# Logistische Regression
# I.

## Odds,
## Logits,
## Odds Ratios,
## Log Odds Ratios

# Logistische Regression

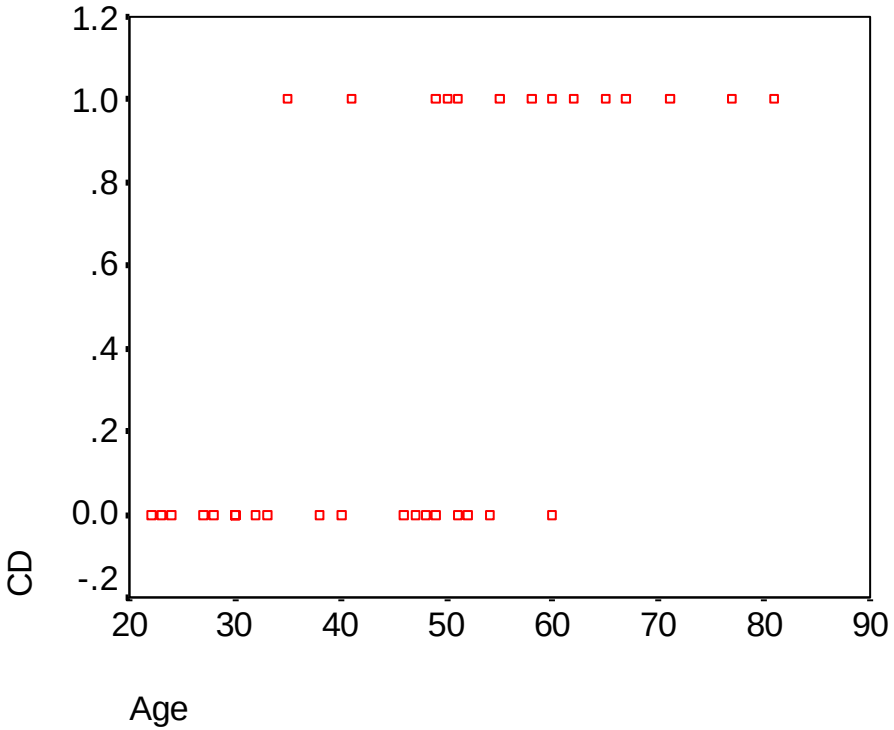
**Tabelle 2 Alter und Symptome von Herz-/Kreislauferkrankung(CD)**

| Alter | CD | Alter | CD | Alter | CD |
|---|---|---|---|---|---|
| 22 | 0 | 40 | 0 | 54 | 0 |
| 23 | 0 | 41 | 1 | 55 | 1 |
| 24 | 0 | 46 | 0 | 58 | 1 |
| 27 | 0 | 47 | 0 | 60 | 1 |
| 28 | 0 | 48 | 0 | 60 | 0 |
| 30 | 0 | 49 | 1 | 62 | 1 |
| 30 | 0 | 49 | 0 | 65 | 1 |
| 32 | 0 | 50 | 1 | 67 | 1 |
| 33 | 0 | 51 | 0 | 71 | 1 |
| 35 | 1 | 51 | 1 | 77 | 1 |
| 38 | 0 | 52 | 0 | 81 | 1 |

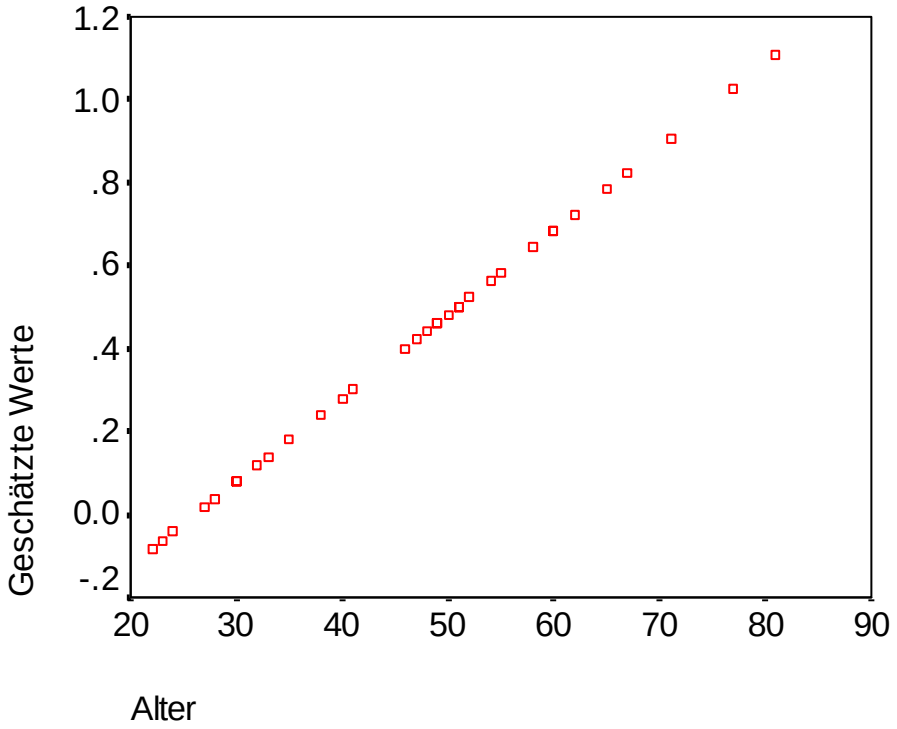
Quelle:http://www.tulane.edu/~PsycStat/ruscher/logistic.ppt

