

Lesson 27

Professor Bradley Warner

October 28, 2016

Objectives:

1. Construct and interpret Wald, Score, Wilson, and Clopper-Pearson confidence intervals for a proportion
2. Conduct and interpret a paired hypothesis test using t-test, sign test, and confidence intervals
3. Know the assumptions of the paired t-test

Confidence Intervals for Proportions In Chapter 2 we made hypothesis tests about π the proportions of successes in a binomial trial. In this lesson we will generate confidence intervals.

To put things in context we will use an example. In game 7 of the 2014 World Series, the National League champion San Francisco Giants' pitcher Madison Bumgarner threw 50 strikes in 68 pitches. Let's use this data to estimate his true unknown proportion of strikes.

The first interval uses the normal approximation to the binomial. The confidence interval then is

$$\hat{\pi} \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

In R, the limits of the confidence interval are:

```
(phat<-50/68)
```

```
## [1] 0.7352941
```

```
phat+c(1,-1)*qnorm(.025)*sqrt(phat*(1-phat)/68)
```

```
## [1] 0.6304351 0.8401532
```

If I wanted to get fancy, I could write my own function in r;

```
wald.ci=function(x,n,conf.level=.95,alternative = c("two.sided", "less", "greater")){
  DNAME <- deparse(substitute(x))
  DNAME <- paste(DNAME, "successes and", deparse(substitute(n)), "trials")
  alternative <- match.arg(alternative)
  CONFINT <- switch(alternative,
    two.sided = x/n +
      c(-1,1)*qnorm(1-(1-conf.level)/2)*sqrt(x/n*(1-x/n)/n),
    less = c(-Inf, x/n+qnorm(conf.level)*sqrt(x/n*(1-x/n)/n)),
    greater = c(x/n-qnorm(conf.level)*sqrt(x/n*(1-x/n)/n), Inf)
  )
  attr(CONFINT,"conf.level")<-conf.level
  structure(list(conf.int=CONFINT,data.name=DNAME),class="htest")
}
```

```
wald.ci(50,68)
```

```
##  
##  
##  
## data: 50 successes and 68 trials  
##  
## 95 percent confidence interval:  
## 0.6304351 0.8401532
```

The score interval works by inverting the hypothesis test

```
(phat+qnorm(.025)^2/(2*68)+c(1,-1)*qnorm(.025)*sqrt(phat*(1-phat)/68+qnorm(.025)^2/(4*68^2)))/(1+qnorm(.025)^2/(2*68))
```

```
## [1] 0.6199227 0.8255026
```

This is also what we get from

```
prop.test(50,68,correct=FALSE)
```

```
##  
## 1-sample proportions test without continuity correction  
##  
## data: 50 out of 68, null probability 0.5  
## X-squared = 15.059, df = 1, p-value = 0.0001042  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.6199227 0.8255026  
## sample estimates:  
## p  
## 0.7352941
```

The Wilson is a plus 4 and we get as follows

```
(phat<-52/72)
```

```
## [1] 0.7222222
```

```
phat+c(1,-1)*qnorm(.025)*sqrt(phat*(1-phat)/72)
```

```
## [1] 0.6187638 0.8256807
```

The Clopper-Pearson comes from looking at a beta distribution. Here is my code

```
CPCI<-function(x, n, alternative = c("two.sided", "upper", "lower"),conf.level = 0.95, ... ){  
  alternative <- match.arg(alternative)  
  phat <- (x)/(n)
```

```

out <- list(method="Clopper-Pearson Confidence Interval")
class(out) <- 'htest'

out$parameter <- c("Sample size"=n,"Number of Successes" = x)

out$conf.int <- switch(alternative,
  two.sided = c(qbeta((1-conf.level)/2,x,n-x+1),qbeta(1-(1-conf.level)/2,x+1,n-x)),
  upper = c(0, qbeta(conf.level,x+1,n-x)),
  lower = c(qbeta(1-conf.level,x,n-x+1), 1)
)
attr(out$conf.int, "conf.level") <- conf.level

out$statistic <- phat
out$data.name <- "Data entered as summary numbers"
names(out$statistic) <- "Estimated probability of success"

return(out)
}

