Tarea 2

Valentina Gonzalez -Andres Felipe Beltran Rodriguez

11/8/2021

# Primer punto

1. Los agricultores de la sabana de Bogotá realizaron un estudio respecto a la aplicación de fungicidas y abonos orgánicos y su relación con la presencia de daños por patógenos. Se utilizó una encesta para recoger la información. La encuesta tenía las siuguientes tres preguntas:
2. Qué tipo de abono orgánico utiliza?
   1. Gallinaza
   2. Compost
   3. Cascarilla de arroz+desechos vegetales
   4. Guano
3. Qué tipo de aplicación de fungicidas utiliza?
   1. Metil Tiofanato 70% (Enovit Metil)
   2. Belanaxil 8% (Galben M)
   3. Azufre 80% (Sufrevit)
4. Cómo considera la presencia de daños por patógenos?
   1. Alto
   2. Bajo

Los encuestados únicamente escogían un tipo de fungicida y un tipo de abono orgánico.

Todos los encuestados respondieron todas las preguntas. Los resultados encontrados fueron los siguientes:

1. El 12% de las personas utilizan Metil Tiofanato 70% y Cascarilla de arroz+desechos vegetales.
2. De las personas que prefieren Cascarilla de arroz+desechos vegetales el 50% encuentran la presencia de daños por patógenos alta independiente del tipo de aplicación de fungicidas.
3. El 9% de las personas utilizan Belanaxil 8% y Cascarilla de arroz+desechos vegetales.
4. El 40% de los que utilizan Guano, utilizan Metil tiofanato 70%.
5. El 40% de las personas que utiliza Compost encuentran la presencia de daños por patógenos alta sin importar el tipo de fungicidas que se aplique.
6. El 20% de las personas utilizan Gallinaza.
7. De los que utilizan Gallinaza, el 40% utiliza Metil Tiofanato 70%.
8. De los que utilizan Gallinaza, el 40% utiliza Belanaxil 8%.
9. El 15% de las personas utiliza Compost y Metil Tiofanato 70%.
10. El 9% de las personas utiliza Azufre 80% y Cascarilla de arroz+desechos vegetales.
11. El 30% de las personas que utiliza Gallinaza encuentran la presencia de daños por patógenos alta sin importar el tipo de fungicida.
12. De los que utilizan Guano nadie utiliza Azufre 80%.
13. El 6% de las personas utilizan Compost y Belanaxil 8%.
14. De las personas que utilizan Guano el 60% encuentran lapresencia de daños por patógenos alta sin importar el tipo de fungicidas.
15. El 30% de las personas prefieren Compost.
16. Construya el árbol de la situación anterior y calcule todas las probabilidades en todas las intersecciones y en todas las ramas del árbol.

prob\_data <- data.frame(pathString = c(  
 'Gallinaza',  
 'Compost',  
 'Cascaras',  
 'Guano',  
 'Gallinaza/Me',  
 'Gallinaza/Be',  
 'Gallinaza/Az',  
 'Compost/Me',  
 'Compost/Be',  
 'Compost/Az',  
 'Cascaras/Me',  
 'Cascaras/Be',  
 'Cascaras/Az',  
 'Guano/Me',  
 'Guano/Be',  
 'Guano/Az',  
 'Gallinaza/Me/A',  
 'Gallinaza/Me/B',  
 'Gallinaza/Be/A',  
 'Gallinaza/Be/B',  
 'Gallinaza/Az/A',  
 'Gallinaza/Az/B',  
 'Compost/Me/A',  
 'Compost/Me/B',  
 'Compost/Be/A',  
 'Compost/Be/B',  
 'Compost/Az/A',  
 'Compost/Az/B',  
 'Cascaras/Me/A',  
 'Cascaras/Me/B',  
 'Cascaras/Be/A',  
 'Cascaras/Be/B',  
 'Cascaras/Az/A',  
 'Cascaras/Az/B',  
 'Guano/Me/A',  
 'Guano/Me/B',  
 'Guano/Be/A',  
 'Guano/Be/B',  
 'Guano/Az/A',  
 'Guano/Az/B'  
 ),  
 prob = c(0.2,  
 0.3,  
 0.3,  
 0.2,  
 0.4,  
 0.4,  
 0.2,  
 0.5,  
 0.2,  
 0.3,  
 0.4,  
 0.3,  
 0.3,  
 0.4,  
 0.6,  
 0,  
 0.3,  
 0.7,  
 0.3,  
 0.7,  
 0.3,  
 0.7,  
 0.4,  
 0.6,  
 0.4,  
 0.6,  
 0.4,  
 0.6,  
 0.5,  
 0.5,  
 0.5,  
 0.5,  
 0.5,  
 0.5,  
 0.6,  
 0.4,  
 0.6,  
 0.4,  
 0.6,  
 0.4  
 ))  
# knitr::kable(prob\_data)

