1.使用Hoeffding's檢定發現綜觀這18年來，僅臺南市、高雄市、新竹縣、澎湖縣、新竹市的出生率跟平均薪資成相關(不獨立)，同時這些縣市的薪資與出生率皆呈現負相關，另外，也可以發現從2011年之後整體薪資與出生率皆呈現正相關。

2.若改變資料，看新資是否高於平均與出生率是否高於平均的關係，可以經由mcnemar檢定發現從2005年開始，有顯著一致性

3.將縣市區分為南北，北部為台中以北，南部為台中以南，並扣除花東，由無母數曼尼檢定可以知道，在出生率的部分，每一年皆呈現南北台灣的出生率一致，可見出生率不受是否為南北縣市所影響

4.使用無母數Wilcoxon pair檢定可以發現2016年的出生率明顯較1999-2007、2012低，卻也明顯較2010高

pay <- read.csv("M:/106s/cloud/hw1/hw1/aa.csv", header=F)

colnames(pay) <- c("year",cityname)

birth <- read.csv("M:/106s/cloud/hw1/hw1/a2.csv", header=T)

pay <- pay[, -(23:24)]

birth <- birth[2:19, 1:22]

install.packages("testforDEP",repos="http://cran.us.r-project.org")

library(testforDEP)

cor\_city <- c()

cor\_year <- c()

col\_h <- c()

hoe\_p\_city <- c()

hoe\_d\_city <- c()

hoe\_p\_year <- c()

hoe\_d\_year <- c()

mar\_p <- c()

mann\_birth\_p <- c()

mann\_birth\_w <- c()

wil\_pair\_p <- c()

wil\_pair\_w <- c()

all\_agr <- 0

all\_neg <- 0

p\_agr\_b\_neg <- 0

p\_neg\_b\_agr <- 0

cityname <- c("區域別總計","新北市","臺北市","桃園市","臺中市","臺南市","高雄市","宜蘭縣","新竹縣","苗栗縣","彰化縣","南投縣","雲林縣","嘉義縣","屏東縣","臺東縣","花蓮縣","澎湖縣","基隆市","新竹市","嘉義市")

#test1-1

windows()

plot(1, type="n", xaxt="n", xlab="", ylab="p-value", main = "p-value for hoeffding test for 18 years", ylim=c(0, 1), xlim=c(0, 45))

abline(h=0.05)

for (i in 2:22){

hoe\_p\_city[i-1] <- hoeffd(pay [, i], birth[, i])$P[1, 2]

hoe\_d\_city[i-1] <- hoeffd(pay [, i], birth[, i])$D[1, 2]

cor\_city[i-1] <- cor(as.numeric(pay [, i]), as.numeric(birth[, i]))

col\_h[i-1] <- ifelse(hoe\_p\_city[i-1] <= 0.05, "red", "black")

points(x=2\*i-1, y=hoe\_p\_city[i-1], col=col\_h[i-1])

text(x=2\*i-1, y=hoe\_p\_city[i-1]-0.03, labels=cityname[i-1], cex=0.7, col=col\_h[i-1])

}

windows()

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("hoeffding test and cor btw birth and pay", "for all the cities", sep="\n"), ylim=c(0, 22), xlim=c(0, 4))

abline(v=1:3)

abline(h=1:21)

for (i in 1:21){

text(0.5, i-0.5, labels=cityname[i], col = col\_h[i])

text(1.5, i-0.5, labels=round(hoe\_d\_city[i], 4), col=col\_h[i])

text(2.5, i-0.5, labels=round(hoe\_p\_city[i], 4), col=col\_h[i])

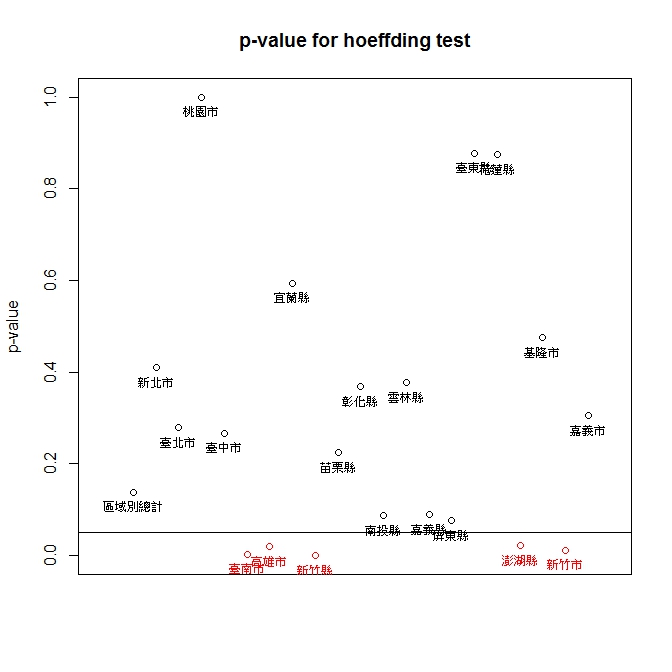
text(3.5, i-0.5, labels=round(cor\_city[i], 4), col=col\_h[i])

