

Aufgaben Klausur Statistik WiSe 2014/15 1.

Termin (gesamt: 40 Punkte)

Aufgabe 1 (20 Punkte)

Aufgabe 1: variable/Prädiktor 1 = soziale Situation (x). Kodiert in: "Situation1": situation mit den ausprägungen x1: alleine in einem raum, x2 "Situation2": mit einer anderen person zusammen in einem raum, x3 "Situation3": zusammen mit mehreren personen in einem raum. es soll die (konzentrations)-leistung (metrisch, ein höherer Wert gleich höhere Leistung) gemessen werden unter den verschiedenen bedingungen und geschaut werden ob es unterschiede gibt. zusätzlich wird die variable alter erhoben um zu sehen ob diese einen einfluss auf die leistung hat. situation x1 wurde als referenzgruppe gewählt (dummy-kodierung)

j

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | 26.8447 | 9.1800 | 2.924 | 0.00518 | ** |
| bedingung2 | -15.3207 | 6.8240 | -2.245 | 0.02921 | * |
| bedingung3 | -1.9582 | 6.7090 | -0.292 | 0.77159 | |
| alter | 1.0558 | 0.1881 | 5.613 | 8.67e-07 | *** |

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

http://s4.postimg.org/e8688yk8t/modellschaetzung_aufg1.png

(hab versucht, die Tabelle so gut wie es geht nachzubasteln, passt nicht 100%-ig aber müsste gehen damit ;)

Ja, passt ungefähr. Ich glaube Pr bei alter war ~0.3

- Welches Skalenniveau hat die Variable "soziale Situation"? (1 Punkt)
- Wie ?kodiert? man die Dummyvariable? Wie interpretiert man sie dann? (~3,5 Punkte)
- Interpretieren sie die Ergebnisse des Outputs formal und inhaltlich bezüglich der Fragestellung (hier gab's 9 Punkte, das weiß ich noch genau, weil ich nicht wusste, wo die herkommen sollen :-D)
- Berechnen sie einen Wert für eine Person unter "Situation1" (gab's das noch einzeln als Aufgabe? Ich hab' nur 2 Werte gerechnet: Für Sit. 1 und Sit. 2, jeweils 30 Jahre alt)
 - Berechnen sie einen Wert für eine Person unter "Situation2" die 30 Jahre ist
 - Berechnen sie einen Wert für eine Person unter "Situation1" die 30 Jahre alt ist
- Was ist die Nullhypothese der Untersuchung oben bezüglich Alter? Wie wäre die Nullhypothese wenn man von einem negativen Einfluss von Alter auf die Konzentration ausgeht? Was würde sich im Output ändern? Welche Parameter würden sich wie verändern?

Lösungen: Aufgabe 1: (bitte verbessern, falls falsch bzw. ergänzen)

a) ordinal, da man sie in eine Reihenfolge bringen kann.

b) X_1 : sonst=0, mit einer Person=1; X_2 : sonst (hier zählen ja auch die aus den anderen Kategorien als "0")=0, mit mehreren Personen=1. man hat also 2 Dummies. man interpretiert sie NUR in Bezug zur Referenzgruppe (und die ist "allein im Raum")

c) zur Interpretation: Situation 2 ist signifikant, bedeutet dass Personen die mit einer weiteren Person zusammen in einem Raum sind schlechter Leistung erbringen als Personen die alleine in einem Raum sind. weil der Zusammenhang negativ ist. ob man alleine oder in Situation 2 ist hat also einen Effekt auf die Leistung. zu Alter: Alter hat einen positiven Einfluss auf die Leistung, bedeutet dass je älter eine Person ist, desto höher waren die Leistungen! (hab' ich auch so)

d) *was genau habt ihr bei Aufgabe 1 gerechnet? Situation 1: ist nur estimate? Situation 2: estimate+ bzw- Wert von Situation 2? und bei dem Alter 30 einfach beide Rechnungen * 30???*

- --> Regressionsgleichung aufstellen: $y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3$
(Die Betas nimmt man ja dann aus der Tabelle. Da er ja gesagt hat, dass auch die nicht-signifikanten mit rein kommen, hauen wir die also auch mit rein!)
- Person in Situation 1: Bei X_1 "0" einsetzen, bei X_2 "0" einsetzen, bei X_3 "30" einsetzen ich hab die direkt weggelassen. Mist (Muss man glaub ich auch nicht, wurde ja nicht gefordert. Ich find's nur dann übersichtlicher).
- Person in Situation 2: Das gleiche wie oben, außer dass für X_1 "1" eingesetzt wird.

e)

