

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Exercise 1: What is the total number of observations in this dataset?

Answer:

Total number of observations in dataset is 63.

Script:

```
# Number of observations is the number of rows in the dataset
mydata=read.csv("http://tinyurl.com/birthdata1")
numberOfObservations = nrow(mydata)
cat("Number of observations in dataset: ", numberOfObservations)
```

Console:

```
> # Number of observations is the number of rows in the dataset
> mydata=read.csv("http://tinyurl.com/birthdata1")
> numberOfObservations = nrow(mydata)
> cat("Number of observations in dataset: ", numberOfObservations)
Number of observations in dataset: 63
```

Exercise 2: What range years are included in this dataset?

Answer:

The years in the dataset range from 1940 to 2002.

Script:

```
# range() returns a vector containing the minimum and maximum of all the given arguments.
mydata=read.csv("http://tinyurl.com/birthdata1")
yearRange = range(mydata$year)
cat("Range of years in the dataset is: ", yearRange)
```

Console:

```
> # range() returns a vector containing the minimum and maximum of all the given arguments.
> mydata=read.csv("http://tinyurl.com/birthdata1")
> yearRange = range(mydata$year)
> cat("Range of years in the dataset is: ", yearRange)
Range of years in the dataset is: 1940 2002
```

Exercise 3: What command would you use to extract just the counts of girls born?

- (a) mydata\$boys
- (b) mydata\$girls
- (c) girls
- (d) mydata[girls]
- (e) \$girls

Answer:

mydata\$girls

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Script:

```
# We can access the data in a single column of a data frame
# separately using a command of the form dataframe$varname.
mydata=read.csv("http://tinyurl.com/birthdata1")
mydata$girls
```

Console:

```
> # we can access the data in a single column of a data frame
> # separately using a command of the form dataframe$varname.
> mydata=read.csv("http://tinyurl.com/birthdata1")
> mydata$girls
 [1] 1148715 1223693 1364631 1427901 1359499 1330869 1597452 1800064 1721216 1733177 1730594 1827830
[13] 1875724 1900322 1958294 1973576 2029502 2074824 2051266 2071158 2078142 2082052 2034896 1996388
[25] 1967328 1833304 1760412 1717571 1705238 1753634 1816008 1733060 1588484 1528639 1537844 1531063
[37] 1543352 1620716 1623885 1703131 1759642 1768966 1794861 1773380 1789651 1832578 1831679 1858241
[49] 1907086 1971468 2028717 2009389 1982917 1951379 1930178 1903234 1901014 1895298 1925348 1932563
[61] 1981845 1968011 1963747
```

Exercise 4: What is the mean (average) of the girl births?

Answer:

The mean(average) of the girl births is

Script:

```
# mean() is the generic function for the (trimmed) arithmetic mean.
mydata=read.csv("http://tinyurl.com/birthdata1")
mean = mean(mydata$girls)
cat("The mean(average) of the girl births is: ", mean)
```

Console:

```
> # mean() is the generic function for the (trimmed) arithmetic mean.
> mydata=read.csv("http://tinyurl.com/birthdata1")
> mean = mean(mydata$girls)
> cat("The mean(average) of the girl births is: ", mean)
The mean(average) of the girl births is: 1793915
```

Exercise 5: Is there an apparent trend in the number of girls born over the years? How would you describe it? Answer with just a few sentences.

Answer:

Yes, there is an apparent trend in the number of girls born over the years. The number of girls born each year progressively increases till it reaches a ceiling value IN 1960 and then declines till it hits a local minima value between 1970 and 1980.