## Creando variables de ayuda

El objetivo de este paso es generar nuevas variables a partir de las entradas originales que van a ayudarnos a definir la estructura requerida para diobujar el arbol.

1. tree\_level : el nivel de la rama en la cual se va a dibujar una probabilidad en especifico
2. tree\_group: El nombre de la primera rama que se usa para buscar la probabilidad madre.

3 node\_type = un nombre unico para construir componentes a medida para la visualizacion.

También creamos la variable llamadamax\_tree\_level qUe nos dice el numero total de ramas en el arbol.

library(data.tree)  
library(dplyr)  
library(stringr)  
library(DiagrammeR)  
prob\_data <- prob\_data %>% mutate(  
 tree\_level = str\_count(string = pathString,   
 pattern = '/')+1,  
 tree\_group = str\_replace(string = pathString,   
 pattern = '/.\*',  
 replacement = ''),  
 node\_type = 'decision\_node')  
  
max\_tree\_level <- max(prob\_data$tree\_level, na.rm = T)

## Encontrando las probabilidades padre

para determinar la probabilidad acumulada para un resultado dado, es necesario multiplicar las probabilidades en ramas secundarias por las probabilidades de las ramas madres asociadas a cada una de ellas.

Para hacer esto, creamos un nuevo data\_frame parent\_lookup, que contiene todas las probabilidades que vienen de nuestra fuente de datos.

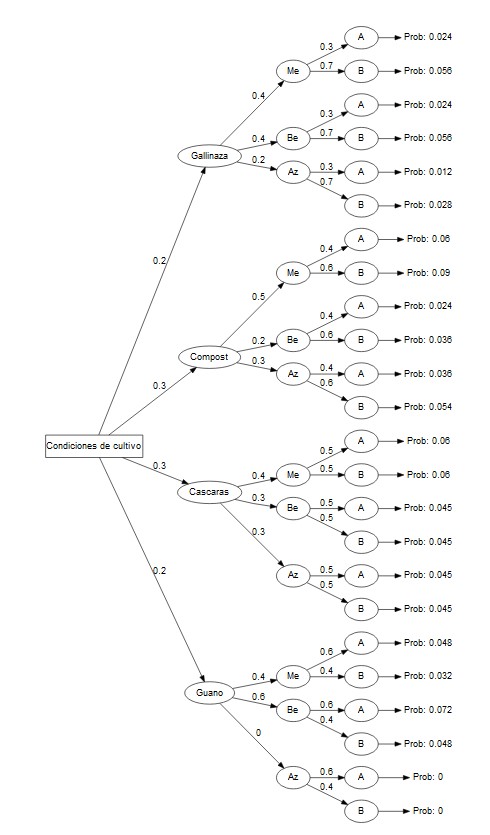
Podemos entonces iterar a traves de los cuatro niveles, tomando las probabilidades de todas las ramas madre. Al final podemos calcular la probabilidad acumulada overall\_prob al multiplicar a lo largo de todas probabilidades a traves de una secuencia de ramas.

