Peter von der Lippe (November 2013)

Analyse der Panelmortalität mit der Logistischen Regression

Der folgende Text war ursprünglich konzipiert für eine Unterrichtung im Rahmen des wiss. Beirats des ZiPP (Praxispanel des Zentralinstituts für die kassenärztliche Versorgung). Ich habe hierüber in Sitzungen des wiss. Beirats im März, Juni und November in Berlin vorgetragen. Das primär im Folgenden behandelte Anwendungsbeispiel betrifft also die im Rahmen einer Wiederholungsbefragung (Panel) befragten Arztpraxen. Alles was hier gesagt wird gilt natürlich ganz genauso für andere Erhebungseinheiten (z.B. Haushalte).

1. Grundfragen einer Untersuchung der Panelmortalität (panel attrition)

Eine nicht zu unterschätzende Schwierigkeit ist die

- *Identifikation* der Fälle von echter Panelmortalität (PM) unter den ausgetretener Erhebungseinheiten (dies verlangt eine Abgrenzung gegenüber unechten Abgängen wie z.B. Abgang durch Praxisaufgabe; das ist "unecht", weil nicht mehr existierende Praxen – im Unterschied zu attriters – auch nicht mehr zur Zielgesamtheit gehören), und die
- *Differenzierung potenzieller Einflussfaktoren* für die PM nach ihrem Ausmaß (ihrer Relevanz) und danach, ob sie zufällige oder systematische Einflüsse darstellen.

Der zweite Punkt verlangt die *Bildung von Hypothesen* über mögliche Bestimmungsfaktoren der PM und setzt eine entsprechend differenzierte Befragung voraus (man kann die Einheiten [Arztpraxen] nur nach den Merkmalen differenzieren, die auch erhoben worden sind). Ist ein Einfluss (Bestimmungsfaktor) des Austritts *systematisch*, d.h. korreliert er mit Untersuchungsmerkmalen, so kann dies eine Verzerrung (bias) erzeugen, es sei denn, er ist hinsichtlich seines Ausmaßes vernachlässigbar. Es kann z.B. sein dass, wie gelegentlich vermutet, gerade solche Praxen bei der t-ten Erhebungswelle austreten, die vorher (in Welle t-1) besonders hohe Einnahmen hatten. Damit würden dann die (durchschnittlichen) Einnahmen μ_t in der t-ten Welle mit dem entsprechenden Stichprobenwert $\hat{\mu}_t = \overline{x}_t$ unterschätzt.

Das vorrangige Interesse einer Analyse der PM kann darin bestehen,

- entweder zu prüfen, ob der Verdacht einer bias berechtigt sein könnte: ist das *Ausmaß* (die Relevanz) der PM *gering* und sind die Austritte weitgehend *zufällig* (*nichtsignifikant*), so ist dies im Sinne der Zielsetzung positiv zu bewerten, weil dann ja Hypothesen, wie die oben dargestellte (Praxen mit besonders hohen Einnahmen scheiden signifikant häufiger aus) nicht zutreffend sein dürften und sich auch komplizierte weitere Verfahren erübrigen, wie z.B. Schätzwerte für die fehlenden Angaben der ausgetretenen Einheiten zu finden;[1]
- oder Fälle von *signifikant* häufigerem (wahrscheinlicherem) Austreten – also von *systematischen* Faktoren beim Austrittsverhalten – zu identifizieren, um gezielt Maßnahmen gegen PM zu ergreifen.

2. Möglichkeiten und Grenzen einer deskriptiven Analyse der Panelmortalität

Was Ausmaß (Relevanz) und Charakter (zufällig vs. systematisch) möglicher Ursachen der PM betrifft, so geben schon beschreibende Statistiken über die *Häufigkeit* des Ausscheidens[2] aus

[1] Wenn außerdem auch noch hinzukommt, dass die Struktur der Masse neu hinzugekommener Einheiten ähnlich der Masse der ausgeschiedenen Einheiten ist, kann man argumentieren, dass Panelmortalität für die Ergebnisse eines Panels kein gravierendes Problem darstellt. Das betrifft natürlich nur solche Panel, wie das ZiPP, bei denen bei jeder Erhebungswelle jeweils auch Neueintritte in das Panel möglich sind.

[2] Ich verwende hier "ausscheiden", "ausfallen", "austreten" und "abgehen" als synonym.

dem Panel differenziert *nach Merkmalen der ausgetretenen Praxen* wichtige Hinweise. Dabei wird jeder potenzielle Einflussfaktor einzeln für sich betrachtet.

Die Frage ist dabei stets, ob das Austrittsrisiko einer nach Merkmalen der Praxen gebildeten Gruppe über- oder unterdurchschnittlich ist (d.h. es sind dann bedingte [durch die Gruppenzugehörigkeit] relative Austrittshäufigkeiten zu betrachten, oder es sind – was auf das Gleiche hinausläuft – Assoziationsmaße[3] zu berechnen). Eine solche rein deskriptive Betrachtung der Bestimmungsfaktoren der PM hat jedoch deutlich ihre Grenzen, denn

- sie wird kompliziert und unhandlich, oder gar undurchführbar, wenn man *gleichzeitig mehrere Einflussfaktoren* auf die PM betrachten will (man müsste Gruppen auf der Basis von Merkmalskombinationen mit $K \geq 2$ Merkmalen bilden [z.B. drei Merkmale: ländliche hausärztliche Gemeinschaftspraxis] und miteinander vergleichen, wobei die Gruppen dann i.d.R. meist sehr klein werden), und
- die Betrachtung ist *nicht modellgestützt* im Unterschied zur Untersuchung der Austrittswahrscheinlichkeit und deren Determinanten mit dem stochastischen Modell der "logistischen Regression", in dem die (transformierte) Austrittswahrscheinlichkeit eine Funktion bestimmter "erklärender" Variablen $X_1, X_2, \ldots, X_K$ ist (die metrisch skaliert sein können oder aber auch auf geringerem Skalenniveau gemessen sein können, z.B. dichotom sein können).

Der Vorteil einer Analyse auf Basis eines stochastischen Modells[4] ist, dass nicht nur Hypothesen über Parameter des Modells geprüft werden können, sondern dass ein solches Modell auch – wegen der Zufallsvariable U in der Regressionsgleichung – im Hinblick darauf beurteilt werden kann, ob es den Daten besser oder schlechter angepasst ist. Man kann also bessere und schlechtere Modelle unterscheiden und ein Modell schrittweise verbessern. Es ist wichtig, beide Aspekte zu unterscheiden

- Hypothesen über Parameter (wie meist H_0: $\beta_k = 0$, also kein Einfluss von X_k) und
- die Güte der Anpassung des Modells insgesamt.

