

**Análisis estadístico e inferencial  
para el estudio de casos de COVID-19  
con relación a la obesidad, diabetes y tabaquismo**



*Jair Jaime Hernández González*

## Índice

Introducción.....	3
Población objetivo y método de muestreo.....	3
Estadísticas descriptivas para las variables Sexo y Edad al 7 de febrero 2023	
Diabetes.....	4
Obesidad.....	8
Tabaquismo.....	12
Análisis inferencial para las variables Sexo y Edad al 9 de mayo 2023	
Diabetes.....	16
Obesidad.....	22
Tabaquismo.....	28
Conclusiones.....	34
Bibliografía.....	34

## Introducción

Durante este proyecto se estudiarán los datos obtenidos de los registros de covid-19 de 2023 en el estado de Veracruz y se visualizara en base a estos tres factores muy relevantes, los cuales son; obesidad, tabaquismo y diabetes. Con estas variables analizaremos que porcentaje de nuestra muestra sufrió de estos padecimientos mientras se encontraba bajo la enfermedad, observaremos su sexo y en que rangos de edad se puede apreciar una mayor relevancia en lo que se refiere a cada padecimiento en cuestión. Las variables que se analizarán de cada población (obesidad, tabaquismo y diabetes) serán; edad y sexo. A continuación, se presenta las estadísticas descriptivas para cada variable incluyendo tablas, gráficos y su respectiva información para apreciar de manera adecuada cada uno de los factores. A su vez realizaremos un análisis inferencial para las poblaciones ya mencionadas en las cuales observaremos su proporción respecto a las variables de sexo, como se comporta la media y su desviación estándar, además de incluir los cálculos de sus respectivos limites de confianza. Todo esto manejando pruebas de hipótesis con su grafica para comparar con la media y varianza de la población general en México.

## Población objetivo y método de muestreo

Nuestra población objetivo son todas aquellas personas infectadas con covid-19 durante este 2023 en el estado de Veracruz, de donde tomaremos 3 subgrupos los cuales son obesidad, tabaquismo y diabetes. Además, utilizaremos las columnas de sexo y edad de la base de datos como variables. Utilizaremos la columna de fecha de ingreso para seleccionar solamente los observados en 2023 y la de estado de residencia para filtrar solamente nuestro estado deseado. Nuestro tipo de muestreo se asemeja a un conglomerado o por racimos.

Variable	Tipo
Sexo	Cualitativa nominal
Edad	Cuantitativa de razon

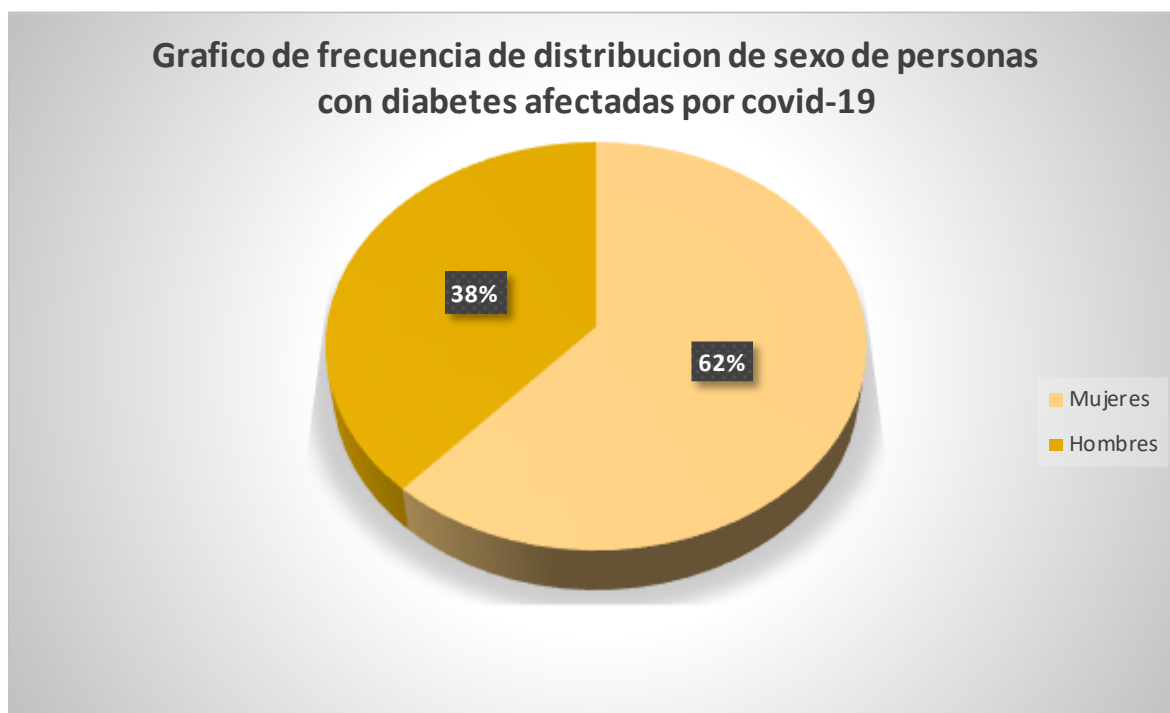
## Estadísticas descriptivas para las variables Sexo y Edad al 7 de febrero 2023

### Diabetes

*Tabla 1. Frecuencias de la variable sexo para diabetes.*

Numero de clase	Frecuencia absoluta	Frecuencia relativa
Mujeres	207	61.61%
Hombres	129	38.39%

En esta tabla se puede apreciar que hay un porcentaje bastante mayor de mujeres con diabetes que se han diagnosticado con covid-19 durante estas primeras semanas del 2023 en el estado de Veracruz.



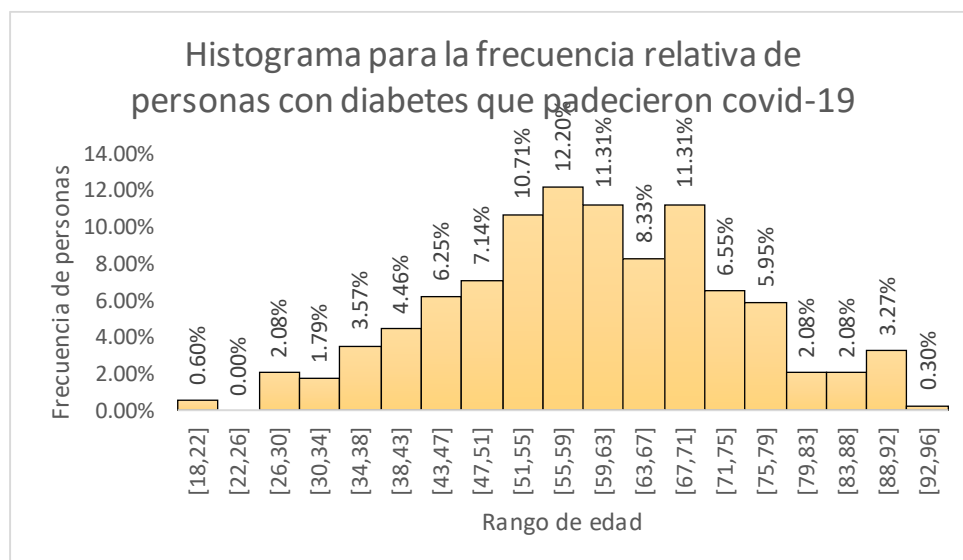
*Grafica 1. Frecuencia de la variable sexo para diabetes.*

Basado en la tabla 1 este grafico representa la gran diferencia que hay en los porcentajes de hombres y mujeres que padecen diabetes y resultaron positivos a covid-19.

Tabla 2. Frecuencias de la variable edad para diabetes.

