

CAUSAL CONTRASTS FROM EXPERIMENTAL STUDY DATA

In a randomized experimental study, inference concerning the causal effect of binary variable Z on Y can be made by direct comparison of sample averages. Suppose that $Z \sim \text{Bernoulli}(p)$ for $0 < p < 1$. Two estimators of the causal contrast

$$\delta = \mathbb{E}[Y(1) - Y(0)]$$

can be proposed:

$$\tilde{\delta} = \frac{1}{np} \sum_{i=1}^n Z_i Y_i - \frac{1}{n(1-p)} \sum_{i=1}^n (1 - Z_i) Y_i \quad (1)$$

and

$$\hat{\delta} = \frac{1}{n\hat{p}} \sum_{i=1}^n Z_i Y_i - \frac{1}{n(1-\hat{p})} \sum_{i=1}^n (1 - Z_i) Y_i \quad (2)$$

where

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n Z_i.$$

It can be verified that (2) has lower variance than (1). In the simulation below, we assume that

$$X \sim \text{Normal}(1, 1)$$

and

$$Y|X = x, Z = z \sim \text{Normal}(0.5x + \delta z, 1)$$

so that

$$\mathbb{E}_{Y|X,Z}[Y|X = x, Z = z] = 0.5x + \delta z$$

and

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 1] - \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 0] = \delta.$$

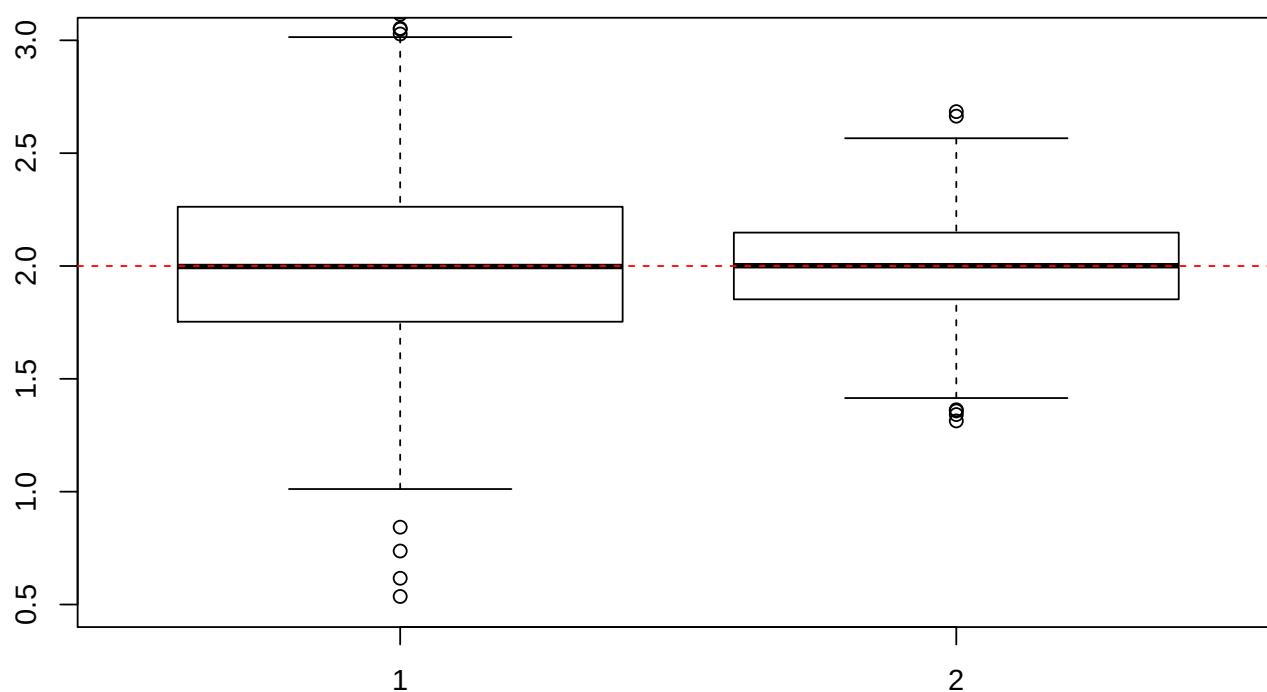
```

set.seed(23)
nreps<-1000;n<-100;p<-0.5;delta<-2
ests.mat<-matrix(0,nrow=nreps,ncol=2)
for(irep in 1:nreps){
  X<-rnorm(n,1,1)
  Z<-rbinom(n,1,p)
  Y<-rnorm(n,delta*Z+0.5*X,1)
  p.hat<-mean(Z)
  ests.mat[irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
  ests.mat[irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
}
var(ests.mat[,1])/var(ests.mat[,2])

+ [1] 3.042542

par(mar=c(4,2,1,1))
boxplot(ests.mat,ylim=range(0.5,3));abline(h=delta,lty=2,col='red')

