

# Pol Sci 630: Problem Set 2 - Solutions

Anh Le

September 9, 2016

**Insert your comments on the assignment that you are grading above the solution in bold and red text. For example write: "GRADER COMMENT: everything is correct! - 4/4 Points" Also briefly point out which, if any, problems were not solved correctly and what the mistake was.**

## 1 Expected Value and Its Properties

**a.**

(1/4 point) (DeGroot, p. 216) Suppose that one word is to be selected at random from the sentence ‘the girl put on her beautiful red hat’. If  $X$  denotes the number of letters in the word that is selected, what is the value of  $E(X)$ ?

**Solution**

As the number of letters in a word,  $X$  can take on following values:  $x \in \{2, 3, 4, 9\}$ , with probability as follows:

$$P(X = 2) = \frac{1}{8} \quad (1 \text{ word ("on") out of 8 words in the sentence}) \quad (1)$$

$$P(X = 3) = \frac{5}{8} \quad (2)$$

$$P(X = 4) = \frac{1}{8} \quad (3)$$

$$P(X = 9) = \frac{1}{8} \quad (4)$$

Therefore,

$$E(X) = \sum_{all x_i} x_i P(X = x_i) = 3.75$$

**b.**

(2/4 point) (Degroot p. 216) Suppose that one letter is to be selected at random from the 30 letters in the sentence given in Exercise 4. If Y denotes the number of letters in the word in which the selected letter appears, what is the value of E(Y)?

**Solution**

Y can take on values  $y \in \{2, 3, 4, 9\}$  with probability as follows:

$$P(Y = 2) = \frac{2}{30} \quad \text{O,N} \quad (5)$$

$$P(Y = 3) = \frac{15}{30} \quad \text{T,H,E, P,U,T, H,E,R, R,E,D, H,A,T} \quad (6)$$

$$P(Y = 4) = \frac{4}{30} \quad \text{G,I,R,L} \quad (7)$$

$$P(Y = 9) = \frac{9}{30} \quad \text{B,E,A,U,T,I,F,U,L} \quad (8)$$

Therefore,

$$E(Y) = \sum_{all y_i} y_i P(Y = y_i) = \frac{73}{15} = 4.867$$

## 2 Plotting distribution

For this problem, you'll need to Google some R techniques (e.g. side-by-side / overlapping plot). Also, label the axes and the plots accordingly.

