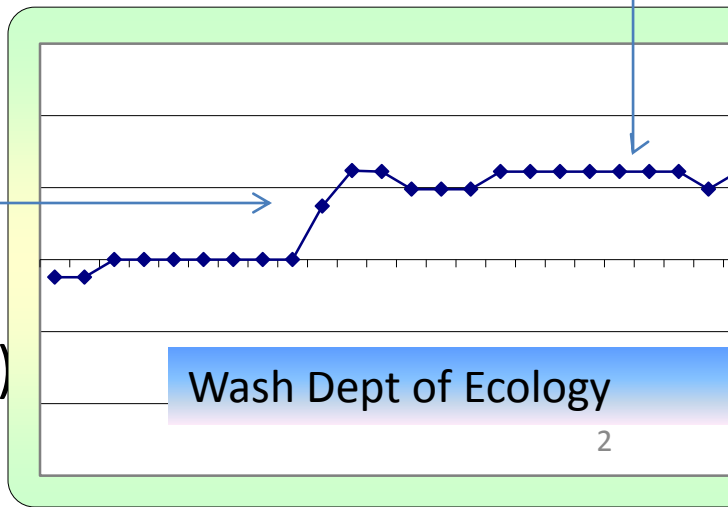
Appendix A: Calculations for Data Quality Assessment (sec. 4-5) aka

What Is Reality?

- **1-pt QC check statistics**
- **Precision calcs**
- **Bias calcs**

Stats are designed to show us how far from the TRUTH we might be.

- ***Measurement Error***
  - Presented as a fraction of the "truth" (e.g., 10% off)
- ***Precision***
  - Random error
  - "wiggle" inherent in system
  - Estimated by (1) repeated measurements of "known," and/or (2) side-by-side measurements of the same thing
  - Some imprecision is unavoidable
- ***Bias***
  - Systematic error
  - "jump" consistently high or low
  - bias can be eliminated (in theory)

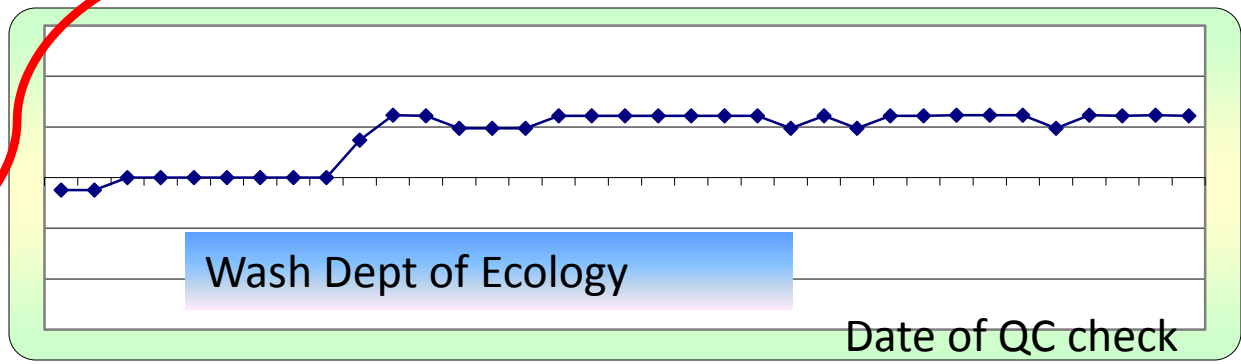


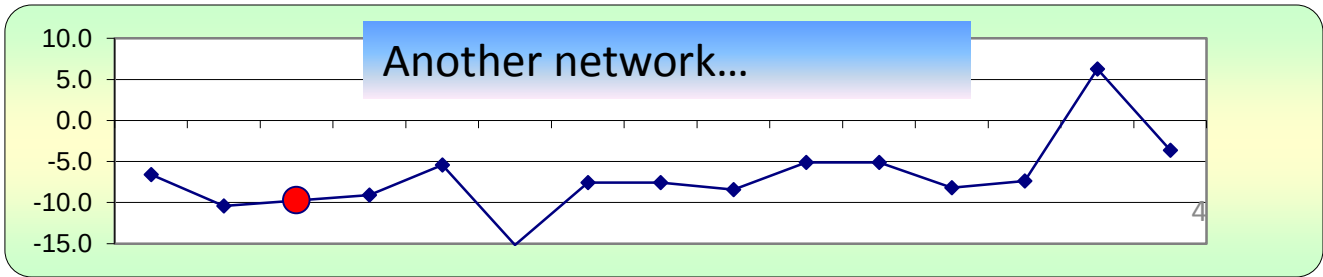
1-pt QC 0^3 check data, in AQS:

Meas Val (Y)	Audit Val (X)
98	90
87	90
79	90
79	90
81	90
80	90
82	90
94	90
96	90
97	90
97	90
98	90
80	90

d-sub-i = d_i = diff/known

- Routine QC checks used to estimate BOTH
- Both come from **d-sub-i**
  - Bias
  - Precision
- sometimes it's obvious
- Sometimes it's not:

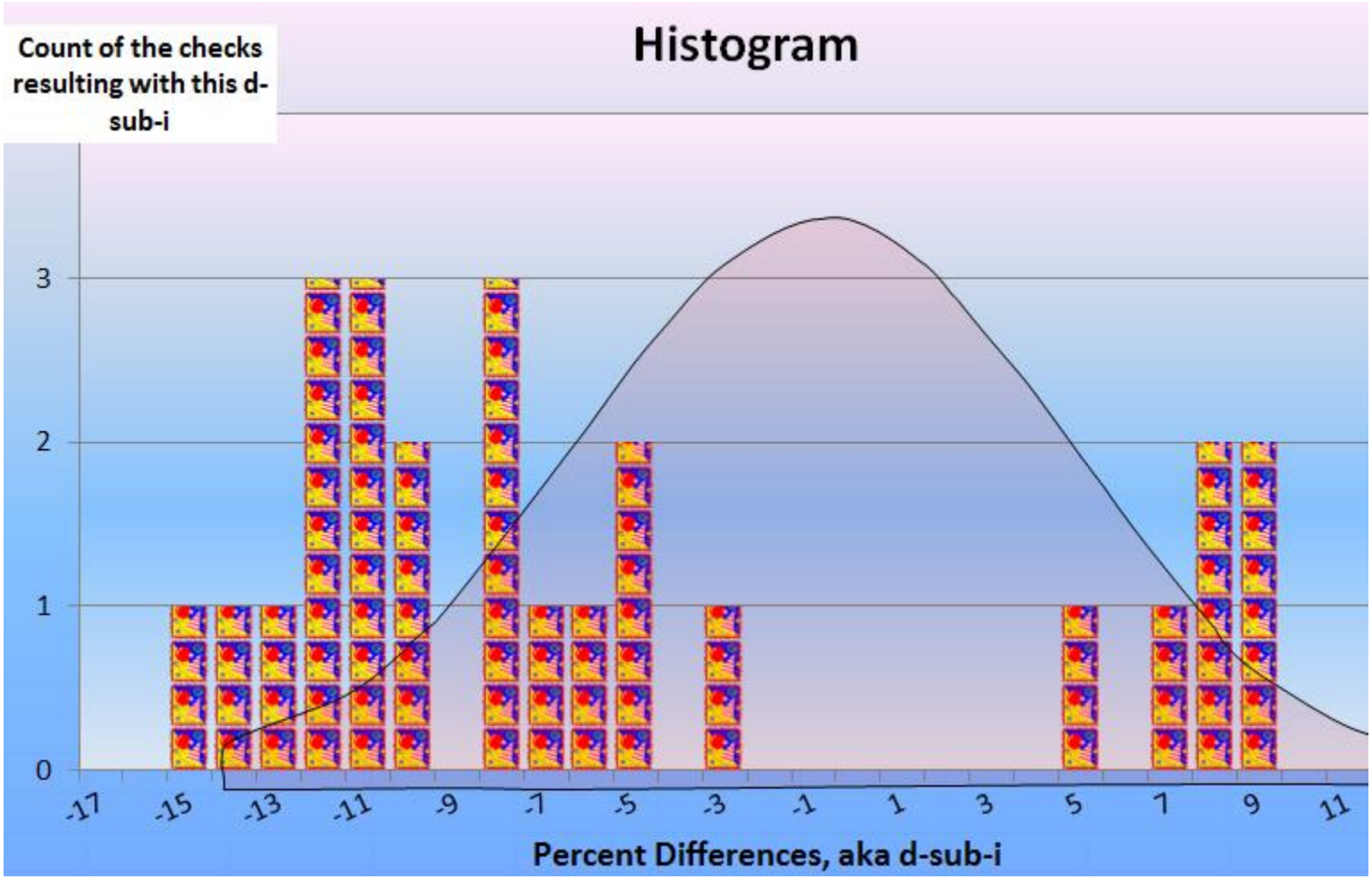




Meas Val (Y)	Audit Val (X)	d-sub-i
98	90	9
87	90	-3
79	90	-12
79	90	-12
81	90	-10
80	90	-11
82	90	-9
94	90	4
96	90	7
97	90	8
97	90	8
98	90	9
80	90	-11
80	90	-11
84	90	-7

