

# Analysis of Economic Mobility

Bonald So

Western: Longitude < -109

Eastern: Longitude > -85

From the summary statistics below, we can see that Eastern Mobility has a wider range than Western Mobility. But they have similar width of IQR. And from the histograms, we can also find that they are both right skewed.

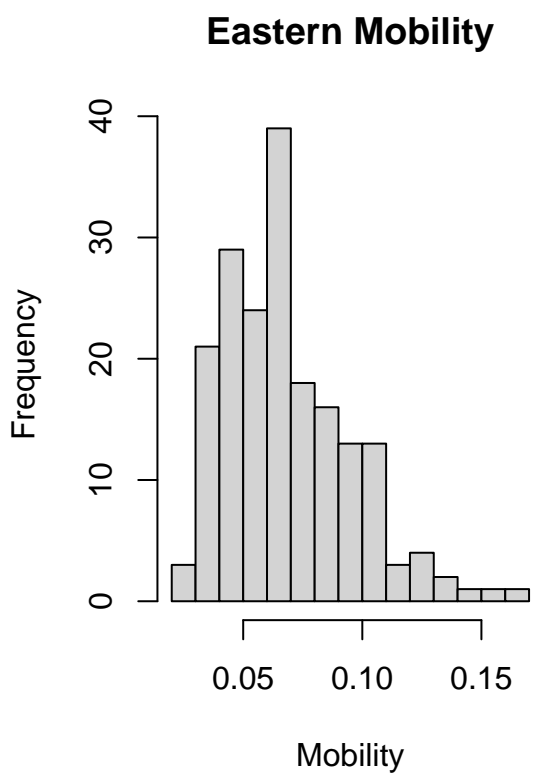
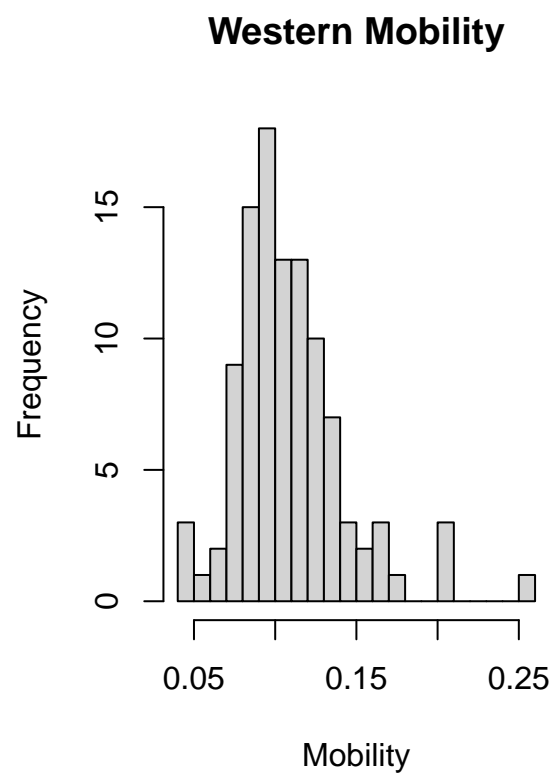
```
data <- read.csv("EconomicMobility.csv", header = TRUE)
western=subset(data, data$Longitude < -109)
eastern=subset(data, data$Longitude > -85)
summary(western$Mobility)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04106 0.08680 0.10484 0.10882 0.12410 0.25338
```

```
summary(eastern$Mobility)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02295 0.04908 0.06443 0.06829 0.08223 0.16298
```

```
par(mfrow=c(1, 2))
wm<-western$Mobility
em<-eastern$Mobility
hist(western$Mobility, breaks=20,
main = "Western Mobility", xlab="Mobility",
col="lightgrey")
hist(eastern$Mobility, breaks=20,
main = "Eastern Mobility", xlab="Mobility",
col="lightgrey")
```



