

Profesor Coordinador: Mario Castillo

Profesores: Astrid Bernal, Carlos Castellanos, Fabio Lagos, María Alejandra López, Gonzalo Torres, Hernando Mutis.

Segundo semestre de 2015

Complementaria 10

Estadística descriptiva (Excel y SPSS)

Descargue de Sicua el archivo de Excel con el nombre “**Tabla datos complementaria 10**”. Allí encontrará cuatro indicadores medioambientales para 148 países del mundo correspondientes al año 2014. Dichos indicadores son los siguientes:

- **Índice de Desempeño Ambiental:** medido en una escala de 0 a 100.
- **Emisión anual de CO₂ por persona** (en toneladas).
- **Nivel de contaminación por CO₂ por persona** (Bajo, Medio, Alto, Muy Alto, depende de la emisión anual de CO₂).
- **Porcentaje de especies en peligro de extinción.**

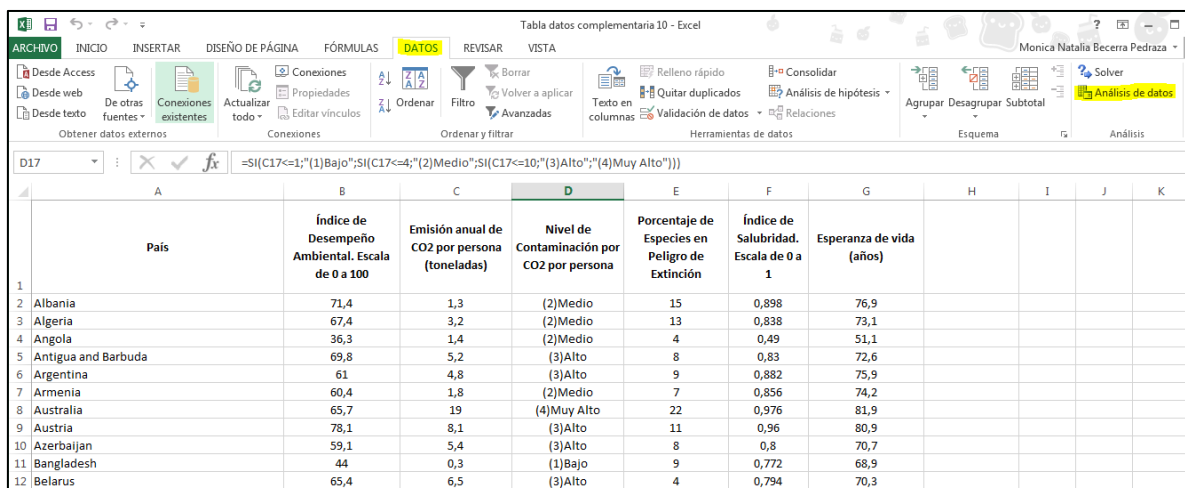
Con base en esta información de solución a los siguientes literales:

- Realice un resumen de las principales estadísticas descriptivas correspondientes a la variable **Índice de Desempeño Ambiental**.

Opción 1: Excel

Estadísticas Descriptivas: Datos / Análisis de Datos / Estadística Descriptiva.

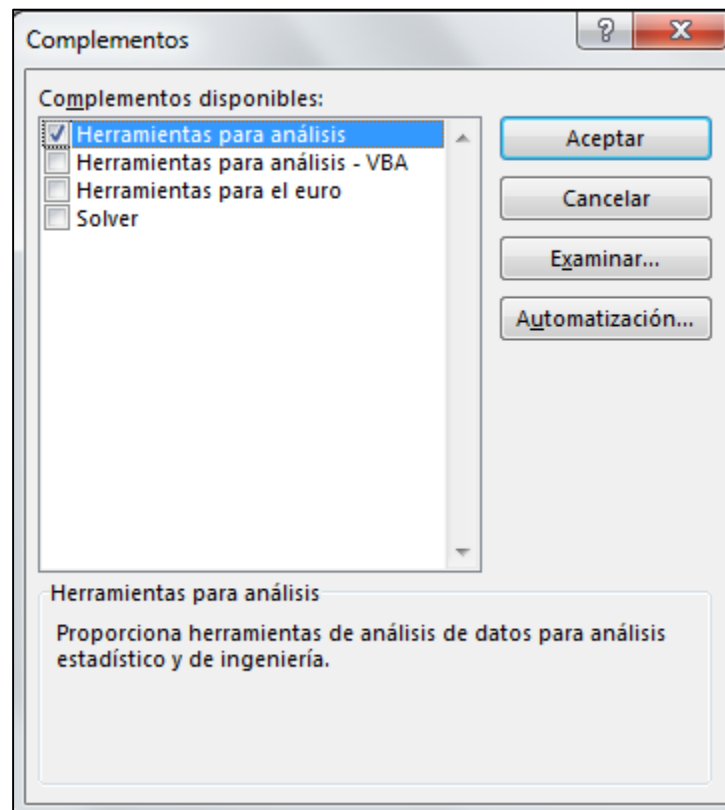
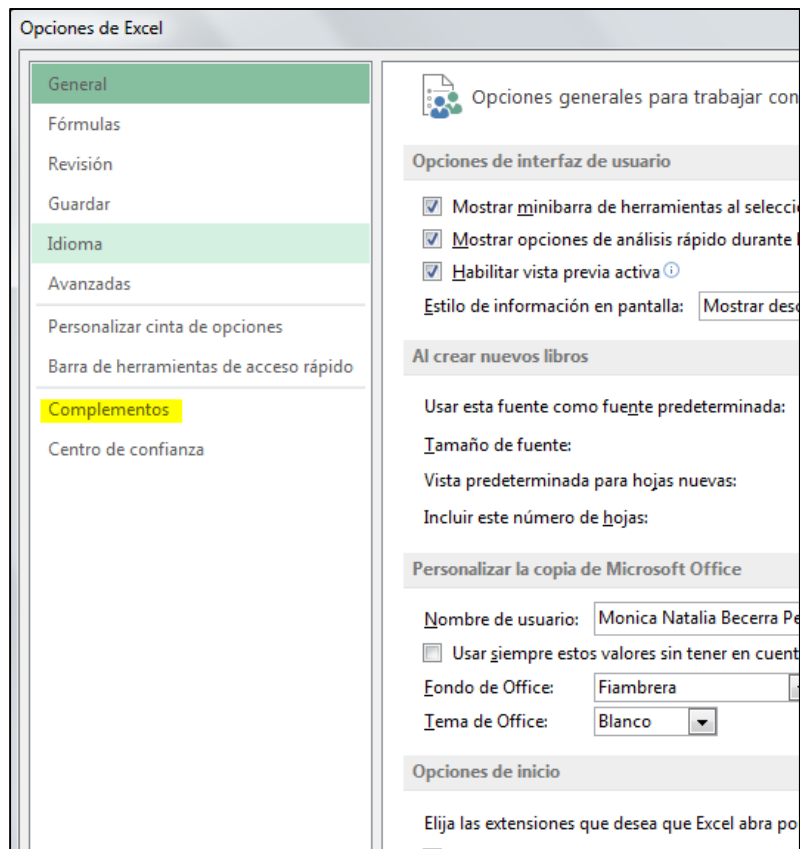
Paso 1: Verificar en el menú Datos de Excel, que este habilitada la opción Análisis de Datos. Si no está habilitada realizar el paso 2.