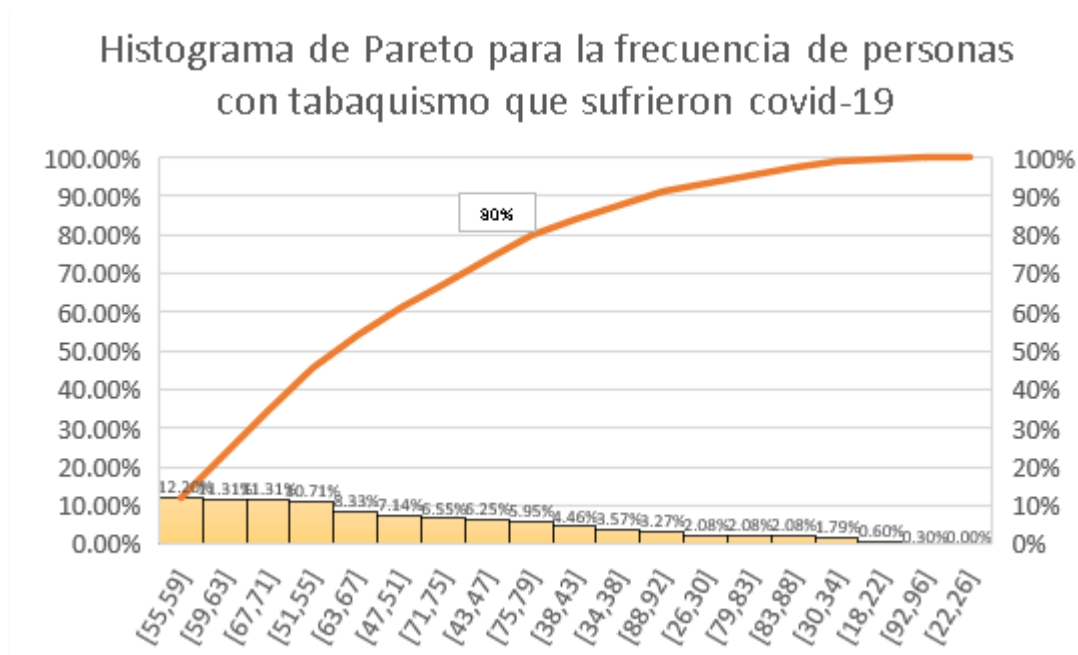
Numero de clase	Clase	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada absoluta	Frecuencia acumulada absoluta
1	[18,22]	2	0.60%	2	0.60%
2	[22,26]	0	0.00%	2	0.60%
3	[26,30]	7	2.08%	9	2.68%
4	[30,34]	6	1.79%	15	4.46%
5	[34,38]	12	3.57%	27	8.04%
6	[38,43]	15	4.46%	42	12.50%
7	[43,47]	21	6.25%	63	18.75%
8	[47,51]	24	7.14%	87	25.89%
9	[51,55]	36	10.71%	123	36.61%
10	[55,59]	41	12.20%	164	48.81%
11	[59,63]	38	11.31%	202	60.12%
12	[63,67]	28	8.33%	230	68.45%
13	[67,71]	38	11.31%	268	79.76%
14	[71,75]	22	6.55%	290	86.31%
15	[75,79]	20	5.95%	310	92.26%
16	[79,83]	7	2.08%	317	94.35%
17	[83,88]	7	2.08%	324	96.43%
18	[88,92]	11	3.27%	335	99.70%
19	[92,96]	1	0.30%	336	100.00%

Esta tabla muestra las frecuencias relativas y absolutas de las edades de personas que fueron registradas con covid-19 y diabetes, separadas en 19 clases.



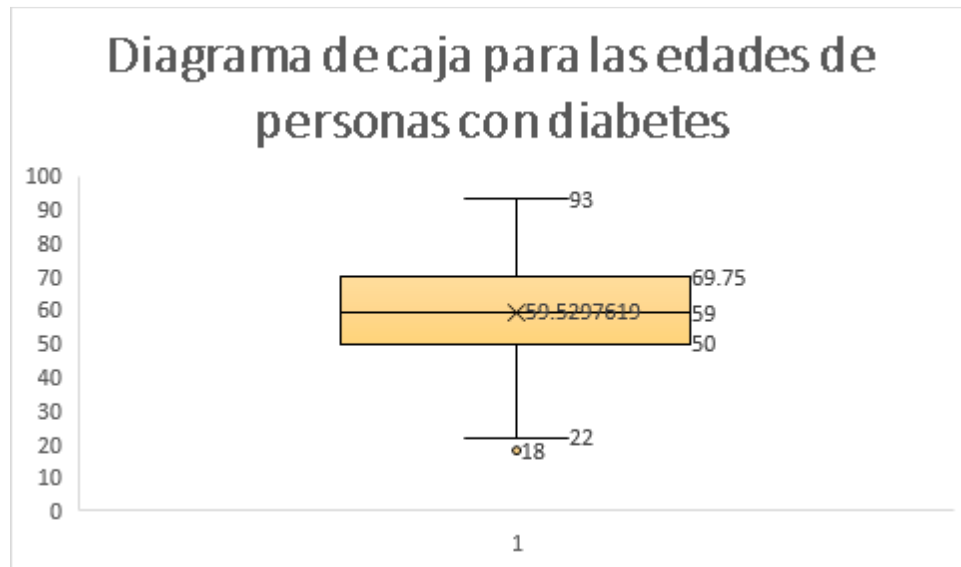
Grafica 2. Histograma de frecuencias de la variable edad para diabetes.

Se puede apreciar que la frecuencia de la edad tiene una distribución que se asemeja a la normal, es notable que las personas entre 55 y 59 años son las que más padecen de diabetes, y que en el rango de 22 a 26 años no existe persona con esta condición en nuestra muestra.



*Grafica 3. Pareto de la variable edad para diabetes.*

El 80% de las personas con diabetes están acumuladas en solo 9 clases, en las cuales podemos observar que forman un rango de 43 a 79 años. Cerca del 80% de las personas con diabetes están centradas solo en un 47% de nuestro rango de edades.



*Grafica 4. Diagrama de caja para la variable edad para diabetes.*

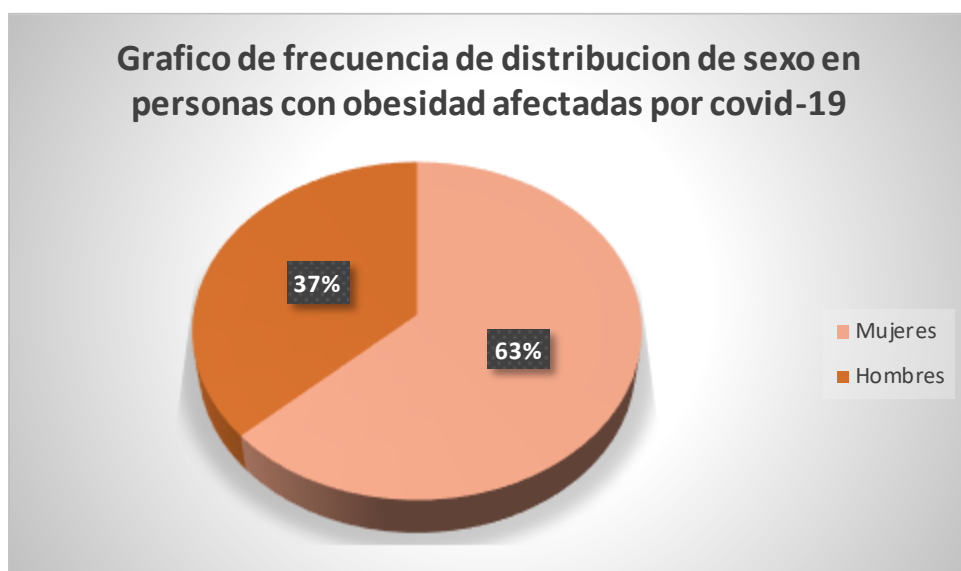
La diferencia que existe entre cuartiles es muy pequeña llegando apenas a 9, y podemos apreciar que la media muestral es casi exacta con la mediana. Sin embargo, existe una ligera variabilidad mayor en las edades de 59 a 69.75 años.

## Obesidad

*Tabla 3 Frecuencias de la variable sexo para obesidad.*

Numero de clase	Frecuencia absoluta	Frecuencia relativa
Mujeres	199	63.17%
Hombres	116	36.83%

Muestra las frecuencias de hombres y mujeres que fueron registrados con covid-19 y padecen obesidad.



