Teil 2: Arbeiten mit einem Tabellenkalkulationsprogramm: Regression und Kurvenanpassung CBL

2.1 Einführung

2.1.1 *Füllen Sie im unten stehenden Abschnitt die Lücken, indem Sie die notwendigen Informationen aus Abb. 1 und 2 entnehmen.*

Man erkennt anhand Abb. 1, dass ein Schiff mit einer Breite von 20 Metern ungefähr eine Länge von …………… Metern haben müsste. Aus Abb. 2 kann man herleiten, dass männliche Personen im Alter von 30 Jahren im Schnitt ca.………….. Kilo wiegen.

Statt wie Sie mithilfe eines Diagramms eine **graphische Bestimmung** mit dem Auge durchzuführen kommt man zu präziseren Ergebnissen, wenn man den Wert mittels der mathematischen Funktion berechnet. Ziel der ***Regressionsanalyse*** ist es, eine mathematische Modellfunktion f(x), herauszufinden, die den Zusammenhang zwischen den beiden Größen beschreibt. Bei der Ermittlung der Modellfunktion wird diese so angepasst, dass die Abstände zu den Messpunkten insgesamt minimal wird. Streng mathematisch betrachtet, wird die Summe der quadratischen Abweichungen minimiert. Mithilfe der Modellfunktionen kann man dann innerhalb des modellierten Datenbereichs anhand der einen Größe, die andere berechnen.

Mithilfe diverser Computerprogramme oder Apps, beispielsweise dem Tabellenkalkulationsprogramm *LibreOffice CALC*, lässt sich die am besten passende Funktionsgleichung bestimmen. Man gibt dem Programm die allgemeine Form der Funktionsgleichung der Modellfunktion vor. In der Praxis ist das sehr häufig eine linearer Modellfunktion, f(x) = a·x + b, man sprich dann von einer **linearen Regression**. Das Programm ermittelt dann die am besten passenden Parameter für die Steigung (a) und den y-Achsenabschnitt (b).

Bei einer **numerischen Bestimmung** gibt man dem Programm Startwerte für die Parameter der Modellfunktion f(x) vor. Das Programm verändert jetzt ausgehend von den Startwerten die Parameter in kleinen Schritten. Es berechnet dann das Ergebnisse von f(x) und vergleicht mit den tatsächlichen Messwerten. Je kleiner diese Unterschiede ausfallen, desto besser sind die gerade benutzten Parameter. Nach vielen Schritten hat es die optimalen Parameter gefunden. Eine *nummerische Bestimmung* ermittelt durch schrittweise Variation und Ausprobieren die optimalen Parameter.

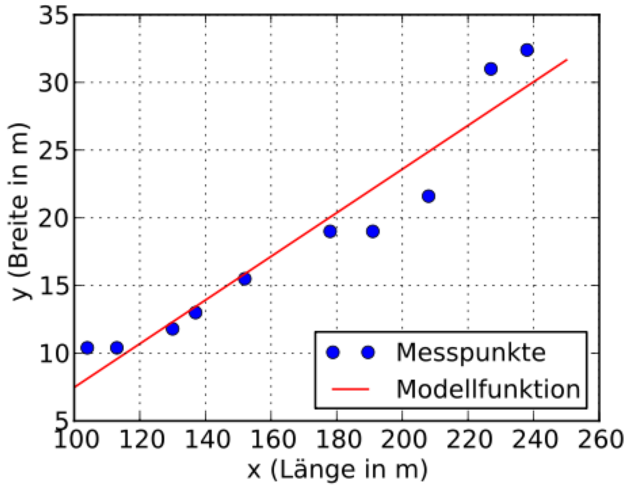


Abb. 2.1: *xy-Diagramm (Streudiagramm) von Längen und Breiten von 10 zufällig ausgewählten Schiffen mit eingezeichneter linearer Modellfunktion.* Quelle: wikipedia.de