Es ist durchaus möglich, dass viele oder alle Bestimmungsfaktoren (im Sinne von Regressoren $X_1, X_2, \ldots, X_K$ in einer geschätzten Regressionsgleichung) "relevant" sein können (im Sinne von signifikanten Regressionskoeffizienten $\beta_1, \beta_2, \ldots, \beta_K$), die Anpassung des Modells insgesamt aber gleichwohl unbefriedigend sein kann, so dass man also evtl. noch nicht das "richtige" Modell gefunden hat und es mit einer anderen Menge (Auswahl) von Regressoren versuchen muss.[5]

Andererseits ist zu bedenken, dass bei der Schätzung der Parameter der Modelle Annahmen (i.d.R. über die Verteilungen der Störgrößen U_i)[6] getroffen werden (damit die Schätzungen bestimmten Gütekriterien genügen) die erfüllt sein können, oder auch nicht, und die somit getestet werden sollten (vgl. Abschn. 4.5). Die H_0 bedeutet bei diesen Tests i.d.R., dass die betreffende Modellannahme erfüllt ist, so dass man – anders als bei den Tests der Koeffizienten – an der Annahme von H_0, nicht wie sonst gewohnt an deren Ablehnung (= "signifikant") interessiert ist.

---

[3] Man spricht von Assoziation statt Korrelation, wenn die Variablen X und Y beide dichotom sind (Y = 1 bedeutet "Austritt" [attrition] und Y = 0 keine attrition, die Praxis ist also ein "panelist" geblieben).

[4] Man sollte bedenken, dass es auch nichtstochastische Modelle gibt, ohne die bestimmte Messungen gar nicht möglich sind (z.B. die Messung der Lebenserwartung e_0 auf der Grundlage des [nichtstochastischen] Modells der stationären Bevölkerung). Man kann ja nicht Säuglinge befragen, wie lange sie glauben, noch zu leben.

[5] Der umgekehrte Fall (nichtsignifikante Regressoren aber unbefriedigende Anpassung insgesamt) ist häufiger und ein Indiz für Kollinearität.

[6] Genau genommen sind es Verteilungen der Zufallsvariablen $U_1, U_2, \ldots, U_n$, aber weil man unabhängige "Züge" aus *identischen* Verteilungen annimmt, kann man auch von *der* Zufallsvariable U und ihrer Verteilung sprechen.

3. Warum Schätzung der Austrittswahrscheinlichkeit π mit der logistischen Regression?

Die logistische Regression [das Logit-Modell] ist erforderlich, weil die zu erklärende (abhängige) Zufalls-Variable Y (Teilnahme am Panel in Welle t) dichotom (binär) ist und die darauf bezogene Wahrscheinlichkeit $\pi = P(y = 1)$ zwischen 0 und 1 liegt. Es kann nicht einfach wie folgt eine *lineare* Schätzgleichung (linear probability model) für die Austrittswahrscheinlichkeit der *i-ten Einheit* in *Erhebungswelle t* aufgrund von Merkmalen der Einheit in t-1 geschätzt werden als $\hat{\pi}_{i,t} = \beta_0 + \beta_1 x_{1,i,t-1} + ... + \beta_k x_{K,i,t-1}$; denn:

- die rechte Seite der Gleichung (die Merkmale [Regressoren] x_1, …, x_K können dort metrisch skaliert oder kategorial sein) kann auch Werte ergeben, die größer als 1 oder kleiner als 0 sein können, aber die geschätzte Wahrscheinlichkeit $\hat{\pi}_i$ muss zwischen 0 und 1 liegen[7] und
- die Störgrößen können nicht für alle i die gleiche Varianz haben (nicht homoskedastisch sein).

Das Problem wird mit dem folgenden Trick gelöst: Man transformiert $\hat{\pi}_i$ auf der linken Seite, so dass nicht dort nicht mehr eine Wahrscheinlichkeit mit $0 \le \pi_i \le 1$ steht, sondern sog Logits λ_i (d.h. logarithmierte odds)[8], die wie die rechte Seite von Gl. 1 Werte zwischen $-\infty$ und $+\infty$ annehmen können. Die logistische Regression unterscheidet sich also von der "normalen" Regression nur dadurch, dass man sich auf der linken Seite der Transformationsfunktion (link function)

$$\pi_i \rightarrow \lambda_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln(\omega_i) \text{ bedient.}$$

4. Interpretation der Schätzergebnisse einer logistischen Regression

4.1. Interpretation der Regressionskoeffizienten

a) Regressionsgleichung und logistische Kurve

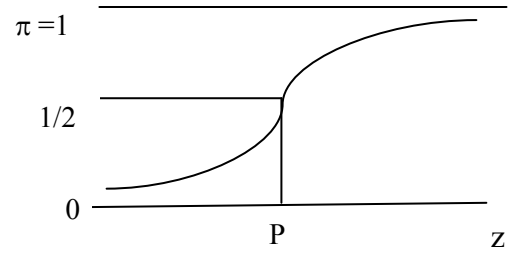
Geschätzt wird das Modell der Gl. 3 in der Übersicht1[9] also $\hat{\lambda}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... = z_i$. Aber weil auf der linken Seite Logits λ stehen haben die geschätzten Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_k$ (k = 1, …, K bei den Regressoren x_1, x_2, …, x_K) auf der rechten Seite nicht die übliche Interpretation. Für die Interpretation muss man alles quasi zurücktransformieren auf die Ebene der odds mit

(2) $$\hat{\omega}_i = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ...\right) = e^{\hat{\beta}_i} e^{\hat{\beta}_1 x_{1i}} e^{\hat{\beta}_2 x_{2i}} ...$$

bzw. auf die Ebene der Wahrscheinlichkeiten mit

(1a) $$\hat{\pi}_i = \frac{\exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ...\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ...\right)} = = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}}$$

was die bekannte S-förmige (sogmoide) "logistische Kurve" darstellt, wenn man sie mit z auf der Abszisse und $\hat{\pi}_{i,t}$ auf der Ordinate zeichnet.



Die Kurve nähert sich der Nulllinie ($\hat{\pi}_i = 0$) bei $z \rightarrow -\infty$ und der Linie $\hat{\pi}_i = 1$ bei $z \rightarrow +\infty$. Es ist leicht zu sehen, dass die Kurve einen Wendepunkt bei z = 0 (Punkt P) hat, wobei dann wegen $e^0 = 1$ die Wahrscheinlichkeit $\hat{\pi}_i = \frac{e^0}{1+e^0} = 0{,}5$ ist.

