

Evaluation of sample processing protocols for the analysis of pumpkin leaf metabolites



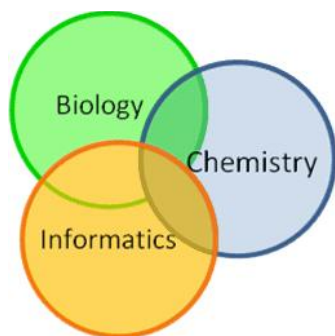
Goals: Compare different extraction and drying protocols to identify the “optimal” sample processing approach

Topics:

1. Data quality overview
2. Statistical comparisons
3. Power analysis

Statistics





Data Quality Overview

	Extraction	Treatment
100% MeOH	: 6	fresh frozen:18
ACN:IPA:H2O (3:3:2)	: 6	lyophilized : 6
MeOH:CHCl3:H2O (5:2:2)	:12	

	Extraction_Treatment
100% MeOH _ fresh frozen	:6
ACN:IPA:H2O (3:3:2) _ fresh frozen	:6
MeOH:CHCl3:H2O (5:2:2) _ fresh frozen	:6
MeOH:CHCl3:H2O (5:2:2) _ lyophilized	:6

Goal: Calculate and visualize the summary statistics for each metabolite/treatment (Use DATA: **Pumpkin data 1.csv**)

Calculate:

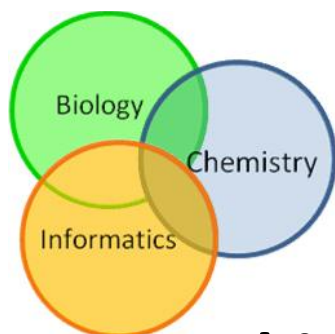
1. Mean and standard deviation (sd)
2. The percent relative standard deviation, %RSD, $(sd/mean)*100$

Visualize:

1. The relationship between mean vs. sd, mean and %RSD
2. Compare mean metabolite values for all treatments

Exercises:

1. Describe the relationship between analyte mean and sd, mean and %RSD?
2. Describe what constitutes an “optimal” method?
3. Which extraction/treatment should be chosen to process further samples?



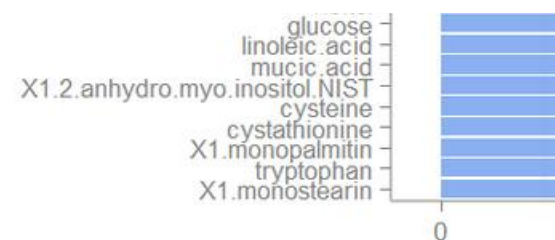
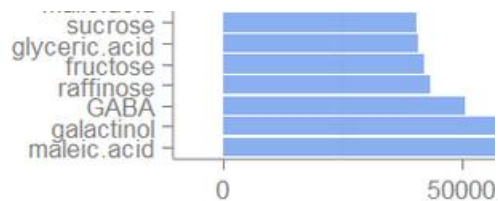
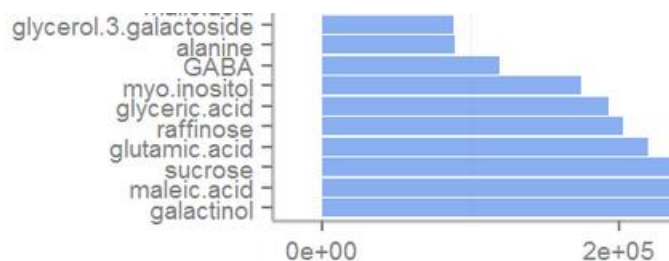
Question:

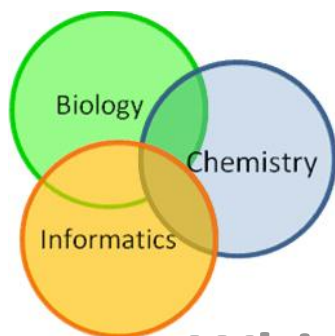


Which metabolite has the largest mean, sd and RSD for the 100% MeOH_fresh frozen extraction method?