**a.**

(1/4 point) Download a variable you are interested in, using WDI. Plot the histogram, density plot, boxplot, and normal quantile plot.

```
# install.packages("WDI")
library(WDI)
```

```

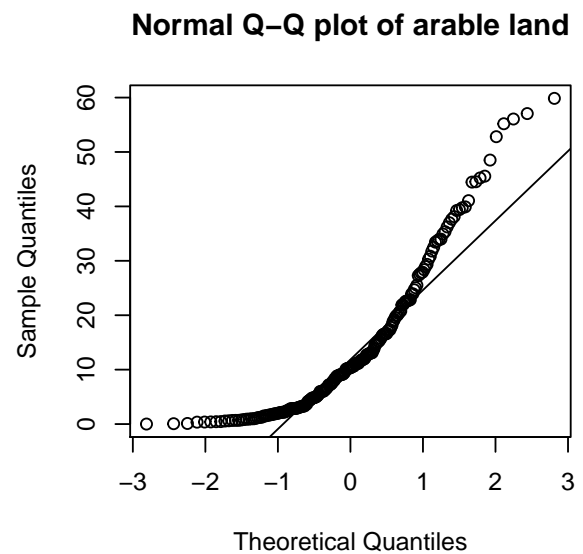
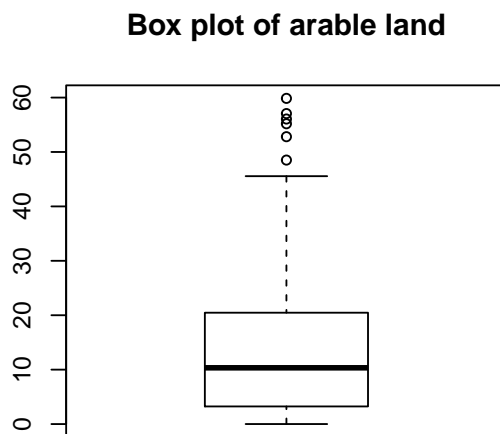
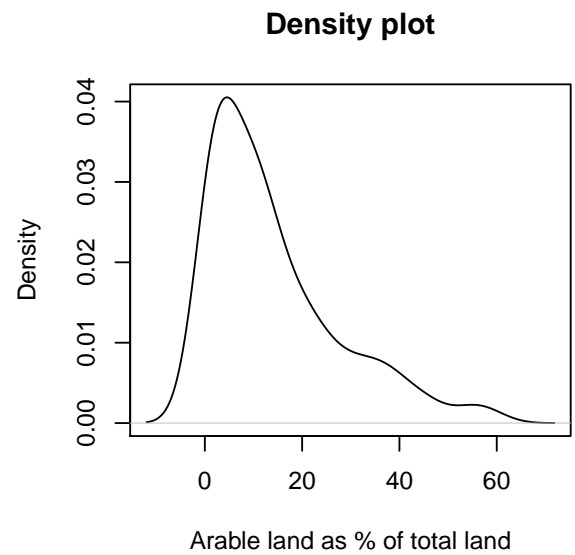
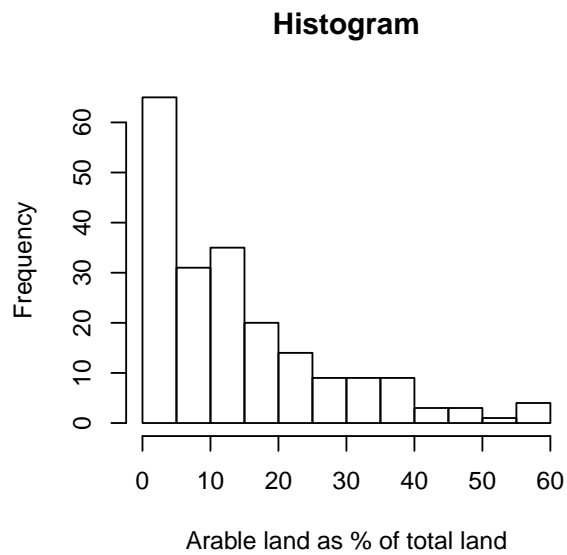
## Loading required package: RJSONIO

d_land <- WDI(indicator = c("AG.LND.ARBL.ZS", "NY.GDP.PCAP.KD"),
             start=2010, end=2010, extra=TRUE)
d_land <- d_land[d_land$region != "Aggregates", ]

# Rename column
colnames(d_land)[colnames(d_land) == "AG.LND.ARBL.ZS"] <- "arable_land_pct"
colnames(d_land)[colnames(d_land) == "NY.GDP.PCAP.KD"] <- "gdp_percapita"

xlabel <- "Arable land as % of total land"
par(mfrow=c(2, 2))
hist(d_land$arable_land_pct, main = "Histogram", xlab = xlabel)
plot(density(d_land$arable_land_pct, na.rm = TRUE), main = "Density plot", xlab = xlabel)
boxplot(d_land$arable_land_pct, main = "Box plot of arable land")
qqnorm(d_land$arable_land_pct, main = "Normal Q-Q plot of arable land")
qqline(d_land$arable_land_pct)

```



b.

(1/4 point) Plot the density plots of that variable for Europe and Asia, 1) side by side (Hint: `par(mfrow=c(?, ?))`), and 2) overlapping in the same plot.

