

Trendstudie von Fuchs et al. 1995

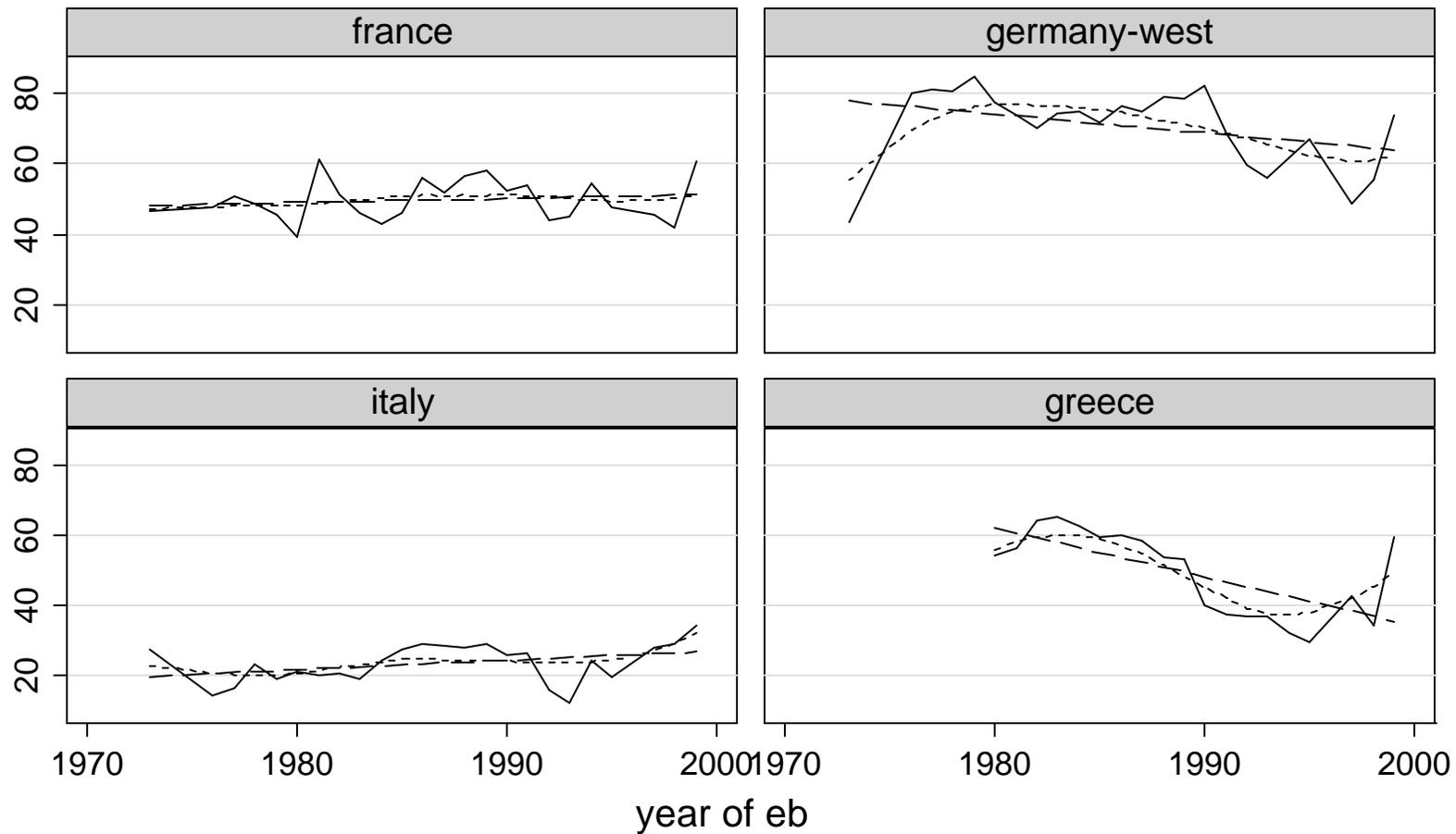
- Fortschreibung
- Einteilung nach Niveau und Varianz der Unterstützung
- Prüfung auf nicht-lineare Trends
- Analyse auf Individualebene (erfordert Logit-Analyse)