# Beobachtete vw. geschätzte Werte auf der Basis eines linearen Regressionsmodells für eine dichotome abhängige Variable
# Beispiel: CHD und Alter

## Beobachtete Werte




## Lineare Regression



# Probleme bei linearer Regression mit dichotomer abhängiger Variable

1. Kleinste Quadrate Regression basiert auf Normalverteilten Fehlertermen

Bei dichotomer abhängiger Variable kann der Fehlerterm nur zwei Werte annehmen (richtig, falsch)

Folge: Hypothesentests können ungültig sein

2. Vorhergesagten Werte können größer als “eins” und kleiner als “null” sein

# Wahrscheinlichkeiten und Odds

Wahrscheinlichkeit  Herz-/Kreislauferkrankung
Beispiel CD
CD=0:P= 0.58 (=19/33) Wahrscheinlichkeit keine HK
CD=1:P= 0.42 (=14/33) Wahrscheinlichkeit HK

Odds Herz/Kreislauferkrankung:
Wahrscheinlichkeit, dass etwas wahr ist dividiert durch die Wahrscheinlichkeit, dass es nicht wahr ist
Beispiel CD
Odds=(P/1-P)
Odds=0.42/0.58=0.75 Odds HK

# Odds in einer 2x2 Tabelle

| | Raucher | Nicht Raucher |
|---|---|---|
| gestorben | $p_1$ (0.30) | $p_2$ (0.25) |
| überlebt | $1-p_1$ (0.70) | $1-p_2$ (0.75) |

