Tests

Difference

```r
x <- sample(1:6, 10, replace = T)
y <- x2 # x == y?
y <- sample(1:6, 10, replace = T) # x == y?

x <- sample(1:6, 100, replace = T)
y <- sample(1:6, 100, replace = T) # x == y?
y2 <- sample(1:6, 100, replace = T, prob = c(1, 1, 1, 1, 1, 2)) # x == y?
y3 <- sample(1:6, 100, replace = T, prob = c(100, 100, 100, 100, 100, 101))
```


Difference

- test (검정)
- 2 그룹 비교(t-test), 여러 그룹 비교(ANOVA)
- t-test 설명
- ANOVA 설명 1, 2, 3, 4
- chisq 설명

t-test in R

```r
A <- round(rnorm(100, mean = 100, sd = 5))
B <- round(rnorm(100, mean = 100 + 5, sd = 5))

t.test(A, B)
boxplot(A, B)
```


Anova in R

```r
# A, B
C <- round(rnorm(100, mean = 100, sd = 5))

scores <- data.frame(
    group = rep(c('A', 'B', 'C'), each = 100),
    value = c(A, B, C)
)
# head(scores)
# tail(scores)
res <- aov(value ~ group, scores)
summary(res)

boxplot(value ~ group, scores)
```


Chisq in R

```r
chisq.test(c(23,16,14,19,28))

chisq.test(c(33,16,14,19,28))
```


Regression

$y = Ax + B$

- Linear Regression
- Multiple Regression
- Logistic Regression

Linear Regression

```r
set.seed(123)
x <- 1:100
y <- 5 + 3*x + rnorm(100, mean=0, sd=20)  # y = 5 + 3x + error

df1 <- data.frame(x, y)

model_linear <- lm(y ~ x, data = df1) # LM
summary(model_linear)

plot(df1$x, df1$y, main="Simple Linear Regression", xlab="x", ylab="y")
abline(model_linear, col="red", lwd=2)
```


Multiple Regression

```r
set.seed(123)
x1 <- rnorm(100, mean=50, sd=10)
x2 <- rnorm(100, mean=30, sd=5)
y <- 10 + 2*x1 + 3*x2 + rnorm(100, mean=0, sd=15)
df2 <- data.frame(x1, x2, y)

model_multiple <- lm(y ~ x1 + x2, data = df2) ## LM
summary(model_multiple)
```

Logistic Regression

```r
set.seed(123)
x1 <- rnorm(100, mean=5, sd=2)
x2 <- rnorm(100, mean=10, sd=3)

prob <- 1 / (1 + exp(-( -5 + 0.8*x1 + 0.5*x2 )))  # 로지스틱 함수
y <- rbinom(100, size=1, prob=prob) ### 1 / 0

df3 <- data.frame(x1, x2, y)

model_logistic <- glm(y ~ x1 + x2, data = df3, family = binomial)
summary(model_logistic)

df3$pred <- predict(model_logistic, type="response")[1:10]
head(df3, 10)
```


Q&A?

Assignments

1. rnorm을 사용하여 2개의 데이터 그룹 생성. (n = 30)
    - N(3, 2) & N(10, 2)
2. 두 그룹간 평균 차이 유의성 확인

EXTRA 3. 위 데이터를 weight_loss, group(0, 1)으로 데이터 프레임 생성 후 선형 회귀 분석