# Chapter 6 Lab

*Tianwei Liu*

*10/30/19*

## Preparation

```
require(knitr)
require(haven)
require(car)
require(AER)
library(stargazer)

opts_chunk$set(echo = TRUE)
options(digits = 6)
```

```
load("Chapter6_Lab.RData")
```

We will use General Social Survey data for this lab. The key variables we will use are

- happy: 3 very happy, 2 pretty happy, 1 not too happy
- married: 1 for currently married (based on dta$marital, which is 1 for married, 2 for widowed, 3 for divorced, 4 for separated, 5 for never married)
- sex: 1 for men, 2 for women
- edcat: 1 for less than high school, 2 for high school, 3 for some college and 4 for college graduate
- race: 1 for white, 2 for black, 3 for other races (self-identified, first race mentioned)

**(a) Use OLS to estimate the difference in happiness (the "happy" variable) for married versus unmarried people. Then use the t test function to assess the difference in happiness by married status (see the Computing Corner in Chapter 6 of the book). Discuss similarities and differences in OLS model and t test function results**

```
reg1 <- lm(happy ~ married, data = dta)
summary (reg1)
```

```
##
## Call:
## lm(formula = happy ~ married, data = dta)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3227 -0.3227 -0.0249  0.6773  0.9751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.02490    0.00603   335.6   <2e-16 ***
## married      0.29781    0.00762    39.1   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.598 on 26346 degrees of freedom
##   (2442 observations deleted due to missingness)
## Multiple R-squared:  0.0548, Adjusted R-squared:  0.0547
## F-statistic: 1.53e+03 on 1 and 26346 DF,  p-value: <2e-16
```

```r
t.test(happy ~ married, data = dta, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  happy by married
## t = -39.08, df = 26346, p-value <2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.312748 -0.282872
## sample estimates:
## mean in group 0 mean in group 1
##         2.02490         2.32271
```

```r
summary(reg1)$coefficients[1] + summary(reg1)$coefficients[2]
```

```
## [1] 2.32271
```

The t-test function gives the same result as our OLS model.

Since "married" is a binary variable, in the regression model of "happy" on "married", the intercept represents the average happiness of unmarried people, and the slope coefficient stands for the mean difference in happiness between married and unmarried people. We see that the difference in the t test is the same as the coefficient on married in the OLS model and that the t statistic and p-value are the same in both methods.

**(b) Use OLS with robust standard errors to estimate the difference in happiness for married and unmarried people. Then use the t test function with unequal variances to assess the difference in happiness by marital status. Discuss similarities and differences in OLS model and t test function results.**

```r
coeftest(reg1, vcov. = vcovHC(reg1, type = "HC1"))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.024903   0.006071  333.54   <2e-16 ***
## married     0.297810   0.007641   38.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
t.test(happy ~ married, data = dta, var.equal = FALSE) # t-test with unequal variance
```

```
##
##  Welch Two Sample t-test
##
## data:  happy by married
## t = -38.98, df = 20513, p-value <2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.312787 -0.282834
## sample estimates:
## mean in group 0 mean in group 1
##         2.02490         2.32271
```

The results are consistent across two methods we used above.

Using robust standard errors (which in this case equals to assuming heteroskedasticity) does not change the estimated difference of means between married and unmarried people. It does change the standard errors, but does so in the same way that the t test with unequal variance does, producing the same t statistics and t tests across the OLS and t test methods, as before.

**(c) Create an interaction between married and age. Estimate a model that explains happiness as a function of age and marital status, allowing for the age effect to differ according to marital status. (For simplicity, we use only married and unmarried for marital status.) What is the effect of age for unmarried people? For married people?**

```r
dta$marriedage <- dta$married * dta$age
reg2 <- lm (happy ~ age + married + marriedage, data = dta)
summary(reg2)
```

```
##
## Call:
## lm(formula = happy ~ age + married + marriedage, data = dta)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3292 -0.3241 -0.0483  0.6745  1.0222
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.130031   0.027026   78.81  < 2e-16 ***
## age         -0.002820   0.000707   -3.99  6.6e-05 ***
## married      0.210952   0.035104    6.01  1.9e-09 ***
## marriedage   0.002350   0.000904    2.60   0.0093 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.598 on 26344 degrees of freedom
##    (2442 observations deleted due to missingness)
## Multiple R-squared:  0.0554, Adjusted R-squared:  0.0553
## F-statistic:  515 on 3 and 26344 DF,  p-value: <2e-16
```

The model is : Happiness $= B0 + B1*$ age $+ B2 *$ married $+ B3 *$ age $*$ married $+$ error_term In the model, the effect of age is -0.00282 (B3) for unmmarried people and $B1 + B3 = -0.00282 + 0.002350 = -0.00047$ for married people. The intercept is 2.13 (B0) for unmarried people and $B0 + B2 = 2.13 + 0.2109 = 2.3409$ for married people.