**Odds als Raucher zu sterben:**

**$odds(1)= p_1/(1-p_1) = .30/.70=0.43$**

**Odds als Nichtraucher zu sterben:**

**$odds(2)= p_2/(1-p_2)= .25./75=0.33$**

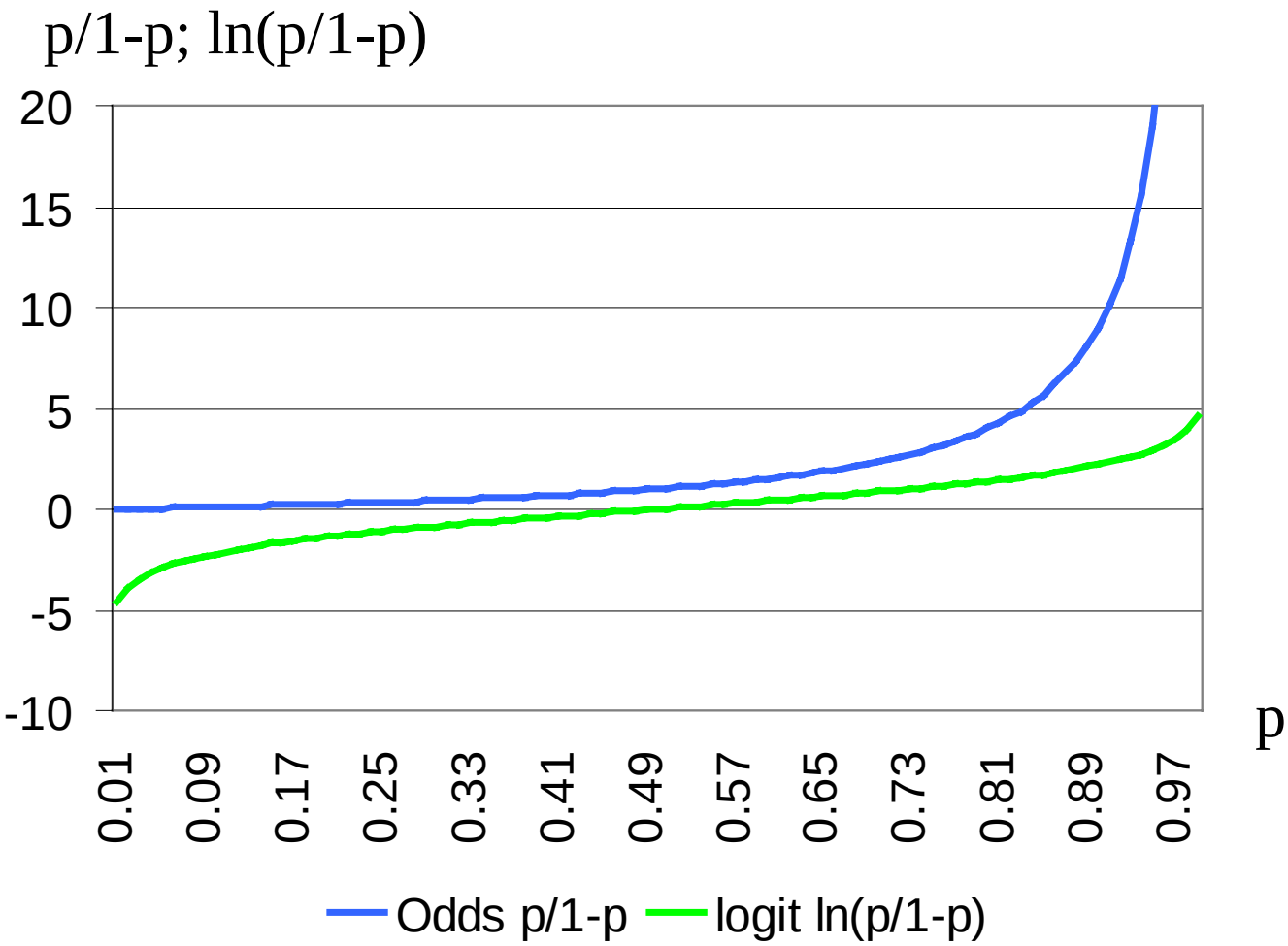
# Logit Transformation

$$Logit = \ln\left(\frac{p}{1-p}\right)$$

| p | .10 | .20 | .30 | .40 | .50 | .60 | .70 | .80 | .90 |
|---|---|---|---|---|---|---|---|---|---|
| 1-p | .90 | .80 | .70 | .60 | .50 | .40 | .30 | .20 | .10 |
| Odds | 0.11 | 0.25 | 0.43 | 0.67 | 1.00 | 1.50 | 2.33 | 4.00 | 9.00 |
| Logit | -2.20 | -1.39 | -0.85 | -0.41 | 0.00 | 0.41 | 0.85 | 1.39 | 2.20 |

- Logit ist symmetrisch um 0 (p = .50)
- Je extremer die Wahrscheinlichkeit p von .50 abweicht, desto stärker verändert sich der Logit
- Für sehr große Logits nähert sich p null bzw. eins an, ohne jedoch diese Werte zu erreichen
- Daher befinden sich die Wahrscheinlichkeiten p auch für sehr große Logits immer in den Schranken von null und eins

# Wertebereich p, odds und logits




p: 0<=p<=1

odds: 0,+∞

logit: -∞, +∞

# Transformation Odds in Logits und zurück

$$Logit = \ln\left(\frac{p}{1-p}\right)$$

| p | .10 | .20 | .30 | .40 | .50 | .60 | .70 | .80 | .90 |
|---|---|---|---|---|---|---|---|---|---|
| 1-p | .90 | .80 | .70 | .60 | .50 | .40 | .30 | .20 | .10 |
| Odds | 0.11 | 0.25 | 0.43 | 0.67 | 1.00 | 1.50 | 2.33 | 4.00 | 9.00 |
| Logit | -2.20 | -1.39 | -0.85 | -0.41 | 0.00 | 0.41 | 0.85 | 1.39 | 2.20 |

Bsp: p = 0.20; 1-p = 0.80

Odds = p/1-p = 0.20/0.80 = 0.25

Logit = ln(Odds) = ln(0.25) = -1.386

Odds = exp(Logit)=exp(-1.386)=0.25

exp...Euler'sche Zahl=2.71828

# Odds Ratio in einer 2x2 Tabelle

| | Raucher | Nicht Raucher |
|---|---|---|
| gestorben | $p_1$ (0.30) | $p_2$ (0.25) |
| überlebt | $1-p_1$ (0.70) | $1-p_2$ (0.75) |

**Odds als Raucher zu sterben:**

**$odds(1)= p_1/(1-p_1) = .30/.70=0.43$**

**Odds als Nichtraucher zu sterben:**

**$odds(2)= p_2/(1-p_2)= .25./75=0.33$**

# Odds Ratio in einer 2x2 Tabelle

**Odds ratio (1):**

**Quotient aus odds(1) und odds(2)**

**Quotient: odds als Raucher zur sterben zu odds als Nichtraucher zu sterben.**

**Odds ratio (1)= $p_1/(1-p_1)$ / $p_2/(1-p_2)$**

**= .43 / .33 =1.29**

Das Risiko eines Rauchers zu sterben ist um 29% höher, als das Risiko eines Nichtrauchers zu sterben.

Nichtraucher: Referenzgruppe

# Odds Ratio in einer 2x2 Tabelle

**Odds ratio (2):**

**Quotient aus odds(2) und odds(1)**

**Quotient: odds als Nichtraucher zur sterben zu Odds als Raucher zu sterben.**

**Odds Ratio (2) = $p_2/(1-p_2)$ / $p_1/(1-p_1)$**

**= .33 / .43 = 0.77**

Das Risiko eines Nichtrauchers zu sterben ist um 23% niedriger, als das Risiko eines Rauchers zu sterben.

Raucher: Referenzgruppe

# Odds, Odds Ratio

**Der Odds**

- Wahrscheinlichkeit zur Gegenwahrscheinlichkeit

  odds(1)= $p_1/(1-p_1)$ = .30/.70=0.43 Odds Raucher zu sterben

  odds(2)= $p_2/(1-p_2)$= .25/.75=0.33 Odds Nichtraucher zu sterben

**Der LN(Odds)**

LN(odds(1))= LN(0.43)= -0.84

LN(odds(2))= LN(0.33)= -1.11

**Der Odds Ratio**

- Der Quotient aus zwei Odds

  Odds ratio (1) = odds(1)/odds(2)= 1.29 (RF Nichtraucher)

  Odds ratio (2) =odds(2)/odds(1)= 0.77 (RF Raucher)

**Der LN(Odds Ratio)**

- Der natürliche Logarithmus des Odds Ratios

  LN (Odds ratio 1) = 0.25 (RF Nichtraucher)

  LN (Odds ratio 2) = -0.25 (RF Raucher)

