實驗設計 HW2

統碩一 106354012 曹立諭

2018/3/20

1. Exercise 2.4

Describe the benefits and risks of using these five methods.

As part of a larger experiment, Dale (1992) looked at six samples of a wetland soil undergoing a simulated snowmelt. Three were randomly selected for treatment with a neutral pH snowmelt; the other three got a reduced pH snowmelt. The observed response was the number of Copepoda removed from each microcosm during the first 14 days of snowmelt.

Reduced pH			Neutral pH		
256	159	149	54	123	248

Using randomizationmethods, test the null hypothesis that the two treatments have equal average numbers of Copepoda versus a two-sided alternative.

H_0: 兩試驗沒有差別

H_1: reject H_0

假設$\alpha = 0.05$

我們用A與B兩組資料的差異和作為統計量

```
#Permutation test
#原本資料計算
A<-c(256,159,149);A
```

```
## [1] 256 159 149
```

```
B<-c(54,123,247);B
```

```
## [1]  54 123 247
```

```
D<-A[]-B[];D<-matrix(D,ncol = length(A))
#A與B資料的差異和
D_sum<-apply(D,1,sum)
#各種排列數差異和計算
W<-c(A,B)
W<-matrix(data = W,ncol =length(W) )
W_sum<-apply(W,1,sum)
cb_W<- combn(W,3)
cb_A_sum<-apply(cb_W,2,sum)
cb_B_sum=NULL
n=NULL
for(i in 1:choose(length(W),length(W)/2)){
  n=W_sum-cb_A_sum[i]
  cb_B_sum=c(cb_B_sum,n)
}
all_d_sum<-cb_A_sum-cb_B_sum
#將所有排列差異和按照順序列出
all_d_sum <- matrix(sort(all_d_sum),ncol = 5,byrow = T)
print(all_d_sum)
```

-336	-316	-264	-140	-126
-122	-88	-70	-68	-50
50	68	70	88	122
126	140	264	316	336

共20組排列差異總和

統碩一 106354012 曹立諭


```
#計算P值(計算有幾種可能使得兩者的差異絕對值不小於觀察差異)
p_value<-2*sum(1*(D_sum<=all_d_sum)/choose(length(W),length(W)/2))
p_value
```

```
## [1] 0.4
```

接著亂數模擬1000次

```
s_d=NULL
P_hat_value=NULL
sd_p_hat=NULL
for (j in 1:1000) {
  s_d=NULL
  for(i in 1:1000){
   x <- sample(W,3)
   sx <- sum(x)
   sy <- sum(W)-sum(x)
   sd <- sx-sy
   s_d=c(s_d,sd)
}
p_hat_value <- 2*sum(1*(D_sum<=s_d[-1]))/1000
P_hat_value = c(P_hat_value,p_hat_value)
}
sort(P_hat_value)
```

```
P_hat_value <- sum(P_hat_value)/1000
sd_p_hat <- (p_hat_value*(1-p_hat_value))/1000
print(paste('P hat value :' , P_hat_value))
```

```
## [1] "P hat value : 0.399372"
```

```
print(paste('stand error of p hat :' ,sd_p_hat))
```

```
## [1] "stand error of p hat : 0.000229264"
```

P_hat_value=0.399372

由圖可知我們 P值>0.05，我們不拒絕虛無假設，及兩試驗沒有差別。

2018/3/20

2. $E(\hat{\sigma}^2) = \sigma^2$

We know $Y_{ij} \sim N(u_i, \sigma^2)$ $\quad E(Y_{ij}^2) = u_i^2 + \sigma^2$

$Y_{i}. \sim N\left(u_i, \frac{\sigma^2}{n_i}\right)$ $\quad E(\bar{Y}_{i.}^2) = u_i^2 + \frac{\sigma^2}{n_i}$

One remaining parameter to estimate is

$$\mathrm{Var}(\epsilon_{ij}) = \sigma^2 \quad , j = 1, \ . \ . \ . \ , n_i \quad , i = 1, \ . \ . \ . \ , g$$

Consider

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{N-g} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i}.\right)^2}{N-g}$$

In both the separate means and single means models, $\hat{\sigma}$ is unbiased for σ^2, i.e.

$$E(\hat{\sigma}^2) = E\left(\frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i}.\right)^2}{N-g}\right)$$

$$= E\left(\frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij}^2 - 2Y_{ij}\bar{Y}_{i.} + \bar{Y}_{i}.^2)}{N-g}\right)$$

$$= E\left(\frac{\sum_{i=1}^{g}(\sum_{j=1}^{n_i} Y_{ij}^2 - n_i\bar{Y}_{i.}^{\ 2})}{N-g}\right)$$

$$= \frac{\sum_{i=1}^{g} n_i(u_i^2 + \sigma^2) - \sum_{i=1}^{g} n_i(u_i^2 + \frac{\sigma^2}{n_i})}{N-g}$$

$$= \frac{\sum_{i=1}^{g} n_i\,\sigma^2 - \sum_{i=1}^{g}\sigma^2}{N-g} \quad (\because n_1 + n_2 + 、、、 + n_g = N)$$

$$= \sigma^2$$

實驗設計 HW2

統碩一 106354012 曹立諭

2018/3/20

3. ANOVA

Consider the following decomposition

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i\cdot} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i\cdot}).$$

Further,taking squares,

其中

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i} 2(\bar{y}_{i\cdot} - \bar{y}_{..})(y_{ij} - \bar{y}_{i\cdot}) = 2\sum_{i=1}^{g}(\bar{y}_{i\cdot} - \bar{y}_{..})\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot}) = 0$$

$$SST = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij} - \overline{y_{i\cdot}})^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left((\bar{y}_{i\cdot} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i\cdot})\right)^2$$

$$= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y}_{..})^2 + 2(\bar{y}_{i\cdot} - \bar{y}_{..})(y_{ij} - \bar{y}_{i\cdot}) + (y_{ij} - \bar{y}_{i\cdot})^2$$

$$= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y}_{..})^2 + (y_{ij} - \bar{y}_{i\cdot})^2$$

$$= \sum_{i=1}^{g} n_i(\bar{y}_i - \bar{y}_{..})^2 + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$$

$$= \sum_{i=1}^{g} n_i\hat{\alpha}_i^{\,2} + \mathrm{SS_E}$$

$$\therefore SS_T = SS_{Trt} + SS_E$$