Primero, obtenemos las distintas probabilidades para poder encontrar las probabilidades del nodo madre:

parent\_lookup <- prob\_data %>% distinct(pathString, prob)

Luego iteramos a lo largo de la primera capa para obtener todas las probabilidades madre.

for (i in 1:(max\_tree\_level - 1)) { # loop through all tree layers to get all immidiate parent probabilities (to calculate cumulative prob)  
   
 names(parent\_lookup)[1] <-paste0("parent",i)  
 names(parent\_lookup)[2] <-paste0("parent\_prob",i)  
   
 for (j in 1:i) {  
   
 if (j == 1) prob\_data[[paste0("parent",i)]] <- sub("/[^/]+$", "", prob\_data$pathString)  
 else if (j > 1) prob\_data[[paste0("parent",i)]] <- sub("/[^/]+$", "", prob\_data[[paste0("parent",i)]])  
 }  
   
 prob\_data <- prob\_data %>% left\_join(parent\_lookup, by = paste0("parent",i))  
   
}  
  
  
prob\_data$overall\_prob <- apply(prob\_data %>% select(contains("prob")) , 1, prod, na.rm = T) # calculate cumulative probability   
  
terminal\_data <- prob\_data %>% filter(tree\_level == max\_tree\_level) %>% # create new rows that will display terminal/final step calulcations on the tree  
 mutate(node\_type = 'terminal',  
 pathString = paste0(pathString, "/overall"),  
 prob = NA,  
 tree\_level = max\_tree\_level + 1)  
  
start\_node <- "Condiciones de cultivo" # name the root node  
  
prob\_data = bind\_rows(prob\_data, terminal\_data) %>% # bind everything together   
 mutate(pathString = paste0(start\_node,"/",pathString),  
 overall\_prob = ifelse(node\_type == 'terminal', overall\_prob, NA),  
 prob\_rank = rank(-overall\_prob, ties.method = "min", na.last = "keep"))  
  
prob\_data = bind\_rows(prob\_data, data.frame(pathString = start\_node, node\_type = 'start', tree\_level = 0)) %>% # add one new row to serve as the start node label  
 select(-contains("parent"))  
  
make\_my\_tree <- function(mydf, display\_level = NULL, show\_rank = FALSE, direction = "LR") {  
   
 if (!is.null(display\_level) ) {  
 mydf <- mydf %>% filter(tree\_level <= display\_level)  
   
 }  
   
 mytree <- as.Node(mydf)   
   
 GetEdgeLabel <- function(node) switch(node$node\_type, node$prob)  
   
 GetNodeShape <- function(node) switch(node$node\_type, start = "box", node\_decision = "circle", terminal = "none")  
   
   
 GetNodeLabel <- function(node) switch(node$node\_type,   
 terminal = ifelse(show\_rank == TRUE, paste0("Prob: ", node$overall\_prob,"\nRank: ", node$prob\_rank),  
 paste0("Prob: ", node$overall\_prob)),  
 node$node\_name)  
   
 SetEdgeStyle(mytree, fontname = 'helvetica', label = GetEdgeLabel)  
   
 SetNodeStyle(mytree, fontname = 'helvetica', label = GetNodeLabel, shape = GetNodeShape)  
   
 SetGraphStyle(mytree, rankdir = direction)   
   
 plot(mytree)  
   
}  
  
  
make\_my\_tree(prob\_data)



Arbol de probabilidades para el punto 1

1. Qué porcentaje de los que encuentran la presencia de daños por patógenos alta utiliza Metil Tiofanato 70%.

Primero encontramos la probabilidad de encontrar la presencia de daños por patógenos alta:

sum(prob\_data$overall\_prob[c(which(grepl('(?=.\*/A/overall)',prob\_data$pathString,perl =T)))])

## [1] 0.45

Luego, calculamos la probabilidad de encontrar daño por patogenos alta cuando se usa Metil Tiofanato 70%

sum(prob\_data$overall\_prob[c(which(grepl('(?=.\*/Me/A/overall)',prob\_data$pathString,perl =T)))])

## [1] 0.192

Luego, el cociente:

0.192/0.45

## [1] 0.4266667

La probabilidad de encontrar la presencia de daños por patógenos alta al usar metil tiofanato 70% es del 42.7%.

