# BIOS 662 Fall 2018

# Power and Sample Size, Part II

David Couper, Ph.D.

david_couper@unc.edu

or

couper@bios.unc.edu

https://sakai.unc.edu/portal

# Outline

- Two sample: Continuous
- Two sample: Binary
- Case-control studies
- Estimating power with simulations

# Two Sample Test: Continuous Outcome

- Hypotheses

$$H_0 : \mu_1 = \mu_2 \ \text{ versus } \ H_A : \mu_1 \neq \mu_2$$

- Assume homogeneity of variance, $\sigma^2$ known, normality/CLT

- Then

$$N = \frac{2\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} = 2\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\Delta}\right)^2$$

- Note that there are $N$ observations in *each* group so that the total sample size is $2N$

# Two Sample Test: Example

- A drug company wants to compare 2 drugs for lowering LDL-cholesterol
- Previous studies have found $\sigma^2 = 25^2 = 625$
- A difference of 15 mg/dl is considered to be clinically meaningful
- For $\alpha = 0.05$ (2-sided) and $1 - \beta = 0.9$

$$N = \frac{2(625)(1.96 + 1.28)^2}{225} \approx 59$$

- So 118 patients are needed for the study

## Two Sample Test: $\boldsymbol{\sigma}$ Unknown

- What if the variance is not known?
- For $N_1 = N_2 = N$, one can show that

$$\frac{\overline{Y}_1 - \overline{Y}_2}{s_p\sqrt{\frac{2}{N}}} \sim t_{2N-2,\lambda}$$

where

$$\lambda = \Delta\sqrt{N/2}$$

# Two Sample Test, $\sigma$ Unknown: R

```
# by hand
> 1-pt(qt(0.975,116), 116, 15/25*sqrt(59/2))
[1] 0.8982732

> power.t.test(59, delta=15, sd=25)

     Two-sample t test power calculation

              n = 59
          delta = 15
             sd = 25
      sig.level = 0.05
          power = 0.8982732
    alternative = two.sided

 NOTE: n is number in *each* group
```

# Two Sample Test, $\sigma$ Unknown: SAS

```
proc power;
   twosamplemeans
   meandiff = 15
   ntotal   = 118
   stddev   = 25
   power    = .;
```

```
Two-sample t Test for Mean Difference

Distribution                  Normal
Method                         Exact
Mean Difference                   15
Standard Deviation                25
Total Sample Size                118
Number of Sides                    2
Null Difference                    0
Alpha                           0.05

Computed Power

0.898
```

## Two Sample Test, $\boldsymbol{\sigma}$ Unknown

- Given $\beta$, solve for $N$

$$1 - \beta = \Pr[T \geq t_{2N-2,0;1-\alpha/2}]$$

where $T \sim t_{2N-2,\Delta\sqrt{N/2}}$

- For example, suppose $\beta = 0.1$, $\Delta = 0.5$; numerical search in R:

```
> N <- 50; 1-pt(qt(0.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.6968888
> N <- 90; 1-pt(qt(0.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.9155872
> N <- 86; 1-pt(qt(0.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.9032299
> N <- 85; 1-pt(qt(0.975,2*N-2), 2*N-2, 1/2*sqrt(N/2))
[1] 0.899894
```

- So $N = 86$

# Two Sample Test, $\sigma$ Unknown: R

```
> power.t.test(power=0.9, delta=0.5)

     Two-sample t test power calculation

              n = 85.03129
          delta = 0.5
             sd = 1
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

# Two Sample Test, $\sigma$ Unknown: SAS

```
proc power;
   twosamplemeans
   meandiff = 15
   ntotal   = .
   stddev   = 30
   power    = 0.9;

Two-sample t Test for Mean Difference

Distribution                Normal
Method                       Exact
Mean Difference                 15
Standard Deviation              30
Nominal Power                  0.9
Number of Sides                  2
Null Difference                  0