Niveau und Varianz für vier Länder

- Voraussetzung: Zufriedenheit wird dichotomisiert (zufrieden / nicht zufrieden)
- Betrachtung des Anteils der Zufriedenen per Jahr / Land (Aggregation)
- Ca. 30 „Fälle“ pro Land

nation	mean	sd
france	49.9	6.0
germany-west	70.8	11.0
italy	23.4	5.6
greece	49.2	12.1

Trends in vier Ländern (Aggregat)



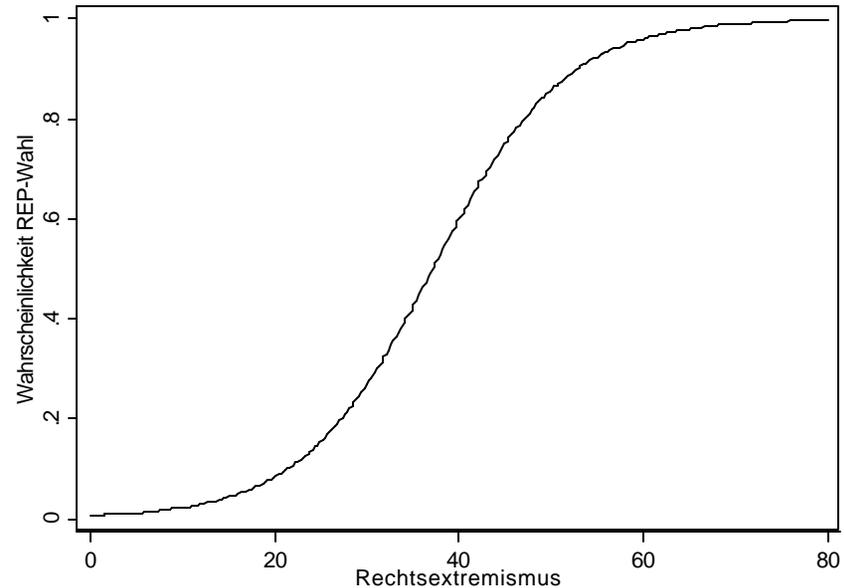
Graphs by nation

Analyse auf der Individualebene

- Abhängige Variable wird weiter als dichotom (1=zufrieden, 0=unzufrieden) betrachtet
- Modelliert wird die Wahrscheinlichkeit,
 - daß ein beliebiger Befragter zufrieden ist
 - in Abhängigkeit von der Zeit
 - von der Nation
 - weitere theoretisch interessante Kovariaten könnten hinzugefügt werden

Probleme der linearen Regression

- Viele interessante abhängige Variablen in Politikwissenschaft dichotom oder dichotomisierbar (Nichtwahl, Wahl einer rechten Partei etc.)
- 0/1-Kodierung der abhängigen Variablen, Interpretation als *Wahrscheinlichkeit (erwartete relative Häufigkeit)*
- Im Prinzip akzeptabel, aber
 - Wahrscheinlichkeiten müssen zwischen 0 und 1 liegen
 - manchmal ist S-förmige Beziehung plausibler als lineare



Grundgedanke der logistischen Regression (Logit-Analyse)

- Wahrscheinlichkeiten haben Wertebereich 0-1
- Abhängige Variable wird deshalb transformiert:
- Zunächst werden statt Wahrscheinlichkeiten „Odds“ betrachtet, das sind Brüche aus einer Wahrscheinlichkeit und ihrer Gegenwahrscheinlichkeit:
 $p \rightarrow p/1-p$. Diese haben einen Wertebereich von 0 bis (fast) $+\infty$
- Diese Odds werden in einem zweiten Schritt logarithmiert:
 $p/1-p \rightarrow \ln(p/1-p)$ (natürlicher Logarithmus = Umkehrfunktion zur Exponentialfunktion auf Basis von e (2.718...)) .

