Musterlösung Übungsaufgabe 6

Statistische Modellbildung II

9.Dezember 2017

Aufgabe 1

Bestimmen Sie den minimalen Stichprobenumfang für eine Variablenbeziehung in der Höhe von ca. f^2=0.1. Die Variablenbeziehung soll in einem Regressionsmodell mit 20 weiteren Kontrollvariablen mit einer Power von 0.8 und einem Signifikanzniveau von 95% (bzw. Irrtumswahrscheinlichkeit 0.05) getestet werden. Stellen Sie Ihren Denk-/Rechenvorgang dar.

Die Teststärke ist eine Funktion der drei Faktoren Signifikanzniveau, geschätzte Effektstärke und Stichprobenumfang. Deswegen lässt sich der notwendige Stichprobenumfang aus einer vorab festgelegten Teststärke, einem bestimmten Signifikanzniveau und einer erwarteten Effektstärke ableiten (a-priori-Analyse).

Folgende Werte sind gegeben:

- Variablenbeziehung: $f^2 = 0.1$
- Anzahl der Kontrollvariablen: 20
- Teststärke: 80%
- Signifikanzniveau: 95% (Irrtumswahrscheinlichkeit $\alpha = 0.05$)

Gesucht:

- N (Stichprobenumfang)

Nun ist in einem ersten Schritt den Nonzentralitätsparameter λ zu berechnen. Ist dieser berechnet sind sämtliche Größen vorhanden um die umgeformte Gleichung zur Berechnung von N auflösen zu können.

Der Nonzentralitätsparameter λ ergibt sich aus einer Teststärkentabelle für die Analyse mit Alpha = 0.05, gemäß dem gewählten Signifikanzniveau. Unser **u**, also die Anzahl unabhängiger Modellvariablen, beträgt 21. Somit betrachten wir in der Tetstärkentabelle die Zeile mit u = 21 bzw. 20.

v ergibt sich aus der folgenden Gleichung:

$$v = N - u - 1$$

Da hier N bestimmt werden soll, können wir v nicht berechnen (was sich aber auch vernachlässigen lässt in diesem Fall).

In der entsprechenden Zeile der Teststärkentabelle wird nun der erste Wert gesucht, bzw. der kleinste Wert, der die geforderte Teststärke von 80%(= 0.8) erstmalig überschreitet. Ist dieser gefunden, kann die notwendige Stichprobengröße abgeleitet werden, aus der umgeformten Gleichung:

$$N = \frac{\lambda}{f^2}$$

Für den Nonzentralitätsparameter wird auf diese Weise ein Wert von 24 aus der Teststärkentabelle ermittelt (u = 20). Somit ergibt sich unter den angeführten Randbedingungen eine optimale Stichprobengröße von 240, um mit einer Wahrscheinlichkeit von 80 Prozent einen signifikanten Effekt zu entdecken.

$$N = \frac{24}{0.1} = 240$$

Tipp: siehe Urban/Mayerl 2011: 159f.

Aufgabe 2

Was ist unter einem Konfidenzintervall zu verstehen und welchen "praktischen" Nutzen hat es. Gehen Sie bei Ihrer Antwort auf die berechneten Grenzen des Konfidenzintervalls ein.

Ein Konfidenzintervall gibt den Bereich der jeweiligen Wahrscheinlichkeitsverteilung von b an, in dem b mit einer bestimmten Wahrscheinlichkeit zu finden ist. Als Verteilungsmodell wird hier nicht die Standardnormalverteilung, sondern die t-Verteilung genutzt. Um das Konfidenzintervall eines b-Koeffizienten zu ermitteln, muss zuerst der Standardfehler von b sowie die Freiheitsgrade berechnet werden. Der t-Wert kann dann aus der t-Tabelle entnommen werden. Sind alle Werte bekannt, können sie in die Gleichung $KI_{95} = b \pm t_{df} \times SE_b$ eingesetzt werden. So lässt sich sagen, dass b mit 95%iger Wahrscheinlichkeit im Verteilungsbereich zwischen den errechneten Grenzen liegt.

In Übungsaufgabe 5 lag das Konfidenzintervall für bild_tert mit 95%iger Wahrscheinlichkeit zwischen -0,451 und 1,241 (5% Irrtumswahrscheinlichkeit). Da das Konfidenzintervall die **0** einschließt, lässt sich klar sagen, dass der Effekt nicht signifikant ist.

Aufgabe 3

Welche Form von Fehlschluss wird durch ein niedriges Signifikanzniveau "begünstigt"?

Bei einem niedrigen Signifikanzniveau z.b. von 95% auf 90% ist die Wahrscheinlichkeit höher den Fehler 1. Art zu begehen und somit fälschlicherweise von einem Zusammenhang auszugehen, obwohl tatsächlich keiner besteht.

Aufgabe 4

In welchen Fällen ist es sinnvoll das Signifikanzniveau höher anzusetzen als 95%?

Mit steigendem Stichprobenumfang werden auch substanziell unbedeutende Effekte signifikant. Daher könnte es bei großen Stichproben sinnvoll sein, dass Signifikanzniveau höher anzusetzen.