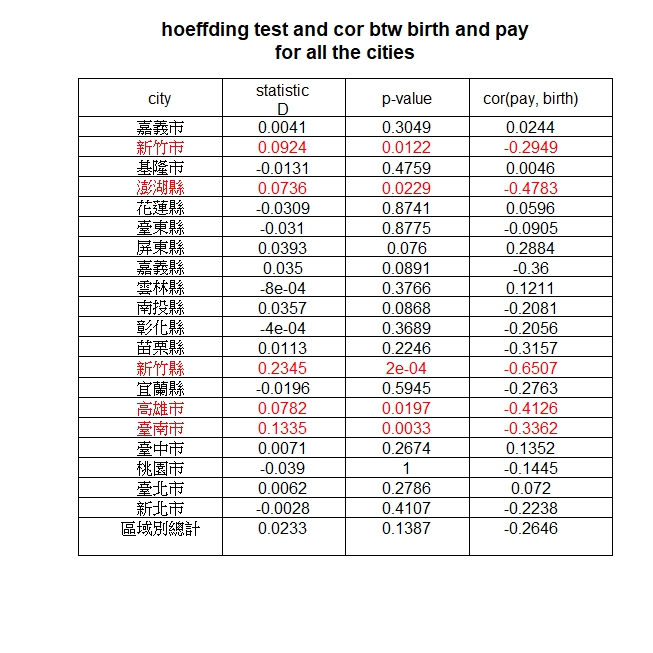
}

text(0.5, 21.9, labels="city")

text(1.5, 21.9, labels=paste("statistic" ,"D", sep="\n"))

text(2.5, 21.9, labels="p-value")

text(3.5, 21.9, labels="cor(pay, birth)")



#test1-2

for (i in 1:18){

hoe\_p\_year[i] <- hoeffd(as.numeric(pay[i, -1]), as.numeric(birth[i, -1]))$P[1, 2]

hoe\_d\_year[i] <- hoeffd(as.numeric(pay[i, -1]), as.numeric(birth[i, -1]))$D[1, 2]

cor\_year[i] <- cor(as.numeric(pay[i, -1]), as.numeric(birth[i, -1]))

}

hoe\_p\_year <- round(hoe\_p\_year, 4)

hoe\_d\_year <- round(hoe\_d\_year, 4)

cor\_year <- round(cor\_year, 4)

windows()

col\_h\_year <- ifelse(hoe\_p\_year <= 0.05, "red", "black")

plot(x=as.numeric(pay[, 1]),hoe\_p\_year, xlab="", ylab="p-value", main = "p-value for hoeffding test for 18 years", ylim=c(0, 1), xlim=c(1998, 2017), col=col\_h\_year)

abline(h=0.05)

windows()

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("hoeffding test and cor btw birth and pay", "for all years", sep="\n"), ylim=c(0, 19), xlim=c(0, 4))

abline(v=1:3)

abline(h=1:18)

for(i in 1:18){

text(0.5, i-0.5, labels = i+1998, col=col\_h\_year[i])

text(1.5, i-0.5, labels = hoe\_d\_year[i] , col=col\_h\_year[i])

text(2.5, i-0.5, labels = hoe\_p\_year[i] , col=col\_h\_year[i])

text(3.5, i-0.5, labels = cor\_year[i] , col=col\_h\_year[i])