# Interpretation Odds ratio und LN(Odds ratio)

| | Anteilswerte p | | Odds | | Odds Ratio | ln (Odds ratio) |
|---|---|---|---|---|---|---|
| | Raucher | Nicht Raucher | Raucher | Nicht Raucher | | |
| Beispiel 1 | | | | | | |
| gestorben | 0.3 | 0.25 | 0.429 | 0.333 | 1.286 | 0.251 |
| überlebt | 0.7 | 0.75 | 2.333 | 3.000 | 0.778 | -0.251 |
| Beispiel 2 | | | | | | |
| gestorben | 0.9 | 0.1 | 9.000 | 0.111 | 81.000 | 4.394 |
| überlebt | 0.1 | 0.9 | 0.111 | 9.000 | 0.012 | -4.394 |
| Beispiel 2 | | | | | | |
| gestorben | 0.5 | 0.5 | 1.000 | 1.000 | 1.000 | 0.000 |
| überlebt | 0.5 | 0.5 | 1.000 | 1.000 | 1.000 | 0.000 |

Odds ratio (OR):
1. OR=1, kein Zusammenhang
2. OR>1, positiver Zusammenhang
3. OR<1, negativer Zusammenhang
4. Schief verteilt

Ln(Odds ratio) (LN(OR)):
1. LN(OR=0), kein Zusammenhang
2. LN(OR>0), positiver Zusammenhang
3. LN(OR)<0, negativer Zusammenhang
4. symmetrisch um Null verteilt