Computed N Total

Actual        N
 Power    Total
 0.903      172
```

# Two Sample Test: Binary Outcome

- Hypotheses

$$H_0 : \pi_1 = \pi_2 \ \text{ versus } \ H_A : \pi_1 \neq \pi_2$$

- Then

$$N \approx \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\pi_1 - \pi_2)^2}$$

where

$$\sigma^2 = (\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2))/2$$

see page 161 of the text

- Again there are $N$ observations in each group so that the total sample size is $2N$

# Two Sample Test: Binary Outcome

- Suppose $\pi_1 = 0.2727$, $\pi_2 = 0.2$, $\alpha = 0.05$ (two-sided), $1 - \beta = 0.9$. Then

$$N \approx \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\pi_1 - \pi_2)^2} = 712$$

- SAS

```
proc power;
   twosamplefreq
   refp   = 0.2
   pdiff  = 0.0727
   ntotal = .
   power  = 0.9;
```

# Two Sample Test: Binary Outcome

```
Pearson Chi-square Test for Two Proportions

               Fixed Scenario Elements

Distribution                              Asymptotic normal
Method                                 Normal approximation
Reference (Group 1) Proportion                          0.2
Proportion Difference                                0.0727
Nominal Power                                           0.9
Number of Sides                                           2
Null Proportion Difference                                0
Alpha                                                  0.05
Group 1 Weight                                            1
Group 2 Weight                                            1