Grundgedanke der logistischen Regression (Logit-Analyse)

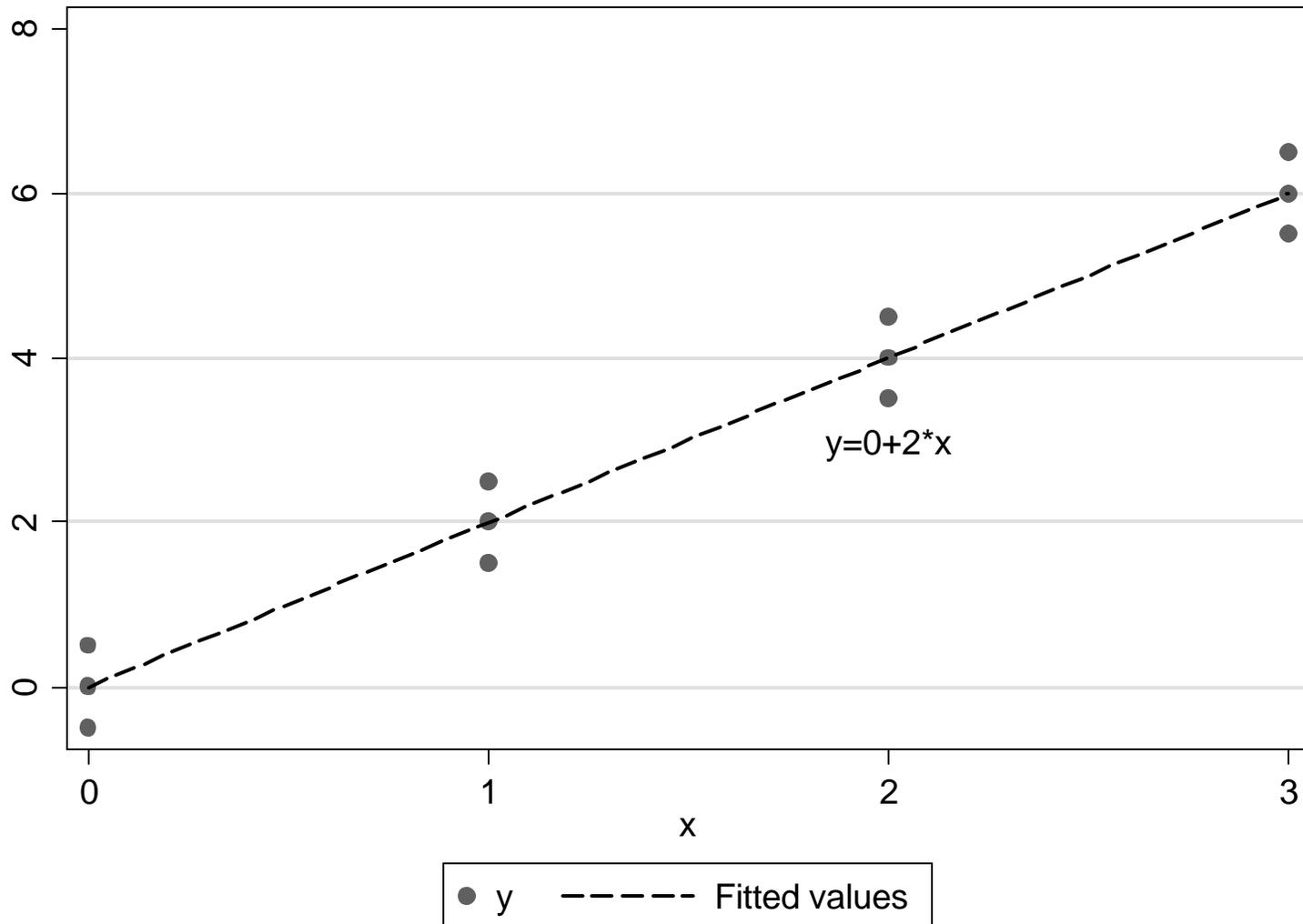
- Diese Größe wird auch als „logit“ bezeichnet. Logits haben einen Wertebereich von (fast) $-\infty$ bis (fast) $+\infty$
- Als unabhängige Variablen kommen wie bei der linearen Regression intervallskalierte oder kategoriale Merkmale in Frage
- Modell ist ebenfalls multivariat darstellbar und linear in den Logits: $\text{logit}(y) = a + b_1 x_1 - b_2 x_2 \dots$
- Wenn man sich für die abhängige Variable in der ursprünglichen Form interessiert, muß die Transformation rückgängig gemacht werden. Diese Beziehung ist nicht-linear