# Logistische Regression
# II.

# Modell
# Kategorielle Variablen
# Interpretation der Parameter

# Logistische Regression

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$$

Abhängige Variable
logit

=

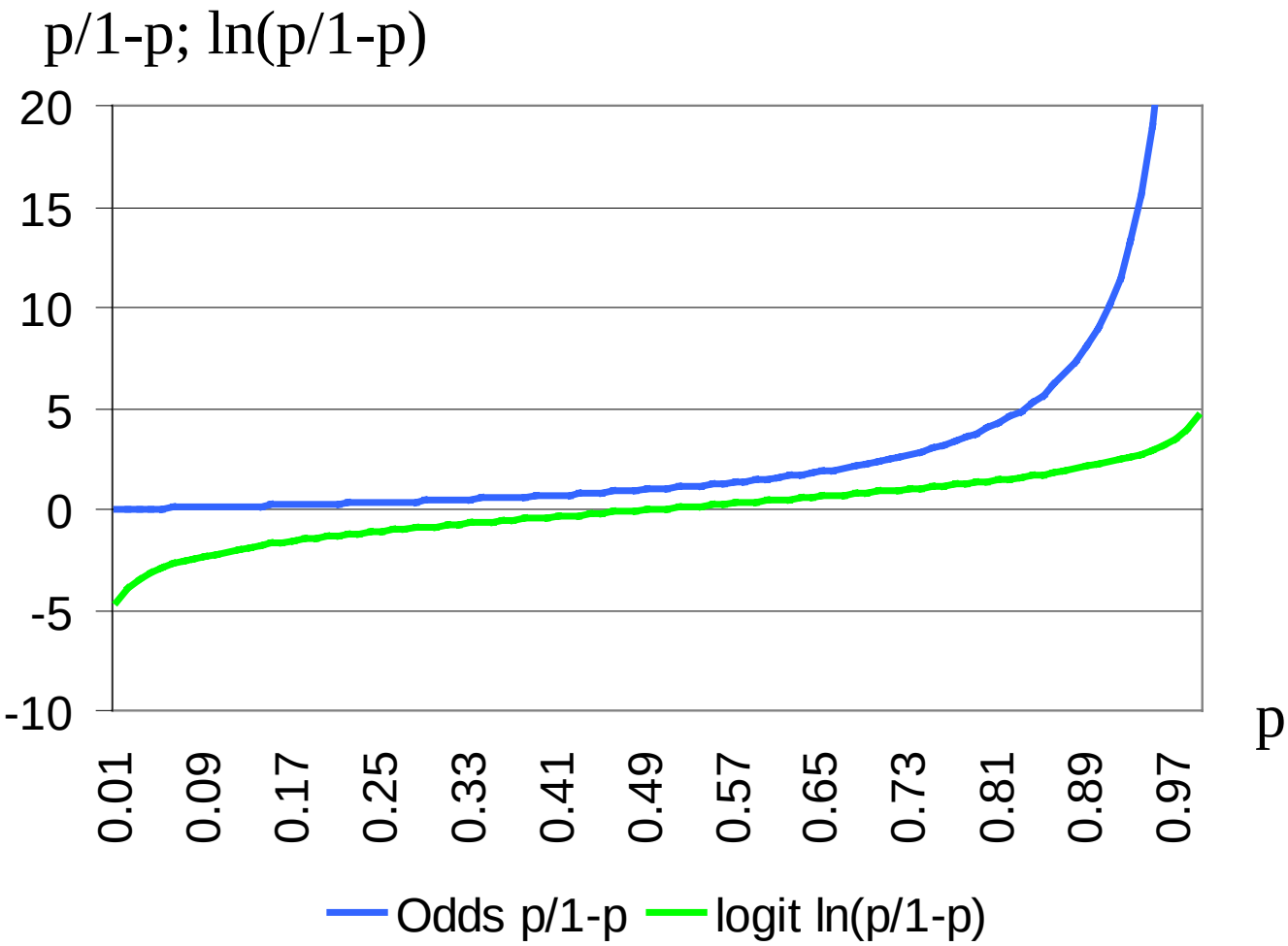
Unabhängige Variablen:
$x_1$ .. $x_k$

Parameterwerte:
$\beta_0$ .. $\beta_k$

# Abhängige Variable
# logit

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right]$$

# Wertebereich p, odds und logits




p: 0<=p<=1

odds: 0,+∞

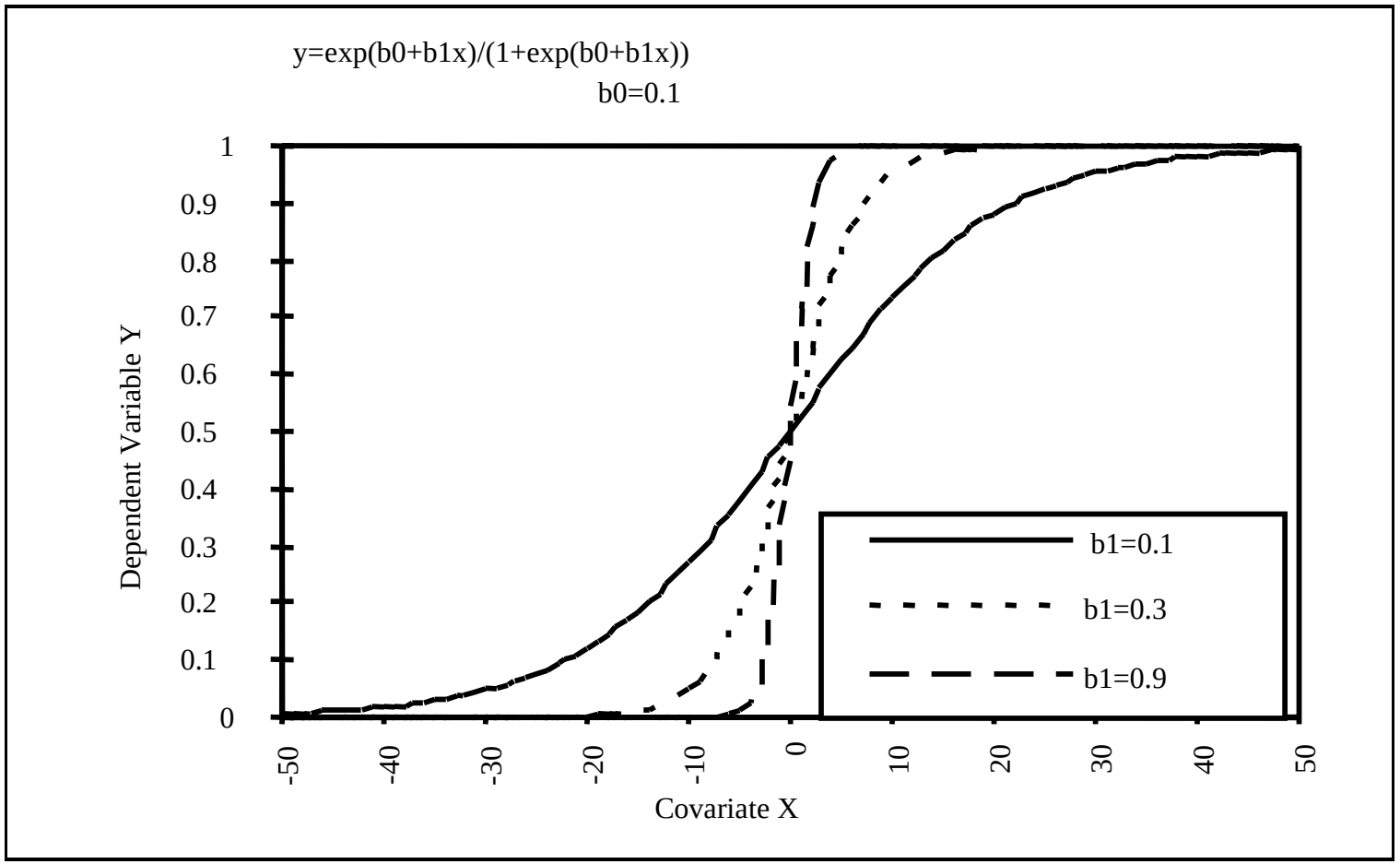
logit: -∞, +∞

# Logistische Regression

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 \ldots + \beta_k x_k)}$$

Logistische Verteilung

# Logistische Verteilung



# Unabhängige Variablen

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$$

# Kategorielle unabhängige Variablen = Dummy Variablen

**Kodierung von Bildung mit Hochschule als Referenzgruppe**

| Bildung | Dummy Variablen | | | |
|---|---|---|---|---|
| | D1 | D2 | D3 | D4 |
| Hochschule | 0 | 0 | 0 | 0 |
| Abitur | 1 | 0 | 0 | 0 |
| Fachschule | 0 | 1 | 0 | 0 |
| Lehre | 0 | 0 | 1 | 0 |
| Pflichtschule | 0 | 0 | 0 | 1 |

Referenzgruppe wird immer ausgelassen in der Kodierung

# Datenstruktur:
## Sterblichkeit in Abhängigkeit von Alter

| age | survive | count |
|---|---|---|
| 35-39 | 1 | 84 |
| 35-39 | 1 | 5 |
| 35-39 | 0 | 4 |
| 35-39 | 1 | 1 |
| 35-39 | 1 | 1 |
| 35-39 | 0 | 2 |
| 35-39 | 1 | 4 |
| 35-39 | 1 | 6 |
| 35-39 | 0 | 1 |
| 35-39 | 1 | 23 |
| 35-39 | 1 | 3 |
| 40-44 | 1 | 2 |
| 40-44 | 0 | 1 |
| 40-44 | 1 | 1 |

1. age: kategoriell

2. survive: 1.. gestorben
   0.. überlebt

3. count: Anzahl der Personen
   SPSS Syntax:
   WEIGHTBY count .

