Homework 4

Dominic LaRoche

Wednesday, April 29, 2015

1. [10pts] Is there a difference in knowledge between the two arms following the intervention? Use an unadjusted mixed model on the individual level data to determine this, while accounting for the clustering design. Show relevant code and output. Write your answer in a few sentences appropriate for a journal. Include both your statistical methods (briefly) and the results.

m1<-lmer(kscore~arm+(1|school),data=share)  
KRsumm<-lmerTest::summary(m1,ddf="Kenward-Roger")  
KR.t<-qt(.975,df=KRsumm$coef[2,3])  
KRsumm$coef[2,1]

## [1] 0.5374958

KR.95CI<-round(c(KRsumm$coef[2,1]-KR.t\*KRsumm$coef[2,2],KRsumm$coef[2,1]+KR.t\*KRsumm$coef[2,2]),2)  
KR.95CI

## [1] 0.18 0.89

We tested for a difference in mean knowlege between the two arms post-intervention using a linear mixed-effects model, with a random intercept for each school, to account for the clustering of individuals within each school. We found a significantly higher mean level of knowledge among those in the intervention arm with a difference of 0.54 points (95% Confidence Interval; 0.18, 0.89).

1. [4pts] Using your model from 1, what is the intracluster correlation?

schoolVar<-summary(m1)$varcor$school[1]  
errorVar<-(summary(m1)$sigma)^2  
icc<-schoolVar/(schoolVar+errorVar)  
icc

## [1] 0.02801898

The ICC for this data is 0.028.

1. [4pts] What is the design effect for this study? What does this mean?

#calculate adjusted mean cluster size  
n<-dim(share)[1]  
m.prime<-(1/length(levels(factor(share$school))))\*(n-(sum(summary(factor(share$school))^2)/n))  
#calculate design effect  
des.eff<-1+(m.prime-1)\*icc  
des.eff

## [1] 7.241192

The design effect for this study is 7.24. This represents the relative increase in variance of the overall mean that is induced by the clustering design.

1. [10pts] Analyze this study at the cluster level as outlined in the Wears paper. Show relevant code and output. Write your answer in a few sentences appropriate for a journal. Include both your statistical methods (briefly) and the results. Compare your answer to the results you found by analyzing at the individual level. SAS users should use the ods output statement in a proc means. Stata users should use the collapse command.

#aggregate responses at the cluster level  
clust.level<-aggregate(share$kscore, list(school=share$school, arm=share$arm), mean, na.rm=T)  
m2<-lm(x~arm,data=clust.level)#fit a linear model (equivalent to a t-test)  
summ2<-summary(m2)  
summ2

##   
## Call:  
## lm(formula = x ~ arm, data = clust.level)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.9055 -0.3075 0.0596 0.3325 0.6011   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.1178 0.1285 32.05 < 2e-16 \*\*\*  
## armIntervention 0.5150 0.1782 2.89 0.00826 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4451 on 23 degrees of freedom  
## Multiple R-squared: 0.2664, Adjusted R-squared: 0.2345   
## F-statistic: 8.353 on 1 and 23 DF, p-value: 0.008257

m2.95CI<-round(c(summ2$coef[2,1]-qt(.975,df=summ2$df[2])\*summ2$coef[2,2], summ2$coef[2,1]+qt(.975,df=summ2$df[2])\*summ2$coef[2,2]),2)

We tested for a difference in post-intervention knowledge between the two arms by testing for a difference in the mean knowledge among schools that recieved the intervention versus the mean knowledge among schools that received the control. To do this we calculated the mean knowledge at each school and compared the schools on the intervention to those which were not with a t-test. We found significantly higher knowledge in schools that received the intervention compared to control schools with a difference of 0.51 (95% CI; 0.15, 0.88). This confidence interval is slightly wider than that acheived by the mixed model using individual data due to the slightly larger standard errors on the estimates (the df are roughly the same). Moreover, the point estimate appears to be biased towards 0, which might be die to the uneven cluster sizes.

5.[5pts] Your research team is designing a study to investigate a new pain regimen, as compared to usual care, in a two arm RCT. Calculate the required sample size to detect a standardized effect size of 0.2 (a small effect) with 80% power, using a 2:1 allocation to intervention and control respectively.