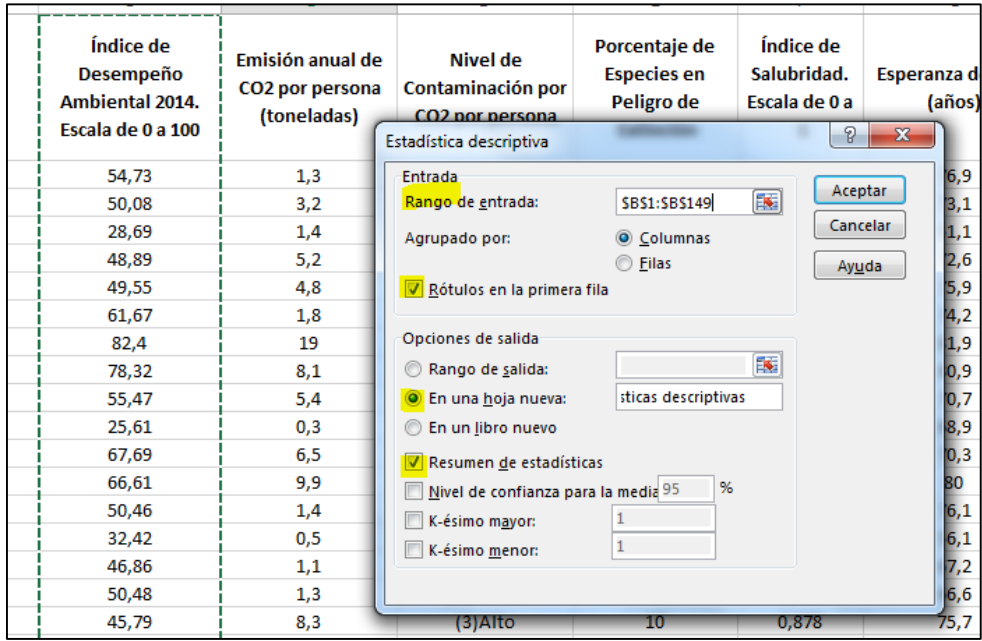
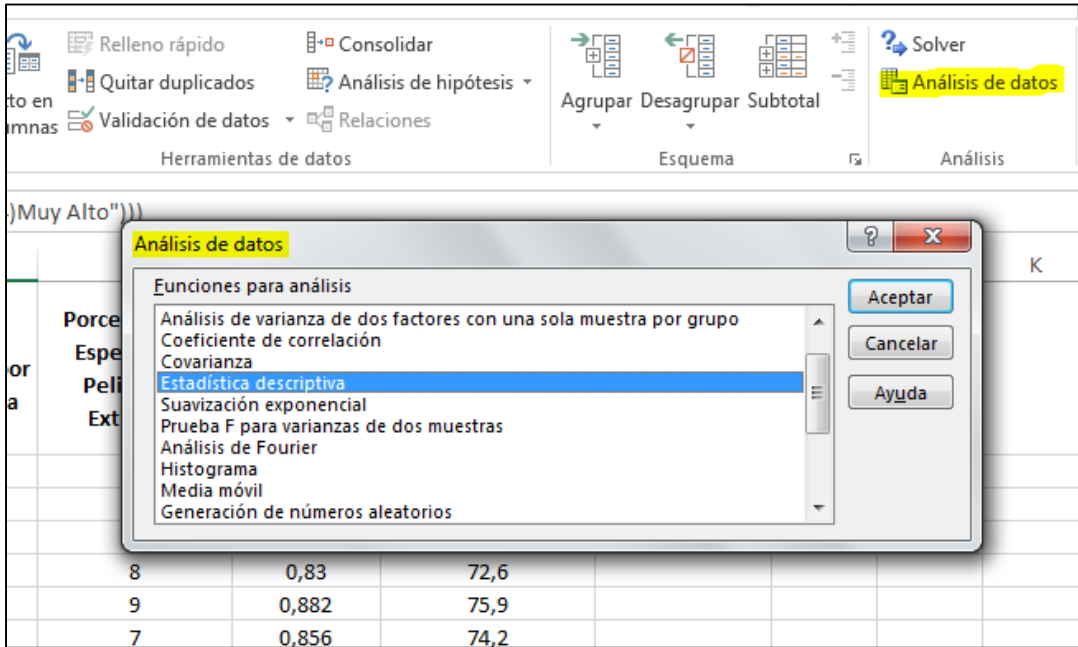
	A	B	C	D	E	F	G	H	I	J	K
	País	Índice de Desempeño Ambiental. Escala de 0 a 100	Emisión anual de CO2 por persona (toneladas)	Nivel de Contaminación por CO2 por persona	Porcentaje de Especies en Peligro de Extinción	Índice de Salubridad. Escala de 0 a 1	Esperanza de vida (años)				
1											
2	Albania	71,4	1,3	(2)Medio	15	0,898	76,9				
3	Algeria	67,4	3,2	(2)Medio	13	0,838	73,1				
4	Angola	36,3	1,4	(2)Medio	4	0,49	51,1				
5	Antigua and Barbuda	69,8	5,2	(3)Alto	8	0,83	72,6				
6	Argentina	61	4,8	(3)Alto	9	0,882	75,9				
7	Armenia	60,4	1,8	(2)Medio	7	0,856	74,2				
8	Australia	65,7	19	(4)Muy Alto	22	0,976	81,9				
9	Austria	78,1	8,1	(3)Alto	11	0,96	80,9				
10	Azerbaijan	59,1	5,4	(3)Alto	8	0,8	70,7				
11	Bangladesh	44	0,3	(1)Bajo	9	0,772	68,9				
12	Belarus	65,4	6,5	(3)Alto	4	0,794	70,3				

Paso 2: Habilitar la opción Análisis de Datos en Excel. Ir a:
Archivo/Opciones/Complementos/Complementos de Excel/Ir/Herramientas para análisis.





Paso 3: Seleccionar la opción de Análisis de datos, para obtener los datos de la estadística descriptiva del índice de desempeño ambiental.



	A	B
1	Índice de Desempeño Ambiental 2014. Escala de 0 a 100	
2		
3	Media	50,88101351
4	Error típico	1,351440271
5	Mediana	50,365
6	Moda	63,79
7	Desviación estándar	16,44098048
8	Varianza de la muestra	270,3058391
9	Curtosis	-0,797931485
10	Coeficiente de asimetría	0,211806573
11	Rango	69,24
12	Mínimo	18,43
13	Máximo	87,67
14	Suma	7530,39
15	Cuenta	148
16		
17		

Opción 2: SPSS

Paso1: Copiar la base de datos en SPSS.

Paso 2: Ir a Analizar/Estadísticos descriptivos/Frecuencias



Paso 3. Elegir los estadísticos de interés, tendencia central (media, mediana, moda, suma), Dispersión (desviación estándar, varianza, rango, mínimo, máximo, media de error estándar), Distribución (asimetría, curtosis).

Frecuencias: Estadísticos

Valores percentiles

☐ Cuartiles

☐ Puntos de corte para: 10 grupos iguales

☐ Percentiles:

Añadir

Cambiar

Eliminar

Tendencia central

☒ Media

☒ Mediana

☒ Moda

☒ Suma

☐ Los valores son puntos medios de grupos

Dispersión

☒ Desviación estándar

☒ Varianza

☒ Rango

☒ Mínimo

☒ Máximo

☒ Media de error estándar

Distribución

☒ Asimetría

☒ Curtosis

Continuar Cancelar Ayuda

Estadísticos

Índice

N	Válido	148
	Perdidos	0
Media		50,88
Error estándar de la media		1,351
Mediana		50,36
Moda		64
Desviación estándar		16,441
Varianza		270,306
Asimetría		,212
Error estándar de asimetría		,199
Curtosis		-,798
Error estándar de curtosis		,396
Rango		69
Mínimo		18
Máximo		88
Suma		7530

Conclusión: Los países tienen en promedio un índice de desempeño ambiental de 50.88 puntos. La moda es de 63.79 puntos, que es el valor mayor frecuencia. La mediana corresponde a 50.365 que es el valor para el cual la cantidad de países que tienen un índice menor 50.365 es igual a la cantidad de países que tienen un índice mayor a 50.365.