	100% MeOH _ fresh frozen mean +/- std dev
zymosterol	78.7 +/- 19
xylose	4280 +/- 340
xylonic.acid	937 +/- 250
valine	16400 +/- 1400
urea	1390 +/- 460
tyrosine	3760 +/- 450
tyramine	7290 +/- 910

Statistics



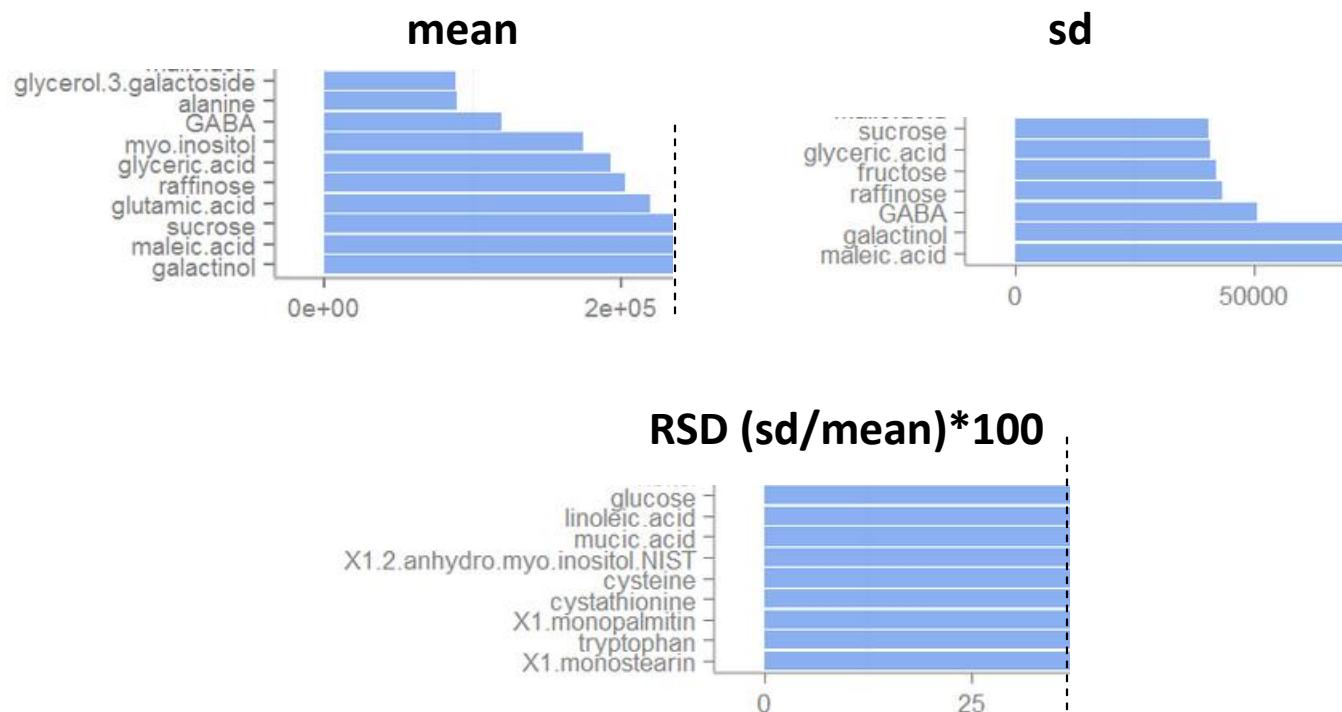


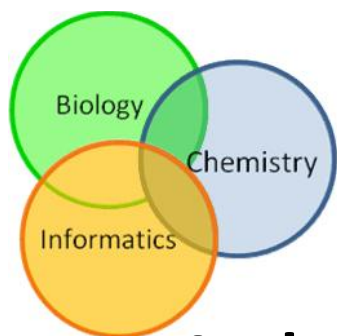
Answer:



Which metabolite has the largest mean, sd and RSD for the 100% MeOH_fresh frozen extraction method?

