

**Análisis Estadístico Descriptivo: Fumadores, Correlación Entre Edad y Cigarros Fumados
al Día, Diferencias de Consumo Entre Hombres y Mujeres**

Armando Delfín, Eduardo Longaart, y Sonia Eveligret

Universidad Central De Venezuela

Departamento de Estadística y Probabilidad

Computación I

Jesús Ochoa

12/7/2024

Nota del Autor

Correspondence concerning this article should be addressed to Armando Delfín

Resumen

En este trabajo de investigacion se realiza un análisis descriptivo sobre una base de datos que contiene un numero de observaciones plasmadas en un registro con 3900 hombres y mujeres que son fumadores, a traves del uso de graficos y de estadisticas descriptivas, para obtener informacion y posteriormente hacer un estudio que nos permita conocer si existe una relacion entre la edad de las personas, su sexo y el numero de cigarros que consumen por dia. Lo que sea desea tener como resultado de este trabajo de investigacion es una estimacion basica enfocada a todos lo registros que contienen informacion de las personas que son fumadoras regulares, y de forma practica a traves del uso de graficos explicar las caracteristicas que posee una persona fumadora acorde a su sexo, edad y la cantidad de cigarros que fumen por dia

Palabras Claves: keyword1, keyword2, keyword3

Análisis Estadístico Descriptivo: Fumadores, Correlación Entre Edad y Cigarros Fumados al Día, Diferencias de Consumo Entre Hombres y Mujeres

¿De que manera se manejo la base de datos suministrada?

El analisis exploratorio de datos es una forma de analizar datos definidos por Jonh W. Tukey, el cual es el tratamiento estadistico al que se somete las muestras recogidas durante un proceso de investigacion en cualquier campo cientifico

Antes de manejar la base de datos, conviene analizar los datos que se utilizaran. Esto permite observar las características fundamentales de los mismos y comprender la estructura del conjunto de los datos, identificando la variable que se tiene como objetivo y explorar las posibles técnicas de modelado. (Tukey, J.W, 1977)

En este trabajo de investigacion se extrajo toda la informacion necesaria de la base de datos de la siguiente manera:

Lo primero que se llevo a cabo fue observar como se comportaban los datos y si dentro de la base de datos habian registros con valores nulos (En el caso de la columna que contiene a la variable “cigs_per_day” se encontraban 14 registros con valores nulos). Una vez sabiendo todo esto lo siguiente que se realizo fue la depuracion de la data, filtrando solo los registros que son relevantes los cuales son aquellos que contienen a las personas que son fumadoras.

Luego de hacer el proceso de depuracion, se estudiaron a las variables “sex” y “age” para determinar que tanto abarcaban estas dos variables dentro de la base de datos, ordenandolas de forma ascendente para hacer mas facil el manejo de dichas variables dentro de la base de datos.

Al momento de finalizar todo el proceso de filtrado, depuracion y ordenamiento de los registros se procedio a realizar graficos que facilitaran las interpretaciones de los datos y asi poder transmitir de forma clara todos los resultados obtenidos de la investigacion.

Análisis Descriptivo

A continuacion, en las siguientes tablas se muestran estadísticas descriptivas para las variables en estudio, Toda la informacion respecto a los datos a manejar fue extraida de <https://www.kaggle.com/datasets/jacepranter/smoker-health-data>, de la cual se estudia solo a los

1932 fumadores activos de entre los 3900 registros que incluyen ex-fumadores.

- En promedio, los fumadores consumen 18 cigarros diariamente, con una desviación estándar de 9 cigarros. Menos del 25% de ellos fuma más de 20 cigarros (un paquete) por día. La distribución es asimétrica positiva, y leptocúrtica.
- En promedio, los fumadores tienen 48 años de edad, con una desviación estándar de 8 años. El 75% de ellos tiene 53 años o menos. La distribución es asimétrica positiva, y leptocúrtica.
- De todas las personas en observación, 1105 son de sexo masculino, y 827 de sexo femenino. Representando el 57% y 43% respectivamente.