# SPSS Syntax

```
WEIGHT by COUNT.

LOGISTIC REGRESSION VAR=survive

  /METHOD=ENTER age

  /CONTRAST (age)=Indicator(1).
```

Dummy Kodierung
1. Kategorie: Referenzgruppe

Abhängige Variable survive

0 überlebt

1 gestorben

Unabhängige Variable age

**AGE**

| | | Häufigkeit | Prozent | Gültige Prozente | Kumulierte Prozente |
|---|---|---|---|---|---|
| Gültig | 35-39 | 241426 | 20,5 | 20,5 | 20,5 |
| | 40-44 | 248388 | 21,1 | 21,1 | 41,5 |
| | 45-49 | 200114 | 17,0 | 17,0 | 58,5 |
| | 50-54 | 224376 | 19,0 | 19,0 | 77,5 |
| | 55-59 | 265165 | 22,5 | 22,5 | 100,0 |
| | Gesamt | 1179469 | 100,0 | 100,0 | |

# SPSS Output

**Variablen in der Gleichung**

| | | Regressions koeffizientB | Standardf ehler | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Schritt 1[a] | AGE | | | 1101,007 | 4 | ,000 | |
| | AGE(1) | ,289 | ,078 | 13,840 | 1 | ,000 | 1,335 |
| | AGE(2) | ,815 | ,073 | 123,584 | 1 | ,000 | 2,259 |
| | AGE(3) | 1,224 | ,068 | 324,924 | 1 | ,000 | 3,401 |
| | AGE(4) | 1,631 | ,064 | 644,390 | 1 | ,000 | 5,109 |
| | Konstante | -6,737 | ,059 | 12971,259 | 1 | ,000 | ,001 |

a. In Schritt 1 eingegebene Variablen: AGE.

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$$

$$\left[\frac{\pi(x)}{1-\pi(x)}\right] = e^{\beta_0} * e^{\beta_1 x_1} * e^{\beta_2 x_2} * \ldots\ldots * e^{\beta_k x_k}$$

# Interpretation der Parameterwerte

## Referenzgruppe (RF) ist jüngste Altersgruppe 35-39

**Variablen in der Gleichung**

| | | Regressionskoeffizient B | Standardfehler | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Schritt 1[a] | AGE | | | 1101,007 | 4 | ,000 | |
| | AGE(1) | ,289 | ,078 | 13,840 | 1 | ,000 | 1,335 |
| | AGE(2) | ,815 | ,073 | 123,584 | 1 | ,000 | 2,259 |
| | AGE(3) | 1,224 | ,068 | 324,924 | 1 | ,000 | 3,401 |
| | AGE(4) | 1,631 | ,064 | 644,390 | 1 | ,000 | 5,109 |
| | Konstante | -6,737 | ,059 | 12971,259 | 1 | ,000 | ,001 |

a. In Schritt 1 eingegebene Variablen: AGE.

0: kein Effekt, gleiches Risiko wie in RF

>0: Risiko, dass $\pi(x)=1$(zu sterben) ist höher als in RF

<0: Risiko, dass $\pi(x)=1$(zu sterben) ist niedriger als in RF

# Interpretation der Parameterwerte

## Referenzgruppe (RF) ist jüngste Altersgruppe 35-39

**Variablen in der Gleichung**

| | | Regressions koeffizientB | Standardf ehler | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Schritt 1[a] | AGE | | | 1101,007 | 4 | ,000 | |
| | AGE(1) | ,289 | ,078 | 13,840 | 1 | ,000 | 1,335 |
| | AGE(2) | ,815 | ,073 | 123,584 | 1 | ,000 | 2,259 |
| | AGE(3) | 1,224 | ,068 | 324,924 | 1 | ,000 | 3,401 |
| | AGE(4) | 1,631 | ,064 | 644,390 | 1 | ,000 | 5,109 |
| | Konstante | -6,737 | ,059 | 12971,259 | 1 | ,000 | ,001 |

a. In Schritt 1 eingegebene Variablen: AGE.

1: kein Effekt, gleiches Risiko wie RF

>1: Risiko, dass $\pi(x)=1$(zu sterben) ist höher als in RF

Age(1) hat ein um 33.5% höheres Risiko zu sterben als RF

<1: Risiko, dass $\pi(x)=1$(zu sterben) ist niedriger als in RF

# Interpretation der Parameterwerte

## Referenzgruppe (RF) ist älteste Altersgruppe 55-59

**Variablen in der Gleichung**

| | | Regressions koeffizientB | Standardf ehler | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Schritt 1[a] | AGE | | | 1101,007 | 4 | ,000 | |
| | AGE(1) | -1,631 | ,064 | 644,390 | 1 | ,000 | ,196 |
| | AGE(2) | -1,342 | ,056 | 566,466 | 1 | ,000 | ,261 |
| | AGE(3) | -,816 | ,050 | 266,037 | 1 | ,000 | ,442 |
| | AGE(4) | -,407 | ,042 | 94,998 | 1 | ,000 | ,666 |
| | Konstante | -5,106 | ,025 | 41403,541 | 1 | ,000 | ,006 |

a. In Schritt 1 eingegebene Variablen: AGE.

