

4 Regression und Curve Fitting

4.1 Linear Standard Curves

Datensatz: Lineare_Regression_Rohdaten.pzf
Lineare_Regression_Projekt.pzf

Step by Step: Linear standard curves.pdf

Eine lineare Regression ist die einfachste Art einer Regression. Es wird versucht den gerichteten Zusammenhang zwischen zwei kontinuierlichen Variablen zu quantifizieren. Dabei wird auch auf die Anpassungsgüte des Modells geachtet. Eine der beiden Variablen wird als unabhängige Variable bezeichnet, die andere als abhängige. Die Richtung der Wirkung ist somit klar. Die Unabhängige beeinflusst die Abhängige.

Ziel der Entwicklung eines Regressionsmodells ist auch die Schätzung unbekannter Daten. In nachfolgendem Beispiel werden verschiedene Proteinkonzentrationen mit einem optischen Verfahren gemessen. Als Messwert wird die „Optical Density“ erfasst. Aus den ersten fünf Zeilen wird das Regressionsmodell entwickelt, mit dem die unbekannt Proteinkonzentrationen geschätzt werden können. In der Laborpraxis entsprechen die ersten fünf Zeilen einem Standard um das Messgerät zu kalibrieren.


Statistische Hintergründe zur linearen Regression liefert das PRISM RegressionBook im Kapitel 7.

Öffnen Sie bitte den oben benannten Rohdatensatz oder geben Sie die entsprechenden Daten in ein neues Projekt ein. Überlegen Sie sich an Hand der Informationen aus Kapitel 2.3 welches Datentabellen- und Graphen-Format Sie benötigen. Die Form der Datentabelle gibt zumindest einen kleinen Hinweis.

Tabelle 13: Daten für eine lineare Regression

Protein [mg/mL]	Optical Density
0.05	0.097
0.10	0.183
0.20	0.290
0.30	0.371
0.40	0.462
	0.141
	0.190
	0.258
	0.423

Als Datenstruktur müssen Sie eine XY-Organisation auswählen. Die Datentabelle weißt keine Messwiderholungen der Optical Density für die einzelnen Proteinkonzentrationen aus.

Nach Eingabe der Daten wählen Sie „Analyze -> XY Analyses -> Linear regression“. Im sich öffnenden Dialog (vgl. Abbildung 56) können Sie weitere Parameter festlegen.

Sie sollten auswählen, dass unbekannte Daten geschätzt werden sollen.

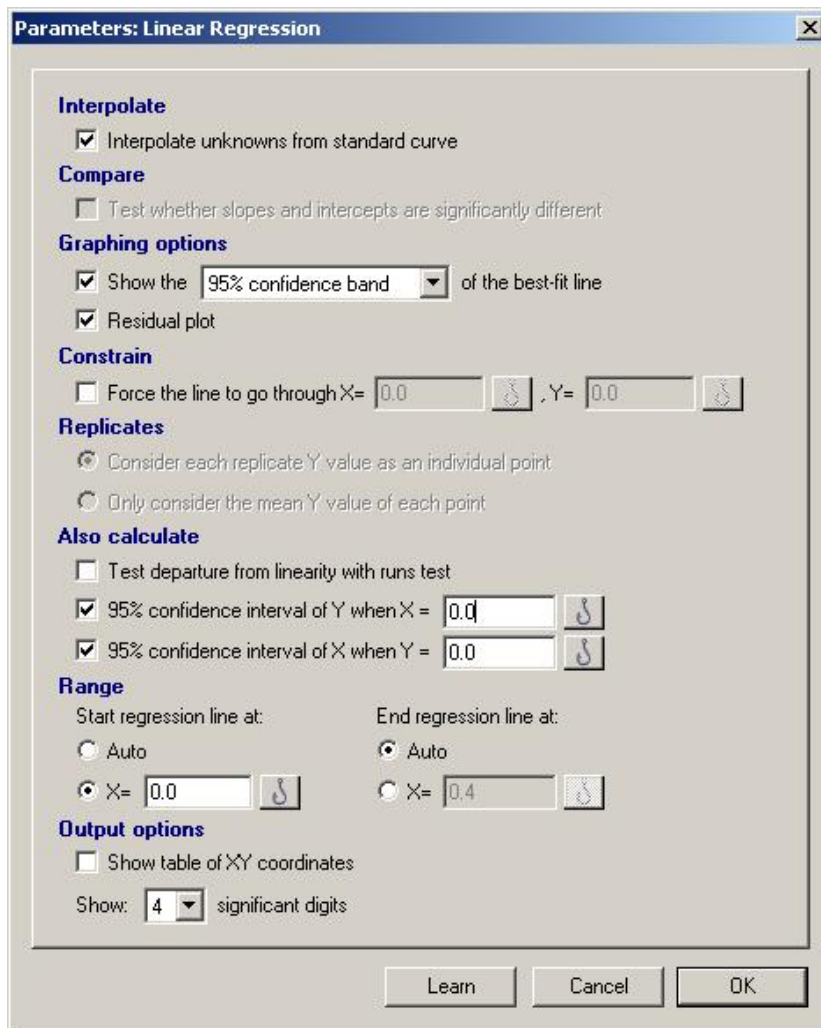


Abbildung 56: Optionen einer linearen Regression

Das „95% confidence band“ zeigt den Vertrauensbereich der Regression.