Tabla 1

Estadística Descriptiva Sobre las Variables Cuantitativas

Variable	Promedio	D. Estandar	Q1	Q2	Q3	C.S	Kurtosis
C.D	17.9	9.43	10	20	20	0.07	2.30
EDAD	47.7	7.97	41	46	53	0.51	2.40

Tabla 2

Frecuencias Relativas de los Fumadores Respecto a su Sexo

Sexo	Frecuencia		
	Absoluta	Frecuencia Relativa	Frecuencia Relativa Porcentual
M	1105	0.57	57%
F	827	0.43	43%
TOTAL	1932	1	1

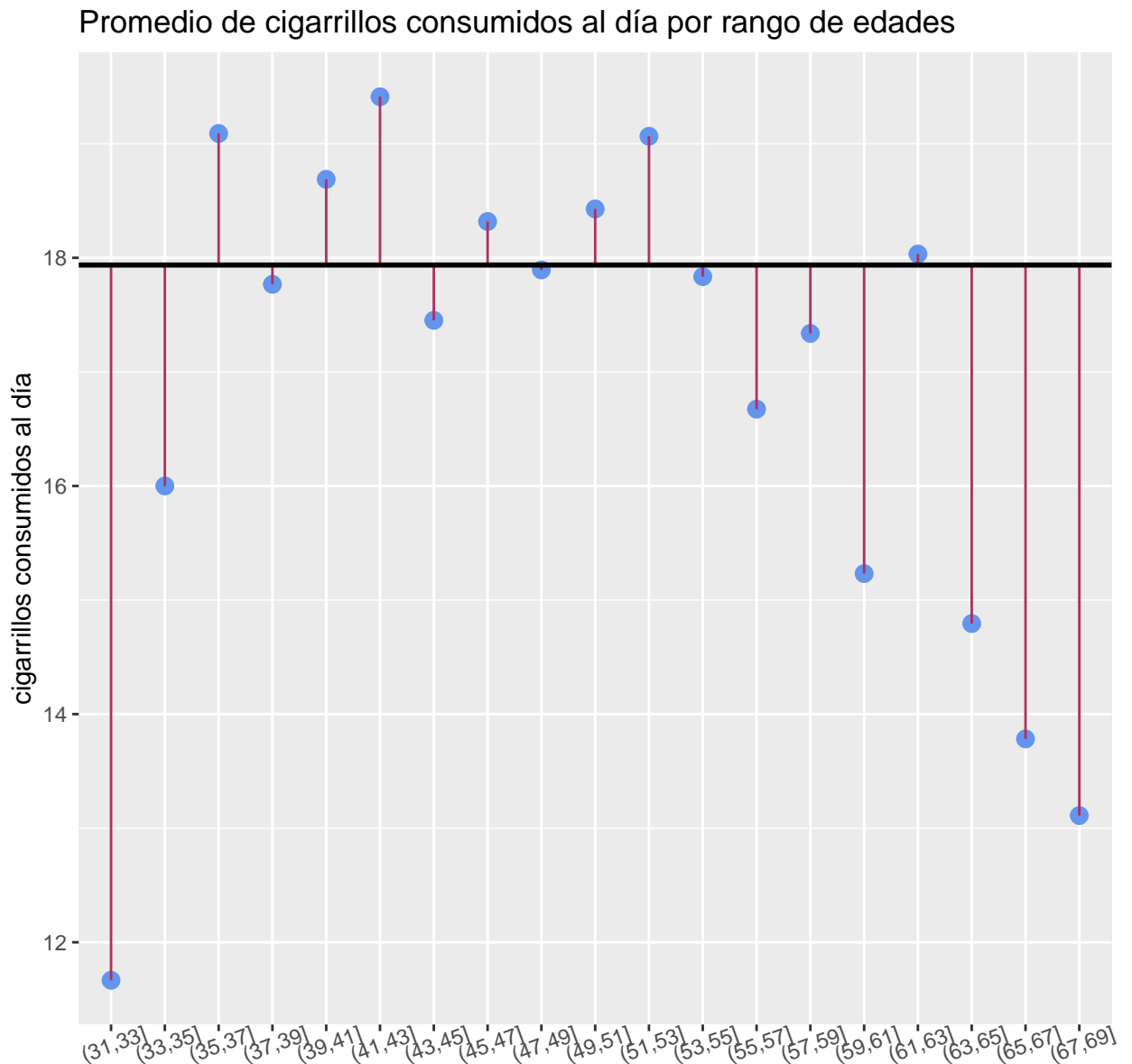
Análisis Bivariante: ¿Existe Relación Entre la Edad de los Fumadores y la Cantidad de Cigarros que Consumen al día?

