Exam Practic 2

Nick Thompson

10/13/2016

Pull in Data

```
setwd("~/Documents/GitHubRepo/729_Reed_MLE_git/Exam/CCES")
#data <- read.csv(file = "https://raw.githubusercontent.com/Neilblund/729A/master/data/voterid.csv", he
#or
# data <- read.dta(file = "CCES.RData")
#or
# data <- read.dta13(file = "CCES.RData")
# save(data, file = "data.RData")
load("CCES.RData")
data <- x
#View(data)
```

Introduction:

```
#descriptive statistics for all variables
#stargazer(voterid, type = 'text')
# run probit, show results
data <- na.omit(data)
#View(data) - here we put the point prediction that we're looking for.
# If you're using mean, then keep mean.
# If you're using a SD up and down from the mean, then use that.
# If you're using a different range, then use that.
data$var1 = data$age
data$var2 = data$income
data$var3 = as.factor(data$race)
data$var4 = data$divergence
data$var5 = data$educ
data$var6 = data$female
data$var7 = data$competitive
data$var8 = as.factor(data$st_id)
data$var9 <- data$voted
data$mean_var1 <- mean(data$var1, na.rm=T)
data$sd_var1 <- 0.5*sd(data$var1)
data$med_var1 = median(data$var1)
```

Descriptive statistics

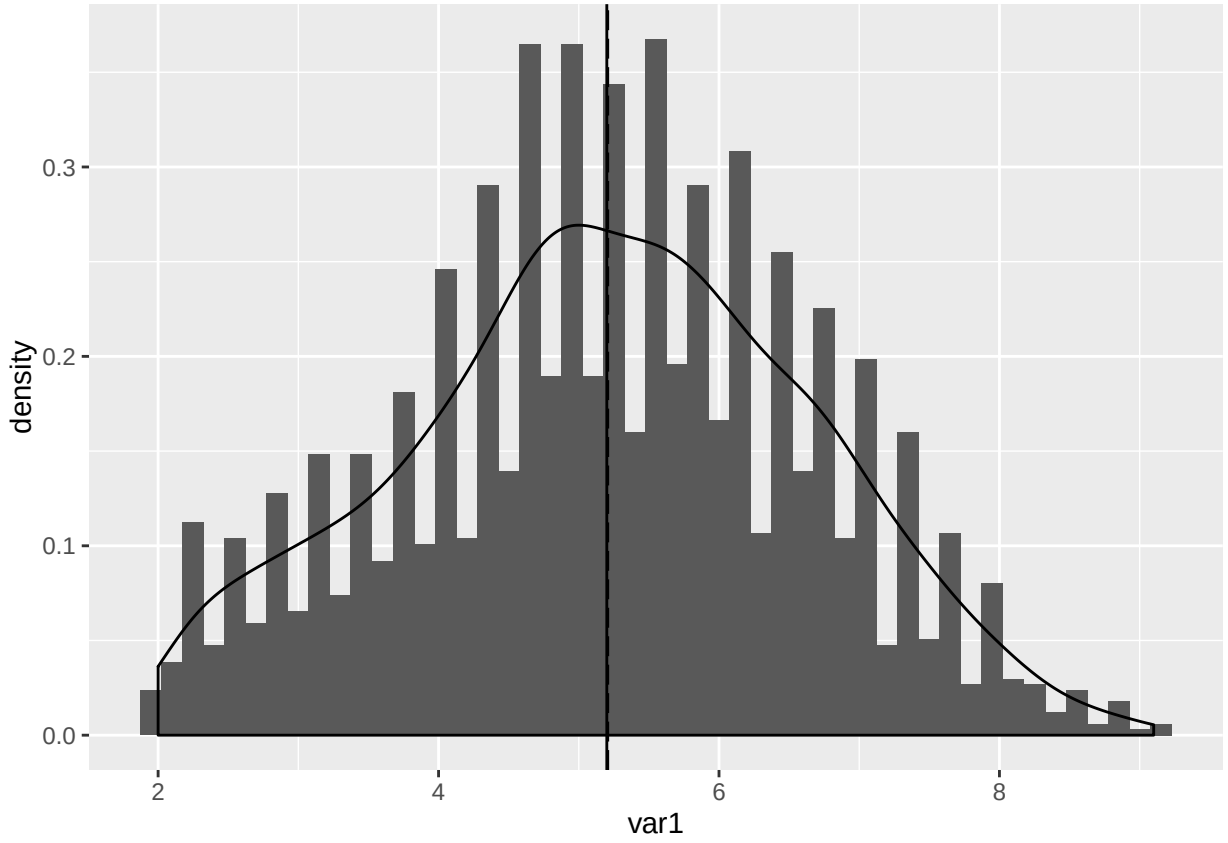
Plot a histogram to see what the data looks like. Identify skewness for determining if using mean, median, or tail.

```
g <- ggplot(data,aes(x=var1))
# adjust the binwidth to the desired width so the histogram tells you something.
# the geom_vlines give you lines to mark the mean and median.
```

```
# Consider using the median if the skewness > 0.5 or skewness < -0.5, but use your
# own discretion.  If the -0.5 < skewness < 0.5 then use the mean.
skewness(data$var1)
```

```
## [1] -0.04902077
```

```
# kurtosis(data$var1) # if you care about kurtosis
# anova(model_1p) # if for some reason you want ANOVA
g + geom_histogram(aes(y=..density..),binwidth = .15) +
  geom_density() +
  geom_vline(xintercept = data$mean_var1,linetype='longdash') +
  geom_vline(xintercept = data$med_var1)
```




```
data$var1_obs_low <- data$mean_var1 - data$sd_var1
data$var1_obs_high <- data$mean_var1 + data$sd_var1
```

Calculate the average effect of `variable name` using observed values.

The logit model.

```
# run logit, show results
model_1l <-glm(voted~age+income+
                 as.factor(race)+divergence+educ+
                 female+competitive+as.factor(st_id),
               family="binomial",data=data)
```

A table of the logit model.

```
#summary(model_1l)
stargazer(model_1l,header=F) # if you want to see a print out in your console,
```

```
                              # then after model_1l typt this
# ,type='text',
```

The logit predicted probabilities.

Calculate the average effects with a logit model.

A way to do this with probit.

A table printout of the probit model.

Calculate the predicted probabilities in the probit model.

A summary of the probit test.

Calculate the average effects and difference for a probit model.

Summary: Interpreting the Coefficients, include the AIC

Log Likelihood

Deviances

Bayes

AIC

Simulations:

Violin Plots of the Simulation

Table 1:

	Dependent variable:
	voted
age	0.258^{***}
	(0.046)
income	0.070^{***}
	(0.022)
as.factor(race)2	-0.368
	(0.224)
as.factor(race)3	-0.577^{*}
	(0.296)
as.factor(race)4	-0.199
	(0.334)
divergence	-0.775^{**}
	(0.325)
educ	0.084
	(0.058)
female	-0.231^{*}
	(0.135)
competitive	0.096^{**}
	(0.048)
as.factor(st_id)2	-0.695
	(0.844)
as.factor(st_id)4	-7.123^{***}
	(1.147)
as.factor(st_id)5	-0.398
	(0.588)
as.factor(st_id)11	-1.312
	(0.821)
as.factor(st_id)14	-1.204^{*}
	(0.625)
as.factor(st_id)15	-0.635
	(0.583)
as.factor(st_id)16	-0.834
	(0.647)
as.factor(st_id)21	-2.479^{***}
	(0.611)
as.factor(st_id)22	-5.059^{***}
	(0.686)
as.factor(st_id)23	-0.627