# STEPS TO TREAT DATA FOR CALIBRATION CURVE BY LINEAR REGRESSION

170524 /jp

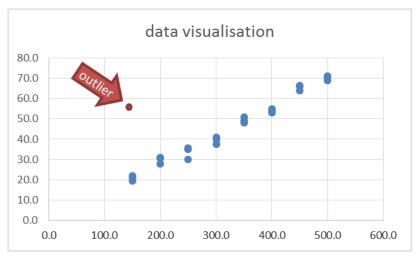
### 1/ DATA VISUALISATION

Input is a table of X-Y data:

c(mg/L)	signal (mV)		
150.0	22.0	21.0	19.5
200.0	30.5	31.0	28.0
250.0	35.0	35.5	30.0
300.0	40.0	41.0	37.5
350.0	51.0	49.0	48.0
400.0	55.0	53.0	54.0
450.0	64.0	66.5	66.0
500.0	70.5	71.0	<u>69.0</u>

The independent variable (X) are: concentrations of standard (mg/l, ug/ml or mol/l), volumes of standard, milligrams or millimols of standard etc. The number of the levels is typically 5-7. Since experiments should be replicated (at least 3x), first question is: how many points should be shown in a graph? The answer is: **all data measured** should be in the graph (N=24). **Do not average** the values at the same x-levels (rows)! This approach would hide outliers and also decrease degree of freedom! If replicates are present, rewrite the data into two columns (to make a graph in Excel easy) and plot a scattered (x-y) graph. The visualisation helps to reveal obvious outliers (gross errors) that should be removed.

L	
150.0	22.0
200.0	30.5
250.0	35.0
300.0	40.0
350.0	51.0
400.0	55.0
450.0	64.0
500.0	70.5
150.0	21.0
200.0	31.0
250.0	35.5
300.0	41.0
350.0	49.0
400.0	53.0
450.0	66.5
500.0	71.0
150.0	19.5
200.0	28.0
250.0	30.0
300.0	37.5
350.0	48.0
400.0	54.0
450.0	66.0
500.0	69.0



Note: in this step it become also clear that there is a strong correlation between Y and X, which is not surprising: since the beginning we intended to construct a calibration curve! Therefore, it is not necessary to prove a strong correlation according to Spearman.

#### 2/ THE FIRST MODEL OF LINEAR REGRESSION Y = b \* X + a

Our ultimate goal is to find an equation of a line that fits best the points, which will be called "calibration line". For the first estimation we may use either right button click on the graph points (Add trendline + linear + show the equation in the graph – check the red arrows below)

