**Vargas Carrillo Raul Paulino**

**Manejo de datos en R con dplyr**

Obtener la información utilizando la librería dplyr en R deberá adjuntar tanto el código utilizado como los resultados obtenidos.

La base de datos “Frijol.xlsx”, utilizada para este estudio, contiene información sobre el uso de tecnología y servicios en el sector agrícola. Estos datos han sido obtenidos del Sistema de Información Agropecuaria y Pesquera (SIAP) de la SAGARPA, y corresponden a los años agrícolas 2017, 2018 y 2019. Las variables de estudio incluidas son las siguientes:

Cultivo: Tipo de cultivo (Frijol)

Variedad: Variedad del cultivo.

Año: Año agrícola (2015,2018 y2019)

Estado

Municipio

Modalidad: Riego (Requieren el suministro de importantes cantidades de agua a través de diversos métodos artificiales de riego) o Temporal (Depende del comportamiento de las lluvias durante el ciclo de producción y de la capacidad del suelo para captar el agua y conservar la humedad).

SupSembrada: Superficie en la cual ya se depositó la semilla de cualquier cultivo, previa preparación del suelo (ha).

SupCosechada: Es el área de la cual se obtuvo producción (ha).

SupFertilizada: Superficie en la que se aplica, al menos una vez, algún químico para el mejoramiento del desarrollo del cultivo (ha).

SupMecanizada: Superficie en la que se utiliza, al menos una vez, algún tipo de maquinaria y equipo en la siembra y cosecha de la superficie cultivada (ha).

SupSanidadVegetal: Superficie sembrada en la cual se realizan actividades orientadas a la prevención, control y erradicación de plagas o enfermedades que afectan los cultivos(ha).

SupSemillaMejorada: Superficie sembrada con semilla obtenida de una modificación genética natural y que no lleva ninguna alteración transgénica (ha).

SupSemillaCriolla: Área sembrada en la cual se utiliza semilla recolectada o comprada de cosechas anteriores (ha).

ProdObt: Es la cantidad de producto que se levantó en determinada superficie cosechada (ton).

Utilizando la librería dplyr obtenga la siguiente información:

1. Obtener una tabla con las variables: estado, municipio, modalidad y variedad donde la modalidad de riego es por temporal para el municipio Acatzingo.

> Frijol %>%

+ filter(Modalidad == 'Temporal' & NomMunicipio == 'Acatzingo') %>%

+ select(NomEstado, NomMunicipio, Modalidad, Variedad)

# A tibble: 9 × 4

NomEstado NomMunicipio Modalidad Variedad

*<chr>* *<chr>* *<chr>* *<chr>*

1 Puebla Acatzingo Temporal Frijol ayocote

2 Puebla Acatzingo Temporal Frijol ayocote

3 Puebla Acatzingo Temporal Frijol ayocote

4 Puebla Acatzingo Temporal Frijol flor de mayo

5 Puebla Acatzingo Temporal Frijol flor de mayo

6 Puebla Acatzingo Temporal Frijol flor de mayo

7 Puebla Acatzingo Temporal Frijol negro jamapa

8 Puebla Acatzingo Temporal Frijol negro jamapa

9 Puebla Acatzingo Temporal Frijol negro jamapa

1. Obtener la tabla con las variables: Nombre del estado, Superficie sembrada, variedad, año agrícola y modalidad de riego donde la superficie sembrada es mayor a 20000 hectáreas en el estado de Sinaloa.

> Frijol %>%

+ filter(SuperficieSembrada > 20000 & NomEstado == 'Sinaloa') %>%

+ select(NomEstado, SuperficieSembrada, Variedad, AñoAgricola, Modalidad)

# A tibble: 4 × 5

NomEstado SuperficieSembrada Variedad AñoAgricola Modalidad

*<chr>* *<dbl>* *<chr>* *<dbl>* *<chr>*

1 Sinaloa 29911. Frijol azufrado 2017 Riego general

2 Sinaloa 23063. Frijol azufrado 2018 Riego gravedad total

3 Sinaloa 31391. Frijol azufrado 2018 Riego gravedad total

4 Sinaloa 24905. Frijol azufrado 2019 Riego gravedad total

1. Obtener el valor máximo y mínimo de la superficie cosechada para la variedad “Frijol bayo”.

> Frijol %>%

+ filter( Variedad== 'Frijol bayo') %>%

+ summarise(max(SuperficieCosechada), min(SuperficieCosechada))

# A tibble: 1 × 2

`max(SuperficieCosechada)` `min(SuperficieCosechada)`

*<dbl>* *<dbl>*

1 8505 2

1. Obtener una tabla con la variables: superficie sembrada, superficie cosechada y variedad donde la superficie sembrada sea diferente a la superficie cosechada.

