02441 Applied Statistics and Statistical Software

Exercise 1D - laborforce

This dataset contains the labor force participation rate (LFPR) of women in 19 cities in the United States in each of two years (1968 and 1972). The data help to measure the growing presence of women in the labor force over this period.

Variable name	Description
city	City in the US
1972	Labor Force Participation rate of women in 1972
1968	Labor Force Participation rate of women in 1968

1. Compare LFPR rates in the two years with a pooled t-test since the United States did not change much from 1968 to 1972

Start by loading data:

```
lf <- read.table('labor.txt', header = TRUE)
```

If we suspect that the variance is roughly the same for the groups of 1968 and 1972 we can use a Pooled t-test in R

```
var.test(lf$x1972, lf$x1968)
```

```
##
##  F test to compare two variances
##
## data:  lf$x1972 and lf$x1968
## F = 1.0841, num df = 18, denom df = 18, p-value = 0.8658
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4176788 2.8139438
## sample estimates:
## ratio of variances
##           1.084124
```

```
t.test(lf$x1972, lf$x1968, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  lf$x1972 and lf$x1968
## t = 1.4959, df = 36, p-value = 0.1434
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01198526  0.07935368
## sample estimates:
## mean of x mean of y
## 0.5268421 0.4931579
```

The pooled t-test indicates that the quantitative presence of women in labor force is not significantly different in the years 1972 and 1968.

2. Compare LFPR rates in the two years with a paired t-test

Paired T-test in R

```
t.test(lf$x1972, lf$x1968, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  lf$x1972 and lf$x1968
## t = 2.4577, df = 18, p-value = 0.02435
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.004889895 0.062478527
## sample estimates:
## mean of the differences
##              0.03368421
```

The paired t-test indicates that the quantitative presence of women in labor force is significantly different in the years 1972 and 1968.

3. Which of the two methods you have used do you find most appropriate?

The paired t-test is more powerful. Every city is its own control, meaning that cities can perform very different overall, but when taking city dependent differences into account the difference can be highly significant.

4. If you have found a significant change - how big is this change then?

The quantitative presence of women in labor force decreased from 1968 to 1972 by 0.03 units (95% CI:[0.0049;0.0625])