ORIGIN ≡ 0

Fisher's Exact Test

Fisher's Exact Test may be used for 2 X 2 contingency tables that fail the criterion for use of the Normal Approximation.

Assumptions:

- **Observed values $X_1, X_2, X_3, \ldots X_{n1}$ are a random sample**
  Observed values $Y_1, Y_2, Y_3, \ldots Y_{n2}$ .are a random sample.

Contingency Table		
a	b	*a+b*
c	d	*c+d*
a+c	*b+d*	***n***

Model:

Let Probabilities:
- $P_i = P(X=i)$
- $P_j = P(Y=j)$
- $P_{ij} = (X=i, Y=j)$

Criterion for Normal Approximation:

- **IF expected values in each cell $E_{ij} \geq 5$ THEN Approximation may be used OTHERWISE use Exact Test e.g., Fisher's Exact Test**

Zar Example 23.4 & 24.20

Contingency Table		
12	7	*19*
2	9	*11*
14	*16*	***30***

$a := 12 \quad b := 7 \qquad a! = 4.79 \times 10^{8} \qquad b! = 5040$

$c := 2 \quad d := 9 \qquad c! = 2 \qquad d! = 3.6288 \times 10^{5}$

$n := a + b + c + d \qquad n = 30 \qquad n! = 2.6525 \times 10^{32}$

$R := \begin{pmatrix} a + b \\ c + d \end{pmatrix} \qquad R = \begin{pmatrix} 19 \\ 11 \end{pmatrix} \qquad C := \begin{pmatrix} a + c \\ b + d \end{pmatrix} \qquad C = \begin{pmatrix} 14 \\ 16 \end{pmatrix}$

^ Row and Column totals

Expected Values		
8.867	10.133	
5.131	5.867	

Normal Approximation would be OK, but let's use Fisher's Exact Test anyway... >

Fisher's Exact Test:

Enumerate all Possible Contingency Tables:

- **Enumerate more extreme 2X2 Contingency tables with identical row and column totals as the observed table.**
- **Calculate the exact probability of each table based on the Hypergeometric Distribution.**

Hypergeometric Probability of Observed 2X2 contingency table:

$T := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad T = \begin{pmatrix} 12 & 7 \\ 2 & 9 \end{pmatrix}$

$P_T := \frac{(a + b)! \cdot (c + d)! \cdot (a + c)! \cdot (b + d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!} \qquad P_T = 0.01906$ **< a,b,c,d,n are specified as in the table above.**

One-tailed Case:

Enumeration of More Extreme Contingency Tables:

Since Zar fixes row and column totals in his example he calculates only two more extreme cases:

Case A:

$a := 13 \quad b := 6 \quad c := 1 \quad d := 10$

$TA := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TA = \begin{pmatrix} 13 & 6 \\ 1 & 10 \end{pmatrix}$

Cumulative Probability:

$P_{TA} := \dfrac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!} \qquad P_{TA} = 0.00205 \qquad P_{TA} = 0.00205231$

Case B:

$a := 14 \quad b := 5 \quad c := 0 \quad d := 11$

$TB := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TB = \begin{pmatrix} 14 & 5 \\ 0 & 11 \end{pmatrix}$

$P_{TB} := \dfrac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!} \qquad P_{TB} = 0.00007996 \qquad P_{TA} + P_{TB} = 0.00213227$

Summation of Probabilities:

$P := \begin{pmatrix} P_T \\ P_{TA} \\ P_{TB} \end{pmatrix} \qquad P = \begin{pmatrix} 0.01905714 \\ 0.00205231 \\ 0.00007996 \end{pmatrix}$

$\sum P = 0.021189$ **< this is the probability of obtaining a table as extreme as the one observed in the direction of H_1.**

Hypotheses:

H_0: $p_{ij} = (P_i)(P_j)$ **< That is, variables X & Y are independent!**

H_1: $P_{1,j} > P_{2,j}$ **< One sided test**

^ probability of row 1 is greater than row 2 in contingency table

Decision Rule:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

IF $\Sigma P < \alpha$ THEN REJECT H_0 OTHERWISE ACCEPT H_0

$\sum P = 0.02119$ **therefore reject H_O**

Two-tailed Case:

Enumeration of More Extreme Contingency Tables:

Case C: **Cumulative Probability:**

$a := 3 \quad b := 16 \quad c := 11 \quad d := 0$

$TC := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TC = \begin{pmatrix} 3 & 16 \\ 11 & 0 \end{pmatrix}$