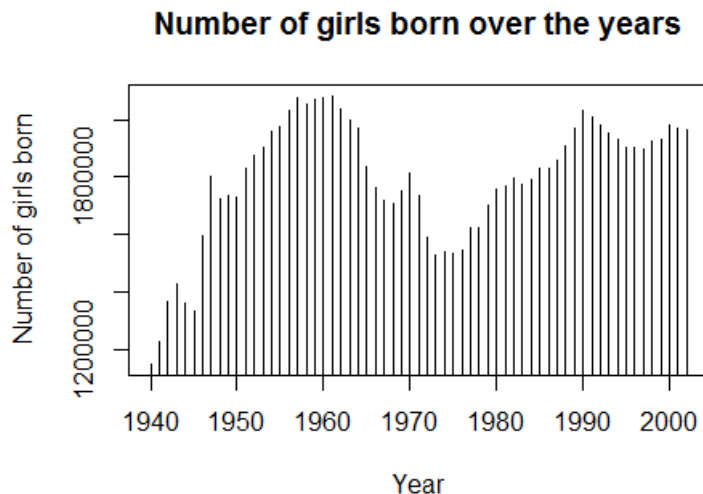
Script:

```
# plot() is the generic function for plotting of R objects
mydata=read.csv("http://tinyurl.com/birthdata1")
plot(x = mydata$year, y = mydata$girls, type="h", main = "Number
      of girls born over the years", xlab = "Year", ylab = "Number of girls born")
```

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Plot:

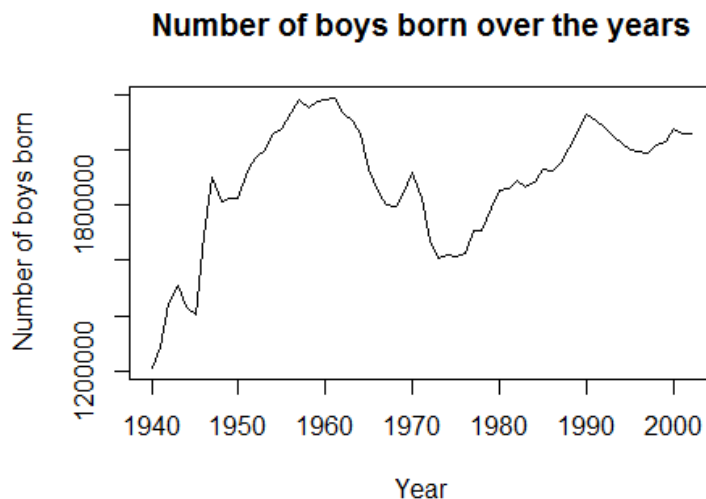


Exercise 6: Create a similar time plot for the boy data using connected lines.

Script:

```
# plot() is the generic function for plotting of R objects
# type = "l" is used for lines
mydata=read.csv("http://tinyurl.com/birthdata1")
plot(x = mydata$year, y = mydata$boys, type="l",
     main = "Number of boys born over the years",
     xlab = "Year", ylab = "Number of boys born")
```

Plot:



Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Exercise 7: What is the minimum of the total number of births? Which year did this happen?

Answer:

Minimum of the total births is 2360399 and this happened in 1940

Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
totalBirths = mydata$boys + mydata$girls
minTotalBirths = min(totalBirths)
minTotalBirthsIndex = match(minTotalBirths, totalBirths)
minTotalBirthsYear = mydata$year[minTotalBirthsIndex]
cat("Minimum of the total births is ",minTotalBirths, " and this happened in ",
    minTotalBirthsYear)
```

Console:

```
> mydata=read.csv("http://tinyurl.com/birthdata1")
> totalBirths = mydata$boys + mydata$girls
> minTotalBirths = min(totalBirths)
> minTotalBirthsIndex = match(minTotalBirths, totalBirths)
> minTotalBirthsYear = mydata$year[minTotalBirthsIndex]
> cat("Minimum of the total births is ",minTotalBirths, " and this happened in ",
+     minTotalBirthsYear)
Minimum of the total births is 2360399 and this happened in 1940
```

Exercise 8: What is the maximum of the total number of births? Which year did this happen?

Answer:

Maximum of the total births is 4268326 and this happened in 1961

Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
totalBirths = mydata$boys + mydata$girls
maxTotalBirths = max(totalBirths)
maxTotalBirthsIndex = match(maxTotalBirths, totalBirths)
maxTotalBirthsYear = mydata$year[maxTotalBirthsIndex]
cat("Maximum of the total births is ",maxTotalBirths, " and this happened in ",
    maxTotalBirthsYear)
```

Console:

```
> mydata=read.csv("http://tinyurl.com/birthdata1")
> totalBirths = mydata$boys + mydata$girls
> maxTotalBirths = max(totalBirths)
> maxTotalBirthsIndex = match(maxTotalBirths, totalBirths)
> maxTotalBirthsYear = mydata$year[maxTotalBirthsIndex]
> cat("Maximum of the total births is ",maxTotalBirths, " and this happened in ",
+     maxTotalBirthsYear)
Maximum of the total births is 4268326 and this happened in 1961
```

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Exercise 9: What are the max and min of the proportion of boys born over time?