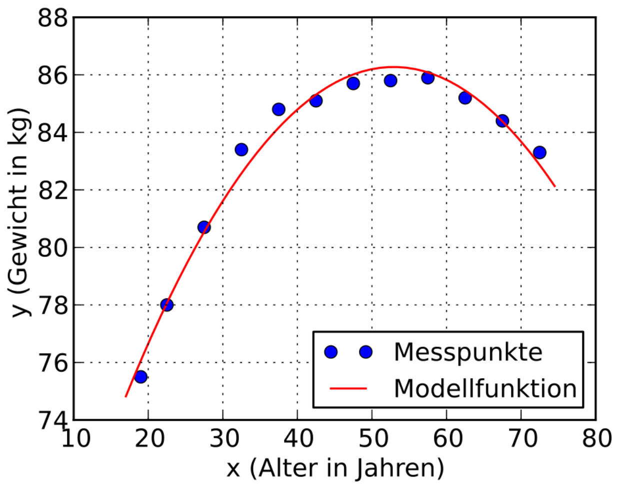


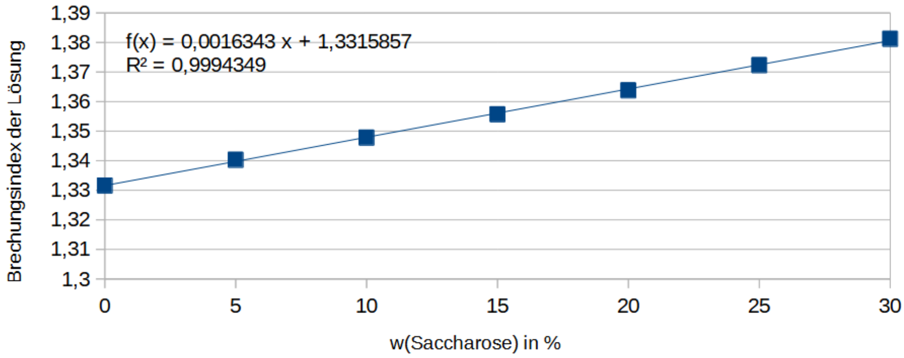
Abb. 2.2: *xy-Diagramm: Durchschnittliches Gewicht von Männern nach Alter mit eingezeichneter parabelförmiger Modellfunktion Quelle: wikipedia.de*

2.2. Lineare Regression/Kalibrierung

Die Anzeige der meisten Analyseinstrumente gibt meistens ein Signal aus, dass linear vom Gehalt abhängt.

1. Die Kalibrierung eines Refraktometers liefert beispielsweise folgende Ergebnisse:

w(Sac.) in %	**Brechungsindex** (Messwert)
0	1,3316
5	1,3403
10	1,3478
15	1,3557
20	1,3638
25	1,3723
30	1,3812



2.2.1 Berechnen Sie den Gehalt einer Probe (gerundet auf zwei Nachkommastellen), wenn der Brechungsindex 1,3511 beträgt.

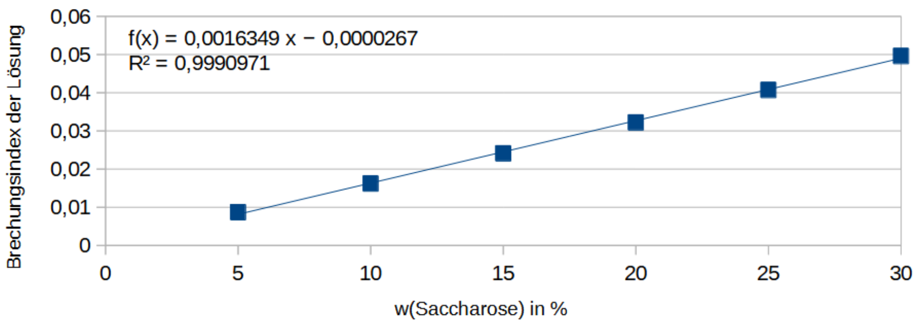
Man erkennt am Diagramm, dass der Saccharosegehalt und der Brechungsindex *keine* Proportionalität zueinander aufweisen. Verdoppelt man beispielsweise den Saccharose-gehalt von 5% auf 10%, so nimmt der Brechungsindex nicht auf das Doppelte zu, sondern nur von ca. 1,35 auf ca. 1,36. Weiterhin stellt man fest, dass reines Wasser, also 0% Saccharosegehlt. nicht das Messsignal 0 ergibt, sondern 1,3330 beträgt.

Zur einfacheren Interpretation zieht man es häufig vor, einen linearen Zusammenhang zu erhalten, bei dem ein Gehalt von Null, ein Messsignal von Null erzeugt. Dann sind Messsignal und Gehalt zueinander **proportional**: Liegt beispielsweise der doppelte Gehalt vor, so ist auch das Ergebnis ungefähr doppelt so hoch. Viertelt man den Gehalt, so liefert das Gerät ein Ergebnis, dass nur ca. ein Viertel des ursprünglichen Wertes beträgt. Dies führt auch zu einer Kalibriergerade , die keinen oder nur eine kleinen y-Achsenabschnitt besitzt. Hierfür kann man so verfahren, dass man eine Lösung mit einem Gehalt von Null **(Blindprobe)** misst und das Messergebnis **(Leerwert, Blindwert, Reagenzienleerwert)** von den anderen Mess-werten abzieht. Die meisten Messinstrumente lassen es zu, dass der Blindwert auf Tastendruck auf Null gestellt wird. So entfällt das manuelle Abziehen des Wertes von den anderen Messergebnissen.

Das **Einstellen des Reagenzienleerwertes auf Null** wird auch häufig ***„Messen gegen die Referenz“*** oder ***„Referenz auf Null stellen“*** o.ä. bezeichnet. Im Biolabor-jargon spricht man auch von ***„blanken“***, weil eine Blind-probe ohne Gehalt im Englischen *blank solution* heißt.

