

Statistische Modellbildung II (WS18/9)

Übungsblatt 1

Arieta Thaqi (Matrikelnr.: 2962887)

1. Aufgabe

1.a

Man bezeichnet die Kontrolle der Drittvariablen auch als Auspartialisierung. Dabei wird bei einem sequentiellen Vorgehen, die Bereinigung einer unabhängigen Variablen um die Einflüsse der anderen Variablen vorgenommen. Je mehr Anteile auspartialisiert werden, die mit der abhängigen Variable unkorreliert sind, desto stärker kommt der eigentliche Effekt einer unabhängigen Variable zum Vorschein.

1.b

In multivariaten Modellen ist es oft der Fall, dass unabhängige Variablen in einem Modell erst dann statistische Signifikanz aufweisen, wenn weitere sinnvolle erklärende Variablen in das Modell miteinbezogen werden. Das Modell kann so an Qualität und Erklärungskraft gewinnen. Das lässt sich auch am nachfolgenden Beispiel des bivariaten Zusammenhangs zwischen der x-Variable Alter und der y-Variable Einkommen verdeutlichen. Hier besteht im Grunde kein Zusammenhang. Zwar steigt das Einkommen in gewisser Maßen mit dem Alter an. Jedoch fällt es auch wieder, wenn man beispielsweise im Rentenalter angekommen ist. Hier werden weitere Erklärungsfaktoren benötigt, wie die Bildung. Mit Einbezug der Variable Bildung wird Alter signifikant. Da mit steigenden Jahren auch das Bildungsniveau ansteigt, was letztendlich das Einkommen beeinflusst. Das wird auch in der folgenden Aufgabe ersichtlich.

2. Aufgabe

Syntax:

*Alter.

```
FREQUENCIES v84.
```

```
compute alter_neu = v84 - 18 .
```

```
formats alter_neu (f1.0).
```

```
VARIABLE LABELS alter_neu "Alter0".
```

```
FREQUENCIES alter_neu.
```

*Schulabschluss recodieren.

```
FREQUENCIES v86
```

```
/STATISTICS all.
```

```
recode v86 (1=0) (2=1) (3=2) (4=3)(5=4) (else=SYSMIS) into schulabschluss_neu.
```

```
formats schulabschluss_neu (f1.0).
```

```
VARIABLE LABELS schulabschluss_neu "Bildung".
```

```
VALUE LABELS schulabschluss_neu 0 "Kein Schulabschluss" 1 "HS" 2 "RS" 3 "FHR" 4 "Abi".
```

```
FREQUENCIES schulabschluss_neu.
```

*Geschlecht recodieren.

```
FREQUENCIES v81.
```

```
recode v81 (2=0) (1=1) (else =SYSMIS) into geschlecht_neu.
```

```
formats geschlecht_neu (f1.0).
```

```
VARIABLE LABELS geschlecht_neu "Geschlecht recodiert".
```

```
VALUE LABELS geschlecht_neu 0 "weibliche" 1 "männlich".
```

3. Aufgabe

Für y-Variable Nettoeinkommen habe ich die Variable v419 verwendet.

Modell a: Einkommen auf Alter

Beta = 0,04 Signifikanzniveau: 95%

Im bivariaten Modell besteht kein Zusammenhang zwischen den beiden Variablen.
Zusammenhang ist signifikant.

Modell b: Einkommen auf Bildung

Beta: 0,190 Signifikanzniveau: 99,9%

Im bivariaten Modell besteht ein (mäßiger) Zusammenhang zwischen den beiden Variablen. Der Zusammenhang ist hochsignifikant. Mit jedem steigenden Bildungsgrad (1 Einheit nach rechts auf der x-Achse) steigt das Einkommen um ca. 247€ an.

