

Computerarithmetik und Rechenverfahren

5. Praktikum: Lineare Ausgleichsrechnung

5.1. Regression: Modellierung

- Analysieren Sie die Datensätze im G.R.I.P.S. zu diesem Praktikum: Stellen Sie die Daten graphisch dar, und definieren Sie ein geeignetes Modell. Alle Daten sind in allen Komponenten mit additivem normalverteiltem Fehler mit Mittelwert 0, aber unterschiedlicher Varianz gestört.
- Stellen Sie Modellgleichungen mit linearen Parametern auf.
- Berechnen Sie mit MATLAB die Lösung des linearen Ausgleichsproblems.
- Stellen Sie Messdaten und Ausgleichslösung graphisch dar, und bewerten Sie die Lösung.

Die Werte in den .mat-Files sind nicht sortiert. Sortieren von Vektoren kann MATLAB, allerdings müssen Sie yMeasure entsprechend den xMeasure mitsortieren:

```
[xS, I] = sort(xMeasure);
yS= yMeasure(I);
```

5.2. Einstellungstest 1: Regression

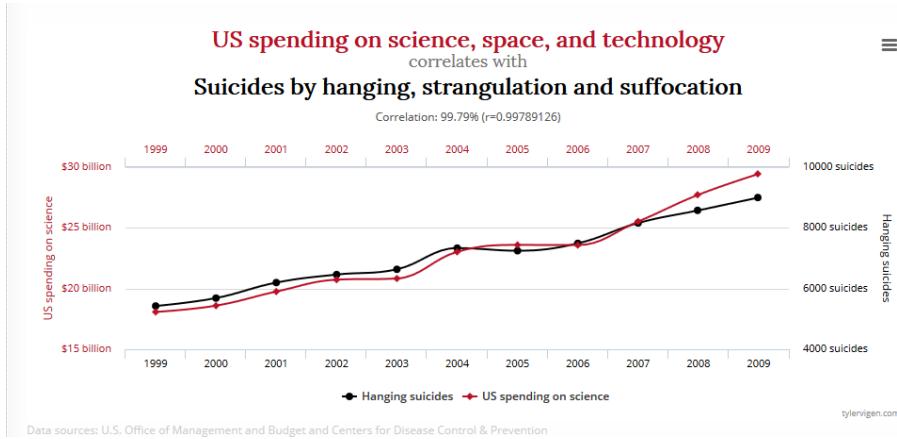
Setzen Sie die Folgen aus einem Einstellungstest fort, indem Sie ein lineares Modell $a_n = a_0 + c \cdot n$ annehmen! Bestimmen Sie a_0 und c über ein lineares Ausgleichsproblem. Errechnen Sie daraus den Wert a_7 , und vergleichen Sie mit dem Wert, den Excel über den drag-Operator liefert.

2	6	10	14	18	22
2	3	5	7	11	13
1	2	4	8	16	32

5.3. ☺Regressionsgerade: spurious correlations

Auf <http://tylervigen.com/spurious-correlations> finden sich Daten, die hohe Korrelation ohne Kausalität (?) aufweisen.

Laden Sie die Daten zum Zusammenhang zwischen Selbstmorden und Raumfahrtbudget in den Workspace. Hier sind Selbstmorde und Raumfahrtbudget als Funktion der Zeit gegeben.



Plotten Sie die Daten

- als Funktion der Zeit
- in der Form "Raumfahrtbudget = f(Selbstmorde)"
- und die zugehörige Regressionsgerade.

5.4. Normalengleichung für Regressionsgerade

- a) Berechnen Sie *ohne Rechnerunterstützung* die Ausgleichsgerade an folgende Messpunkte:

i	1	2	3	4	5
x_i	2	4	6	8	10
y_i	2	4	3	5	6

- b) Bestätigen Sie Ihre Rechnung mit MATLAB.
c) Stellen Sie die Ergebnisse graphisch dar.