1. Cuál es la probabilidad de que al tomar una persona al azar que utiliza Belanaxil 8%, esta persona utilice Guano y encuentre la presencia de daños por patógenos baja.

Primero calculamos el total de los que usan Belanaxil 8%

sum(prob\_data$overall\_prob[c(which(grepl('(?=.\*/Be/A/overall|/Be/B/overall)',prob\_data$pathString,perl =T)))])

## [1] 0.35

Y luego el porcentaje de los que usan belanaxil 8%, guano, y encuentran la presencia por daños de patogenos baja:

sum(prob\_data$overall\_prob[c(which(grepl('(?=.\*/Guano/Be/B/overall)',prob\_data$pathString,perl =T)))])

## [1] 0.048

Luego, calculamos el cociente:

0.048/0.35

## [1] 0.1371429

Cuál es la probabilidad de que al tomar una persona al azar que utiliza Belanaxil 8%, esta persona utilice Guano y encuentre la presencia de daños por patógenos baja.

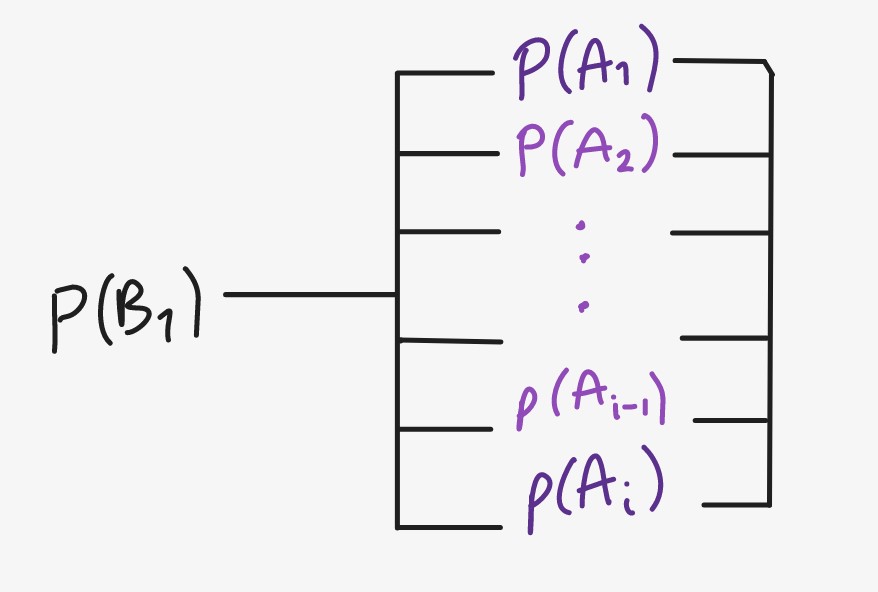
# Segundo punto

1. Una forma de incrementar la probabilidad de rendimiento de un cultivo (conocida como la confiabilidad del sistema de riego) es mediante un riego eficiente incrementando el número de asperjadores del sistema de riego.

Supóngase que un agricultor desea una probabilidad no menor de 0.99999 de que el sistema de riego funcione adecuadamente con un rendimiento por encima del promedio.

¿Cuántos asperjadores deben configurarse en paralelo, para alcanzar esta confiabilidad de operación, si se sabe que la probabilidad de que uno cualquiera de los asperjadores funcione adecuadamente es de 0.95? (supóngase que los asperjadores funcionan independiente entre sí).