2.2. a) Geben Sie bei dem Beispiel von oben bleibend (Aufgabe 1), die Messwerte nach Einstellung des Leerwertes auf Null, in untenstehender Tabelle an.

w(Sac.) in %	Brechungsindex
5	
10	
15	
20	
25	
30	



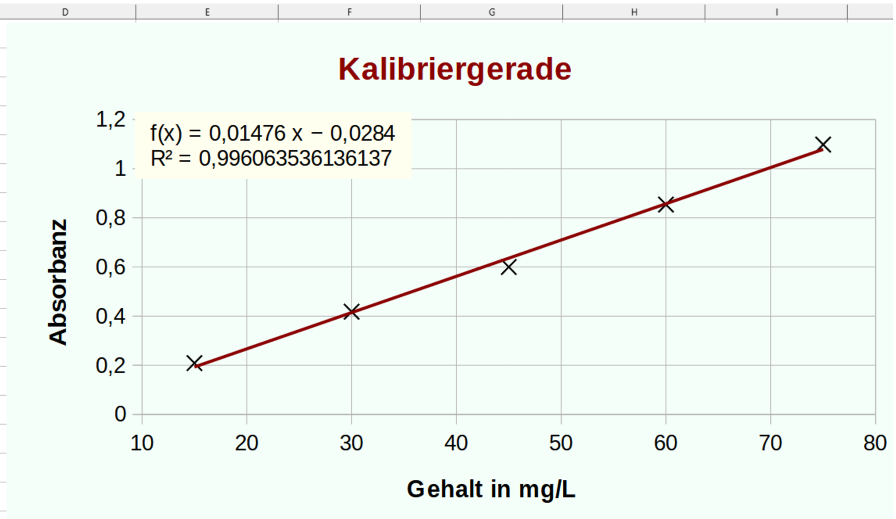
b) Die Probe zeigt jetzt ein Brechungsindex von 0,0195. Zeigen Sie, dass dasselbe Ergebnis wie oben resultiert (auf zwei Nachkommastellen genau)

2.2.1 Regressionsgerade/Kalibriergerade erstellen mit LibreOffice Calc

Lernvideo für diese Aufgabe hier: https://youtu.be/YYPCKRYQf-w

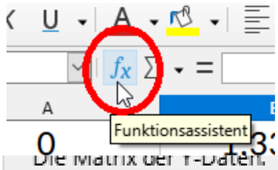
Beispiel einer Kalibrierung (incl. Regressionsgerade):

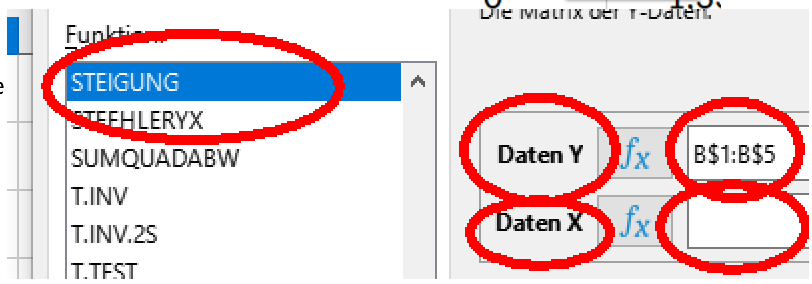
	A	B	C
1	**Bezeichnung**	**Gehalt in mg/L**	**Absorbanz**
2	Kalibr_1	15	0,209
3	Kalibr_2	30	0,418
4	Kalibr_3	45	0,6
5	Kalibr_4	60	0,854
6	Kalibr_5	75	1,098
7			
8			
9		Absorbanz der Probe	0,548
10			
11			
12			
13	Steigung	0,01476	
14	Achsenabschnitt	-0,0284	
15			
16			
17	**Gehalt der Probe in mg/L:**		**39,05**



2.2 Vorgehensweise zur Erstellung mit LibreOffice Calc

- Tabelle mit den Messpunkten. Spalte links: Gehalt (z.B. mg/L oder % o.ä.). Spalte rechts daneben: Messwert (z.B. Absorbanz, Dichte. Das Zahlenfeld mit der gedrückter Maus markieren. ⇒ Zahlenfelst ist jetzt blau unterlegt.
- Menü → Einfügen → Diagramm erstellen → Diagrammtyp: **XY-Streudiagramm** → Fertigstellen.
- Doppelklick auf Diagramm, so dass von kleinen **schwarze Miniquadraten** umrahmt: 
- **Rechts**klick auf einen der Datenpunkte → **Trendlinie einfügen. Geradengleichung** und **Bestimmtheitsmaß (R^2)** anzeigen lassen (anklicken). R^2 ist ein Maß, wie gut der lineare Zusammenhang ist. Idealwert bei absolut linearem Zusammenhang: 1,0000.
- Weitere nützliche Funktionen in **Libreoffice Calc** sind **STEIGUNG (SLOPE*)** und **ACHSENABSCHNITT (INTERCEPT*)**, mit denen man sich diese Werte der Geradengleichung in Zellen anzeigen lassen kann. So können Sie z.B. zum Weiterrechnen benutzt werden.
  - Auf Zelle gehen, in der die Steigung angezeigt werden soll. **Funktionsassistent** starten und STEIGUNG (Slope) auswählen.
  - Daten Y und Daten X durch Markieren der entsprechenden Zahlenspalte mit der Maus übernehmen lassen.
  - Auf Zelle gehen, in der der y-Achsenabshcnitt angezeigt werden sollen und mit der Funktion ACHSENABSCHNITT genauso verfahren.