Computed N Total

Actual       N
 Power   Total

 0.900    1432
```

# Two Sample Test: Binary Outcome

- Why the difference? SAS uses a different approximation, which we now derive (cf. Fleiss, 1981)

- For $N_1 = N_2 = N$, Pearson's chi-square test statistic is equivalent to

$$Z = \frac{p_2 - p_1}{\sqrt{2\bar{p}\bar{q}/N}}$$

  where $\bar{p} = (p_1 + p_2)/2$, $\bar{q} = 1 - \bar{p}$

- Without loss of generality, consider the alternative $\pi_2 - \pi_1 = \delta_A > 0$.

## Two Sample Test: Binary Outcome

- Power to detect $\delta_A$ using a two-sided test is

$$\begin{aligned}\Pr[Z > z_{1-\alpha/2} \mid \delta_A] + \Pr[Z < z_{\alpha/2} \mid \delta_A] \\ \approx \Pr[Z > z_{1-\alpha/2} \mid \delta_A]\end{aligned}$$

- We need to know the distribution of $Z$ under $H_A$

$$E(p_2 - p_1) = \delta_A$$

$$\mathrm{Var}(p_2 - p_1) = \frac{\pi_2(1-\pi_2)}{N} + \frac{\pi_1(1-\pi_1)}{N}$$

## Two Sample Test: Binary Outcome

- So

$$
\begin{aligned}
1-\beta &= \Pr\left[\frac{p_2-p_1}{\sqrt{\frac{2\bar{p}\bar{q}}{N}}} > z_{1-\alpha/2} \mid \delta_A\right] \\
&= \Pr\left[p_2-p_1 > z_{1-\alpha/2}\sqrt{\frac{2\bar{p}\bar{q}}{N}} \mid \delta_A\right] \\
&= \Pr\left[\frac{(p_2-p_1)-\delta_A}{\sqrt{\mathrm{Var}(p_2-p_1)}} > \frac{z_{1-\alpha/2}\sqrt{\frac{2\bar{p}\bar{q}}{N}}-\delta_A}{\sqrt{\mathrm{Var}(p_2-p_1)}} \mid \delta_A\right]
\end{aligned}
$$

# Two Sample Test: Binary Outcome

- Implying

$$-z_{1-\beta} = \frac{z_{1-\alpha/2}\sqrt{2\bar{p}\bar{q}/N} - \delta_A}{\sqrt{\mathrm{Var}(p_2 - p_1)}}$$

- Using $\bar{p}\bar{q} \approx \bar{\pi}(1-\bar{\pi})$ where $\bar{\pi} = (\pi_1 + \pi_2)/2$ yields

$$z_{1-\beta}\sqrt{\mathrm{Var}(p_2 - p_1)} + z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})/N} = \delta_A$$

## Two Sample Test: Binary Outcome

- Therefore

$$\frac{z_{1-\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}+z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})}}{\delta_A}=\sqrt{N}$$

- Thus, the sample size required per arm to detect $\delta_A$ with power $1-\beta$ is

$$\frac{\left(z_{1-\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}+z_{1-\alpha/2}\sqrt{2\bar{\pi}(1-\bar{\pi})}\right)^2}{\delta_A^2}$$

# Two Sample Test: Binary Outcome

- In R by hand

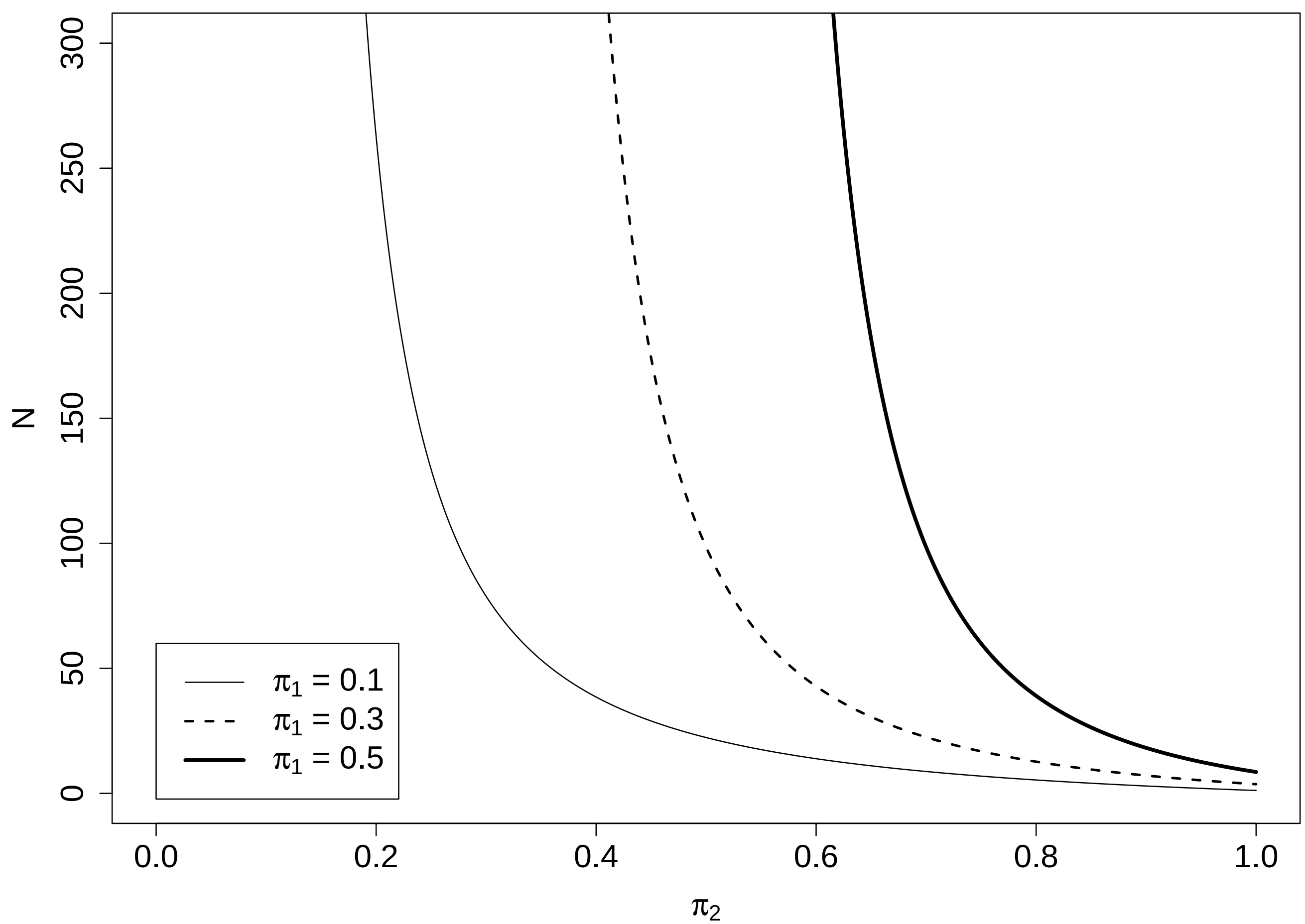
```
# sample size formula for comparing two
# binomial proportions based on Fleiss (second edition) page 41

