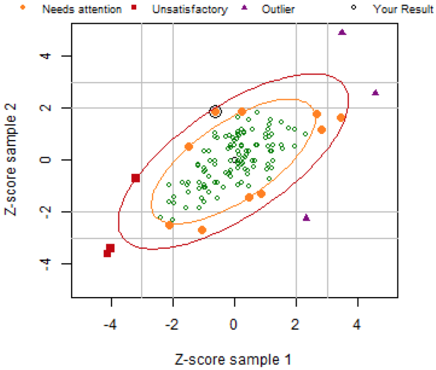


Case study: 2022-M2 Factor VIII Inhibitor



2022-S1 PT percentage (issue of bad pipette)





Conclusions

- The z-scores from the EQA reports provide valuable information at both **univariate** and **bivariate** level.
- EQA scores are snapshots of the quality in your lab, but IQC provides a video of this story. Use state of the art tools to improve your ICQ and EQA results will improve!
- Statistical Process Control/Monitoring provides several tools that can be very helpful not only for identifying problems in the ICQ/EQA analysis, but also providing feedback, useful to the root cause analysis.

Acknowledgments

The authors would like to thank:

- **Dr Frédéric Sobas** and all **Lyon Hemostasis team** of Hospices Civils de Lyon in France, for their motivation in using newly developed state of the art methods in quality monitoring.

The End

Thank you!