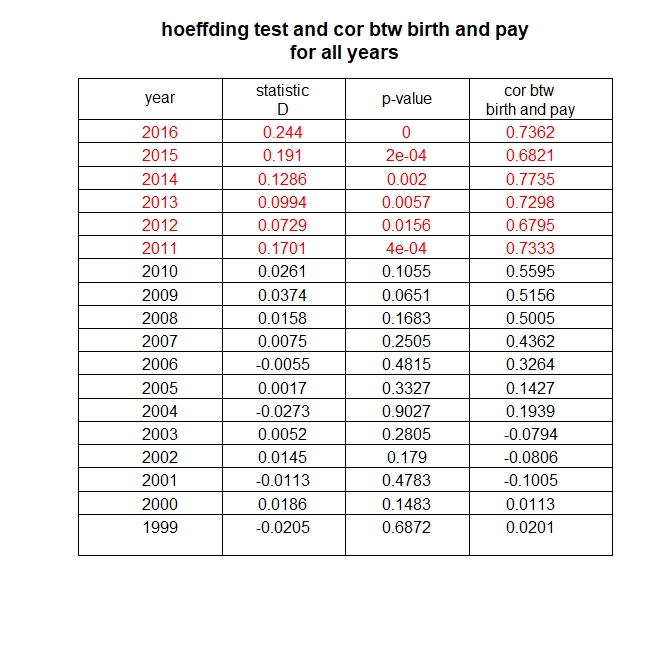
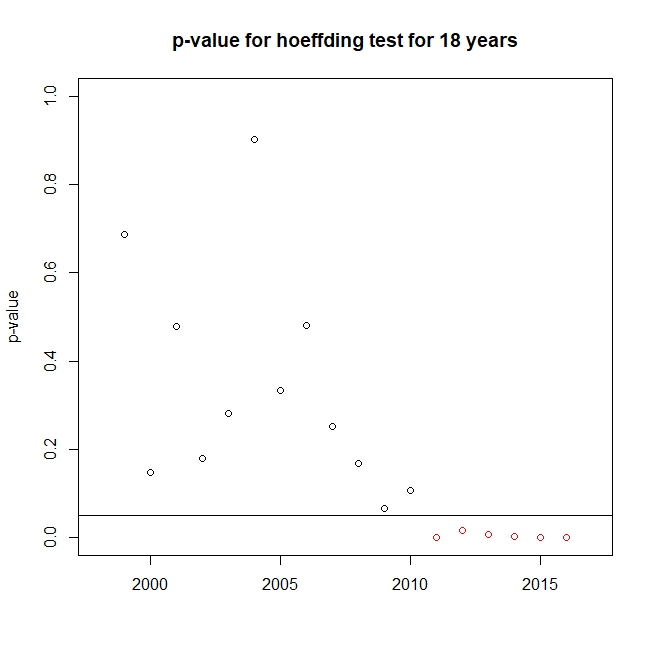
}

text(0.5, 18.9, labels = "year")

text(1.5, 18.9, labels = paste("statistic" ,"D", sep="\n"))

text(2.5, 18.9, labels = "p-value")

text(3.5, 18.9, labels = paste("cor btw ", "birth and pay", sep="\n"))



#test1-3

windows()

par(mfrow=c(1,2))

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("cor btw birth and pay", paste("for the cities", "that birth and pay are depandent",sep="\n"), sep="\n"), ylim=c(0, 6), xlim=c(0, 1))

abline(h=c(1:5))

abline(v=0.5)

for(i in 1:5){

id\_city <- c(7, 8, 10, 19, 21)

id\_city\_name <- c("臺南市", "高雄市", "新竹縣", "澎湖縣", "新竹市")

cor\_city[i] <- round(cor(pay [, id\_city[i]], birth[, id\_city[i]]), 4)

text(0.75, i-0.5, labels = cor\_city[i])

text(0.25, i-0.5, labels = id\_city\_name [i])

}

text(0.75, 5.5, labels = paste("cor btw ", "birth and pay", sep="\n"))

text(0.25, 5.5, labels = "city")

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("cor btw birth and pay", paste("for the years", "that birth and pay are depandent",sep="\n"), sep="\n"), ylim=c(0, 7), xlim=c(0, 1))

abline(h=c(1:6))

abline(v=0.5)

for(i in 1:6){

cor\_year[i] <- round(cor(as.numeric(pay[i+12, -1]), as.numeric(birth[i+12, -1])), 4)