5.5. Reale Daten: Kinobesucher

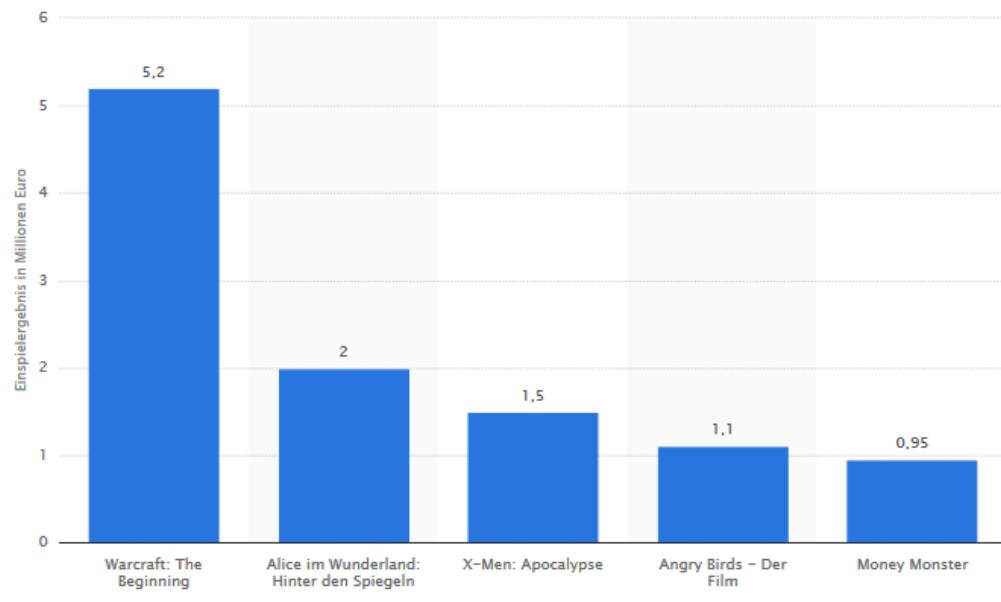
Unter <http://de.statista.com/> finden Sie aufbereitete Daten des Statistischen Bundesamtes usw., die Sie im Netz der OTH Regensburg kostenlos downloaden können.

Sie können Suchbegriffe eingeben, und Daten als Excel-Tabelle downloaden. Laden Sie z.B. von

<http://de.statista.com/statistik/daten/studie/232862/umfrage/kinofilme-nach-umsatz-am-vergangenen-wochenende-weltweit/>

Schätzen Sie eine Potenzgesetz der Form $C \cdot n^a$, dabei n die Position in der Rangliste, C, a unbekannte Parameter. Untersuchen Sie, wie start C und a von Woche zu Woche variieren.

Einspielergebnis der erfolgreichsten Kinofilme am vergangenen Wochenende (26.05. bis 29.05.2016) in Deutschland (in Millionen Euro)



© Statista 2016

Weitere Informationen:

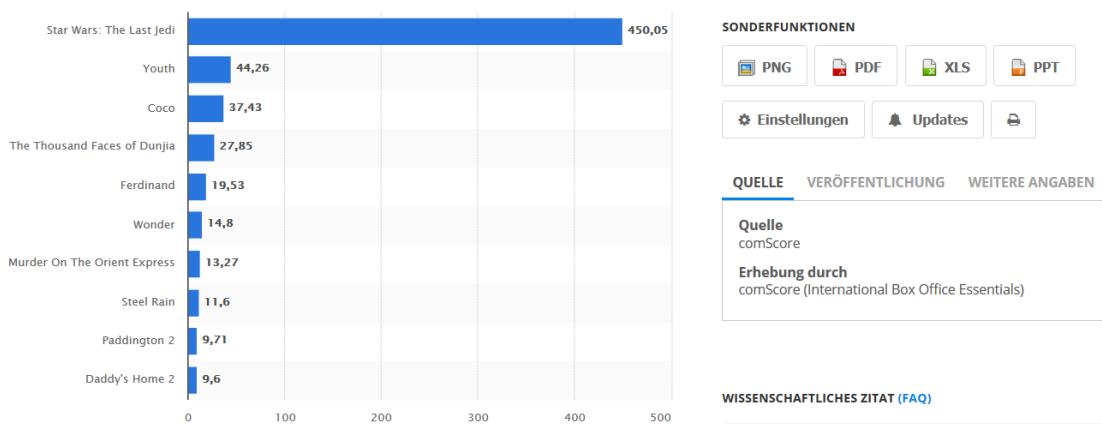
Deutschland; 26.05.2016 bis 29.05.2016

Quelle:

G+J Entertainment Media (Blickpunkt:Film)

Sorry, die Daten von 2017 taugen nicht besonders gut zu $C \cdot n^\alpha$: Star Wars ist immer eine Singularität.

Umsatzstärkste Kinofilme am vergangenen Wochenende (14.12. bis 17.12.2017) weltweit (in Millionen US-Dollar)



5.6. Reale Daten: Pendelversuch mit PhyPhox



Mit dem SmartPhone soll die Länge eines Pendels ermittelt werden.

- Legen Sie dazu das SmartPhone parallel zu den Seiten ausgerichtet auf die Sitzfläche. Setzen Sie eine Person so auf die Sitzfläche, dass das SmartPhone fest eingeklemmt ist. Starten Sie mit PhyPhox eine Aufzeichnung der Beschleunigungssensordaten ohne g . Geben Sie der schaukelnden Person einen geeigneten Anfangsimpuls. Stoppen Sie nach ca. 10 Schaukelschwingungen die Aufzeichnung.