* Definimos la probabilidad que se debe hallar como P(B) B: el conjunto de aspersores funciona adecuadamente con un rendimiento por encima del promedio.



Entonces,

utilizando la ley de deMorgan:

Ya que los eventos son independientes:

restamos 1 a ambos lados

multiplicamos por -1 a ambos lados:

usamos logaritmo para encontrar la potencia k:

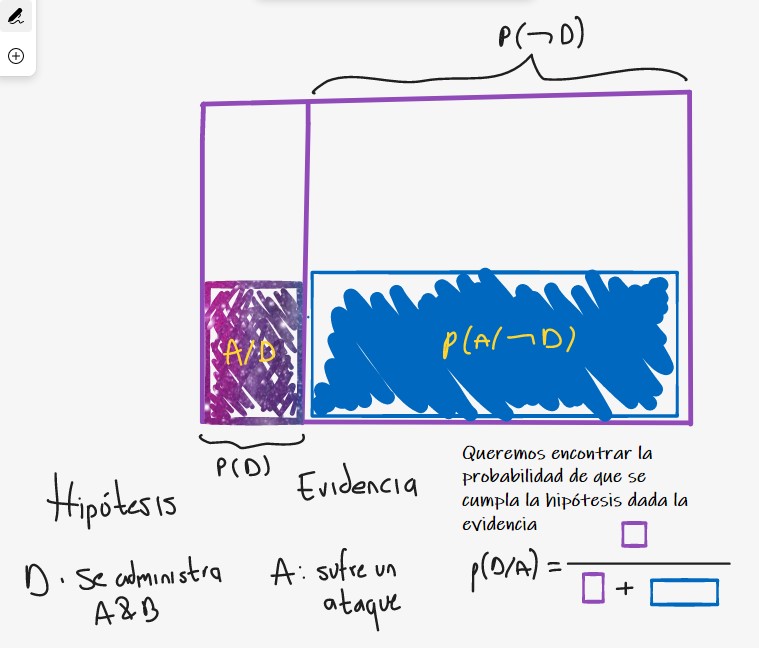
Es posible utilizar 4 asperjadores para asegurar una probabilidad de que todos funcionen por encima del promedio mayor a 0.99999.

# Tercer punto

1. Un médico veterinario ha decidido recetar dos nuevos medicamentos a 100 animales enfermos del corazón y lo hace en la forma siguiente: 35 reciben el medicamento A, 35 reciben el medicamento B y 30 ambos medicamentos. El fármaco A reduce la probabilidad de ataque cardiaco en 25%, el B lo hace en un 30% y los dos medicamentos,si se toman juntos, funcionan de manera independiente. Se escogió a los 100 animales de forma que cada uno tuviera 70% de probabilidad de sufrir un ataque cardiaco. Si un animal seleccionado aleatoriamente sufre un ataque cardiaco.¿ Cuál es la probabilidad de que le haya administrado los dos medicamentos?

A: Sufrir un ataque cardiaco B: Recibe el medicamento A C: Recibe el medicamento B D: recibe A & B

Nuestras probabilidades a priori son:



Ahora debemos hallar las probabilidades condionales, tenemos todas excepto la probabilidad de ver la evidencia dado que se cumpla la hipotesis de que tomo ambos medicamentos.

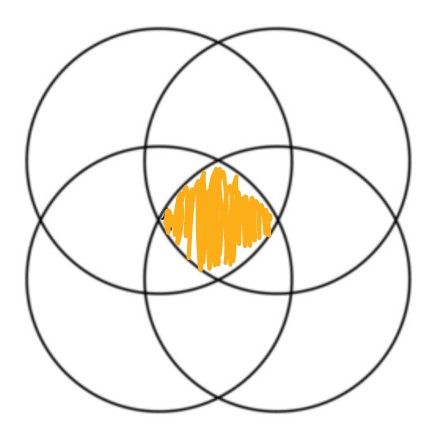
de los perros que sufrieron un infarto la probabilidad de haber recibido los dos medicamentos es 0.24