1: kein Effekt, gleiches Risiko wie RF

>1: Risiko, dass $\pi(x)=1$(zu sterben) ist höher als in RF

<1: Risiko, dass $\pi(x)=1$(zu sterben) ist niedriger als in RF

Age(1) hat ein um 80% niedrigeres Risiko zu sterben als die RF

(1-exp(B))*100

# Logistische Regression
# III.

# Parameter Interpretation
# Maximum Likelihood Schätzung
# Modell Testen

# Parameter Interpretation

# Odds Ratio in einer 2x2 Tabelle

| | Raucher | Nicht Raucher |
|---|---|---|
| gestorben | $p_1$ (0.30) | $p_2$ (0.25) |
| überlebt | $1-p_1$ (0.70) | $1-p_2$ (0.75) |

**Odds als Raucher zu sterben:**

**odds(1)= $p_1/(1-p_1)$ = .30/.70=0.43**

**Odds als Nichtraucher zu sterben:**

**odds(2)= $p_2/(1-p_2)$= .25./75=0.33**

**Odds ratio (1)= $p_1/(1-p_1)$ / $p_2/(1-p_2)$**

**.43 / .33 =1.29**

# Logistische Regression mit einer unabhängigen Variable (2x2 Tabelle)

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta x_1)}$$

Y=1…gestorben
Y=0…überlebt

X=1…Raucher
X=0…Nicht-Raucher

# Logistische Regression mit einer unabhängigen Variable

| Outcome Variable Y | Raucher x=1 | Nicht-Raucher x=0 |
|---|---|---|
| gestorben y=1 | $\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$ | $\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$ |
| überlebt y=0 | $1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$ | $1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$ |
| Total | 1.00 | 1.00 |

Independent Variable X

# Logistische Regression mit einer unabhängigen Variable

Log odds ratio

$$\ln \Psi = \ln\left[\frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))}\right]$$

Einsetzen aus Tabelle

$$\Psi = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}$$

$$= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

$\beta_1$ ist der Logarithmus des Odds ratios
$\exp(\beta_1)$ ist der Odds ratio

# Maximum Likelihood Schätzung der Parameter

# Maximum Likelihood Schätzung der Parameter

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$$

Parameterwerte $\beta_0$, $\beta_1$,.., $\beta_k$ so zu waehlen, dass die beobachtete Stichprobenvereilung am wahrscheinlichsten ist.

# Beispiel ML Schaetzung:

Von 10 untersuchten Personen haben 5 Symptome einer Herz-/Kreislauferkrankung.
Fuer welchen Wert ist das Zustandekommen der Stichprobenverteilung (x=5 Kranke auf n=10 Beobachtete) am wahrscheinlichsten?

$$P(x) = \frac{n!}{x!(n-x)!}\left[P^{x}(1-P)^{n-x}\right]$$

| P | P(x) |
|---|---|
| 0,1 | 0,001488 |
| 0,15 | 0,008491 |
| 0,2 | 0,026424 |
| 0,25 | 0,058399 |
| 0,3 | 0,102919 |
| 0,35 | 0,15357 |
| 0,4 | 0,200658 |
| 0,45 | 0,234033 |
| 0,5 | 0,246094 |
| 0,55 | 0,234033 |
| 0,6 | 0,200658 |
| 0,65 | 0,15357 |
| 0,7 | 0,102919 |
| 0,75 | 0,058399 |
| 0,8 | 0,026424 |
| 0,85 | 0,008491 |
| 0,9 | 0,001488 |
| 0,95 | 6,09E-05 |

# Schaetzen der Parameter:

Likelihood Funktion (LF)

$$LF = \Pi \left\{ P_i^{Y_i} (1 - P_i)^{1-Y_i} \right\} \qquad P = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}$$

$Y_i$.. Outcome Variable;
z.B. 0 wenn ueberlebt und 1 wenn gestorben

Jene Parameterwerte $\beta_0, \beta_1, \ldots, \beta_n$ sollen gesucht werden, die die Likelihoodfunktion LF maximieren.

# Schaetzen der Parameter:

Likelihood Funktion (LF)

$$LF = \Pi \left\{ P_i^{Y_i} (1 - P_i)^{1-Y_i} \right\} \qquad P = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}$$

$Y_i$.. Outcome Variable;
z.B. 0 wenn ueberlebt und 1 wenn gestorben

Jene Parameterwerte $\beta_0, \beta_1, \ldots, \beta_n$ sollen gesucht werden, die die Likelihoodfunktion LF maximieren.

# Schaetzen der Parameter:

Log Likelihood Funktion (LN (LF))

$$LN \quad LF \quad = \sum \left\{ \left[ Y_i LN \quad P_i \right] + \left[ (1 - Y_i) LN \quad (1 - P_i) \right] \right\}$$

$Y_i$.. Outcome Variable;
z.B. 0 wenn ueberlebt und 1 wenn gestorben

$$P_i = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n}}$$

soll maximiert werden.