[7] Wir lassen im Folgenden zur Vereinfachung das Subskript t wegen (also π_i statt π_{it}).
[8] bei einem "odd" ω_{it} (Wettverhältnis) hat man einen Wertebereich von $0 \le \omega_{i,t} < \infty$.
[9] Siehe nächste Seite.

Übersicht 1

Abhängige Variable (**linke Seite** der Regressionsgleichung; i ist die beobachtete Einheit [Praxis])			Regressoren X (**unabhängige Variablen**) und **rechte Seite** der Regressionsgleichung;ohne Erwähnung der Störgröße)
y-Variable	Wertebereich	Bedeutung	Gleichung (Regressoren x_1, x_2, …)
(1) $\hat{\pi}_i$ (oder p_i)	$0 \le \pi \le 1$	$\pi = P(Y = 1)$ Wahrscheinlichkeit für einen Austritt (Y = 1) $1 - \pi = P(Y = 0)$ Wahrscheinlichkeit für Verbleib im Panel (Y = 0)	π **linear** (in x-Variablen) zu schätzen mit $\hat{\pi}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots$ macht aus verschiedenen Gründen keinen Sinn. Wie effektiv geschätzt wird zeigt Gl. (3). Das läuft hinaus auf (1a) $\hat{\pi}_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots)}$ $= \frac{e^{z_i}}{1 + e^{z_i}}$, eine **nichtlineare** Gleichung
(2) $\hat{\omega}_i = \frac{\hat{\pi}_i}{1 - \hat{\pi}_i}$	$0 \le \omega < +\infty$	die ω_i (i = 1, …,n) heißen **odds** (Wettverhältnis) oder **odds ratios** $\pi = 0{,}2$ heißt $\omega = 0{,}2/0{,}8$ oder: die Wetten "stehen 2 zu 8"	(2) $\hat{\omega}_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots) = e^{z_i}$ Rückrechnung zur Wahrscheinlichkeit (2a) $\hat{\pi}_i = \frac{\hat{\omega}_i}{1 + \hat{\omega}_i}$ (keine Proportionalität)*
(3) $\hat{\lambda}_i = \ln(\hat{\omega}_i)$	$-\infty < \lambda < +\infty$	die Größen λ (logarithmierte odds) heißen **Logits**	(3a) $\hat{\lambda}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots = z_i$ Rückrechnung zur Wahrscheinlichkeit (1a) $\hat{\pi}_i = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-\hat{\lambda}_i}}$ was bekannt ist als "logistische Funktion"**

* Eine Ver-ε-fachung der odds hat somit einen unterschiedlichen Effekt auf die Wahrscheinlichkeit π, je nachdem wie groß diese (π) ist. Beispiel: Bei einer Verdoppelung ($\varepsilon = 2$), so dass dann gilt $\omega_2 = 2\omega_1$, gilt für π_2 im Verhältnis zu π_1?

$\pi_1 = 0{,}2$ entspricht $\omega_1 = ¼$ Verdoppelung heißt $\omega_2 = ½$ und dem entspricht $\pi_2 = 1/3$, das 1,666 fache von 0,2
$\pi_1 = 0{,}5$ entspricht $\omega_1 = 1$ Verdoppelung heißt $\omega_2 = 2$ und dem entspricht $\pi_2 = 2/3$, das 1,333 fache von 0,5
$\pi_1 = 0{,}8$ entspricht $\omega_1 = 4$ Verdoppelung heißt $\omega_2 = 8$ und dem entspricht $\pi_2 = 8/9$, das 1,111 fache von 0,8

** oder auch als Verteilungsfunktion $F(z) = P(Z \le z)$ der logistischen (Wahrscheinlichkeits-)Verteilung er Zufallsvariable Z, für die gilt $-\infty < z < +\infty$. Nach Gl. (3a) ist $\pi = ½$ wenn $z = 0$ ist und π nähert sich asymptotisch 0 bzw. 1 wenn $z \to -\infty$ bzw. $z \to +\infty$. Die (stetige) logistische Verteilung ist somit (wie die Standardnormalverteilung) symmetrisch um z = 0. Das der **Logit** Analyse entsprechende Modell mit der Verteilungsfunktion der Standardnormalverteilung statt der logistischen Verteilung) heißt **Probit** Modell.

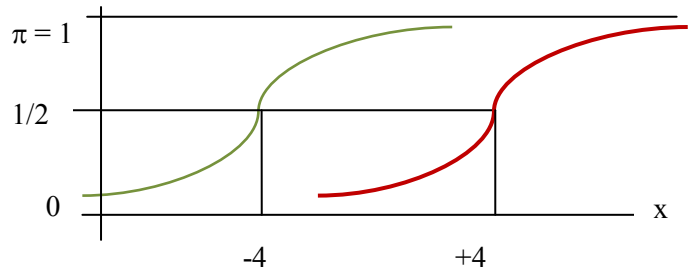
Es mag nützlich sein, für einige Werte Vergleiche zwischen Wahrscheinlichkeit (π) odd ratio (ω) und logit (λ) anzustellen:

π	0,1	0,3	0,4	0,5	0,6	0,7	0,9
ω	1/9	3/7 = 0,43	4/6 = 2/3	1	6/4 = 1,5	7/3 = 2,33	9
λ	- 2,197	- 0,847	- 0,405	0	0,405	0,847	2,197

b) Das intercept β_0 und exp(β_0)

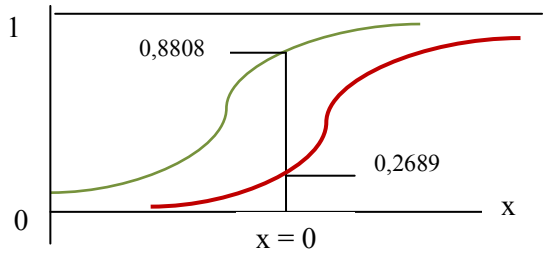
Nehmen wir zur Vereinfachung nur einen Regressor $x = x_1$ an und die logistische Funktion

$\hat{\pi}_i = \frac{1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-(\beta_0 + 5x_i)}}$ dann wirkt sich β_0 auf die Lage des Punkt P (wo $z = \beta_0 + 5x = 0$ ist) so aus:



$\beta_0 = +20 \Rightarrow$ **P liegt bei x = -4**
$\beta_0 = 0 \quad \Rightarrow$ P liegt bei x = 0
$\beta_0 = -20 \Rightarrow$ P liegt bei x = +4
Negative Werte für β_0 bedeuten, dass x entsprechend größer sein muss, um eine Wahrscheinlichkeit bestimmter Höhe zu erreichen.

