Project1

1/15/2017

```r
setwd("~/Documents/Biostatistics/Collaboration II")
#setwd("~/Documents/BIOS7352/Project1")

totalData <- readRDS("dataForABP_CBF_2017-01-11.rds")
library(Hmisc)
library(knitr)

crvdata <- totalData[totalData$asl.reac.usable == 1 & totalData$enrolled.dx.factor != "Dementia",
                c(1:18, 61, grep("asl.reac", names(totalData)),
                  grep("ma.", names(totalData)))]
names(crvdata) <- sub("asl.reac.", "", names(crvdata), fixed = TRUE)
names(crvdata) <- sub(".hct", "", names(crvdata), fixed = TRUE)
names(crvdata) <- sub("ma.", "", names(crvdata), fixed = TRUE)

exclude <- totalData[totalData$asl.reac.usable == 0 | totalData$enrolled.dx.factor == "Dementia",
                c(1:18, 61, grep("asl.reac", names(totalData)),
                  grep("ma.", names(totalData)))]
names(exclude) <- sub("asl.reac.", "", names(exclude), fixed = TRUE)
names(exclude) <- sub(".hct", "", names(exclude), fixed = TRUE)
names(exclude) <- sub("ma.", "", names(exclude), fixed = TRUE)


#totalData$asl.reac.usable == 1,


cats <- names(crvdata)[2:19]
comparison <- c(c(), c(), c(),c())
is <- length(crvdata$map.id)
xs <- length(exclude$map.id)
for (cat in cats){
  if (is.factor(crvdata[,cat])){
    chiData <- rbind(cbind(crvdata[,cat],rep("is", length(crvdata[,cat]))),
                  cbind(exclude[,cat],rep("xs", length(exclude[,cat]))))
    pp <- chisq.test(table(chiData[,1], chiData[,2]))$p.value
    comparison <- rbind(comparison, c(label(crvdata[,cat]),'','', round(pp,4)))
    for (lev in levels(crvdata[,cat])){
      comparison <- rbind(comparison, c(paste("--",lev ),
                  paste(s <- sum(crvdata[,cat]==lev, na.rm=T), " (", round(s*100/is),
                        "%)", sep=""),
                  paste(s <- sum(exclude[,cat]==lev, na.rm=T), " (", round(s*100/xs),
                        "%)", sep=""),""))
    }
    next
  }
  anovaData <- as.data.frame(rbind(cbind(crvdata[,cat],rep("is", length(crvdata[,cat]))),
                  cbind(exclude[,cat],rep("xs", length(exclude[,cat])))))
  anovaData[,1] <- as.numeric(as.character(anovaData[,1]))
  pp <- kruskal.test(anovaData[,1] ~ anovaData[,2])$p.value
  comparison <- rbind(c(label(crvdata[,cat]),
```

```
                paste(round(mean(crvdata[,cat], na.rm=T),1)," (",
                      round(sd(crvdata[,cat],na.rm=T),1), ")",sep=""),
                paste(round(mean(exclude[,cat], na.rm=T),1)," (",
                      round(sd(exclude[,cat],na.rm=T),1), ")",sep=""),
                round(pp,4)), comparison)
}
comparison <- as.data.frame(comparison[,c(1,3,2,4)])
colnames(comparison) <- c("Variable","Excluded", "Analyzed Data", "P-Value")
kable(comparison, width=3,
      caption=paste("Comparison of Demographics for Excluded & Included Data, w/ N=",
                    length(exclude$map.id), " and N=",length(crvdata$map.id),
                                    " respectively", sep=""))
```

Table 1: Comparison of Demographics for Excluded & Included Data, w/ N=113 and N=223 respectively