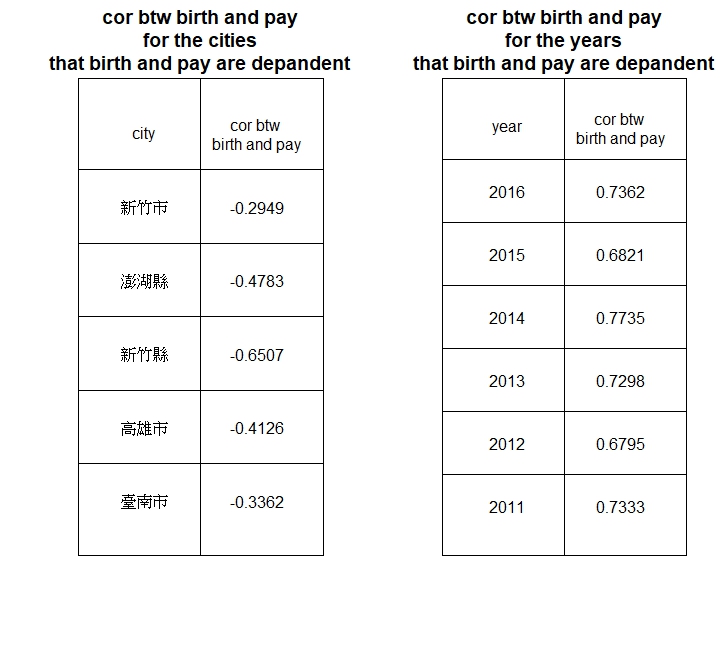
text(0.75, i-0.5, labels = cor\_year[i])

text(0.25, i-0.5, labels = i+2010)

}

text(0.75, 6.5, labels = paste("cor btw ", "birth and pay", sep="\n"))

text(0.25, 6.5, labels = "year")



windows()

plot(1, xaxt="n", yaxt="n" ,type="n", main = paste("2016 table for agreement","p-value for Mcnemar test is 0.617", sep = "\n") , ylim=c(0, 1), xlim=c(0, 1), frame = T, ylab="payment isn't higher than country v.s. payment is higher than country", xlab="birth isn't higher than country v.s. birth is higher than country")

abline(v=0.5)

abline(h=0.5)

for (i in 1:18){

pay\_compare <- as.vector(ifelse(pay[i, -(1:2)]>pay[i, 2], 1, 0))

birth\_compare <- as.vector(ifelse(birth[i, -(1:2)]>birth[i, 2], 1, 0))

mar\_p[i] <- mcnemar.test(pay\_compare, birth\_compare )$p.value

}

all\_agr <- 0

all\_neg <- 0

p\_agr\_b\_neg <- 0

p\_neg\_b\_agr <- 0

for (i in 1:20){

if (pay\_compare[i]==1&birth\_compare[i]==1) (all\_agr <- all\_agr+1)else (all\_agr <- all\_agr)

if (pay\_compare[i]==0&birth\_compare[i]==0) (all\_neg <- all\_neg+1)else (all\_neg <- all\_neg)

if (pay\_compare[i]==1&birth\_compare[i]==0) (p\_agr\_b\_neg <- p\_agr\_b\_neg +1)else (p\_agr\_b\_neg <- p\_agr\_b\_neg )

if (pay\_compare[i]==0&birth\_compare[i]==1) (p\_neg\_b\_agr <- p\_neg\_b\_agr+1)else (p\_neg\_b\_agr <- p\_neg\_b\_agr)

}

text(0.25, 0.25, labels = all\_neg)

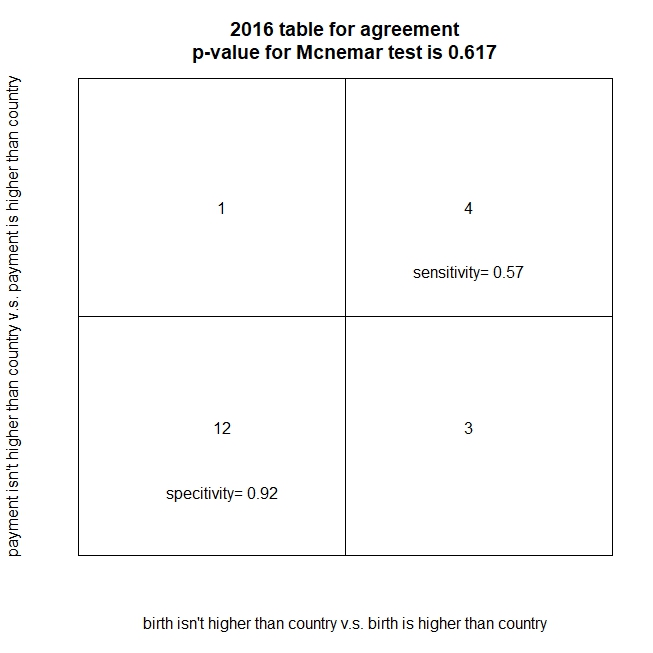
text(0.75, 0.25, labels = p\_neg\_b\_agr)

