

Woche 2

Benjamin Schlegel

1 Streudiagramm mit Regressionslinie

Material

- Video: <https://youtu.be/N8Xcc0utib0>
- Fürs Daten einlesen (und rekodieren): *R for Social Scientists*, Kapitel 3
- Für ggplot2: *R for Social Scientists*, Kapitel 4.4

Übung

1. Erstelle ein Streudiagramm mit einer linearen Regressionslinie, welches den Zusammenhang zwischen dem Ja-Anteil der 2. Welle des Schweizer Fernsehens und dem Ja-Anteil an der Urne darstellt. Verwende dazu den Datensatz **aggregat2.dta** mit den Variablen **srf2ja** und **acceptpo**. Filtere zuerst die Daten, um nur diejenigen Datenpunkte zu behalten, bei denen das SRF Umfrageergebnisse publiziert hat. Wie viele dieser Datenpunkte sind im Datensatz vorhanden?
2. Erstelle ein Streudiagramm mit einer linearen Regressionslinie, welches den Zusammenhang zwischen BIP pro Kopf und dem Demokratiegrad zeigt. Wende dazu den Datensatz **freedom_house_index.csv** mit den Variablen **GDPPC** und **Democratic_Score**.

Lernziel

- Du solltest in der Lage sein, selber eine solche Grafik zu erstellen ohne noch viel überlegen zu müssen.

2 Univariate Lineares Modell

Material

- Video: <https://youtu.be/dV-yEvR3kkU>
- *Regression Analysis: A Primer for Social Sciences*, Kapitel 1
- *R for Social Scientists*, Kapitel 6.1

Übungen

3. Welche der folgenden Aussagen sind richtig und welche falsch?

Aussage	richtig	falsch
Bei einer linearen Regression muss die abhängige Variable normalverteilt sein.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Bei einer linearen Regression muss die Ordinate kontinuierlich sein.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Ein anderer Name für die abhängige Variable ist <i>exogene Variable</i> .	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Aussage	richtig	falsch
Bei der endogenen Variable ist das Messniveau egal.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

4. Welche der folgenden Aussagen sind richtig und welche falsch?

Aussage	richtig	falsch
β_1 ist der erwartete Wert der Outcome Variable, wenn der Prädiktor den Wert 0 hat.	<input type="checkbox"/>	<input type="checkbox"/>
β_0 ist der erwartete Wert des unabhängigen Variable, wenn die abhängige Variable den Wert 0 hat.	<input type="checkbox"/>	<input type="checkbox"/>
Die Abszisse ist ein anderer Begriff für den Achsenabschnitt.	<input type="checkbox"/>	<input type="checkbox"/>
Der Intercept ist ein anderer Begriff für die Steigung.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

5. Welches der folgenden Modelle stellt ein einfaches univariates lineares Modell dar?

- $x_i = \beta_0 + \beta_1 y_i$
- $y_0 = \beta_i + \beta_i x_1 + \epsilon_i$
- $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$
- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- $y_i = \beta_1 x_i + \epsilon_i$

6. Welcher der folgenden Punkte ist nicht im Fehlerterm abgebildet?

- Paul kreuzt bei der Umfrage ausversehen die falsche Antwort an.
- Frau Müller gab bei einer Telefonbefragung an, dass sie an der Befragung nicht teilnehmen möchte. Deshalb ist sie im Datensatz nicht enthalten.
- Der Forscher Franz hat ein Modell gerechnet, welches als unabhängige Variable Anzahl Störche enthält und als abhängige Variable Anzahl Neugeborene in einem Land, dabei hat er jedoch vergessen, die grösses das Landes ins Modell einzubeziehen.
- Der 80-jährige Bruno hat an keiner Abstimmung in den letzten 10 Jahren teilgenommen, obwohl ältere Männer die höchste Beteiligungsrate haben.
- Bei einer Telefonbefragung schreibt die Callcentermitarbeiterin Petra das falsche Geburtsdatum ins Formular, weil sie wegen dem Dialekt den Befragten am anderen Ende der Leitung falsch verstanden hat.

Lernziele

- Du verstehst das univariate lineare Regressionsmodell mit seinen Komponenten:
 - abhängige Variable (AV) [Ordinate, Outcome, Regressand, endogene Variable]: kontinuierlich und normalverteilt
 - unabhängige Variable (UV) [Abszisse, Prädiktor, Regressor, exogene Variable]: keine Anforderungen
 - Achsenabschnitt [Intercept / Konstante]: Erwartungswert von y bei x=0
 - Steigungskoeffizient [Slope]: Erwartete Änderung von y, wenn x+1
 - Fehlerterm [Störterm, Residuum]: Abweichung vom Idealwert durch Zufalls im menschlichen Verhalten, Fehlerspezifikation des Modells und Messfehler

3 Schätzverfahren für eine lineare Regression: Warum verwenden wir OLS?

Material:

- Video: <https://youtu.be/6v5Fp9z72Pg>

Übungen

7. Welche der folgenden Aussagen sind richtig und welche falsch?

Aussage	richtig	falsch
Effizienz ist ein Gütekriterium für Schätzverfahren.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Effizienz bedeutet, dass bei grössere Stichproben der geschätzte Parameter eher mit dem realen Wert übereinstimmt.	<input type="checkbox"/>	<input type="checkbox"/>
Konsistenz ist eine Annahme von OLS	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Erwartungstreu bedeutet, dass der Erwartungswert mit dem wahren Wert übereinstimmt.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

