刘程华 2018011687 计91

1 (a) H_0: there is a significant difference in treatment effects

H_1: there is no significant difference in treatment effects

Test statistic:

$$T = \mathrm{Max}\left([\max(M_A, M_B, M_C) - \min(M_A, M_B, M_C)], [\max(F_A, F_B, F_C) - \min(F_A, F_B, F_C)]\right)$$

其中 M_A 表示 A 药物治疗 male 的效果平均值. 考虑到药效在男女间有明显差异, 我们认为性别不可交换, T 为 A, B, C 三者对男性效果的极差和 A, B, C 三者对女性效果的极差的较大者

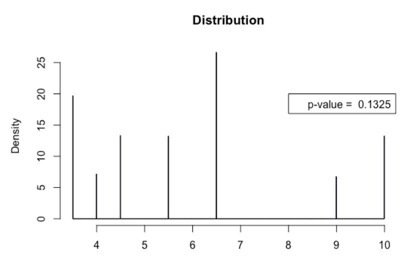
(b) 情况总数为 $(C_6^2 C_4^2 \cdot C_2^2) \cdot C_6^2 C_4^2 C_2^2) = 8100$

种情况 (因为每组数据代表一个个体, 我们认为数值相同的也具有差异性)

我们随机模拟 5000 次来得到 T 的近似分布.

事实上此题样本量小, 我们可以遍历得到准确分布

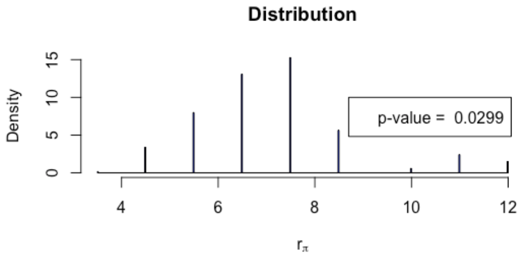
(c)



在 $\alpha = 0.1$ 的情况下, 我们拒绝 H_0

#附: 事实上此题目是模糊的, 上面我们理解差异指 A, B, C 药物之间的差异。差异也可以指治疗效果对性别的差异。把这种也考虑在内, 我们不妨令 T 为 $M_A, M_B, M_C, F_A, F_B, F_C$ 六者的极差, 性别可交换, 此时共有 $C_{12}^2 C_{10}^2 C_8^2 C_6^2 C_4^2 C_2^2 = 7484400$ 种情况

我们随机模拟 100000 次来得到 T 的近似分布.



在 $\alpha=0.05$ 水平下 我们拒绝 H_0

2. (a)

使用置换检验 在 X、Y 样本总体中随机抽取 n_1 得到 X'，剩下的 Y'

计算 $T'=\overline{X'}-\overline{Y'}$ 多次重复 可以得到 T 的数值分布.

可由此计算 p-value

(b)

$$T=\frac{S}{n_1}-\frac{\sum Y_i}{n_2}=\frac{S}{n_1}-\frac{SUM-S}{n_2}$$

$$=-\frac{SUM}{n_2}+\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S \quad \text{其中 } SUM=\sum_{i=1}^{n_1}X_i+\sum_{i=1}^{n_2}Y_i$$

在置换检验中是恒定的

是 S 的线性变换，T 的数值分布是线性变换的，p-value 相同，检验是等价的

3.

Z-test: $H_0: \mu=\mu_0 \quad H_1: \mu\neq\mu_0$

$$U=\frac{\bar{X}-\mu_0}{\sigma}\cdot\sqrt{n}\sim N(0,1)$$

$$D=\{u \mid |u|>c\}=\{u \mid u<-c, u>c\} \quad (\text{设 } c>0)$$

$$\text{Power}=P(u\in D)=P\left(\left(\frac{\bar{X}-\mu}{\sigma}\cdot\sqrt{n}+\frac{\mu-\mu_0}{\sigma}\cdot\sqrt{n}\right)\in D\right) \quad \left(\text{设 } a=\frac{\mu-\mu_0}{\sigma}\cdot\sqrt{n}\right)$$