Utilizando la varianza y la desviación estándar se puede calcular el coeficiente de variación, así:

$$CV = \frac{\sigma}{\mu}$$

La curtosis es una medida de forma que hace referencia a la elevación o achatamiento de una distribución, generalmente se compara con una distribución normal. Para el caso del Índice de desempeño ambiental, la curtosis menor a cero indica que la distribución es plana y tiene colas cortas. Existen tres tipos de distribuciones de acuerdo a su curtosis: mesocúrtica, leptocúrtica y platicúrtica, con niveles de elevación alto, medio y reducido, respectivamente.

El coeficiente de asimetría indica cuán simétrica o asimétrica es una distribución. Por ejemplo, una distribución normal es simétrica, de manera que el valor de su asimetría es cero. En este caso el índice de desempeño ambiental tiene un valor de asimetría positivo mayor a cero, que indica que la gráfica de su distribución tiene una cola positiva orientada hacia la derecha.

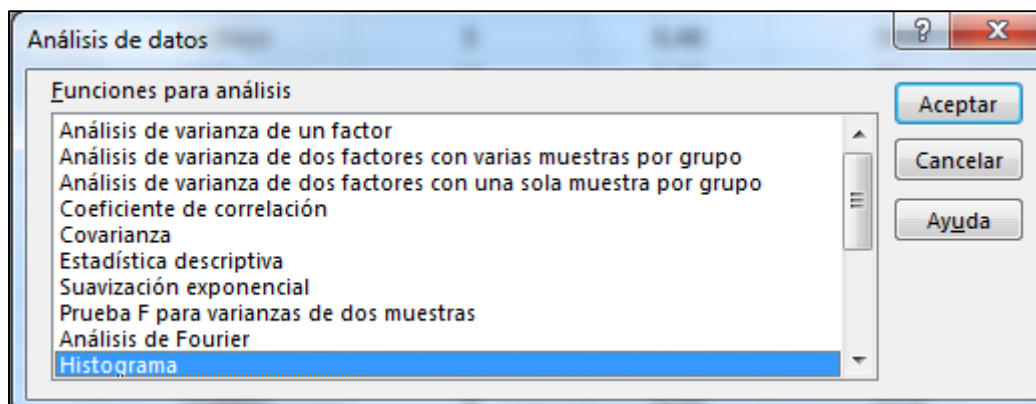
- b. Realice un histograma para la variable **Índice de desempeño ambiental** y otro para la variable **Nivel de contaminación por CO₂ por persona (Bajo, Medio, Alto, Muy Alto)**. ¿Qué puede concluir?

NOTA: El nivel de contaminación depende de la Emisión anual de CO₂ de la siguiente forma:

Nivel	Emisión anual CO ₂ por persona (toneladas)
Bajo	[0,1]
Medio	(1,4]
Alto	(4,10]
Muy Alto	(10,infinito)


Opción 1: Excel


Paso 1: Hacer el histograma de la variable Índice de desempeño ambiental con los datos de la tabla y generar el histograma.



Histograma


Entrada

Rango de entrada: 

Rango de clases: 

☒ Rótulos

Opciones de salida

☐ Rango de salida: 

☒ En una hoja nueva:

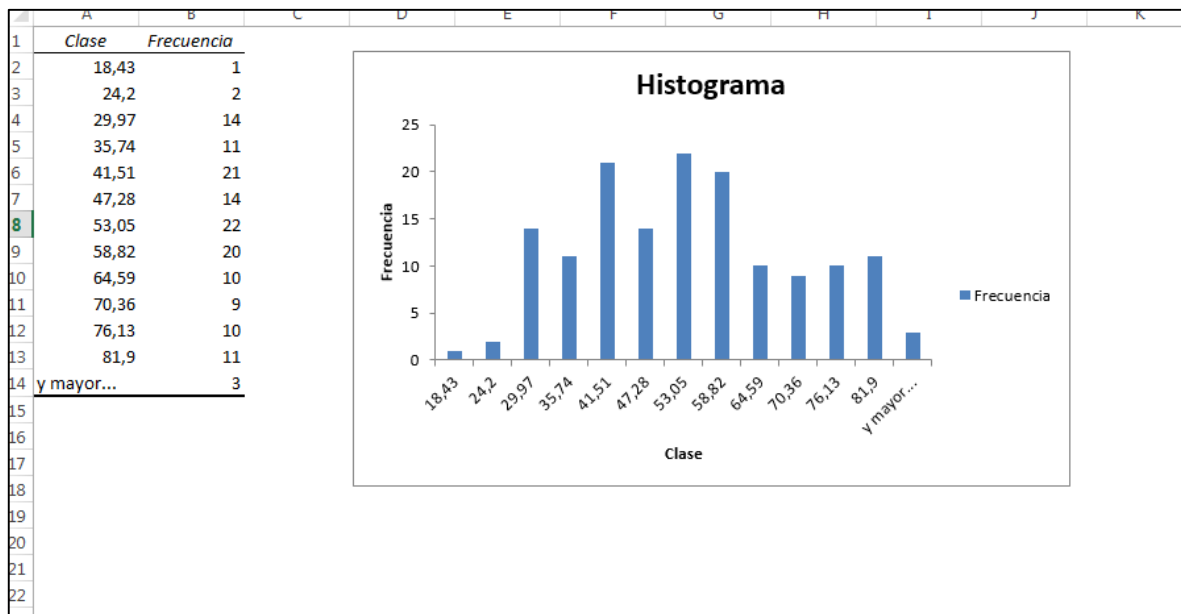
☐ En un libro nuevo

☐ Pareto (Histograma ordenado)

