

Anexo 1. Ejemplo del trabajo de la actividad 2: Estadística descriptiva para variable cuantitativas

Estudiante de biología.

Actividades 2. Estadística descriptiva para variable cuantitativas

Estudiante de Biología
Área Biológica-Agropecuaria

Hipótesis de investigación

La longitud del sépalo y del pétalo de las tres especies de la flor iris (*setosa*, *virginica* y *versicolor*) se relaciona con la anchura del sépalo y del pétalo de estas mismas.
La relación que existe entre largo del pétalo y ancho del pétalo, así como la relación entre largo de sépalo y ancho del sépalo será mayor que la que exista entre otra combinación de variables.

Objetivo

El objetivo en términos estadísticos es relacionar el largo del sépalo, así como el largo del pétalo con el ancho del sépalo y del pétalo en cada una de las especies de flor iris (*setosa*, *virginica* y *versicolor*).

Metodología

Se cuenta con una tabla de datos llamada Iris, obtenida de manera secundaria, con 150 casos (individuos muestreados: 50 organismos de la especie *setosa*, 50 de la especie *virginica* y 50 de *versicolor*) y 5 variables a considerar (especie, largo del sépalo, ancho del sépalo, largo del pétalo y ancho del pétalo). La primera variable, especie, es categórica o cualitativa en escala de medición de tipo nominal, mientras que las últimas 4 corresponden a variables de tipo cuantitativo o numérico, son continuas y su escala de medición es de razón. Los cálculos estadísticos aplicados a mi tabla de datos fueron: media, mediana, desviación estándar y rango intercuartílico para cada una de las variables, exceptuando a la variable categórica "especie". También se determinó el valor de correlación de Pearson entre las variables cuantitativas. Dichos cálculos, así como los gráficos para visualizar los datos, se realizaron mediante el uso de la herramienta estadística Rstudio Ver. 2021.09.1+372

Resultados

Medidas de tendencia central y medidas de dispersión

Largo del sépalo

La media o el promedio obtenido para la variable largo del sépalo es de 5.84 cm y la mediana de 5.8 cm; ambos valores son muy parecidos por lo que no se presenta un sesgo significativo, sin embargo, en la Figura 1 se puede apreciar que este sesgo está ligeramente inclinado hacia la derecha. La variable corre en un rango de 4.3 cm que es el límite inferior hasta 7.9 cm como límite superior, mientras que la medida de la desviación estándar 0.82 nos indica una baja dispersión de los datos con respecto a la media.
En la Figura 1 se observa una distribución en forma de campana un tanto asimétrica. Vemos que la mayor concentración de datos está situada al centro entre 5.5 cm y 6.5 cm, siendo los límites inferior 4.3 cm y superior 7.9 cm los que presentan una menor acumulación.

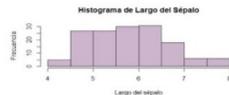


Figura 1. Largo del sépalo en individuos de las especies *setosa*, *virginica* y *versicolor*. Datos obtenidos de: Conjunto de datos flor iris.

Ancho del sépalo

Para el ancho del sépalo se presentan valores atípicos por lo que se considera a la mediana como medida de tendencia central, con un valor de 3 cm. En la Figura 2 se observa una muy ligera

dispersión hacia los valores más altos y una distribución casi simétrica. Se señala en el diagrama que el rango intercuartílico se encuentra entre el cuartil 1 (Q1) 2.8 cm y el cuartil 3 (Q3) 3.3 cm. Por debajo del mínimo o límite inferior existe un dato atípico con un valor de 2 cm, mientras que por arriba del máximo se presentan tres datos atípicos. La máxima concentración se encuentra entre los valores 3 cm y 4.5 cm Ubicados en el tercer cuartil.

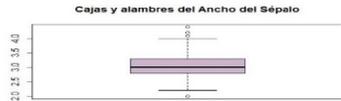


Figura 2. Ancho del sépalo en individuos de las especies *setosa*, *virginica* y *versicolor*. Datos obtenidos de: Conjunto de datos flor iris.

Largo del pétalo

La medida de la mediana para el largo del pétalo es de 4.35 cm y la media de 3.75 cm, lo que indica un sesgo. Podemos ver, de manera muy interesante que los datos se encuentran divididos en dos poblaciones. La Figura 3 muestra que en la segunda población se presenta una distribución un tanto asimétrica, con un sesgo que se inclina ligeramente a la derecha, es decir, es un sesgo positivo. La mayor frecuencia de los datos se acumula entre los valores 4 cm y 5 cm.

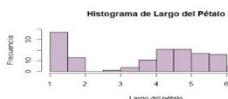


Figura 3. Largo del pétalo en individuos de las especies *setosa*, *virginica* y *versicolor*. Datos obtenidos de: Conjunto de datos flor iris.

Ancho del pétalo

En la Figura 4, referente al ancho del pétalo, el valor de la mediana es de 1.3 cm. La distribución es totalmente asimétrica, con una dispersión, como en los casos anteriores, hacia los valores mayores, en este caso superior a 1.8 cm que es el valor del tercer cuartil (Q3).