# Cuarto punto

1. Se instalaron cuatro anuncios panorámicos en la carretera a la entrada de la ciudad para hacerle publicidad a un nuevo fertilizante para los cultivos y se sabe por experiencia la probabilidad de que cada anuncio sea visto por un conductor escogido aleatoriamente. La probabilidad de que el primer anuncio sea visto por un conductor es de 075, la probabilidad de que el segundo sea visto es de 0.82, la probabilidad para el tercero es de 0.87 y la del cuarto es de 0.9. Si el evento consiste en que el conductor vea uno cualquiera de los anuncios es independiente de si ha visto o no los demás. ¿Cuál es la probabilidad de que?

PA<-0.75  
PB<-0.82  
PC<-0.87  
PD<-0.9

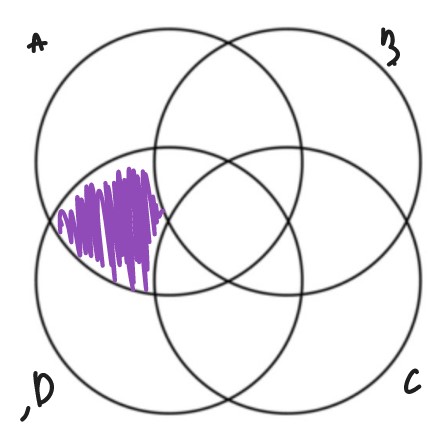
1. Los cuatro anuncios sean vistos por un conductor escogido aleatoriamente?



PA\*PB\*PC\*PD

## [1] 0.481545

* La probabilidad de que los cuatro anuncios sean vistos por un conductor escogido aleatoriamente es de 0.481

1. El primero y cuarto anuncios sean vistos, sin que el segundo y tercero sean notados? 
2. Exactamente uno de los anuncios sea visto?

PA+PB+PC+PD-(PA\*PB)-(PA\*PC)-(PA\*PD)-(PB\*PC)-(PB\*PD)-(PC\*PD)+(PA\*PB\*PC)+(PA\*PB\*PD)+(PB\*PC\*PD)+(PA\*PC\*PD)-(PA\*PB\*PC\*PD)-(PA\*PB\*PC\*PD)

## [1] 0.51787

* La probabilidad de que exactamente uno de los anuncios sea visto es de 0.518

1. Ninguno de los anuncios sea visto?

1-(PA+PB+PC+PD-(PA\*PB)-(PA\*PC)-(PA\*PD)-(PB\*PC)-(PB\*PD)-(PC\*PD)+(PA\*PB\*PC)+(PA\*PB\*PD)+(PB\*PC\*PD)+(PA\*PC\*PD)-(PA\*PB\*PC\*PD))

## [1] 0.000585

* La probabilidad de que ninguno de los avisos sea visto es de 0.00058

1. El tercero y cuarto anuncios no sean vistos?

1-(PC+PD-(PC\*PD))

## [1] 0.013

* La probabilidad de que el tercero y cuarto anuncios no sean vistos es de 0.987

# Quinto punto

1. Corpoica acaba de ganar una licitación para tres trabajos similares. La experiencia indica que la probabilidad de que un determinado trabajo se demore más de lo planeado es 0.4. Si el ritmo al cual se trabaja es independiente.

PB1<-0.4  
PB2<-0.4  
PB3<-0.4

1. ¿Cuál es la probabilidad de que los tres trabajos se terminen en el tiempo planeado?

(1-PB1)\*(1-PB2)\*(1-PB3)

## [1] 0.216

* La probabilidad de de que los tres trabajos se terminen en el tiempo planeado es de 0.72

b .¿Cuál es la probabilidad de que los tres trabajos tengan retrazo?

PB1\*PB2\*PB3

## [1] 0.064

* La probabilidad de de que los tres trabajos tengan retraso es de 0.064

1. ¿Cuál es la probabilidad de que al menos uno tenga retrazo?