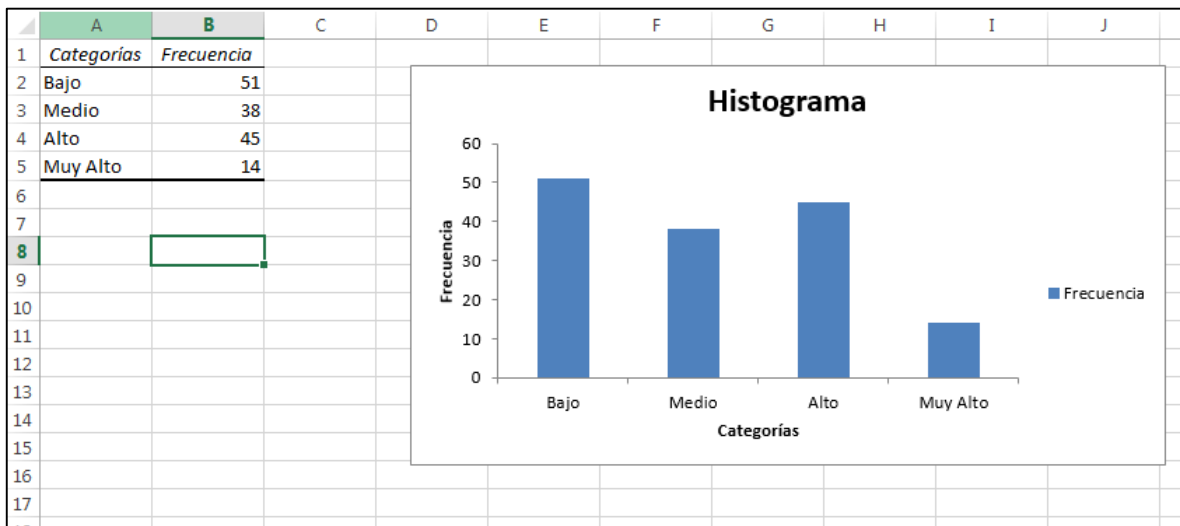
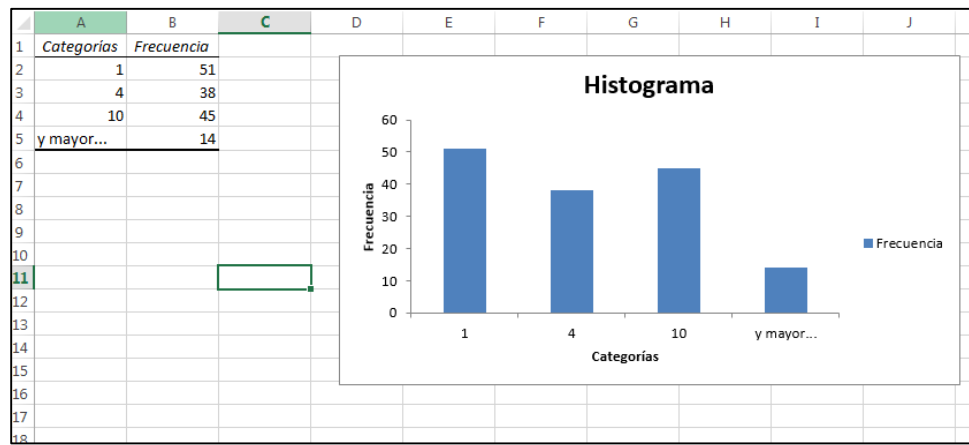
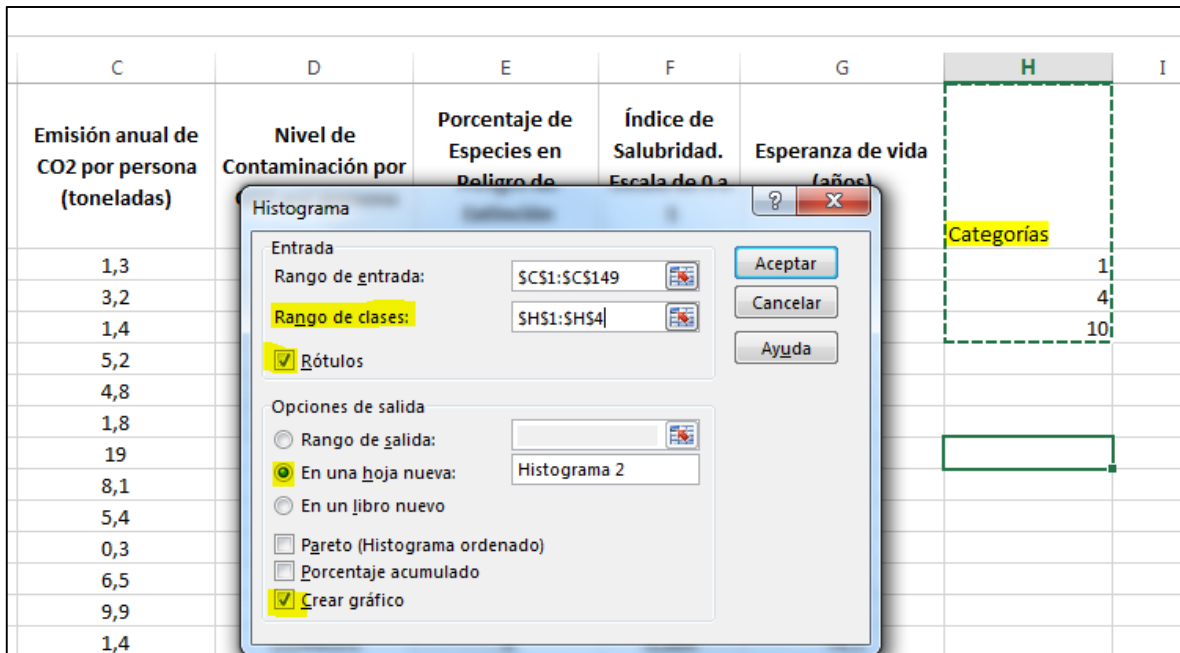
☐ Porcentaje acumulado

☒ Crear gráfico

Aceptar Cancelar Ayuda



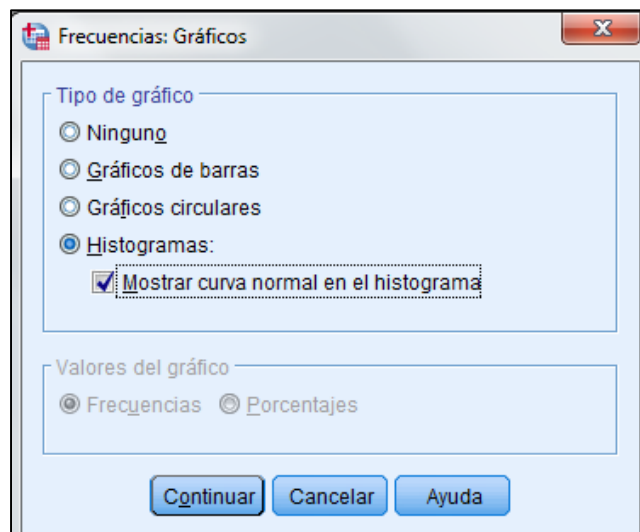
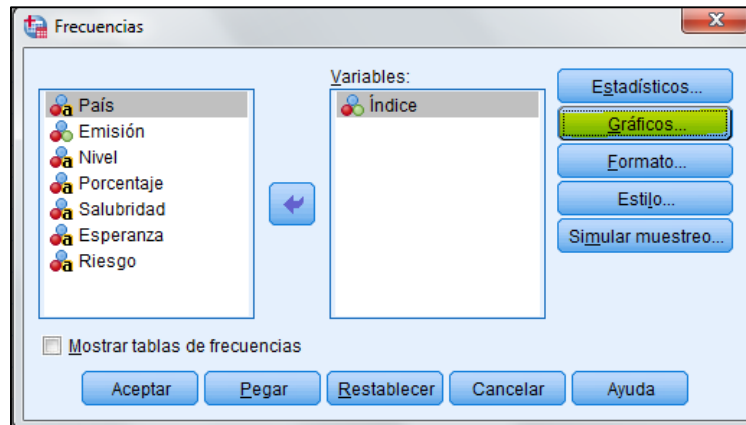
Para realizar el histograma del nivel de contaminación, es necesario ingresar los valores límite que separan cada categoría. En el caso de la categoría Muy Alto (10 a infinito), no se pone en Excel, dado que Excel ya asume que el límite de la última categoría va hasta infinito.

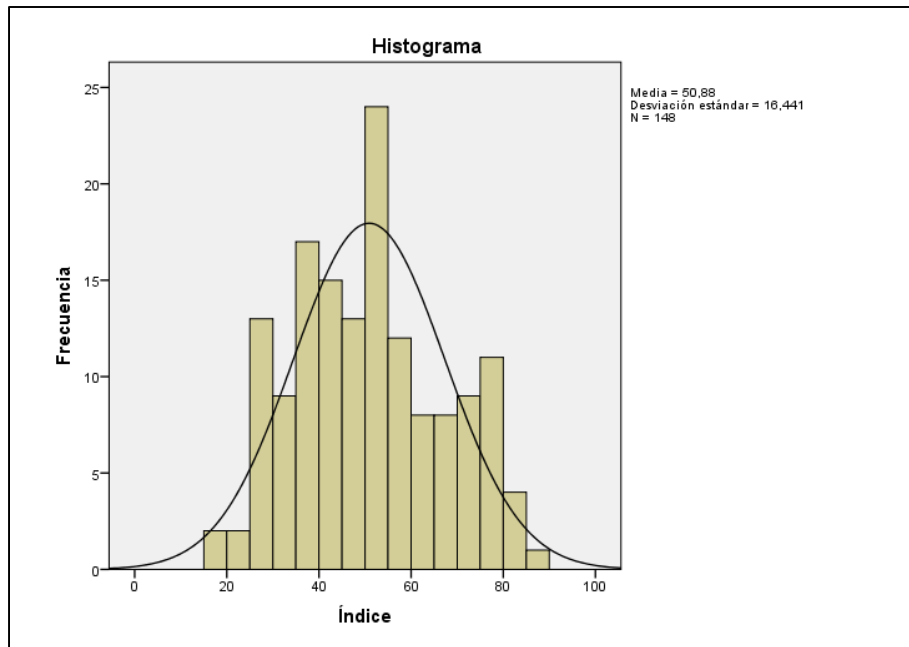


Opción 2. SPSS

Paso1: Copiar la base de datos en SPSS.