*Grafica 5. Frecuencia de la variable sexo para obesidad.*

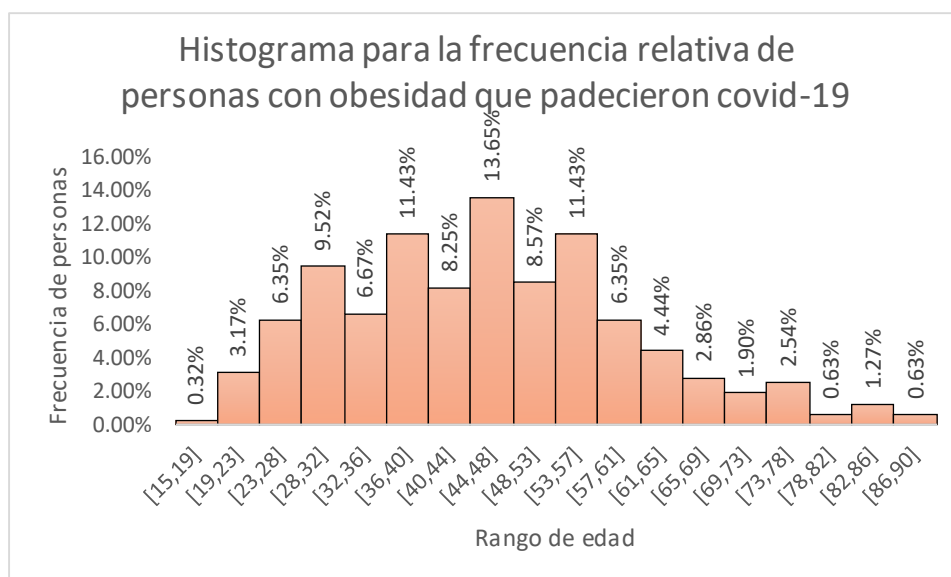
Muestra gráficamente la gran diferencia que existe en esta población entre hombres y mujeres, en el estado de Veracruz el 63% de los registrados con covid-19 y padecen de obesidad son mujeres.



Tabla 4. Frecuencias de la variable edad para obesidad.

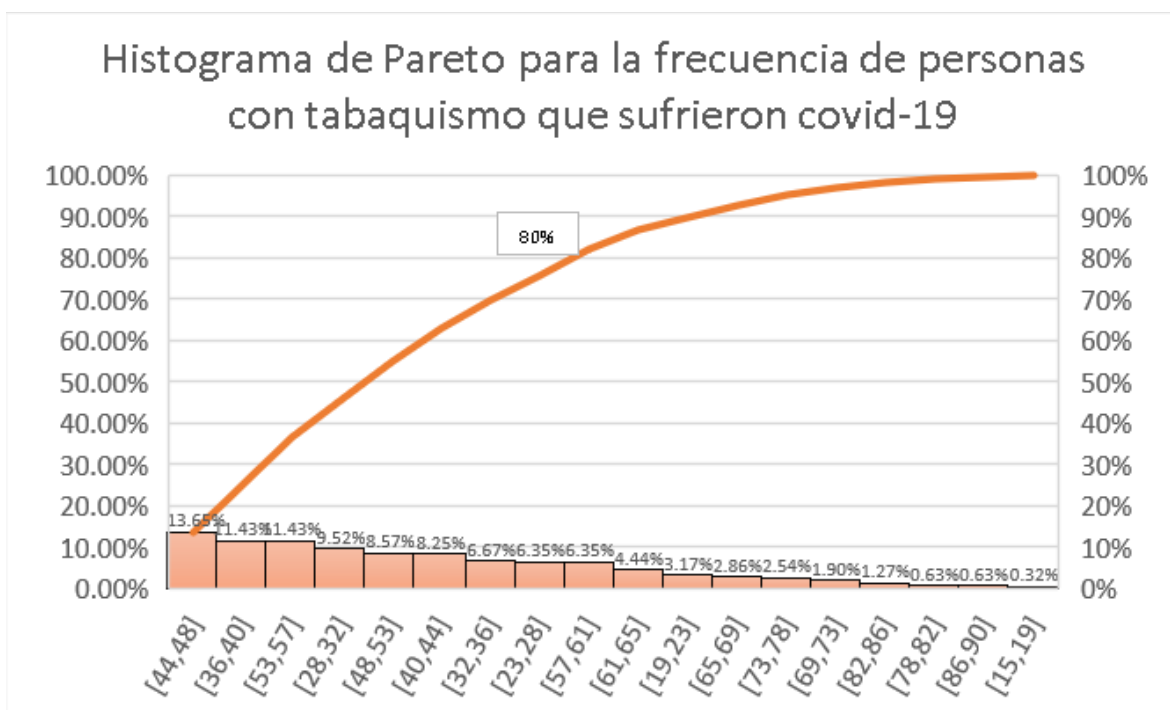
Numero de clase	Clase	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada absoluta	Frecuencia acumulada absoluta
1	[15,19]	1	0.32%	1	0.32%
2	[19,23]	10	3.17%	11	3.49%
3	[23,28]	20	6.35%	31	9.84%
4	[28,32]	30	9.52%	61	19.37%
5	[32,36]	21	6.67%	82	26.03%
6	[36,40]	36	11.43%	118	37.46%
7	[40,44]	26	8.25%	144	45.71%
8	[44,48]	43	13.65%	187	59.37%
9	[48,53]	27	8.57%	214	67.94%
10	[53,57]	36	11.43%	250	79.37%
11	[57,61]	20	6.35%	270	85.71%
12	[61,65]	14	4.44%	284	90.16%
13	[65,69]	9	2.86%	293	93.02%
14	[69,73]	6	1.90%	299	94.92%
15	[73,78]	8	2.54%	307	97.46%
16	[78,82]	2	0.63%	309	98.10%
17	[82,86]	4	1.27%	313	99.37%
18	[86,90]	2	0.63%	315	100.00%

Muestra las frecuencias de los afectados por covid-19 que padecen de obesidad, se puede observar que hay gran variedad de rangos de edad en esta población contando con 18 clases.



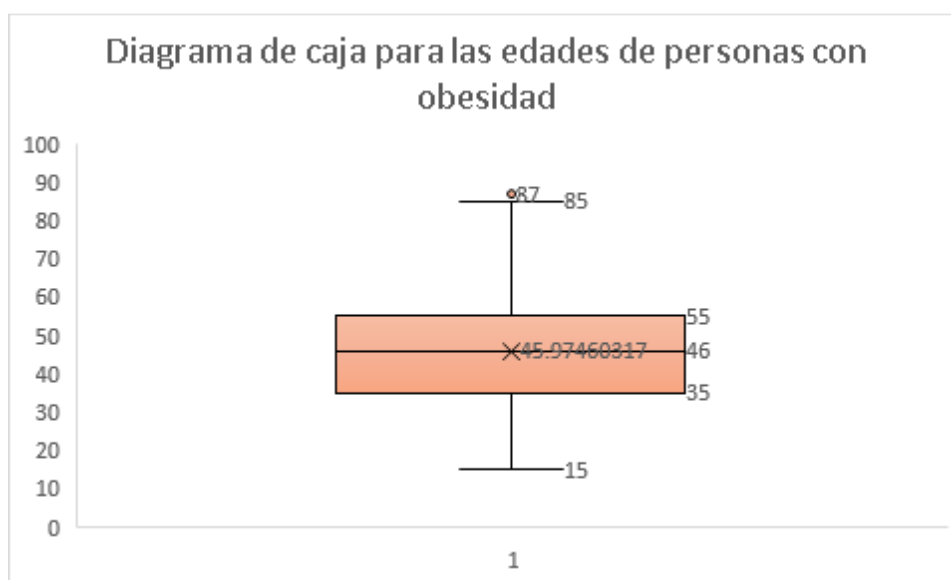
Grafica 6. Histograma de frecuencias de la variable edad para obesidad.

En este histograma observamos que hay una mayor cantidad de personas con obesidad en los rangos de 44 a 48 años, sin embargo, las clases adyacentes a esta no se quedan atrás y se reducen mientras más aumenta la edad.



Grafica 7. Pareto de la variable edad para obesidad.

Podemos observar que el 80% de las personas con obesidad que fueron muestreadas corresponden al 44% de un rango de edad.



Grafica 8. Diagrama de caja para la variable edad para obesidad.

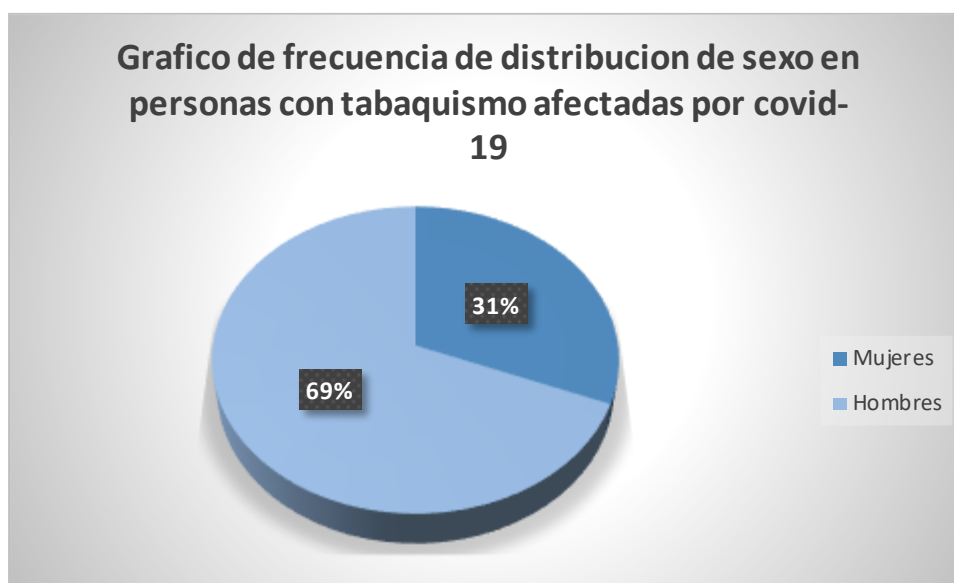
Se observa una mínima diferencia entre los cuartiles dos y cuatro con respecto al segundo, y este mismo es muy cercano a la media muestral. Podemos decir que el 50% de las personas con obesidad afectadas por covid-19 están entre los 15 y 46 años, y el otro 50% está entre los 46 y 85.

