

## Logistic Regression

### Coronary Heart Disease<sup>1</sup> (CHD)

File: **chdage.sav**

**Pilot study:**

Hypothetical data from an explorative study at a hospital.

100 patients randomly drawn from the hospital's database were checked for coronary heart disease symptoms.

**Objective:**

Try to quantify the risk of coronary heart disease symptoms.

Use Crosstable and Logistic Regression techniques.

### High School and Beyond<sup>2</sup>

File: **hsb2.sav**

**Sample Survey:**

200 students drawn randomly from the first wave of High School and Beyond (HSB), a longitudinal study of American youth conducted by the National Opinion Research Center on behalf of the National Center for Education Statistics (NCES). Data were collected from 58,270 high school students (28,240 seniors and 30,030 sophomores) and 1,015 secondary schools in the spring of 1980.

**Objective:**

Try to determine factors that are associated with a high write score ( $\geq 52$ )

Start with female and read

---

<sup>1</sup> Example: Hosmer and Lemeshow (2000) Applied Logistic Regression: Second Edition, pp.1-10

<sup>2</sup> Example from UCLA . Data: <http://www.ats.ucla.edu/stat/stata/notes/hsb2.dta>  
see full discussion for this example at UCLA: [http://www.ats.ucla.edu/stat/stata/seminars/stata\\_logistic/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/default.htm)

## Artificial Respiration<sup>3</sup>

File: **ArtificialRespiration.xls**

### Explorative study:

1989, in January 63 of 131 patients died while treated at the intensive care ward under artificial respiration.

### Objective:

Try to generate a hypothesis to be tested in a later study.

Variable	Variable label
patient	Patient ID
result	Result
gender	Gender
age	Age
height	Height
cause	Cause
duration	Duration of artificial respiration (h)
oxy_conc	Oxygen concentration
intens	Intensity of artificial respiration

Value	Value label
result 0	died
1	survived
gender 1	male
2	female
cause 1	Pneumonia
2	Accident
3	other

<sup>3</sup> Modified example from: Bühl A (2006). SPSS 14. pp. 378-381

## Low birth weight study<sup>4</sup>

File: **bwght mod learn.sav**

### Sampling:

Data from an observational study at a hospital in 1989  
 All births of a year were recorded (n=1388)

### Objective:

Clarify the influence of smoking on the birth weight of a child.

codebook:

Variable	Label	Measurement Level
faminc	1988 family income, \$1000s	Scale
cigtax	cig. tax in home state, 1988	Scale
cigprice	cig. price in home state, 1988	Scale
bwght	birth weight, ounces	Scale
fatheduc	father's yrs of educ	Scale
motheduc	mother's yrs of educ	Scale
parity	birth order of child	Scale
male	=1 if male child	Nominal
white	=1 if white	Nominal
cigs	cigs smoked per day while preg	Scale
packs	packs smoked per day while preg	Scale

Value	Label
male 0	female child
1	male child
white 0	non white
1	white

1. Create new variables:

Variable	Label	Measurement Level
bwgr	birth weight, gramm	Scale
smoked	mother smoked while preg	Nominal
lowbwg	low birth weight (<=2500 gr)	Nominal

Value	Label
smoked 0	no
1	yes
lowbwg 0	no
1	yes

2. Explore whether smokers and non smokers are comparable in respect to other factors that might influence the birth weight
3. Explore the correlations between all factors
4. Perform an ANOVA
5. Perform an Multiple Regression (include collinearity tests)
6. Perform a Logistic Regression

<sup>4</sup> Modified example from: Wooldridge, J. (2003) Introductory econometrics: a modern approach.  
 1 ounce = 28,349523 gr  
 Normal birthweight values (male av 3400gr; female av 3200 gr)  
 Low birthweight: 2500gr