Paso 2: Ir a Analizar/Estadísticos descriptivos/Frecuencias/Gráficos





Paso 3: Concluir.

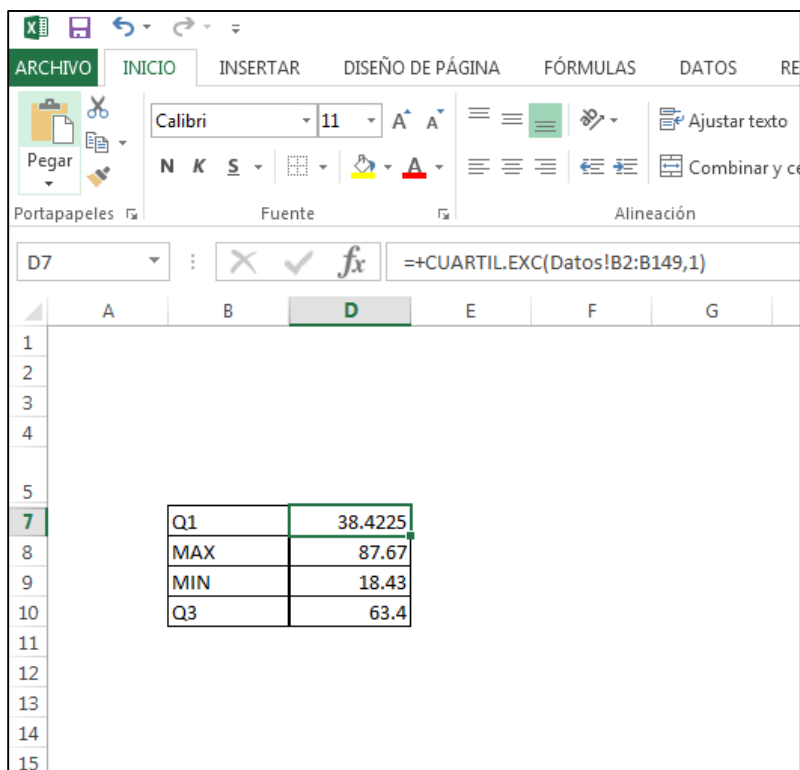
A partir del primer histograma realizado podemos concluir, que no hay simetría en la distribución de las frecuencias del índice de desempeño ambiental de los 148 países. Los índices ambientales más frecuentes son 53.5, 58.82 y 41.51. También podemos observar que tan solo un país tiene el índice de desempeño ambiental más bajo (18.43).

El segundo histograma de categorías nos permite observar que 14 países tienen nivel Muy Alto de contaminación. En el caso de nivel bajo, medio y alto, se tiene como resultado 51, 38 y 48 países por cada categoría respectivamente.

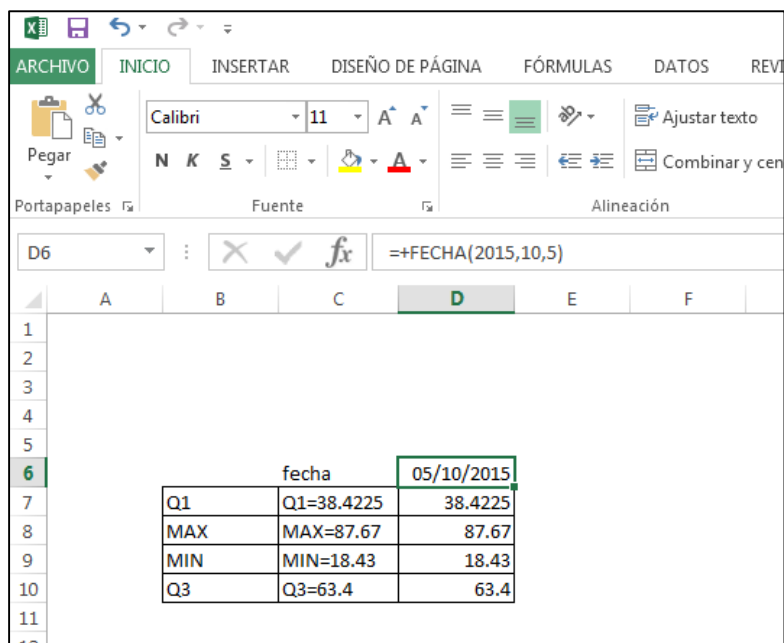
c. Realice un diagrama de caja para la variable **Índice de Desempeño Ambiental**

Opción 1: Excel

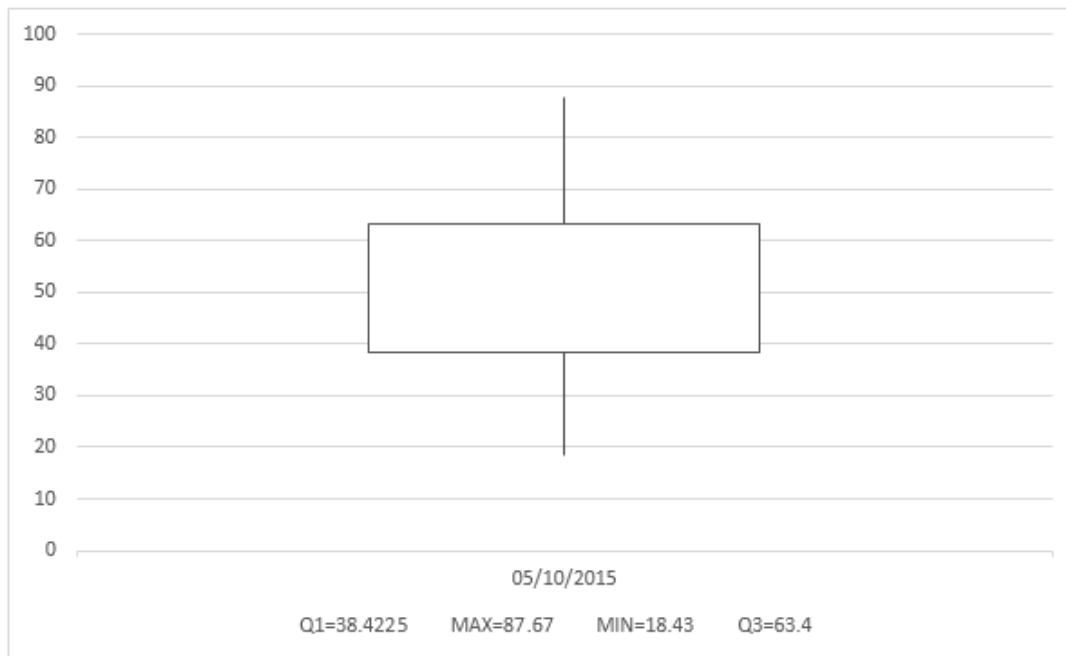
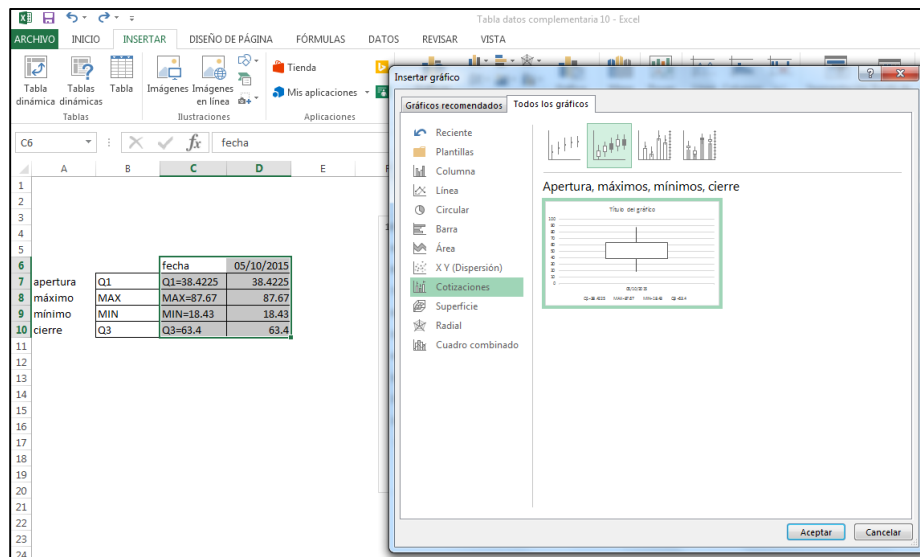
Paso 1: Crear una tabla con los valores de los cuartiles Q1 y Q3 (25, 75), máximo y mínimo usando las funciones de Excel: CUARTIL.EXC, MIN, MAX, para la columna de la variable índice de desempeño ambiental. No olvide ordenar previamente los datos de la columna índice de desempeño de menor a mayor.