A través del análisis descriptivo bivariante (coeficiente de correlación de Pearson) y el uso de gráficos se determinó que existe correlación inversa entre el número de cigarros que fuma una persona y su edad

Covarianza	Índice de Correlación
-5.30	-0.07

La covarianza entre las dos variables, edad, y cigarros consumidos al día, tienen un valor de covarianza negativo, lo que significa que tienen una relación negativa o inversa. A su vez, el índice de correlación, el cual nos indica la fuerza de la relación negativa antes descubierta es de -0.07. Teniendo en cuenta que una relación inversa puede llegar a un máximo de -1, esta correlación parece ser muy tenue.

Para observar estas dos variables gráficamente, nos referimos a la siguiente figura

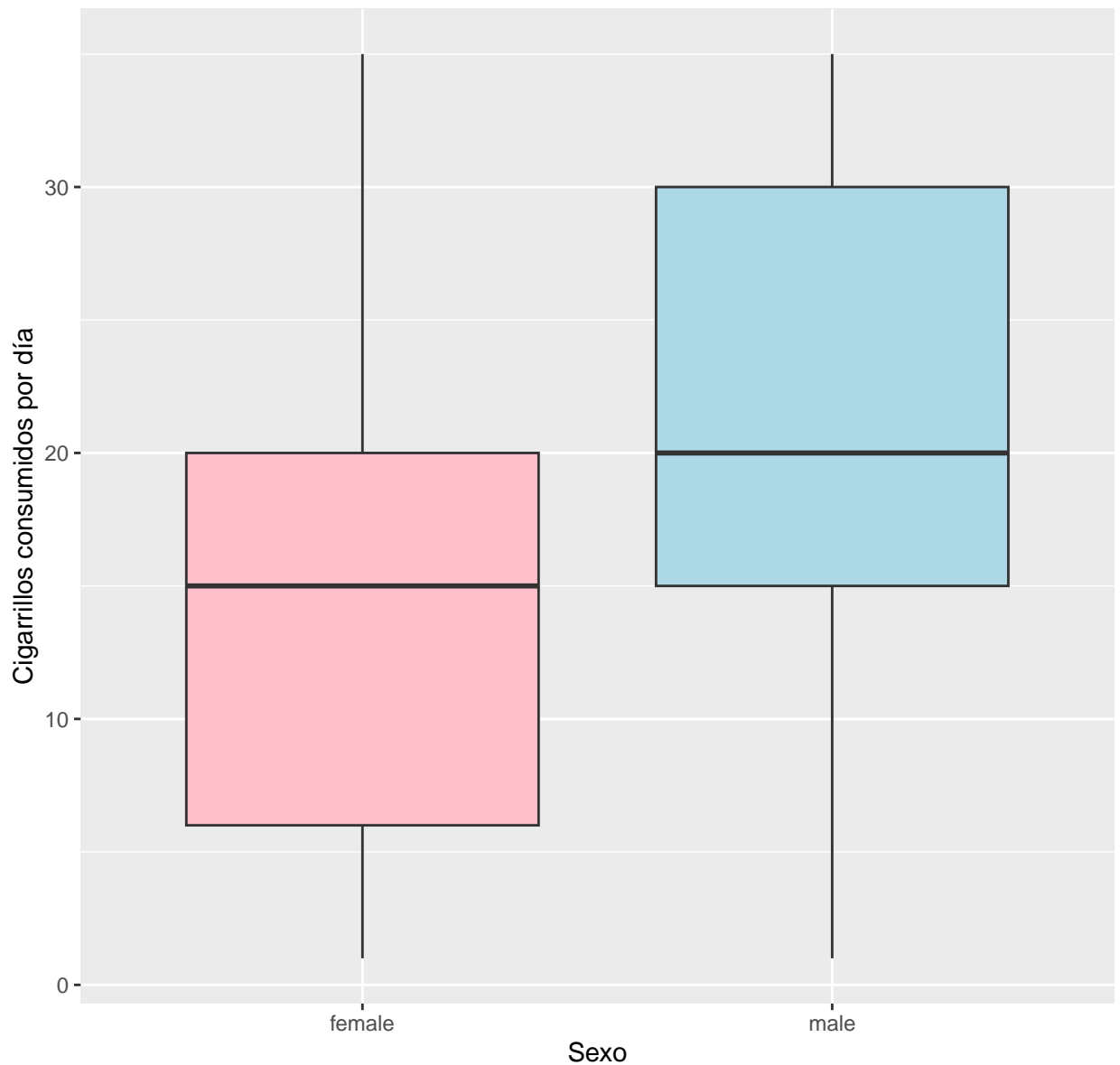


En este gráfico se muestra la correlación lineal, viendo como se distancian las medias aritméticas de cada intervalo de edad respecto a la media general extraída en la primera tabla, 17.9 cigarrillos de media. Observamos que las medias más altas se encuentran en la parte izquiera del gráfico, es decir que le pertenecen en los intervalos de edades que comprenden a las personas más relativamente jóvenes en estudio. A partir de los 55 años, se puede apreciar un declive lineal en cuanto a los promedios de cigarrillos fumados al día, lo que se corresponde con las medidas numéricas descriptivas de covarianza y correlación negativa halladas previamente.

Cantidad de cigarros que consumen las personas dependiendo de su sexo

De manera visual indicaremos que sexo dentro de la base datos consumen una mayor cantidad de cigarros por día, a través del análisis descriptivo y el gráfico que nos indica que generalmente el sexo mas propenso a consumir una mayor cantidad de cigarros al dia son los hombres mas que el grupo conformado por las mujeres.

