Fünf Theorien zur Unterstützung der Europäischen Integration (Gabel 1998)

- Was sind „Dummies“?
- Mögliches Problem: Kollinearität

Dummy-Variablen

- Kategoriale Größen (Land, Geschlecht, Konfession u.ä.) können in Regressionsmodellen als unabhängige Variablen eingesetzt werden
- Dichotome Variablen (z.B. männliches Geschlecht) werden auf 0/1 kodiert: 1 = Merkmal liegt vor; 0= Merkmal liegt nicht vor
- Beispiel: Konservatismus= 55,9-2*männlich. 0/1 Variablen werden als „Dummies“ bezeichnet

Dummy-Variablen

- Nominalskalierte Merkmale mit mehr als zwei Ausprägungen (katholisch, protestantisch, keine Konfession) werden durch *mehrere* Dummies erfaßt.
- Dabei muß eine beliebige Kategorie ausgelassen werden („Referenzkategorie); ansonsten ist das Modell nicht schätzbar (Kollinearität)
- Für ein Merkmal mit drei Ausprägungen werden deshalb nur zwei Dummies benötigt (der Wert des potentiellen dritten Dummies ist durch die ersten beiden schon festgelegt)
- Konservatismus=54,5+1,2*katholisch+0,6*protestantisch

Dummy-Kodierung: Konfession

Konfession	Dummy: Kath.	Dummy: Prot.
keine	0	0
Katholisch	1	0
Protestantisch	0	1
Logisch ausgeschlossen	1	1

Orthogonalität

- Bei experimentellen Designs werden die unabhängigen Variablen von der Forscherin gesetzt
- Beispiel Aggressivität durch Alkohol/Horrorfilme:
  - Gruppe 1: kein Treatment (Kontrolle)
  - Gruppe 2: nur Alkohol
  - Gruppe 3: nur Film
  - Gruppe 4: Film + Alkohol
  - Zufällige Aufteilung auf Gruppen, Vorher-Nachher-Messung (y=Differenz) mit psychometrischer Skala
- $y=a+b_1*Alkohol+b_2*Film+(b_3Alkohol*Film)$
- b_1 erfaßt den Effekt von Alkohol, wenn die Variable „Film“ (und alle anderen Einflüsse) konstant gehalten werden

Orthogonalität

	Horrorfilm		
Alkohol	0	1	Total
0	10	10	20
1	10	10	20
Total	20	20	40

- Alle Zellen/Kombinationen gleichmäßig besetzt
- Keine Korrelation zwischen den beiden unabhängigen Variablen
- Keine Hintergrundvariable, die deren Ausprägung beeinflußt

Kollinearität

- Bei Umfragedaten (ex-post-facto Design) in der Regel Korrelation zwischen unabhängigen Variablen
- Manche Kombinationen von Ausprägungen erstens nicht beobachtet, zweitens empirisch unplausibel/unmöglich (angelernter Arbeiter mit Hochschulabschluß)
- Lineare Beziehungen zwischen den unabhängigen Variablen werden als (Multi-) Kollinearität bezeichnet
- Typisch für Umfragedaten z.B. enge Beziehungen zwischen Schulabschluß, Beruf, Einkommen und politischen Einstellungen

Kollinearität

- Perfekte Kollinearität
  - $x_1=a+b*x_2$ | $R^2=1$
  - Fehler
    - Intrinsische Beziehung zwischen zwei Variablen z.B. Geburtsjahr und Alter in Jahren (bei einer Querschnittsbefragung)
    - Bei Dummies: Referenzkategorie durch zusätzlichen Dummy repräsentiert
    - Interaktionseffekte
  - Zahl der Fälle < Zahl der Variablen
- Hohe Kollinearität
  - $x_1=a+b*x_2$ | $R^2=>0.9$
  - Standardfehler werden sehr groß = Schätzung schwanken sehr stark über Stichproben hinweg
  - Die Schätzungen für den Koeffizienten einer Variablen werden davon beeinflußt, welche anderen Variablen im Modell enthalten sind
  - Interpretationsprobleme: Kann man sich überhaupt vorstellen, daß die übrigen Variablen konstant gehalten werden?

Beispiel

	1. Aggr.	2. Aggr.	3. Aggr. (coll.)	4. Aggr. (coll.)
Alkohol	2.40	2.40	2.57	1.77
Horrorfilm	2.96		3.12	
Constant	-0.02	1.45	-0.15	1.93
N	40	40	32	32
R-squared	0.82	0.33	0.79	0.20

- 1.+2. Parameter für „Alkohol“ identisch
- 3.+4. Parameter kleiner, wenn „Film“ nicht berücksichtigt
- Grund: negative (-.25) Korrelation zwischen Alkohol und Film

Kollinearität

- Diagnose
  - Regression einer unabhängigen Variable auf eine (Kollinearität) oder *alle* (Multikollinearität) anderen unabhängigen Variablen
  - $1\text{-}R^2$=tolerance; Faustregel tol >0.1
  - 1/tol.=VIF; Faustregel VIF <10
- Maßnahmen
  - Alternative Kodierung (bei Interaktionseffekten)
  - Mehr Fälle, möglichst mit „ungewöhnlichen“ Kombinationen der unabhängigen Variablen
  - (theoretisch begründeter) Ausschluß von unabhängigen Variablen
  - Zusammenfassung der hochkorrelierten Variablen zu einem Index/Faktor
  - Fortgeschrittene Methoden