ss_fleiss <- function(pi1,pi2,alpha,power){
   q1 <- 1-pi1
   q2 <- 1-pi2
   pbar <- (pi1+pi2)/2
   qbar <- 1-pbar
   num <- qnorm(1-alpha/2)*sqrt(2*pbar*qbar)+qnorm(power)*sqrt(pi1*q1+pi2*q2)
   den <- (pi2-pi1)
   (num/den)^2
}

> ss_fleiss(0.2,0.2727,0.05,0.9)
[1] 715.5618
```

## Graphical Summary

Sample size (per arm) for comparing $\pi_1$ against $\pi_2$ with $\alpha = 0.05$ (one-sided) and 90% power



# Case-Control: Binary Exposure

- Hypotheses

$$H_0 : \text{OR} = 1 \quad \text{vs.} \quad H_A : \text{OR} \neq 1$$

$$\text{OR} = \frac{\text{odds(disease} \mid \text{exposed)}}{\text{odds(disease} \mid \text{unexposed)}}$$

- Recall

| | Disease | No disease |
|---|---|---|
| Exposed | $\pi_{11}$ | $\pi_{12}$ |
| Unexposed | $\pi_{21}$ | $\pi_{22}$ |

# Case-Control: Binary Exposure

$$
\begin{aligned}
\text{OR} &= \frac{\text{odds(disease} \mid \text{exposed)}}{\text{odds(disease} \mid \text{unexposed)}} \\
&= \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} \\
&= \frac{\pi_{11}/\pi_{21}}{\pi_{12}/\pi_{22}} \\
&= \frac{\text{odds(exposed} \mid \text{disease)}}{\text{odds(exposed} \mid \text{no disease)}}
\end{aligned}
$$

# Case-Control: Binary Exposure

- Hypotheses

$$H_0 : \mathrm{OR} = 1 \ \text{ vs. } \ H_A : \mathrm{OR} \neq 1$$

- From the previous page

$$\mathrm{OR} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

  where $\pi_1 = \Pr(\text{exposed} \mid \text{case})$,
  $\pi_2 = \Pr(\text{exposed} \mid \text{control})$

- For specified OR and $\pi_2$ we can determine $\pi_1$

$$\pi_1 = \frac{\pi_2 \, \mathrm{OR}}{1 + \pi_2(\mathrm{OR} - 1)}$$

## Case-Control, Binary Exposure: Example

- Cases: Neural tube defect babies

  Controls: Normal babies

  Exposure: Self reported dieting to lose weight during first trimester

- It is estimated that 20% of women will diet to lose weight during pregnancy; $\pi_2 = 0.2$

- The investigator wants to be able to detect $\mathrm{OR} = 1.5$

## Case-Control, Binary Exposure: Example cont.

- First determine the corresponding value of $\pi_1$:

$$\pi_1 = \frac{0.2(1.5)}{1 + 0.2(0.5)} = 0.2727$$

- For $\pi_1 = 0.2727$, $\pi_2 = 0.2$, $\alpha = 0.05$ (two-sided), and $1 - \beta = 0.9$, the two sample binary outcome formula yields a sample size of $N = 716$ cases and $N = 716$ controls

# Case-Control Sample Size

- Cases are often harder to obtain than controls
- How many controls per case?
- Continuous exposure model
- Discrete exposure model

# Case-Control: Continuous Exposure

- Ury (Biometrics 1975)
- Cases

$$Y_{1i} = \mu_i + \delta + \epsilon_{1i}; \quad i = 1, \ldots, N$$

- Controls ($k$ for each case)

$$Y_{2ij} = \mu_i + \epsilon_{2ij}; \quad j = 1, \ldots, k$$

- Assume $\epsilon_{1i}, \epsilon_{2ij}$ iid with

$$E(\epsilon_{1i}) = E(\epsilon_{2ij}) = 0$$

$$\mathrm{Var}(\epsilon_{1i}) = \mathrm{Var}(\epsilon_{2ij}) = \sigma^2$$

# Case-Control: Continuous Exposure

- Let

$$\bar{Y}_{2i} = \frac{1}{k}\sum_{j=1}^{k} Y_{2ij}$$

- Then a consistent and unbiased estimator of the exposure effect is

$$\hat{\delta}_k = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - \bar{Y}_{2i}) \equiv \bar{Y}_1 - \bar{Y}_2$$

## Case-Control: Continuous Exposure

- By independence and homogeneity of variance assumptions

$$\text{Var}(\bar{Y}_1) = \frac{\sigma^2}{N} \quad \text{and} \quad \text{Var}(\bar{Y}_2) = \frac{\sigma^2}{kN}$$

- Therefore

$$\text{Var}(\hat{\delta}_k) = \frac{\sigma^2}{N}\left(\frac{k+1}{k}\right)$$

- For $k = 1$,

$$\text{Var}(\hat{\delta}_1) = \frac{2\sigma^2}{N}$$

## Case-Control: Continuous Exposure

- Relative efficiency

