

Statistics with R

Drew Schmidt

04/2/2015

Throughout these exercises, you will need to have the following data in your R session:

```
workshop <- c(1, 2, 1, 2, 1, 2, 1, 2)
workshop <- factor(workshop)

gender    <- c("f", "f", "f", NA, "m", "m", "m", "m")
gender    <- factor(gender)

q1 <- c(1, 2, 2, 3, 4, 5, 5, 4)
q2 <- c(1, 1, 2, 1, 5, 4, 3, 5)
q3 <- c(5, 4, 4, NA, 2, 5, 4, 5)
q4 <- c(1, 1, 3, 3, 4, 5, 4, 5)

df <- data.frame(workshop, gender, q1, q2, q3, q4)
```

Basics

1. Inspect the `df` object. Try various commands (`print`, `str`, `class`, ...).
2. Examine the `colnames()` of `df`. (Try `colnames(df)`).
3. Get a `summary()` of `df`.
4. Use the `na.exclude()` function to remove NA's from `df`. Store the result in an object named `df_nona`.

Basic Plots

We can access an individual column of a dataframe via the `$` operator. So you could enter `df_nona$gender` into an R session and have only that variable returned.

5. Generate a barplot of gender using the `barplot()` command.
6. Generate a stacked barplot of gender by workshop using the `plot()` command. (Hint: `plot(df_nona$gender, df_nona$workshop)`).

Basic Statistics

7. Fit a linear model via:

```
mymodel <- lm(q4 ~ q1 + q2 + q3, data=df_nona)
```

8. Inspect the model object in various ways.
9. Perform an analysis of variance (Hint: `anova(mymodel)`).
10. See what happens when you call `plot(mymodel)`.

11. Drop the first two columns and store it in a new dataframe called `df2`. One possible way to do this is:

```
df2 <- df_nona[, 3:6]
```

12. Compute the principal components of the data in `df2` (Hint: `prcomp()`).
13. Compute the covariance matrix of the data in `df2` (Hint: `cov()`).