Statistics





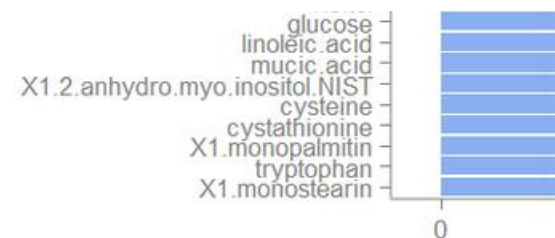
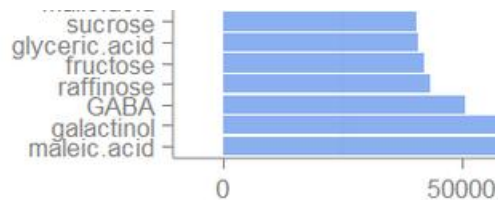
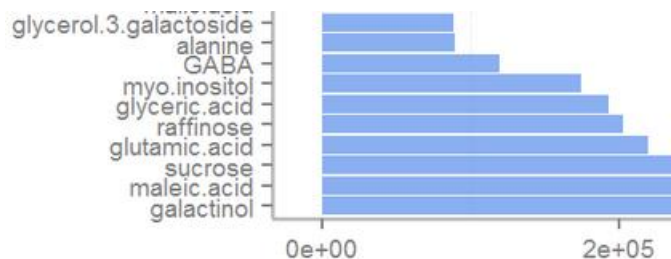
Question:

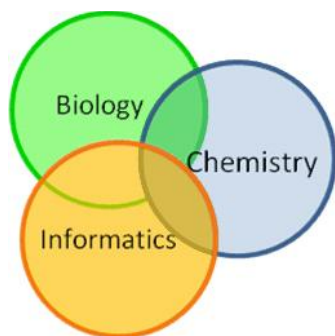


Is there a relationship between mean and sd, mean and RSD?

	100% MeOH _ fresh frozen mean +/- std dev
zymosterol	78.7 +/- 19
xylose	4280 +/- 340
xylonic.acid	937 +/- 250
valine	16400 +/- 1400
urea	1390 +/- 460
tyrosine	3760 +/- 450
tyramine	7290 +/- 910