- Die 1. Partiellem Ableitungen bilden
- Gleichungssystem gleich Null setzen
- Gleichung loesen

Kann nicht analytisch geloest werden
1. Startwerte, 2. Newton-Raphson Algorithmus

# Modell Testen

# Guete des logistischen Regressionsmodells

1. Je groesser die LN LF (je naeher bei Null)
   desto besser reproduzieren die Schaetzer der Parameterwerte die Stichprobenverteilung (desto besser das Modell)

Problem: LN LF haengt von Stichprobengroesse und Anzahl der Parameter ab

Log-Likelihood Test

Vergleich LN LF aktuelles Modell mit Baseline Modell (Modell ohne abhaengigen Variablen aber mit Konstante)

Nullhypothese: Alle Parameterwerte der Kovariaten sind gleich Null

# Guete des logistischen Regressionsmodells

Log-Likelihood Test

1. Vergleich LN LF aktuelles Modell mit Baseline Modell (Modell ohne abhaengigen Variablen aber mit Konstante)

Nullhypothese: Alle Parameterwerte der Kovariaten sind gleich Null

---

**$G=-2(LN_0-LN_1)$**

**$LN_0$ Modell ohne Kovariaten $LN_1$ Modell mit Kovariaten**

**G ist $\chi^2$ verteilt**

**DF=Anzahl der Kovariaten**

# Guete des logistischen Regressionsmodells

Log-Likelihood Test

2. Vergleich LN LF aktuelles Modell mit vorhergehendem Modell

Nullhypothese: Alle Parameterwerte der neu integrierten Kovariaten sind gleich Null

**$G=-2(LN_n-LN_{n+k})$**

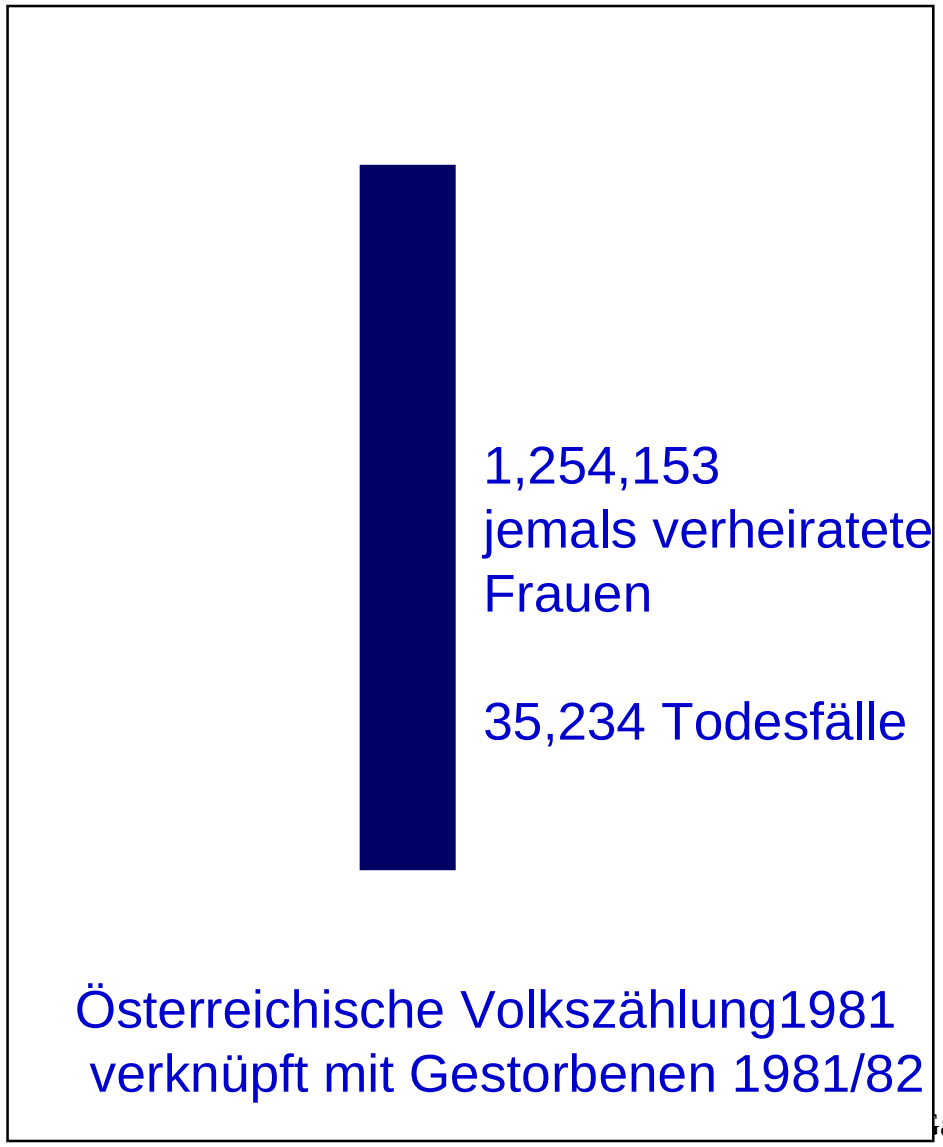
**$LN_n$ Modell mit n Kovariaten, $LN_{n+k}$ Modell mit n+k Kovariaten**

**G ist $\chi^2$ verteilt**

**DF=Anzahl der k neu integrierten Kovariaten**

# SPSS

# Oesterreichische Volkszaehlung und Gestorbenen Daten




F35icdn.sav
F60icdn.sav
M35icdn.sav
M60icdn.sav

Label file:
Value Labels Österreichische Gestorbenen Daten.doc

# Ausblick

- Interaktionseffekte
- Wie gehe ich meine Studie an?
- 4 Gruppen: (1) Bildung
  (2) Soziooekonomischer Status
  (3) Kinder
  (4) Familienstand

4. Wie gehe ich meine Modellierung an
5. Literatur

Praesentation der Ergebnisse und Diskussion der Angewandten Modellierungsstrategien in den letzten beiden Einheiten