

## Übung zur Vorlesung „Empirische Ökonomie 1“

# Übungsblatt 9: Binäre abhängige Variablen

Allgemeiner Hinweis: Bitte verwenden Sie für alle Regressionen „robuste“ Standardfehler. Wir verwenden den Gretl-Beispieldatensatz `data12-1` von Ramanathan. Er enthält Angaben zu einer Zufallsstichprobe von 60 Bewerbern für ein Medizinstudium in den USA, von denen nur die Hälfte angenommen wurde. Betrachten Sie diese Stichprobe als „groß“.

Folgende Variablen sind im Datensatz enthalten:

<b>Accept</b>	Dummy: = 1 wenn angenommen, = 0 sonst
<b>GPA</b>	Durchschnittsnote im Bachelor (je höher, desto besser)
<b>Age</b>	Alter [Jahre]
<b>Gender</b>	Dummy: = 1 wenn männlich, = 0 wenn weiblich

Ergebnisse des Aufnahmetests nach Teilbereichen:

<b>Bio</b>	Biologie
<b>Chem</b>	Chemie
<b>Phy</b>	Physik
<b>Red</b>	Leseverständnis
<b>Prb</b>	Problemlösung
<b>Qnt</b>	Quantitativ / Mathematik

### Aufgabe 1:

Ihre Schwester hatte am Tag des Aufnahmetests Schwierigkeiten, sich zu konzentrieren. Nun befürchtet sie, dass ein schlechtes Ergebnis im Teilbereich Bio ihre Wahrscheinlichkeit angenommen zu werden mindert.

- (a) Schätzen Sie zunächst das lineare Wahrscheinlichkeitsmodell

$$\text{Accept}_i = \beta_0 + \beta_1 \text{Bio}_i + u_i$$

und gehen Sie davon aus, dass die Annahmen A1–A3 des linearen Regressionsmodells erfüllt sind.

Speichern Sie die vorhergesagten Werte  $\widehat{\text{Accept}}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{Bio}_i$ , indem Sie im Fenster mit den Regressionsergebnissen im Menü „Speichern“ den Punkt „geschätzte Werte“ auswählen. Nennen Sie die neue Variable unter „Name der Variable“ `ydach_ols`.

Testen Sie die Nullhypothese „Das Testergebnis im Teilgebiet Bio beeinflusst nicht die Wahrscheinlichkeit zum Medizinstudium zugelassen zu werden“ zu einem Signifikanzniveau von 5%. Was ist das Ergebnis?

- (b) Welchen Wert der abhängigen Variable würden Sie aufgrund Ihrer Schätzung in Teilaufgabe (a) bei `Bio = 2` vorhersagen?

(c) Schätzen Sie nun ein Probit-Modell der Form

$$\Pr(\text{Accept}_i = 1 | \text{Bio}_i) = \Phi(\beta_0 + \beta_1 \text{Bio}_i).$$

Wählen Sie dazu im Menüpunkt „Modell“ → „Nichtlineare Modelle“ den Punkt „Probit“.

Testen Sie die Nullhypothese  $H_0 : \beta_1 = 0$  bei einem Signifikanzniveau von 5%. Was ist das Ergebnis?

Speichern Sie die vorhergesagten Wahrscheinlichkeiten

$$\Pr(\widehat{\text{Accept}}_i = 1 | \text{Bio}_i) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \text{Bio}_i)$$

als neue Variable (wie oben erklärt) unter dem Namen `ydach_probit`.

- (d) Erzeugen Sie ein Streudiagramm mit `Bio` auf der X-Achse und den beiden neu erzeugten Variablen `ydach_ols` und `ydach_probit` auf der Y-Achse. Was ist in etwa die durch das Probit-Modell vorhergesagte Aufnahmewahrscheinlichkeit bei `Bio = 2`?
- (e) Leiten Sie den geschätzten marginalen Effekt von `Bio` auf die Aufnahmewahrscheinlichkeit am Mittelwert von `Bio` in der Stichprobe her und bestimmen Sie diesen anhand des GRETL-Outputs.

## Aufgabe 2:

Betrachten Sie ein allgemeines Probit-Modell der Form

$$\Pr(Y = 1 | X_1, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$$

mit  $k \geq 2$ , wobei alle Parameter  $\beta_0, \dots, \beta_k$  ungleich 0 sind.

Leiten Sie allgemein den marginalen Effekt von  $X_1$  auf die Wahrscheinlichkeit, dass die abhängige Variable den Wert Null annimmt, her und beschreiben Sie dessen Eigenschaften. Wie sieht in diesem Fall der marginale Effekt von  $X_1$  auf die Wahrscheinlichkeit, dass die abhängige Variable den Wert Null annimmt, am Mittel aus?

### Aufgabe 3:

Schätzen Sie nun ein Probit-Modell mit der abhängigen Variablen `Accept`, bei der Sie alle anderen ursprünglich im Datensatz enthaltenen Variablen als erklärende Variablen aufnehmen.

- (a) Testen Sie jeweils einzeln (mit einem t-Test) die folgenden Nullhypothesen zum 10% Signifikanzniveau:

„Bio, Chem, Phy bzw. GPA hat keinen Einfluss auf `Accept`.“

Was ist jeweils das Ergebnis?

- (b) Testen Sie die Nullhypothese „Die Testergebnisse in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik spielen bei der Aufnahmeentscheidung keine Rolle“ zu einem Signifikanzniveau von 1%. Was ist das Ergebnis?

Hinweis: Tests gemeinsamer Hypothesen können auch z.B. nach dem Probit-Modell analog zu den  $F$ -Tests nach OLS durchgeführt werden. Achten Sie einfach wieder auf den  $p$ -Wert.

- (c) Testen Sie die Nullhypothese „Der Einfluss der Testergebnisse in den Teilbereichen Biologie und Chemie auf die Aufnahmeentscheidung ist gleich“ zu einem Signifikanzniveau von 5%. Was ist das Ergebnis?