Confidence interval - proportion

In this back page lab we are going to establish a confidence interval for a proportion. The background to this is that we are publishing a paper which describes the highly pathogenic avian influenza outbreak which occurred in the Klein Karoo in ostriches during 2011. One of our epidemiologic variables we want to include is a proportion of farms within our control area that ended up being positive. This will hopefully help future epidemiologists as a between farm prevalence is often used in working out a sampling frame for a surveillance strategy. To work out the proportion is very easy (positive farms divided by the population of farms at risk) However, in order to show how confident we are that our between farm prevalence is accurate we wish to add a 95% confidence interval into our proportion because we know that we sampled the majority of farms in the area and we believe our sample strategy was complete enough for an accurate estimate of between farm prevalence. The R code below is what we used to establish our 95% confidence interval of our between farm prevalence for high pathogenic avian influenza in the control zone we established in Oudtshoorn.

Epi Lab color code

Software/Packages/Add-ins required

Software/Packages/Add-ins recommended

Description text

R code to copy/paste into console

R code to copy/paste into console that needs adjustment to your personal workspace

Websites where you can download requirements

Lab #2 requirements

- R - http://cran.r-project.org/bin/windows/base/
- R Studio - www.rstudio.com/ide/download/desktop
- prevalence R package
- Internet connection
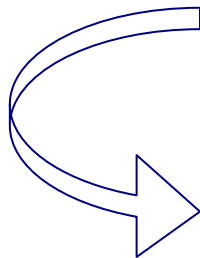- JAGS - http://sourceforge.net/projects/mcmc-jags/

The code

```
#remember that you can just copy and paste the blue lines of data into your R Studio console
#we import the dataset which is a list of farms, their intermediate disease status and their final status based on whether
#they were within the control area we were evaluating. In the dataset I have omitted the column names to illustrate how
#to allocate column names to a dataset. Below we import the dataset and allocate it to a variable called x.
x <- read.csv('http://www.jdata.co.za/backpagelabs/backpagelabs_jdg_ci.txt', header=F)
#set the column names - first column is a reference number per farm, intermediate is the TRUE/FALSE status of disease
#and final status is the disease status of only those farms in our control area
colnames(x)<-c("Ref","Intermediate","FinalStatus")
#now view the top and bottom 6 rows of data in the x variable we have allocated the data to
head(x)
tail(x)
#note that the last two farms, while positive were no in our control area, so now we must exclude them from our analysis
#we use the na.omit function for this purpose and we make a new data set called finalstatuslist
#PLEASE NOTE - the way this seems to work well in R is if the empty data is represented in your source data as NA (not N/A
#or by a blank entry)
finalstatuslist<-na.omit(x$FinalStatus)
#lets see how many rows of data were in our original imported set - should be 248 farms
summary(x)
#now we look at how many rows are in our data where NA has been omitted - should be 246 farms
length(finalstatuslist); summary(finalstatuslist)
#for the denominator for prevalence we need the population at risk (PAR) so lets make this variable
PAR<-length(finalstatuslist)
PAR
```

```
#so our total population at risk is 246 farms
#now we need the number of positive farms for our numerator data
POS<-sum(finalstatuslist == "Positive")
#this code essentially sums the events that are Positive (each positive is taken as 1) in our final data set
POS
#so our number of Positive farms totals 40
#a basic prevalence is therefore calculated by:
POS/PAR
# now what this lab is for - the 95% confidence interval. An easy (there are others) way of getting confidence interval
#data for a proportion is by using a function propCI from a package called "prevalence"
#note that this requires the "prevalence" package but also you'll need to install a program called JAGS from the internet
#install it from the website listed under LAB REQUIREMENTS.
#if you haven't installed the prevalence package yet then type this into your R console
install.packages("prevalence")
#now to load the newly installed package
library(prevalence)
#Now for working out the confidence interval - we use the function propCI
propCI(x=POS,n=PAR)
# Here the positive total is the POS data variable we made, the total sampled is the PAR data variable. So you'll see that
#5 different CI's are given. We wont go into it here but they all have differing reasons for being used. Because our
#sample size is relatively big the different CI methods have very similar CI's of between 11.64% and 21.47%.
#for our research we would use the WALD method so I will be using the 4th row of information
#lets try to isolate the row of info we will be using
propCI(x=POS,n=PAR)[4,]
#in summary - x = 40 and is our positive farms, n = 246 and is our population at risk, p = prevalence of 0.1626 (so
16.26%) with a CI of between 0.1164 (so 11.6%) and 20.87%.Our confidence interval confidence level is 95% which is a
standard but it can be changed if you wish
#for our publication we will say that the between farm prevalence of highly pathogenic avian influenza within the
controlled area was 16.26% (95 Conf: 11.6%-20.87%)
```

The output

```
     x   n          p     method level     lower      upper
1   40 246  0.1626016 agresti.coull  0.95 0.1214537  0.2141249
2   40 246  0.1626016        exact  0.95 0.1187775  0.2147668
3   40 246  0.1626016     jeffreys  0.95 0.1205557  0.2125299
4   40 246  0.1626016         wald  0.95 0.1164901  0.2087131
5   40 246  0.1626016       wilson  0.95 0.1217406  0.2138381


     x   n          p method level     lower      upper
4   40 246  0.1626016   wald  0.95 0.1164901  0.2087131
```

Citations

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical
Computing, Vienna, Austria. URL http://www.R-project.org/

Brecht Devleesschauwer, Paul Torgerson, Johannes Charlier, Bruno Levecke, Nicolas Praet, Pierre Dorny, Dirk
Berkvens and Niko Speybroeck (2013). prevalence: Tools for prevalence assessment studies. R package version 0.2.0.
http://cran.r-project.org/package=prevalence