

Questions & Answers Statistik II

Tehilla Ostrovsky

2023-07-10

Q1

Interpretationen von Konfidenzintervallen für MLR mit diskreten Variablen:

Das Modell:

$$Y_i = \alpha + \beta_1 \times X_i + \beta_2 \times D_i + \beta_3 \times (X_i \times D_i) + \epsilon_i$$

β_1 der erwartete Zuwachs in Y , wenn X sich um eine Einheit erhöht in der Referenzkategorie.

(a) KI für β_2 wenn $KI_{low} = -12.09$ und $KI_{upper} = 25.67$

Die plausiblen Werte für die Differenz in den erwarteten Werten von Y von Personen in der Referenzkategorie und Personen in der anderen Kategorie liegen zwischen -12.09 und 25.67 , wenn $X_i = 0$ ist.

Da dieser Wertebereich die Null enthält, gehen wir davon aus, dass sich die Mittelwerte der AV der Populationen (Referenzkategorie und anderen Kategorie) nicht signifikant unterscheiden.

(b) KI für β_2 wenn $KI_{low} = -12.09$ und $KI_{upper} = -25.67$

Die plausiblen Werte für die Differenz in den erwarteten Werten von Y von Personen in der Referenzkategorie und Personen in der anderen Kategorie liegen zwischen -12.09 und -25.67 , wenn $X_i = 0$ ist. Da das KI im negativen Bereich liegt, können wir ableiten, dass die Personen in der anderen Kategorie kleinere Werte haben im Vergleich zur Referenzkategorie.

(c) KI für β_2 wenn $KI_{low} = +12.09$ und $KI_{upper} = +25.67$

Die plausiblen Werte für die Differenz in den erwarteten Werten von Y von Personen in der Referenzkategorie und Personen in der anderen Kategorie liegen zwischen $+12.09$ und $+25.67$, wenn $X_i = 0$ ist. Da das KI im positiven Bereich liegt, können wir ableiten, dass die Personen in der anderen Kategorie höhere Werte haben im Vergleich zur Referenzkategorie.

Q2

VIF - Erklärung

$$VIF = \frac{1}{1 - r_j^2}$$

Mit VIF können wir die Frage beantworten: Um welchen Faktor ist die Varianz von β_j größer, wenn die Korrelation zwischen den Prädiktoren ungleich Null ist im Vergleich zu wenn die Korrelation Null ist.

Wenn $r^2 \rightarrow 1$ wird $VIF \sim \frac{1}{0}$

Wenn $r^2 \rightarrow 0$ wird $VIF \sim \frac{1}{1}$

Q3

Übungsblatt 10 / VL 10

Interpretation einer Stetigen variable in Regressionsmodelle mit diskreten Prädiktoren und Interaktionen

X_i ist eine Zufallsvariable und beschreibt den Wert einer zufälligen Person i auf der stetigen Prädiktorvariable (Anzahl Stunden)

Schätzwert für β_2 , den erwarteten Unterschied zwischen Personen von der Kategorie mit der Ausprägung und die aus der kategorie ohne die Ausprägung, die den gleichen Wert X_i haben.

Schätzwert für β_1 , den erwarteten Zuwachs in der AV, wenn X_i sich um eine Einheit erhöht.

$\beta_1 + \beta_3$ Steigungsparameter in der anderen Kategorie; erwarteter Zuwachs in Y , wenn X_1 sich um eine Einheit erhöht.

Q4

Wann werden η^2 vs. η_{partiell}^2 verwendet (wurde in the VL nicht besprochen und deshalb für die Prüfung weniger relevant)?

A4

In einem Design mit 2 Faktoren (A, B) interessieren wir uns ausschliesslich für den Einfluss/Effekt von Faktor A auf die AV (z.B. A = treatment Variable, B = Geschlecht), dann möchten wir wissen, wie gross der Effekt von A auf die AV ist, unter der Kontrolle des Effekts von B auf die AV. Im Beispiel, möchten wir wissen, wie gross der Treatment-Effekt ist, wenn wir berücksichtigen, dass die AV auch wegen des Geschlechts variiert (siehe, VL 6, Folie #45).