- $H_0 = \beta_3 = 0$ kein Einfluss von Alter auf Leistung?
- $H_0: \beta_3 \geq 0$, $H_1: \beta_3 < 0$.
- Der Output ändert sich nur darin, dass sich der p-Wert vom Alter ändert. Weil die gerichtete Hypothese sozusagen in die falsche Richtung geht, halbiert sich der p-Wert nicht einfach nur, sondern man muss Eins minus den halbierten alten p-Wert rechnen:
- $p_{\text{neu}} = 1 - (p_{\text{alt}}/2)$. Ich glaub da kam irgendwas in Richtung 0,98 raus, jedenfalls so in der Größenordnung. Entsprechend auch keinen signifikanten Effekt mehr.

Aufgabe 2 (12 Punkte)

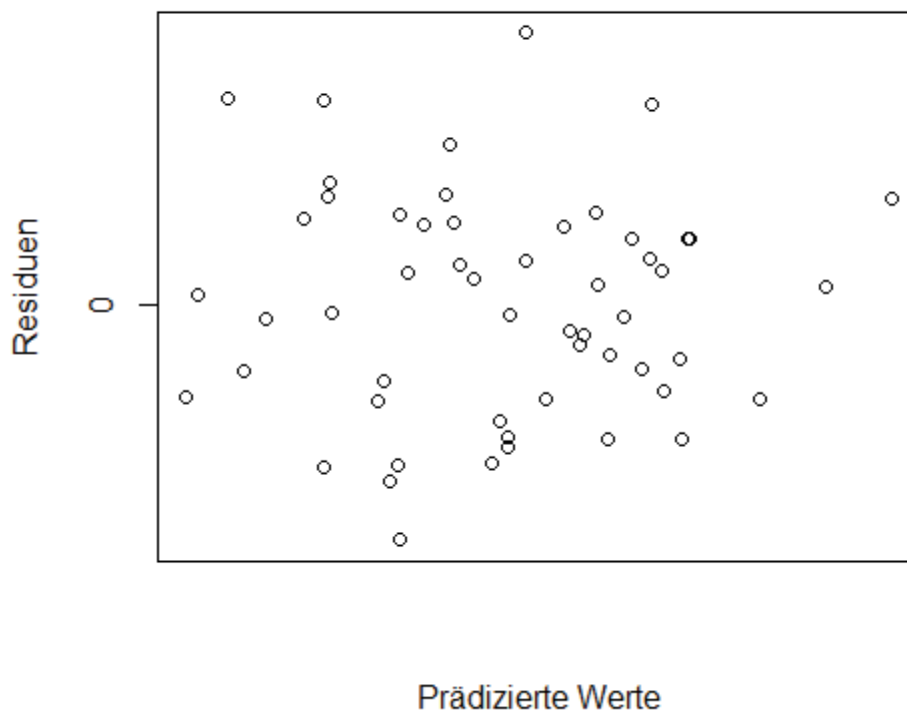
Es soll untersucht werden, ob die Ergebnisse aus Aufgabe 1 anhand einer studierenden SP repliziert werden kann. Dies ist aber nicht der Fall. Sie möchten herausfinden, warum dies so ist und lassen sich deswegen die Streudiagramme der Residuen gegenüber den prädizierten Werten ausgeben. (also x-Achse die Y-Dachs und Y-Achse die Residuen).

Vergleichen Sie die Diagramme von Analyse 1 und Analyse 2.

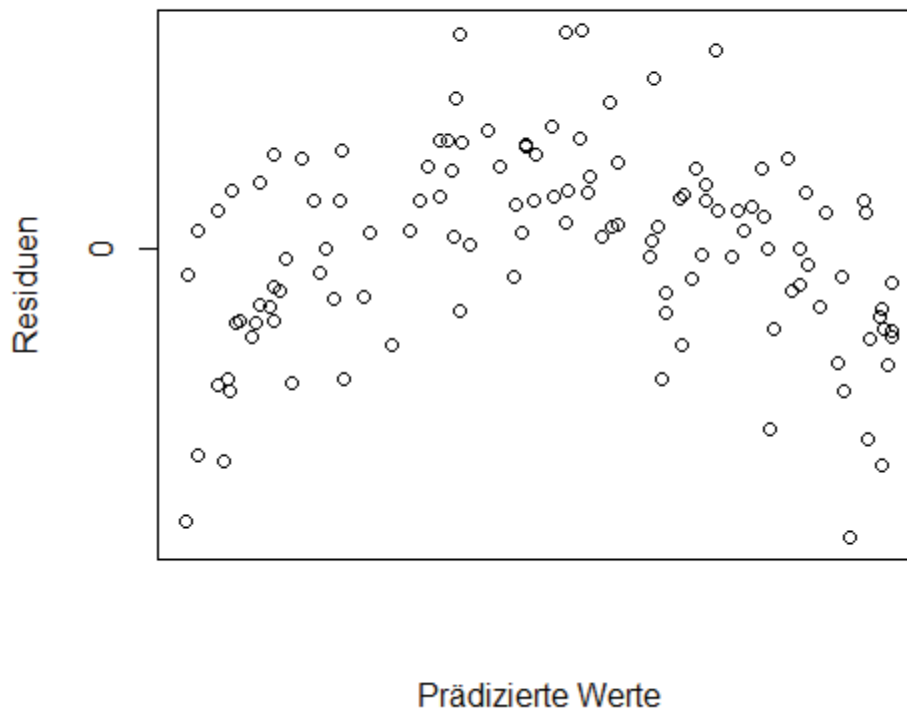
Dann kamen da 2 Streudiagramme, bei dem einen (Analyse 1) war eine unsystematische Punktewolke zu erkennen und bei dem anderen (Analyse 2) die Punkte in einer Art umgekehrter Parabel verlaufend.

