# Class 19: The Last One

Sean Westwood

## Practice

You are responsible for reviewing terms from the book and lecture. Today we will practice some analysis. You will work on the practice by yourself. I will answer questions as if this were an exam.

**Time management will be important during the final. You should be able to answer the questions below in less than 60 minutes**

### Sex Ratio and the Price of Agricultural Crops in China

In this exercise, we consider the effect of a change in the price of agricultural goods whose production and cultivation are dominated by either men or women.

This exercise is based on: Qian, Nancy. 2008. "Missing Women and the Price of Tea in China:The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics* 123(3): 1251–85.

Our data come from China, where centrally planned production targets during the Maoist era led to changes in the prices of major staple crops. We focus here on tea, the production and cultivation of which required a large female labor force, as well as orchard fruits, for which the labor force was overwhelmingly male. We use price increases brought on by government policy change in 1979 as a proxy for increases in sex-specific income, and ask the following question: Do changes in sex-specific income alter the incentives for Chinese families to have children of one gender over another?

| Name | Description |
|---|---|
| `birpop` | Birth population in a given year |
| `biryr` | Year of cohort (birth year) |
| `cashcrop` | Amount of cash crops planted in county |
| `orch` | Amount of orchard-type crops planted in county |

| Name | Description |
|---|---|
| teasown | Amount of tea sown in county |
| sex | Proportion of males in birth cohort |
| post | Indicator variable for introduction of price reforms (0 before and 1 after) |
| han | Proportion Han Chinese |

```
women <- read.csv("chinawomen.csv", header = TRUE)
```

## Question 1

1.1. Compute the mean sex ratio (proportion of males) in 1985 separately for tea-producing (>0)/non-orchard (==0) regions and orchard(>0)/non-tea-producing (==0) regions, then take the difference in means for the two groups.

```
women$teadum <- 0
women$orchdum <- 0
women$teadum[women$teasown > 0 & women$orch == 0] <- 1
women$orchdum[women$orch > 0 & women$teasown == 0] <- 1

women.after <- women[women$biryr == 1985, ]

mean(women.after$sex[women.after$teadum == 1]) - mean(women.after$sex[women.after$orchdum ==
    1])
```

```
[1] 0.006962526
```

1.2. Now compute the 95% CIs as well as a the p-value for this difference.

```
t.test(women.after$sex[women.after$teadum == 1], women.after$sex[women.after$orchdum ==
    1])
```

```
    Welch Two Sample t-test

data:  women.after$sex[women.after$teadum == 1] and women.after$sex[women.after$orchdum == 1]
t = 0.45608, df = 35.177, p-value = 0.6511
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02402376  0.03794881
```

```
sample estimates:
mean of x mean of y
0.5305382 0.5235757
```

1.3. Repeat the analysis above (1.1 and 1.2) for 1983.

```r
women.before <- women[women$biryr == 1983, ]

mean(women.before$sex[women.before$teadum == 1]) - mean(women.before$sex[women.before$orchdum
    1])
```

```
[1] 0.01764056
```

```r
t.test(women.before$sex[women.before$teadum == 1], women.before$sex[women.before$orchdum ==
    1])
```

```
    Welch Two Sample t-test

data:  women.before$sex[women.before$teadum == 1] and women.before$sex[women.before$orchdum =
t = 0.83641, df = 33.781, p-value = 0.4088
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02523132  0.06051243
sample estimates:
mean of x mean of y
0.5390109 0.5213703
```

## Question 2

2.1. Compute the mean difference in sex between the two types of regions for each of the years between 1981 and 1989. In which year was the absolute difference the largest? Hint: you can use the abs() function.