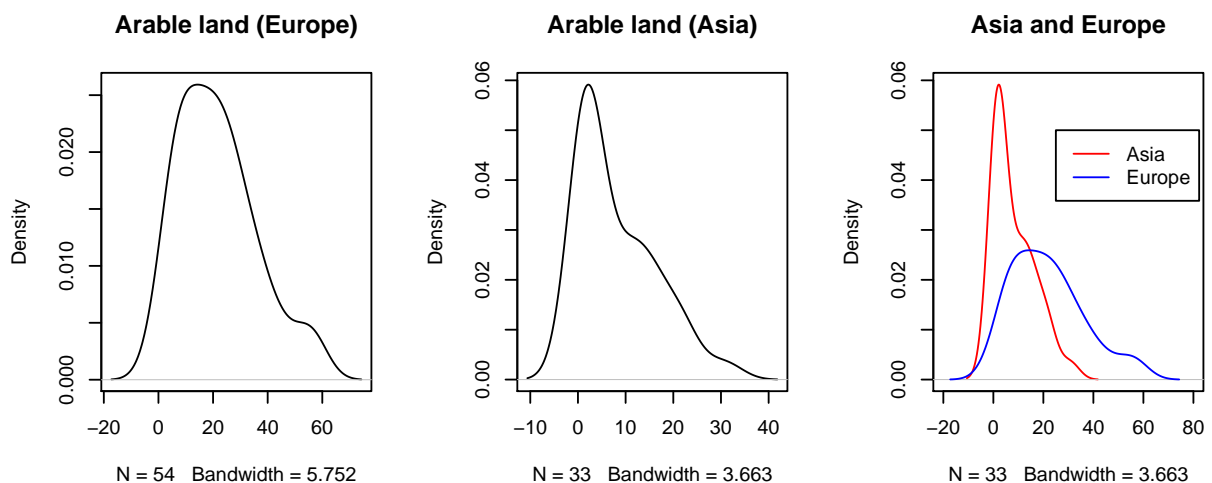
```
par(mfrow=c(1, 3))
europe_density <- density(
  d_land[d_land$region == "Europe & Central Asia (all income levels)", "arable_land_pct"]
```

```

na.rm=TRUE)
asia_density <- density(
  d_land[d_land$region == "East Asia & Pacific (all income levels)", "arable_land_pct"],
  na.rm=TRUE)
plot(europe_density, main = "Arable land (Europe)")
plot(asia_density, main = "Arable land (Asia)")

# Overlaying
plot(asia_density, xlim = c(-20, 80), col='red', main = "Asia and Europe")
lines(europe_density, col='blue')
legend(25, .05, c("Asia", "Europe"),
  lty=c(1,1), # gives the legend appropriate symbols (lines)
  lwd=c(1,1), col=c("red", "blue"))