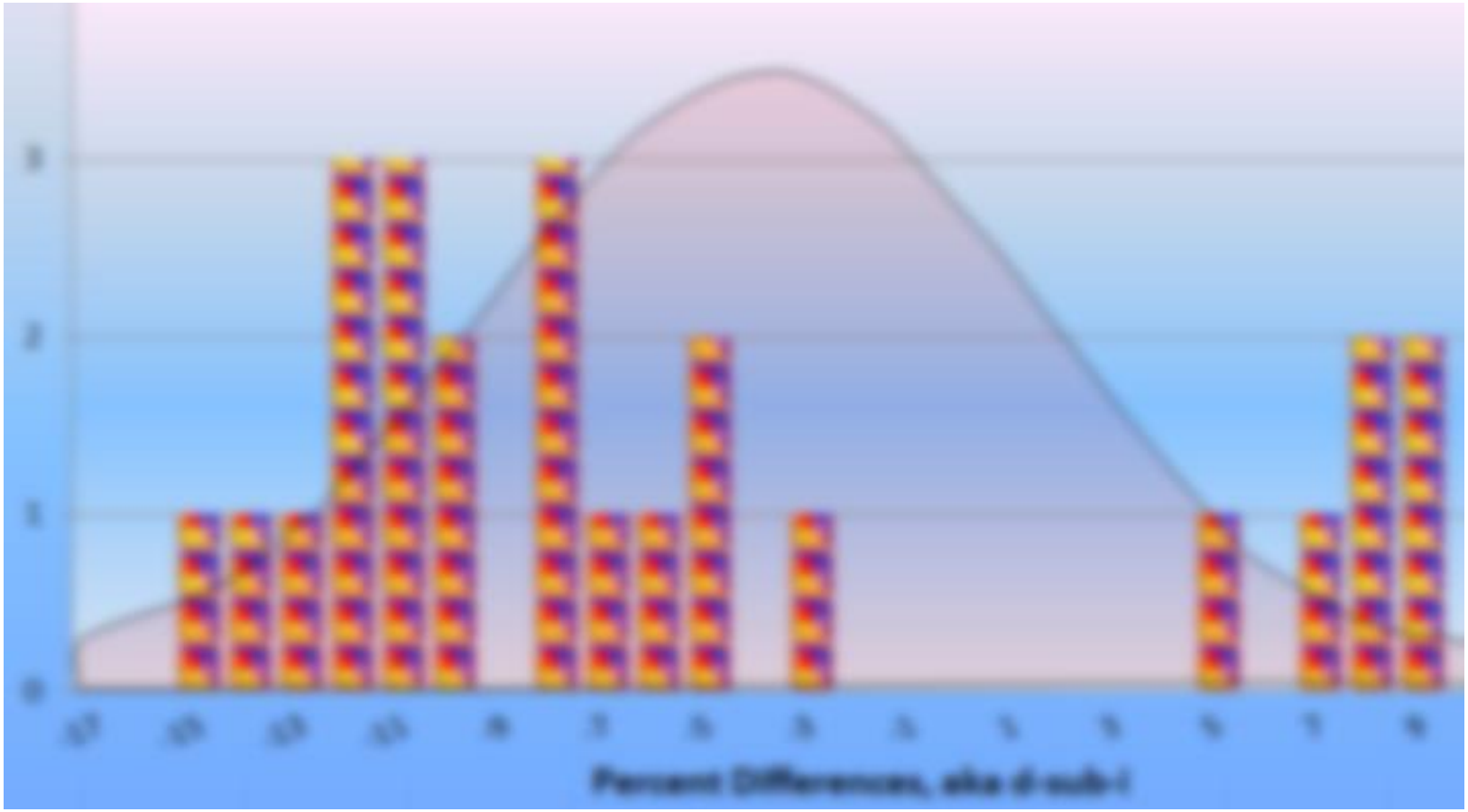
- **d-sub-i values represent:**
  - All of the measurements' error during that day, week, month, quarter
    - The QC checks are supposed to be "randomized" so that they are a sample, or subset, of the whole universe of possible QC checks (the population), and then represent the population of QC checks you could do at any time
  - As a proportion of the "truth," so "truth" is always on the bottom (diff/known; so error is quantified as a fraction of the truth so we can imagine it, e.g., 10%)
  - "error" = distance from truth at that moment

d (Eqn. 1)
-16
-14
-13
-12
-12
-12
-11
-11
-11
-10
-10
-9
-9
-9
-8
-7
-6
-6
-3
4



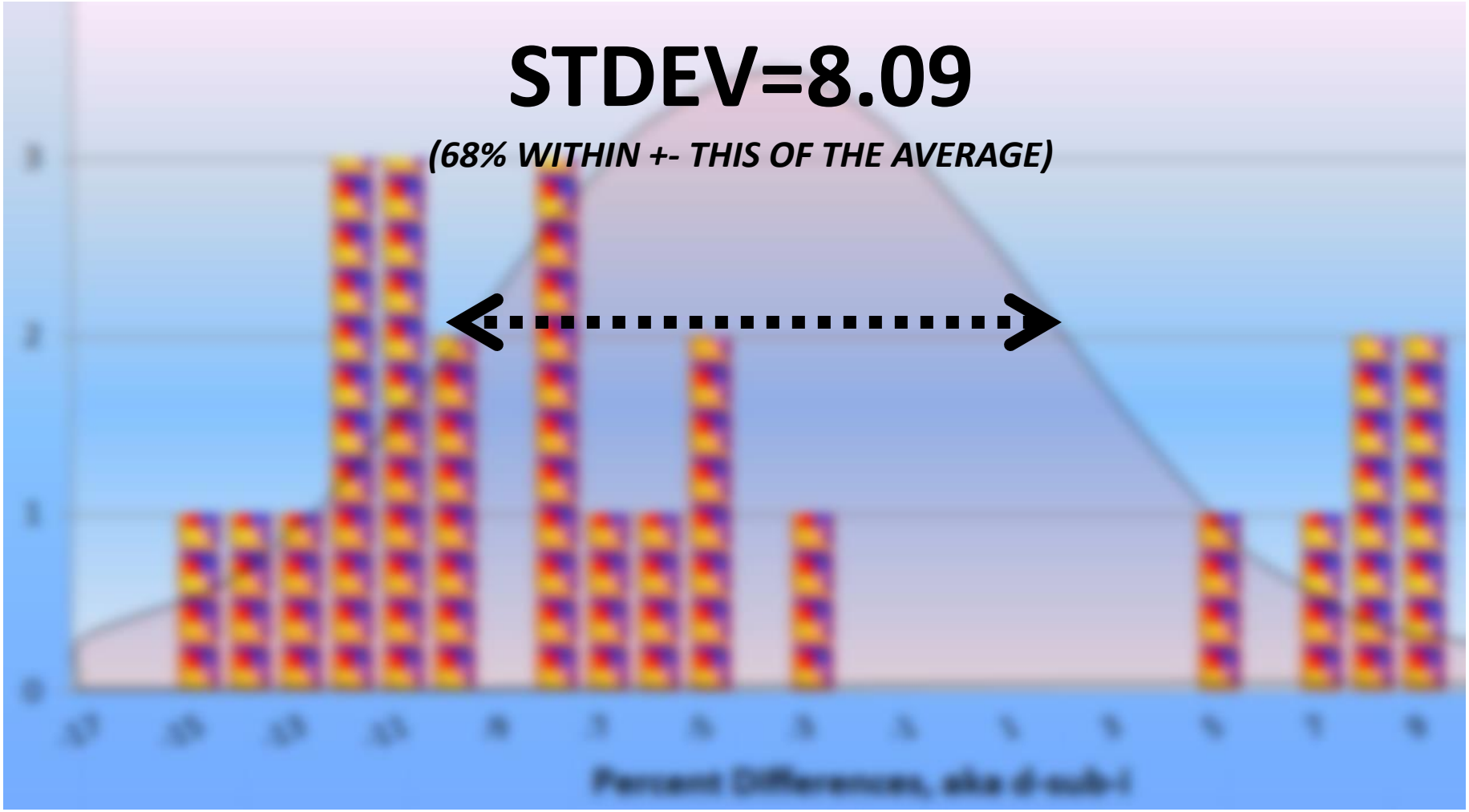
How can we apply these results to get bias and precision for ALL our measurements of ozone with this analyzer during this time period?