3

**(d) Estimate separate models explaining happiness in terms of age for married and unmarried people. Comment on similarities and differences compared to results immediately above for (i) the estimated effect of age and (ii) the intercept.**

```
reg2.unmarried <- lm(happy ~ age, data = dta[dta$married == 0,])
reg2.married <- lm(happy ~ age, data = dta[dta$married == 1,])
```

```
summary (reg2.unmarried)
```

```
##
## Call:
## lm(formula = happy ~ age, data = dta[dta$married == 0, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0595 -0.0539 -0.0257  0.0110  1.0222
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.130031   0.027184   78.35  < 2e-16 ***
## age         -0.002820   0.000711   -3.97  7.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.602 on 9836 degrees of freedom
##   (1057 observations deleted due to missingness)
## Multiple R-squared:  0.0016, Adjusted R-squared:  0.0015
## F-statistic: 15.7 on 1 and 9836 DF,  p-value: 7.32e-05
```

We see the same results as above: the effect of age is -0.00282 for unmarried people, with an intercept of 2.130.

```
summary(reg2.married)
```

```
##
## Call:
## lm(formula = happy ~ age, data = dta[dta$married == 1, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.329 -0.325 -0.319  0.675  0.684
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.340983   0.022324  104.86   <2e-16 ***
## age         -0.000470   0.000561   -0.84      0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.596 on 16508 degrees of freedom
##   (1389 observations deleted due to missingness)
## Multiple R-squared:  4.24e-05,   Adjusted R-squared:  -1.82e-05
## F-statistic:  0.7 on 1 and 16508 DF,  p-value: 0.403
```

The effect of age for married people is -0.000469, with and intercept of 2.3409.

This is a general phenomenon: if we include all variables interacted with a dummy variable for a given group, we will get the same results as when estimating a model limited only to that group.

**(e) Marianne Bertrand wrote an article called "Work on Women's Work is Never Done: Career, Family, and the Well-Being of College-Educated Women" published in the *American Economic Review: Papers & Proceedings* 2013, 103(3): 244–250. She analyzed the effect of careers and family on college-educated women.**

**She defines a career variable that is 1 if someone's earnings are above the twenty-fifth percentile in the relevant year and age group.**

**Estimate a model in which happiness is a function of career, being married and an interaction of career and married ("careermarried"). To match Betrand's analysis, limit the data to only to college educated (dta$educat==4) women (dta$sex==2).**

**Interpret the estimated average happiness for the four types of women implied by this analysis.**

```
reg3 <- lm(happy  ~ career + married + careermarried, data=dta[dta$sex==2 & dta$educat==4,])
summary (reg3)
```

```
##
## Call:
## lm(formula = happy ~ career + married + careermarried, data = dta[dta$sex ==
##     2 & dta$educat == 4, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.421 -0.421 -0.125  0.579  0.875
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.1247     0.0200  106.41   <2e-16 ***
## career         0.0920     0.0298    3.08   0.0021 **
## married        0.2968     0.0250   11.85   <2e-16 ***
## careermarried -0.0992     0.0402   -2.47   0.0137 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.574 on 3595 degrees of freedom
##   (400 observations deleted due to missingness)
## Multiple R-squared:  0.0476, Adjusted R-squared:  0.0469
## F-statistic:   60 on 3 and 3595 DF,  p-value: <2e-16
```

- College-educated women with no careers and not married have average happiness of 2.12.
- College-educated women with careers and not married have average happiness of $2.12 + 0.09 = 2.21$.

- College-educated women with career equal zero and married have average happiness of $2.12 + 0.29 = 2.31$.

- College-educated women with careers and married have average happiness of $2.12 + 0.092 + 0.29 - 0.099 = 2.30$.