> Frijol %>%

+ filter(SuperficieSembrada != SuperficieCosechada) %>%

+ select(SuperficieSembrada, SuperficieCosechada, Variedad)

# A tibble: 741 × 3

SuperficieSembrada SuperficieCosechada Variedad

*<dbl>* *<dbl>* *<chr>*

1 92 42 Frijol ayocote

2 438 419. Frijol azufrado

3 150 113 Frijol azufrado

4 2847 2782 Frijol azufrado

5 725 636. Frijol azufrado

6 986 873. Frijol azufrado

7 4625 4223 Frijol azufrado

8 1565 1560 Frijol azufrado

9 3122 3117 Frijol azufrado

10 3651 3634 Frijol azufrado

# ℹ 731 more rows

1. Obtener el promedio de la superficie sembrada por estado.

> Frijol %>%

+ group\_by(NomEstado) %>%

+ summarise(mean(SuperficieSembrada))

# A tibble: 32 × 2

NomEstado `mean(SuperficieSembrada)`

*<chr>* *<dbl>*

1 Aguascalientes 411.

2 Baja California 5

3 Baja California Sur 219.

4 Campeche 257.

5 Chiapas 942.

6 Chihuahua 1185.

7 Ciudad de México 11.4

8 Coahuila 104.

9 Colima 4.08

10 Durango 3083.

# ℹ 22 more rows

1. Obtener máximo y promedio para cada una de las variedades de frijol.
2. Utilizando la función “mutate” crear una nueva variable con Rendimiento Obtenido (Productividad obtenida dividida por la superficie cosechada Ton/ha) para cada una de las observaciones del conjunto de datos.

> frijol <- Frijol %>%

+ mutate(RenObt = ProducciónObtenida/SuperficieCosechada)

> frijol$RenObt

[1] 0.9698462 0.9199511 0.6800000 0.9803063 0.9396454 0.3929808 0.9910591 0.9501686

[9] 0.4700000 0.8400000 0.6800000 0.7000000 0.8000000 0.7600000 0.7600000 0.6600000

[17] 0.8500000 0.8900000 0.8800000 0.7000000 0.9000000 0.4200000 0.7890000 0.8162983

[25] 0.8200000 0.4130435 0.8026316 0.2900000 0.8030303 0.8036364 0.8151282 0.8103333

[33] 0.8900000 1.0000000 0.6500000 0.9700000 1.0200000 1.0900000 0.5100000 1.1800000

[41] 0.4700000 1.0500000 1.0400000 0.9000000 1.0900000 0.9800000 1.0199639 0.2641860

[49] 0.3428571 0.3200000 0.3173554 1.4098361 1.1300643 1.2101367 1.1100000 1.2799862

[57] 1.1198971 1.3100858 0.7799228 1.3600000 1.1200000 1.1400000 0.8000000 1.9000000

[65] 1.7000000 1.5400000 0.6800000 1.2200000 1.1200000 1.2500000 1.2500000 0.8900000

[73] 1.1800000 0.9100000 1.1800000 1.1300071 1.3000000 1.2000031 0.7980000 1.0000000

[81] 0.9000000 0.8200000 1.2499971 1.1000000 0.8353846 1.9265469 1.6808496 1.9681667

[89] 0.6000000 1.5946829 1.7501075 2.1100000 0.8500000 2.0410549 1.9700818 2.0000000

[97] 1.9600000 0.9000000 1.7307674 1.5000000 2.0243162 1.5000000 1.0000000 1.1500000

[105] 1.1000000 0.9700000 0.4000000 0.4000000 0.4000000 0.4201058 0.4299020 0.3900000

[113] 0.4100000 0.3800000 1.4221176 1.1694461 1.2783266 1.0840000 1.2907179 1.1862410

[121] 1.3392330 0.7934401 1.2000000 1.3000000 1.2500000 1.0500000 0.9800000 1.4375000

[129] 0.8400000 1.1100000 0.9800000 1.0600000 0.9400000 0.7400000 0.9100000 0.7500000

[137] 0.9829874 0.5259834 1.6000000 0.8250032 0.4789973 1.1000000 1.2500000 0.8815385

[145] 0.7885714 0.8800000 0.6500000 0.9100000 1.7599530 1.5093480 1.8000000 0.1800000

[153] 0.8000000 1.7996289 1.7464200 1.1000000 1.6000000 1.2000000 1.9393630 1.6289416

…

1. Calcula el número de observaciones en la variable “Superficie Con Sanidad Vegetal” que tienen valores NA (Not Available). (Utiliza la función table con la opción useNA = c(”always”))

> table(Frijol$SuperficieConSanidadVegetal, useNA = "always")

…

…

34000 35952 45000 47381 56419 58500 77361 92000 **<NA>**

1 1 1 1 1 1 1 1 **8**

1. Por el método de imputación de sustitución por media remplaza los valores “NA” de la variable “Superficie Con Sanidad Vegetal”.

media <- mean(Frijol$SuperficieConSanidadVegetal, na.rm = T)

Frijol$SuperficieConSanidadVegetal <- replace\_na(Frijol$SuperficieConSanidadVegetal, media)

mean(Frijol$SuperficieConSanidadVegetal)

> media <- mean(Frijol$SuperficieConSanidadVegetal, na.rm = T)

> media

[1] 196.9693

> Frijol$SuperficieConSanidadVegetal <- replace\_na(Frijol$SuperficieConSanidadVegetal, media)

> mean(Frijol$SuperficieConSanidadVegetal)

[1] 196.9693