Common Sta 101 Commands for R

1 inference function

Use the following command to load the inference function:

```r
source("http://stat.duke.edu/courses/Fall12/sta101.001/labs/inference.R")
```

```r
inference(data, group, est, type, method, null, alternative, success, order, conflevel, siglevel, nsim)
```

- `data` = response variable, categorical or numerical variable
- `group` = explanatory variable, categorical (optional)
- `est` = "mean", "median", or "proportion"
- `type` = "ci" for confidence interval, or "ht" for hypothesis test
- `method` = "theoretical" or "simulation"
- `null` = (optional) null value, does not need to be defined for chi-square or ANOVA
- `alternative` = (optional) "less", "greater", or "twosided"
- `success` = (optional) if data is categorical, the name of the level that is defined as success
- `order` = (optional) if group is defined, the order in which to subtract the groups
- `conflevel` = (optional) for confidence intervals, 0.95 by default
- `siglevel` = (optional) for hypothesis testing, 0.05 by default
- `nsim` = (optional) number of simulations, 10000 by default

2 One quantitative variable

Summary statistics

```r
summary(x)
```

- `most summary statistics at once`

```r
mean(x)
```

- `na.rm = TRUE to get rid of NA values`

```r
median(x)
```

- `na.rm = TRUE to get rid of NA values`

```r
sd(x)
```

- `na.rm = TRUE to get rid of NA values`

Visualization

```r
hist(x)
```

```r
boxplot(x)
```
# horizontal = TRUE for horizontal plot

```r
qqnorm(x)
qqline(x)
```

# for normal probability plot and straight line

## 3 One categorical variable

### Summary statistics

```r
table(x)
```

### Visualization

```r
barplot(table(x))
```

## 4 Two quantitative variables

Note: Out of scope for project 1.

### Summary statistics

```r
cor(x, y)
```

# use = "complete.obs" to get rid of NA values

### Visualization

```r
plot(y ~ x)
```

## 5 Two categorical variables

### Summary statistics

```r
table(x, y)
```
Visualization

```r
barplot(table(x,y))
# beside = TRUE for side-by-side barplot
# legend = TRUE to include a color legend
mosaicplot(table(x,y))
```

6 One categorical and one quantitative variable

\( y = \text{quantitative} \)
\( x = \text{categorical} \)

Summary statistics

```r
by(y, x, summary)
# summary by group
by(y, x, mean)
# mean by group
# na.rm = TRUE to get rid of NA values
by(y, x, sd)
# sd by group
# na.rm = TRUE to get rid of NA values
```

Visualization

```r
boxplot(y ~ x)
```

7 Subsetting

```r
subset(dataname, !is.na(x))
# the data set "dataname", but only cases for which x is not NA
subset(dataname, x == "levelA")
# the data set "dataname", but only cases for which x is equal to "levelA"
x[!is.na(x)]
# the variable x, but only cases for which x is not NA
y[!is.na(x)]
# the variable y, but only cases for which x is not NA
x[x < 30]
# the variable x, but only cases for which x is less than 30
x[x != "levelA"]
# the variable x, but only cases for which x does not equal "levelA"
```
8 Probability distributions

\texttt{pnorm(q, mean, sd)}
\begin{itemize}
  \item # calculate area under the normal curve below q
  \item # for a normal distribution with given mean and sd
\end{itemize}
\texttt{dnorm(x, mean, sd)}
\begin{itemize}
  \item # calculate the normal probability density at x (can be a vector)
  \item # for a normal distribution with given mean and sd,
  \item # useful for plotting a normal curve over a histogram
\end{itemize}
\texttt{dbinom(x, size, prob)}
\begin{itemize}
  \item # calculate the probability for x successes in size trials,
  \item # where probability of success is prob
\end{itemize}

9 Plotting lines

\texttt{abline(h = value)}
\begin{itemize}
  \item # add a horizontal line to an existing plot
\end{itemize}
\texttt{abline(v = value)}
\begin{itemize}
  \item # add a vertical line to an existing plot
\end{itemize}
\texttt{abline(lm(y~x))}
\begin{itemize}
  \item # overlays linear regression line on the scatterplot of y vs. x,
  \item # only works if plot(y ~ x) ran first
\end{itemize}

10 Sampling

\texttt{sample(x, size, replace = FALSE)}
\begin{itemize}
  \item # sample from x size number of elements without replacement (default)
  \item # to sample with replacement replace = TRUE
\end{itemize}

11 Plotting options

These arguments can be passed to the \texttt{plot}, or \texttt{hist}, or other similar functions. To learn more about all plotting parameters, type \texttt{?par}.

\texttt{main = "main title"}
\item # title of plot, to be placed in the top center
\texttt{xlab = "x-axis label"}
# x-axis label
ylab = "y-axis label"
    # y-axis label
xlim = c(min, max)
    # x-axis limits
ylim = c(min, max)
    # y-axis limits