```

```
CPCI(50,68)
```

```

##
## Clopper-Pearson Confidence Interval
##
## data: Data entered as summary numbers
## Estimated probability of success = 0.73529, Sample size = 68,
## Number of Successes = 50
## 95 percent confidence interval:
## 0.6142896 0.8349615

```

But this is what `binom.test` returns

```
binom.test(50,68)
```

```

##
## Exact binomial test
##
## data: 50 and 68
## number of successes = 50, number of trials = 68, p-value =
## 0.0001308
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.6142896 0.8349615
## sample estimates:
## probability of success
## 0.7352941

```

The book mentions that the best comprise in terms of coverage is the score interval.

Paired Tests

```
## Warning: package 'fastR' was built under R version 3.2.2
```

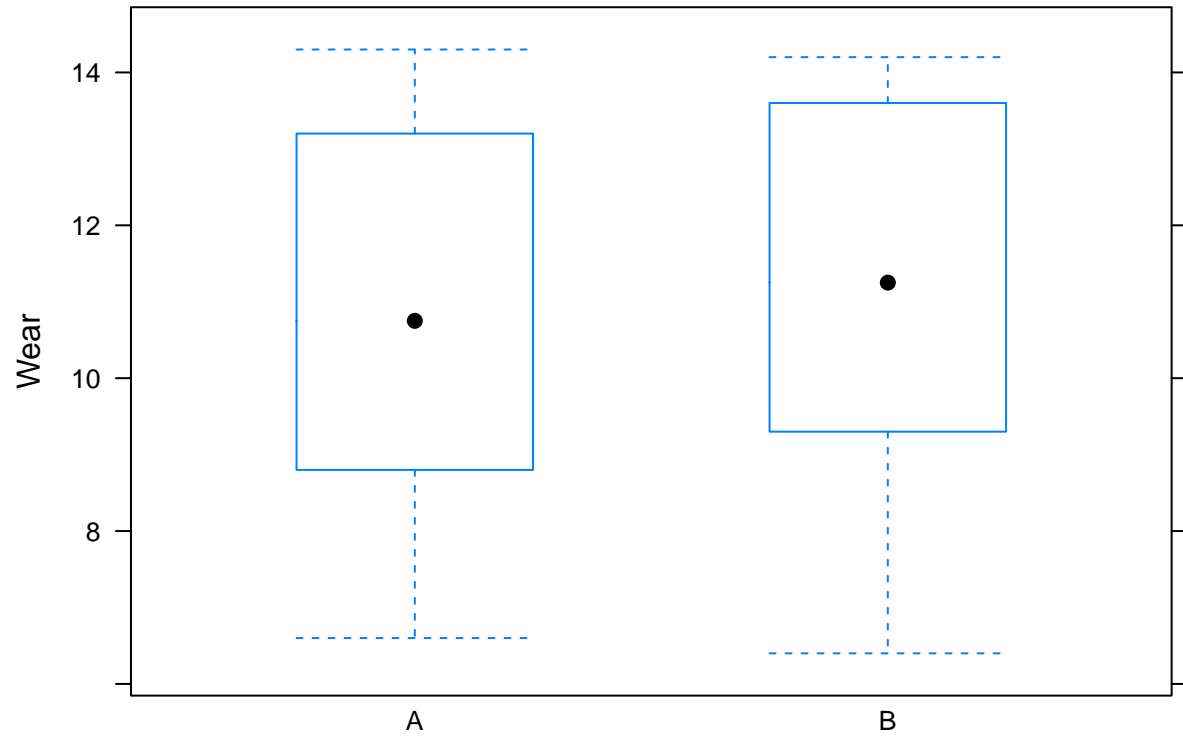
Suppose I want to test the wear of a material on the sole of a shoe. I put the shoes on 20 boys. Here is the data

```
Shoe
```

```
##      Boy Wear Material
## 1      1 13.2         A
## 2      2  8.2         A
## 3      3 10.9         A
## 4      4 14.3         A
## 5      5 10.7         A
## 6      6  6.6         A
## 7      7  9.5         A
## 8      8 10.8         A
## 9      9  8.8         A
## 10     10 13.3         A
## 11      1 14.0         B
## 12      2  8.8         B
## 13      3 11.2         B
## 14      4 14.2         B
## 15      5 11.8         B
## 16      6  6.4         B
## 17      7  9.8         B
## 18      8 11.3         B
## 19      9  9.3         B
## 20     10 13.6         B
```

