

T-Tests (Part - 2)

Example: Comparing Crime Rates in Southern states and other states

Data

Consider the UScrime dataset distributed with the package MASS.

Data Description

Criminologists are interested in the effect of punishment regimes on crime rates. This dataset contains information about the effect of punishment regimes on crime rates in 47 US states in 1960.

Variables

Prob: the probability of imprisonment

So : a categorical variable (indicator variable) for Southern states

U1 : the unemployment rate for urban males, aged 14-24

U2 : the unemployment rate for urban males, aged 35-39

Additional Variables:

(<https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/UScrime.html>)

Example: Compare the probability of imprisonment in Southern states and other states.

Here, the outcome variable is continuous (probability Prob of imprisonment).

Here, the grouping variable is categorical (Southern states versus other states).

```
library(MASS)
library(psych)
```

```
attach(UScrime)
```

Explore the data

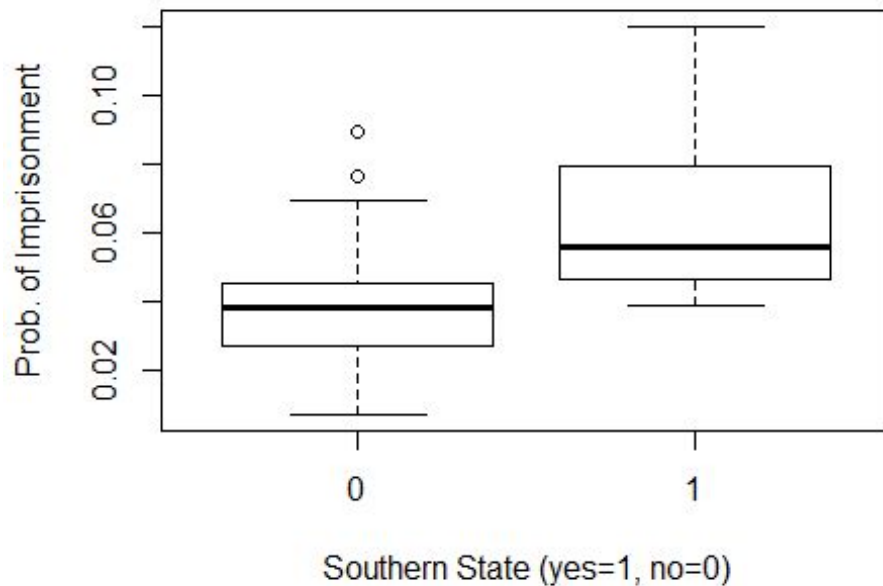
```
table(So)
```

```
## So
##  0  1
## 31 16
```

```
summary(Prob)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00690 0.03270 0.04210 0.04709 0.05445 0.11980
```

```
boxplot(Prob ~ So, data=UScrime, xlab="Southern State (yes=1, no=0)",
ylab="Prob. of Imprisonment")
```



```
aggregate(UScrime$Prob, by=list(UScrime$So), FUN=mean)
```

```
##   Group.1      x
## 1      0 0.03851265
## 2      1 0.06371269
```

Hypothesis 1

H1: The probability of going to jail in a southern state is higher than the probability of jail in other states

Independent t-test

A two-group independent t-test can be used to test the hypothesis that the two population means are equal. Here, we assume that the two groups are independent and that the data is sampled from normal populations.

Format

The format of the R command is either `t.test(y ~ x, data)` where `y` is numeric and `x` is a dichotomous variable, or `t.test(y1, y2)` where `y1` and `y2` are numeric vectors (the outcome variable for each group)

Technical Points

- The default test assumes unequal variance.
- We can add a `var.equal=TRUE` option to specify equal variances and a pooled variance estimate.

- It applies the Welch degrees-of-freedom modification.
- By default, a two-tailed alternative is assumed (i.e. the means differ, but the direction is not specified).
- We can add the option `alternative="less"` or `alternative="greater"` to specify a directional test.

Testing Hypothesis 1

The following code compares Southern (group 1) and non-Southern (group 0) states on the probability of imprisonment, using a two-tailed test without the assumption of equal variances:

```
t.test(Prob ~ So, data=UScrime)

##
##  Welch Two Sample t-test
##
## data:  Prob by So
## t = -3.8954, df = 24.925, p-value = 0.0006506
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03852569 -0.01187439
## sample estimates:
## mean in group 0 mean in group 1
##      0.03851265      0.06371269
```

Result of Independent t-test

Based on the above output of the t-test, we can reject the hypothesis that Southern states and non-Southern states have equal probabilities of imprisonment ($p < 0.001$)