Ergänzung: Bei Analyse 1 reichten die Residuen Werte von -40 bis +40.

- Bei Analyse 2 von -150 bis +100.



http://s12.postimg.org/nx2mcniv1/resid_plot_1.png



http://s12.postimg.org/dyhmq69fh/resid_plot_2.png

Aufgabe a) Sind anhand der Streudiagramme mögliche Erklärungen für die unterschiedlichen Ergebnisse der Untersuchungen erkennbar? Wenn ja, was ist der Unterschied und wie ist er erklärbar?

Aufgabe b) Was könnte eine tiefere Ursache für die unterschiedlichen Ergebnisse sein? (Bei Teilaufgabe B gab's weniger Punkte, wenn ich mich richtig erinnere war die Aufteilung 8 für (a) und 4 für (b) ???)

Lösungen Aufgabe2:

a)

- bei Streudiagramm der Analyse 2 Autokorrelation (also Korr. der Fehler) erkennbar, in Analyse 1 sind die Voraussetzungen (GMA) augenscheinlich erfüllt
- dadurch sind zwar Betas noch erwartungstreu geschätzt, aber die Standardfehler und Signifikanzen nicht, deshalb evtl. in zweiter Analyse keine (signifikante) Replikation. <- Ich würde sagen das stimmt, wenn Autokorrelation alleine auftritt. In diesem Beispiel hier ist Autokorrelation aber meiner Meinung nach eher ein "Symptom" für eine falsche Spezifikation des Modells (nicht linearen Zusammenhang nicht ins Modell aufgenommen). Dadurch sind dann schon die Betas verzerrt. <-- Aber ich dachte, das wäre nur der Fall bei einer Verletzung der GMA2, was doch hier nicht sicher festgestellt werden kann?
- "Wie ist er erklärbar?": So ne Autokorrelation kann z.B. durch Modellfehlspezifikation kommen, um das zu erkennen müsste man aber Streudiagramm von y auf x betrachten (Hier noch ergänzen!)

b)

- Hier hab' ich die verschiedenen Gründe aufgeschrieben, wie eine Autokorrelation zustande kommen kann (Modellfehlspezifikation; fehlende Variable, die einen Einfluss auf Y hat; Messwdh. [das kann ja ausgeschlossen werden]) und ein Beispiel gegeben: Modellfehlspez.: In Population evtl. nichtlinearer Zusammenhang. Fehlende Variable: Z.B. Die Vertrautheit der Personen, mit denen die Leute in einem Raum waren; Konkurrenzverhalten in den einzelnen Kleingruppen.....
- Könnte mir hier auch gut vorstellen, dass er darauf hinaus wollte, dass es zwei unterschiedliche Stichproben waren. Die erste eine Zufallsstichprobe aus deutschen Haushalten und die zweite nur Psychostudenten.. Evtl. sind die Zusammenhänge bei Psychos anders, z.B. könnte sich bei denen das Alter nicht linear, sondern kurvilinear auswirken, was zu den Grafiken passen würde.

Aufgabe 3 (8 Punkte)

a) Annahme Fehler & Kovariablen sind unabhängig und Erwartungswert der Fehler ist 0.

1.) Ein Beispiel bei dem die Unabhängigkeitsannahme nicht erfüllt ist?

2.) was folgt für die KQ-Schätzer Eigenschaften, wenn verletzt?

c) Was ist die Konsequenz wenn die Fehler untereinander korrelieren?