Distribución de consumo diario de cigarrillos respecto al sexo



Este gráfico emplea el poderoso ‘five point summary’ (Tukey, 1977) a través de modelos de cajas y bigotes. Podemos observar , que el 3er cuartil de la distribución de las mujeres

fumadoras es igual a la mediana de los hombres; 20 cigarros al día. De esto podemos derivar que menos del 25% de las mujeres fumadoras en observación consumen cantidades mayores a las de un paquete estándar de 20 cigarros, contrastado con que, por el otro lado, el 50% de los hombres observados fuman 20 cigarros o más. Fumar más de un paquete por día parece ser mucho más común entre los hombres fumadores observados.

Conclusión

Dentro de este análisis estadístico descriptivo se han obtenido como resultados:

- Una correlación lineal inversa en cuanto a la edad de los fumadores y la cantidad de cigarros que consumen por día. Esto quiere decir que a medida que la edad de los fumadores observados se incrementaba, en promedio, podíamos ver un ligero decrecimiento en la cantidad de cigarros consumidos al día. Puede ser visto también desde la perspectiva opuesta, a medida que la edad decrecía, la media de cigarros consumidos al día tendía a incrementarse.
- Una disparidad considerable entre el consumo de cigarros por día entre hombres y mujeres fumadores, con una mayor proporción de hombres, el 50%, fumando más de un paquete (20 cigarrillos) por día, cantidades que a su vez, solo son correspondidas por menos de una cuarta parte, 25%, de las mujeres fumadoras observadas.

Así, se ha podido dar respuesta a las preguntas planteadas en este trabajo, pudiendo resumir en que, en promedio, de entre 1932 fumadores observados: los hombres fuman más que las mujeres, y los de una edad determinada suelen fumar más que las personas mayores a ellos.

Results

Discussion

Limitations and Future Directions

Conclusion

References

Apéndice

Title for Appendix