Das Residual Plot visualisiert die Abweichung der beobachteten von den erwarteten Werten über den Vorhersagebereich. Das Plot sollte keine Systematik zeigen.

Außerdem sollten Sie festlegen, dass die Regressionsgerade ab $X = 0$ gezeichnet werden soll. Falls Sie dieses nicht tun, würde sie ab dem niedrigsten X-Wert gezeichnet werden.

Ein Klick auf OK erzeugt zum einen ein Tabellenoutput unter „Results“ als auch eine Regressionslinie im Graphen.

Im Resultsbereich wird eine Tabelle mit den Modellparametern und eine andere Tabelle mit den geschätzten Daten erzeugt.

Im der Resulttabelle (vgl. Tabelle 14) müssen in jedem Fall zwei Dinge interpretiert werden.

Die Modellgüte (Goodness of Fit) gibt an wie gut die Regressionsgerade die zu beschreibenden Daten trifft. Der Wert von r^2 liegt im Bereich von > 0 und < 1 . Werte von über 0,8 können in der Regel als gut gelten. Letztlich ist es abhängig von Ziel, welches mit der Regression verfolgt wird.

Über die Steigung wird die Frage geklärt, ob eine Änderung der unabhängigen Variablen überhaupt zu einer Änderung der abhängigen führt. Ohne dass dieser Effekt vorhanden ist macht ein Regressionsmodell keinen Sinn, die Variablen hängen nicht voneinander ab. Die Steigung kann sowohl positiv als auch negativ sein. Es wird geprüft ob sie signifikant unterschiedlich von 0 ist (Is slope significantly non-zero?). Ist sie unterschiedlich von 0 ist ein Effekt vorhanden.

Der „Runs Test“ liefert eine Information darüber, ob die Annahme eines linearen Zusammenhangs haltbar ist und prüft die Abweichung von der Linearitätsannahme. Im vorliegenden Fall kann von einer Linearität ausgegangen werden.

Tabelle 14: Modellparameter lineare Regression

		Optical Density
1	Best-fit values	
2	Slope	1.008 ± 0.06857
3	Y-intercept when X=0.0	0.06901 ± 0.01686
4	X-intercept when Y=0.0	-0.06849
5	1/slope	0.9925
6	95% Confidence Intervals	
7	Slope	0.7894 to 1.226
8	Goodness of Fit	
9	r ²	0.9863
10	Sy.x	0.01963
11	Is slope significantly non-zero?	
12	F	215.9
13	DFn, DFd	1.000, 3.000
14	P value	0.0007
15	Deviation from zero?	Significant
16	Data	
17	Number of X values	5
18	Maximum number of Y replicates	1
19	Total number of values	5
20	Number of missing values	0
21	Runs test	
22	Points above line	2
23	Points below line	3
24	Number of runs	3
25	P value (runs test)	0.5000
26	Deviation from linearity	Not Significant

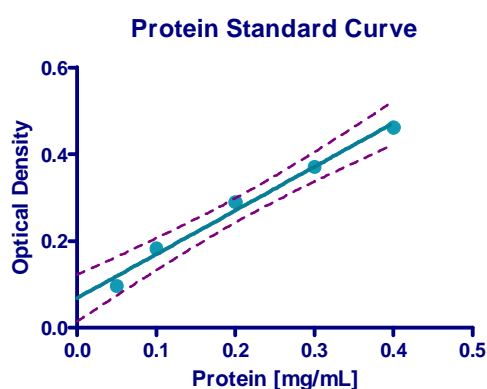


Abbildung 57: Graph lineare Regression

Das obige Modell liefert $r^2 = 0.986$ und eine höchst signifikanten Unterschied der Steigung von 0. Das Modell ist als durchaus gut zu bezeichnen. Nebenstehender Graph bestätigt die genannten Zahlen recht gut.

Es gibt nun die Möglichkeit die geschätzten Daten sowohl als Tabelle als auch graphisch dem Graphen hinzuzufügen.

Die Tabelle mit den geschätzten Werten können Sie über „Copy and Paste“ in den Graphen einfügen.

Markieren Sie dazu den Datenbereich in der Tabelle „Interpolated X values ...“ und wählen Sie über die rechte Maustaste „Copy“, wechseln Sie danach zurück zum Graphen und wählen Sie über die rechte Maustaste „Paste Table“. Schieben Sie Tabelle an die gewünschte Stelle.

