Factor Variables

Kei Sakamoto

```r
data(CPS1985,package="AER")#borrowing from Applied Econometrics with R
```

Table of categories and frequencies for two factor variables(gender and occupation)

```r
table(CPS1985$gender)
```

```
##
##   male female
##    289    245
```

```r
table(CPS1985$occupation)
```

```
##
##      worker  technical    services      office       sales  management
##         156         105          83          97          38          55
```

Directly using factor variables in regression formula(automatically select male and worker as base categories)

```r
lm(log(wage) ~ education+experience+gender+occupation, data=CPS1985)
```

```
##
## Call:
## lm(formula = log(wage) ~ education + experience + gender + occupation,
##
##     data = CPS1985)
##
## Coefficients:
##        (Intercept)             education              experience
##            0.97629               0.07586                 0.01188
##        genderfemale     occupationtechnical     occupationservices
##           -0.22385               0.14246                -0.21004
##     occupationoffice        occupationsales   occupationmanagement
##           -0.05477              -0.20757                 0.15254
```

base category からの intercept の差がそれぞれの coefficient に現れる。
male からの差、worker からの差。interaction 入れてないので slope が
pararel という implicit な assumption が impose されてる…。

Manually redefine the reference category(reselect base category)
```
CPS1985$gender <- relevel(CPS1985$gender,"female")
CPS1985$occupation <- relevel(CPS1985$occupation,"management")
```

Rerun regression
```
lm(log(wage) ~ education+experience+gender+occupation, data=CPS1985)

##
## Call:
## lm(formula = log(wage) ~ education + experience + gender + occupation,

##     data = CPS1985)
##
## Coefficients:
##       (Intercept)            education           experience
##           0.90498              0.07586              0.01188
##        gendermale      occupationworker  occupationtechnical
##           0.22385             -0.15254             -0.01009
##  occupationservices      occupationoffice       occupationsales
##          -0.36259             -0.20731             -0.36011
```

female からの差、management からの差に切り替わる。education や
experience の coef は変わっていないのは当然で、male の coef は
female の coef の-1 倍になっているあたりも当然。