If I simply look at the two materials it does not appear there is much of a difference

```
bwplot(Wear~Material,Shoe)
```

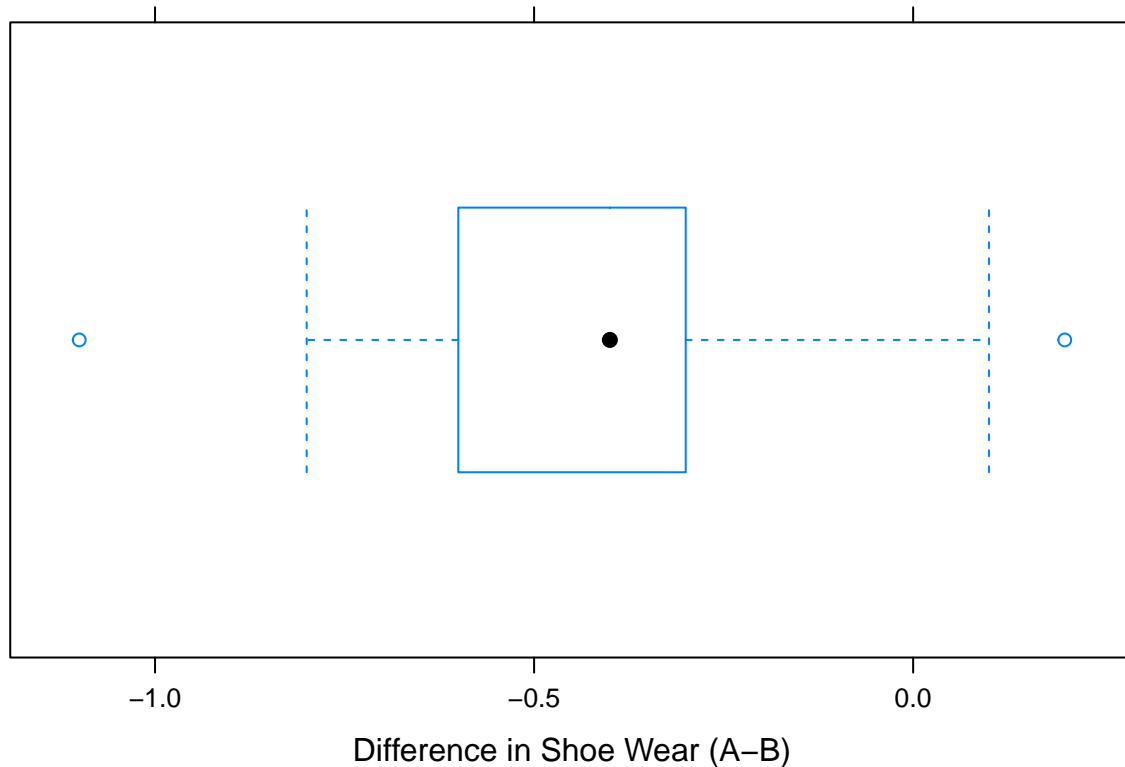


This is because there is so much variability in wear from boy to boy and this variance masks the difference in the materials. Thus the experiments put each type of material on each of the shoes for the boys. Thus one shoe had type A and the other type B. This is called a blocked design or paired design. Now I look at the difference in wear within the boys. This data looks like

```
Shoe[1:10,2]-Shoe[11:20,2]
```

```
## [1] -0.8 -0.6 -0.3  0.1 -1.1  0.2 -0.3 -0.5 -0.5 -0.3
```

```
bwplot(~Shoe[1:10,2]-Shoe[11:20,2],xlab="Difference in Shoe Wear (A-B)")
```