Below are the values of standard deviation(SD), coefficient of variation(CV) and skewness(SK) of western (first row) and eastern (second row) mobility respectively.

```
Attributes<-function(y) {  
  N<-length(y)  
  ybar<-mean(y)  
  sd<-sqrt(sum((y - ybar)^2) / N)  
  cv<-sd/ybar  
  sk<-sum((y - ybar)^3) / N / sd^3  
  return (c(sd, cv, sk))  
}  
wmatt<-Attributes(wm)  
ematt<-Attributes(em)  
wmatt
```

```
## [1] 0.03423649 0.31460789 1.26503176
```

```
ematt
```

```
## [1] 0.02641876 0.38686741 0.91628878
```

```

mixRandomly <- function(pop) {
  pop1 <- pop$pop1
  n_pop1 <- nrow(pop1)

  pop2 <- pop$pop2
  n_pop2 <- nrow(pop2)

  mix <- rbind(pop1, pop2)
  select4pop1 <- sample(1:(n_pop1 + n_pop2), n_pop1, replace = FALSE)

  new_pop1 <- mix[select4pop1, ]
  new_pop2 <- mix[-select4pop1, ]
  list(pop1 = new_pop1, pop2 = new_pop2)
}

getSD <- function(variate) {
  function(pop) {
    p1 = Attributes(pop$pop1[, variate])
    p2 = Attributes(pop$pop2[, variate])
    return (p1[1]/p2[1])
  }
}

getCV <- function(variate) {
  function(pop) {
    p1 = Attributes(pop$pop1[, variate])
    p2 = Attributes(pop$pop2[, variate])
    return (p1[2]/p2[2])
  }
}

getSK <- function(variate) {
  function(pop) {
    p1 = Attributes(pop$pop1[, variate])
    p2 = Attributes(pop$pop2[, variate])
    return (p1[3]/p2[3])
  }
}

ratioSD<-getSD("Mobility")
ratioCV<-getCV("Mobility")
ratioSK<-getSK("Mobility")
pop<-list(pop1=eastern, pop2=western)
sdratio <- sapply(1:1000, FUN = function(...) {
  ratioSD(mixRandomly(pop))
})
cvratio <- sapply(1:1000, FUN = function(...) {
  ratioCV(mixRandomly(pop))
})
skratio <- sapply(1:1000, FUN = function(...) {
  ratioSK(mixRandomly(pop))
})
sdpval<-mean(abs(sdratio - 1) >= abs(wmatt[1]/ematt[1] - 1))

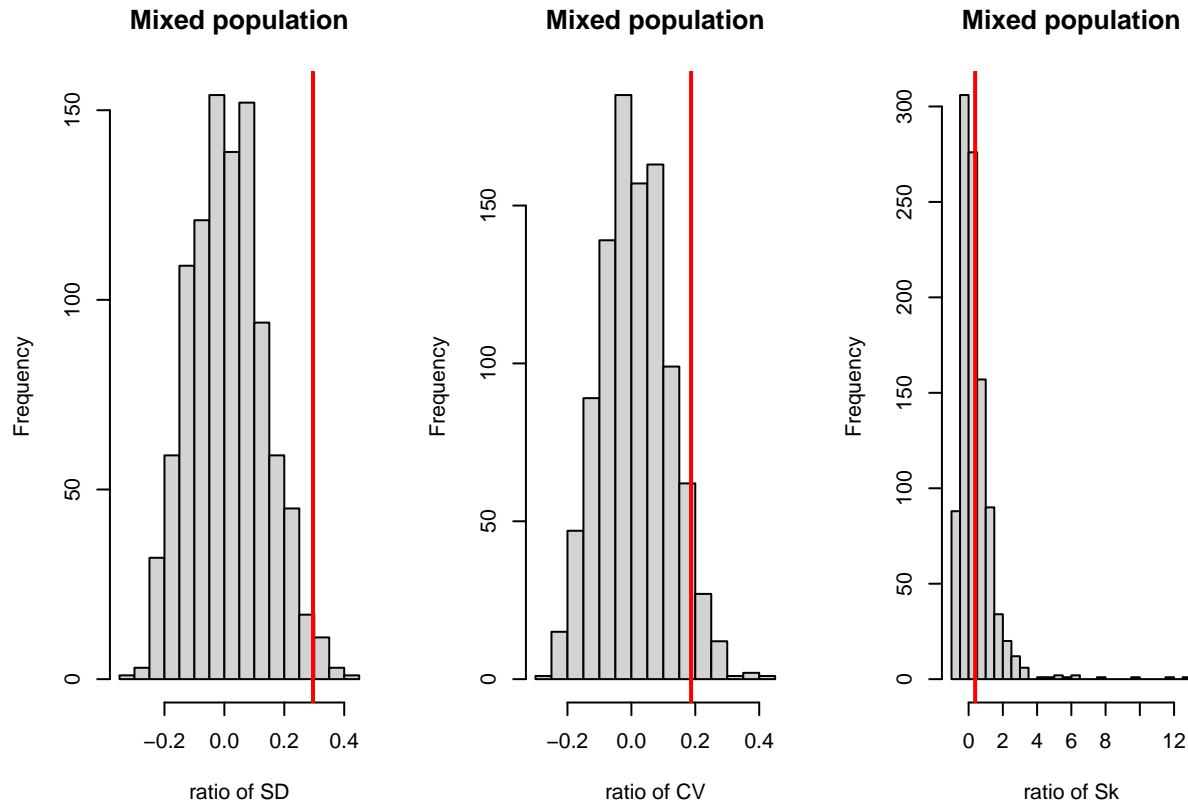
```

```

cvpval<-mean(abs(cvratio - 1) >= abs(wmatt[2]/ematt[2] - 1))
skpval<-mean(abs(skratio - 1) >= abs(wmatt[3]/ematt[3] - 1))

par(mfrow=c(1,3))
hist(sdratio - 1, breaks=20,
main = "Mixed population", xlab="ratio of SD",
col="lightgrey")
abline(v = abs(wmatt[1]/ematt[1] - 1), col = "red", lwd = 2)
hist(cvratio - 1, breaks=20,
main = "Mixed population", xlab="ratio of CV",
col="lightgrey")
abline(v = abs(wmatt[2]/ematt[2] - 1), col = "red", lwd = 2)
hist(skratio - 1, breaks=20,
main = "Mixed population", xlab="ratio of Sk",
col="lightgrey")
abline(v = abs(wmatt[3]/ematt[3] - 1), col = "red", lwd = 2)