Um die geschätzten Daten in graphischer Form müssen Sie die interpolierten Daten als neue Datenreihe hinzufügen. Die Vorgehensweise ist grundsätzlich in Kapitel 2.4 beschrieben.

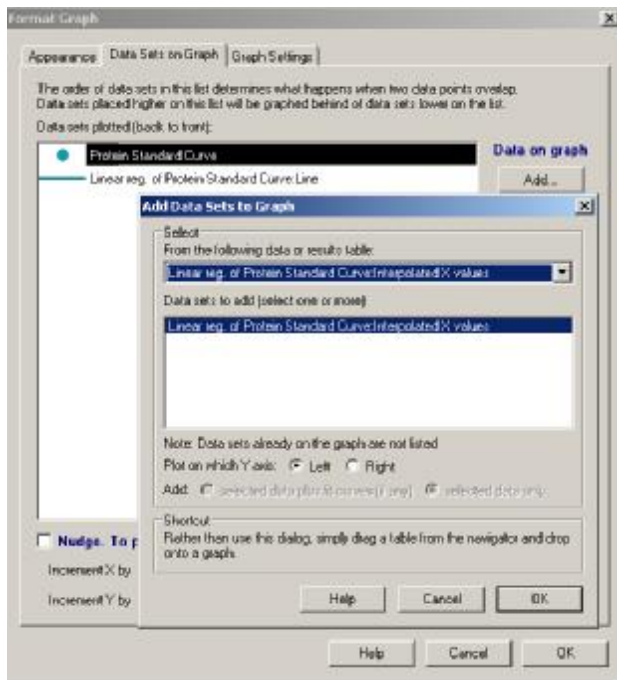




Abbildung 58: Add Data Set

Wechseln Sie bitte zum Graphen und

wählen Sie „Change ->  (= Add Data Sets)“. Wählen Sie wie nebenstehend den Datensatz mit den interpolierten Daten aus.

Die neuen Daten erscheinen als farbige Zeichen im Plot.

Über „Change ->  (= Format Graph (symbols, bars, ...))“ können Sie die Darstellungseigenschaften des eingefügten Datensatzes ändern.

Im „Format Graph“ Dialog müssen Sie im Pull-Down-Menü den die Datenreihe der Interpolierten Daten anwählen.

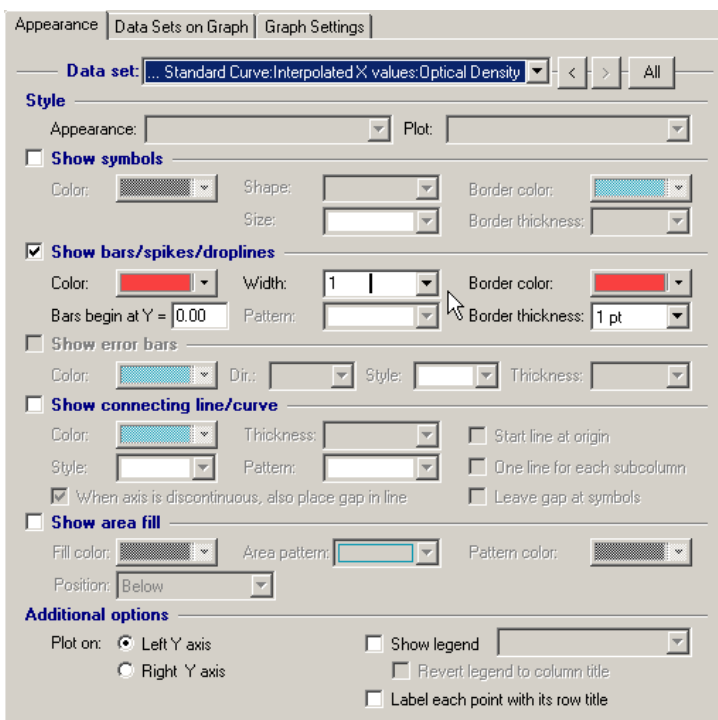


Abbildung 59: Change Symbols

Nebenstehende Auswahl erzeugt eine farbige Linie. Außerdem sollten Sie das Anzeigen der Legende abwählen.

Nachfolgender Graph ist das Ergebnis.

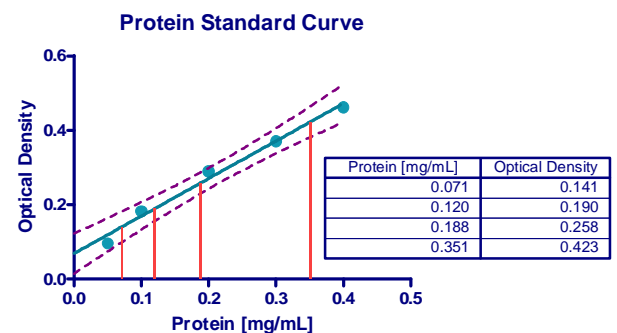


Abbildung 60: Guter Regressionsgraph