## Tabaquismo

*Tabla 5. Frecuencias de la variable sexo para tabaquismo.*

Numero de clase	Frecuencia absoluta	Frecuencia relativa
Mujeres	32	31.07%
Hombres	71	68.93%

Muestra la frecuencia absoluta y relativa de la variable sexo de personas que presentaron covid-19 y tienen tabaquismo en el estado de Veracruz.



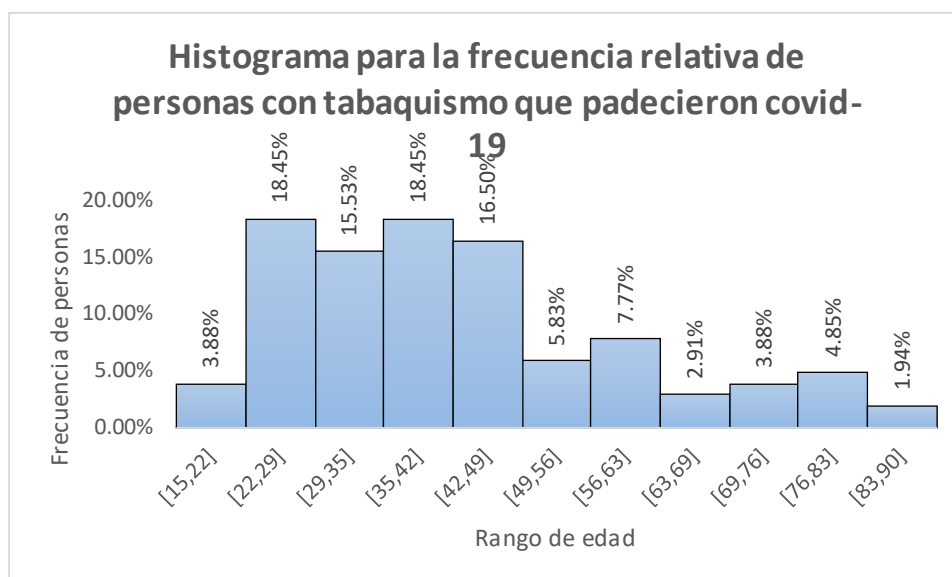
*Grafica 9. Frecuencia de la variable sexo para tabaquismo.*

Se puede observar que la mayoría de las personas que tiene tabaquismo en el estado de Veracruz son hombres con el 69%.

Tabla 6. Frecuencias de la variable edad para tabaquismo.

Numero de clase	Clase	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada absoluta	Frecuencia acumulada absoluta
1	[15,22]	4	3.88%	4	3.88%
2	[22,29]	19	18.45%	23	22.33%
3	[29,35]	16	15.53%	39	37.86%
4	[35,42]	19	18.45%	58	56.31%
5	[42,49]	17	16.50%	75	72.82%
6	[49,56]	6	5.83%	81	78.64%
7	[56,63]	8	7.77%	89	86.41%
8	[63,69]	3	2.91%	92	89.32%
9	[69,76]	4	3.88%	96	93.20%
10	[76,83]	5	4.85%	101	98.06%
11	[83,90]	2	1.94%	103	100.00%

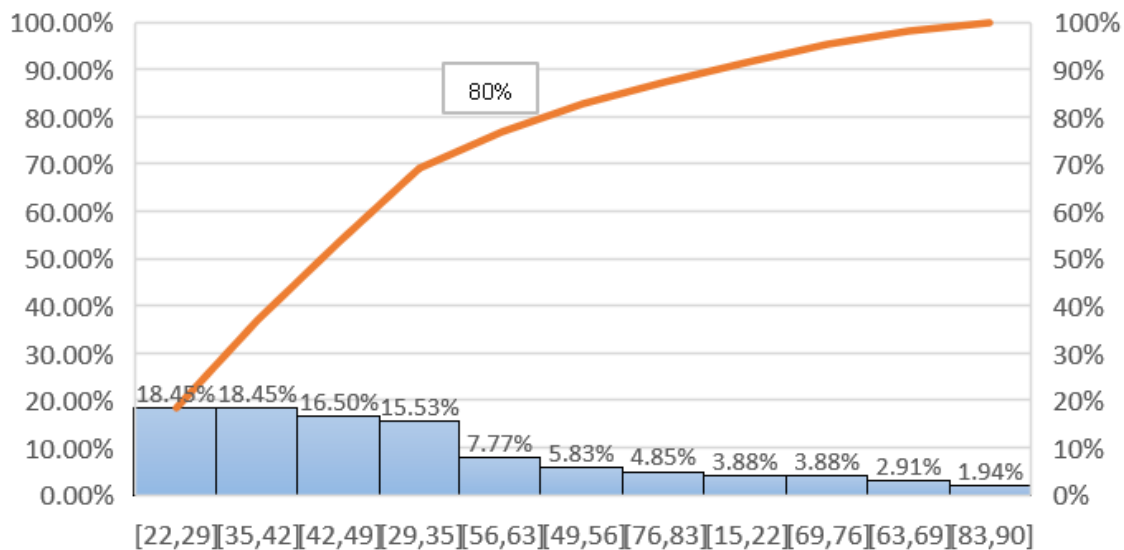
Muestra las frecuencias de la variable edad, se puede apreciar que en nuestra muestra la mayoría de las personas que tienen tabaquismo están en 4 clases de las edades de 22 a 49 años.



Grafica 10. Histograma de frecuencias de la variable edad para tabaquismo.

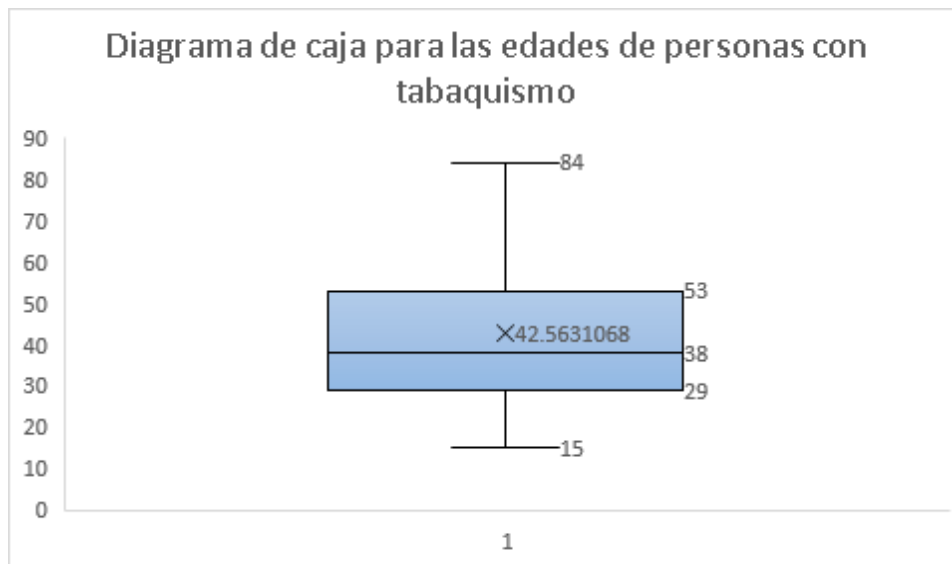
En este histograma se puede observar que los datos son bastante variados casi siendo indiferente la edad para el tabaquismo presentado en nuestra muestra, sin embargo, se centran en 4 categorías que tienen un alto porcentaje.

### Histograma de Pareto para la frecuencia de personas con tabaquismo que sufrieron covid-19



*Grafica 11. Pareto de la variable edad para tabaquismo.*

En este grafico los rangos con mayor relevancia son de rangos más bajos y se puede decir que cerca del 80% de las personas que tienen tabaquismo están acumuladas en solo un 45% de nuestro rango de edad muestreado.



