

# Questions for final exam

*Lakshmi Arbatti*

*Sunday, March 29, 2015*

## Lecture 8a

What does the ggplot function **geom\_point** do?

- The function is used to create scatterplots.
- This function is helpful to get the geometric points in a dataset.
- This function can be used to create a histogram.
- This function does not exist at all.

## Lecture 8b

In statistics, what is the meaning of **multicollinearity**?

- Its a phenomenon in which two or more predictor variables in a regression model are highly correlated.
- Its a model in which many values are linear.
- There is no such thing as multicollinearity in statistics.
- It is a model in which there is no relationship between multiple variables.

## Lecture 9a

Given the piece of code below:

```
N <- 100
df <- data.frame(
  var1 = runif(N, min=0, max=10),
  var2 = sample(letters[1:5], N, replace=T)
)
kable(head(df))
```

Which of the variables declared above are categorical?

- var2 is the categorical variable
- var1 is the categorical variable
- The sample does not have any categorical variable

## Lecture 9b

What does **floor(2.9)** return?

- Returns the number 2
- Throws an error since the function floor does not exist in R
- Rounds the number 2.9 to 3
- Returns the number 2.9

## Lecture 10a

What does **readRDS()** function do?

- Reads a binary file into a dataframe
- Reads a csv data set
- There is no such function in R

## Lecture 10b

Which function returns the column names of a dataframe?

- names()
- getcols()
- readRDS()
- readdata()

## Lecture 11a

What is **Centrality** in the igraph package?

- Degree of the graph
- The central point in a graph
- A point in the graph
- There is no such term in igraph package

## Lecture 11b

What is **Vertex and edge betweenness()** in the igraph package?

- The number of geodesics (shortest paths) going through a vertex or an edge
- The distance between two points in a graph
- Does not really mean anything

## Lecture 12a

What is the equation for a line?

- $y = mx + b$ , where  $y$  is the y intercept,  $m$  is the slope and  $b$  is the constant
- $y = mx + b$ , where  $y$  is the name of the line,  $m$  is the mean
- $a + b + c = 0$
- None of the choices

## Lecture 12b

Which plotting function adds one or more straight lines through a current plot?

- `abline()`
- `addline()`
- `moreline()`
- None of the choices

## Lecture 13a

What does the generic **group\_by** function do?

- The function groups a table by one or more variables
- The function is used in logistic regression to group similar labels
- The function groups different variables into a single variable
- There is no such function in R

## Lecture 13b

In the following piece of code, what is the **cut** function used for?

```
N <- 10
age <- runif(N,7,10.5)
grade <- cut(age,breaks = 7:11,labels = 2:5,right =TRUE)
```

- To convert numeric values in the vector `age` to factors and store the values in the vector `grade`
- To cut and paste values from the vector `age` into vector `grade`
- To cut values from the vector `age` and store them in the environment variables
- **cut** throws an error

## Lecture 14a

What is the **manipulate** function useful for?

- The **manipulate** function can be used to create interactive plots with slider, picker, checkbox or button
- The **manipulate** function can be used to manipulate a data frame
- The **manipulate** function can be used to change the data in a database table
- The **manipulate** function doesn't really do anything

## Lecture 14b

Which function can be used to fit **Generalized Linear Models**

- The **glm()** function
- The **lm()** function
- The **gen()** function

## Lecture 15a

In statistics, what is **homoscedasticity**?

- A sequence or vector is **homoscedastic** if the variables in the sequence or vector have finite variance
- **homoscedasticity** is the science of measuring the coefficients in a dataset
- There is no such term as **homoscedasticity** in statistics
- A sequence is **homoscedastic** if the variables in the sequence have no or unequal variance

## Lecture 15b

Based on the video by Andrew Ng on learning curves, which of the following is a true statement for High Bias algorithms?

- If a learning algorithm is suffering from high bias, getting more data will not (by itself) help much
- High bias algorithms can easily be resolved with small data samples
- High bias algorithms are biased to sensitivity
- High bias algorithms are biased to specificity