(Alternative: Binden Sie das SmartPhone an ein Stück Paketband, vgl. das PhyPhox-Video zum Experiment "Fadenpendel". Nachteil. Das SmartPhone wird sich auch um die Längsachse drehen; dieser Effekt muss zusätzlich berücksichtigt werden.)

In jedem Fall: es ist IHR SmartPhone. Wenn Sie keinen Versuch machen wollen, verwenden Sie die Daten aus dem GRIPS.

- Finden Sie durch graphische Darstellung heraus, welche Koordinatenrichtung(en) den Hauptteil der Schwingung abbilden (Routine phyphoxEinlesenDarstellen aus früherem Praktikum). Grundsätzlich findet die Schwingung näherungsweise in einer Ebene statt, so dass Sie 2 Richtungen brauchen. Bei leichtem Schaukeln reicht aber eine Richtung aus, die andere enthält mehr Rauschen als Information.
- Bestimmen Sie den zeitlichen Anteil der Aufzeichnung, der dem leicht abklingenden Sinus der Schaukelbewegung entspricht. Hilfreiches Kommando, wenn Sie z.B. den Index ermitteln wollen, der dem Zeitpunkt 8 sec entspricht: `find(t > 8,1)`. Schneiden Sie diesen Teil aus dem Vektor heraus.

(Ja, das geht teilweise in PhyPhox auch über das Menü Zeitautomatik. Aber wir wollen es ja selber programmieren...)

- Laut Theorie, vgl. Buch von Dahmen, Reusken, Beispiel 6.1, ergibt sich eine exponentiell gedämpfte Sinusschwingung; diese lässt sich beschreiben durch ein Modell

$$y(t) = A \exp(-\delta t) \cdot \sin(\omega t + \varphi)$$

mit Amplitude A , Abklingkoeffizient δ , Frequenz ω und Phase φ . Für die Verwendung der Funktionen der MATLAB Optimization Toolbox müssen die gesuchten Parameter in einen Vektor $x = [A, \delta, \omega, \varphi]$ zusammengefasst werden, also

$$y(t) = x_1 \exp(-x_2 t) \cdot \sin(x_3 t + x_4)$$

Schreiben Sie eine Funktion, die diesen Zusammenhang darstellt.

- e) Verwenden Sie die Funktion `lsqnonlin` aus der Optimization Toolbox, um optimale Parameter im kleinsten-Quadrat-Sinn zu finden. Dazu brauchen Sie einen Startwert, den Sie aus der Graphik schätzen können. Das Argument von `lsqnonlin` stellt die in der $\|\cdot\|_2$ zu minimierende Funktion dar.
- f) Stellen Sie das optimierte Modell graphisch im Vergleich zu den Rohdaten dar.

5.7. * Korrelation \neq Kausalität, nochmal**

Erzeugen Sie zufällige Vektoren der Länge 100 und bestimmen Sie deren Korrelation.

Achtung: Zufallszahlen werden in allen Programmiersprachen als Pseudo-Zufallszahlen über einen deterministischen Algorithmus ausgehend von einem *seed* berechnet. Dieser *seed* ist bei MATLAB bei jedem Hochlauf derselbe. Wenn Sie und ein anderer Nutzer MATLAB starten, oder Sie an verschiedenen Tagen, erhalten Sie dieselben "Zufallszahlen" – um ggf. Ergebnisse reproduzieren zu können.

Erarbeiten Sie mit der Dokumentation die Bedeutung des Codes. Zufällige Daten haben "meist" Korrelation nahe 0, statistisch sind aber beliebig große Ausreißer zu ± 1 hin möglich. Schätzen Sie vorab die maximale Korrelation bei 10000 Zufallsvektoren in \mathbb{R}^4 .

```
% The default settings are the Mersenne Twister with seed 0.
rng('default');
len = 40; N = 10000;
data = rand(len,N);
C = corr(data);
% max(max(abs(C))) == 1, da Diagonaleinträge alle 1
CC = C - diag(diag(C));
[maxVal, maxInd] = max(abs(CC(:)))
[X, Y]=ind2sub(size(CC),maxInd)
CC(X,Y)
```

Der *seed* und der konkrete Pseudo-Zufallszahlengenerator können über `rng` ausgewählt werden.

5.8. * Korrelation \neq Kausalität, nochmal**

Lesen Sie in der Süddeutschen Zeitung vom 1.12.2018 den Artikel *Steile Thesen, nichts gewesen* zu statistisch signifikanten, aber nicht reproduzierbaren Ergebnissen in wissenschaftlichen psychologischen Studien.

Elektronischer Zugriff über <https://www.oth-regensburg.de/hochschule/aktuelles/einzelansicht/news/lesen-mal-ganz-anders.html>