*Grafica 12. Diagrama de caja para la variable edad para tabaquismo.*

En este diagrama notamos como la mediana es bastante más alejada de la media muestral. Además, el tercer cuartil es notablemente mayor que los anteriores,

podemos decir que existe una mayor variabilidad en los datos en las edades de 38 a 53 años

*Tabla 7. Medidas de centralidad para la variable edad.*

Medidas de centralidad para la variable EDAD			
Estadísticos	Diabetes	Obesidad	Tabaquismo
Media	59.53	45.97	42.56
Mediana	59	46	38
Moda	58	53	43

Muestra las medidas de centralidad para la edad. Se observa que los datos de la media, mediana y moda en la población de diabetes son muy parecidos, podría decirse que los datos de frecuencia se distribuyen casi simétricamente, mientras que en obesidad la moda es algo más elevada que las otras dos medidas, y en tabaquismo la mediana es algo menor que las otras, por lo que tiene mayor variabilidad en sus frecuencias.

## Análisis inferencial para las variables SEXO y EDAD al 9 de mayo 2023

### Diabetes

Para todas las pruebas de hipótesis correspondientes a diabetes se utilizará un nivel de significancia del 5%, y una confianza del 95% para los intervalos requeridos.

### Pruebas de hipótesis para la proporción de hombres

Suponemos muestra grande proveniente de una población normal.

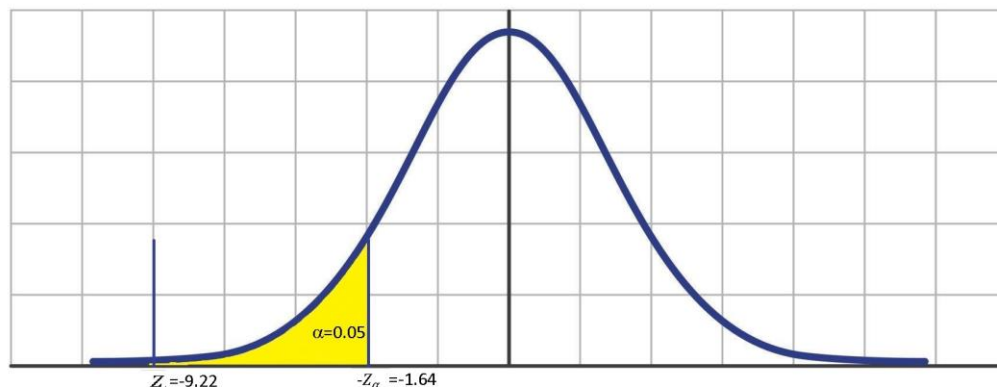
$$H_0: p \geq 50\% \text{ vs } H_1: p < 50\%$$

*Tabla 8. Medidas para las pruebas de hipótesis en proporción de hombres con diabetes.*

Estadístico de prueba	Valores críticos	P-valor
-9.22	-1.64	1.40E-20

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la proporción de hombres con diabetes.

GRAFICO DE LA REGION DE RECHAZO PARA PA PROPORCION DE HOMBRES CON DIABETES



*Grafica 13. Región de rechazo para la proporción de hombres con diabetes de cota izquierda.*

Podemos visualizar la región de rechazo de cota izquierda para la proporción de hombres con diabetes en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_{\alpha}$ . Como  $-9.22 < -1.64$  se rechaza  $H_0$ , entonces la proporción es menor a 50% con una significancia del 5%. Los datos afirman que la proporción de hombres con diabetes es menor a 50% con un error del 5%.



### Intervalo de confianza para la proporción de hombres

Suponemos una muestra de una población normal grande.

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$0.39 - 1.96 \sqrt{\frac{0.39(1-0.39)}{1871}} < p < 0.39 + 1.96 \sqrt{\frac{0.39(1-0.39)}{1871}}$$
$$0.37 < p < 0.42$$

Al sustituir en la formula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.37 y 0.42, lo cual afirma que la proporción de hombres con diabetes es menor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

### Pruebas de hipótesis para la proporción de mujeres

Suponemos muestra grande proveniente de una población normal.

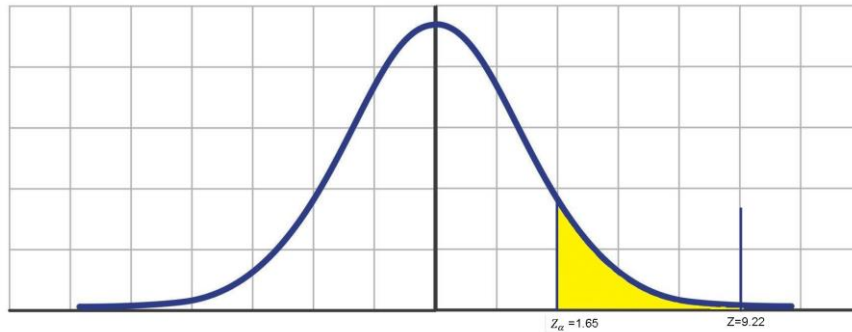
$$H_0: p \leq 50\% \text{ vs } H_1: p > 50\%$$

*Tabla 9. Medidas para las pruebas de hipótesis en proporción de mujeres con diabetes.*

Estadístico de prueba	Valores críticos	P-valor
9.22	1.64	0

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota derecha para la proporción de mujeres con diabetes.

GRAFICO DE LA REGION DE RECHAZO DE LA PROPORCION DE MUJERES CON DIABETES



*Grafica 14. Región de rechazo para la proporción de mujeres con diabetes de cota derecha.*

Podemos visualizar la región de rechazo de cota derecha para la proporción de mujeres con diabetes en una distribución normal.

Se rechaza  $H_0$ , si  $Z > Z_{\alpha}$ , Como  $9.22 > 1.64$  se rechaza  $H_0$ , la proporción es mayor a 50% con una significancia del 5%. Los datos muestran que la proporción de mujeres con diabetes es mayor a 50% con un error del 5%.

### Intervalo de confianza para la proporción de mujeres

Suponemos una muestra de una población normal grande.

$$\begin{aligned} \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ 0.61 - 1.96 \sqrt{\frac{0.61(1-0.61)}{1871}} < p < 0.61 + 1.96 \sqrt{\frac{0.61(1-0.61)}{1871}} \\ 0.58 < p < 0.63 \end{aligned}$$

Al sustituir en la fórmula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.58 y 0.63, lo cual afirma que la proporción de mujeres con diabetes es mayor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

## Media de la edad

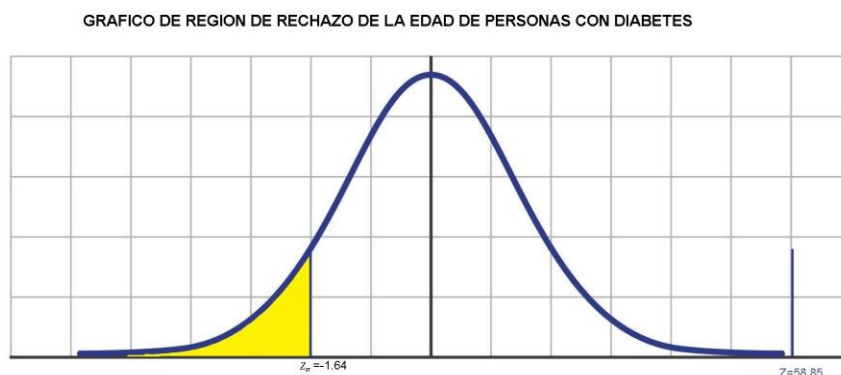
Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

$$H_0: \mu \geq 40 \text{ vs } H_1: \mu < 40$$

Tabla 10. Medidas para las pruebas de hipótesis en la media para la edad de personas con diabetes.

Estadístico de prueba	Valores críticos	P-valor
-1.64	58.85	1

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la media de edad de personas con diabetes.



Grafica 15. Región de rechazo para la media de edad de personas con diabetes de cota izquierda.

Podemos visualizar la región de rechazo de cota izquierda para la media de edad de personas con diabetes en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_{\alpha}$ , Como  $58.85 > -1.64$  no se rechaza  $H_0$  entonces la media es mayor a 40 con una significancia del 5%. Los datos muestran que la media de edad de personas con diabetes es mayor a 40 con un error del 5%.