We assume that these results, and their distribution, is representative of all the QC checks we could have done:

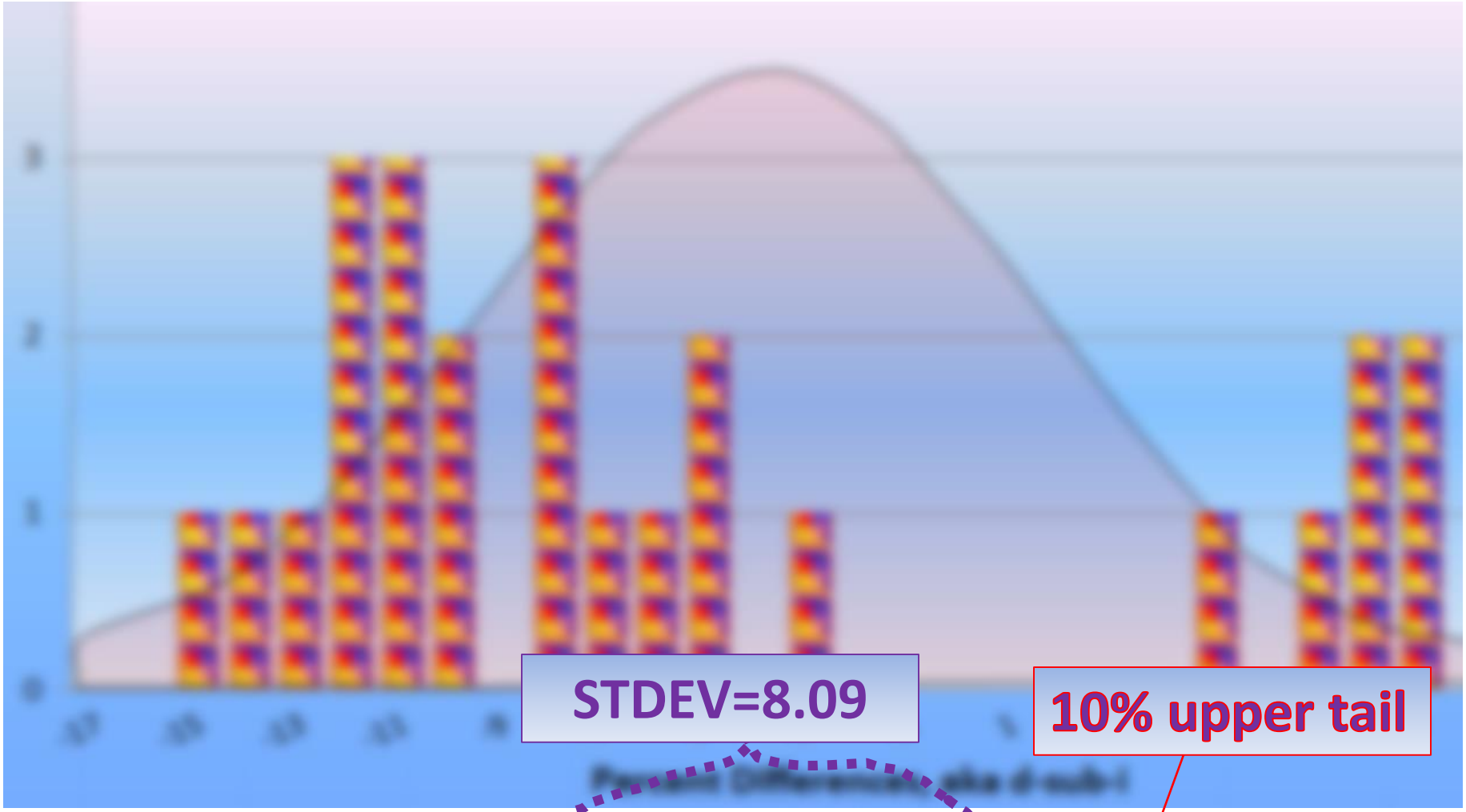


There's a reason no x-axis units

The goal for acceptable measurement uncertainty is defined for O3 precision as an upper **90 percent confidence limit for the coefficient variation (CV) of 7%**



- But we do not care about the low-imprecision tail
- Only care about the extreme tail of high imprecision
- Want to be able to say "**90% confident that your precision is less than** this value"

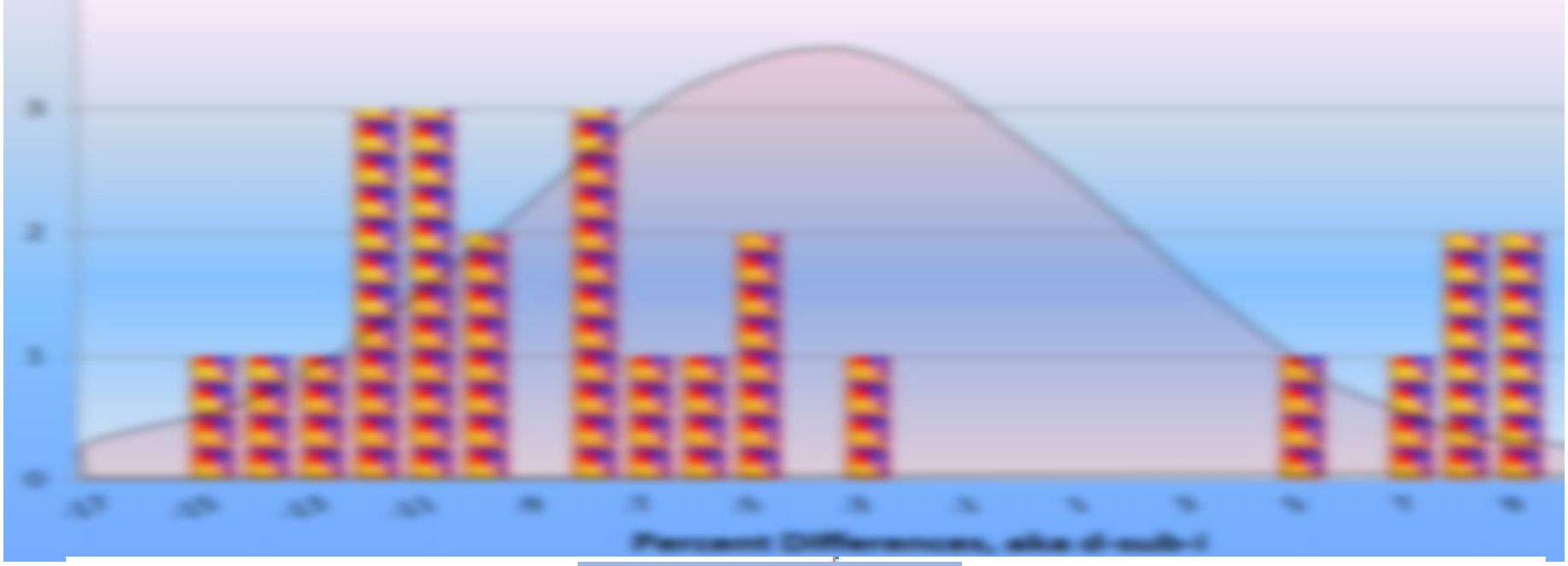


STDEV=8.09

10% upper tail

CFR eq'n 2:

$$\text{Precision\_Estimate} = \sqrt{\frac{n \bullet \sum_{i=1}^{n} d_i^2 - (\sum_{i=1}^{n} d_i)^2}{n(n-1)}} \bullet \sqrt{\frac{n-1}{\chi^2_{0.1,n-1}}}$$



STDEV=8.09

10% upper tail

$$Precision\_Estimate = \sqrt{\frac{n \bullet \sum_{i-1}^{n} d_i^2 - (\sum_{i-1}^{n} d_i)^2}{n(n-1)}} \bullet \sqrt{\frac{n-1}{\chi^2_{0.1,n-1}}}$$

chi-sqrd(90%) = CHIINV(0.9,n) = 15.65

then 8.09 x SQRT(n-1/15.65) = 10.01 %

Use the DASC Tool to Understand Your QC Checks and Audit Results (like EPA does)

- Calculations of measurement uncertainty are carried out by EPA, *and* PQAOs should report the data for all measurement quality checks
- YOU do these calculations and charts easily, and save yourself time, money, and embarrassment

We will review each in both the DASC tool and the AMP256 report

First, what is the DASC tool?