```

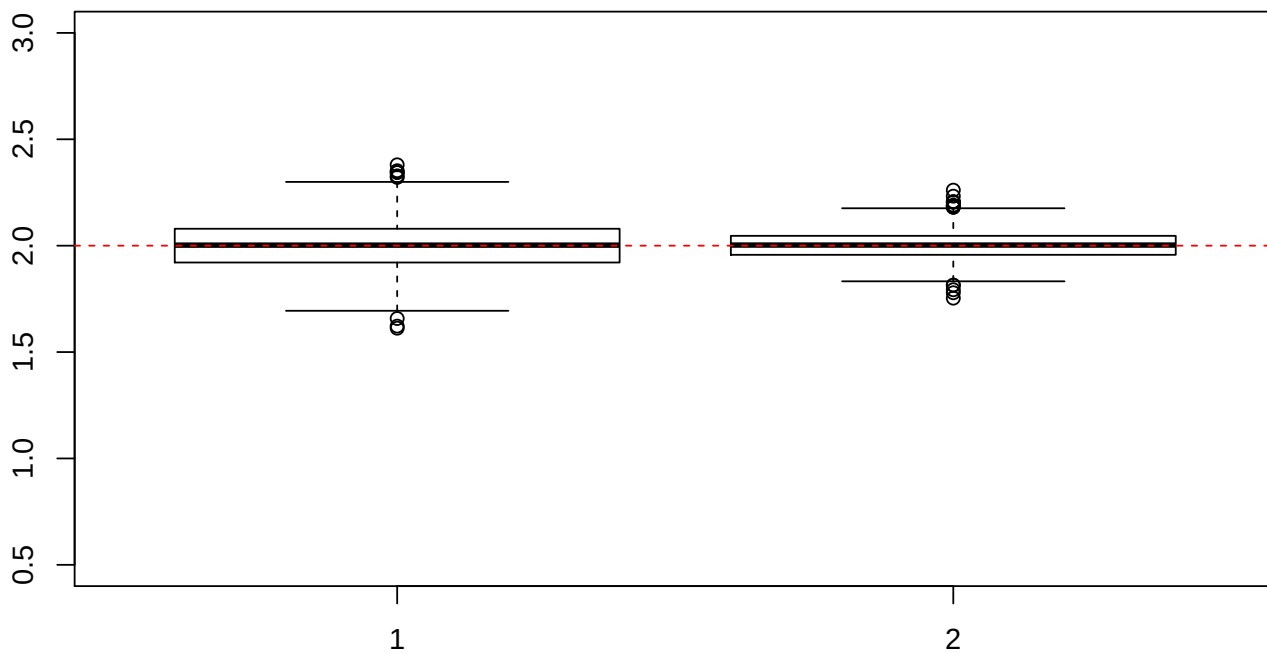


There is a smaller difference if n is larger:

```
nreps<-1000;n<-1000;p<-0.5;delta<-2
ests.mat<-matrix(0,nrow=nreps,ncol=2)
for(irep in 1:nreps){
  X<-rnorm(n,1,1)
  Z<-rbinom(n,1,p)
  Y<-rnorm(n,delta*Z+0.5*X,1)
  p.hat<-mean(Z)
  ests.mat[irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
  ests.mat[irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
}
var(ests.mat[,1])/var(ests.mat[,2])

+ [1] 2.854795

par(mar=c(4,2,2,1))
boxplot(ests.mat,ylim=range(0.5,3));abline(h=delta,lty=2,col='red')
```

