

Review

Descriptive Statistics

- Read the data into R. (`read.table()` or `read.csv ()`)

- Objects:

vector, matrix, factor, list, and data frame

- Logic operation

`>, <, >=, <=, !=, &, |`

- How to extract subset of data frame

- How to create new variable
-

R programming

Basic programming skill (Write a function)

```
Myfunction<-function()  
{  
  return() }
```

Control structure:

if else for while

Useful command

**mean(...), sd(...), sum(...), sample(...), table(...),
apply(...), tapply(...), which(...), sort(...) rank(...),
sort.list(), rep(), seq()**

Graph

- scatter plot (`plot(,type="",color="....")`)
- Histogram (`hist(...)`)
- Box plot (`boxplot(...)`)
- QQ plot (`qqnorm(...). qqline(...)`)
- Pie plot (`pie(...)`)
- Bar chart (`barplot(...)`)
- Plot a function(`plot(x,y,type='l')`)

Comment on the graph!

Probability distribution

- Probability distribution

`dnorm()` `pnorm()` `qnorm()` `rnorm()`

`dunif()` `punif()` `qunif()` `runif()`

`dbinom()` `pbinom()` `qbinom()` `rbinom()`

`dt ()` `pt ()` `qt ()` `rt ()`

`df()` `pf()` `qf()` `rf()`

`dexp()` `pexp()` `qexp()` `rexp()`

Probability distribution estimation

- Moment method

- $$\mu_1 = M_1 \quad \mu_2 = M_2 \quad \mu_3 = M_3 \quad \text{so on...}$$
$$E(X) = \frac{1}{n} \sum_{i=1}^n X_i \quad E(X^2) = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad E(X^3) = \frac{1}{n} \sum_{i=1}^n X_i^3$$

Continue this until there are enough equations to solve for the unknown parameters

- Maximum likelihood estimation

(Define the log likelihood function

Perform the optimization to determine the MLE estimation `nlm()` `optimize()`

Hypothesis Test

- One sample mean t-test
- One sample proportion test
- Two sample mean t-test
- Two sample proportion test
- The paired t -test
- Analysis of variance (One way and two way ANOVA)

(`t.test(...)`, `prop.test(...)`, `oneway.test()`
`anova(lm())`)

Decision and conclusion!!

Multiple Linear Regression

- Regression equation. Interpretation of the coefficients.
- Correlation matrix, Interpretation
- R^2 , Interpretation
- ANOVA table, Interpretation
- Hypothesis test on each coefficients
- Global test of hypothesis on all coefficients
- Prediction based on regression equation
- Diagnostic Plots
- Possible transformation
- Possible interaction
- Model comparison
- Stepwise regression
- Model evaluation and cross-validation

Interpretation!!!!

```
cor(...)  
lm(...)  
summary(...)  
anova(...)  
fitted(...)  
resid (...)  
predict(...)  
plot()  
rstudent()  
lm.influence()  
influence.measures()  
step()
```

Chi-square test

- Construct contingency table
- A Chi-Square Test for Independence

```
M<-table()  
Xsq <- chisq.test(M))  
Xsq$observed  
Xsq$expected
```