Eine andere Art, sich dies anschaulich zu machen, ist zu fragen, wie groß $\hat{\pi}$ bei einer einfachen Regression (mit einem Regressor x) ist, wenn x = 0 ist, so dass z = β0 ist (wie groß β_1 ist spielt nur eine Rolle für der Steilheit der Kurve)



$\beta_0 = +2 \Rightarrow$ Höhe $e^2/(1+e^2) = 0{,}88079$
$\beta_0 = 1 \quad \Rightarrow$ Höhe $0{,}73105 = e/(1+e)$
$\beta_0 = 0 \quad \Rightarrow$ Höhe $e^0/(1+e^0) = 1/(1+1) = 0{,}5$
$\beta_0 = -1 \Rightarrow$ Höhe $e^{-1}/(1+e^{-1}) = 0{,}26894$
je kleiner β_0 und damit auch z ist, desto weiter rechts liegt die Kurve $e^{-2}/(1+e^{-2}) = 0{,}119$; $e^{-3}/(1+e^{-3}) = 0{,}047$

c) Die Größen exp (β_k) für die Regressoren x_k und die Gestalt logistische Kurve

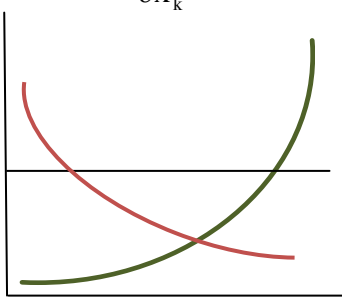
Die Größen $\hat{\beta}_k$ wirken sich auf die Steilheit der Funktion (in der Umgebung von $\pi = ½$) aus und damit auch darauf, wie geringe Veränderungen bezüglich x_k große (bei großem $\hat{\beta}_k$ und damit steiler Kurve) Veränderungen von π bewirken (oder kleine Veränderungen, wenn $\hat{\beta}_k$ klein ist). Die Idealvorstellung eines Regressors ist ja, dass eine Schwelle $\tilde{x}_k$ existiert, bei der die Wahrscheinlichkeit einen Sprung (senkrecht von 0 auf 1) macht

wenn $x_k < \tilde{x}_k$ dann $\pi = 0$ (kein Austritt aus dem Panel)
wenn $x_k \geq \tilde{x}_k$ dann $\pi = 1$ (mit Sicherheit Austritt aus dem Panel)

und es wäre dann mit x_k eine fehlerfreie Zuordnung einer Einheit zu den beiden Gruppen attriter und non-attriter möglich, das Modell wäre also "perfekt" (vgl. Abschn. 4.4: wie wahrscheinlich Fehler bei der Klassifikation sind, ist auch ein Kriterium bei Beurteilung der Güte der Anpassung).

Die Größen $\exp(\hat{\beta}_k) = e^{\hat{\beta}_k}$ (bei den k = 1, …, K Regressoren) heißen ***Effektkoeffizient*** und $e^{\hat{\beta}_k} = 1{,}5$ besagt, dass sich z.B. bei Übergang von $x_k = 0$ zu $x_k = 1$ (dichotomes Merkmal X_k) oder bei der isolierten Erhöhung von X_k um eine Einheit wegen $\frac{\partial \hat{\omega}}{\partial x_k} = e^{\hat{\beta}_k}$ ***die odds*** um 50% erhöhen, also $\hat{\omega}_1/\hat{\omega}_0 = 1{,}5$. Es gilt also:

	Effekt bzgl. $\hat{\omega}_i$
$\hat{\beta} < 0$ negativ	$\exp(\hat{\beta}) < 1$ (verringernd)
$\hat{\beta} = 0$	$\exp(\hat{\beta}) = 1$
$\hat{\beta} > 0$ positiv	$\exp(\hat{\beta}) > 1$ (vergrößernd)



Ordinate exp(βx), Abszisse x.
rot **β negativ**, grün **β positiv**

Bei einer Waagrechten, wie im Falle von $\beta_k = 0$ bewirkt Vergrößerung von x_k keine Vergrößerung von ω und damit auch keine der Wahrscheinlichkeit.

Die Steigung der Kurve $\pi_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}}$ ist $\beta_1\pi(1-\pi)$, und dies am größten bei $\pi = ½$,[10] denn bekanntlich gilt für die Varianz $V = \pi(1-\pi)$, dass $0 \le V \le ¼$ ist (bei $\pi = ½$ wirken sich also Veränderungen in x am stärksten aus auf Vergrößerung bzw. Verringerung der Wahrscheinlichkeit).

Was die Vergrößerung oder Verringerung der odds für die Veränderung der Wahrscheinlichkeit π (also π_1 in Relation zu π_0) bedeutet, hängt – wie in Fußnote * der Übersicht 1 erwähnt – ab vom Ausgangsniveau π_0.[11] Wegen $\hat{\pi}_i = \frac{\hat{\omega}_i}{1+\hat{\omega}_i}$ gilt aber bei eine Ver-ψ-fachung der odds von $\hat{\omega}_0$ zu $\hat{\omega}_1 = \psi\hat{\omega}_0$ für die Veränderung der Wahrscheinlichkeit

$$(4) \qquad \frac{\hat{\pi}_1}{\hat{\pi}_0} = \frac{\psi\hat{\omega}_0}{1+\psi\hat{\omega}_0}\frac{1+\hat{\omega}_0}{\hat{\omega}_0} = 1+\frac{\psi-1}{1+\psi\hat{\omega}_0},$$

so dass eine Vergrößerung (ψ-1 > 0) der odds auch eine Vergrößerung (aber eben nicht eine proportionale, also eine ψ-fache) der Wahrscheinlichkeit bedeutet und $\psi < 1$ auch eine entsprechende Verringerung.

Die Effektkoeffizienten stellen also Vervielfachungen (wenn $\exp(\beta_j) \neq 1$) der odds (nicht der Wahrscheinlichkeiten) dar und man könnte deshalb bei ihnen auch von ***odds ratios*** sprechen.[12]

4.2. Schätzung der Regressionskoeffizienten einer logistischen Regression

Das Modell der logistischen Regression (Logit Modell) ist nichtlinear *in den Parametern* (das ist schätztechnisch der schwierigere Fall, denn ein Modell, das nichtlinear *in den Variablen* ist, kann durch Variablensubstitution oder Variablentransformation linearisiert werden [es ist damit immer noch – wie man sagt – "intrinsisch linear"]).