Answer:

The max proportion of boys over time is 0.5142562 and the min is 0.5111665

Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
boysBirths = mydata$boys
totalBirths = mydata$boys + mydata$girls
proportionOfBoys = boysBirths/totalBirths
maxProportionOfBoys= max(proportionOfBoys)
minProportionOfBoys= min(proportionOfBoys)
cat("The max proportion of boys over time is ", maxProportionOfBoys,
    " and the min is ",minProportionOfBoys )
```

Console:

```
> mydata=read.csv("http://tinyurl.com/birthdata1")
> boysBirths = mydata$boys
> totalBirths = mydata$boys + mydata$girls
> proportionOfBoys = boysBirths/totalBirths
> maxProportionOfBoys= max(proportionOfBoys)
> minProportionOfBoys= min(proportionOfBoys)
> cat("The max proportion of boys over time is ", maxProportionOfBoys,
+     " and the min is ",minProportionOfBoys )
The max proportion of boys over time is  0.5142562  and the min is  0.5111665
```

Exercise 10: Make a plot of the proportion of boys over time.

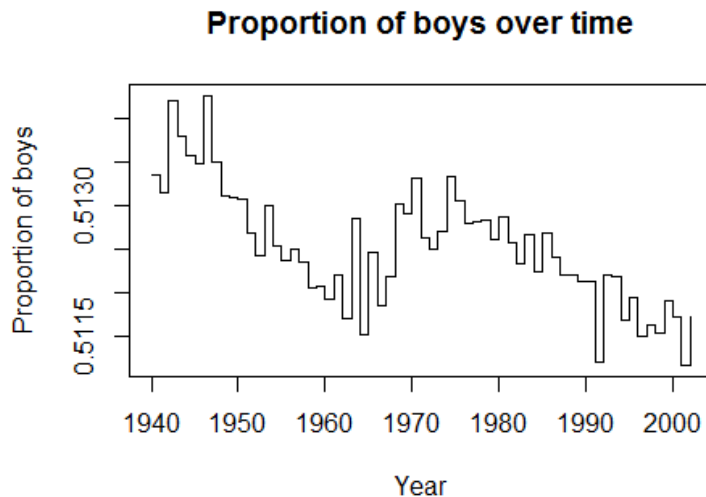
Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
boysBirths = mydata$boys
totalBirths = mydata$boys + mydata$girls
proportionOfBoys = boysBirths/totalBirths
plot(mydata$year, proportionOfBoys, type = "s",
     main = "Proportion of boys over time",
     xlab = "Year", ylab = "Proportion of boys")
```

Plot:

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics



Exercise 11: Briefly comment on the plot of the proportion of boys over time.

Answer:

The proportion of boys over time ranges from 0.5111665 to 0.5142562. While the proportion does change over time, it always exceeds the proportion of girls over time.

Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
comparisonResults = mydata$boys > mydata$girls
table(comparisonResults)
```

Console:

```
> mydata=read.csv("http://tinyurl.com/birthdata1")
> comparisonResults = mydata$boys > mydata$girls
> table(comparisonResults)
comparisonResults
TRUE
63
```

Exercise 12: Which statement is true?

- (a) Every year there are more girls born than boys.
- (b) Every year there are more boys born than girls.
- (c) Half of the years there are more boys born, and the other half more girls born.

Answer:

- (b) Every year there are more boys born than girls.

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Script:

```
# The only response for mydata$boys > mydata$girls is TRUE
mydata=read.csv("http://tinyurl.com/birthdata1")
comparisonResults = mydata$boys > mydata$girls
table(comparisonResults)
```

Console:

```
> # The only response for mydata$boys > mydata$girls is TRUE
> mydata=read.csv("http://tinyurl.com/birthdata1")
> comparisonResults = mydata$boys > mydata$girls
> table(comparisonResults)
comparisonResults
TRUE
 63
```

Exercise 13: Make a plot that displays the boy-to-girl ratio for every year in the data set.

What do you see?

- (a) There appears to be no trend in the boy-to-girl ratio from 1940 to 2002.
- (b) There is initially an increase in boy-to-girl ratio, which peaks around 1960. After 1960 there is a decrease in the boy-to-girl ratio, but the number begins to increase in the mid 1970s.
- (c) There is initially a decrease in the boy-to-girl ratio, and then an increase between 1960 and 1970, followed by a decrease.
- (d) The boy-to-girl ratio has increased over time.
- (e) There is an initial decrease in the boy-to-girl ratio born but this number appears to level around 1960 and remain constant since then.

Answer:

(c) There is initially a decrease in the boy-to-girl ratio, and then an increase between 1960 and 1970, followed by a decrease.

Script:

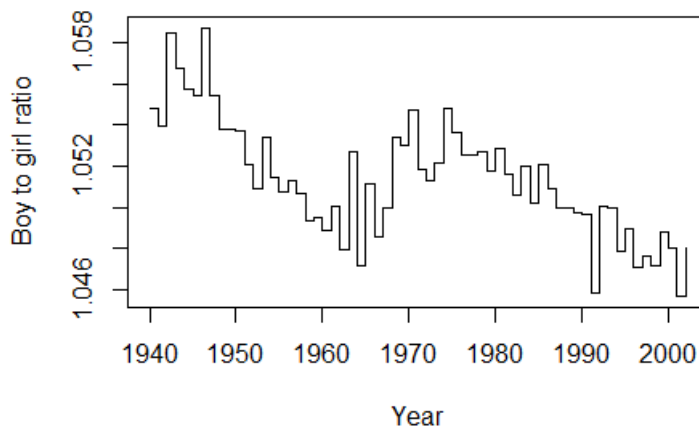
```
mydata=read.csv("http://tinyurl.com/birthdata1")
boyToGirlRatio = mydata$boys/mydata$girls
plot(mydata$year, boyToGirlRatio, type = 's',
     main = "Boy to girl ratio over time",
     xlab = "Year", ylab = "Boy to girl ratio")
```

Plot:

Homework 0

STAT 104 - Introduction to Quantitative Methods for Economics

Boy to girl ratio over time



Exercise 14: Calculate absolute differences between number of boys and girls born in each year and determine which year out of the present data had the biggest absolute difference in the number of girls and number of boys born.

Answer:

Year that had the biggest absolute difference in the number of girls and number of boys born: 1963

Script:

```
mydata=read.csv("http://tinyurl.com/birthdata1")
differences = mydata$boys - mydata$girls
absoluteDifferences = abs(differences)
maxAbsoluteDifference = max(absoluteDifferences)
maxAbsoluteDifferenceIndex = match(maxAbsoluteDifference, absoluteDifferences)
maxAbsoluteDifferenceYear = mydata$year[maxAbsoluteDifferenceIndex]
cat("Year that had the biggest absolute difference
    in the number of girls and number of boys born: "
    ,maxAbsoluteDifferenceYear)
```

Console:

```
> mydata=read.csv("http://tinyurl.com/birthdata1")
> differences = mydata$boys - mydata$girls
> absoluteDifferences = abs(differences)
> maxAbsoluteDifference = max(absoluteDifferences)
> maxAbsoluteDifferenceIndex = match(maxAbsoluteDifference, absoluteDifferences)
> maxAbsoluteDifferenceYear = mydata$year[maxAbsoluteDifferenceIndex]
> cat("Year that had the biggest absolute difference
+     in the number of girls and number of boys born: "
+     ,maxAbsoluteDifferenceYear)
Year that had the biggest absolute difference
in the number of girls and number of boys born: 1963
```