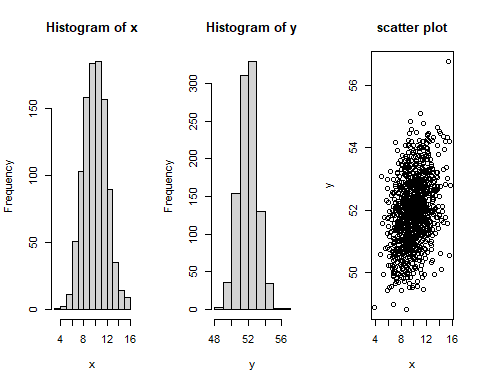
Inference\_supplementary doc

TY Kwon

x<-rnorm(1000,10,2)  
y<-0.2\*x+50+rnorm(1000,0,1)  
par(mfrow=c(1,3))  
hist(x)  
hist(y)  
plot(y~x, main="scatter plot")



# 모수모형은 다음과 같다.

# y=50+0.2x+u

model<-lm(y~x)  
summary(model)

##   
## Call:  
## lm(formula = y ~ x)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.9402 -0.7474 -0.0022 0.6970 3.7308   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 50.05137 0.16756 298.70 <2e-16 \*\*\*  
## x 0.19276 0.01651 11.68 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.028 on 998 degrees of freedom  
## Multiple R-squared: 0.1202, Adjusted R-squared: 0.1193   
## F-statistic: 136.4 on 1 and 998 DF, p-value: < 2.2e-16

#모집단 1000개의 개체들을 모두 관찰할 수 없기 때문에 모집단에서 30개의 개체를 임의로 추출하여 표본을 구성하였다.

obs<-c(1:1000)  
s<-sample(obs,30)  
s

## [1] 919 315 774 485 127 260 97 354 891 296 598 210 644 738 122 350 535 746 69  
## [20] 889 67 954 815 617 784 408 292 600 305 947

# 추출된 위의 30개의 개체들 자료를 이용하여 표본회귀모형을 추정하였다.

model.sample<-lm(y[s]~x[s])  
summary(model.sample)

##   
## Call:  
## lm(formula = y[s] ~ x[s])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.8470 -0.6572 0.3167 0.8489 1.9804   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 50.1872 1.0353 48.475 <2e-16 \*\*\*  
## x[s] 0.1912 0.1049 1.823 0.079 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.114 on 28 degrees of freedom  
## Multiple R-squared: 0.1061, Adjusted R-squared: 0.07416   
## F-statistic: 3.323 on 1 and 28 DF, p-value: 0.07902

# 위의 표본말고 다시 추출하여 표본을 다시 구성한다면?

# 결과는 달라질것..

s1<-sample(obs,30)  
s1

## [1] 830 666 111 72 786 743 827 602 57 205 636 276 102 17 850 708 212 887 929  
## [20] 651 342 564 669 922 846 551 416 615 787 373

model.s2<-lm(y[s1]~x[s1])  
summary(model.s2)

##   
## Call:  
## lm(formula = y[s1] ~ x[s1])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.93746 -0.54731 -0.02367 0.60236 2.10370   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47.84658 0.97126 49.262 < 2e-16 \*\*\*  
## x[s1] 0.40860 0.09807 4.167 0.000268 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.029 on 28 degrees of freedom  
## Multiple R-squared: 0.3827, Adjusted R-squared: 0.3607   
## F-statistic: 17.36 on 1 and 28 DF, p-value: 0.0002685

# 1. 표본으로 추출된 개체들의 구성이 달라지면 표본회귀모형을 이용하여 모회귀모형을 추정하고자 할때 그 추정치(OLS)가 달라진다.

# 2. 운이 나쁘다면 모수모형과 굉장히 다른 표본 모형을 얻을 수 있다.

# 3. 운이 좋다면 모수모형과 굉장히 유사한 모형을 관찰할 수 있다.

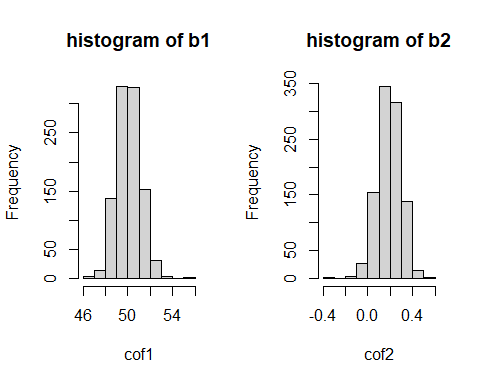
# 4.표본모형을 구성하는 회귀계수들은 “확률변수” 이다.

# 5.그렇다면 표본모형을 구성하는 회귀계수 (b1,b2)의 확률분포는?

# 6. 반복실험 해보자!!

# 7. 표본모형의 회귀계수의 확률분포 알아보기 : 30개씩 추출-추정 반복하여 추정된 회귀계수들의 분포를 살펴보자 (히스토그램을 그려보자)

cof1<- matrix(0,1000,1)  
cof2<-matrix(0,1000,1)  
  
for ( i in 1:1000){  
 s<-sample(obs,30)  
 model<-lm(y[s]~x[s])  
 cof1[i]<-model$coefficients[1]  
 cof2[i]<-model$coefficients[2]  
}  
  
par(mfrow=c(1,2))  
hist(cof1,main="histogram of b1")  
hist(cof2,main="histogram of b2")



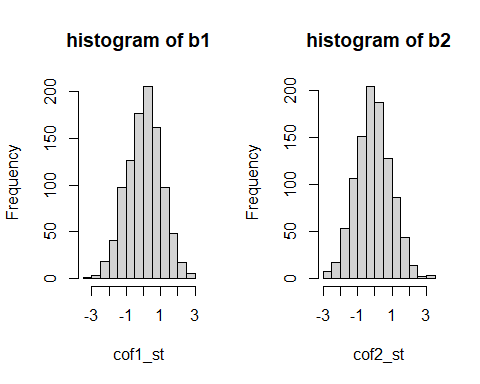
summary(cof1)

## V1   
## Min. :46.73   
## 1st Qu.:49.36   
## Median :50.04   
## Mean :50.08   
## 3rd Qu.:50.80   
## Max. :55.04

summary(cof2)

## V1   
## Min. :-0.3157   
## 1st Qu.: 0.1223   
## Median : 0.1926   
## Mean : 0.1905   
## 3rd Qu.: 0.2653   
## Max. : 0.5282

cof1\_st<- matrix(0,1000,1)  
cof2\_st<-matrix(0,1000,1)  
  
for ( i in 1:1000){  
 s<-sample(obs,30)  
 model<-lm(y[s]~x[s])  
 cof1\_st[i]<-model$coefficients[1]  
 cof2\_st[i]<-model$coefficients[2]  
}  
  
cof1\_st<- (cof1\_st-50)/sd(cof1\_st)  
cof2\_st<- (cof2\_st-0.2)/sd(cof2\_st)  
  
  
par(mfrow=c(1,2))  
hist(cof1\_st,main="histogram of b1")  
hist(cof2\_st,main="histogram of b2")



# 8. 어떠한 경우이던지 간에(모수값이 어떠한 값이던지 간에) 위의 0을 중심으로 가장 높고 점차 낮아지는 형태의 분포모양을 갖게 되며

## 이 모양이 모두 동일하다.(단, 꼬리의 두터움이 자료의 수와 X변수의 수에 따라서만 달라짐) :

# –> t분포 모양 (degree of freedom (자유도)에 따라 꼬리의 두터움의 정도만 달라짐!!! )