Paso 2: Insertar una celda con la fecha, y rotular los nombres de las estadísticas de acuerdo a como se quiere que aparezcan en la gráfica.



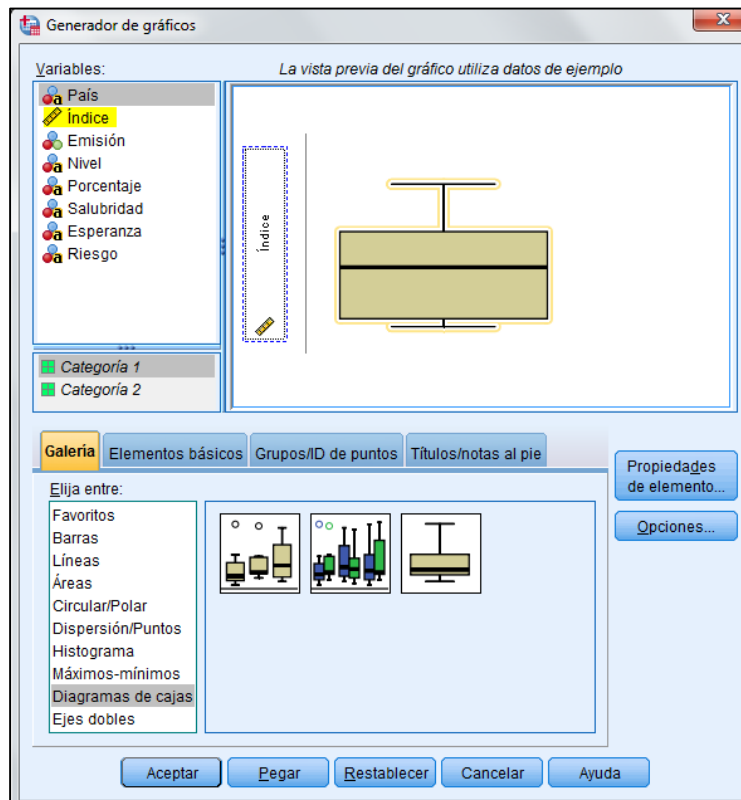
Paso 2: Seleccionar el rango de datos incluyendo la celda fecha y los rótulos, e insertar un gráfico de cotización apertura, máximos, mínimos, cierre.

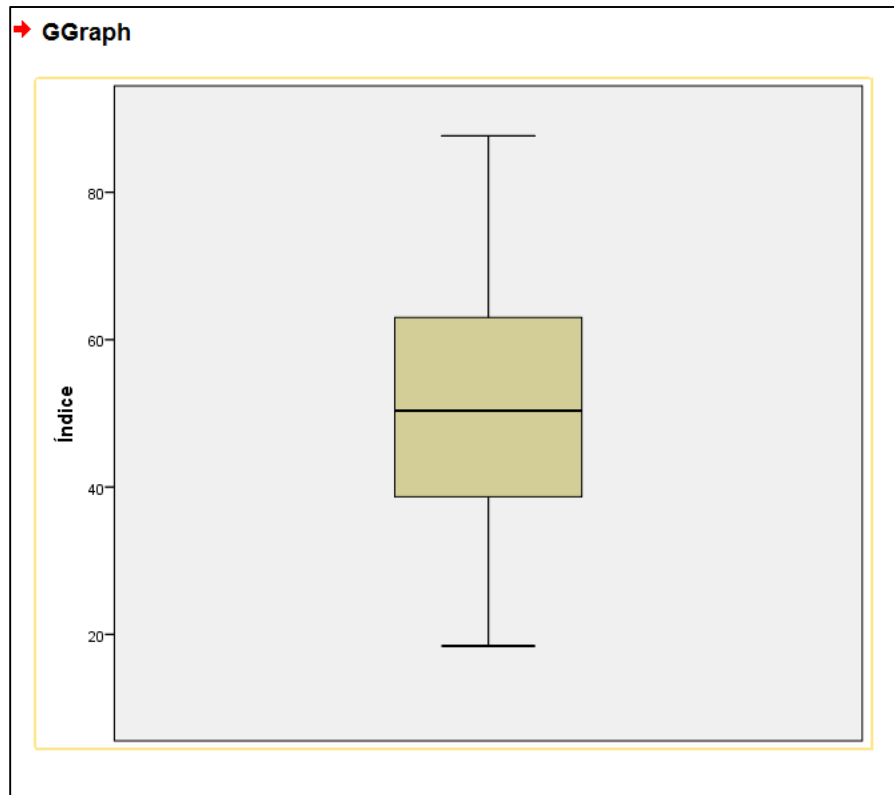


Opción 2: SPSS

Paso1: Copiar la base de datos en SPSS.

Paso 2: Ir a Gráficos/ Generador de gráficos/ Diagramas de caja

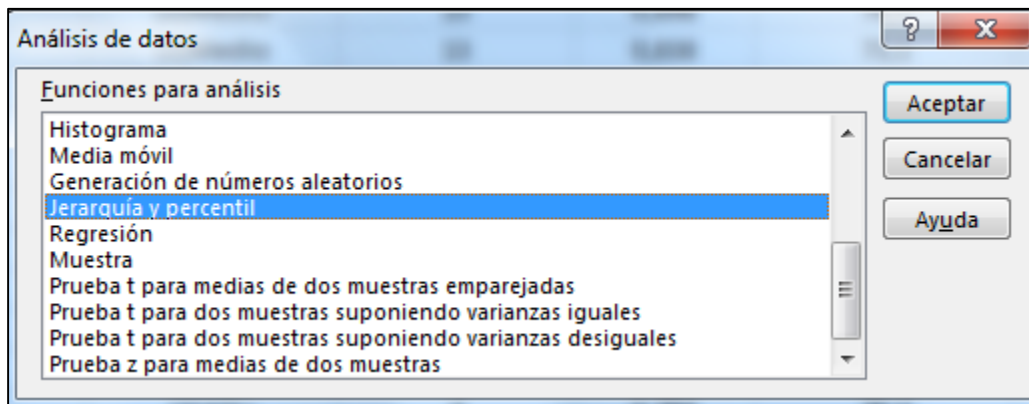




- d. Realice un análisis de percentiles para la variable **Índice de Desempeño Ambiental**. ¿Cuáles son los valores correspondientes a los percentiles 20.4, 75.5, y 95.2%? ¿A qué países corresponden estos valores?