Modell c: Einkommen auf Geschlecht

Beta: 0,235 Signifikanzniveau: 99,9% a: 1208€ b: 734€

Im bivariaten Modell besteht ein (mäßiger) Zusammenhang zwischen den beiden Variablen. Der Zusammenhang ist hochsignifikant. Frauen verdienen im Durchschnitt 1208€. Männer verdienen im Unterschied zu Frauen durchschnittlich 734€ mehr. Hier muss bedacht werden, dass Frauen häufiger in Teilzeit arbeiten, als Männer, wodurch die Ergebnisse verzerrt sind.

Modell ab: Einkommen auf Alter und Bildung

Beta(Alter): 0,092 Signifikanzniveau: 99,9%

Beta(Bildung): 0,214 Signifikanzniveau: 99,9%

Im multivariaten Modell wird Alter durch Hinzunahme von Bildung im Vergleich zum bivariaten Modell hochsignifikant. Der Zusammenhang von Alter und Einkommen steigt auch minimal an. Auch der Einfluss der Bildung auf das Einkommen steigt an. Die Variablen Alter und Bildung korrelieren hoch miteinander. Die Variablen gewinnen beide an Erklärungskraft, weil sie sich gegenseitig bedingen. Am Beispiel des Studenten kann das sehr gut veranschaulicht werden. Ein Student ist zwar hochgebildet, verdient jedoch sehr wenig. Nimmt man die Variable Alter hinzu, lässt sich erkennen, dass das Einkommen neben dem Bildungsgrad auch vom Alter abhängig ist, da Berufserfahrung auch eine Bestimmgröße des Einkommens ist.

Modell abc: Einkommen auf Alter, Bildung und Geschlecht.

Beta(Alter): 0,095 Signifikanzniveau: 99,9%

Beta(Bildung): 0,222 Signifikanzniveau: 99,9%

Beta(Geschlecht): 0,246 Signifikanzniveau: 99,9%

Im Modell ab und abc zeigt sich, dass die UV Alter an Signifikanz gewinnt, was auf den Effekt der Auspartialisierung durch Hinzunahme der weiteren UV Bildung und Geschlecht zurückzuführen ist. In Modell abc sind alle Variablen hochsignifikant. Zudem nimmt die Stärke des Zusammenhangs zwischen Alter und Einkommen von Modell ab zu Modell abc kaum zu. Die Stärke der Korrelation der UV Bildung mit Einkommen bleibt zwischen Modell b und Modell abc in etwa gleich und steigt nur gering an. Im Vergleich aller UV in Modell abc zeigt sich, dass das Geschlecht den stärksten Einfluss auf das Einkommen hat.

Aufgabe 3b.

Modellzusammenfassung

| Modell | R | R-Quadrat | Korrigiertes R-Quadrat | Standardfehler des Schätzers |
|--------|-------------------|-----------|------------------------|------------------------------|
| 1 | ,035 ^a | ,001 | ,001 | 1564,565 |
| 2 | ,209 ^b | ,044 | ,043 | 1531,034 |
| 3 | ,319 ^c | ,102 | ,101 | 1484,222 |

a. Einflußvariablen : (Konstante), Alter0

b. Einflußvariablen : (Konstante), Alter0, Bildung

c. Einflußvariablen : (Konstante), Alter0, Bildung, Geschlecht_r

Vergleicht man die drei Modelle, dann lässt sich erkennen, dass durch die zusätzliche Berücksichtigung der beiden Variablen Bildung und Geschlecht das R^2 ansteigt. Selbst bei Betrachtung des korrigierten R^2 , ist dieser Effekt zu erkennen und basiert demnach nicht nur auf der Hinzunahme weiterer Variablen, da R^2 nicht verschlechtert werden kann. Das bivariate Modell kann nur 0,1% der Varianz „erklären“. Im Vergleich dazu, kann unter Berücksichtigung der Bildung bereits 4% der Varianz „erklärt“ werden. Im dritten Modell kann sogar ca. 10% der modellgebundene Varianz erklärt werden. Der Standardfehler sinkt unter Berücksichtigung der zusätzlichen Variablen leicht.