Figura 4. Ancho del pétalo en individuos de las especies *setosa*, *virginica* y *versicolor*. Datos obtenidos de: Conjunto de datos flor iris.

Ancho del sépalo para cada especie (*setosa*, *virginica* y *versicolor*)

En la Figura 5 se presentan a las tres especies comportándose de manera distinta a la una de las otras. Por ejemplo, vemos que tanto *setosa* como *virginica* presentan datos atípicos por encima y por debajo de sus límites, y si bien, *setosa* tiene el ancho del sépalo mayor a las otras dos especies también presenta un dispersión débil hacia valores altos, mientras que *versicolor*, por otro lado, presenta un sesgo hacia valores inferiores a la mediana.

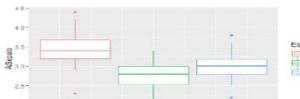


Figura 5. Ancho del sépalo en individuos separados por especies: *setosa*, *virginica* y *versicolor*. Datos obtenidos de: Conjunto de datos flor iris.

Tabla 1. Matriz de correlación de Pearson. Datos obtenidos de: Conjunto de datos flor iris.

	Largo sépalo	Ancho sépalo	Largo pétalo	Ancho pétalo
Largo sépalo	1.00	-0.11	0.87	0.81
Ancho sépalo	-0.11	1.00	-0.42	-0.36
Largo pétalo	0.87	-0.42	1.00	0.86
Ancho pétalo	0.81	-0.36	0.86	1.00

De acuerdo con la tabla anterior podemos interpretar que casi para todas las correlaciones hay un valor bastante significativo, a excepción del resultado conocido entre "Largo del sépalo" y "Ancho del sépalo" cuyo valor fue de -0.11, muy cercano al 0, lo que indicaría ausencia de relación lineal entre las variables. En la Figura 6 se presenta el diagrama de dispersión para esta correlación, vemos que los datos se encuentran bastante dispersos y aparentemente no siguen un patrón, pero si observamos a la especie *setosa* (color rojo) esta presenta lo que podría considerarse una relación positiva entre las variables. Contrastado con lo mencionado, el valor de correlación más alto fue de 0.96 entre el ancho del pétalo y el largo del pétalo, indicando así una correlación positiva muy alta. Es decir, que mientras el ancho del pétalo aumenta el largo del pétalo también. Hay que resaltar que nuestros datos se dividen en dos poblaciones, donde *setosa* es la especie con pétalos más cortos y delgados a comparación de *versicolor* y *virginica*, mostrándose un comportamiento bien diferenciado para cada especie (Fig. 7).



Figura 6. Diagrama de dispersión para las variables Ancho del sépalo y Largo del sépalo. Datos obtenidos de: Conjunto de datos flor iris.



Figura 7. Diagrama de dispersión para las variables Ancho del pétalo y Largo del pétalo. Datos obtenidos de: Conjunto de datos flor iris.

Aunque el valor de la correlación de Pearson entre el ancho del pétalo y el ancho del sépalo es significativo con un valor de -0.36 lo que indicaría una correlación negativa, al analizar el diagrama (Fig. 8) se presenta un caso completamente distinto. Si existiera una relación entre estas dos variables para las especies *versicolor* y *virginica*, dicha relación sería positiva, por la inclinación que presenta la línea imaginaria. Las especies *versicolor* y *virginica* presentan una relación positiva, aumentando el largo del pétalo al mismo tiempo que aumenta el largo del sépalo. El valor de la relación es de 0.87 y vemos que los datos se encuentran menos dispersos entre ellos (Fig. 9).

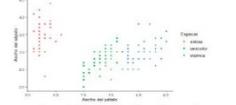


Figura 8. Diagrama de dispersión para las variables Ancho del pétalo y Ancho del sépalo. Datos obtenidos de: Conjunto de datos flor iris.

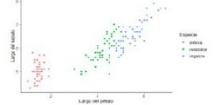


Figura 9. Diagrama de dispersión para las variables Largo del pétalo y Largo del sépalo. Datos obtenidos de: Conjunto de datos flor iris.

En las figuras 8 y 9, 10 y 11 *setosa* nuevamente se presenta como una población aislada y sin relación lineal entre sus variables. Finalmente, en la figura 10, cuyo valor de correlación fue de -0.42, de nuevo no se visualiza la relación inversa en el gráfico, además de que los datos se presentan muy dispersos, semejantes a la figura 11, pero en este caso sí se puede llegar a interpretar una relación positiva entre las variables ancho del pétalo y largo del sépalo, considerando un poco la inclinación de la línea y que el valor de correlación de Pearson fue de 0.81.

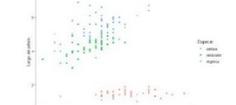


Figura 10. Diagrama de dispersión para las variables Ancho del sépalo y Largo del pétalo. Datos obtenidos de: Conjunto de datos flor iris.



Figura 11. Diagrama de dispersión para las variables Ancho del pétalo y Largo del sépalo. Datos obtenidos de: Conjunto de datos flor iris.

Referencias

Equipo RStudio (2021). Rstudio desarrollo integrado para R. Studio, PBC, Boston, MA, URL <http://www.rstudio.com/>