Die Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_k$ werden deshalb meist mit der Maximum Likelihood (ML) Methode geschätzt, weil man ja nicht einfach mit y-Werten von 0 und 1 operieren kann. Die Grundgesamtheit ist zweipunktverteilt (= binomialverteilt mit n = 1) mit P(y=1) = π und P(y=0) 1-π. Dann ist die Likelihood L für eine Stichprobe von n = 5 mit den folgenden Einheiten 1, 1, 0, 0, 1 (n_1 = 3 mal y = 1 und $n_o = n - n_1$ = 2 mal y = 0) bei gegebenem π wegen

$\pi^1(1-\pi)^0 = \pi$	$\pi^1(1-\pi)^0$	$\pi^0(1-\pi)^1 = 1-\pi$	$\pi^0(1-\pi)^1$	$\pi^1(1-\pi)^0$	also mit **ππ(1-π)(1-π)π** oder allgemein $\pi^{n1}(1-\pi)^{n0}$
Einheit 1	Einheit 2	Einheit 3	Einheit 4	Einheit 5	
attriter	attriter	non attriter	non attriter	attriter	

zu bestimmen mit $L = \binom{n}{n_1}\pi^{n_1}(1-\pi)^{n_0}$, im konkreten Fall wäre das $\binom{5}{3}\pi^3(1-\pi)^2$. Das Maximum dieser Likelihoodfunktion wäre bei $\pi = 3/5 = 0{,}6$ gegeben. Bei $\binom{5}{3} = \binom{5}{2} = 10$ erhält man für L in Abhängigkeit von dem angenommenen π die Werte

$\pi = 0{,}5$	$\pi = 0{,}6$	$\pi = 0{,}7$
0,3125	0,3456	0,3087

[10] und am geringsten in der Nähe von $\pi = 0$ und $\pi = 1$.

[11] Eine gleich große Veränderung der odds kann bezüglich der Wahrscheinlichkeiten also etwas sehr Verschiedenes bedeuten. Das sieht man schon daran, dass die odds zwischen 0 und + ∞ liegen, die Wahrscheinlichkeiten aber zwischen 0 und 1.

[12] Der Ausdruck wird aber – wie in der Übersicht 1 erwähnt – auch für die odds selber gebraucht, weil ja auch sie *Verhältnisse* (von Wahrscheinlichkeiten) sind.

Der Wert $\hat{\pi}^{ML} = 0{,}6 = n_1/n$ ist dann der Maximum Likelihood Schätzer für π (Bei der Likelihoodfunktion lässt man die $\binom{n}{n_1}$, was hier 10 beträgt, weg und der Funktion $L_0 = \pi^{n_1}(1-\pi)^{n_0}$ werden wir gleich wieder begegnen.

Wenn man nun berücksichtigt, dass π abhängt von ("erklärt" wird durch) z, was wiederum eine lineare Funktion der Regressoren $x_1, x_2, \ldots$ ist, dann ergibt sich für

m Produkt Dabei gilt für π und 1- π Einheiten i in Abhängigkeit von den Merkmalswerten (bei den x-Variablen dieser Einheit): $\pi_i = \frac{e^{z_i}}{1+e^{z_i}} = \frac{1}{1+e^{-z_i}}$ und $1-\pi_i = 1-\frac{1}{1+e^{-z_i}} = \frac{e^{-z_i}}{1+e^{z_i}}$ und die Likelihoodfunktion wäre bei der obigen Stichprobe von 5 Einheiten (jede Einheit i hat ja ihr eigenes z_i weil sich ja die Werte der Regressoren hier jeweils unterscheiden):

(5) $$L_1 = \left(\frac{1}{1+e^{-z_1}}\right)\left(\frac{1}{1+e^{-z_2}}\right)\left(\frac{e^{-z_3}}{1+e^{-z_3}}\right)\left(\frac{e^{-z_4}}{1+e^{-z_4}}\right)\left(\frac{1}{1+e^{-z_5}}\right) \text{ statt } L_0 = \pi^{n_1}(1-\pi)^{n_0}.$$

Dieses L_1 ist die Likelihoodfunktion für das zu schätzende ("saturierte", nicht restringierte) Modell und das ist zu unterscheiden von der Likelihoodfunktion

- L_0 für das restringierte Modell, d.h. bei H_0: $\beta_1 = \beta_2 = \ldots = \beta_K = 0$ wobei π ohne irgendwelche Regressoren geschätzt wird, d.h. als Konstante β_0, denn bei Geltung von H_0 ist ja $\hat{\pi} = \beta_0$ und allein aufgrund der relativen Häufigkeiten, mit denen man aus dem Panel ausscheidet (n_1/n) bzw. nicht ausscheidet (n_0/n) bestimmt. Man erhält dann

(5a) $$L_0 = \left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0} \text{ und } LL_0 = n_1 \ln\left(\frac{n_1}{n}\right) + n_0 \ln\left(\frac{n_0}{n}\right)$$ für die logarithmierte restringierte Likelihoodfunktion,[13] und

- $L_{max} = 1$ für das perfekt angepasste Modell, bei dem $\hat{\pi}_i = 1$ ist wenn die Einheit i zu y = 1 gehört (attriter ist) und $\hat{\pi}_i = 0$ ist, wenn i non-attriter ist. Die Interpretation dieses Falls ist jedoch schwierig, denn, wie die Gl. 5 für L_1 zeigt, muss dann für alle n_1 attriters (i = 1, …, n_1) einheitlich gelten $\frac{1}{1+e^{-z_i}} = 1$, was verlangt $e^{-zi} \to \infty$ für alle n_1 attriters (wenn alle attriters richtig eingeordnet werden, dann automatisch auch alle n_0 non-attriters).[14]

LL_0 ist ohne jede Schätzung einer Regressionsfunktion aus den Daten leicht zu errechnen, wenn bekannt ist, wie viele Einheiten jeweils in die beiden Gruppen fallen, wie groß also n_1 und $n_0 = n - n_1$ ist. Lmax ist stets 1 (wie das konkret bei einem Modell erreicht werden soll ist aber offen). Erst bei L_1 und LL_1 ist eine Regressionsfunktion zu schätzen und dabei kommt es auf die Merkmalswerte $x_{1i}, x_{2i}, \ldots, x_{Ki}$ der Einheiten i = 1, …, n an

Da ein restringiertes Modell nie besser angepasst sein kann als ein nichtrestringiertes Modell und L eine Wahrscheinlichkeit ($0 \le L \le 1$) bezeichnet, muss gelten $L_0 \le L_1$ (und damit ist dann auch $-LL_0 \ge -LL_1$). Darauf beruhen einige Maße der goodness of fit.