Fehler in Übungsblättern

Korrekturen in den Übungsblättern

Übungsblatt 8 - Aufgabe 2

Fehler: Identifikation von Ausreißern – Cook's Distance Plot. Hier sind **2** Versuchspersonen als Ausreißer identifiziert.

Korrektur: Identifikation von Ausreißern – Cook's Distance Plot. Hier sind **3** Versuchspersonen als Ausreißer identifiziert.

Übungsblatt 9 - Aufgabe 5

Die Berechnung von $4/n$ aus dem Plot wurde nicht korrekt berechnet.

Korrektur:

Da $n = 30$ sollte die Berechnung vom Cook's distance = $4/30$.

Weil dieser Wert im Übungsblatt nicht richtig berechnet wurde, liegt den Cut-off Wert (horizontale Linie) an der falschen Stelle.

Dont worry! In der Prüfung wird dies auf jeden Fall auf dem richtigen Platz liegen und die Interpretation wird "alle Punkte/Beobachtungen", die oberhalb dieses Cut-offs liegen, werden als Ausreisser bezeichnet.

Reminder:

Cut-off-Wert für Cooks distance.

Cook's Distanz gibt für jede Beobachtung an, wie stark sich die Vorhersagewerte aller Beobachtungen ändern würden, wenn die Beobachtung aus der Analyse ausgeschlossen wird.

- Je grösser die Cook's Distanz, desto höher der Einflusswert der Beobachtung.
- Als Cut-Off-Wert für einen Wert mit hohem Einfluss wird ein Wert von $4/n$ vorgeschlagen, wobei n die Stichprobengröße ist.

Übungsblatt 9 - Aufgabe 4

Wie würdest du in diesem Modell ein β_{kind} von 0 interpretieren?

Korrekte Interpretation:

Bei konstanter Anzahl gelesener Bücher erwartet man einen Anstieg von 0 PUNKTEN IN DER AKADEMISCHEN LEISTUNG. Das heißt, bei konstanter Anzahl gelesener Bücher hängt die AKADEMISCHE LEISTUNG nicht mehr von der ANZAHL DER KINDER PRO KLASSE ab.

Übungsblatt 10 - Aufgabe 5

Die Interpretation von den KI (Konfidenzintervalle) war nicht korrekt beschrieben.

(a) KI α

Wir gehen davon aus, dass das durchschnittliche monatliche Einkommen von Personen mit Abschluss „Gymnasium“ zwischen 1278.308 und 7471.692 Euro liegen könnte.

Da dieser Wertebereich die Null nicht enthält, gehen wir davon aus, dass in der Population einen Zusammenhang besteht zwischen Abschluss „Gymnasium“ und dem monatlichen Einkommen.

(b)

KI β_{BA} :

Die plausiblen Werte für die Differenz in den erwarteten Werten von Y von Personen in der Referenzkategorie (Gymnasium) und Personen in der anderen Kategorie (BA) liegen zwischen -3371.941 und 5688.608 , wenn $X_i = 0$ ist.

Da dieser Wertebereich die Null enthält, gehen wir davon aus, dass sich die Mittelwerte der Durchschnittlichen Monatslohn in der Populationen (Gymnasium und BA) nicht signifikant unterscheiden.

(c) KI β_{MA}

Die plausiblen Werte für die Differenz in den erwarteten Werten von Y von Personen in der Referenzkategorie (Gymnasium) und Personen in der anderen Kategorie (BA) liegen zwischen -1771.941 und 7688.608 , wenn $X_i = 0$ ist.

Da dieser Wertebereich die Null enthält, gehen wir davon aus, dass sich die Mittelwerte der Durchschnittlichen Monatslohn in der Populationen (Gymnasium und BA) nicht signifikant unterscheiden.

Übungsblatt 8, Aufgabe 5

Fehler: β_z wurde nicht quadriert.

Korrigierte Berechnung:

$$f^2 = \frac{\beta_z^2}{1 - \beta_z^2} = \frac{.45^2}{1 - .45^2} = \frac{.2025}{1 - .2025} = .254$$