Opción 1: Excel Análisis de Percentiles: Datos/ Análisis de Datos/ Jerarquía y Percentil

Paso 1: Generar los datos para los percentiles, mediante la herramienta de análisis de datos de Excel, seleccionando el rango de datos de interés:



Jerarquía y percentil

Entrada

Rango de entrada:

Agrupado por: ☒ Columnas ☐ Filas

☒ Rótulos en la primera fila

Opciones de salida

☐ Rango de salida:

☒ En una hoja nueva:

☐ En un libro nuevo

Aceptar Cancelar Ayuda

Se obtienen los siguientes resultados:

	A	B	C	D	E	F	G
1	Posición	Índice de Desempeño Ambiental 2014. Escala de 0 a 100	Jerarquía	Porcentaje			
2	129	87,67	1	100,00%			
3	82	83,29	2	99,30%			
4	7	82,4	3	98,60%			
5	38	81,44	4	97,90%			
6	54	80,47	5	97,20%			
7	123	79,79	6	96,50%			
8	8	78,32	7	95,90%			
9	128	78,09	8	95,20%			
10	103	78,04	9	94,50%			
11	98	77,75	10	93,80%			
12	140	77,35	11	93,10%			
13	39	76,92	12	92,50%			
14	120	76,43	13	91,80%			
15	99	76,41	14	91,10%			
16	112	75,8	15	90,40%			
17	49	75,72	16	89,70%			
18	68	74,67	17	89,10%			
19	46	74,66	18	88,40%			
20	119	74,45	19	87,70%			
21	70	74,36	20	87,00%			
22	56	73,28	21	86,30%			
23	25	73,14	22	85,70%			

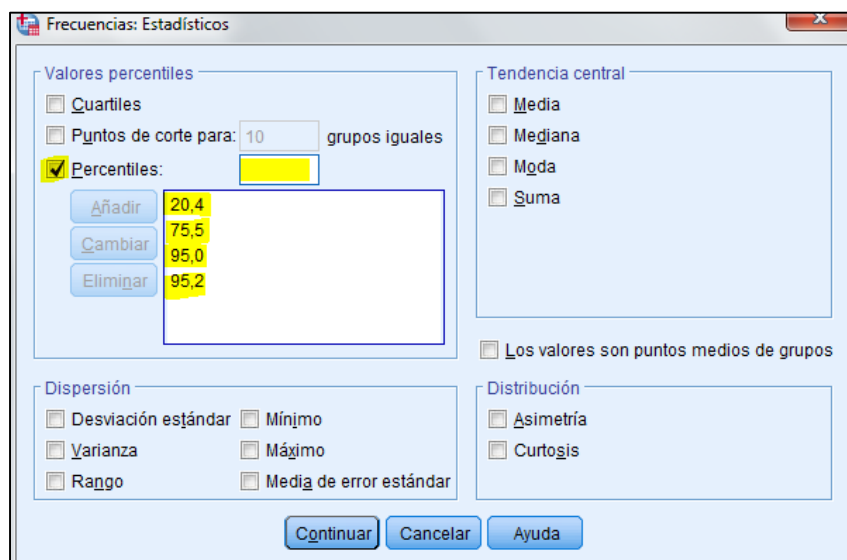
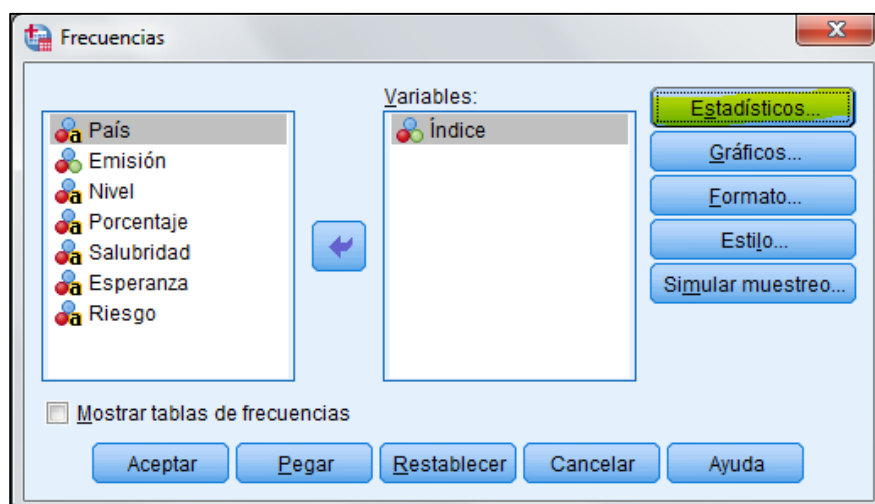
Paso 2: Ordenar los datos de la tabla de manera decreciente, y generar los percentiles correspondientes.

Percentil	Índice	Posición	País
20.4	36.28	101	Nigeria
75.5	63.79	122	Korea
95.2	78.09	128	Sweden

Opción 2: SPSS

Paso1: Copiar la base de datos en SPSS.

Paso 2: Ir a Analizar/Estadísticos descriptivos/Frecuencias/Estadísticos



Estadísticos

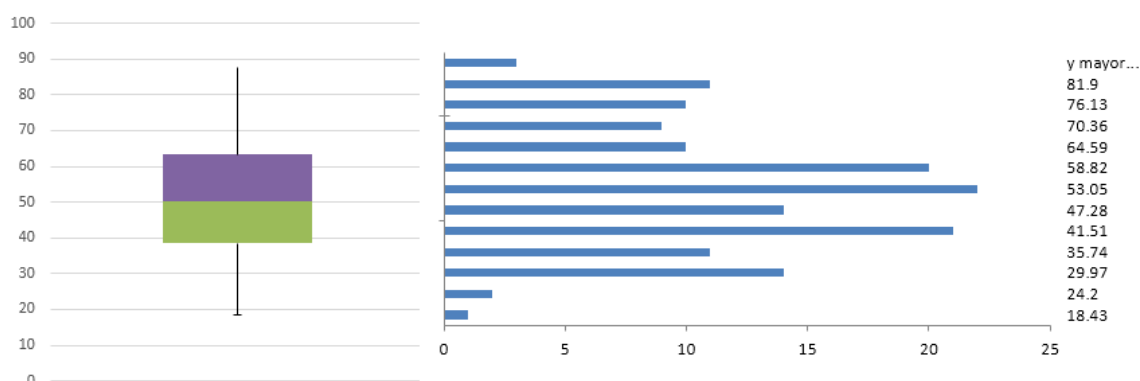
Índice		
N	Válido	148
	Perdidos	0
Percentiles	20,4	36,23
	75,5	63,79
	95	78,22
	95,2	78,29

Para el caso del percentil 20.4% se puede interpretar de la siguiente forma: de los 147 países, el 20.1% tienen un Índice de Desempeños Ambiental inferior al Índice de Nigeria, es decir,