Aufgabe:

a) *Erstellen Sie ein entsprechendes Kalibrierdiagramm mit Ihren Daten aus dem Praktikum (Alternative auf Lehreranweisung hin: Daten vom Beispiel oben nutzen).*

b) *Berechnen Sie den Gehalt der Probe mit dem Taschenrechner und der Geradengleichung der Näherungsfunktion.*

**Je nach installierter Sprache werden die deutschen oder die in Klammern angegeben englischen Funktionsnamen angezeigt.*

2.2.2 Regressionsgerade (Kalibriergerade) ermitteln mit Taschenrechner

Die meisten wissensch. Taschenrechner können die Geradengleichung einer Regressionsgeraden (Steigung + y-Achsenabschnitt) anhand von Wertepaaren ermitteln.

Ist Ihr Taschenrechner nicht dabei? Informieren Sie den Lehrer und suchen Sie selbst unter ***youtube.de*** mit Taschenrechner-Modell und „Regression“ als Stichworte.

Prüfen Sie für Ihr Taschenrechner-Modell die Anleitung anhand folgender Werte:

- **Beispiel für X/Y-Wertepaare: 2,5/0,858 5,0/1,753 7,5/2,591 10,0/3,515 12,5/4,510**
- **Ergebnisse:** Achsenabschnitt ≈ -0,07440 Steigung ≈ 0,36264, ⇒ y ≈ 0,36264·x – 0,07440 R^2 ≈ 0,999047

2.2.2.1 Casio, z.B. fx-85, fx-991 etc.

- Mit MODE in den STAT-Modus wechseln. Anschließend *A+BX* auswählen (Geradengleichung)
- Wertepaare eingeben. Am Ende mit der AC-Taste („All Clear“) die Dateneingabe beenden.
- Mit SHIFT 1 in in das Statistik-Menü wechseln und *Reg* für Regressionsanalyse auswählen. Im Untermenü kann man sich A (Achsenabschnitt) und B (Steigung) anzeigen lassen. Quadriert man den Wert von r, erhält man das Bestimmtheitsmaß R^2 anzeigen lassen.

2.2.2.2 Casio **ClasWiz**-Modelle

- Mit MODE oder MENU den STATISTIK-Modus aufrufen . Anschließend *A+BX* auswählen (Geradengleichung).
- Wertepaare eingeben. Am Ende mit der AC-Taste („All Clear“) die Dateneingabe beenden.
- Unter Optn „Regression“ auswählen. Werte werden angezeigt. Quadriert man den Wert von r, erhält man das Bestimmtheitsmaß R^2.

2.2.2.3: Casio-Modelle der Reihe **MS**:, z.B. fx-82MS, fx-991MS, fx-100MS, fx-350MS

- Speicher leeren/Reset (clear): „CLR“ durch SHIFT MODE auswählen. Danach mit MODE **REG**ression auswählen und im Untermenu dann LINeare Regression.
- Wertepaare eingeben: Zuerst X-Wert tippen, unmittelbar danach als Trennsymbol das Hochkomma: **,** (neben „M+“). Jetzt den dazugehörigen Y-Wert eintippen. Unmittelbar danach mit M+ das Wertepaar in den Speicher (**M**emory) befördern. Evtl. wird danach die Anzahl der vorhandenen Wertepaare im Speicher angezeigt. Für alle Wertepaare wiederholen!
- Mit SHIFT 2 in das Statistik-Menü (*S-VAR*) wechseln.Dort *A* (Achsenabschnitt) und *B* (Steigung) auswählen und mit =-Taste anzeigen lassen. Mit AC Anzeige beenden und ggf. wieder ins *S-VAR*-Menü wechseln. Quadriert man den Wert von r, erhält man das Bestimmtheitsmaß R^2.

2.2.2.4 Texas Instruments, z.B. TI-30X Plus

- Mit data erst die Werteliste eingeben. Alle x-Werte in einer Spalte (z.B. in L1) eingeben, die dazugehörigen y-Werte in der Nachbarspalte (L2). Jede Eingabe mit enter quittieren. Werte löschen mit delete möglich.
- Mit 2nd data ins Statistikmenu . *LinReg* aufrufen.
- Im sich öffnenden Menu überprüfen ob die betreffende Daten-Spalten auch vorausgewählt wurde
- FRQ-Zeile (Häufigkeit) so lassen wie sie ist. Auf *CALC* scrollen und enter drücken.

2.2.2.5 Sharp – EL-9950

- STAT drücken. Wertepaare in L1 und L2 eingeben. Dann mit *„Quit“* beenden: 2ndF + CL.
- STAT drücken und ***„REG“*** auswählen (Regressionsanalyse). *A+BX* auswählen (Geradengleichung). ENTER

2.2.2.6 Sharp - EL-531

- Mit MODE den STATISTIK-Modus aufrufen (1 drücken) . Dann *„LINE“* auswählen (linear).
- Wertepaare eingeben: X-Wert des ersten Wertepaares eingeben und STO drücken (store). Jetzt dazugehörigen y-Wert eingeben und M+ drücken. (Anzeige der erfassten Datenpunkte: „DATA SET = 1“). Alle weiteren Wertepaare genauso eingeben: Nach X-Wert STO drücken, nach dazugehörigem y-Wert M+.
- ON/C drücken („clear“). Achsenabschnitt a anzeigen: RCL (
  Steigung b anzeigen: RCL)
- r anzeigen: ALPHA ÷ = . Für das Bestimmtheitsmaß (r^2) muss dieser Wert noch quadriert werden.
- Mit MODE wieder in den normalen Modus zurückkehren.