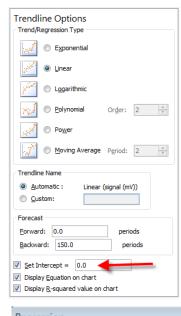
Trendline Options Trend/Regression Type	linear regression curve					
Trend/Regression Type <ul> <li>Exponential</li> <li>Exponential<th>80 70 60 50 40 30 20 10 -10 9 100</th><th>y = 0.1417:</th><th>-</th><th></th><th>••</th><th>600</th></li></ul>	80 70 60 50 40 30 20 10 -10 9 100	y = 0.1417:	-		••	600
Custom:						
Forecast Eorward: 0.0 periods Backward: 150.0 periods	or using a dialo	og Data - I	Data ar	alysis -	- Regress	sion:
Set Intercept = 0.0						
	C1.11		IT.			
Display Equation on chart	SUI	MMARY OUTPU	JT			
	SU				=correl()	
Display Equation on chart		MMARY OUTPU Regression Sta			=correl() 0.99133811	
<ul> <li>Display Equation on chart</li> <li>Display <u>R</u>-squared value on chart</li> </ul>	Mu	Regression Sto	atistics			
Display Equation on chart     Display R-squared value on chart	Mu R S	Regression Sto Itiple R	atistics 0.991338 0.982751			
Display Equation on chart     Display R-squared value on chart Regression Input	Mu R S Adj	Regression Sta Itiple R quare usted R Square ndard Error	atistics 0.991338 0.982751 0.981967 2.245787			
Display Equation on chart     Display R-squared value on chart     degression	Mu R S Adj	<i>Regression Sta</i> Itiple R quare usted R Square	atistics 0.991338 0.982751 0.981967			
Display Equation on chart     Display R-squared value on chart Regression Input	Mu R S Adj Sta	Regression Sta Itiple R quare usted R Square ndard Error	atistics 0.991338 0.982751 0.981967 2.245787			
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:	Mu R S Adj Sta	Regression Sta Itiple R quare usted R Square ndard Error servations	atistics 0.991338 0.982751 0.981967 2.245787	SS		F .
☑ Display Equation on chart         ☑ Display R-squared value on chart         ☑ Regression         Input         Input Y Range:	Mu R S Adj Sta Ob:	Regression Sta Itiple R quare usted R Square ndard Error servations	atistics 0.991338 0.982751 0.981967 2.245787 24 df	<u>SS</u> 6321.875	0.99133811	
Display Equation on chart         Display R-squared value on chart         Input         Input Y Range:         \$\$\$4\$\$11:\$\$\$35         Input X Range:	Mu R S Adj Sta Ob: AN Reg	Regression Sta Iltiple R quare usted R Square ndard Error servations OVA	atistics 0.991338 0.982751 0.981967 2.245787 24 df 1	6321.875	0.99133811 MS	
☑ Display Equation on chart         ☑ Display R-squared value on chart         ☑ Display R-squared value on chart         Input         Input Y Range:       \$B\$11:\$B\$35         Input X Range:       \$A\$11:\$A\$35         ☑ Labels       Constant is Zero	Mu R S Adj Sta Ob: AN Reg	Regression Sta Iltiple R quare usted R Square ndard Error servations OVA gression sidual	atistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22	6321.875	0.99133811 	
☑ Display Equation on chart         ☑ Display R-squared value on chart         ☑ Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         ¥A\$11:\$A\$35         ☑ Labels	Mu R S Adj Sta Ob Essi AN Reg Res Tot	Regression Sta Iltiple R quare iusted R Square ndard Error servations OVA gression sidual tal	atistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23	6321.875 110.9583 6432.833	0.99133811 MS 6321.875 5.04356061	1253.455
☑ Display Equation on chart         ☑ Display R_squared value on chart         ☑ Display R_squared value on chart         Regression         Input         Input Y Range:       \$B\$11:\$B\$35         Input X Range:       \$A\$11:\$A\$35         ☑ Labels       Constant is Zero         ☑ Confidence Level:       95         %	Mu R S Adj Sta Ob: ES AN Reg Res Tot	Regression Sta Iltiple R quare usted R Square ndard Error servations OVA gression sidual tal	ntistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23 coefficients	6321.875 110.9583 6432.833 andard Erro	0.99133811 MS 6321.875 5.04356061 t Stat	1253.45 P-value
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output options         Input Range:         \$F\$39	Mu R S Adj Sta Ob: ES AN Reg Res Tot	Regression Sta Itiple R quare usted R Square ndard Error servations OVA gression sidual tal	ntistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23 officients -0.70833	6321.875 110.9583 6432.833 andard Erro 1.378892	0.99133811 MS 6321.875 5.04356061 t Stat -0.5136976	1253.455 P-value L 0.612584
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output options <ul> <li>Qutput Range:</li> <li>\$F\$39</li> <li>New Worksheet Ply:</li> <li></li> </ul> <li>Input Support Support</li>	Mu R S Adj Sta Ob: ES AN Reg Res Tot	Regression Sta Iltiple R quare usted R Square ndard Error servations OVA gression sidual tal	ntistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23 officients -0.70833	6321.875 110.9583 6432.833 andard Erro 1.378892	0.99133811 MS 6321.875 5.04356061 t Stat	1253.455 P-value L 0.612584
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output Range:         \$F\$39         New Worksheet Ply:         New Workbook	Mu R S Adj Sta Ob: ES AN Reg Res Tot	Regression Sta Iltiple R quare usted R Square ndard Error servations OVA gression sidual tal Co ercept ng/L)	atistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23 coefficients -0.70833 0.141667	6321.875 110.9583 6432.833 andard Erro 1.378892 0.004001	0.99133811 MS 6321.875 5.04356061 t Stat -0.5136976 35.4041629 -T.INV.2T(0.05	1253.455 P-value L 0.612584 6.82E-21 N-2)= 2.07387
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output options         ● Qutput Range:         \$F\$39         ○ New Worksheet Ply:         ○ New Workbook         Residuals	Mu R S Adj Sta Ob ES AN Reg Res Tot	Regression Sta Itiple R quare usted R Square ndard Error servations OVA gression sidual tal	atistics 0.991338 0.982751 0.981967 2.245787 24 df 1 22 23 coefficients -0.70833 0.141667	6321.875 110.9583 6432.833 andard Erro 1.378892 0.004001	0.99133811 <i>MS</i> 6321.875 5.04356061 <i>t Stat</i> -0.5136976 35.4041629	1253.455 P-value L 0.612584 6.82E-21 N-2)= 2.07387
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output Range:         \$F\$39         New Worksheet Ply:         New Workbook	Mu R S Adj Sta Ob ES AN Reg Res Tot Con Con Con Con Con Con Con Con Con Con	Regression State         Iltiple R         quare         usted R Square         ndard Error         servations         OVA         gression         sidual         tal         carcept         ng/L)         = $\frac{ a }{s_a} = \frac{0.7083}{1.3789}$	$\begin{array}{c} \text{atistics} \\ 0.991338 \\ 0.982751 \\ 0.981967 \\ 2.245787 \\ 24 \\ \hline \\ df \\ 1 \\ 22 \\ 23 \\ \hline \\ efficients \\ -0.70833 \\ 0.141667 \\ \hline \\ \frac{3}{9} = 0.5137 \\ \end{array}$	6321.875 110.9583 6432.833 andard Erro 1.378892 0.004001 , t(crit.)= t(crit.)=	0.99133811 MS 6321.875 5.04356061 t Stat -0.5136976 35.4041629 -T.INV.2T(0.05 -TINV(0.05;N-2	1253.455 P-value L 0.612584 6.82E-21 (N-2)= 2.07387 )= 2.07387
✓ Display Equation on chart         Display R-squared value on chart         Regression         Input         Input Y Range:         \$B\$11:\$B\$35         Input X Range:         \$Labels         Confidence Level:         95         %         Output options         ● Qutput Range:         \$F\$39         New Worksheet Ply:         New Workbook         Residuals         Residuals	Mu R S Adj Sta Ob Sta Ob AN Reg Tot Tot S S Tot S S S S S S S S S S S S S S S S S S S	Regression Statute Itiple R quare usted R Square ndard Error servations OVA gression sidual tal construction ercept ng/L) = $\frac{ a }{s_a} = \frac{0.7083}{1.3789}$	$\frac{atistics}{0.991338}$ $0.982751$ $0.981967$ $2.245787$ $24$ $df$ $1$ $22$ $23$ $0efficients$ $-0.70833$ $0.141667$ $\frac{3}{9} = 0.5137$ $recomi$	6321.875 110.9583 6432.833 andard Erro 1.378892 0.004001 , t(crit.)= t(crit.)=	MS 6321.875 5.04356061 t Stat -0.5136976 35.4041629 T.INV(0.05;N-2 d becaus	1253.455 P-value L 0.612584 6.82E-21 N-2)= 2.07387