Statistics



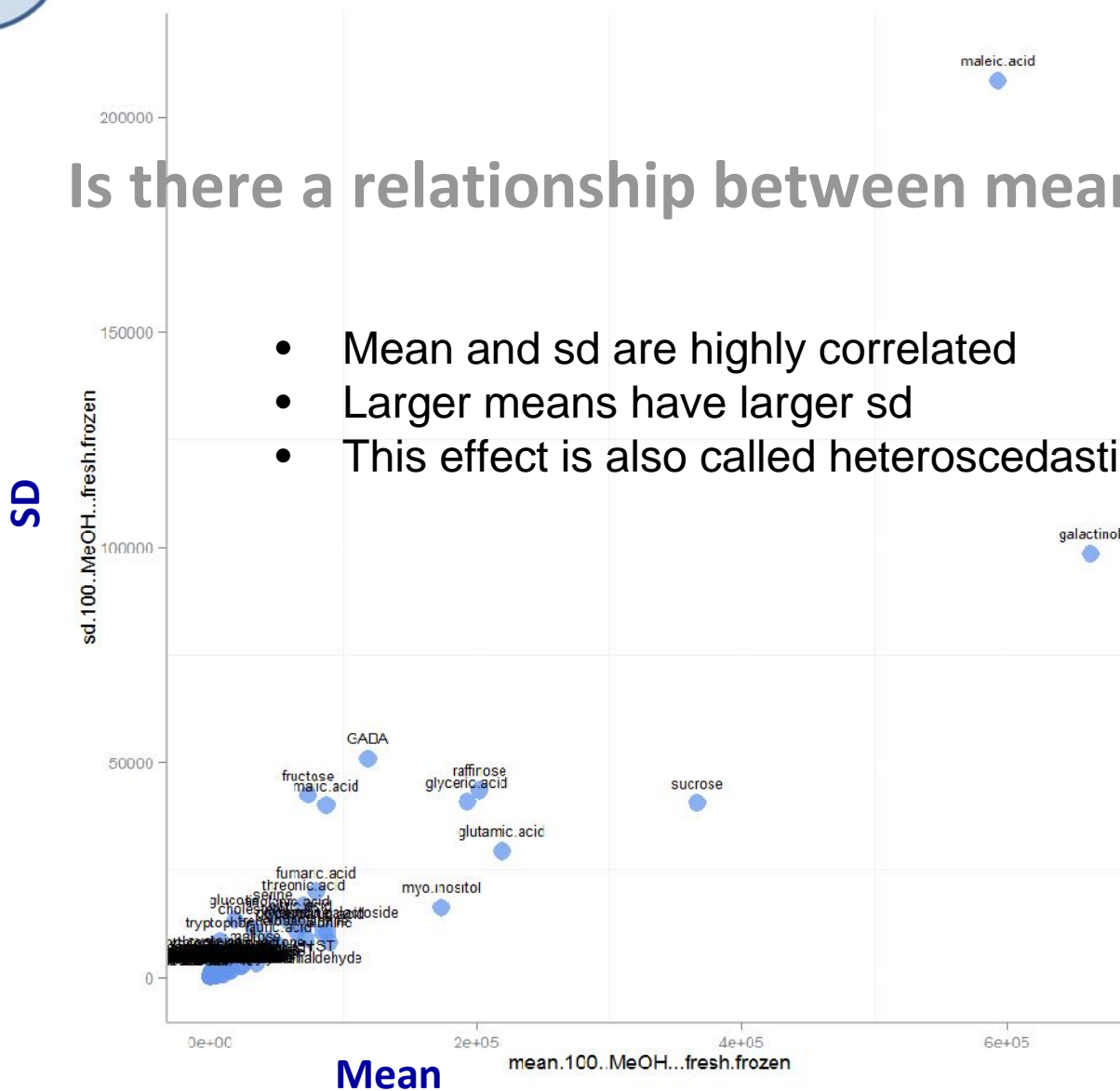


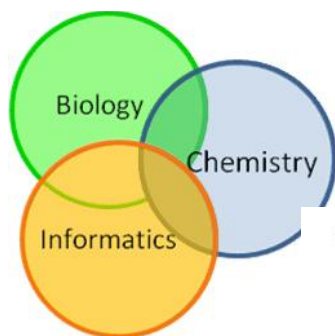
Answer:



Is there a relationship between mean and sd?

- Mean and sd are highly correlated
- Larger means have larger sd
- This effect is also called heteroscedasticity