text(0.25, 0.75, labels = p\_agr\_b\_neg )

text(0.75, 0.75, labels = all\_agr)

text(0.25, 0.1, labels=paste("specitivity=", round(all\_neg/(all\_neg+p\_agr\_b\_neg ), 2)))

text(0.75, 0.6, labels=paste("sensitivity=", round(all\_agr/(all\_agr+p\_neg\_b\_agr ), 2)))



windows()

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = "statistics about Mcnemar test", ylim=c(0, 19), xlim=c(0, 7))

abline(v=1:6)

abline(h=1:18)

for (i in 1:18){

all\_agr <- 0

all\_neg <- 0

p\_agr\_b\_neg <- 0

p\_neg\_b\_agr <- 0

pay\_compare <- as.vector(ifelse(pay[i, -(1:2)]>pay[i, 2], 1, 0))

birth\_compare <- as.vector(ifelse(birth[i, -(1:2)]>birth[i, 2], 1, 0))

for (j in 1:20){

if (pay\_compare[j]==1&birth\_compare[j]==1) (all\_agr <- all\_agr+1)else (all\_agr <- all\_agr)

if (pay\_compare[j]==0&birth\_compare[j]==0) (all\_neg <- all\_neg+1)else (all\_neg <- all\_neg)

if (pay\_compare[j]==1&birth\_compare[j]==0) (p\_agr\_b\_neg <- p\_agr\_b\_neg +1)else (p\_agr\_b\_neg <- p\_agr\_b\_neg )

if (pay\_compare[j]==0&birth\_compare[j]==1) (p\_neg\_b\_agr <- p\_neg\_b\_agr+1)else (p\_neg\_b\_agr <- p\_neg\_b\_agr)

}

mar\_p[i] <- round(mcnemar.test(pay\_compare, birth\_compare )$p.value, 4)

mar\_m[i] <- round(mcnemar.test(pay\_compare, birth\_compare )$statistic, 4)

col\_mar[i] <- if (mar\_p[i] <0.05) ("red") else ("black")

text(0.5, i-0.5, labels=as.vector(pay[, 1])[i], col = col\_mar)

text(1.5, i-0.5, labels=(all\_agr+all\_neg), col=col\_mar)

text(2.5, i-0.5, labels=(p\_agr\_b\_neg+p\_neg\_b\_agr), col=col\_mar)

text(3.5, i-0.5, labels=round(all\_agr/(all\_agr+p\_neg\_b\_agr), 4), col=col\_mar)

text(4.5, i-0.5, labels=round(all\_neg/(all\_neg+p\_agr\_b\_neg), 4), col=col\_mar)

text(5.5, i-0.5, labels=mar\_m[i], col=col\_mar)

text(6.5, i-0.5, labels=mar\_p[i], col=col\_mar)

}

text(0.5, 18.75, labels="year")

text(1.5, 18.75, labels="agreement", cex=0.8)

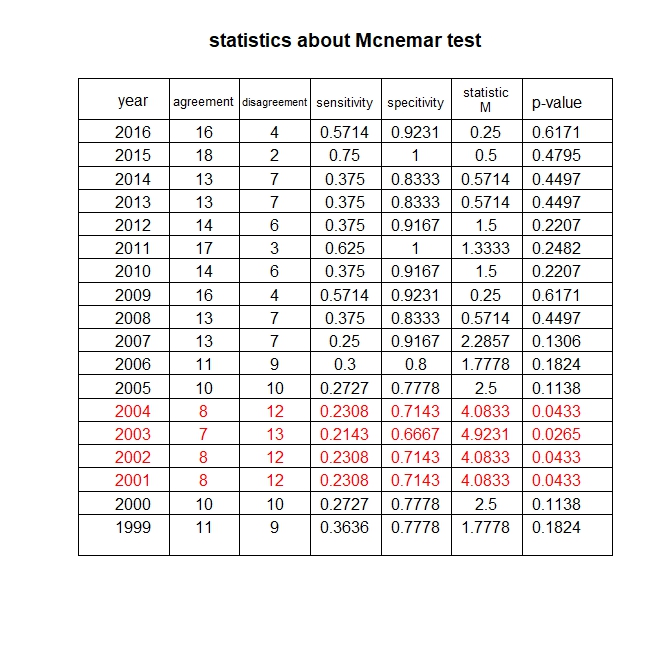
text(2.5, 18.75, labels="disagreement", cex=0.7)

text(3.5, 18.75, labels="sensitivity", cex=0.8)

text(4.5, 18.75, labels="specitivity", cex=0.8)

