

1. Section 2.3.1 Basic Commands

The function "summary" is a useful one that, when applied to a vector of numbers, returns the minimum, lower quartile, median, mean, upper quartile, and maximum. After entering `set.seed(3)`, `y=rnorm(100)`, and `summary(y)`, what is the median of y? (Make sure not to enter any commands between "set.seed(3)" and "y=rnorm(100)" and round your answer to 4 decimal places.)

Solution: 0.0342

```
set.seed(3)
y = rnorm(100)
summary(y)
median(y)
```

2. Section 2.3.2 Graphics

Use the help topic for "plot" to find out which argument creates a subtitle at the bottom of the plot.

- a) description
- b) footer
- c) foot
- d) subtitle
- e) sub

Solution: E

```
?plot
# sub: a sub title for the plot: see title.
```

3. Section 2.3.3 Indexing Data

If an n by p matrix X contains a dataset with observations as the n rows and variables as the p columns (i.e., the usual arrangement of a data matrix), which R command returns a vector containing the values of the 2nd variable for all n observations?

- a) X[2]
- b) X[2,]
- c) X[,2]
- d) X[2,n]
- e) X[n,2]

Solution: C

```
X <- matrix(1:16, 4, 4)
X[, 2]
```

4. Section 2.3.4 Loading Data

After loading a data frame into R it is helpful to use the "str" function to display the structure of the data frame, in particular the variable types. For example, decimal values are of numeric type ("num"), integers are of integer type ("int"), and character variables are converted to factors by "read.table" unless "stringsAsFactors=T" is specified. Note that R determines whether a quantitative variable is imported as "num" or "int" depending on how the values are recorded in the data set. For example, if a variable consists only of integer values, it will be imported as "int." However, if each value is recorded with a zero in the first decimal place (i.e., 1.0 instead of 1) then it will be imported as "num." For the 8 quantitative variables in the Auto.data data set, how many are imported as integer type rather than numeric type? (Hint: be sure to clear your workspace before loading the Auto.data data set.)

Solution: 3

```
str(Auto)
```

```
> str(Auto)
'data.frame': 392 obs. of 9 variables:
 $ mpg      : num  18 15 18 16 17 15 14 14 14 15 ...
 $ cylinders : int   8  8  8  8  8  8  8  8  8  8 ...
 $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
 $ horsepower  : num  130 165 150 150 140 198 220 215 225 190 ...
 $ weight      : num  3504 3693 3436 3433 3449 ...
 $ acceleration: num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
 $ year        : int   70 70 70 70 70 70 70 70 70 70 ...
 $ origin      : int   1  1  1  1  1  1  1  1  1  1 ...
 $ name        : Factor w/ 304 levels "amc ambassador brougham",...
 - attr(*, "na.action")=Class 'omit' Named int [1:5] 33 127 331 3
 .. ..- attr(*, "names")= chr [1:5] "33" "127" "331" "337" ...
```

5. Section 2.3.5 Additional Graphical and Numerical Summaries

Based on the boxplots of miles per gallon by number of cylinders for the Auto.data data set, order the number of cylinders according to the median miles per gallon from highest to lowest.

- a) 3
- b) 4
- c) 5
- d) 6
- e) 8

Solution: Order = B (Highest), C (Second Highest), A (Middle), D (Second-Lowest), E (Lowest)

```
attach(Auto)
cylinders=as.factor(cylinders)
plot(cylinders, mpg)
# From the resulting boxplots the medians are ordered 4, 5, 3, 6, 8.
```

6. Section 2.3.5 Additional Graphical and Numerical Summaries

How would you describe the histogram of miles per gallon for the Auto Data set?

- a) Symmetric and bell-shaped (i.e., “normal”)
- b) Symmetric but not bell-shaped
- c) Slightly skewed with a long left tail
- d) Slightly skewed with a long right tail
- e) Bimodal

Solution: D

```
hist(mpg)
# The resulting histogram looks slightly skewed with a long right tail.
```

7. Section 2.3.5 Additional Graphical and Numerical Summaries

The scatterplot matrix of miles per gallon, displacement, horsepower, weight, and acceleration reveals reasonably strong, negative, nonlinear associations between miles per gallon and each of three of the other four variables. Which of the variables does *not* display much of an association with miles per gallon?

- a) Displacement
- b) Horsepower
- c) Weight
- d) Acceleration

Solution: D

```
pairs(~ mpg + displacement + horsepower + weight + acceleration, Auto)
# Of the four scatterplots in the top row (with mpg on the vertical axis), only the scatterplot of
acceleration versus mpg in the top right doesn't show much evidence of an association.
```

8. Section 2.3.5 Additional Graphical and Numerical Summaries

In the scatterplot of horsepower versus miles per gallon, one car "sticks out" with a relatively high miles per gallon of more than 30 given its moderate horsepower of about 130. What is this car? (Use the "identify" function.)

- a) Cadillac Eldorado
- b) Chevrolet Citation
- c) Datsun 200sx
- d) Datsun 280-zx
- e) Mazda GLC

Solution: D

```
plot(horsepower,mpg)
identify(horsepower,mpg,name)
# The car that "sticks out" is observation #334, which is a Datsun 280-zx.
```