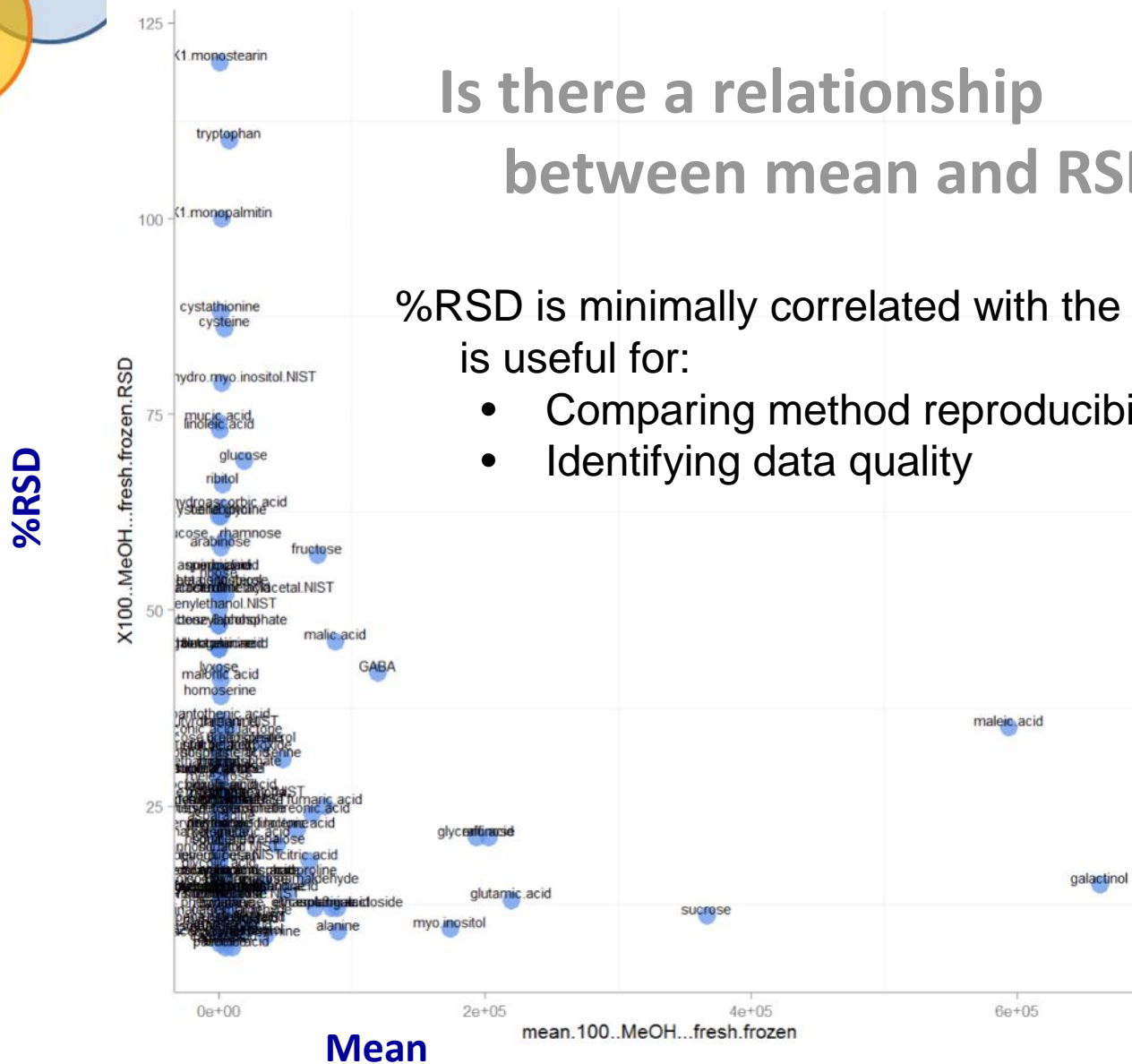
Answer



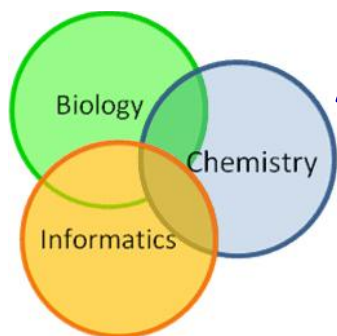
Is there a relationship
between mean and RSD?

%RSD is minimally correlated with the mean, and
is useful for:

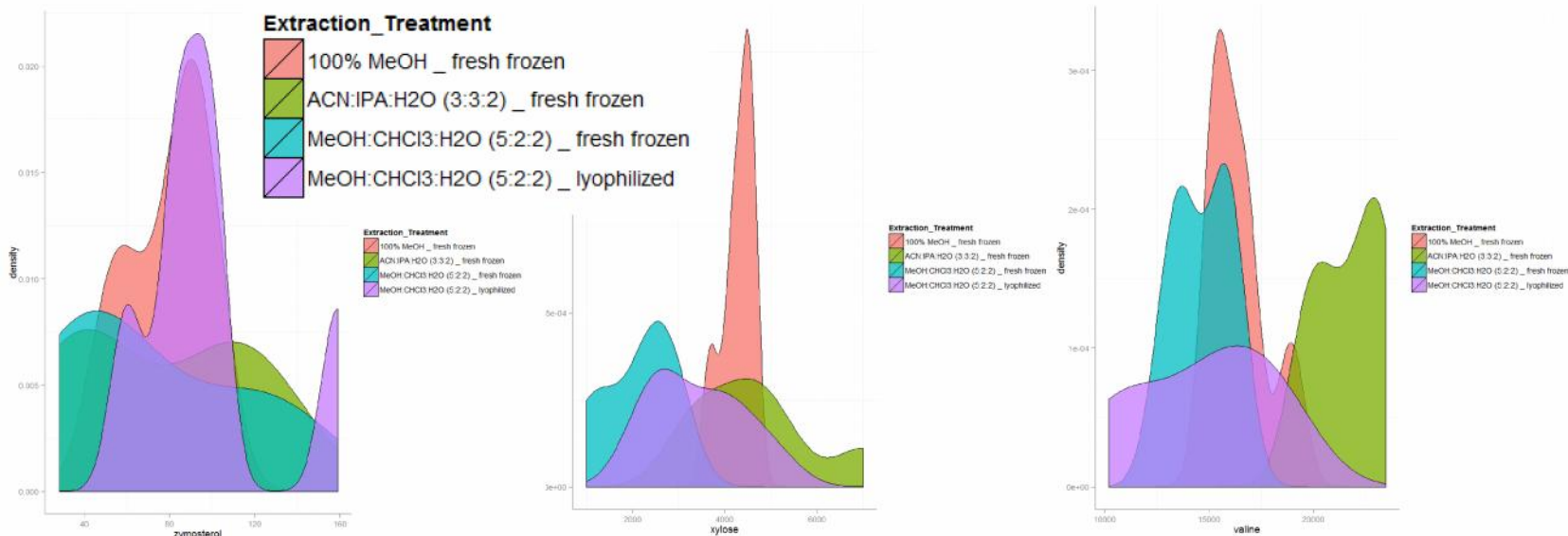
- Comparing method reproducibility
- Identifying data quality