8. Welche der folgenden Formeln beschreibt OLS?

- $\sum_i |\epsilon_i|$
- $\sum_i \epsilon_i^2$
- $\sum_i x_i = 0$
- $\sum_i \epsilon_i = 0$
- $\sum_i x_i^2$

9. Welches der folgenden Punkte ist KEINE Annahme von OLS?

- Linearität
- keine perfekte Multikollinearität
- **x und y korrelieren nicht**
- Homoskedastizität
- Zufallsauswahl

Lernziele

- Du kennst die drei Gütekriterien eines Schätzverfahrens
- Du verstehst, warum die Methode der kleinsten Quadrate verwendet wird und nicht der Betrag oder die Summe aller Fehlerterme = 0 genommen wird
- Du kennst die Annahmen von OLS (mehr Details werden später in diesem Kurs folgen)

4 Selber eine univariates lineare Regression schätzen mit der Methode der kleinsten Quadrate

Material

- Video: <https://youtu.be/oTKuYXzemxs>
- Video: https://youtu.be/ghE_ILJBjmM
- *Regression Analysis: A Primer for Social Sciences*, Kapitel 2 und 3

Übungen

10. Berechne β_0 und β_1 für Aufgabe 1 mit Hilfe der Formeln in R und vergleiche sie mit deiner in Aufgabe 1 erstellten Grafik. Überprüfe anschliessend deine Berechnung, in dem du das Modell in R mit der `lm()` Funktion überprüfst.

Lernziele

- Du kennst und verstehst die Formeln für β_0 und β_1 und kannst sie in R selber anwenden.
- Du kannst eine univariate lineare Regression in R mit Hilfe der Funktion `lm()` durchführen.
- Du verstehst die Formeln für die Standardabweichung und kannst sie nachschlagen.

5 Ein univariates lineares Modell schätzen und interpretieren

Material

- <https://youtu.be/uYo-XC86-IU>

Übung

11. Schau dir den summary-output des linearen Regressionsmodells aus Aufgabe 10 an. Welche der folgenden Aussagen sind richtig und welche sind falsch.

Aussage	richtig	falsch
Die 2. Welle von SRF erklärt rund 70% der Varianz des Abstimmungsentscheids.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Pro Prozentpunkt in der das Abstimmungsergebnis höher ist, ist die 2. Welle von SRF im Durschnitt um 1.15 Prozentpunkte höher.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Der Achsenabschnitt liegt bei -6.56 und es macht Sinn, ihn zu interpretieren.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Der Standardfehler von β_1 beträgt 0.072.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

12. Schätze ein Modell für die Beziehung in Aufgabe 2. Interpretiere das Modell-Ergebnis.

Lernziele

- Du kannst ein univariates lineares Modell in R schätzen und das Ergebnis (β s, s.e., R^2) interpretieren.

6 Hypothesen testen

Material

- Video: https://youtu.be/_4_msnKJaIw
- Artikel: Spektrum - Gehirn&Geist - Revolution von unten - Methodenkritik

Übungen

13. Welche der folgenden Aussagen sind richtig und welche falsch?

Aussage	richtig	falsch
Uns Ziel ist es, die Nullhypothese zu verwerfen.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Je mehr Sterne ein Ergebnis hat, desto signifikanter ist es.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Ist der p-Wert grösser als 0.05, können wir die Nullhypothese verwerfen.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Theorie ist nice to have, aber kann auch weggelassen werden, wenn wir Hypothesen prüfen wollen.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

14. Welche der folgenden Aussagen sind richtig und welche falsch?

Aussage	richtig	falsch
Wenn ein Ergebnis nicht signifikant ist, dann sollten wir einfach die Stichprobe anpassen und so lange rumspielen, bis es signifikant ist.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Wir sollten die Hypothese erst aufstellen, nach wir die Resultate gesehen haben. Sonst könnte es ja sein, dass wir sie ablehnen müssen und dann ein Nullresultat haben und es nirgends veröffentlichen können. Und: Dass wir das gemacht haben, sollten wir natürlich	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Nicht signifikante Resultate werden von den guten Journalis genauso oft publiziert, wie signifikante.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Ein gutes Journal hat einen Peer Review Prozess. Dort kontrollieren andere Wissenschaftler anonym, ob das eingereichte Paper den wissenschaftlichen Standards stand hält und geben allenfalls Verbesserungsvorschläge.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

15. Überprüfe folgende Hypothese: *Je demokratischer ein Land ist, desto höher ist sein BIP.* (Theorie: In einem freieren Land gründen Personen lieber eine Firma und tragen so zur Produktivität bei als in einer Diktatur, wo sie jederzeit enteignet werden können. Deshalb ist das BIP pro Kopf in diesen Ländern höher.)

Lernziele

- Du bist in der Lage, selber eine Hypothese aufzustellen und zu testen. Du weisst, dass dabei eine theoretische Herangehensweise fundamental ist.
- Du weisst, was eine Nullhypothese und eine alternative Hypothese ist.
- Du kennst die Probleme in der Wissenschaft wie P-Hacking, Publikationsbias und HARKING.
- Du weisst was Peer Review und Impact-Faktor bedeuten.