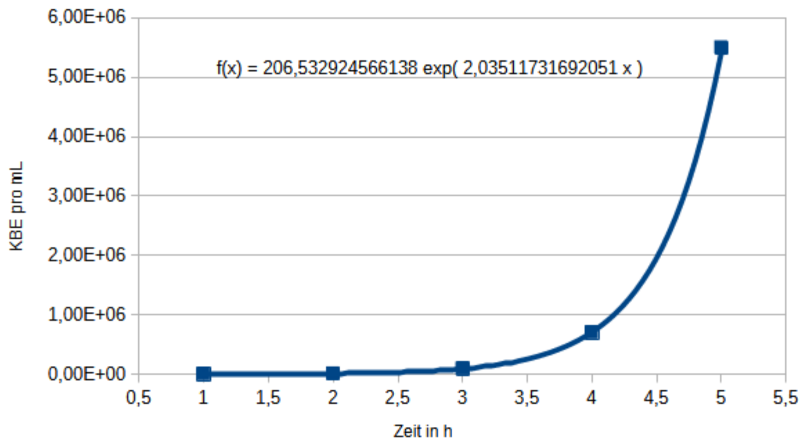
2.3 Logarithmische Diagramme und exponentielle Zusammenhänge mit dem Tabellenkalkulationsprogramm

Manchmal ist der Zusammenhang zwischen zwei Größen nicht linear. Das bekannteste Beispiel ist das exponentielle Wachstum der Bakterien auf einer Kulturplatte.

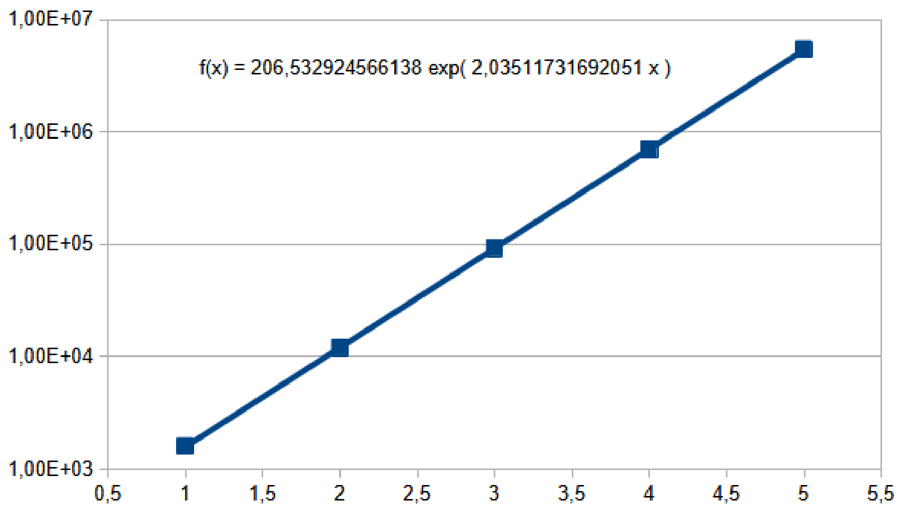
1. Übernehmen Sie in das Tabellenkalkulationsprogramm folgende Tabelle.

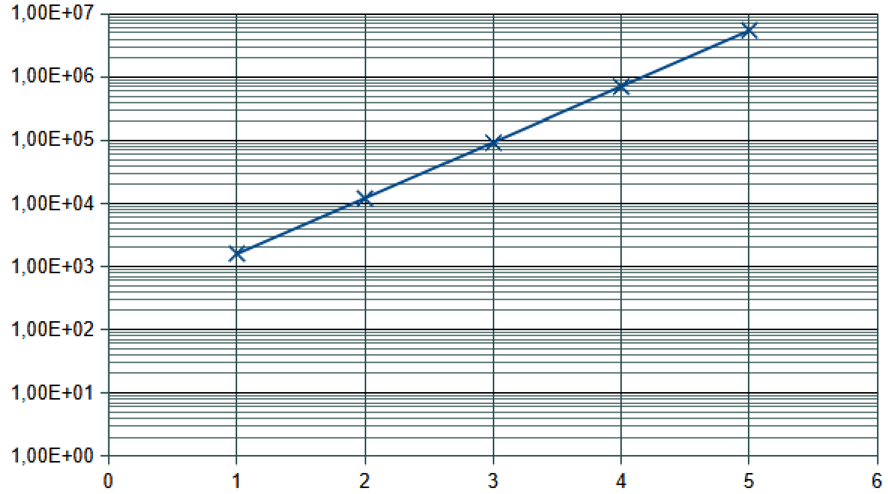
   Zeit (in h)/ KBE (pro mL): 1 / 1600 2 / 12000 3 / 92000 4 / 700000 5 / 5500000
2. Lassen Sie in der Spalte rechts daneben auch den Logarithmus der KBE berechnen.
3. Erstellen Sie zuerst Auftragung der KBE gegen die Zeit mit linearer Achsenskalierung. Lassen Sie sich die <u>exponentielle</u> Näherungsfunktion anzeigen und übernehmen Sie diese hier auf dem Papier. **Hinweis: Es gibt zwei verschiedene Schreibweisen: exp (….·x)= $e^{\dots\cdot x}$**