Wenn die logarithmierte Likelihoodfunktion LL_1 bezüglich der β-Koeffizienten maximiert wird entsteht ein nichtlineares Gleichungssystem, das iterativ (bis sich LL nicht mehr vergrößern lässt) gelöst wird (Newton-Raphson-Algorithmus). Die so gewonnene ML- Schätzung ist konsis-

[13] Bei n = 10 Einheiten und davon 2 attriters erhält man $LL_0 = 2 \cdot \ln(0{,}2) + 8 \cdot \ln(0{,}8) = -3{,}219 - (-1{,}785) = -5{,}004$. Auch L_0 ist nach der obigen Formel für diese Zahlen leicht zu rechnen $0{,}2^2 0{,}8^8 = 0{,}00671$ (wie eine Wahrscheinlichkeit liegt dies zwischen 0 und 1). Man sieht auch leicht, dass in der Tat $\ln(0{,}00671) = -5{,}004$ ist.

[14] Das Problem mit einer Likelihood von 1 ist, dass sich π nur asymptotisch (bei $z \to \infty$) dem Wert 1 nähert.

tent, effizient, aber nur asymptotisch erwartungstreu und asymptotisch normalverteilt (d.h. sie erlaubt einen t-Test)

4.3. Signifikanz der Regressionskoeffizienten

Natürlich kann man auch hier die Koeffizienten auf "Signifikanz" testen, also z.B. die Hypothese H_0: $\beta_k = 0$ und damit $e^{\beta_k} = 1$ (kein Einfluss des Regressors x_k) prüfen. Ist z.B. der Koeffizient $\beta_1 = 0$ (nichtsignifikant), so bewirkt $x_1 \neq 0$ für die Variable x_1 wegen $\exp(\beta_1 x_{1i}) = e^0 = 1$ und

$$(1a) \qquad \hat{\pi}_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ...)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ...)} = \frac{\exp(\beta_0 + \beta_2 x_{2i} + ...)}{1 + \exp(\beta_0 + \beta_2 x_{2i} + ...)}$$

keine Veränderung der Wahrscheinlichkeit des Austritts der Einheit i aus dem Panel, egal wie groß x_{ki} ist. Die Variable x_k ist insofern "irrelevant" für die Erklärung der PM.

Das Testen der Koeffizienten $\hat{\beta}_k$ (also der Erklärungsbeiträge der "predictors", genauer gesagt der "changes in the *logits*", nicht wie $\exp(\hat{\beta}_k)$ changes in the odds) geschieht mit dem ***Wald Test***. Die Prüfgröße $w = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})} = \frac{\hat{\beta}^2}{\hat{\sigma}^2_{\hat{\beta}}}$ entspricht t^2 beim bekannten t-Test in der "normalen" Regressionsanalyse und sie ist χ^2 verteilt mit K (Anzahl der Regressoren, also 1 bei einfacher Regression) Freiheitsgraden. Unter "Sig." wird in den entsprechenden Programmen (etwa SPSS) der p-Wert (= prob-value, oder deutsch "Überschreitungswahrscheinlichkeit") angeben. Wenn der p-Wert kleiner ist als das Signifikanzniveau α (z.B. $p = 0{,}03$ gegenüber $\alpha = 0{,}05$) ist, dann ist der Regressionskoeffizient "signifikant" auf dem Niveau α (also auf dem 5% Niveau).

Auf W baut auch das Bayesian Information Criterion BIC auf, es gilt BIC = W – ln(n) wobei n die Anzahl der Befragten ist. Der (positive) Einfluss des Regressors x_k gilt als

schwach bei $0 < BIC < 2$, stark bei $6 < BIC < 10$ und sehr stark bei $BIC > 10$.

4.4. Maße der Güte der Anpassung eines Modells (einer Regressionsfunktion)

Die "logistische Regression" ist von der Fragestellung her, nicht aber hinsichtlich der Voraussetzungen und der Entscheidungsregel mit der Diskriminanzanalyse verwandt.[15] Bei der einfachen Diskriminanzanalyse gilt es, solche Gewichte β_0, β_1, …, β_K einer Linearkombination der x-Werte $\tilde{x}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_K x_{Ki}$ zu finden, dergestalt, dass eine Einheit i aufgrund ihres Werts $\tilde{x}_i$ mit möglichst wenig Fehlzuweisungen einer von zwei vorgegebenen Klassen zugeordnet wird.

Klassifizierungstabelle

"predicted" = Zuordnung nach dem cut-value 0,5, also
$\hat{\pi}_i \leq 0{,}5 \Rightarrow$ non-attriter (y = 0), $\hat{\pi}_i > 0{,}5 \Rightarrow$ attriter (y = 1),

observed (tatsächlich)	predicted	
	y = 1	y = 0
y = 1	**TP** true positive	**FN** false negative
y = 0	**FP** false positive	**TN** true negative

T = TP+TN = number of correct predictions (grüne Felder) F = FP + FN
n = T + F = number of observations
count R^2= T/n (oder % satz der Richtigen)
liegt eher im gewohnten Bereich von R^2 während Mc Fadden's Pseudo R^2 viel kleiner ist

[15] Dort ist die Funktion mit der eine Zuordnung zu einer der beiden vorgegebenen Klassen erfolgt eine *Linear*kombination der x-Variablen, so das Modell mit dem oben kritisierten linear probability model verwandt ist

SPSS liefert (bei einem konkreten Problem mit Antwortausfällen [panel attrition] die folgende Tabelle

Klassifizierungstabelle				
		Vorhergesagt		
		Antwortausfaelle		Prozentsatz der Richtigen
Beobachtet		nein,	ja,	
	nein,	2283	30	98,7
	ja,	952	42	4,2
	Gesamtprozentsatz			70,3

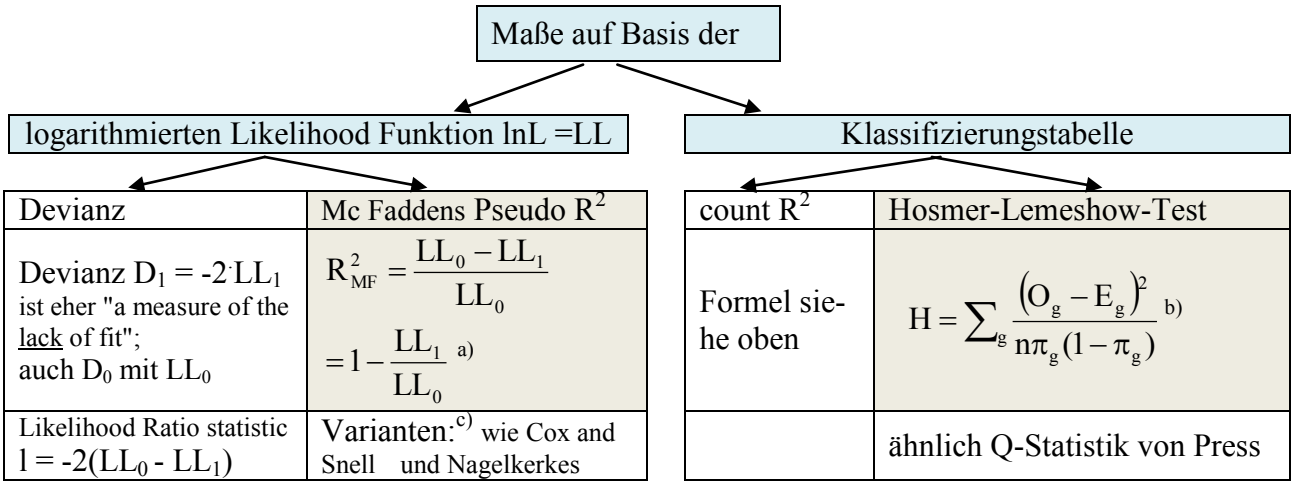
dem entspricht in unserer Notation eine Tabelle