Statistics



**Analytical variance can be used to
select optimal workup methods
and flag problem analytes.**

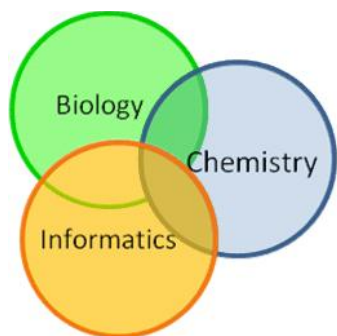


Statistics

zymosterol
xylose
xylonic.acid
valine
urea
tyrosine
tyramine

100% MeOH _ fresh frozen mean +/- std dev

zymosterol	78.7 +/- 19
xylose	4280 +/- 340
xylonic.acid	937 +/- 250
valine	16400 +/- 1400
urea	1390 +/- 460
tyrosine	3760 +/- 450
tyramine	7290 +/- 910



Data quality



Below
LOQ
(sensitivity)

%RSD

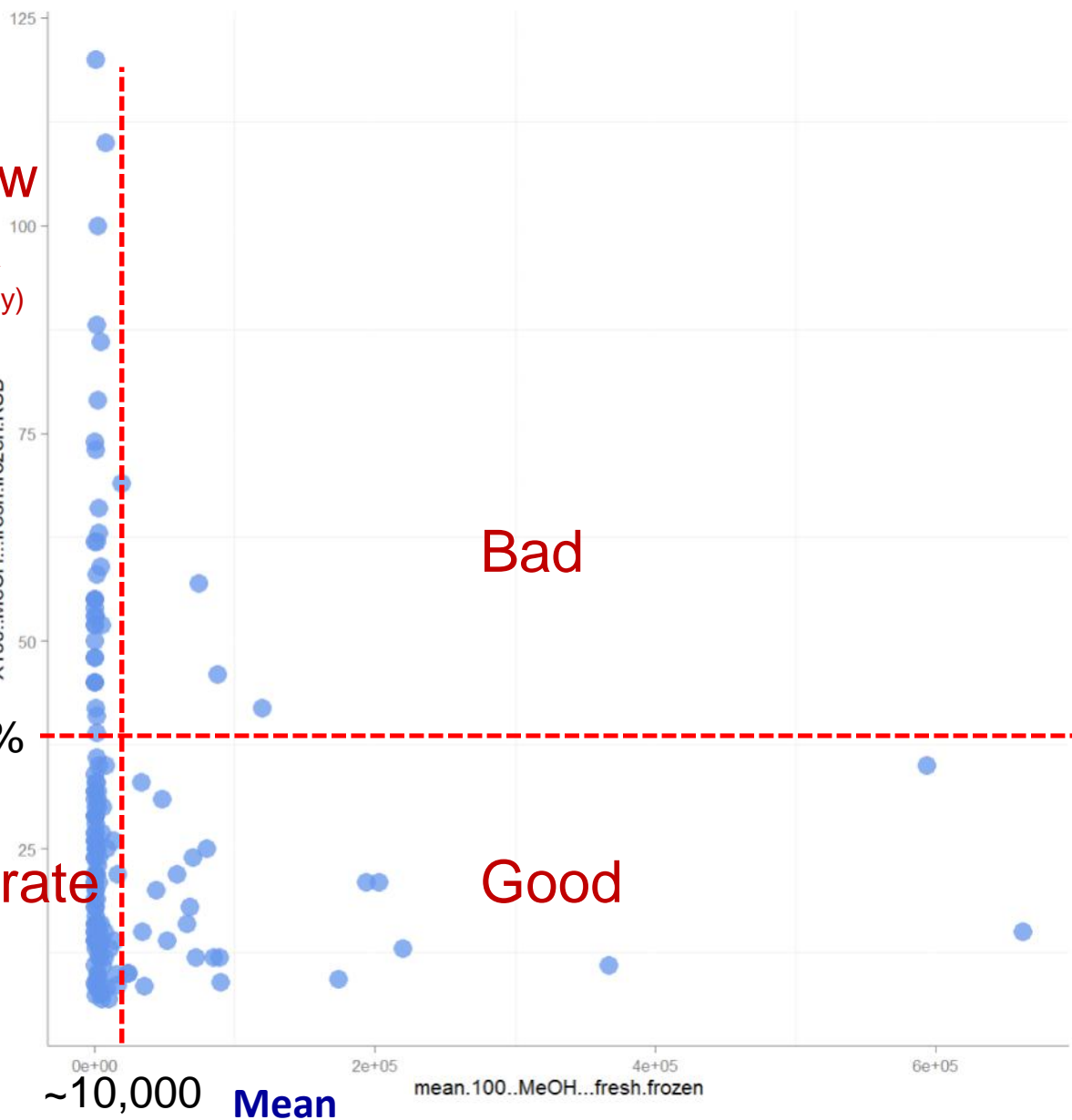
~40%

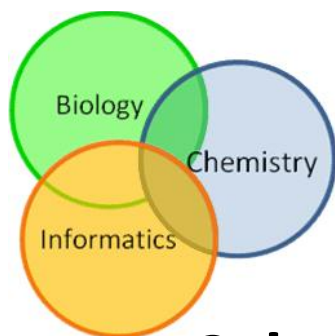
Moderate

Bad

Good

Statistics



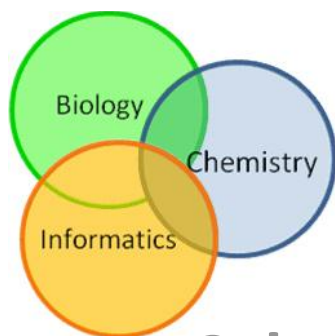


Question:



Calculate the mean RSD across all analytes for each workup method

- %RSD (also called the coefficient of variation or CV) is the sd (variation) scaled by the mean (magnitude).
- Removes the relationship between variation and magnitude
- Provides a single value which can be used to compare the variation of a measurement among different treatments/samples



Answer:

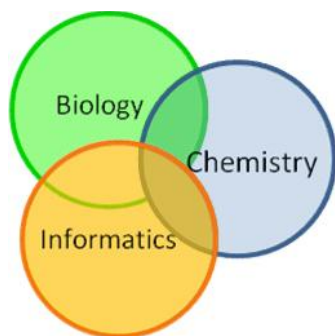


Calculate the mean RSD across all analytes for each workup method

RSD mean +/- sd and median (min, max)