f(x) =



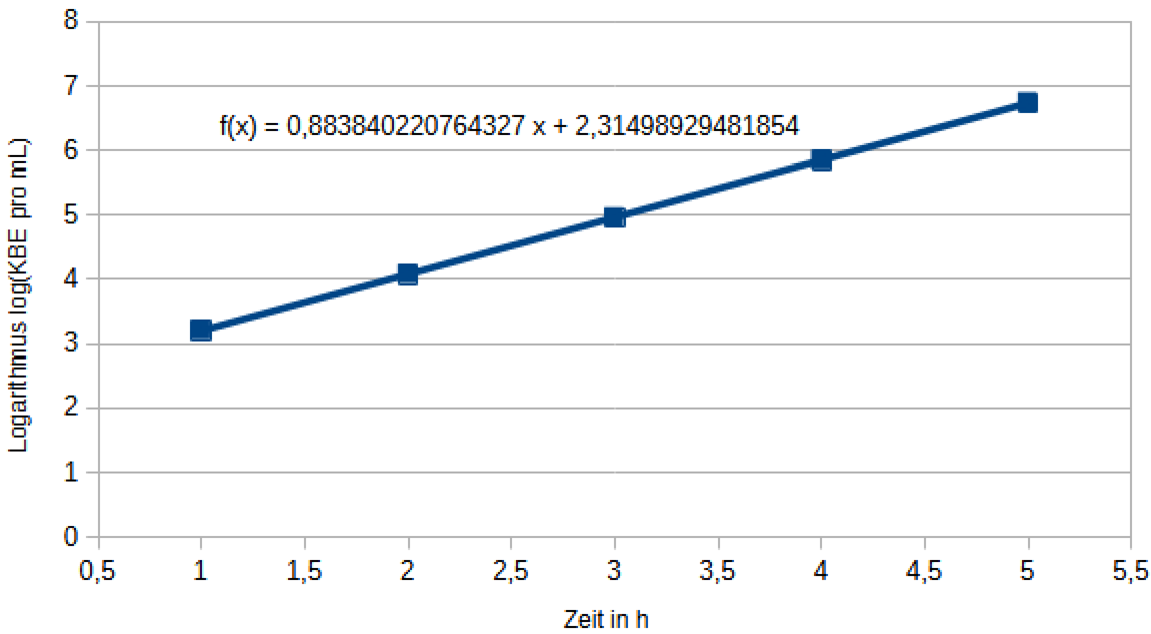
4. Duplizieren Sie das Diagramm zwei mal mit *copy-and-paste*. Formatieren Sie die y-Achse logarithmisch, indem Sie z.B. auf die entsprechende Achse doppelklicken und/oder dabei die rechte Maustaste nutzen. Erzeugen Sie <u>möglichst detailliert</u> folgende Diagramme. Achten Sie z.B. auf die Zahlenformat und wo die y-Achse jeweils beginnt. Ergänzen Sie auf den Achsen zusätzlich sinnvolle Achsentitel.





5. Tragen Sie in einem weiteren Diagramm den Logarithmus **lg(KBE)** gegen die Zeit auf und lassen sich die Näherungsgerade anzeigen. Das Diagramm soll möglichst detailliert so aussehen. Übernehmen Sie zusätzlich die Näherungsgerade hier auf das Papier.

f(x) =



6. Drucken Sie das Blatt aus, dass alle Diagramme auf einer Seite sind.

7. Berechnen Sie auf der Rückseite des Ausdrucks:

Nach welcher Zeit beträgt die Anzahl KBE pro mL ca. 2000000? Geben Sie das Ergebnis anschließend auch in Minuten an! Zeigen Sie, das mit beiden Näherungsfunktionen dasselbe Ergebnis resultiert!

2.4 Nicht-lineare Kurvenanpassung mithilfe der Option „SOLVER“

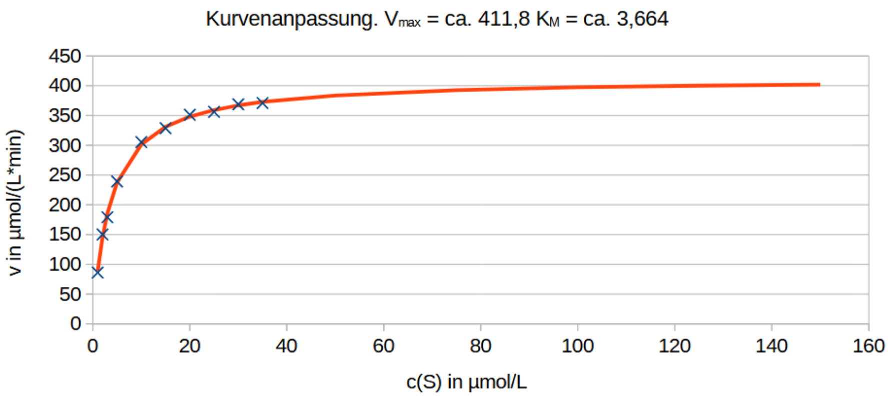
Tabellenkalkulationsprogramme können auch eine Kurvenanpassung auch mit beliebigen Funktionsformen durchführen.