## Intervalo de confianza para la media de edad

Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

$$\bar{X} - z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{X} + z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$$
$$59.74 - 1.96 \left( \frac{14.51}{\sqrt{1871}} \right) < \mu < 59.74 + 1.96 \left( \frac{14.51}{\sqrt{1871}} \right)$$

$$59.08 < \mu < 60.39$$

Al sustituir en la fórmula de intervalos de confianza para medias vemos que la media esta entre 59.08 y 60.39, lo cual afirma que la media de edad de personas con diabetes se encuentra entre 59.08 años y 60.39 años con una confiabilidad del 95%.

### Desviación estándar de la Edad

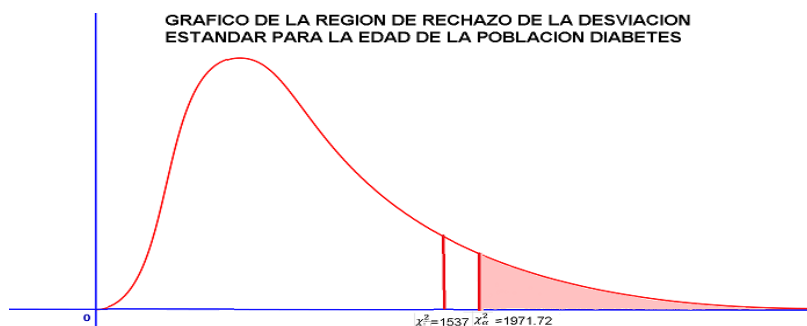
Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$H_0: \sigma \leq 16 \text{ vs } H_1: \sigma > 16$$

*Tabla 11. Medidas para las pruebas de hipótesis en la desviación estándar para la edad de personas con diabetes.*

Estadístico de prueba	Valores críticos	P-valor
1537.27	1971.72	1

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota derecha para la desviación estándar de la edad de personas con diabetes.



*Grafica 16. Región de rechazo para la desviación estándar de edad de personas con diabetes de cota derecha.*

Podemos visualizar la región de rechazo de cota derecha para la media de edad de personas con diabetes en una distribución chi-cuadrada.

Se rechaza  $H_0$ , si  $\chi^2 > \chi^2_{\alpha, n-1}$ . Como  $1537.27 < 1971.72$  no rechazamos  $H_0$ , la desviación estándar es menor a 16 con una significancia del 5%. Los datos afirman que la desviación estándar de las personas con diabetes es menor a 16 años con un error el 5%.

### Intervalo de confianza para la media de edad

Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}$$

$$\frac{(1871 - 1)14.51^2}{1991.75} < \sigma^2 < \frac{(1871 - 1)14.51^2}{1752.04}$$

$$14.06 < \sigma < 14.99$$

Al sustituir en la fórmula de intervalos de confianza para desviaciones estándar vemos que la desviación esta entre 14.06 y 14.99, lo cual afirma que la desviación estándar de la edad de personas con diabetes se encuentra entre 14.06 años y 14.99 años con una confiabilidad del 95%.

## Obesidad

Para todas las pruebas de hipótesis correspondientes a obesidad se utilizará un nivel de significancia del 5%, y una confianza del 95% para los intervalos requeridos.

### Pruebas de hipótesis para la proporción de hombres

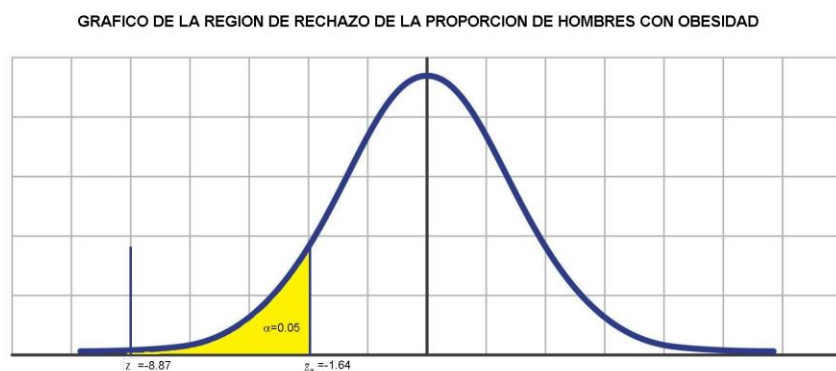
Suponemos muestra grande proveniente de una población normal.

$$H_0: p \geq 50\% \text{ vs } H_1: p < 50\%$$

*Tabla 12. Medidas para las pruebas de hipótesis en proporción de hombres con obesidad.*

Estadístico de prueba	Valores críticos	P-valor
-8.87	-1.64	3.5144E-19

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la proporción de hombres con obesidad.



*Grafica 17. Región de rechazo para la proporción de hombres con obesidad de cota izquierda.*

Podemos visualizar la región de rechazo de cota izquierda para la proporción de hombres con diabetes en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_\alpha$ , Como  $-8.87 < -1.64$  se rechaza  $H_0$ , entonces la proporción es menor a 50% con una significancia del 5%. Los datos afirman que la proporción de hombres con obesidad es menor a 50% con un error del 5%.

### Intervalo de confianza para la proporción de hombres

Suponemos una muestra de una población normal grande.

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.37 - 1.96 \sqrt{\frac{0.37(1 - 0.37)}{1105}} < p < 0.37 + 1.96 \sqrt{\frac{0.37(1 - 0.37)}{1105}}$$

$$0.34 < p < 0.39$$

Al sustituir en la fórmula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.34 y 0.39, lo cual afirma que la proporción de hombres con obesidad es menor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

### Pruebas de hipótesis para la proporción de mujeres

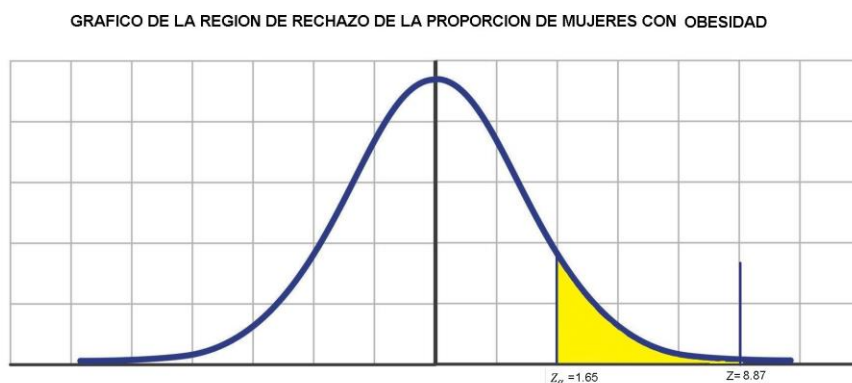
Suponemos muestra grande proveniente de una población normal.

$$H_0: p \leq 50\% \text{ vs } H_1: p > 50\%$$

*Tabla 13. Medidas para las pruebas de hipótesis en proporción de mujeres con obesidad.*

Estadístico de prueba	Valores críticos	P-valor
8.87	1.64	0

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota derecha para la proporción de mujeres con obesidad.



*Grafica 18. Región de rechazo para la proporción de mujeres con obesidad de cota derecha.*

Podemos visualizar la región de rechazo de cota derecha para la proporción de mujeres con obesidad en una distribución normal.

Se rechaza  $H_0$ , si  $Z > Z_{\alpha}$ , Como  $8.87 > 1.64$  se rechaza  $H_0$ , la proporción es mayor a 50% con una significancia del 5%. Los datos muestran que la proporción de mujeres con obesidad es mayor a 50% con un error del 5%.

### Intervalo de confianza para la proporción de mujeres

Suponemos una muestra de una población normal grande.

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.63 - 1.96 \sqrt{\frac{0.63(1-0.63)}{1105}} < p < 0.63 + 1.96 \sqrt{\frac{0.63(1-0.63)}{1105}}$$

$$0.61 < p < 0.66$$

Al sustituir en la fórmula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.61 y 0.66, lo cual afirma que la proporción de mujeres con diabetes es mayor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

### Media de la edad

Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

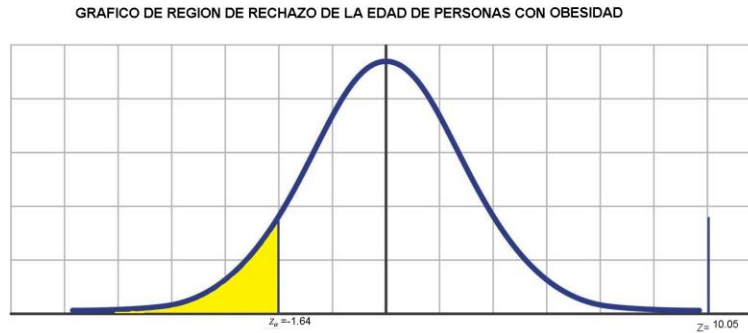
$$H_0: \mu \geq 40 \text{ vs } H_1: \mu < 40$$

*Tabla 14. Medidas para las pruebas de hipótesis en la media para la edad de personas con obesidad.*

Estadístico de prueba	Valores críticos	P-valor
10.05	-1.64	1

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la media de edad de personas con obesidad.