$P_{TC} := \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!}$ $\quad P_{TC} = 0.000006663$ $\quad P_{TC} = 0.00000666$

Case D:

$a := 4 \quad b := 15 \quad c := 10 \quad d := 1$

$TD := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TD = \begin{pmatrix} 4 & 15 \\ 10 & 1 \end{pmatrix}$

$P_{TD} := \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!}$ $\quad P_{TD} = 0.00029319$ $\quad P_{TC} + P_{TD} = 0.00029985$

Case E:

$a := 5 \quad b := 14 \quad c := 9 \quad d := 2$

$TE := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TE = \begin{pmatrix} 5 & 14 \\ 9 & 2 \end{pmatrix}$

$P_{TE} := \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!}$ $\quad P_{TE} = 0.004397801$ $\quad P_{TC} + P_{TD} + P_{TE} = 0.0047$

Case F:

$a := 6 \quad b := 13 \quad c := 8 \quad d := 3$

$TF := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad TF = \begin{pmatrix} 6 & 13 \\ 8 & 3 \end{pmatrix}$

$P_{TF} := \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!}$ $\quad P_{TF} = 0.03078461$ $\quad P_{TC} + P_{TD} + P_{TE} + P_{TF} = 0.0355$

^ Since cumulative probability is greater than the cumulative one-tailed probabilty, Case F is excluded below

Summation of Probabilities:

$$P := \begin{pmatrix} P_T \\ P_{TA} \\ P_{TB} \\ P_{TC} \\ P_{TD} \\ P_{TE} \end{pmatrix} \qquad P = \begin{pmatrix} 0.01905714 \\ 0.00205231 \\ 0.00007996 \\ 0.00000666 \\ 0.00029319 \\ 0.0043978 \end{pmatrix}$$

$\sum P = 0.025887$ **< this is the probability of obtaining a table as extreme as the one observed in either direction.**

Hypotheses:

H_0: $p_{ij} = (P_i)(P_j)$ **< That is, variables X & Y are independent!**

H_1: $p_{ij} <> (P_i)(P_j)$ **< Two sided test**

Decision Rule:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

IF ΣP < α THEN REJECT H_0 OTHERWISE ACCEPT H_0

$\sum P = 0.025887$ **therefore reject H_O**

Prototype in R:

```
#FISHER'S EXACT TEST OF INDEPENDENCE

X=matrix(c(12,7,2,9),nrow=2,byrow=T)
X

#ONE-TAILED CASE:
fisher.test(X,alternative="greater",conf.level=0.95)
```

```
        Fisher's Exact Test for Count Data

data:  X
p-value = 0.02119
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 1.331695      Inf
sample estimates:
odds ratio
  7.166131
```

```
#TWO-TAILED CASE:
fisher.test(X,alternative="two.sided",conf.level=0.95)
```

```
        Fisher's Exact Test for Count Data

data:  X
p-value = 0.02589
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  1.053152 87.138198
sample estimates:
odds ratio
  7.166131
```

Note that R also has a Hypergeometric Distribution function that may be used to calculate probabilities:

```
#HYPERGEOMETRIC DISTRIBUTION IN R:
#ONE TAIL:
X=12:14  #specifies possible values of a
X
m=14  #column total a+c
n=16   #column total b+d
k=19   #row total a+b

dhyper(X,m,n,k)  #point probabilities
phyper(X,m,n,k)  #cumulative probabilities

#OTHER TAIL:
X=3:6   #other values of a

dhyper(X,m,n,k)  #point probabilities
phyper(X,m,n,k)  #cumulative probabilities
```

$$P = \begin{pmatrix} 0.01905714 \\ 0.00205231 \\ 0.00007996 \\ 0.00000666 \\ 0.00029319 \\ 0.0043978 \end{pmatrix}$$

```
> dhyper(X,m,n,k)  #point probabilities
[1] 1.905714e-02 2.052307e-03 7.996002e-05

> phyper(X,m,n,k)  #cumulative probabilities
[1] 0.9978677 0.9999200 1.0000000


> dhyper(X,m,n,k)  #point probabilities
[1] 6.663335e-06 2.931867e-04 4.397801e-03 3.078461e-02

> phyper(X,m,n,k)  #cumulative probabilities
[1] 6.663335e-06 2.998501e-04 4.697651e-03 3.548226e-02
```