Tutorial 6

STAT 3013/8027

---

1. Consider the following simple linear regression model:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\
\epsilon_i &\sim iid \quad n(0, \sigma^2), \quad i = 1, \ldots, n.
\end{aligned}
$$

Ans. See the handwritten pages for $\hat{\beta}_0$ and $\hat{\beta}_1$, as well as their expectations and variances.

```
gdp <- read.csv("gdp2013.csv", header=T)
labor <- read.csv("labor2013.csv", header=T)
D <- merge(gdp, labor, by=c("Country.Name", "Country.Code"))
dim(D)
```

```
## [1] 248   4
```

```
D <- na.omit(D)
dim(D)
```

```
## [1] 206   4
```

```
names(D)[3:4] <- c("gdp", "labor")

##
y <- log(D$gdp)
x <- log(D$labor)



##
S.xy <- sum ( (y-mean(y))* (x-mean(x)) )
S.xx <- sum ( (x-mean(x))^2 )
beta.1.hat <- S.xy/S.xx
beta.1.hat
```

```
## [1] 0.9875253
```

```
beta.0.hat <- mean(y) - beta.1.hat*mean(x)
beta.0.hat
```

```
## [1] 9.669016
```

```
mod <- lm(y ~ x)
summary(mod)
```

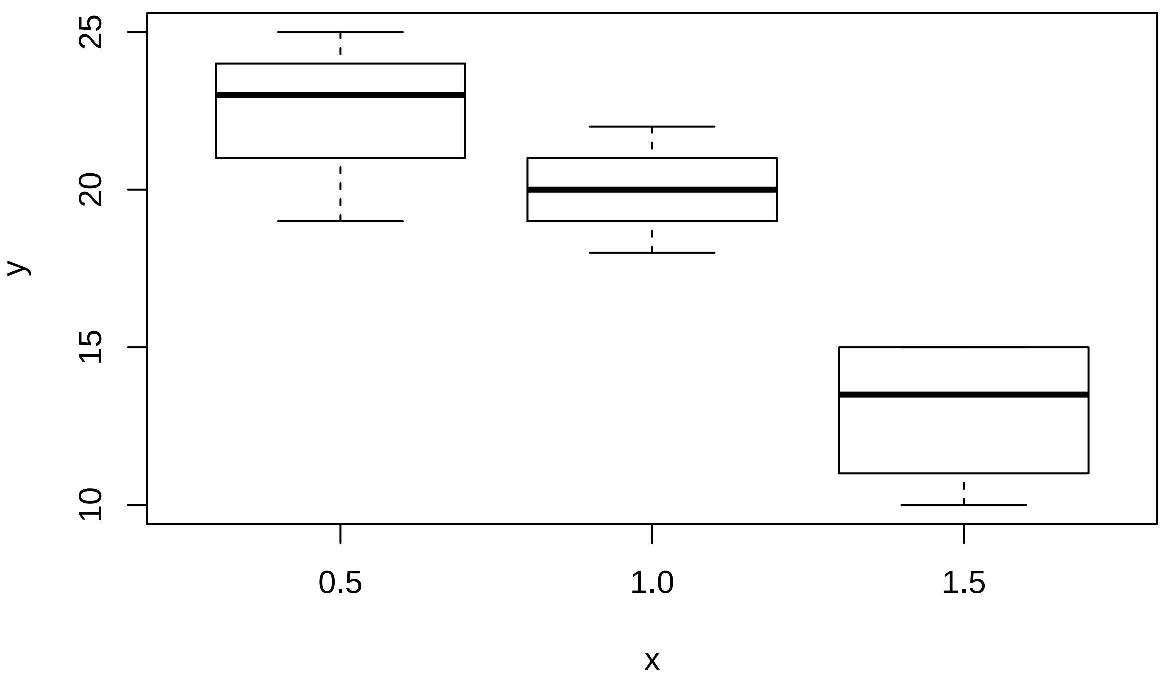
```
##
## Call:
```

```
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2597 -1.0684  0.0685  0.9935  2.8452
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.66902    0.66436   14.55   <2e-16 ***
## x            0.98753    0.04165   23.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.4 on 204 degrees of freedom
## Multiple R-squared:  0.7338, Adjusted R-squared:  0.7325
## F-statistic: 562.2 on 1 and 204 DF,  p-value: < 2.2e-16
```

2. Least-squares estimates for a categorical regression model. **Ans.** See the handwritten pages for the derivation of $\hat{\mu}_j$. We found:

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} y_{i,j} = \bar{y}_{.,j}$$

```
y <- c(21, 23, 19, 24, 25, 23, 19, 21, 20, 18,
       22, 20, 15, 10, 13, 14, 11, 15)
x <- as.factor(c(rep("0.5",6),  rep("1.0",6), rep("1.5",6)))
plot(y ~ x)
```



Note: In our model we assume every y_{ij} has the same variability. A typical concern is whether the variability within each group is similar across the groups. As a quick check we can examine the box part of the box plots to see if they are reasonably similar in spread. It seems roughly OK here, although group 2 has a slightly smaller spread comparatively.

```
mu.hat <- tapply (y, x ,mean)
mu.hat
```

```
##  0.5  1.0  1.5
## 22.5 20.0 13.0
```