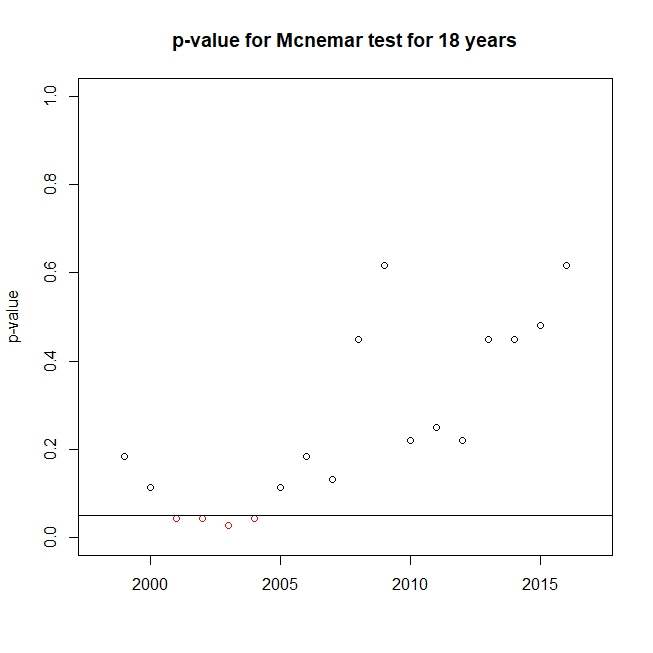
text(5.5, 18.9, labels=paste("statistic", "M", sep="\n"), cex=0.8)

text(6.5, 18.75, labels="p-value")



plot(x=as.numeric(pay[, 1]), y=mar\_p, xlab="", ylab="p-value", main = "p-value for Mcnemar test for 18 years", ylim=c(0, 1), xlim=c(1998, 2017), col=col\_mar)

abline(h=0.05)



#TEST3

birth\_n <- birth[, c(3:6, 9:11, 13, 20:21)]

birth\_s<- birth[, c(7:8, 12, 14:16, 19, 22)]

windows()

for (i in 1:18){

mann\_birth\_p[i] <- wilcox.test(as.numeric(birth\_n[i, ]), as.numeric(birth\_s[i, ]))$p.value

mann\_birth\_w[i] <- wilcox.test(as.numeric(birth\_n[i, ]), as.numeric(birth\_s[i, ]))$statistic

}

mann\_birth\_p <- unlist(mann\_birth\_p)

col\_m <- ifelse(mann\_birth\_p<0.05, "red", "black")

plot(y=mann\_birth\_p, x=as.vector(pay[, 1]), xlab="year", ylab="p-value", main = "p-value for birth location test", ylim=c(0, 1), xlim=c(1998, 2017), col=col\_m)

abline(h=0.05)

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("difference birth rate", paste("btw northern and southern Taiwan", "H0:birth(year,north)=birth(year,south)",sep="\n"), sep="\n"), ylim=c(0, 19), xlim=c(0, 3))

abline(v=1:2)

abline(h=1:18)

for (i in 1:18){

text(0.5, i-0.5, labels=as.vector(pay[, 1])[i], col = col\_m[i])