**(f) The GSS provides a race variable (dta$race). The variable equals 1 for white respondents, 2 for black respondents and 3 for everyone else. Add race to the above model and interpret the coefficients related to race. Estimate another model with a different reference category and explain coefficients across the two models.**

```
reg4 <- lm(happy  ~ career + married + careermarried + factor(race), data=dta[dta$sex==2 & dta$educat==4
summary (reg4)
```

```
##
## Call:
## lm(formula = happy ~ career + married + careermarried + factor(race),
##      data = dta[dta$sex == 2 & dta$educat == 4, ])
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -1.436 -0.433 -0.158   0.564   1.008
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.1575     0.0208  103.87   <2e-16 ***
## career           0.0883     0.0297    2.97    0.003 **
## married          0.2785     0.0252   11.06   <2e-16 ***
## careermarried   -0.0913     0.0401   -2.28    0.023 *
## factor(race)2   -0.1659     0.0327   -5.08    4e-07 ***
## factor(race)3   -0.1005     0.0404   -2.49    0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.572 on 3593 degrees of freedom
##    (400 observations deleted due to missingness)
## Multiple R-squared:  0.0556, Adjusted R-squared:  0.0542
## F-statistic: 42.3 on 5 and 3593 DF,  p-value: <2e-16
```

```
dta$race2 <- relevel(factor(dta$race), ref = "2")
reg5 <- lm(happy  ~ career + married + careermarried + factor(race2), data=dta[dta$sex==2 & dta$educat==
summary(reg5)
```

```
##
## Call:
## lm(formula = happy ~ career + married + careermarried + factor(race2),
##      data = dta[dta$sex == 2 & dta$educat == 4, ])
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -1.436 -0.433 -0.158   0.564   1.008
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.9915     0.0338   58.83   <2e-16 ***
```

```
## career          0.0883      0.0297     2.97     0.003 **
## married          0.2785      0.0252    11.06    <2e-16 ***
## careermarried   -0.0913      0.0401    -2.28     0.023 *
## factor(race2)1   0.1659      0.0327     5.08     4e-07 ***
## factor(race2)3   0.0654      0.0497     1.32     0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.572 on 3593 degrees of freedom
##   (400 observations deleted due to missingness)
## Multiple R-squared:  0.0556, Adjusted R-squared:  0.0542
## F-statistic: 42.3 on 5 and 3593 DF,  p-value: <2e-16
```

Changing the referecen category does not lead to changes in the effects. White college-educted woman are in general the happiest, when holding all else equal. Black college-educated women are 0.17 less happy than white counterparts; other races are reportedly 0.10 less happy. The results are the same when changing the reference category, we just need to add the intercept and the coefficients up to get the same numbers.

**(g): Estimate a model similar to the above, but for a different group of your choice (e.g., limit by race or gender or education) and feel free to include different covariates.**

```
reg6 <- lm(happy  ~ career + married + careermarried + factor(race), data=dta[dta$sex==2 & (dta$educat==
summary (reg6)
```

```
##
## Call:
## lm(formula = happy ~ career + married + careermarried + factor(race),
##     data = dta[dta$sex == 2 & (dta$educat == 4 | dta$educat ==
##         3), ])
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -1.418 -0.404 -0.102  0.596  1.091
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.1023     0.0147  142.90  < 2e-16 ***
## career          0.1033     0.0215    4.81  1.5e-06 ***
## married         0.3020     0.0175   17.24  < 2e-16 ***
## careermarried  -0.0900     0.0294   -3.06   0.0022 **
## factor(race)2  -0.1933     0.0198   -9.77  < 2e-16 ***
## factor(race)3  -0.0772     0.0306   -2.52   0.0116 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.579 on 7258 degrees of freedom
##   (835 observations deleted due to missingness)
## Multiple R-squared:  0.073,  Adjusted R-squared:  0.0724
## F-statistic:  114 on 5 and 7258 DF,  p-value: <2e-16
```

For this model limited by women who have attended at least some college, it tells us that among women who attended at least some coolege, white women are the happiest, black women and women of other races are less happy, by 0.19 and 0.077 respectively.

```
reg7 <- lm(happy  ~ career + married + careermarried + factor(race), data=dta[dta$sex==1 & (dta$educat==
summary (reg7)
```

```
##
## Call:
## lm(formula = happy ~ career + married + careermarried + factor(race),
##     data = dta[dta$sex == 1 & (dta$educat == 4 | dta$educat ==
##         3), ])
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -1.367 -0.367 -0.133  0.632  1.065
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.0469     0.0177  115.50  < 2e-16 ***
## career           0.0858     0.0231    3.71  0.00021 ***
## married          0.2595     0.0228   11.40  < 2e-16 ***
## careermarried   -0.0248     0.0297   -0.83  0.40406
## factor(race)2   -0.1115     0.0252   -4.43  9.5e-06 ***
## factor(race)3   -0.0543     0.0311   -1.74  0.08125 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.568 on 6418 degrees of freedom
##   (650 observations deleted due to missingness)
## Multiple R-squared:  0.0523, Adjusted R-squared:  0.0515
## F-statistic: 70.8 on 5 and 6418 DF,  p-value: <2e-16
```

Similarly, for men who attended at least some college, when holding else equal, whites are the happiest; blacks and men of other races are less happy, by 0.11 and 0.054 respectively.