

<https://ml.azure.com>

<https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/module-reference>

Der Diabetes- Datensatz

Quelle: <https://www4.stat.ncsu.edu/~boos/var.select/diabetes.tab.txt>

Siehe auch: Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) "Least Angle Regression," Annals of Statistics (with discussion), 407-499.

Attribute Information:

- age age in years
- sex Geschlecht
- bmi body mass index
- bp average blood pressure
- tc total serum cholesterol
- ldl low-density lipoproteins
- hdl high-density lipoproteins
- tch total cholesterol / HDL
- ltg possibly log of serum triglycerides level
- glu blood sugar level
- prog progress

Einfache Lineare Regression in Excel

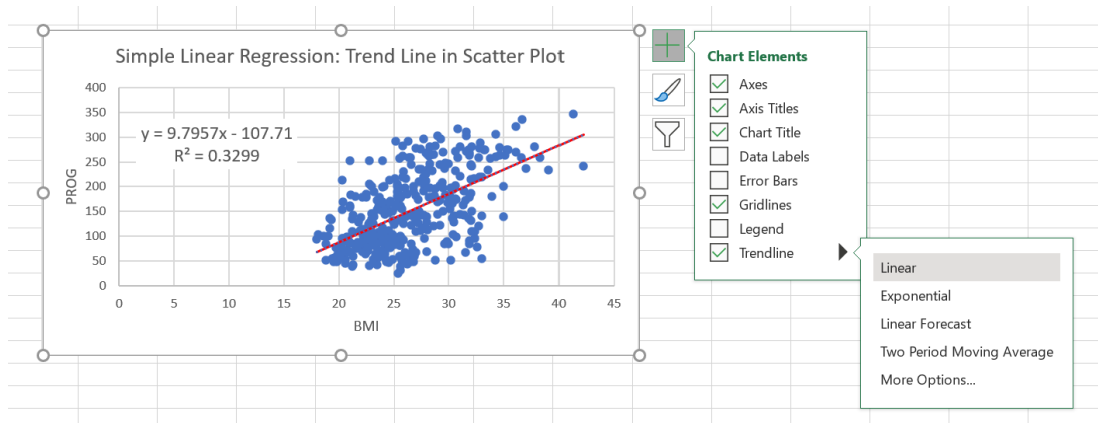
Die Daten in der Datei «Lineare Regression Diabetes.xlsx» enthalten bereits eine Spalte mit der Unterteilung in Trainings- und Testdaten: «train_test_split».

Berechne die lineare Regression des Krankheitsfortschritts «prog» auf den «bmi» in den Trainings-Daten. Am besten in einem separaten Tab «train»:

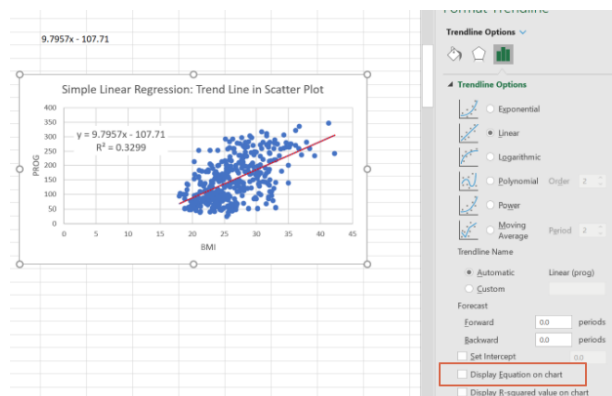
1. Markiere die beiden Spalten
2. Erstelle einen Scatter-Plot

ex	bmi	bp	tc	ldl	hdl	tch	ltg	glu	prog	Column
1	27.8	73	153	104.4	42.00	4	3.4965	73	201	TRUE
1	22.8	101	110	41.6	56.00	2	4.1271	97	40	TRUE
1	35	79.33	195	140.8	42.00	4.64	4.1109	96	140	TRUE
2	34.6	115	204	139.4	36.00	6	4.9628	109	264	TRUE
2	25.9	91	178	106.6	52.00	3	4.5747	75	92	TRUE
2	26.1	85	220	152.4	47.00	5	4.6347	91	139	TRUE
2	25.7	83	181	106.6	66.00	3	3.7377	85	52	TRUE
1	19.5	80	171	85.4	75.00	2	3.9703	80	48	TRUE
2	19.2	81	155	94.6	51.00	3	3.8501	87	116	TRUE