Variable	Excluded	Analyzed Data	P-Value
Diff. in mean SBP, wake - sleep, self-reported periods	14.5 (10.5)	13.4 (9.4)	0.6075
systolic.post.wake.mean minus systolic.pre.wake.mean	11.1 (12.3)	12.3 (12.2)	0.4331
systolic.post.wake.1 minus systolic.pre.wake.1	8.6 (14.2)	8.4 (13.6)	0.8162
ICV (calculated)	1431.2 (146.7)	1360.3 (134.6)	1e-04
Education (years)	16.5 (2.5)	15.5 (2.7)	9e-04
Age at medhx.date, recalculated	73.1 (7.1)	72.8 (7.4)	0.5341
Consensus Decision for Diagnosis			0.3736
– Normal	55 (49%)	121 (54%)	
– MCI	46 (41%)	86 (39%)	
– Dementia	1 (1%)	0 (0%)	
– Ambiguous At Risk	11 (10%)	16 (7%)	
Sex			0
– Male	90 (80%)	109 (49%)	
– Female	23 (20%)	114 (51%)	
Two-level race/ethnicity			0.9012
– Non-Hispanic White	97 (86%)	194 (87%)	
– Other	16 (14%)	29 (13%)	
ApoE4+ (at least one E4 allele)			0.7178
– Yes	41 (36%)	75 (34%)	
– No	72 (64%)	148 (66%)	
Consensus Decision for Diagnosis			0.3736
– Normal	55 (49%)	121 (54%)	
– MCI	46 (41%)	86 (39%)	
– Dementia	1 (1%)	0 (0%)	
– Ambiguous At Risk	11 (10%)	16 (7%)	
Taking at least 1 anti-hypertensive med			0.9461
– Yes	62 (55%)	120 (54%)	
– No	51 (45%)	103 (46%)	
Diabetic, determined by a1c, glucose, and/or rx			0.4304
– Yes	24 (21%)	38 (17%)	
– No	89 (79%)	185 (83%)	
Current smoker (or quit in this or last calendar yr)			0.3542
– Yes	4 (4%)	3 (1%)	
– No	109 (96%)	220 (99%)	
CVD, determined from variables in med hx			0.897
– Yes	3 (3%)	8 (4%)	

Variable	Excluded	Analyzed Data	P-Value
– No	110 (97%)	215 (96%)	
A-fib, determined by med hx and/or echo and/or cmr rhythm			1
– Yes	6 (5%)	13 (6%)	
– No	106 (94%)	209 (94%)	
LV hypertrophy, determined by sex and scaled LV mass			1
– Yes	5 (4%)	11 (5%)	
– No	108 (96%)	211 (95%)	
At least 39 readings AND max gap btw readings is <3h			0.0525
– No	6 (5%)	30 (13%)	
– Yes	91 (81%)	174 (78%)	

Missingness

```r
nrow(totalData)
```

```
## [1] 336
```

```r
#Start with 336
nrow(crvdata)
```

```
## [1] 223
```

```r
#223 left after exclusions

missing <- c(c(), c())
#comparison[,c("Variable", "Analyzed Data")]
for (cat in cats){
    missing <- rbind(missing, c(label(crvdata[,cat]), paste(s <- sum(is.na(crvdata[,cat])), " (", round
 }

missing <- as.data.frame(missing)
colnames(missing) <- c("Variable", "Missingness")

kable(missing, width=3,
      caption=paste("Missingness (N=", nrow(crvdata), ")", sep=""))
```

Table 2: Missingness (N=223)

Variable	Missingness
Consensus Decision for Diagnosis	0 (0%)
Sex	0 (0%)
Two-level race/ethnicity	0 (0%)
ApoE4+ (at least one E4 allele)	0 (0%)
Consensus Decision for Diagnosis	0 (0%)
Taking at least 1 anti-hypertensive med	0 (0%)
Diabetic, determined by a1c, glucose, and/or rx	0 (0%)
Current smoker (or quit in this or last calendar yr)	0 (0%)
CVD, determined from variables in med hx	0 (0%)
A-fib, determined by med hx and/or echo and/or cmr rhythm	1 (0.45%)

Variable	Missingness
LV hypertrophy, determined by sex and scaled LV mass	1 (0.45%)
Age at medhx.date, recalculated	0 (0%)
Education (years)	0 (0%)
ICV (calculated)	0 (0%)
systolic.post.wake.1 minus systolic.pre.wake.1	72 (32.29%)
systolic.post.wake.mean minus systolic.pre.wake.mean	76 (34.08%)
Diff. in mean SBP, wake - sleep, self-reported periods	64 (28.7%)
At least 39 readings AND max gap btw readings is <3h	19 (8.52%)