Ausgangssituation und Aufgabenbeschreibung

In einem Experiment wurden einige XY-Wertepaare erhalten *(siehe rechts)*. Nun sollen die Näherungsfunktion gefunden werden, die den Zusammenhang mathematisch beschreibt. Die Funktion soll folgende Form haben:

$$y=\frac{v_{max}\cdot x}{K_M\cdot x}$$ Michaelis-Menten-Gleichung

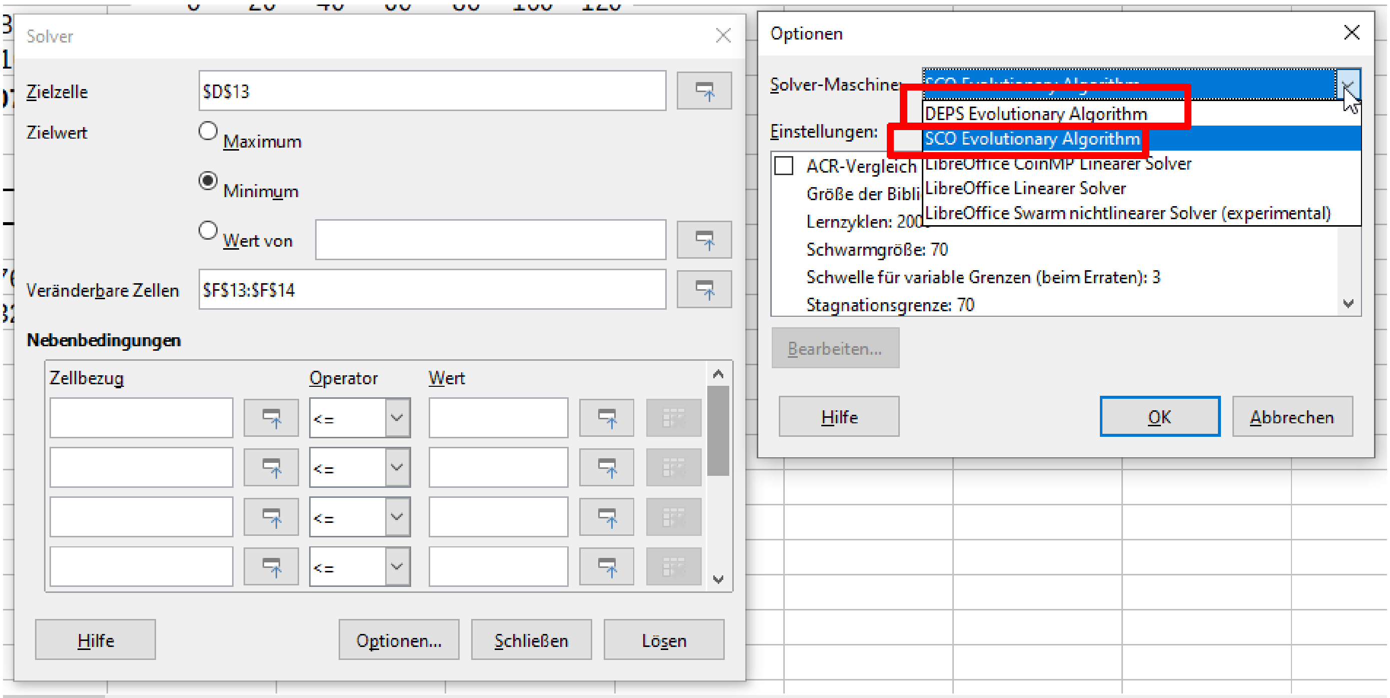
Es sollen also die am besten passenden **Parameter** der Funktion, v_{max} und K_M herausgefunden werden. Das Ergebnis soll grafisch dargestellt werden werden:



x (Substratkonzentration c(S) in µmol/L)	**y (Reaktionsgeschwindigkeit v) in µmol/(L ·min)**
1	86,13
2	150,1
3	179,1
5	239,2
10	305,3
15	328,6
20	351,1
25	356,1
30	368,7
35	370,9

DRINGENDE EMPFEHLUNG. Schritt-für-Schritt-Anleitung als Lernvideo anschauen: https://youtu.be/S2aqNYc8B_A

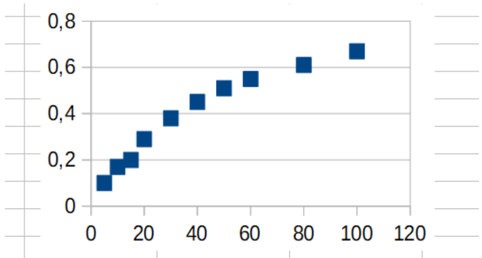
Vergewissern Sie sich zu Beginn, dass ein *Solver für nichtlineare Probleme* installiert ist. Das erkennen Sie daran, dass unter *EXTRAS → SOLVER → OPTIONEN* mehrere Solver-Maschinen zu finden sind. Zumindest einer der beiden rechts eingekästelten Einträge (*siehe Abbildung*)sollte dort zu finden sei. Ist ein solche Solver-Maschine nicht zu finden, kann sie über den Extensions-Manager (auch unter EXTRAS) nachinstallieren werden.