$$y = \frac{e^{(a+b_1x_1+b_2x_2\dots)}}{1 + e^{(a+b_1x_1+b_2x_2\dots)}} \quad 7$$

Interpretation der Koeffizienten

- Bei der linearen Regression ist die Interpretation einfach
 - a ist der erwartete Wert von y , wenn $x_1=0$
 - b_1 entspricht der Veränderung des erwarteten Wertes von y , wenn x_1 um eine Einheit zunimmt.
 - Veränderungen von x und y sind proportional und vom Niveau der unabhängigen Variablen unabhängig
 - Positives Vorzeichen \Leftrightarrow positiver Zusammenhang

Lineare Regression



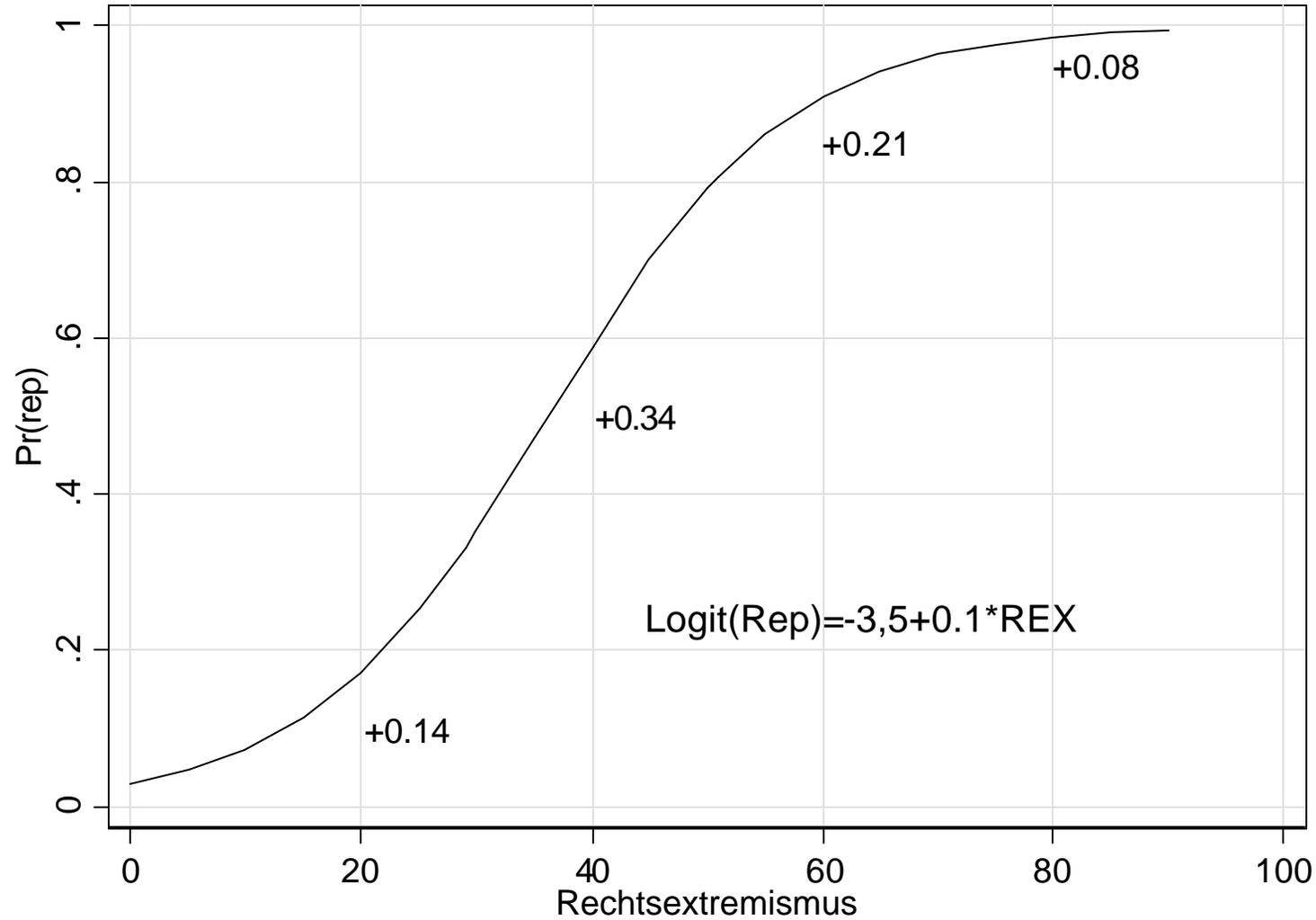
Interpretation der Koeffizienten

- Bei der logistischen Regression ist die Interpretation schwierig:
 - Positives Vorzeichen \Leftrightarrow positiver Zusammenhang (höhere Wahrscheinlichkeit)
 - Koeffizient b_1 beschreibt lineare Veränderungen des *Logits*, wenn x_1 um eine Einheit zunimmt. Bedauerlicherweise sind Veränderungen in den Logits für niemanden intuitiv nachvollziehbar
 - Etwas anschaulicher ist e^{b_1} : Dies ist der multiplikative *Faktor*, um den sich die Odds verändern, wenn x_1 um eine Einheit zunimmt. Auch Odds sind aber nicht sehr anschaulich

Interpretation der Koeffizienten

- Eigentlich interessant sind die geschätzten Wahrscheinlichkeiten. Wegen der S-Form der Kurve aber keine Proportionalität zwischen x und *Wahrscheinlichkeit*: Steigung der Kurve (=Veränderung der Wahrscheinlichkeit) hängt vom Wert von x ab (geringe Steigung bei sehr hohen und sehr niedrigen Werten)