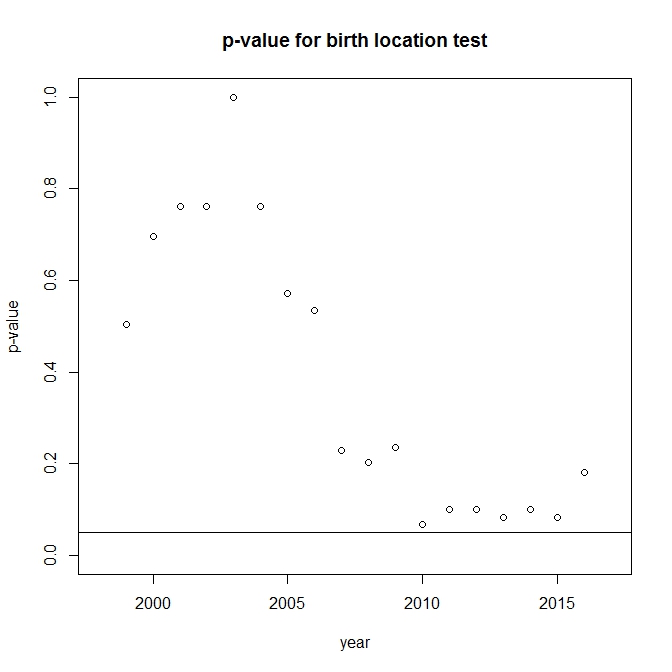
text(1.5, i-0.5, labels=round(mann\_birth\_w[i], 4), col=col\_m[i])

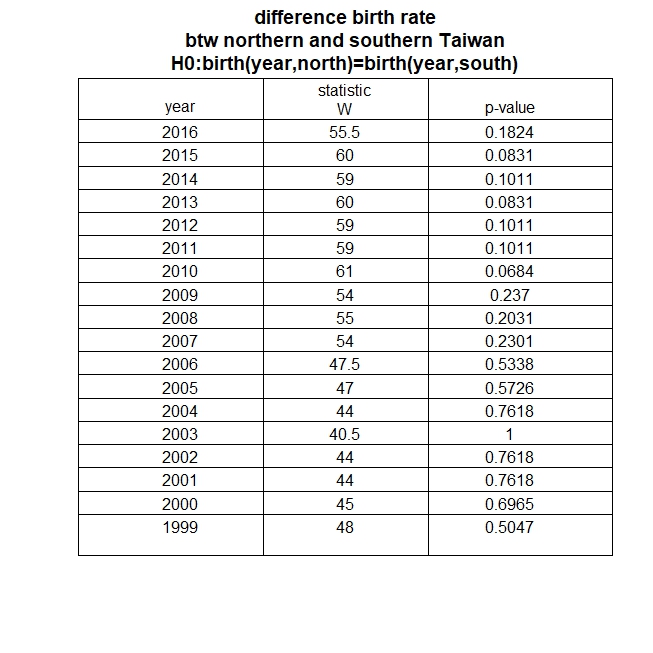
text(2.5, i-0.5, labels=round(mann\_birth\_p[i], 4), col=col\_m[i])

}

text(0.5, 18.5, labels="year")

text(1.5, 18.9, labels=paste("statistic" ,"W", sep="\n"))

text(2.5, 18.5, labels="p-value")



#test4

windows()

for (i in 1:18){

wil\_pair\_p[i] <-ifelse(i < 18, wilcox.test(as.numeric(birth[i, ]), as.numeric(birth[18, ]), paired = T)$p.value, "null")

wil\_pair\_w[i] <-ifelse(i < 18, wilcox.test(as.numeric(birth[i, ]), as.numeric(birth[18, ]), paired = T)$statistic, "null")

}

wil\_pair\_p <- round(as.numeric(wil\_pair\_p ), 4)

wil\_pair\_w <- round(as.numeric(wil\_pair\_w ), 4)

col\_w <- ifelse(wil\_pair\_p<0.05, "red", "black")

plot(y=wil\_pair\_p , x=as.vector(pay[, 1]), xlab="year", ylab="p-value", main = "p-value for birth rate compare to this year test", ylim=c(0, 1), xlim=c(1998, 2017), col=col\_w)

abline(h=0.05)

plot(1, type="n", xaxt="n", yaxt="n", xlab="", ylab="", main = paste("difference birth rate", paste("btw 2016 and other years", "H0:birth(2016)=birth(other year)",sep="\n"), sep="\n"), ylim=c(0, 18), xlim=c(0, 3))

abline(v=1:2)

abline(h=1:17)

for (i in 1:17){

text(0.5, i-0.5, labels=as.vector(pay[, 1])[i], col = col\_w[i])

text(1.5, i-0.5, labels=round(wil\_pair\_w[i], 4), col=col\_w[i])

text(2.5, i-0.5, labels=round(wil\_pair\_p[i], 4), col=col\_w[i])

}

text(0.5, 17.5, labels="year")

text(1.5, 17.9, labels=paste("statistic" ,"W", sep="\n"))

text(2.5, 17.5, labels="p-value")