*Grafica 19. Región de rechazo para la media de edad de personas con obesidad de cota izquierda.*

Podemos visualizar la región de rechazo de cota izquierda para la media de edad de personas con obesidad en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_{\alpha}$ , Como  $10.05 > -1.64$  no se rechaza  $H_0$  entonces la media es mayor a 40 con una significancia del 5%. Los datos muestran que la media de edad de personas con obesidad es mayor a 40 con un error del 5%.

### Intervalo de confianza para la media de edad

Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

$$\begin{aligned} \bar{X} - z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{X} + z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) \\ 44.82 - 1.96 \left( \frac{15.95}{\sqrt{1105}} \right) < \mu < 44.82 + 1.96 \left( \frac{15.95}{\sqrt{1105}} \right) \\ 43.88 < \mu < 45.76 \end{aligned}$$

Al sustituir en la fórmula de intervalos de confianza para medias vemos que la media esta entre 43.88 y 45.76, lo cual afirma que la media de edad de personas con diabetes se encuentra entre 43.88 años y 45.76 años con una confiabilidad del 95%.

### Desviación estándar de la Edad

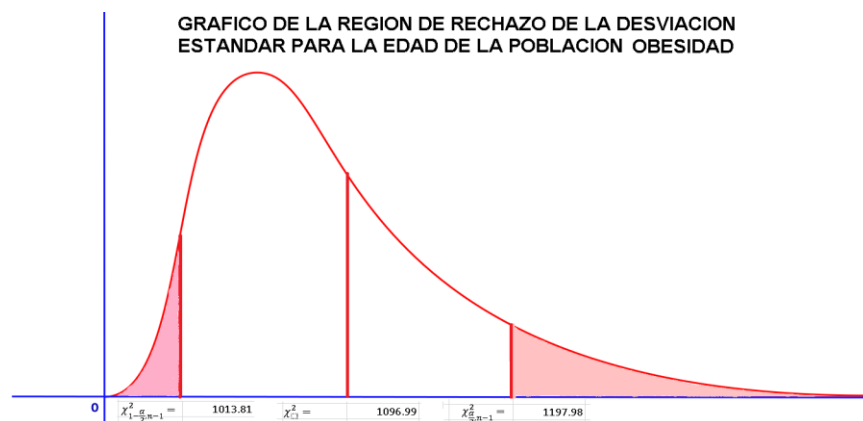
Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$H_0: \sigma = 16 \text{ vs } H_1: \sigma \neq 16$$

Tabla 15. Medidas para las pruebas de hipótesis en la desviación estándar para la edad de personas con obesidad.

Estadístico de prueba	Valores críticos	P-valor
1096.99	1013.81 y 1197.98	1.11

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis bilateral para la desviación estándar de la edad de personas con obesidad.



Grafica 20. Región de rechazo para la desviación estándar de edad de personas con obesidad bilateral.

Podemos visualizar la región de rechazo de cota derecha para la media de edad de personas con diabetes en una distribución chi-cuadrada.

Se rechaza  $H_0$ , si  $\chi^2 > \chi^2_{\frac{\alpha}{2}, n-1}$  o si  $\chi^2 < \chi^2_{1-\frac{\alpha}{2}, n-1}$ . Como  $1013.81 < 1096.99 < 1197.98$ , no rechazamos  $H_0$ , por lo que la desviación estándar es igual a 16 con un error del 5%. Los datos afirman que la desviación estándar de la edad de personas con obesidad es igual a 16 años con un error del 5%.

### Intervalo de confianza para la media de edad

Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}$$

$$\frac{(1105-1)15.95^2}{1197.98} < \sigma^2 < \frac{(1105-1)15.95^2}{1013.81}$$

$$15.31 < \sigma < 16.64$$

Al sustituir en la fórmula de intervalos de confianza para desviaciones estándar vemos que la desviación esta entre 15.31 y 16.64, lo cual afirma que la desviación estándar de la edad de personas con diabetes se encuentra entre 15.31 años y 16.64 años con una confiabilidad del 95%.

## Tabaquismo

Para todas las pruebas de hipótesis correspondientes a diabetes se utilizará un nivel de significancia del 5%, y una confianza del 95% para los intervalos requeridos.

### Pruebas de hipótesis para la proporción de hombres

Suponemos muestra grande proveniente de una población normal.

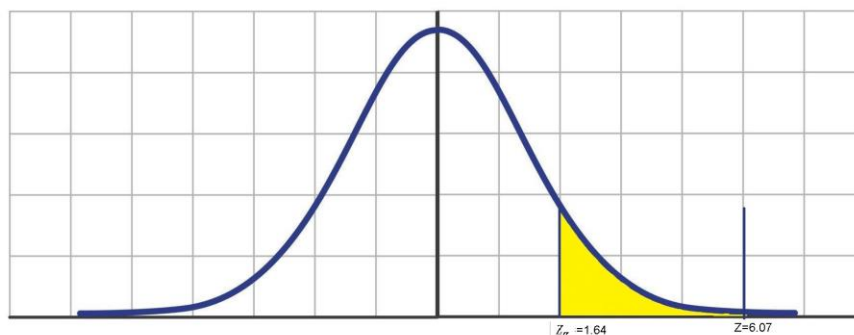
$$H_0: p \leq 50\% \text{ vs } H_1: p > 50\%$$

*Tabla 16. Medidas para las pruebas de hipótesis en proporción de hombres con tabaquismo.*

Estadístico de prueba	Valores críticos	P-valor
6.07	1.64	6.5103E-10

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota derecha para la proporción de hombres con tabaquismo.

GRAFICO DE LA REGION DE RECHAZO DE LA PROPORCION DE HOMBRES CON TABAQUISMO



*Grafica 21. Región de rechazo para la proporción de hombres con tabaquismo de cota derecha.*

Podemos visualizar la región de rechazo de cota derecha para la proporción de hombres con tabaquismo en una distribución normal.

Se rechaza  $H_0$ , si  $Z > Z_{\alpha}$ , Como  $6.07 > 1.64$  se rechaza  $H_0$ , entonces la proporción es mayor a 50% con una significancia del 5%. Los datos afirman que la proporción de hombres con tabaquismo es mayor a 50% con un error del 5%.

### Intervalo de confianza para la proporción de hombres

Suponemos una muestra de una población normal grande.

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.65 - 1.96 \sqrt{\frac{0.65(1-0.65)}{411}} < p < 0.39 + 1.96 \sqrt{\frac{0.65(1-0.65)}{411}}$$

$$0.6 < p < 0.7$$

Al sustituir en la fórmula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.6 y 0.7, lo cual afirma que la proporción de hombres con tabaquismo es mayor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

### Pruebas de hipótesis para la proporción de mujeres

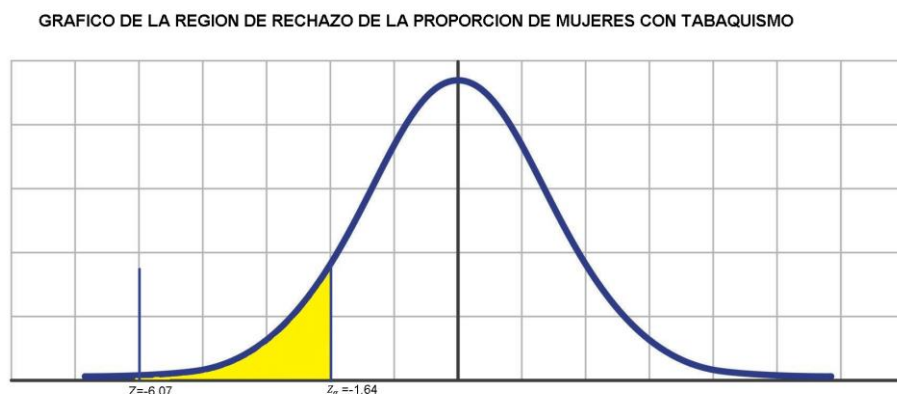
Suponemos muestra grande proveniente de una población normal.

$$H_0: p \geq 50\% \text{ vs } H_1: p < 50\%$$

*Tabla 17. Medidas para las pruebas de hipótesis en proporción de mujeres con tabaquismo.*

Estadístico de prueba	Valores críticos	P-valor
-6.07	-1.64	6.5103E-10

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la proporción de mujeres con tabaquismo.