#create sample size function for comparing means  
ssize<-function(alpha, beta, sigma, k, margin){  
 n2<- (qnorm(1 - alpha/2) + qnorm(1 - beta))^2 \* sigma^2 \* (1 + 1/k)/margin^2  
 n1<-k\*n2  
 return(list(N=ceiling(n1)+ceiling(n2),N1=ceiling(n1),N2=ceiling(n2)))  
}  
Q5<-ssize(alpha=.05, beta=.2, k=2, sigma=1, margin=.2)  
Q5

## $N  
## [1] 884  
##   
## $N1  
## [1] 589  
##   
## $N2  
## [1] 295

We would need 884 total participants to acheive the desired power.

1. [5pts] Replicate the sample size calculation in the SUPPORT trial. You may not be able to get the exact number, but you should be close. (Try to duplicate their results)

#create sample size function for comparing proportions  
pssize<-function(alpha, beta, p1, p2){  
 delta<-abs(p1-p2)  
 n <- (qnorm(1 - alpha/2) + qnorm(1 - beta))^2 \* (p1 \* (1 - p1) + p2 \* (1 - p2))/delta^2  
 N<-2\*n  
 return(ceiling(N))  
}  
N<-pssize(alpha=.02, beta=.18, p1=.47, p2=.37)  
#adjust for twins  
N<-1.12\*N  
#adjust for attrition  
N<-ceiling((.17\*N)+N)  
#using 82% power gets me the closest  
N

## [1] 1329

I got the closest using a power of 82%, which gave a sample size of 1329 individuals.

1. [8 pts] Replicate the original sample size calculations in Livingstone et al (2011). You may not be able to get the exact number, but you should be close.

#calculate pooled variance  
p.sigma<-sqrt(.5\*(3.5^2+4^2))  
#Calculate initial sample size with pooled variance  
careN<-ssize(alpha=.05, beta=.1, sigma=p.sigma, k=1, margin=2)  
careN

## $N  
## [1] 150  
##   
## $N1  
## [1] 75  
##   
## $N2  
## [1] 75

d.eff<-1+(30-1)\*.03  
#account for attrition  
careN<-lapply(careN,FUN=function(x){(x\*.2)+x})  
#correct intervention arm for clustering  
careN$N2<-ceiling(careN$N2\*d.eff)  
careN[2:3]

## $N1  
## [1] 90  
##   
## $N2  
## [1] 169

I got closer on this one with sample sizes in the two arms of 90 and 169.

1. [5 pts] You are writing an NIH R21 grant, which is a pilot and feasibility study mechanism. The feasibility outcomes are the percentage of participants that are screened and found to be eligible, and the percentage of enrolled participants that remain in the study for the entire follow-up time. If the researchers would like their estimates of these two percentages to be within 10% of the “true” values, with 95% confidence, what should the sample size be? Write a few sentences that would be appropriate for the grant proposal.

#using the formula for a 95% confidence interval on a proportion and the fact that the variance is maximized at p=.5 we get:  
PilotN<-ceiling((.5\*.5)/(.1/1.96)^2)  
PilotN

## [1] 97

We determined that we would need a sample size of 90 participants to ensure 95% confidence intervals around each estimate would not exceed 10 percentage points in either direction (resulting in a total maximum 95% CI width of 20 percentage points). We calculated the necesarry sample size by rearranging the 95% CI formula:

to:

by noting that the interval is maximized at and solving for .

1. [6pts] The following formulae are for three alpha spending functions. Assuming theta = 1 and using each of the alpha spending functions below, find the alpha for early stopping when:

#define functions  
OF<-function(t, alpha=.05){  
 Alpha<-2-2\*pnorm(abs(qnorm(alpha/2)/sqrt(t)))  
 return(Alpha)  
}  
Poc<-function(t, alpha=.05){  
 Alpha<-alpha\*log(1+(exp(1)-1)\*t)  
 return(Alpha)  
}

1. one quarter of the information in the trial has been obtained.

OF(.25)

## [1] 8.857544e-05

Poc(.25)

## [1] 0.0178687

.05\*.25

## [1] 0.0125

1. one half of the information in the trial has been obtained.

OF(.5)

## [1] 0.005574597

Poc(.5)

## [1] 0.03100573

.05\*.5

## [1] 0.025

1. three quarters of the information in the trial has been obtained.

OF(.75)

## [1] 0.02362512

Poc(.75)

## [1] 0.04139945

.05\*.75

## [1] 0.0375