

Chapter 2: Basic Statistics and Simple Simulations

I. Statistics

2.1. Statistical inference

1. Probability Distributions:

Example (Splus):

```
pnorm(1.96)
```

```
dnorm(0)
```

```
qnorm(0.975)
```

```
rnorm(100)
```

```
pnorm(1.96,mean=1,sd=2)
```

```
dnorm(0,mean=1,sd=2)
```

```
qnorm(0.975,mean=1,sd=2)
```

```
rnorm(100,mean=1,sd=2)
```

2. One-sample problem:

Example (Splus):

```
### Exploratory Data Analysis (EDA)
```

```
light=c(850,740,900,1070,930,850,950,980,  
      980,880,1000,980,930,650,760,810,1000,1000,960,960)
```

```
par(mfrow=c(2,2))
```

```
tsplot(light)
```

```
hist(light,nclass=5)
```

```
boxplot(light)
```

```
qqnorm(light)
```

```
qqline(light)
```

```
### t Test
```

```
t.test(light,mu=990) # test  $H_0: \mu = 990$  vs.  $H_1: \mu \neq 990$ 
```

```
t.test(light,conf.level=0.9,mu=990) # 90% confidence interval can be  
obtained
```

```
t.test(light,alternative="greater",mu=990) # test  $H_0: \mu \leq 990$  vs.  $H_1: \mu > 990$ 
```

```
help(t.test)
```

```
qt(0.975,19) #  $t_{19,0.025}$ 
```

3. Two-samples problem:

Example:

Protein(H)	134	146	104	119	124	161	107	83	113	129	97	123
Protein(L)	70	118	101	85	107	132	94					

19 rats were divided into two groups, one group with 12 rats and the other with 7 rats. The larger group was given high protein food while the smaller group is given low protein food. The data of the weight gains for these rats under the two diets are given in above table. We now demonstrate the variance test using this data in S-Plus.

(a) Variance test:

Example (Splus):

```
rat1=c(134,146,104,119,124,161,107,83,113,129,97,123)
rat2=c(70,118,101,85,107,132,94)
var.test(rat1, rat2)           #  $H_0: \sigma_1^2 = \sigma_2^2$  vs.  $H_1: \sigma_1^2 \neq \sigma_2^2$ 
var.test(rat1, rat2, conf.level=0.9, alt="g")  #  $H_0: \sigma_1^2 \leq \sigma_2^2$  vs.  $H_1: \sigma_1^2 > \sigma_2^2$ 
qf(0.95, 11, 6)             #  $f_{11, 6, 0.05}$ 
```

(b) t test:

```
t.test(rat1, rat2, var.equal=T)      #  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 \neq \mu_2$ 
t.test(rat1, rat2, alt="g", var.equal=T)  #  $H_0: \mu_1 \leq \mu_2$  vs.  $H_1: \mu_1 > \mu_2$ 
t.test(rat1, rat2)
```

4. One-way ANOVA:

Then, the F statistic for testing $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ is

$$F = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 / k - 1}{\sum_{j=1}^k (n_j - 1) s_j^2 / n_T - k} = \frac{MSB}{MSW}$$

where \bar{x}_j and s_j^2 are the sample means and variances of the jth sample, respectively. Then,

reject H_0 as $F > f_{k-1, n_T - k, \alpha}$.

Example:

Four observations were selected from each of three normal populations with equal variances. The data obtained follow.

Observation	Sample 1	Sample 2	Sample 3
1	20	28	20
2	26	26	19
3	24	31	23
4	22	27	22

- (i) Compute the between-treatments and within-treatments estimates of population variance.
- (ii) Find the F statistic.
- (iii) Test $H_0: \mu_1 = \mu_2 = \mu_3$ at $\alpha = 0.1$.

Example 3 (Splus):

```
sample=c(20,26,24,22,28,26,31,27,20,19,23,22)
group=factor(c(rep("a",4),rep("b",4),rep("c",4)))
result=aov(sample~group)
summary(result)
qf(0.9,2,9)                                #  $f_{2,9,0.1}$ 
```

```
dim(sample)=c(4,3)
xjbar=apply(sample,2,mean)
xbar=mean(sample)
sj2=apply(sample,2,var)
nj=apply(sample,2,length)
msb=sum(nj*(xjbar-xbar)^2)/(3-1)
mse=sum((nj-1)*sj2)/(12-3)
f=msb/mse
f
```

Useful links:

Soft King : <http://www.softking.com.tw/>

Notepad++: <https://notepad-plus-plus.org/>

doPDF : <http://www.dopdf.com/tw/quick-download.php>