Practice Midterm

HS616

March 24, 2015

Question 1

Which of the following equations represents the sensitivity of a test?

- A : sensitivity = number of true positives / number with disease
- B : sensitivity = number of true positives / number of true negatives
- C : sensitivity = number of true negatives / number without disease
- D : sensitivity = number with disease / total population

Question 2

In the following code what does the function xtabs do ?

```
T_shirts <- data.frame(
  sex=sample(c("M","F"), 100, replace=T),
  size=sample(c("L", "M", "S"), 100, replace=T)
)
table(T_shirts)
xtabs(~ sex + size, T_shirts)
```

- A : Introduces equally spaces tabs between columns in the output file
- B : Breaks one data frame into separate dataframes depending on the arguments that are passed to to xtab
- C : Eliminates duplicate data in a table and merges data
- D : Crosstabulates variables with small numbers of unique values

Question 3

What does the following function return?

```
f <- function(x) {
  f <- function(x) {
    f <- function(x) {
      x ^ 2
    }
    f(x) + 1
  }
  f(x) * 2
}
f(10)
```

- A : 441
- B : 202
- C : 200
- D : 40

Question 4

Consider the following code, then select the correct statement regarding it.

```
maxMinusMin <- function(v, ...) max(v, ...) - min(v, ...)
apply(A, 1, maxMinusMin, na.rm=TRUE)
```

- A : If additional parameters are given to the function, they will be passed to `max` and `min`
- B : Typing error
- C : It's an invalid function that will need more parameters
- D : Function is invalid and cannot be executed

Question 5

The 'stingsAsFactors = FALSE' option is useful when reading a data file because:

- A : all choices are correct
- B : even if stingsAsFactors= F, it is easy to convert character data to factors using as.factor()
- C : it allow us to keep character variables as they are rather than convert to factors
- D : the default in R is for columns with character data to be made into factors

Question 6

What data type does apply function return?

- A : Lists
- B : Vectors
- C : All of these answers are correct
- D : Matrices

Question 7

Consider a sequence of 10 coin flips, represented by the string `TTTHTHTTTH`. Which statement gives the total number of different sequences of 10 coin flips that could result in this number of heads?

- A : `sapply(3:10, function(x) factorial(x))`
- B : `integrate(dnorm, -Inf, 0)`
- C : `choose(10,3)`
- D : `factorial(10)/(factorial(4)*factorial(7))`

Question 8

Indentify the distribution type in the following code:

```
x <- seq(0, 4, 0.1)
plot(x, dnorm(x, 2, 0.5), type = "l")
```

- A : Normal
- B : Unified constant
- C : Binomial
- D : Poisson

Question 9

The command `tidyr::gather(df, var, val)` produced the following result:

```
  var val
1   a   1
2   a   2
3   a   3
4   b   1
5   b   2
6   b   3
```

Which answer correctly defines the dataframe `df`?

- A : `df <- data.frame(a=var[1:3], b=val[1:3])`
- B : `df <- data.frame(a=1:3, b=1:3)`
- C : `df <- data.frame(var=rep(c('a','b'), each=3), val=rep(1:3, times=2))`
- D : `df <- data.frame(var=letters[1:3], val=letters[1:3])`

Question 10

Consider the following code:

```
N <-10000
x <- runif(N)
y <- runif(N)
vlength <- sqrt(x^2 +y^2)
in_circle <- vlength < 1
```

Which of the following could be the output of `head(as.integer(in_circle))` ?

- A : `1 1 1 1 1 0`
- B : `0.23, ,0.34, 0.12, 0.45, 0.55, 0.79`
- C : `1 -1 1 0 -1 0`
- D : `TRUE TRUE TRUE TRUE TRUE FALSE`

Question 11

What `SQL` command does matrix multiplication between matrices `A` and `B`?

- A : `SELECT SUM(A.value * B.value) FROM A, B WHERE A.col_num = B.row_num;`
- B : `SELECT A.row_num, B.col_num, SUM(A.value * B.value) AS value FROM A, B WHERE A.col_num = B.row_num GROUP BY A.row_num, B.col_num;`
- C : `SELECT A.row_num, B.col_num AS value FROM A, B WHERE A.col_num = B.row_num;`

Question 12

In database management, what is meant by "Data Aggregation"?

- A : Normalizing the data in a database table
- B : Finding the mean of columns in a database table
- C : The process by which data is gathered and summarized for further statistical analyses
- D : Using an inner join to extract data from a table

Question 13

Every data type is at least a _____

- A : matrix
- B : vector
- C : array
- D : factor

Question 14

When working with databases through R on your local computer, what is the advantage of working with SQLite instead of MySQL?

- A : SQLite is also suitable for a multi-user environment where hundreds of users connect to the database simultaneously.
- B : There isn't an advantage because there is no way to connect to a SQLite database in R.
- C : SQLite uses a flat file, as opposed to requiring a database connection.
- D : There are no packages to connect to a MySQL database in R.

