

Statistics 305: Introduction to Biostatistical Methods for Health Sciences

R Demo for Chapter 17: Correlation

Jinko Graham

Example Data: Child Mortality by Country

- Data on child mortality (number of deaths before age 5 years, per 1000 live births) and percentage of children who are immunized for diptheria, pertussis and tetanus (DPT) from a random sample of 20 countries (see Table 17.2 of text).

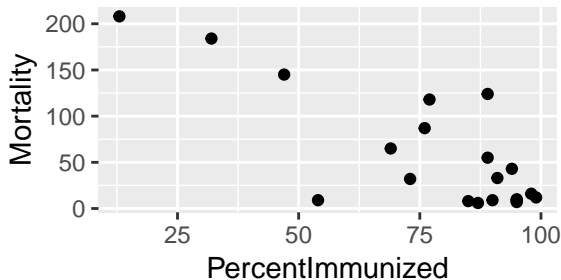
```
uu <- url("http://people.stat.sfu.ca/~jgraham/Teaching/S305_17/Data/mort.txt")
childmort <- read.table(uu,header=TRUE)
head(childmort)
```

##	Nation	PercentImmunized	Mortality	Region
## 1	Bolivia	77	118	SouthAmer
## 2	Brazil	69	65	SouthAmer
## 3	Cambodia	32	184	Asia
## 4	Canada	85	8	NorthAmer
## 5	China	94	43	Asia
## 6	CzechRepub	99	12	Europe

Scatterplots

- ▶ `ggplot()` used before to draw histograms and boxplots.
 - ▶ When given quantitative variables it will draw a scatterplot.
- ▶ The `aes()` specifies which variable to go on the x (horizontal) and y (vertical) axes.
- ▶ `geom_point()` adds the points to the scatterplot.

```
library(ggplot2)
ggplot(childmort, aes(x=PercentImmunized, y=Mortality)) +
  geom_point()
```



Pearson Correlation in R

```
with(childmort,cor(PercentImmunized,Mortality))
```

```
## [1] -0.7910654
```

- ▶ Sample Pearson correlation between PercentImmunized and Mortality is negative: $r = -0.791$.

- ▶ Test if the Pearson population correlation differs from zero
 - ▶ i.e. Test $H_0 : \rho = 0$ vs. $H_a : \rho \neq 0$.

```
with(childmort, cor.test(PercentImmunized, Mortality))
```

```
##  
## Pearson's product-moment correlation  
##  
## data: PercentImmunized and Mortality  
## t = -5.4864, df = 18, p-value = 3.281e-05  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.9137250 -0.5362744  
## sample estimates:  
## cor  
## -0.7910654
```

- ▶ Reject H_0 at level 0.05; there is statistical evidence of correlation between a country's immunization rate and mortality.
 - ▶ In particular, Mortality and PercentImmunized appear to be negatively correlated.

Spearman Correlation in R

```
with(childmort, cor(PercentImmunized, Mortality, method="spearman"))
```

```
## [1] -0.5431913
```

- ▶ Note: Spearman's $r_s = -0.54$ is closer to zero than Pearson's $r = -0.79$.
- ▶ Next, test whether the population Spearman correlation coefficient differs from zero; i.e., test $H_0 : \rho_s = 0$ vs. $H_a : \rho_s \neq 0$.

► Test Spearman's correlation coefficient:

```
with(childmort,cor.test(PercentImmunized,Mortality,  
                        method="spearman",exact=FALSE))
```

```
##  
## Spearman's rank correlation rho  
##  
## data: PercentImmunized and Mortality  
## S = 2052.4, p-value = 0.01332  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## -0.5431913
```

- By default, `cor.test()` uses another null distribution than the t -distribution suggested by the text.
- To disable the default and use the t -distribution to get the p -value as in the text, set `exact=FALSE`.

```
##  
## Spearman's rank correlation rho  
##  
## data: PercentImmunized and Mortality  
## S = 2052.4, p-value = 0.01332  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
##      rho  
## -0.5431913
```

- ▶ There is statistical evidence that the population Spearman correlation is not 0 (at level $\alpha = 0.05$).
- ▶ Mortality and PercentImmunized appear to be negatively correlated, even when we take into account outlying countries through the Spearman's rank-correlation test.