- DASC tool was produced specifically for us to calculate the data assessment statistics in CFR in AMTIC Quality Indicator Assessment Reports (AMP256)
- **http://www.epa.gov/ttn/amtic/qareport.html**
- Easy way to explain and calculate data assessment statistics in CFR
- Excel spreadsheet
- Matches AMP256 (by site)
- Each equation is numbered and matches the numbers in CFR

DASC Tool:

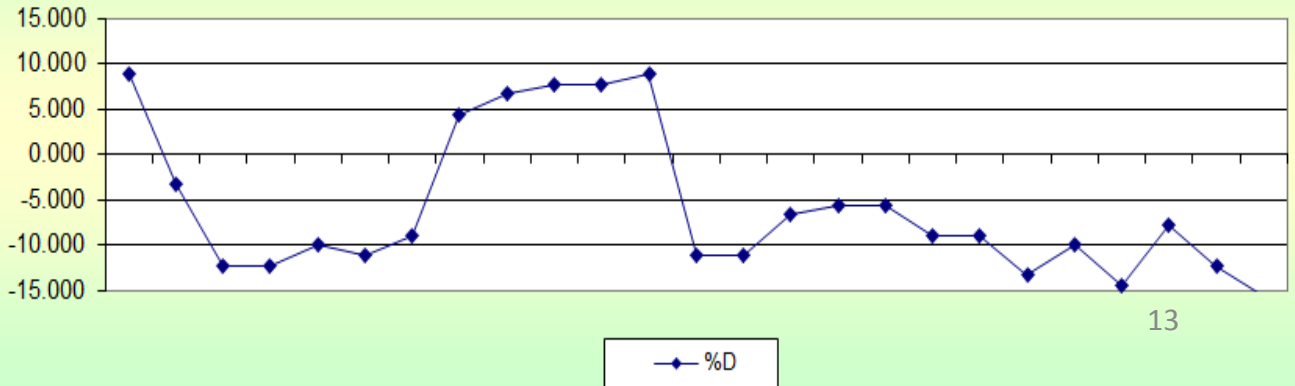
O_3 Assessments

Site ID: {Enter Site ID}		Pollutant type: O_3					CV_{ub} (%)			Bias (%)	
Meas Val (Y)	Audit Val (X)	d (Eqn. 1)	25th Percentile	d^2	|d|	$\|d\|^2$					
0.098	0.09	8.889	-11.111	79.012	8.889	79.012					
0.087	0.09	-3.333	**75th Percentile**	11.111	3.333	11.111	n	S_d	S_{d2}	$\Sigma\|d\|$	"AB" (Eqn 4)
0.079	0.09	-12.222	-3.333	149.383	12.222	149.383	25	8.089	59.010	233.333	9.333
0.079	0.09	-12.222		149.383	12.222	149.383	n-1	Σd	Σd^2	$\Sigma\|d\|^2$	"AS" (Eqn 5)
0.081	0.09	-10.000		100.000	10.000	100.000	24	-144.444	2404.938	2404.938	3.077
0.08	0.09	-11.111		123.457	11.111	123.457					
0.082	0.09	-8.889		79.012	8.889	79.012				Bias (%) (Eqn 3)	Both Signs Positive
0.094	0.09	4.444		19.753	4.444	19.753				10.39	FALSE
0.096	0.09	6.667		44.444	6.667	44.444	CV (%) (Eqn 2)			Signed Bias (%)	Both Signs Negative
0.097	0.09	7.778		60.494	7.778	60.494	10.01			-10.39	TRUE
0.097	0.09	7.778		60.494	7.778	60.494					
0.098	0.09	8.889		79.012	8.889	79.012	Upper Probability Limit			Lower Probability Limit	
0.08	0.09	-11.111		123.457	11.111	123.457	10.08			-21.63	
0.08	0.09	-11.111		123.457	11.111	123.457					
0.084	0.09	-6.667		44.444	6.667	44.444					
0.085	0.09	-5.556		30.864	5.556	30.864					
0.085	0.09	-5.556		30.864	5.556	30.864					
0.082	0.09	-8.889		79.012	8.889	79.012					
0.082	0.09	-8.889									
0.078	0.09	-13.333									
0.081	0.09	-10.000									
0.077	0.09	-14.444									
0.083	0.09	-7.778									
0.079	0.09	-12.222									
0.076	0.09	-15.556									

Return to Main Menu

Print Worksheet

Percent Differences



Precision in DASC = cell i13 = **10.01%**

O_3 Assessments

ant type: O_3					CV_{ub} (%)	Bias
n. 1)	25th Percentile	d^2	|d|	$\|d\|^2$		
8.889	-11.111	79.012	8.889	79.012		
-3.333	**75th Percentile**	11.111	3.333	11.111		
2.222	-3.333	149.383	12.222	149.383		
2.222		149.383	12.222	149.383		
0.000		100.000	10.000	100.000		
1.111		123.457	11.111	123.457		
-8.889		79.012	8.889	79.012		
4.444		19.753	4.444	19.753		
6.667		44.444	6.667	44.444		
7.778		60.494	7.778	60.494		
7.778		60.494	7.778	60.494		
8.889		79.012	8.889	79.012		
1.111		123.457	11.111	123.457		
1.111		123.457	11.111	123.457		

n	S_d	S_{d2}
25	8.089	59.010
n-1	**$\sum d$**	**$\sum d^2$**
24	-144.444	2404.938

CV (%) (Eqn 2)
10.01

Upper Probability Limit	Lowe
10.08	

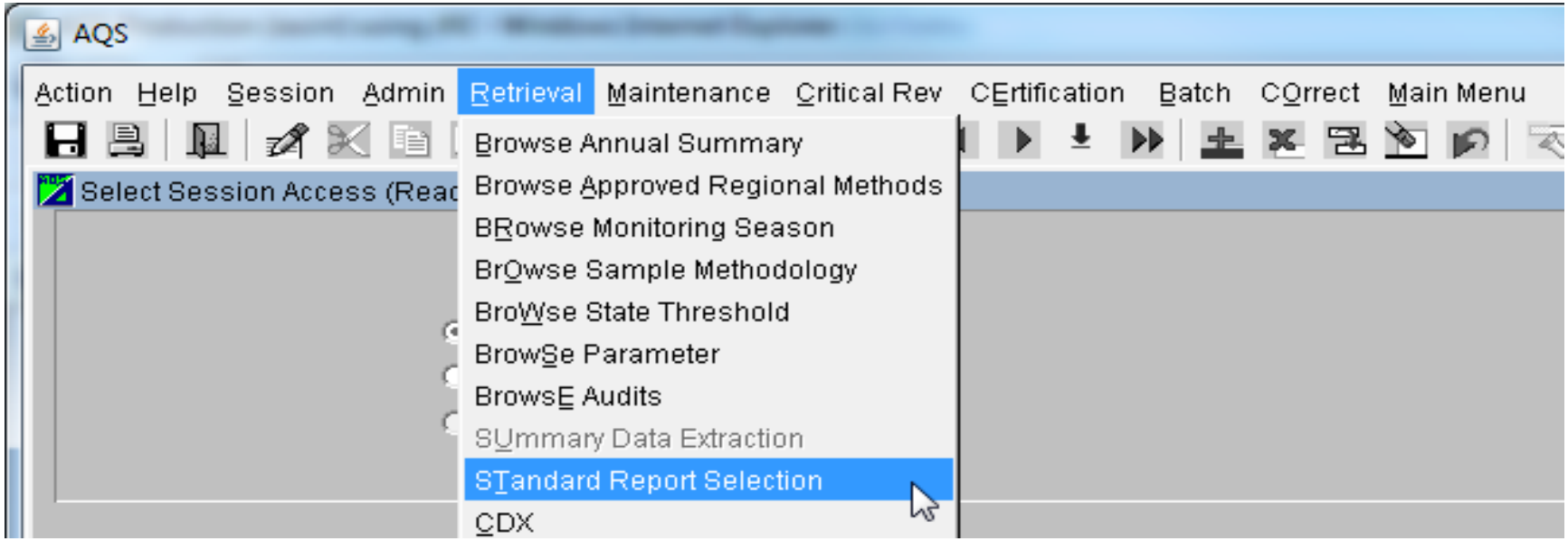
Bias

Signe

-10.3

AMP256-Data Quality Indicators Report

- AQS Standard Report to Compute the Statistics Outlined on 40 CFR Part 58 Appendix A
- Part of the Annual Certification Process to Verify Submission of QA and routine Data to AQS



- **CORRESPONDS to what you can calculate in the DASC spreadsheet, as we will see.**

Does our 10.01% match AMP256?

DATA QUALITY INDICATOR REPORT

One Point Quality Control

PQAO:

Site IDs	POC	MT	Begin Date	End Date	Intervals Required	Valued Intervals	% Complete	CV UB
	1	SP	01-JAN-14	31-DEC-14	12	9	75	10.01
	1	SP	01-JAN-14	31-DEC-14	13	10	77	10.01
MMARY			01-JAN-14	31-DEC-14	25	19	76	10.01
MMARY			01-JAN-14	31-DEC-14	25	19	76	10.01