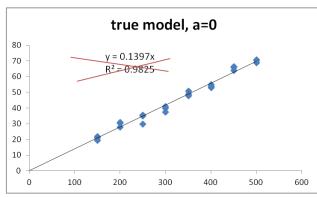
## 3/ IS THE INTERCEPT SIGNIFICANT?

Since our measurements are subjected to random errors, it may happen that another series of data obtained under the same conditions would give an equation of e.g. Y=0.1417\*x + 0.001, and another series for example Y=0.1417\*x - 0.111. This suggests that the value (and sign) of intercept has no meaning. If this is true, very important consequence follows: IT SHOULD BE SET TO ZERO. The table generated by Data analysis gives information, if the intercept -0.7083 of our regression model Y = 0.1417 \* X - 0.7083 is significantly different from 0.

To confirm the null hypothesis a=0 (on the level  $\alpha=0.05$ ), we should compare ratio  $|a|/s_a$  to t(crit.) (see the yellow parts above 0.532<2.0739) or, more simply, compare *P*-value to 0.05 (0.613 > 0.05). In our case, both the values confirm the null hypothesis. An easy rule to remember is: we set the intercept to zero if the *P*-value is higher than 0.05.

In this case we have to FORGET about the first regression model and to CHANGE it to a more simple equation  $Y = b^*X$ , which means that the procedure of finding the true equation has to be repeated with Y(0)=0 (the slope must slightly change, the statistics must also change):





Note that the graphical approach **Add trendline gives a wrong value of R-squared**. Therefore, use Data - Data analysis – Regression instead, which yields both correct R=0.998993 and slope=0.139728:

\$8\$11:\$8\$35	SUMMARY OUTF Regression St Multiple R R Square Adjusted R Square	tatistics 0.998993 0.997986		=correl() 0.99133811
\$A\$11:\$A\$35 💽	Multiple R R Square Adjusted R Squar	0.998993 0.997986		
nstant is <u>Z</u> ero	Multiple R R Square Adjusted R Squar	0.998993 0.997986		
nstant is <u>Z</u> ero	R Square Adjusted R Squar	0.997986		0.99133811
_	Adjusted R Squar			
_		€ 0.954508		
%				
/0	Standard Error	2.209557		
	Observations	24		
	ANOVA			
\$F\$75 🔣		df	SS	MS
	Regression	1	55643.21	55643.2107
	Residual	23	112.2893	4.8821415
	Total	24	55755.5	
Residual Plots		Coefficients	andard Erro	t Stat
	Intercept	0	#N/A	#N/A
	c(mg/L)	0.139728	0.001309	106.758118
	F\$75 ■ Resigual Plots Une Fit Plots	ANOVA     ARegression     Residual     Total     Total     Line Fit Plots     Intercept	ANOVA     A	ANOVA ANOVA ANOVA ANOVA ANOVA AnovA Begression AnovAnovA Begression AnovA

## 4/ CALCULATION OF THE AMOUNT IN AN UNKNOWN

No matter if the intercept was set to zero or not, the final task with calibration curve is to calculate the amount (concentration) in an unknown. We take the correct regression equation and calculate X for a given Y:

**X** = **Y** / **b** (if the intercept was set to zero)

Or, if the intercept a was significant: X = (Y - a) / b

Usually, the unknown was also measured repeatedly; then we find Y as an average of the multiplicates, after testing if there is not an outlier.