Swan capture power analysis

David Wolfson

3/21/2020

Objective: Determine the amount of power provided by various sample sizes of categories of swans.

As we strategize a sampling strategy for the 2020 field season, it's worth considering if capturing a segment of non-breeders and/or cygnets will allow for any statistically significant comparisons. Examples could be the proportion of swans in an age category that migrate or stay resident during the winter, etc.

```
library(pwr)
library(tidyverse)
```

Simulate dataset of different sample sizes for two categories

```
n1<-seq(0, 100, 5)
p1<-seq(0,1, .05)

df<-expand.grid(n1=n1, p1=p1)
df$n2<-100-df$n1
df$p2<-1-df$p1
```

Write function for magnitude of difference of 2 proportions

```
h_func<-function(p1, p2){
  h=abs(2*asin(sqrt(p1))-2*asin(sqrt(p2)))
  return(h)
}

h.temp<-NA
for(i in 1:nrow(df)){
  h.temp<-h_func(df[i,'p1'], df[i,'p2'])
  df[i,5]<-h.temp
}
colnames(df)[5]<-'h'
```

Calculate difference of proportion power calculation for binomial distribution (arcsine transformation) Significance level 0.05

```
df$power<-NA
tmp<-NA
for(i in 1:nrow(df)){
  if(df[i,'n1']>2 & df[i,'n2']>2){
tmp<-pwr.2p2n.test(h = df[i, 'h'], n1 = df[i, 'n1'], n2 = df[i, 'n2'], sig.level = 0.05)
df[i,'power']<-tmp$power
}
}
```

```r
#omit rows with 0's in groups because can't calculate power
df<-df[df$n1!=0,]
df<-df[df$n2!=0,]
```

Ok, now we have power calculations. A good ballpark is 0.8 or higher Next step is to make some plots

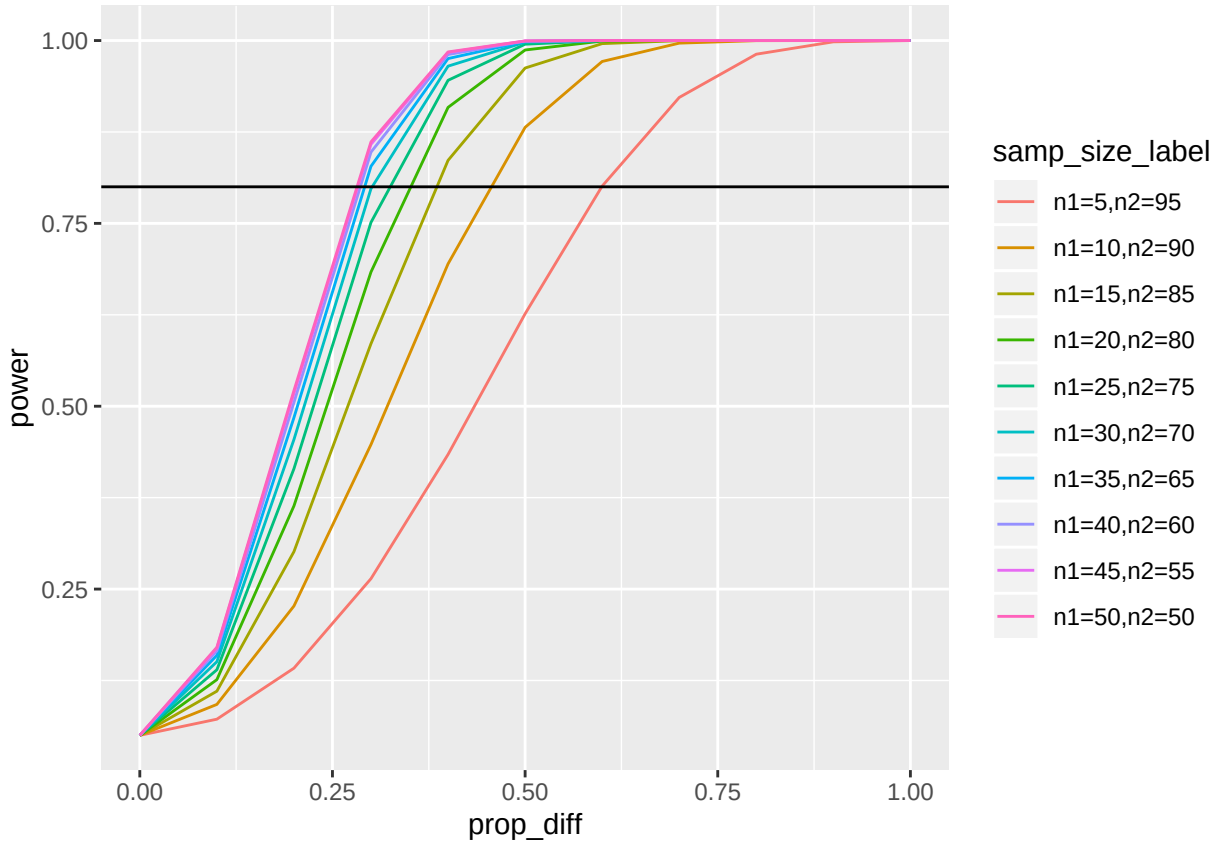
```r
df$samp_size_label<-as.factor(paste(paste0('n1=', df$n1),paste0('n2=', df$n2), sep = ','))
df$prop_diff<-abs(df$p1-df$p2)

#don't need repetitive combinations
labs_to_use<-as.factor(c('n1=5,n2=95', 'n1=10,n2=90', 'n1=15,n2=85', 'n1=20,n2=80',
                         'n1=25,n2=75', 'n1=30,n2=70','n1=35,n2=65', 'n1=40,n2=60',
                         'n1=45,n2=55', 'n1=50,n2=50'))

df2<-df[df$samp_size_label%in%labs_to_use,]

df2$samp_size_label<-fct_relevel(df2$samp_size_label, c('n1=5,n2=95', 'n1=10,n2=90',
                                  'n1=15,n2=85', 'n1=20,n2=80', 'n1=25,n2=75',
                                  'n1=30,n2=70','n1=35,n2=65', 'n1=40,n2=60',
                                  'n1=45,n2=55', 'n1=50,n2=50'))

ggplot(df2, aes(x=prop_diff, y=power,colour=samp_size_label))+
  geom_line()+geom_hline(aes(yintercept=0.8))
```



Kinda hard to think about both sample size and difference in proportions at the same time. Better way to visualize?