$$=P\left(-a+c<\frac{\bar{X}-\mu}{\sigma}\cdot\sqrt{n},\ \frac{\bar{X}-\mu}{\sigma}\cdot\sqrt{n}<-c-a\right)$$

$$=\Phi(-c-a)+1-\Phi(-a+c)$$

Bino-test: $H_0: u = np$ $\quad H_1: u \neq np$

$$u \sim B(n, p)$$

$$D = \{ u \mid u - np > c, \quad np - u > c \}$$

$$\text{Power} = P(u \in D) = P(u > c + np, \ u < c - np)$$

$$= \sum_{i=c+np+1}^{n} C_n^i p^i (1-p)^{n-i} + \sum_{i=1}^{c-np-1} C_n^i p^i (1-p)^{n-i}$$

这里 $c > np$，考虑了双边，以及要注意取整数，需要复杂的分类讨论……

```r
#prob 1
set.seed(123456)
n<-12
nrep<-100000
eff<-c(4,6,6,8,8,10,6,8,8,10,17,17)
eff1<-c(4,6,6,8,8,10)
eff2<-c(6,8,8,10,17,17)
r_perm<-NULL
MA=eff[1:2]
MB=eff[3:4]
MC=eff[5:6]
FA=eff[7:8]
FB=eff[9:10]
FC=eff[11:12]
r_obs<-max(max(mean(MA),mean(MB),mean(MC))-
min(mean(MA),mean(MB),mean(MC)),max(mean(FA),mean(FB),mean(FC))-
min(mean(FA),mean(FB),mean(FC)))
for(i in 1:nrep){

  tmp1<-sample(1:6,6)
  MA=eff1[tmp1[1:2]]
  MB=eff1[tmp1[3:4]]
  MC=eff1[tmp1[5:6]]
  tmp2<-sample(1:6,6)
  FA=eff2[tmp2[1:2]]
  FB=eff2[tmp2[3:4]]
  FC=eff2[tmp2[5:6]]
  r_perm[i]<-max(max(mean(MA),mean(MB),mean(MC))-
min(mean(MA),mean(MB),mean(MC)),max(mean(FA),mean(FB),mean(FC))-
min(mean(FA),mean(FB),mean(FC)))
  }
pval <- length(r_perm[r_perm >= r_obs])/nrep
pval<-round(pval, digits=4)
hist(r_perm, breaks = 500,
     main = "Distribution", col="blue", freq=FALSE, xlab = expression(r[pi]))

legend(8, 20, paste("p-value = ", pval))


#prob 1+
set.seed(123456)
n<-12
nrep<-100000
eff<-c(4,6,6,8,8,10,6,8,8,10,17,17)
r_perm<-NULL
MA=eff[1:2]
```

```
MB=eff[3:4]
MC=eff[5:6]
FA=eff[7:8]
FB=eff[9:10]
FC=eff[11:12]
r_obs<-max(mean(MA),mean(MB),mean(MC),mean(FA),mean(FB),mean(FC))-
min(mean(MA),mean(MB),mean(MC),mean(FA),mean(FB),mean(FC))
for(i in 1:nrep){

  tmp<-sample(1:12,12)
  MA=eff[tmp[1:2]]
  MB=eff[tmp[3:4]]
  MC=eff[tmp[5:6]]
  FA=eff[tmp[7:8]]
  FB=eff[tmp[9:10]]
  FC=eff[tmp[11:12]]
  r_perm[i]<-max(mean(MA),mean(MB),mean(MC),mean(FA),mean(FB),mean(FC))-
min(mean(MA),mean(MB),mean(MC),mean(FA),mean(FB),mean(FC))
  }
pval <- length(r_perm[r_perm >= r_obs])/nrep
pval<-round(pval, digits=4)
hist(r_perm, breaks = 500,
     main = "Distribution", col="blue",freq=FALSE, xlab = expression(r[pi]))

legend(8.7, 10, paste("p-value = ", pval))
```