- Im multivariaten Modell hängt die Veränderung der geschätzten Wahrscheinlichkeit vom Wert *aller* unabhängigen Variablen ab
- Häufig ist es deshalb sinnvoll, sich typische bzw. interessante Konstellationen anzuschauen und für diese durch Einsetzen die geschätzte Wahrscheinlichkeit zu errechnen

Logistische Regression



Anwendung auf Italien/Frankreich

- Unterscheiden sich Niveau und Entwicklung der Zufriedenheit in beiden Ländern?
- Schätzung des Modells
 $\text{logit}(\text{zufrieden}) = a + b_1 * \text{Zeit} + b_2 * \text{Italien} + b_3 * \text{Zeit} * \text{Italien}$

Ergebnis

	zufrieden mit Demokratie
Italy (b_2)	-1.283** (0.039)
Zeit (b_1)	0.005** (0.001)
Zeit*Italien (b_3)	0.006** (0.002)
Konstante (a)	-0.081** (0.025)

Das bedeutet...

- In Frankreich war es zu Beginn der Untersuchung etwas unwahrscheinlicher, zufrieden als unzufrieden zu sein ($e^{-0.81}=0.92$ =Odds zufrieden vs. unzufrieden)
- In Italien war es zu Beginn der Untersuchung erheblich unwahrscheinlicher, zufrieden als unzufrieden zu sein ($e^{-0.81-1.28}=0.25$ = Odds zufrieden vs. unzufrieden)
- In Frankreich nimmt die Wahrscheinlichkeit für Zufriedenheit zu
- In Italien nimmt die Zufriedenheit schneller zu als in Frankreich
- Weitere Größen (Geschlecht, Alter, Bildung, PostMat) können aufgenommen und europaweit oder per Land und/oder über die Zeit getestet werden – alles auf der Individualebene

Geschätzte Anteile „zufrieden“

