Exam 1

Shay Culpepper

10/23/2018

Hypothesis tests

KNN

```r
rm(list = ls(all = TRUE))
setwd("CEXData")
load("cex_2012.RData")
attach(data_cex)

clean_df <- data_cex %>%
  mutate(ent_prop = ENTERTPQ / TOTEXPPQ)

# Calculate 25% quantile
ent_prop_25 <- quantile(clean_df$ent_prop, 0.25, na.rm = TRUE)

# Create tibble to work with
knn_data <- clean_df %>%
  mutate(bottom25 = ent_prop < ent_prop_25) %>%
  select(bottom25,
         BLS_URBN,
         female,
         AGE_REF
       )
cl.column <- knn_data$bottom25

good.obs <- complete.cases(knn_data)
NN_obs <- sum(good.obs)

# select only rows with complete cases
y.data  <- subset(knn_data, good.obs)

## Pick random rows for training data
set.seed(17)

train.obs <- (runif(NN_obs) < 0.8)

train.data <- subset(y.data, train.obs)
test.data  <- subset(y.data, !train.obs)
cl.data    <- cl.column[train.obs]
true.data  <- cl.column[!train.obs]

predicted.ent <- knn(train = train.data[-1],
                     test = test.data[-1],
                     cl = cl.data,
```

```
                    k = 3)
n.correctly.predicted <- sum(predicted.ent == true.data)
correct.rate <- n.correctly.predicted / length(predicted.ent)
print(correct.rate)
```

```
## [1] 0.7551622
```

```
detach(data_cex)
```

OLS

```
rm(list = ls(all = TRUE))
setwd("CEXData")
load("cex_2012.RData")
attach(data_cex)

clean_df <- data_cex %>%
  mutate(ent_prop = ENTERTPQ / TOTEXPPQ) %>%
  data.frame()

model1 <- lm(ent_prop ~
  BLS_URBN  +
  educ_nohs  +
  educ_hs    +
  educ_smcoll+
  educ_as    +
  educ_bach  +
  educ_adv   +
  female +
  ALCBEVPQ +
  AGE_REF
  , clean_df)

stargazer(model1, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Oct 23, 2018 - 11:58:25

```
detach(data_cex)
```

Table 1:

	Dependent variable:
	ent_prop
BLS_URBN2	0.004
	(0.003)
educ_nohs	0.0001
	(0.003)
educ_hs	0.009***
	(0.003)
educ_smcoll	0.014***
	(0.003)
educ_as	0.011***
	(0.003)
educ_bach	0.008***
	(0.003)
educ_adv	0.006*
	(0.003)
female	0.003**
	(0.001)
ALCBEVPQ	0.00001***
	(0.00000)
AGE_REF	−0.0001***
	(0.00004)
Constant	0.048***
	(0.003)
Observations	6,838
R^2	0.011
Adjusted R^2	0.010
Residual Std. Error	0.057 (df = 6827)
F Statistic	7.855*** (df = 10; 6827)
Note:	*p<0.1; **p<0.05; ***p<0.01