The hypothesis test is

$$H_0 : D = 0$$

$$H_A : D \neq 0$$

where D is the difference. If the differences are iid and normal, then I can use a t-test on this data.

First we could get the data in wide format

```
Shoe2<-reshape(Shoe,v.names="Wear",direction="wide",idvar="Boy",timevar="Material")
```

```
t.test(Shoe[1:10,2]-Shoe[11:20,2])
```

```
##
## One Sample t-test
##
## data: Shoe[1:10, 2] - Shoe[11:20, 2]
## t = -3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.6869539 -0.1330461
## sample estimates:
## mean of x
## -0.41
```

```
t.test(Shoe2["Wear.A"]-Shoe2["Wear.B"])
```

```
##
## One Sample t-test
##
## data: Shoe2["Wear.A"] - Shoe2["Wear.B"]
## t = -3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.6869539 -0.1330461
## sample estimates:
## mean of x
## -0.41
```

```
t.test(diff(Shoe[,2],lag=10))
```

```
##
## One Sample t-test
##
## data: diff(Shoe[, 2], lag = 10)
## t = 3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.1330461 0.6869539
## sample estimates:
## mean of x
## 0.41
```

or

```
t.test(Wear~Material,data=Shoe,paired=T)
```

```
##
## Paired t-test
##
## data: Wear by Material
## t = -3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.6869539 -0.1330461
## sample estimates:
## mean of the differences
## -0.41
```

We conclude that there is a difference in the mean differential wear on the materials.

Signed test If the assumption that the differences are normal is not reasonable, we could use the sign test. This test makes on assumptions about the distribution other than it is continuous. The idea is that if the two materials came from the same distribution, then half the time one should be larger than the other. Thus the hypothesis is

$$H_0 : \pi = 0.5$$

$$H_A : \pi \neq 0.5$$

I can test this with the binomial `binom.test`.

```
Shoe[1:10,2]>Shoe[11:20,2]
```

```
## [1] FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
```

```
sum(Shoe[1:10,2]>Shoe[11:20,2])
```

```
## [1] 2
```

```
binom.test(sum(Shoe[1:10,2]>Shoe[11:20,2]),10)
```

```
##
##
##
## data: sum(Shoe[1:10, 2] > Shoe[11:20, 2]) out of 10
## number of successes = 2, number of trials = 10, p-value = 0.1094
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.02521073 0.55609546
## sample estimates:
## probability of success
## 0.2
```

So we fail to reject that there is a difference. Because of the small sample size, the fewer assumptions makes this test have less power.

In the case of ties, most people exclude that observation. Here is my function to do this.

```
sign.test=function(x, y = NULL, md = 0, alternative = c("two.sided", "less", "greater"), conf.level = 0.95)
{
  if(is.null(y)) y=rep(md,length(x))
  if(sum(which(x==y))!=0){
    xx=x
    yy=y
    x=xx[-1*which(xx==yy)]
    y=yy[-1*which(xx==yy)]
  }
  ans=binom.test(sum(x>y),length(x),alternative=alternative,conf.level=conf.level)
  ans$method="Sign Test"
  return(ans)
}
```

```
sign.test(Shoe[1:10,2],Shoe[11:20,2])
```

```
##
## Sign Test
##
```



```
## data:  sum(x > y) out of length(x)
## number of successes = 2, number of trials = 10, p-value = 0.1094
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.02521073 0.55609546
## sample estimates:
## probability of success
##                0.2
```