Question 15

Consider this R code showing two ways of calculating the cost of daily medicine, and select the true statement.

```
price <- c( lisonopril=106/30, crestor=204.00/30,
            clorthiazide=12.10/15, fibrosol=160/30)
dosage_day <- c( lisonopril=3, crestor=0.5,
            clorthiazide=0.5, fibrosol=1)
cost_day_1  = sum(price * dosage_day)
cost_day_1a = price %*% dosage_day
```

- A : R is fun only for statisticians
- B : The Dot product of 2 vectors equals the sum of the element-wise products of the vectors
- C : A diagonal times a vector of that diagonal results in a squared value
- D : A vector times a vector is a scalar

Question 16

What is the result of the following code?

```
A <- matrix(1:4, nrow=1)
A %*% t(A)
```

- A : Aˆ2, a matrix of type double
- B : inv(A), an integer matrix
- C : a 2X2 matrix of type double
- D : 30, a single integer

Question 17

The inverse of matrix A (of size 3x3) is called A_inverse (of size 3x3). Which is the following is FALSE?

- A : 'A * A_inverse' returns an identity matrix (of size 3x3)
- B : 'solve(A_inverse)' returns A
- C : 'solve(A)' returns A_inverse
- D : 'A %*% A_inverse' returns an identity matrix (of size 3x3)

Question 18

Which of the following function keeps track of the function stack and tabulates how much time is spent on each function?

- A : runif()
- B : rnorm()
- C : system.time()
- D : RProf()

Question 19

Which symbol can be used for slicing and extracting data from a vector in R?

- A : [, c( )]
- B : [[c( )]]
- C : $
- D : [ ]

Question 20

The Poisson Distribution is a type of

- A : Cumulative distibution
- B : Continuous Probability Distribution
- C : Discrete Probability Distribution
- D : Random number generation

Question 21

Which statement is true if Q is a 2 by 2 orthogonal matrix?

- A : `all.equal(solve(Q), t(Q))`
- B : `all.equal(Q %*% diag(2), diag(2) %*% t(Q))`
- C : `all.equal( Q %*% diag(2), diag(2) )`
- D : `all.equal( solve(Q) %*% Q, Q )`

Question 22

What R function can be used to generate standard Normal random variables?

- A : pnorm
- B : rnorm
- C : qnorm
- D : dnorm

Question 23

x is a data frame and z is a column of x. Which of the following commands is equivalent to with(x, f(z))?

- A : f(z)
- B : f(x$z)
- C : x$f(z)
- D : It depends.

Question 24

The standard normal distribution has a mean of 0 and a standard deviation of 1, and the area under this curve over all possible x-values is one. What is the area under the curve of a normal probability distribution function with a standard deviation of 2?

- A : 2
- B : 2 pi
- C : 1
- D : 4

Question 25

What is the name of the R function that does the equialent of SQL joins?

- A : sqlJoin
- B : merge
- C : aggregate
- D : join

Question 26

Which of the following is a common problem with messy datasets?

- A : Data is in human readable format
- B : One entity is stored in multiple tables
- C : Primary and foreign keys are well defined
- D : Data is in the third norm form

Question 27

In the following code, what is the type of the variable returned?

```
y <- c(5, 6, 7, 8, NA)
is.na(y)
```

- A : logical
- B : character
- C : numeric
- D : integer

Question 28

What is the correct code for subtracting two dates from one another and then cast the difference to a numeric value?

- A : `as.Date("2014-10-10") - as.Date("2014-10-1") %>% as.numeric`
- B : `(as.Date("2014-10-10") - as.Date("2014-10-1")) %>% as.numeric`
- C : `as.Date %>% ("2014-10-10") - as.Date %>% ("2014-10-1") >%> as.numeric`
- D : `as.numeric %>% (as.Date("2014-10-10" - "2014-10-1" ))`

Question 29

xtab() does the following:

- A : all answers are correct
- B : is similar to table()
- C : can be used to easily generate a sparseMatrix
- D : crosstabulates variables

Question 30

What is the output of the following code?

```
x <- function(numRows=5, numCols=5, probZero=0.7, seed=NULL){
    if(!is.null(seed)) set.seed(seed)
    matrix( rnbinom(numRows * numCols, prob=probZero, size=1), nrow=numRows )
}
```

- A : Generates a random sparse matrix
- B : Generates a random vector
- C : Generates bionomial distribution values and stores them in x
- D : The code does not work

Question 31

How does an ellipsis behave as a function parameter in R?

- A : Each period acts as an anonymous parameter in the function.
- B : It takes an undefined number of arguments and applies them wherever the ellipsis is used in the function, similar to a normal parameter.
- C : It takes each argument passed in by the user and applies them to undefined variables in the function based on order.

Question 32

Which of these lines of code cannot be used to generate a random data set?

- A : `replicate( 100, runif(n=20))`
- B : `z = rnorm(20, mean=10, sd=3)`
- C : `qnorm(c(.05,.95))`
- D : `n <- rnorm(2500, mean=65, sd=4.58)`

Question 33

Generate a sequence from 122 to 154 by intervals of 2.

- A : seq(154,122,2)
- B : seq(2,154,122)
- C : seq(122,154,2)
- D : seq(122,2,154)

Question 34

What is an eigenvector?

- A : a vector which, when multiplied by a square matrix, generates the same value as when the vector is multiplied by its eigenvalue.
- B : a vector with determinant zero.
- C : a vector that Dr. Richard Eigen designed to find a series of velocities relevant in physics.

Question 35

Simulated coin-tossing can be done using different methods. Which of the following will NOT work?

- A : coin <- sample(c("H", "T"), 10, replace = F)
- B : c("H", "T") [1 + rbinom(10, 1, .5)]
- C : rbinom(10, 1, .5)
- D : ifelse(rbinom(10, 1, .5) == 1, "H", "T")