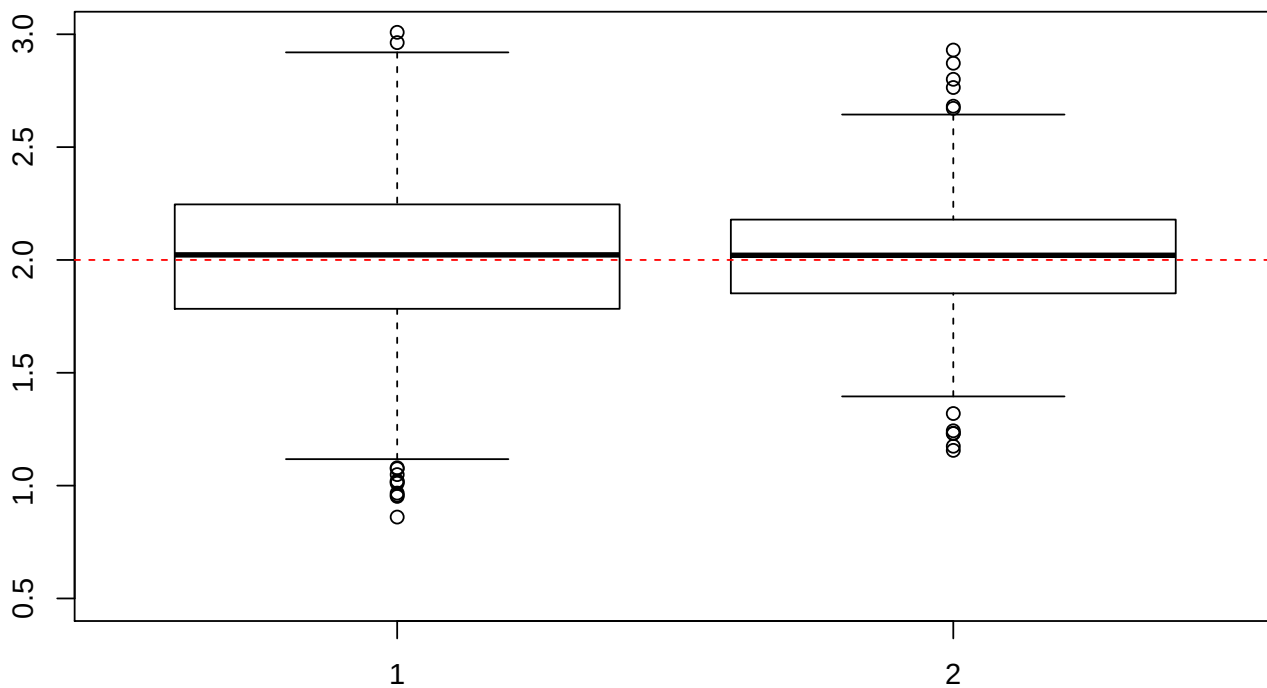


The effect is preserved if p differs from 0.5

```
set.seed(23)
nreps<-1000;n<-100;p<-0.75;delta<-2
ests.mat<-matrix(0,nrow=nreps,ncol=2)
for(irep in 1:nreps){
  X<-rnorm(n,1,1)
  Z<-rbinom(n,1,p)
  Y<-rnorm(n,delta*Z+0.5*X,1)
  p.hat<-mean(Z)
  ests.mat[irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
  ests.mat[irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
}
var(ests.mat[,1])/var(ests.mat[,2])

+ [1] 1.830722

par(mar=c(4,2,1,1))
boxplot(ests.mat,ylim=range(0.5,3));abline(h=delta,lty=2,col='red')
```



```

set.seed(23)
nreps<-1000;n<-100;p<-0.2;delta<-2
ests.mat<-matrix(0,nrow=nreps,ncol=2)
for(irep in 1:nreps){
  X<-rnorm(n,1,1)
  Z<-rbinom(n,1,p)
  Y<-rnorm(n,delta*Z+0.5*X,1)
  p.hat<-mean(Z)
  ests.mat[irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
  ests.mat[irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
}
var(ests.mat[,1])/var(ests.mat[,2])

+ [1] 4.506897

par(mar=c(4,2,1,1))
boxplot(ests.mat,ylim=range(0.5,3));abline(h=delta,lty=2,col='red')

```