PB1+PB2+PB3-(PB1\*PB2)-(PB1\*PB3)-(PB2\*PB3)+(PB1\*PB2\*PB3)

## [1] 0.784

* La probabilidad de de que al menos uno tenga retrazo es de 0.784

# Sexto punto

1. La probabilidad de que una planta de arveja de la variedad 1 sobreviva a la inoculación es de 0.4 y la probabilidad de que una planta de arveja de la variedad 2 sobreviva a la inoculación es de 0.5. La probabilidad de que una planta de arveja de la variedad 1 sobreviva dado que la planta de arveja de la variedad 2 sobrevive es de 0.7. Encuentre la probabilidad de que:

A: Una planta de arveja de la variedad 1 sobreviva a la inoculación

B: Una planta de arveja de la variedad 2 sobreviva a la inoculación

a. Las dos variedades de arveja sobrevivan.

0.7\*0.5

## [1] 0.35

* La probabilidad de que las dos variedades de arveja sobrevivan es de 0.35

1. Si la variedad 1 sobrevive la variedad 2 sobreviva.

0.35/0.4

## [1] 0.875

* La probabilidad de que si la variedad 1 sobrevive la variedad 2 sobreviva es de 0.875

1. Al menos una variedad de las dos sobreviva.

0.5+0.4-0.35

## [1] 0.55

* La probabilidad de que al menos una variedad de las dos sobreviva es de 0.55

# Septimo punto

1. Una finca contiene 20 tipos de insectos y contiene 5 que son dañinos para los cultivos. Se toma 2 insectos al azar sin reemplazo.

Primero se calcula los casos posibles

choose(20,2)

## [1] 190

Hay 190 casos posibles

1. Cuál es la probabilidad de que el segundo insecto sea dañino para el cultivo si el primero lo fue.

(5/20)\*(4/19)

## [1] 0.05263158

La probabilidad de que el primero sea dañino es de 5/20 y de que el segundo escogido sea dañino ya que el primero tambien lo es es de 5/20\*4/19 = 0.0526

1. Cuál es la probabilidad de que los dos insectos sean dañinos para el cultivo?

choose(5,2)/190

## [1] 0.05263158

* La probabilidad de que los dos insectos sean dañinos para el cultivo es de 0.053

1. Cuál es la probabilidad de que ambos insectos no sean dañinos para los cultivos?

choose(15,2)/190

## [1] 0.5526316

* La probabilidad de que ambos insectos no sean dañinos para los cultivos es de 0.553

# Octavo punto

1. La obtención de imágenes por resonancia magnética (MRI, por sus siglas en inglés) es una prueba aceptada no invasiva para evaluar los cambios en el cartílago de las articulaciones. En un estudio para comparar los resultados de la evaluación de MRI con la evaluación quirúrgica artroscópica de desgarres del cartílago en dos sitios de las rodillas de 35 pacientes se obtuvieron las siguientes clasificaciones de los 2x35=70 exámenes. Los desgarres reales se confirmaron mediante el examen quirúrgico artroscópico.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Desgarres | Sin Desgarres | Total |
| MRI postiva | 27 | 0 | 27 |
| MRI negativa | 4 | 39 | 43 |
| Total | 31 | 39 | 70 |

1. Cuál es la probabilidad de que un sitio seleccionado al azar tenga un desgarre y haya sido identificado como tal mediante MRI?

27/70

## [1] 0.3857143

* La probabilidad de que un sitio seleccionado al azar tenga un desgarre y haya sido identificado como tal mediante MRI es de 0.386

1. Cuál es la probabilidad de que un sitio elegido al azar no tenga ningún desgarre y haya sido identificado como un desgarre?