3. Füge eine Trendlinie hinzu



4. Bei den Optionen kann man sich die Gleichung der Ausgleichsgeraden anzeigen lassen etc.

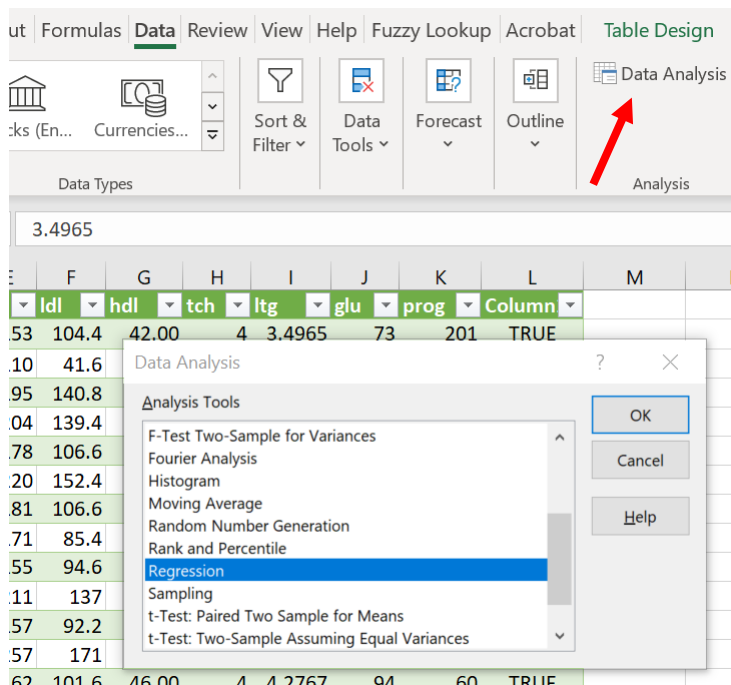


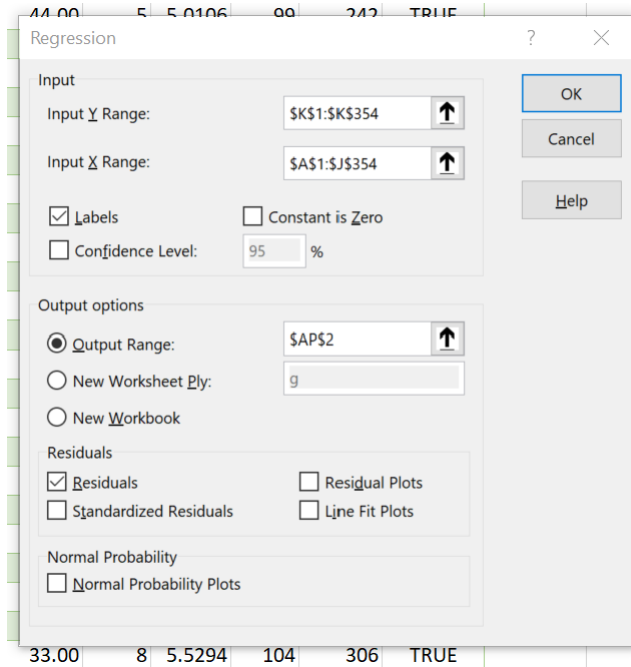
Berechne nun für die Testdaten die Vorhersage des Modells, der Übersichtlichkeit halber am besten in einem separaten Tab»test«.

Berechne den Mean Squared Error, die Root Mean Square Deviation und den Mean Absolute Error auf den Testdate im Tab «Evaluation».

Multiple Linear Regression in Excel

Mithilfe des «Data Analysis»-Toolpacks (siehe Installationsanleitung im nachfolgenden Abschnitt).





Verwenden Sie 'prog' (Progress) im Modell als zu trainierendes Attribut (Y).
Verwenden Sie alle anderen Attribute als Input(X).

Die optimierten Modellparameter werden in einer Tabelle ausgegeben:

	Koeffizienten
Schnittpunkt	-360.934505
age	-0.084473336
sex	-23.42348938
bmi	5.286981847
bp	1.219160671
tc	-1.563699578
ldl	1.167880117
hdl	0.902744613
tch	8.436271965
ltg	71.83262141
glu	0.48984207

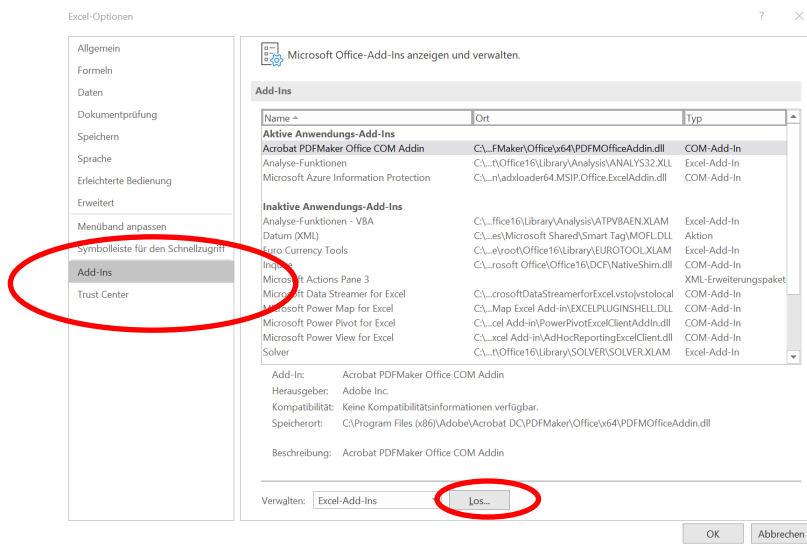
Berechnen Sie nun für die Testdaten die Vorhersage dieses zweiten Modells mithilfe der optimierten Parameter.

Berechnen Sie den Mean Squared Error, die Root Mean Square Deviation und den Mean Absolute Error auf den Testdate im Tab «Evaluation» nun auch für dieses Modell. Vergleichen Sie die Ergebnisse mit denen der einfachen linearen Regression.

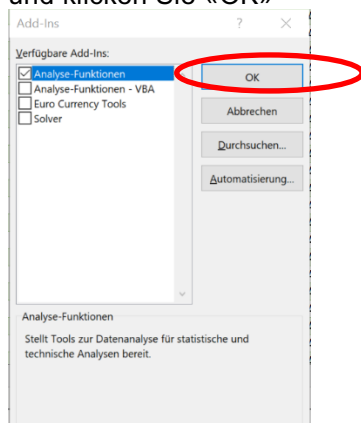
Installation des Excel Add On's «Data Analysis» (Data Analysis ToolPak)

Zur Berechnung von Kennzahlen der Univariaten Datenanalyse

- 1) Wählen Sie Datei -> Optionen -> Add-ins
und «Los...»



- 2) Wählen Sie «Analyse-Funktionen»
und klicken Sie «OK»



Das ToolPak kann nun verwendet werden über «Daten» - «Analyse» -> «Datenanalyse»