```

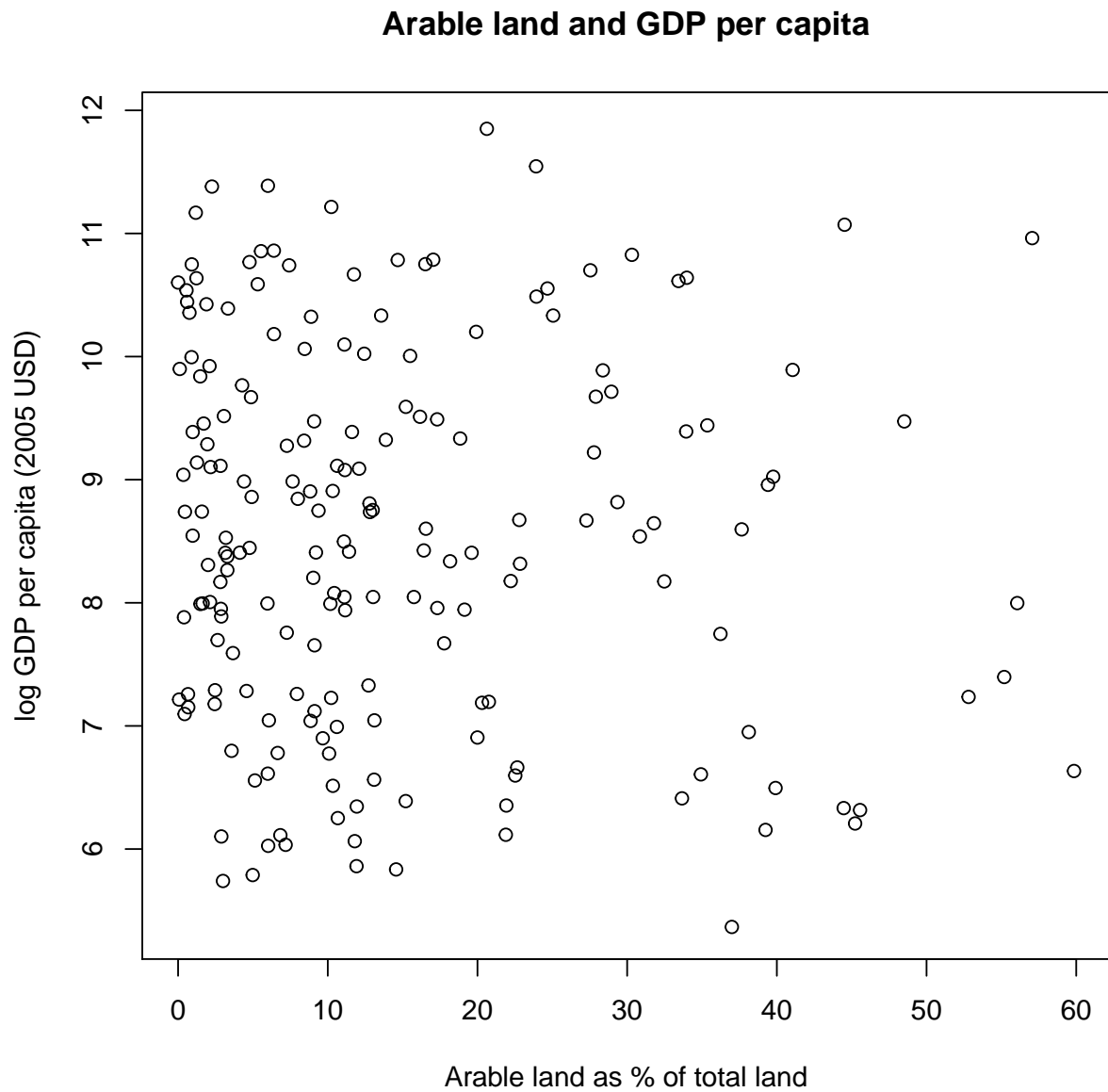


*# Tutorial for legend: <http://www.r-bloggers.com/adding-a-legend-to-a-plot/>*

**c.**

(1/4 point) Draw the scatterplot of that variable against another variable.

```
plot(d_land$arable_land_pct, log(d_land$gdp_percapita),  
     xlab = "Arable land as % of total land",  
     ylab = "log GDP per capita (2005 USD)",  
     main = "Arable land and GDP per capita")
```



d.

(1/4 point) Label the point that represents your country (Hint: Tutorial) and color it red (Some Googling involved)

```
par(mfrow=c(1, 1))
plot(log(gdp_percapita) ~ arable_land_pct,
     data = d_land,
     xlab = "Arable land as % of total land",
     ylab = "log GDP per capita (2005 USD)",
     main = "Arable land and GDP per capita")
d_land_Vietnam <- d_land[d_land$country == "Vietnam", ]
with(d_land_Vietnam,
     text(log(gdp_percapita) ~ arable_land_pct, labels = "Vietnam",
          pos = 3, col = 'red'))
points(d_land_Vietnam$arable_land_pct, log(d_land_Vietnam$gdp_percapita),
       pch = 16, col = 'red')
```

**Arable land and GDP per capita**