	bleiben	ausgetreten	Σ	% Satz
y = 0 (geblieben*)	TN 2283	FP 30	2313	0,987 = TN/Σ
y= 1 (ausgetreten**)	FN 952	TP 42	994	0,042 = TP/Σ

* im Panel geblieben ** aus dem Panel ausgetreten

Man kann mit diesen Zahlenangaben leicht den count R^2 ausrechnen (was auch im SPSS Programm geschieht (siehe rot markierte Selle). Er ist hier (2283+42)/(2313+994) = 2325/3307 = 0,70305, was - wie oben gesagt - von der Größenordnung her eher im gewohnten Bereich von R^2 liegt.

Statt einer Klassifizierungstabelle gibt es auch (siehe unten Hosmer-Lemeshow-Test) die Bildung von G Gruppen und Vergleich der Anzahl der in der Gruppe g = 1,…,G beobachteten Fälle O_g mit der Anzahl der in der Gruppe erwarteten Fälle E_g. [16]



a) "Pseudo" R^2 weil das Maß nicht die übliche Interpretation einer proportional reduction of error (PRE) besitzt und ihm keine Varianzzerlegung zugrundeliegt.

b) H_0: beinhaltet hier keine (nur zufällige) Differenzen zwischen geschätzten und beobachteten Größen; die Prüfgröße H ist χ^2 verteilt sollte H > 0,05 oder besser noch >0,2 sein

c) Formeln (anders als Mc Fadden's Pseudo R^2 hängen sie auch vom Stichprobenumfang n ab):

Cox and Snell:	Nagelkerkes	Cragg and Uhler[17]
$R^2_{CS} = 1 - \left(\frac{L_0}{L_1}\right)^{2/n}$	$R^2_N = \frac{R^2_{CS} - R^2_{MFS}}{1 - (L_0)^{2/n}}$	$R^2_{CU} = \frac{L_1^{2/n} - L_0^{2/n}}{\left(1 - L_0^{2/n}\right) \cdot L_1^{2/n}}$

[16] Bei SPSS werden die Fälle entsprechend ihrer geschätzten Wahrscheinlichkeiten in 10 gleich große Gruppen eingeteilt und in diesen dann die erwarteten mit den beobachteten Häufigkeiten (E_g mit O_g) für das Eintreten von y = 1 verglichen Auf die so gewonnene Kontingenztabelle wird der χ^2 Test angewendet.

[17] erwähnt bei G. S. Maddala, Introduction to Econometrics, 2nd ed.(Prentice Hall, 1992) S. 334.

Zahlenbeispiel: Da L stets eine Wahrscheinlichkeit darstellt ist der logarithmierte Wert LL negativ und bei kleinen Wahrscheinlichkeiten haben wir betragsmäßig große negative Werte, etwa bei L = 0,001 ist LL = - 6,9. Angenommen $L_1 = 0{,}5$ und $L_0 = 0{,}02$, dann ist $LL_1 = -0{,}693$ und $LL_0 = -3{,}912$, so dass man erhält $R^2_{MF} = 0{,}8228$ (eine etwas unrealistische Größenordnung). Die Likelihood Ratio Statistik wäre bei diesen Zahlenbeispielen $-2(LL_0 - LL_1) = 6{,}43775$ (sie ist also nicht auf einen Wertebereich von 0 bis 1 normiert) und L_0/L_1 wie es in der Formel für R^2_{CS} benötigt wird ist 0,02/0,5 = 0,04.

Ist dagegen $L_1 = 0{,}3$ und $L_0 = 0{,}2$ dann ist $R^2_{MF} = 0{,}2519$ (was schon eine etwas realistischere Größenordnung ist) und die Likelihood Ratio Statistik wäre 0,8109 Für L_0/L_1 erhält man 2/3 statt 0,04.

4.5. Prüfung der Modellvoraussetzungen

Abgesehen von der (nichtlinear) transformierten Variable auf der linken Seite unterscheidet sich (3a) $\hat{\lambda}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... = z_i$ nicht von einer "normalen" Regression.

Es müssten also auch die üblichen Annahmen über die Variablen (keine endogenen Regressoren, keine omitted variables usw.) und die Störgrößen (Homoskedastizität, keine Autokorrelation usw.) gelten und statistisch überprüft werden (können). Hierzu habe ich in der Literatur nichts gefunden und die Standard-Statistiksoftware scheint dies auch nicht zu berücksichtigen.

4.6. Ausreißererkennung und stepwise regression

Maßgebend für die Ausreißerdiagnose ist bei SPSS offenbar das $\pm 2\sigma$ Intervall. SPSS arbeitet offenbar bottom up statt top down (in der Terminologie bei L. v. Auer und in meiner Vorlesung in DU: Maurer statt Steinmetz) mit sukzessive kleiner werdenden Devianz bzw. größer werdendem R^2_{CS}. Bei Mitteilung der Schätzergebnisse sollte man angeben, in welcher Reihenfolge die Regressoren in den Ansatz aufgenommen wurden

4.7. Interaktion

Der Koeffizient γ_{21} in der Regression mit Interaktion von 1 und 2 in der Regressionsgleichung $\hat{\lambda}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + .\gamma_{21}(x_{1i} x_{2i}) + ...$ (neben den Haupteffekten gemessen an β_1 und β_2) ist wie folgt zu interpretieren:

(2a) $\hat{\omega}_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_{12}(x_{1i} x_{2i}) + \beta_3 x_{3i} ...) = e^{\beta_0} \cdot e^{\beta_1 x_{1i}} \cdot e^{\beta_2 x_{2i}} \cdot e^{\gamma_{12} x_{1i} x_{2i}} \cdot e^{\beta_3 x_{3i}} \cdot ...$

Bei einem Übergang von x_{2i} zu $(x_{2i} + \Delta)$ verändert sich $\hat{\omega}$ *ceteris paribus*, d.h. wenn alle anderen Einflüsse auf $\hat{\omega}$ können als Konstante betrachtet werden, um den Faktor $\psi = e^{\beta_2 \Delta} \cdot e^{\gamma_{12} x_{1i} \Delta}$, also um mehr (weniger) als bei einer bloßen Änderung von x_{2i} um Δ, wenn γ_{12} positiv (γ_{12} negativ) ist[18]. Wie sich das auf die Veränderung der Wahrscheinlichkeit π von π_0 zu π_1 auswirkt zeigt

(4a) $\frac{\hat{\pi}_1}{\hat{\pi}_0} = 1 + \frac{\psi - 1}{1 + \psi \hat{\omega}_0}$,[19] und es hängt damit davon ab, wie groß $\hat{\omega}_0$ ist.