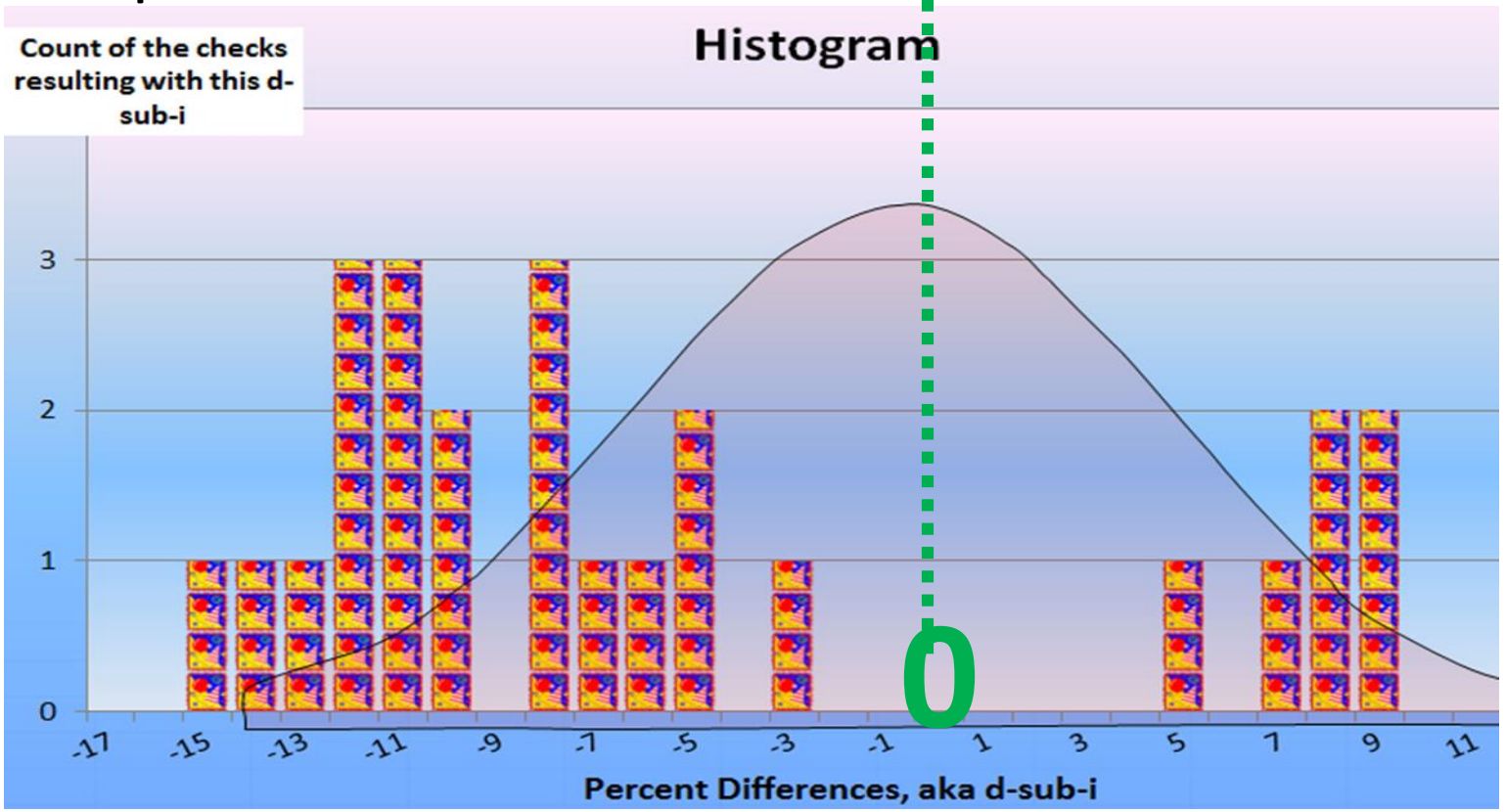
- **90% Confidence Upper Bound of precision is 10.01%**
- **"There is a 90% chance that our precision will not be greater than 10%"**
- **Same as YOU can calculate any time using the DASC**

Summary of precision:

- Calculated from routine QC checks d_i
- Overall upper bound of CV calculated from d_i
- you can be **90% sure that your true precision is less than** this “upper bound of the CV” (eq’n 2)

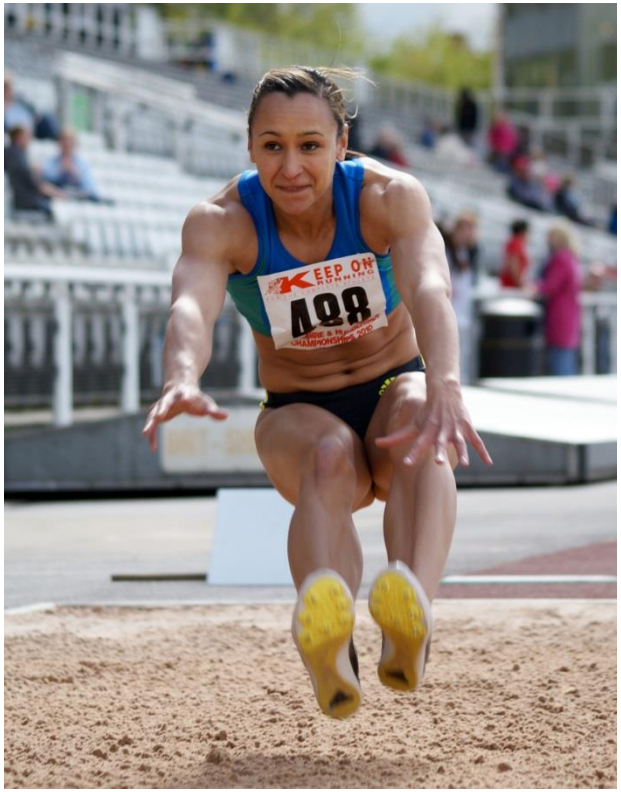
Bias:

- FINALLY look at **where** we are on the x-axis
- (Remember precision only cares about width)
- The goal for acceptable measurement uncertainty for **bias** is an upper 95 percent confidence limit for the absolute bias of 7 percent.



Bias statistics (CFR App A, 4.1.3):

- Remember that bias as well as precision starts from the difference between your instrument's indicated value and the known (audit) value, as
- (meas-known)/known= d_i
- bias (jump) is calculated from d_i
- Bias just *based on* the *AVERAGE* of the d_i with the sign taken into account (if your analyzer is always higher than the known, you have a high (+) bias



Bias in CFR eq'n 3:

$$|bias| = AB + t_{0.95,n-1} \cdot \frac{AS}{\sqrt{n}}$$

AB is the mean of the absolute values of the d_i's = **9.3**

$t_{0.95,n-1}$ is the 95th quantile of a t-distribution
=TINV(2*0.05,n-1) = 1.71

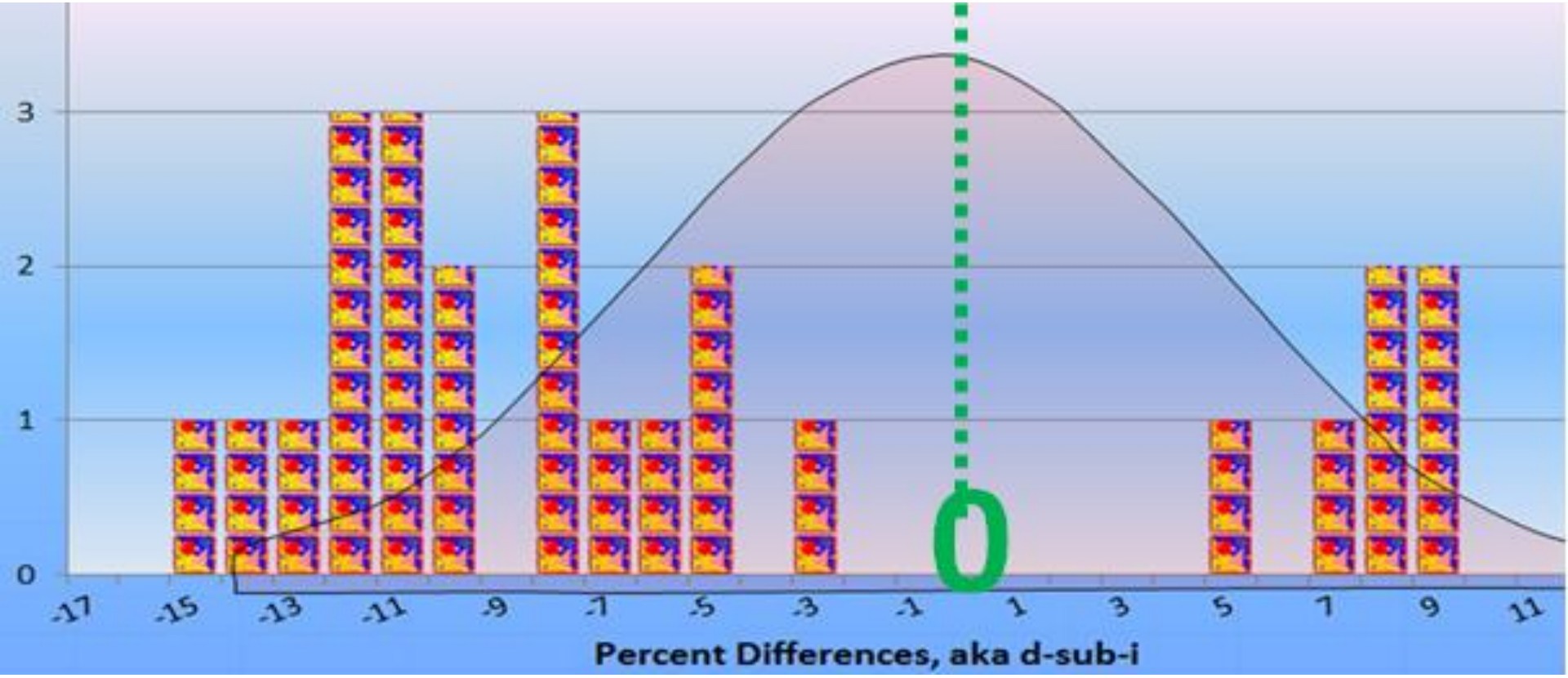
AS is the STDEV of the abs value of these d_i's = **3.08**

So
Abs value of bias = 9.3 + 1.71 * (3.08/sqrt of n)
= 10.39 (!)