	id	mean +/-	std dev
RSD.100..MeOH..fresh.frozen		29.5 +/-	22
RSD.ACN.IPA.H2O..3.3.2..fresh.frozen		31.1 +/-	33
RSD.MeOH.CHCl3.H2O..5.2.2..fresh.frozen		29.6 +/-	19
RSD.MeOH.CHCl3.H2O..5.2.2..lyophilized		35.3 +/-	28
	id	median	(min, max)
RSD.100..MeOH..fresh.frozen		24	(6.9, 120)
RSD.ACN.IPA.H2O..3.3.2..fresh.frozen		20	(5.6, 200)
RSD.MeOH.CHCl3.H2O..5.2.2..fresh.frozen		24	(5.5, 120)
RSD.MeOH.CHCl3.H2O..5.2.2..lyophilized		28.5	(9.8, 180)

Statistics



Selecting the “optimal” method



Optimal can be:

1. Lowest average %RSD for all measurements
2. Lowest %RSD for measurements of interest
3. Largest number of metabolites passing %RSD cutoff
4. Lowest average %RSD for all measurements passing %RSD cutoff

Using strategy #4 for metabolites %RSD 40

	Count	%RSD (mean \pm sd)
RSD.100..MeOH.__.fresh.frozen	113	18.8 \pm 8.4
RSD.ACN.IPA.H2O..3.3.2.__.fresh.frozen	120	17.2 \pm 6.8
RSD.MeOH.CHCl3.H2O..5.2.2.__.fresh.frozen	116	19.9 \pm 8.6
RSD.MeOH.CHCl3.H2O..5.2.2.__.lyophilized	114	23 \pm 7.6



Method #2 (ACN/IPA/water 3:3:2) looks optimal...



%RSD



Log Mean



Analytes with high signal and high %RSD should be further interrogated for explanations of low reproducibility