0/70

## [1] 0

* La probabilidad de que un sitio elegido al azar no tenga ningún desgarre y haya sido identificado como un desgarre es 0

1. Cuál es la probabilidad de que un sitio seleccionado al azar tenga un desgarre y no haya sido identificado por MRI?

4/70

## [1] 0.05714286

* La probabilidad de que un sitio seleccionado al azar tenga un desgarre y no haya sido identificado por MRI es de 0.057

1. Si hay un desgarre, cuál es la probabilidad de una MRI positiva?

27/31

## [1] 0.8709677

* Si hay un desgarre,la probabilidad de una MRI positiva es de 0.871

1. Si hay un desgarre, cuál es la probabilidad de un falso negativo; es decir, una MRI negativa?

4/31

## [1] 0.1290323

* Si hay un desgarre,la probabilidad de una MRI negativa es de 0.129

1. Cuál es la probabilidad de que no se presente desgarre o haya una MRI negativa?

(39/70)+(43/70)-(39/70)

## [1] 0.6142857

* La probabilidad de que no se presente desgarre o haya una MRI negativa es de 0.614

# Noveno punto

1. Se tienen 4 variedades resistentes a una enfermedad, 4 moderadamente susceptibles a la misma enfermedad y 3 variedades completamente susceptibles a tal enfermedad. De estas 11 variedades se desean formar grupos de 4 variedades. De cuántas maneras se pueden escoger:
2. Dos variedades resistentes y dos moderadamente susceptibles.

choose(4,2)\* choose(4,2)

## [1] 36

* Se pueden escoger 36 formas de dos variedades resistentes y dos moderadamente susceptibles

1. Tres resistentes y una completamente susceptible.

choose(4,3)\*choose(3,1)

## [1] 12

* Se pueden escoger 12 formas de tres resistentes y una completamente susceptible

1. Una variedad resistente, una moderadamente susceptible y dos completamente susceptibles.

choose(4,1)\*choose(4,1)\*choose(3,2)

## [1] 48

* Se pueden escoger 48 formas de Una variedad resistente, una moderadamente susceptible y dos completamente susceptibles

# Decimo punto

1. Suponga que le interesa seguir dos rasgos independientes de los guisantes; textura (S=lisa, s=rugosa) y color de la semilla (Y=amarilla, y=verde), en una cruza de segunda generación de padres heterocigotos. Recuerde que la letra mayúscula representa el rasgo dominante. Complete la tabla con los pares de genes para ambos rasgos. Todos los pares posibles son igualmente probables. Color de la semilla Textura de la semilla yy yY Yy YY ss (ss yy) (ss yY) sS Ss SS
2. Qué porcentaje de la descendencia de esta cruza tendrá guisantes amarillos lisos?

9/16\*100

## [1] 56.25

* Un 56.25% de la descendencia de esta cruza tendrá guisantes amarillos lisos

1. Qué proporción de la descendencia tendrá guisantes verdes lisos?

3/16

## [1] 0.1875

%yS = 3/16 = 0.1875 \* La proporcion de la descendencia que tendrá guisantes verdes lisos es de 0.1875

1. Qué proporción de la descendencia tendrá guisantes rugosos amarillos?

3/16

## [1] 0.1875

%Ys = 3/16 = 0.1875 \* La proporcion de la descendencia que tendrá guisantes rugosos amarillos es de 0.1875

1. Qué proporción de la descendencia tendrá guisantes verdes rugosos?

1/16

## [1] 0.0625

%ys = 1/16 = 0.0625 \* La proporcion de la descendencia que tendrá guisantes rugosos verdes es de 0.0625

1. Si una descendencia tiene guisantes amarillos lisos, Cuál es la probabilidad de que lleve un alelo s?

6/9

## [1] 0.6666667

* Si una descendencia tiene guisantes amarillos lisos la probabilidad de que lleve un alelo s es de 0.66

Un alelo s y uno y?

4/9

## [1] 0.4444444

* Si una descendencia tiene guisantes amarillos lisos la probabilidad de que lleve un alelo s y uno y es de 0.44