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1981]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1981]))
```

```
[1] 0.0338322
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1982]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1982]))
```

```
[1] 0.01204438
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1983]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1983]))
```

```
[1] 0.01764056
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1984]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1984]))
```

```
[1] 0.001888236
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1985]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1985]))
```

```
[1] 0.006962526
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1986]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1986]))
```

```
[1] 0.006666828
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1987]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1987]))
```

```
[1] 0.004672121
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1988]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1988]))
```

```
[1] 0.03589176
```

```r
abs(mean(women$sex[women$teadum == 1 & women$biryr == 1989]) -
    mean(women$sex[women$orchdum == 1 & women$biryr == 1989]))
```

```
[1] 0.00964766
```

## Question 3

3.1. Now compare the population of Han Chinese in tea-producing and orchard-producing regions before the policy enactment. Specifically, examine the proportion of Han Chinese in 1978. Estimate the mean difference and 95% confidence intervals for the two regions.

```r
women.1978 <- women[women$biryr == 1978, ]

t.test(women.1978$han[women.1978$teadum == 1], women.1978$han[women.1978$orchdum ==
    1])
```

```

    Welch Two Sample t-test

data:  women.1978$han[women.1978$teadum == 1] and women.1978$han[women.1978$orchdum == 1]
t = 1.9597, df = 35.667, p-value = 0.05788
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.002583257  0.149234096
sample estimates:
mean of x mean of y
0.9523256 0.8790002
```

3.2. Do we reject the null hypothesis of no difference in Han population?

## Question 4

4.1. Compute the mean difference in Han population for each year before the reform in the two regions, i.e., from 1962 until 1978 (a for loop might be easiest, but it is not required).

```r
pre_reform_data <- women[women$biryr >= 1962 & women$biryr <=
    1978, ]

tapply(pre_reform_data$han[pre_reform_data$teadum == 1], pre_reform_data$biryr[pre_reform_da
```

```r
    1], mean) - tapply(pre_reform_data$han[pre_reform_data$orchdum ==
    1], pre_reform_data$biryr[pre_reform_data$orchdum == 1],
    mean)
```

```
      1962       1963       1964       1965       1966       1967       1968
0.06949663 0.06404637 0.06890834 0.07005651 0.06847232 0.07136858 0.06935931
      1969       1970       1971       1972       1973       1974       1975
0.06930467 0.07056496 0.06807626 0.06718398 0.06896967 0.07080438 0.06949562
      1976       1977       1978
0.06470800 0.07813199 0.07332542
```

```r
for (i in 1962:1978) {
    women.s <- women[women$biryr == i, ]
    print(mean(women.s$han[women.s$teadum == 1]) - mean(women.s$han[women.s$orchdum ==
        1]))
}
```

```
[1] 0.06949663
[1] 0.06404637
[1] 0.06890834
[1] 0.07005651
[1] 0.06847232
[1] 0.07136858
[1] 0.06935931
[1] 0.06930467
[1] 0.07056496
[1] 0.06807626
[1] 0.06718398
[1] 0.06896967
[1] 0.07080438
[1] 0.06949562
[1] 0.064708
[1] 0.07813199
[1] 0.07332542
```

## Question 5

5.1. Now, explore the relationship between sex ratios (DV) and post, teasown, orch and biryr. To do this use a linear regression model.

```r
lm(sex ~ post + teasown + orch + biryr, data = women)
```

```
Call:
lm(formula = sex ~ post + teasown + orch + biryr, data = women)

Coefficients:
(Intercept)         post      teasown         orch        biryr
  -2.240238    -0.002178     0.006276     0.002400     0.001393
```

5.2. Given our model, what would we expect the sex ratio to be post-treatment in 1981 for a region with the mean amount of teasown (no orchards)?

```r
-2.240238 + -0.002178 + 0.006276 * mean(women$teasown) + 0.0024 *
    0 + 0.001393 * 1981
```

```
[1] 0.5173229
```

5.3. What is the 95% CI for orch?

```r
confint(lm(sex ~ post + teasown + orch + biryr, data = women))
```

```
                    2.5 %       97.5 %
(Intercept) -2.7547309016 -1.725744805
post        -0.0066535614  0.002298241
teasown     -0.0002229987  0.012775745
orch        -0.0003167762  0.005117342
biryr        0.0011313683  0.001653718
```

## Evaluation

1. On a scale ranging between 1 (Too Hard) and 10 (Too Easy), how was today's class:
2. What was the easiest thing to understand?
3. What was the most difficult thing to understand?
4. How long did you spend on the assignment outside of class?