Die Interpretation von γ_{12} gilt entsprechend bei einer Veränderung von x_{1i} zu $(x_{1i} + \Delta)$ und Konstanz von x_{2i}: ist $\gamma_{12} > 0$ (und damit $e^{\gamma_{12}} \cdot > 1$) verändern sich die odds (immer bezogen auf die Wahrscheinlichkeit, das Panel zu verlassen, also "auszutreten) mehr als allein gemäß β_1(bzw. weniger wenn $\gamma_{12} < 0$ ist).

---

[18] weil dann $K = \exp(\gamma_{12} x_{1i} \Delta) > 1$ bzw. < 1 ist, d.h. $\ln(K) = \gamma_{12} x_{1i} \Delta > 0$ oder < 0 ist, was bei positivem $x_{1i}\Delta$ impliziert, dass dann $\gamma_{12} > 0$ oder < 0 ist.

[19] In der Literatur findet man auch Formeln, in denen γ_{12} als Verhältnis (Bruch) ausgedrückt wird, wobei im Zähler und Nenner jeweils Verhältnisse von bedingten Wahrscheinlichkeiten (für $x_2 = 1$ bzw. $x_2 = 0$ bedingt durch $x_1 = 1$ bzw. $x_1 = 0$) stehen. Ich bin mir nicht sicher, ob so γ_{12} richtig interpretiert ist. Es geht ja immer nur um die *Veränderung* der odds *bezüglich des Austretens aus dem Panel*, nicht bezüglich der Ausprägungen von x_1 oder x_2.

5. Häufige Fehler bei Anwendungen

Ich habe sehr oft gesehen, dass eine logistische Regression gerechnet wurde und dabei einige typische Fehler immer wieder gemacht wurden. Anfänglich hatte ich noch einmal hin und wieder die Autoren darauf per E-Mail aufmerksam gemacht (mich zig-mal entschuldigend: bitte nicht als Kritik auffassen etc.). Ich musste dann aber feststellen, dass ich keine Antwort bekam und offenbar den Autoren nur lästig war, so dass es mir inzwischen ziemlich egal ist, was in Veröffentlichungen alles für ein Quatsch mit der Statistik Software gerechnet wird.

Hier die beiden häufigsten Fehler:

1. Man macht einen Signifikanztest, obgleich die Daten keine Stichprobe sind, sondern faktisch eine Totalerhebung der Grundgesamtheit darstellen. Was für ein abenteuerliches (Un-)Verständnis von "Hypothesen" und "Signifikanz" faktisch hinter solchen Betrachtungen steht[20] habe ich in einem speziellen Text (Nr. 25 auf dieser Seite "Downloads – Allgemein") dargestellt.
2. Die odd ratios, die das Statistik-Programm als Ergebnis ausweist werden als Veränderungen der Wahrscheinlichkeit (statt als Veränderungen der odds) interpretiert. Ist beispielsweise die odds ratio $\exp(\beta_i) = 1{,}2$ (x_i sei z.B. eine dichotome Variable: x_i ist 0 oder 1) wird das interpretiert, als würde sich p um 20% erhöhen wenn $x_i = 1$ statt 0 ist.

   Dass das nicht stimmt steht auch in der Fußnote * unter Übers. 1 (Seite 4).

Wenn $r = \omega_2/\omega_1$ ist, wie verändert sich dann die Wahrscheinlichkeit $p_1 \rightarrow p_2$? Das hängt, wie auf S. 4 schon gezeigt von der Höhe von p_1 ab. Mit $\beta = p_2/p_1$ ist der Zusammenhang

$$\beta = \frac{r}{1 + p_1(r-1)}.$$

Beispiel (vgl. oben Übers. 1):

$r = 2$	$p_1 = 0{,}2$	$\rightarrow \beta = 1{,}667 < r$
	$p_1 = 0{,}8$	$\rightarrow \beta = 1{,}111 < r$
$r = 0{,}8$	$p_1 = 0{,}2$	$\rightarrow \beta = 0{,}833 > r$
	$p_1 = 0{,}8$	$\rightarrow \beta = 0{,}9524 > r$

Wenn $r < 1$ ist $\beta > r$ und wenn $r > 1$ ist $\beta < r$.

[20] Man stellt keine Vermutungen (Hypothesen) auf über etwas, was man bereits kennt. Man hat (Totalerhebung!) $\mu = 10$ festgestellt und sieht nun weil man H_0: $\mu = 0$ testet und mit 5% oder 1% verwerfen kann, dass die Wahrscheinlichkeit nicht dafür spricht, das μ null sein könnte (das, wo doch klar ist, dass μ 10 ist!!). Das alles ist wohl dem Umstand zu verdanken, dass heutzutage eine Arbeit empirisch (statistisch) sein muss und erst dann wissenschaftlich ist wenn in ihr mindestens einmal ein Signifikanztest auftaucht.

Was für ein Unsinn ein Signifikanztest bei einer Vollerhebung ist wird besonders deutlich bei der statistischen Qualitätskontrolle. H_0 ist hier meist "die Ware ist gut". Der Fehler 2. Art (β-Fehler) besteht darin, die falsche H_0 (die Ware/Lieferung ist eigentlich schlecht) anzunehmen (die Lieferung zu akzeptieren). Man nennt das auch das Käuferrisiko. Bei einer Totalerhebung (der Kunde hat alle Stücke geprüft) gibt es aber kein Käuferrisiko (keinen β-Fehler) weil der Kunde ja alle Stücke gesehen hat und einfach nur die guten kauft und die schlechten zur Seite legt und nicht kauft. Man könnte auch sagen: wenn 1- β die power (Macht, Trennschärfe) eines Tests ist (also einen Unterschied zwischen H_0 und H_1 auch zu erkennen), dann ist die power bei einer Totalerhebung maximal (also 1).