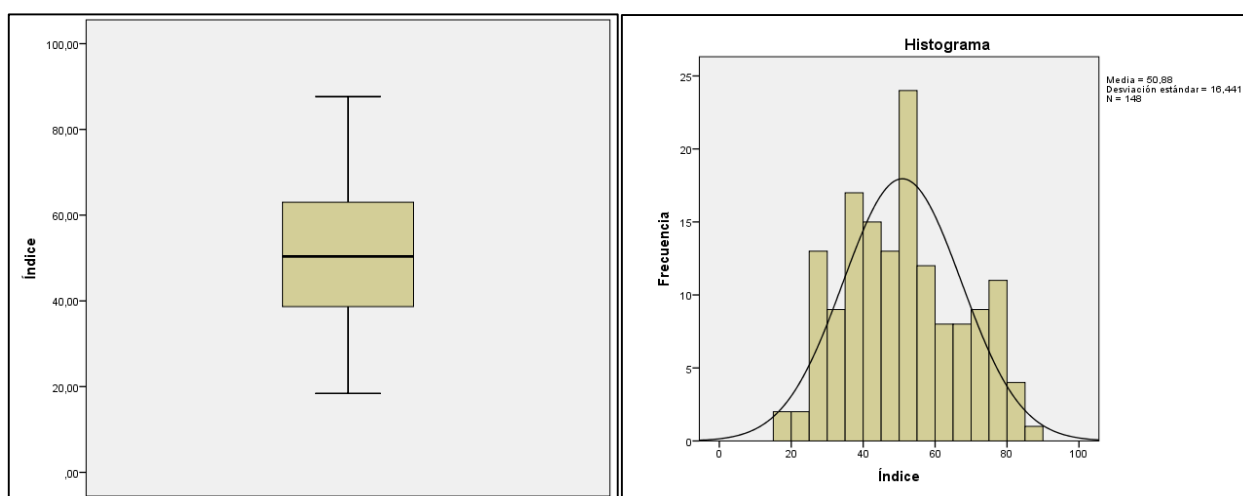
($147 \times 0.204 = 29.99$) 30 países presentan un Índice menor a 36.28. Los demás valores se interpretan de forma similar.

- e. Compare el histograma, los percentiles y el diagrama de caja para la variable índice de eficiencia Ambiental. Concluya brevemente.

En excel: acomodando los gráficos y las escalas de cada figura



En SPSS



N	Válido	148
	Perdidos	23
Percentiles	0	.
	25	38,4225
	50	50,3650
	75	63,4000
	100	87,6700

El histograma representa las frecuencias de la variable índice de desempeño ambiental, y el diagrama de caja representa los tres cuartiles (25,50 y 75) y los valores máximo y mínimo de los datos del índice de desempeño ambiental de los 148 países. Los datos tienen una distribución

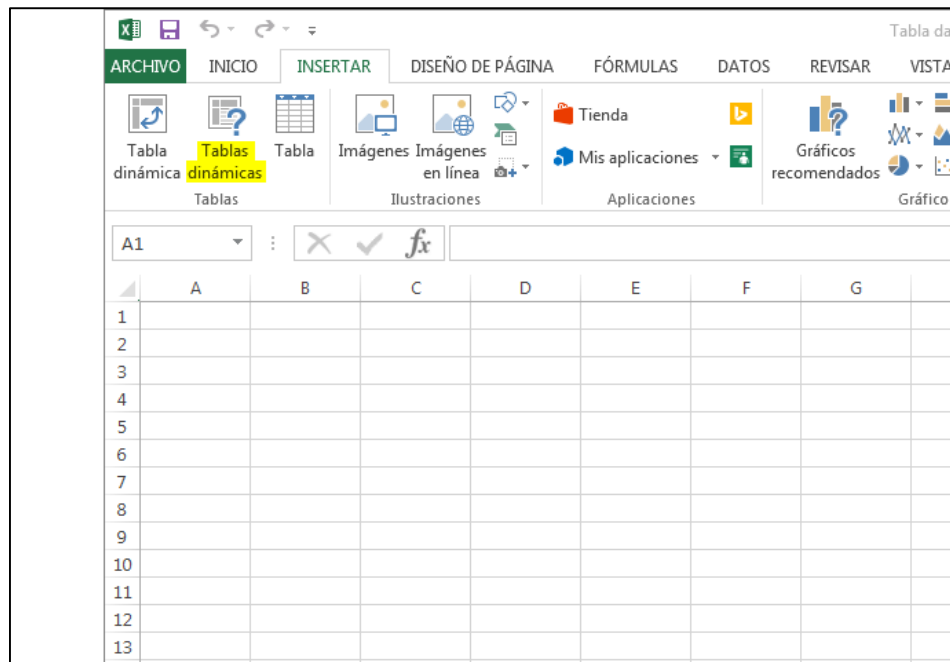
levemente asimétrica hacia la derecha, lo cual se visualiza en el histograma, y también en el diagrama de caja ya que los bigotes (datos por encima del tercer cuartil y por debajo del primer cuartil) y los rectángulos por encima y debajo de la mediana tienen longitudes similares pero están un poco más agrupados en sus valores inferiores. En este caso no hay datos atípicos.

- f. Cree en una nueva columna la variable **Riesgo Especies**, la cual tomará los valores de 0, 1 o 2 dependiendo si el país presenta Bajo, Alto o Muy Alto riesgo de pérdida de especies respectivamente. Esto depende de la variable **Porcentaje de Especies en Peligro de Extinción** de la siguiente forma:

Riesgo Especies	Porcentaje de Especies en Peligro de Extinción
0 (Bajo)	[0,8)
1 (Alto)	[8,20)
2 (Muy Alto)	[20,100]

Realice una tabla de contingencia o tabla dinámica para las variables **Nivel de Contaminación por CO₂ por persona** y **Riesgo Especies**. En esta tabla se quiere analizar el número de países que pertenecen a cada categoría para cada variable. ¿Cuál es la probabilidad de que un país tenga nivel alto de riesgo de pérdida de especies y presente alto nivel de contaminación por CO₂? ¿Qué puede concluir?

Paso 1: Insertar tabla dinámica



Crear tabla dinámica

Seleccione los datos que desea analizar

☒ **Seleccione una tabla o rango**

Tabla o rango:

☐ **Utilice una fuente de datos externa**

Nombre de conexión:

A	B	C	D	E	F	G	H	I
Cuenta de Riesgo de especies	Etiquetas de columna							
Etiquetas de fila	0	1	2	Total general				
(1)Bajo	31	19	1	51				
(2)Medio	15	23		38				
(3)Alto	12	32	1	45				
(4)Muy Alto	8	4	2	14				
Total general	66	78	4	148				

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

☐ País

☐ Índice de Desempeño Ambiental 2014. Escala de 0 a 100

☐ Emisión anual de CO2 por persona (toneladas)

☒ Nivel de Contaminación por CO2 por persona

☐ Porcentaje de Especies en Peligro de Extinción

☐ Índice de Salubridad. Escala de 0 a 1

☐ Esperanza de vida (años)

☒ Riesgo de especies

MÁS TABLAS

Arrastrar campos entre las áreas siguientes:

FILTROS

Nivel de Contaminación por CO2 por pe...

COLUMNAS

Riesgo de especies

FILAS

VALORES

Cuenta de Riesgo de especies

Cuenta de Riesgo de especies				
Etiquetas de columna				
Etiquetas de fila				
	0	1	2	Total general
(1)Bajo	31	19	1	51
(2)Medio	15	23		38
(3)Alto	12	32	1	45
(4)Muy Alto	8	4	2	14
Total general	66	78	4	148

$$P(\text{Alto} \cap 1) = \frac{32}{148} = 0.216$$

Hay 4 países con riesgo muy alto de pérdida de especies, dos de los cuales tienen un muy alto nivel de emisión de CO2. De los 78 países con riesgo medio de perdida de especies, 4 tienen niveles altos de emisión de CO2 y 32 nivel alto de emisiones. En el caso de los 66 países con bajo nivel de riesgo por pérdida de especies, 31 de estos presentan niveles bajos de emisión de CO2.Lo cual podría indicar algún nivel de asociación entre el nivel de emisión de CO2 y el nivel de pérdida de especies.