Thanks Shelly Eberly!

d (Eqn. 1)
9
-3
-12
-12
-10
-11
-9
4
7
8
8
9
-11
-11
-7
-6
-6
-9
-9
-13

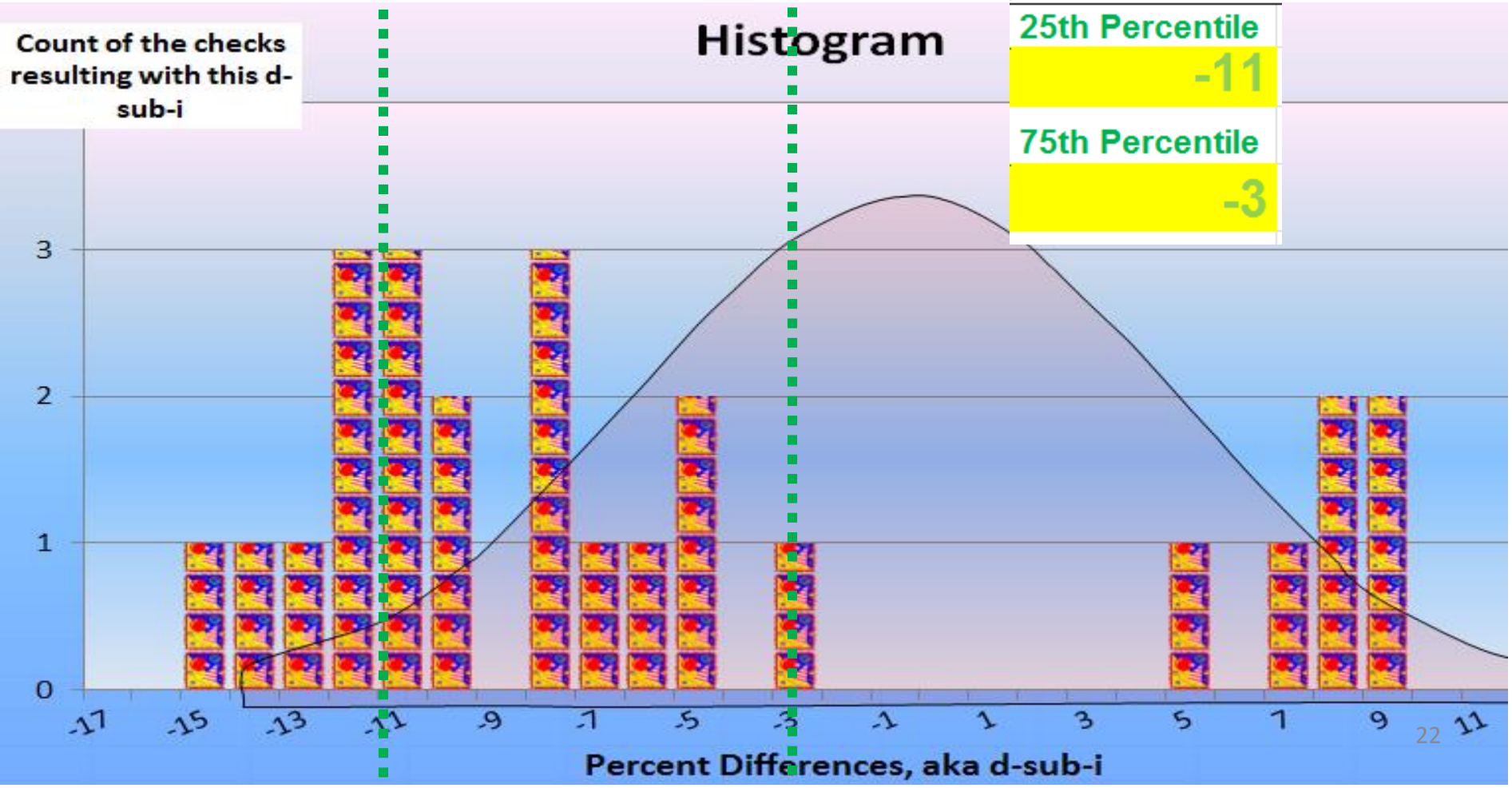
That 10.39 is the abs value of bias, now what's its sign?



- Look at 25% quartile and 75% quartile
- If they straddle zero, bias is unsigned
- If they're both negative, bias is negative
- If they're both positive, bias is positive

Quartiles?

- =QUARTILE(d-sub-i,1) = 25% quartile = -11
- =QUARTILE(d-sub-i,3) = 75% quartile = -3



DASC bias in cell k13:

- Both quartiles are negative
- Bias is negative 10.4 = **-10.4**
- Agrees with DASC:

Pollutant type: O_3					CV_{ub} (%)			Bias (%)	
d (Eqn. 1)	25th Percentile	d^2	|d|	$\|d\|^2$					
9	-11.111	79.012	8.889	79.012					
-3	75th Percentile	11.111	3.333	11.111	n	S_d	S_{d2}	$\Sigma\|d\|$	"AB" (Eqn 4)
-12	-3.333	149.383	12.222	149.383	25	8.089	59.010	233.333	9.333
-12		149.383	12.222	149.383	n-1	Σd	Σd^2	$\Sigma\|d\|^2$	"AS" (Eqn 5)
-10		100.000	10.000	100.000	24	-144.444	2404.938	2404.938	3.077
-11		123.457	11.111	123.457					
-9		79.012	8.889	79.012				Bias (%) (Eqn 3)	Both Signs Positive
4		19.753	4.444	19.753				10.39	FALSE
7		44.444	6.667	44.444	CV (%) (Eqn 2)			Signed Bias (%)	Both Signs Negative
8		60.494	7.778	60.494	10.01			-10.39	TRUE

Does this match AQS standard report AMP256 ?:

DATA QUALITY INDICATOR REPORT

One Point Quality Control

Aug. 2, 2016

App A? Y

POC	MT	Begin Date	End Date	Intervals Required	Valued Intervals	% Complete	CV UB	Bias UB
1	SP	01-JAN-14	31-DEC-14	12	9	75	10.01	- 10.39
1	SP	01-JAN-14	31-DEC-14	13	10	77	10.01	- 10.39
		01-JAN-14	31-DEC-14	25	19	76	10.01	- 10.39
		01-JAN-14	31-DEC-14	25	19	76	10.01	- 10.39

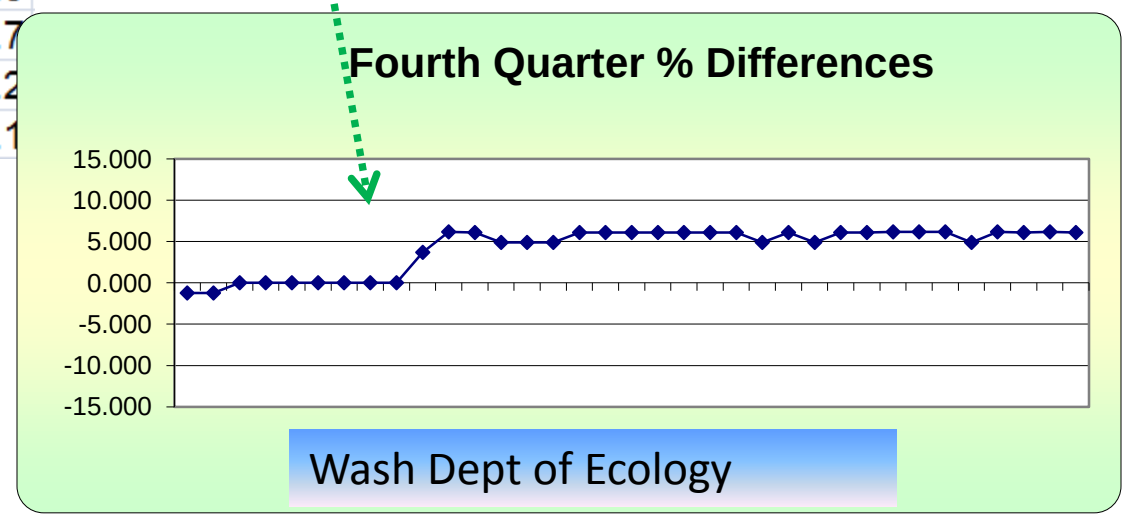
Bias UB *(upper bound of bias)* = **-10.39**

(goal is upper 95 percent confidence limit for the absolute bias of 7 percent)

Both bias and precision are in the same sheet (O3 P&B) in the DASC and use the same input:

Meas Val (Y)	Audit Val (X)	d_i (Eqn. 1)
0.08	0.081	-1.2
0.08	0.081	-1.2
0.081	0.081	0.0
0.081	0.081	0.0
0.081	0.081	0.0
0.081	0.081	0.0
0.081	0.081	0.0
0.081	0.081	0.0
0.081	0.081	0.0
0.084	0.081	3.7
0.086	0.081	6.2
0.087	0.082	6.1