```



```

table <- data.frame("pvalueOfSD" = sdpval,
                    "pvalueOfCV" = cvpval,
                    "pvalueOfSK" = skpval)

library(knitr)
kable(table)

```

pvalueOfSD	pvalueOfCV	pvalueOfSK
0.017	0.085	0.6

Analysis: From the p-value table above, we can see that only p-value of SD is less than 0.05, so it shows strong evidence that Western mobility and Eastern mobility are not the same based on SD of mobility. And p-values of CV and SK are greater than 0.05, so there is no evidence showing against the hypothesis that CV and SK from Western mobility and Eastern mobility are the same based on CV and SK of mobility respectively.

```

calculateSLmulti <- function(pop, discrepancies, M_outer = 1000, M_inner){
  #pop is a list whose two members are two sub-populations

  if (missing(M_inner)) M_inner <- M_outer
  ## Local function to calculate the significance levels
  ## over the discrepancies and return their minimum

  getSLmin <- function(basePop, discrepancies, M) {
    observedVals <- sapply(discrepancies,
                          FUN = function(discrepancy) {discrepancy(basePop)})

    K <- length(discrepancies)

    total <- Reduce(function(counts, i){
      #mixRandomly mixes the two populations randomly, so the new sub-populations are indistinguishable
      NewPop <- mixRandomly(basePop)

      ## calculate the discrepancy and counts
      Map(function(k) {
        Dk <- discrepancies[[k]](NewPop)
        if (Dk >= observedVals[k]) counts[k] <- counts[k] + 1 },
          1:K)
      counts
    },
    1:M, init = numeric(length=K))

    SLs <- total/M
    min(SLs)
  }

  SLmin <- getSLmin(pop, discrepancies, M_inner)

  total <- Reduce(function(count, m){
    basePop <- mixRandomly(pop)
    if (getSLmin(basePop, discrepancies, M_inner) <= SLmin) count + 1 else count
  }, 1:M_outer, init = 0)

  SLstar <- total/M_outer
  SLstar
}

discrepancies <- list(getSD("Mobility"), getCV("Mobility"), getSK("Mobility"))
SLstar=calculateSLmulti(pop, discrepancies, M_outer = 100, M_inner=100)
SLstar

```

```
## [1] 0.1
```

Interpretation of result: since p-value is greater than 0.05, it shows no evidence against the hypothesis that Western mobility and Eastern mobility were different based on SD, CV and Sk of mobility.

The conclusion is consistent with conclusion from the beginning. That is because it takes different discrepancy measures into account and test the hypothesis, which would be more unbiased and thus producing consistent result from what we observed.