$$\text{eff}(\hat{\delta}_1, \hat{\delta}_k) = \frac{\text{Var}(\hat{\delta}_k)}{\text{Var}(\hat{\delta}_1)} = \frac{k+1}{2k} \rightarrow \frac{1}{2} \quad \text{as} \quad k \rightarrow \infty$$

| $k$ | $\text{eff}(\hat{\delta}_1, \hat{\delta}_k)$ |
|---|---|
| 1 | 1.00 |
| 2 | 0.75 |
| 3 | 0.67 |
| 4 | 0.63 |
| 5 | 0.60 |
| 10 | 0.55 |
| $\infty$ | 0.50 |

# Case-Control: Continuous Exposure

- Assuming $N$ large or $\epsilon_{1i}, \epsilon_{2ij} \sim N(0, \sigma^2)$
- Under $H_0 : \delta = 0$

$$Z = \frac{\hat{\delta}_k}{\sqrt{\text{Var}(\hat{\delta}_k)}} \sim N(0, 1)$$

- Under $H_A : \delta = \delta_A > 0$,

$$1 - \beta = \Pr\left[\frac{\hat{\delta}_k - \delta_A}{\sqrt{\text{Var}(\hat{\delta}_k)}} > z_{1-\alpha/2} - \frac{\delta_A}{\sqrt{\text{Var}(\hat{\delta}_k)}}\right]$$

## Case-Control: Continuous Exposure

- Implying

$$-z_{1-\beta} = z_{1-\alpha/2} - \frac{\delta_A}{\sqrt{\mathrm{Var}(\hat{\delta}_k)}}$$

$$(z_{1-\alpha/2} + z_{1-\beta})^2 = \frac{\delta_A^2}{\mathrm{Var}(\hat{\delta}_k)} = \delta_A^2 \frac{Nk}{\sigma^2(k+1)}$$

$$N = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta_A^2} \frac{(k+1)}{2k}$$

## Case-Control: Continuous Exposure

- So, for the two sample problem, compute the usual sample size $N$ per arm assuming an equal sample size per arm
- Multiply $N$ by $(k+1)/(2k)$ to get the number of cases
- Multiply $N$ by $(k+1)/2$ to get the number of controls

## Case-Control: Discrete Exposure

- The same relative efficiency result holds (Ury, Biometrics 1975)
- Here comparing 1:1 vs. $k$:1 controls to cases using a generalization of McNemar/MH
- Same sample size computation; cf. Note 17.2 in the text

## Case-Control, Discrete Exposure: Example

- Suppose that with one control per case, we calculate that 716 cases and 716 controls are needed to achieve a particular $\alpha$ and $\beta$
- Then with 2 controls per case, we need $716 \times 3/4 = 537$ cases and 1074 controls

# Outline

- Two sample: Continuous
- Two sample: Binary
- Case-control studies
- Estimating power with simulations

# Power and Sample Size

- Determination of power / sample size is important for many reasons
- Under-powered: May miss scientifically meaningful differences
- Over-powered: Waste of resources
- Ethics
- How can one compute power / sample size in more complicated situations than those addressed in the notes or text? For example, what is the power of the Kruskal-Wallis test for a fixed sample size?

# Sample Size Calculation by Simulation

- One approach: Conduct a simulation study

1. Simulate a single data set of size $N$ under a particular alternative
2. Evaluate test statistic for the simulated data set; record whether reject $H_0$
3. Repeat steps 1 and 2 multiple times (e.g., 10,000)
4. Compute the proportion of simulated data sets for which $H_0$ is rejected; this is an estimate of power
5. If the estimated power is larger than required, reduce $N$ and repeat steps 1-4; if the estimated power is too low, increase $N$ and repeat steps 1-4

# Simulated Power

- To help determine if the simulation is working correctly, check the following:
  - Simulate data sets under the null. Then the proportion of simulated data sets for which $H_0$ is rejected should approximate the specified type I error rate $\alpha$
  - As one moves away from $H_0$, the estimated power should increase towards 1
- In step 3 on the previous page, use a relatively small number of simulated datasets until close to the desired power, then increase the number of datasets to obtain a more accurate estimate

# Two Sample Test: $\sigma$ Unknown

- Recall