YOU can calculate Bias over any time period using DASC

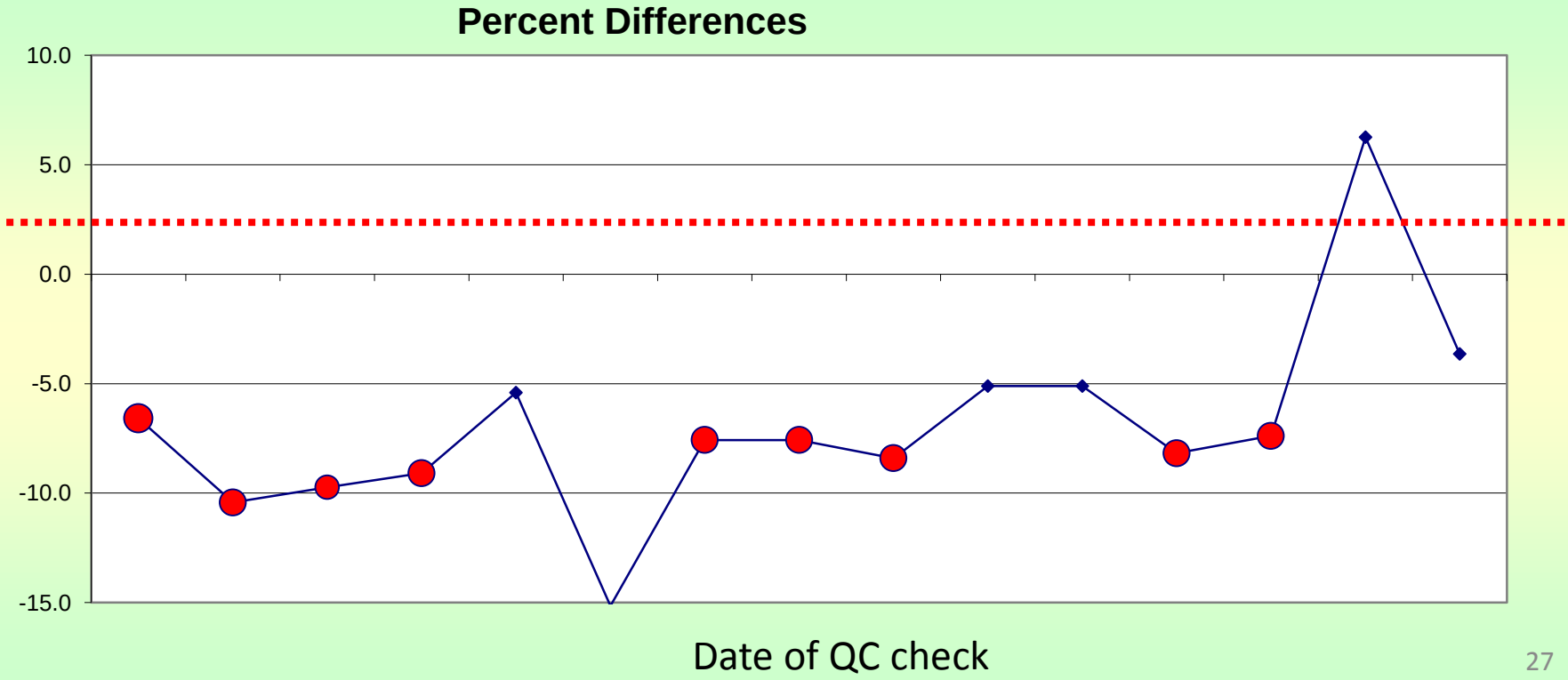


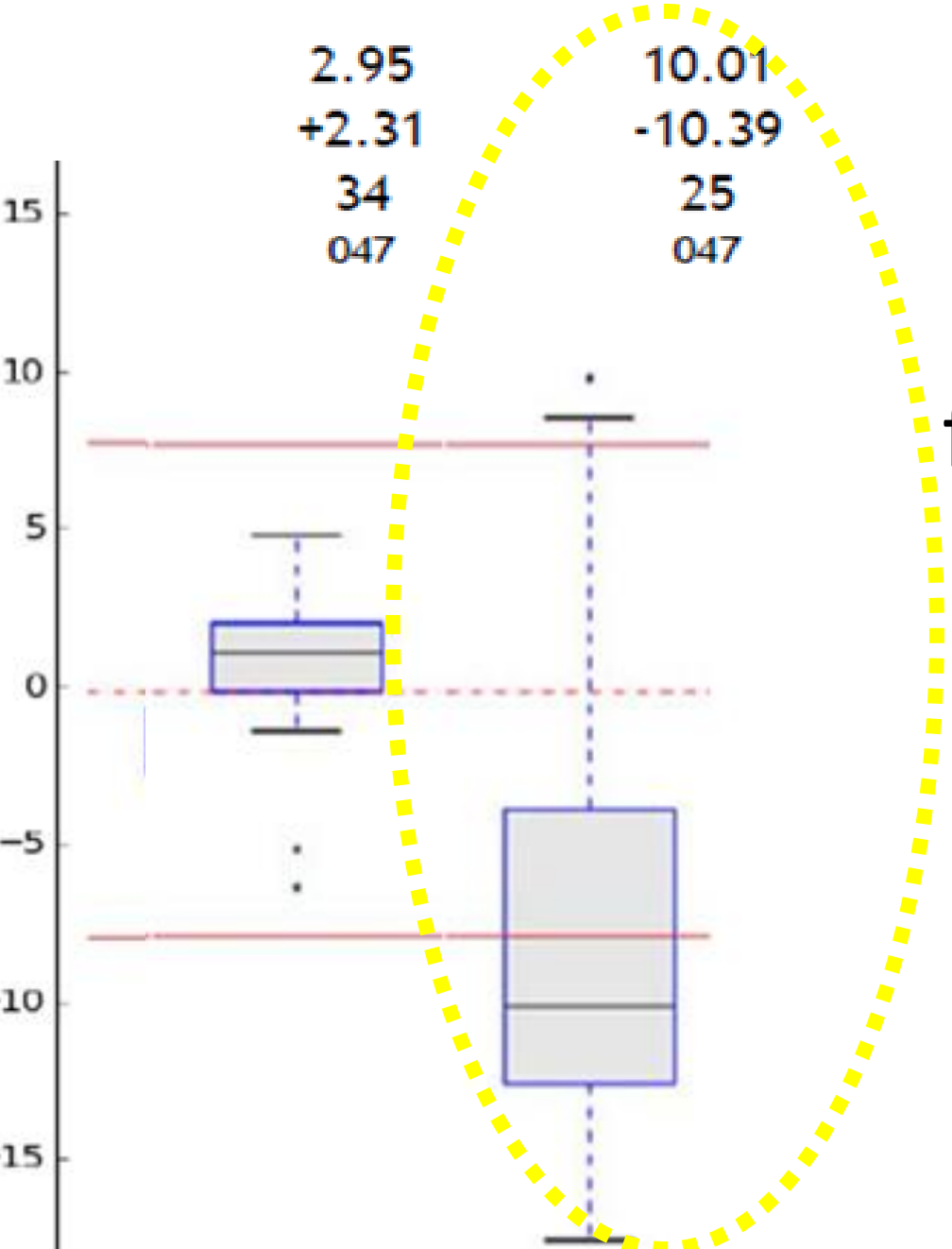
Summary of gas bias:

- Calculated from routine QC checks d_i
- Overall upper limit of **bias** calculated from d_i
- Then look at the sign (and the chart) for whether your analyzer is biased high (+) or low (-)
- We are 95% confident that our 03 bias is less extreme than **-10%**