*Grafica 22. Región de rechazo para la proporción de mujeres con tabaquismo de cota izquierda.*

Podemos visualizar la región de rechazo de cota izquierda para la proporción de mujeres con tabaquismo en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_{\alpha}$ , Como  $-9.22 < -1.64$  se rechaza  $H_0$ , la proporción no es mayor a 50% con una significancia del 5%. Los datos muestran que la proporción de mujeres con tabaquismo es menor a 50% con un error del 5%.

### Intervalo de confianza para la proporción de mujeres

Suponemos una muestra de una población normal grande.

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.35 - 1.96 \sqrt{\frac{0.35(1-0.35)}{411}} < p < 0.35 + 1.96 \sqrt{\frac{0.35(1-0.35)}{411}}$$

$$0.3 < p < 0.4$$

Al sustituir en la fórmula de intervalos de confianza para proporciones vemos que la proporción esta entre 0.3 y 0.4, lo cual afirma que la proporción de mujeres con tabaquismo es menor al 50% y se encuentra dentro de este intervalo con una confiabilidad del 95%.

### Media de la edad

Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

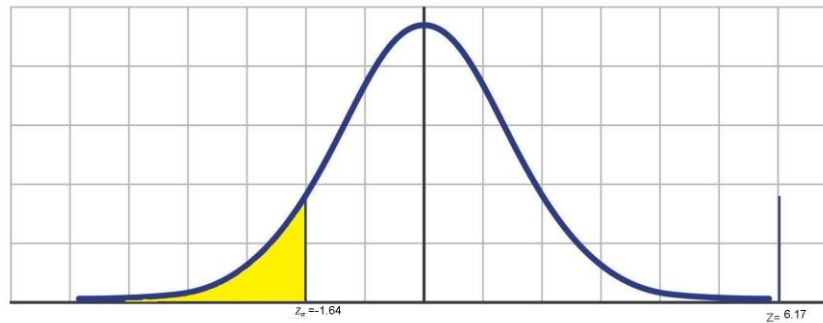
$$H_0: \mu \geq 40 \text{ vs } H_1: \mu < 40$$

*Tabla 18. Medidas para las pruebas de hipótesis en la media para la edad de personas con tabaquismo.*

Estadístico de prueba	Valores críticos	P-valor
6.17	-1.64	1

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la media de edad de personas con tabaquismo.

GRAFICO DE REGION DE RECHAZO DE LA EDAD DE PERSONAS CON TABAQUISMO



*Grafica 23. Región de rechazo para la media de edad de personas con tabaquismo de cota izquierda.*

Podemos visualizar la región de rechazo de cota izquierda para la media de edad de personas con tabaquismo en una distribución normal.

Se rechaza  $H_0$ , si  $Z < Z_\alpha$ , Como  $6.17 > -1.64$  no se rechaza  $H_0$  entonces la media es mayor a 40 con una significancia del 5%. Los datos muestran que la media de edad de personas con tabaquismo es mayor a 40 con un error del 5%.

### Intervalo de confianza para la media de edad

Suponemos muestra grande, por lo que la varianza poblacional se puede sustituir por la varianza muestral.

$$\begin{aligned} \bar{X} - z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) &< \mu < \bar{X} + z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) \\ 45.48 - 1.96 \left( \frac{18.01}{\sqrt{411}} \right) &< \mu < 45.48 + 1.96 \left( \frac{18.01}{\sqrt{411}} \right) \\ 43.74 &< \mu < 47.22 \end{aligned}$$

Al sustituir en la fórmula de intervalos de confianza para medias vemos que la media esta entre 43.74 y 47.22, lo cual afirma que la media de edad de personas con tabaquismo se encuentra entre 43.74 años y 47.22 años con una confiabilidad del 95%.

### Desviación estándar de la Edad

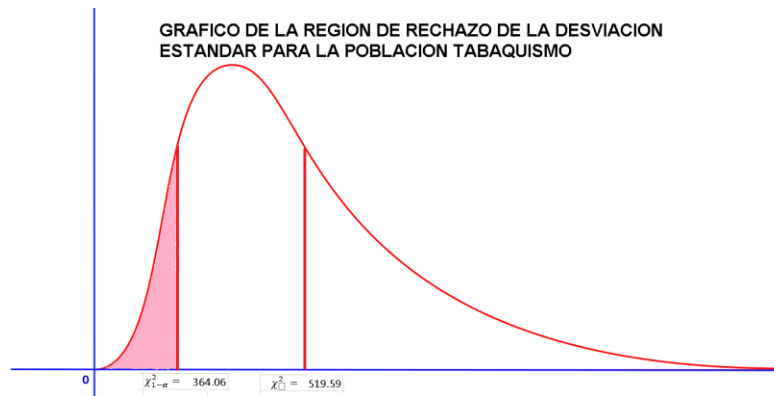
Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$H_0: \sigma \geq 16 \text{ vs } H_1: \sigma < 16$$

Tabla 19. Medidas para las pruebas de hipótesis en la desviación estándar para la edad de personas con tabaquismo.

Estadístico de prueba	Valores críticos	P-valor
519.59	364.06	1

Muestra los datos obtenidos en los cálculos de; estadístico de prueba, valores críticos y p-valor, realizados para las pruebas de hipótesis de cota izquierda para la desviación estándar de la edad de personas con tabaquismo.



Grafica 24. Región de rechazo para la desviación estándar de edad de personas con tabaquismo de cota izquierda.

Podemos visualizar la región de rechazo de cota izquierda para la media de edad de personas con diabetes en una distribución chi-cuadrada.

Se rechaza  $H_0$ , si  $\chi^2 < \chi^2_{1-\alpha, n-1}$ . Como  $519.59 > 364.06$ , ro rechazamos  $H_0$ , la desviación estándar es mayor a 16 con una significancia del 5%. Los datos afirman que la desviación estándar de las personas con diabetes es menor a 16 años con un error el 5%.

### Intervalo de confianza para la media de edad

Suponer una muestra que proviene de una población con distribución normal con varianza conocida.

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}$$

$$\frac{(411-1)18.01^2}{355.79} < \sigma^2 < \frac{(411-1)18.01^2}{467.99}$$

$$16.86 < \sigma < 19.34$$



Al sustituir en la fórmula de intervalos de confianza para desviaciones estándar vemos que la desviación esta entre 16.86 y 19.34, lo cual afirma que la desviación estándar de la edad de personas con diabetes se encuentra entre 16.86 años y 19.34 años con una confiabilidad del 95%.

## Conclusiones

Respecto a la variable sexo podemos decir que en las poblaciones de diabetes y obesidad existe una proporción de más de un 50% de mujeres y caso contrario en tabaquismo donde predominan los hombres. Además, es de resaltar que, de nuestras tres poblaciones, la de tabaquismo es significativamente menor a las otras.

Podemos concluir que en las personas infectadas con coronavirus y que padecen de obesidad, tabaquismo o diabetes, tienen una media de edad superior a los 40 años. Siendo diabetes donde se tiene el intervalo de confianza mas alto estando entre 59.08 y 60.39 años. Podría interpretarse de tal forma que la diabetes es un padecimiento que se presenta con mas frecuencia en personas mayores.

Con relación a la desviación estándar tenemos que en obesidad nuestra desviación es de 16 años, sin embargo, en la población de diabetes fue un poco menor, caso contrario en tabaquismo donde sobrepaso esta medida. En base a esto podemos decir que las edades de personas infectadas con covid-19 y que padecen de tabaquismo son mucho mas variadas que las edades de las otras poblaciones.

Aquí termina el análisis estadístico e inferencial respecto a las poblaciones de tabaquismo, obesidad y diabetes. Durante este proyecto he aprendido mucho respecto al uso de la inferencia para obtener datos que pueden reflejar lo que pasa a gran escala en nuestro país para este caso. También es apreciable cómo se comportan los datos con el tiempo, ya que a pesar de que nuestro análisis estadístico solo tomo un par de semanas del 2023 y el análisis inferencial hasta inicios de mayo, es notable que los datos se comportan de manera muy similar y nos dicen mucho acerca de lo que está pasando en el estado de Veracruz durante este primer trimestre del 2023. Para terminar, agradezco al lector por llegar hasta aquí y espero que sea un contenido útil para conocimiento general o incluso para empezar un análisis propio respecto a alguna perspectiva de interés personal.

## Bibliografía

De Salud, S. (s. f.). Datos Abiertos Dirección General de Epidemiología. gob.mx. <https://www.gob.mx/salud/documentos/datos-abiertos-152127>