```
> power.t.test(59, delta=15, sd=25)

     Two-sample t test power calculation

              n = 59
          delta = 15
             sd = 25
      sig.level = 0.05
          power = 0.8982732
    alternative = two.sided

 NOTE: n is number in *each* group
```

- Let's run a simulation and compare the estimated power to this result

# Simulated Power Using R

```
set.seed(251); n <- 59; sd <- 25
mysim <- function(mdiff,nsims){
        rejects <- 0
        for (ii in 1:nsims){
                y1 <- rnorm(n,0,sd)
                y2 <- rnorm(n,mdiff,sd)
                tt <- t.test(y1,y2,var.equal=T)
                if (tt$p.value<0.05) rejects <- rejects + 1
                }
        print(paste("mdiff:",mdiff,", estimated power:",rejects/nsims))
        }
mysim(0,10000)
mysim(10,100)
mysim(10,100)
mysim(15,10000)
mysim(20,1000)
```

```
[1] "mdiff: 0 , estimated power: 0.0505"
[1] "mdiff: 10 , estimated power: 0.54"
[1] "mdiff: 10 , estimated power: 0.51"
[1] "mdiff: 15 , estimated power: 0.9019"
[1] "mdiff: 20 , estimated power: 0.993"
```

# Simulated Power Using SAS

```
%macro epower(mdiff=,seed=);

%let i=1;   %let n=59;   %let sd=25;   %let nsims=10000;

data;
  %do i = 1 %to &nsims;
    i=&i;
    do j=1 to &n; y=rannor(&seed)*&sd; group=1; output; end;
    do j=1 to &n; y=rannor(&seed)*&sd + &mdiff; group=2; output; end;
  %end;

ods output ttests=ttests;
proc ttest; class group; var y; by i; run;
data ttests; set ttests;
   if method="Pooled";
   reject=0; if Probt<0.05 then reject=1;

proc freq data=ttests; table reject; run;
%mend;

%epower(mdiff=15,seed=97231);
```

# Simulated Power Using SAS cont.

- Reason for:

```
if method="Pooled";
```

- Here's the dataset produced when $i = 1$:

```
Obs   i   Variable   Method          Variances    tValue       DF     Probt

 1    1      y       Pooled           Equal        -3.51      116    0.0006
 2    1      y       Satterthwaite    Unequal      -3.51   115.86    0.0006
```

- Output of the simulation;

```
                    The FREQ Procedure

                                          Cumulative    Cumulative
reject    Frequency      Percent          Frequency       Percent
------------------------------------------------------------
     0         1054        10.54              1054         10.54
     1         8946        89.46             10000        100.00
```