Durchführung (stark verkürzt). **Besser ist: Anleitung im Lernvideo Schritt-für-Schritt durchgehen.**

1. Tragen Sie zuerst die X und Y-Werte in die Tabelle ein und lassen Sie sich das xy-Diagramm anzeigen lassen.

2. Schätzen Sie die Parameter der Funktion (V_{max} und K_M) anhand des Schaubilds **<u>GANZ GROB</u>** ab. Die Werte werden später durch den *SOLVER* automatisch optimiert. Hinweis für die Michaelis-Menten-Gleichung/Substratsättigungskurve: V_{max} ist der maximale y-Wert dem sich die Kurve annähert, im Diagramm rechts z.B. 0,8. K_M ist der x-Wert, bei dem ½ v_{max} erreicht wird. Wenn v_{max} ca. 0,8 ist, dann ist ½ v_{max} ca. 0,4. Der x-Wert dort beträgt, wenn man das Lot fällt, ca. 30 *(siehe Diagramm)*. Tragen Sie die geschätzten Werte (V_{max} und K_M) in zwei beliebige Zellen ein .



Parameter der Funktion	
V_{max} =	0,8
K_M =	30

3. Die Spalte neben „Y-Werte“ beschriften Sie mit ***„v berechnet“.*** Dort lassen Sie sich die v-Werte mit den geschätzten Parametern berechnen. Die Formel entspricht der Modellfunktion, bei uns der Michaelis-Menten-Gleichung.
4. Lassen Sie sich in das gleiche Diagramm nun die dazugehörige Schätzkurve anzeigen: <u>Doppel</u>klicken Sie hierzu auf das Diagramm und klicken Sie dann mit der <u>rechten</u> Maustaste auf einen der Datenpunkte. Im erscheinenden Menü wählen Sie ganz unten ***Datenbereiche***. Gehen Sie auf das Registerblatt Datenreihen und klicken Sie aus ***Hinzufügen***.
5. Bei ***„Bereich für X-Werte“*** klicken Sie auf das Symbol für die Bereichsauswahl und markieren Sie in der Tabelle die Zellen mit den X-Werten.
6. Bei *Y-Werte* wählen Sie mit der gleichen Methode die <u>rechnerischen</u> y-Werte <u>mit Parametern</u>. aus.
7. Wenn Sie nun wieder auf das Diagramm Doppelklicken und mit der <u>rechten</u> Maustaste auf einen der neuen Punkte klicken, können Sie unter ***Datenreihe formatieren…***, die Darstellung so ändern, dass diese zweite Reihe als Kurve dargestellt wird. Hierzu bei Stil ***„Durchgängig“*** auswählen.
8. Spielen Sie jetzt etwas mit den Zahlenwerten der Parameter, um zu sehen, was für ein Einfluss das auf die Kurve des Diagramms im Vergleich zu den echten Datenpunkten hat.
9. Als Nächstes lassen wir den Abstand (Differenz) zwischen dem tatsächlichen Y-Wert und dem mit den Parametern berechneten Y-Wert berechnen. Damit das Ergebnis u.a. stets ein positives Vorzeichen hat, quadrieren wir das Ergebnis, nehmen es also *hoch 2 (^2)*. Ganz unten summieren wir diese Abstände mit der Summenfunktion (Σ) auf.
10. Die Summe der Abstandquadrate ist ein Maß wie gut unsere Näherungsfunktion ist! Spielen Sie erneut etwas mit den Zahlenwerten der Parameter, um zu sehen, was für ein Einfluss das auf diese Kennzahl hat: Je schlechter die Parameter passen, desto größer wird die Kennzahl. Je kleiner diese Kennzahl ist, desto besser ist die Näherung, da dann die Abweichungen zwischen realen und geschätzten y-Werten gering ist. Im nächsten Schritt, lassen wir LibreOffice automatisch so lange die Parameter verändert, bis dieser Zielwert minimal wird.
11. Rufen Sie den Solver auf. → EXTRAS → SOLVER auf. Wählen Sie als ZIELZELLE diejenige aus, die die Summe der quadratischen Abweichungen angibt. Als Zielwert wählen Sie „Minimum“. Als v*eränderbare Zellen* wählen Sie die Zellen aus, in denen v_{max} und K_m hinterlegt sind. Unter OPTIONEN wählen Sie eine der beiden Solver-Maschinen (DEPS oder SCO) aus, die ganz oben in diesem Dokument erwähnt sind (rote Kästen im ersten Bild). Klicken Sie auf ***Lösen*** und schauen Sie wie sich die beiden Parameterzellen verändert/optimiert haben.Klicken Sie aus ***Übernehmen***. Nun wurden die beiden Paramter K_M und V_{max} herausgefunden.