Data analysis with R software

Data-analyysi R-ohjelmistolla

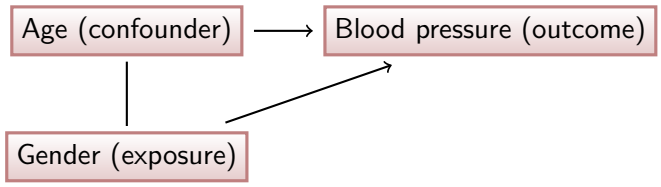
Tommi Härkänen

National Institute for Health and Welfare (THL), Helsinki
E-mail: tommi.harkanen@helsinki.fi

University of Helsinki, February 15, 2012

Contents

Confounding effect



Simulated example

Assume that we have a data set in which

- women have 10 mmHg lower blood pressure (BP) than men (of the same age),
- distribution of age in
  - men is uniform over 50 to 70 years,
  - women is uniform over 50 to 80 years and
- the association of age and BP is +3 mmHg per year.

Descriptive statistics

Overall means:

```
  gender mean.bp mean.age sd.bp sd.age
1 Female   135.7     65.2  27.9    8.7
2   Male   130.1     60.1  20.0    5.8
```

Note here, that the gender differences within age groups are different from the the overall difference:

```
  gender  age.10 mean.bp sd.bp
1 Female [50,60)   105.2  12.6
2 Female [60,70)   135.4  13.4
3 Female [70,80]   165.2  13.3
4   Male [50,60)   114.8  13.0
5   Male [60,70)   145.1  13.2
```

The oldest age group of women has the highest average BP, which has strong influence on the overall means.

Data analysis with R software
└Adjustment for a confounder

Regression analyses

The unadjusted regression analysis:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    135.67       0.34   393.7        0
genderMale      -5.54       0.49   -11.4        0
```

Adjust for the confounder (age):

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -60.32       0.90  -66.86        0
genderMale      9.94       0.21   46.65        0
age             3.01       0.01  220.00        0
```

This gives the correct result on the difference of genders (10).