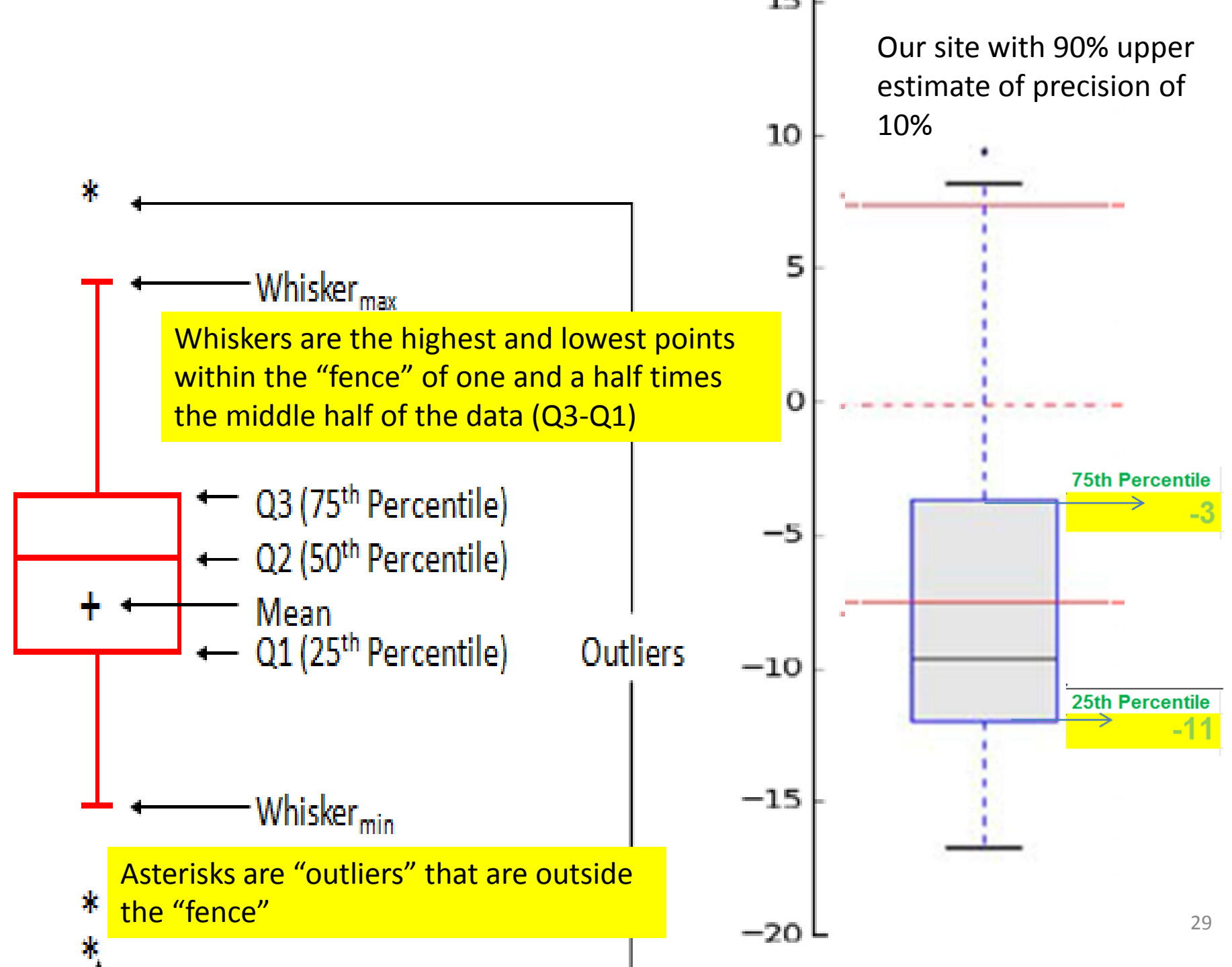
Do I invalidate pollutant data based on d-sub-i?

- Validation tables in QA Handbook:
  - Critical Measurement Quality Objective O3=7%
  - See problems ahead of time by identifying trends in a control chart:





Box and whisker charts show this faster—compare our site on the right with a different CA site


Whisker max
Whiskers are the highest and lowest points within the "fence" of one and a half times the middle half of the data (Q3-Q1)
Q3 (75th Percentile)
Q2 (50th Percentile)
Mean
Q1 (25th Percentile)
Outliers
Whisker min
Asterisks are "outliers" that are outside the "fence"
Our site with 90% upper estimate of precision of 10%
10
5
0
−5
−10
−15
−20
75th Percentile
-3
25th Percentile
-11

$PM_{2.5}$ Precision

- PM2.5 is the same as gaseous, except:
  - d-sub-i are from COLLOCATED, and the known is the average of the two PM2.5, so d-sub-i is
  - (RO-CO)/(avg of RO & CO)
  - Because the known is the avg of 2 measurements, add **SQRT(2)** to the denominator (divide by best estimate of truth)

STDEV

$$CV\_ub = \sqrt{\frac{n \cdot \sum_{i=1}^{n} d_i^{\,2} - \left(\sum_{i=1}^{n} d_i\right)^2}{2n(n-1)}} \cdot \sqrt{\frac{n-1}{\chi^2_{0.1,n-1}}}$$

That's the only difference in the precision stat from gas stats

$PM_{2.5}$ Bias

- PM2.5 bias same as gaseous, except:
  - **known = PEP** audit filter results, so the d-sub-i is the (field-PEP)/PEP
  - Don't take abs value of the d-sub-i
  - D is avg of these d-sub-i values
  - n is # of PEP audits, and if n=3 then t=2.9
    - (as n grows, $t_{0.95}$ goes to 1.65)
  - Use the 25% and 75% quartiles ➔ + or -

Stnd error

$$\text{Upper 90\% Confidence Interval} = D + t_{0.95,df} \cdot \frac{s_d}{\sqrt{n_j}}$$

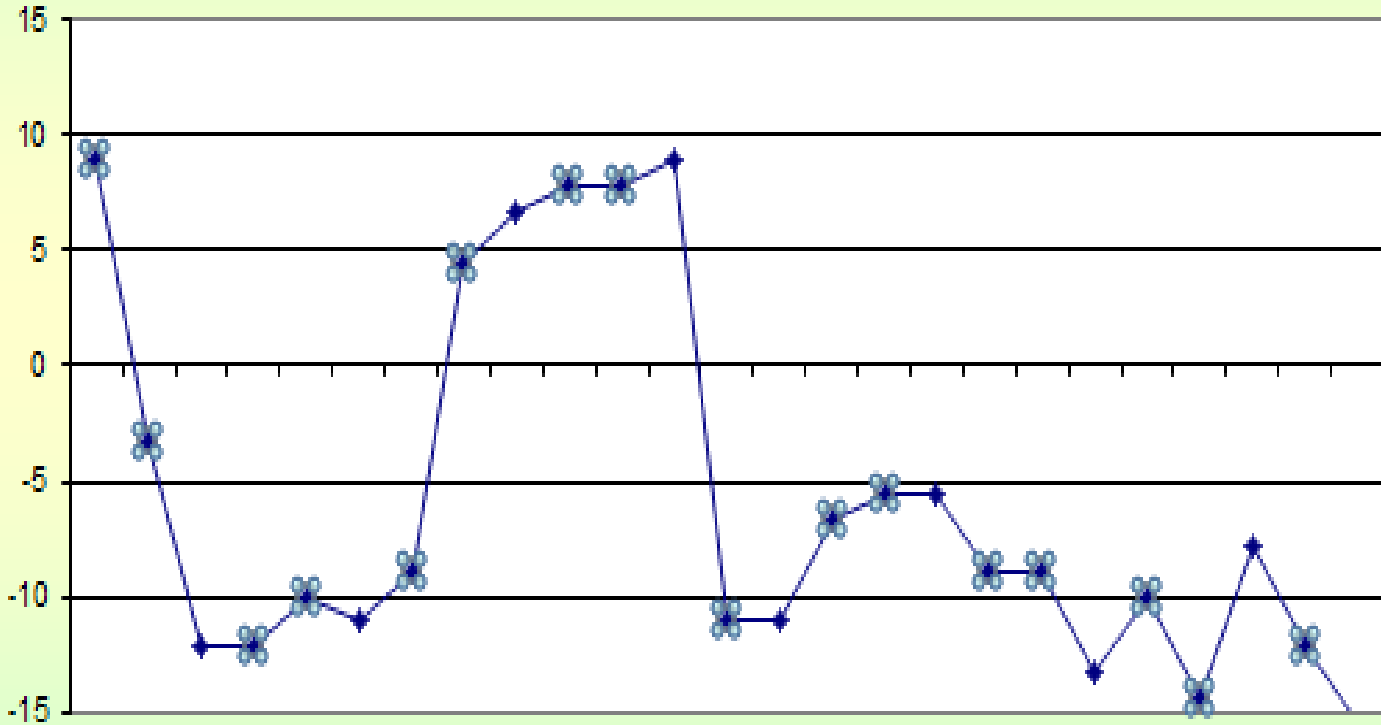
(and the lower confidence interval is D minus t*stnd error!)

PM10 statistics:

- Bias confidence intervals based on monthly flow rate (FR) checks:
  - d-sub-i from FR
  - THEN bias statistics are the same as PM2.5
- Flow rate “acceptability” limits are based on 6-month FR audits (with FR audit device not the same one you use for the monthly):
  - Limit = D +- 1.96 * STDEV

d-sub-i = (sampler-audit_FR)/audit_FR
and D is their average

Meas Val (Y	Audit Val (2	d (Eqn. 1)
98	90	9
87	90	-3
79	90	-12
79	90	-12
81	90	-10
80	90	-11
82	90	-9
94	90	4
96	90	7
97	90	8
97	90	8
98	90	9
80	90	-11
80	90	-11
84	90	-7
85	90	-6
85	90	-6
82	90	-9
82	90	-9
78	90	-13
81	90	-10
77	90	-14
83	90	-8
79	90	-12
76	90	-16



Date of QC check

Do not overlook the built-in control chart, from which you can tell the story of your QC

Thank you!

- Work with Tribal Air Agencies
- Knowledge = Power; Let's Share
  - **http://datatools.tamscenter.com/**
- Melinda Ronca-Battista melinda.ronca-battista@nau.edu; https://www.youtube.com/c/melindaroncabattista