

Data utilizada

Se utilizaron los siguientes datos para realizar los ejercicios.

Grupo de control	Nivel glucosa basal	Nivel glucosa 60 min
1	90	136
1	82	151
1	80	148
1	75	138
1	74	141
1	97	157
1	103	172
1	76	154
1	65	131
1	89	156
1	83	147
1	77	141
1	92	140
1	90	134
1	82	146
1	78	151
1	79	143
1	104	161
1	81	150
1	101	149
1	75	133
1	89	170
1	76	144
1	85	149
1	99	158
1	79	150
1	90	138
1	86	150
1	94	152
1	91	155
1	82	145
1	81	144
1	97	159
1	75	139
1	79	146
2	99	198
2	96	191
2	93	190
2	88	174
2	79	156

2	79	184
2	97	159
2	102	161
2	104	182
2	86	170
2	89	197
2	81	179
2	80	183
2	90	178
2	91	169
2	88	172
2	77	152
2	96	181
2	81	160
2	98	160
2	90	158
2	102	164
2	88	172
2	77	174
2	106	172
2	95	169
2	99	172
2	98	173
2	84	188
2	92	160

Ejercicio 1

a) Obtener, usando algún programa estadístico, las medidas de centralización y dispersión para cada uno de los dos grupos de control para el nivel de glucosa basal, especificando para cada uno de los casos si la media es o no representativa.

Jóvenes

<i>Nivel glucosa basal</i>	
Mean	85.02857143
Standard Error	1.585377358
Median	82
Mode	90
Standard Deviation	9.379218939
Sample Variance	87.9697479
Kurtosis	-0.463968504
Skewness	0.336054724

Range	39
Minimum	65
Maximum	104
Sum	2976
Count	35

<i>Regression Statistics</i>	
Multiple R	0.610539498
R Square	0.372758479
Adjusted R Square	0.35375116
Standard Error	7.67597004
Observations	35

ANOVA

		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression		1	1155.508684	1155.508684	19.61131303	9.8273E-05		
Residual		33	1944.37703	58.92051606				
Total		34	3099.885714					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	95.093	12.004	7.921	0.000	70.669	119.516	70.669	119.516
X Variable 1	0.622	0.140	4.428	0.000	0.336	0.907	0.336	0.907

Para los jóvenes al buscar si la media es representativa se obtuvo un 11.030667%, lo que la ubica como una media representativa.

Adultos

<i>Nivel glucosa basal</i>	
Mean	90.83333333
Standard Error	1.544797225
Median	90.5
Mode	88
Standard Deviation	8.461202871
Sample Variance	71.59195402
Kurtosis	-1.025807155
Skewness	-0.08006821
Range	29
Minimum	77
Maximum	106

Sum	2725
Count	30

Regression Statistics	
Multiple R	0.069970255
R Square	0.004895837
Adjusted R Square	-
Standard Error	12.49622088
Observations	30

ANOVA

		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression		1	21.51165342	21.51165342	0.137757866	0.713313872		
Residual		28	4372.355013	156.1555362				
Total		29	4393.866667					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	164.0207	25.0154	6.5568	0.0000	112.7791	215.2624	112.7791	215.2624
X Variable 1	0.1018	0.2743	0.3712	0.7133	(0.4600)	0.6636	(0.4600)	0.6636

Para los adultos al buscar si la media es representativa se obtuvo un 9.315086%, lo que la ubica como una media representativa.

Para ambos grupos de control sus medias son representativas ya que se encuentran por debajo del 30%. Esto se obtuvo al dividir la desviación estándar con la media.

b) Estudiar la simetría y la curtosis del nivel de glucosa basal en los adultos (grupo de control 2).

Para obtener la curtosis se utilizó la formula en Excel KURT con la data del grupo 2, obteniendo como resultado -1.025807155. Significando que es una distribución plana o platicurtica.

Con respecto a la simetría se utilizó la formula en Excel Skew con la data del grupo 2, obteniendo el resultado -0.08006821. Significa que es una distribución negativa que se sesga a la izquierda.

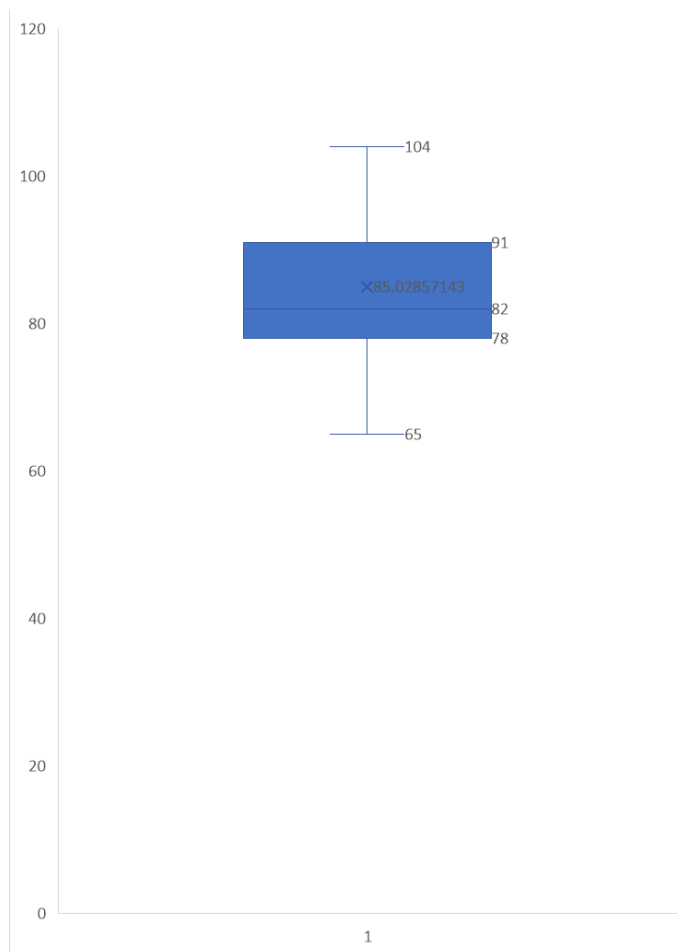
c) Indicar para cada una de las variables de estudio (nivel glucosa basal y nivel glucosa pasados 60 min) y en el grupo de control 1 el valor de los cuartiles y su significado y obtener el box- plot (diagrama de cajas) correspondiente. Estudiar la presencia de valores atípicos.

	Basal	60 minutos
Q1	78.5	141

Q2	82	148
Q3	90.5	153
Q4	104	172

Al utilizar los cuartiles y el boxplot podemos ver que en Basal el Q1 ya entra en boxplot y continua hasta Q3. En otras palabras, los valores típicos se encuentran desde Q1, por lo que cerca del 75% de los datos típicos se encuentran en estos 3 cuartiles. Y esto mismo pasa con el nivel de glucosa luego de 60 minutos. Adicionalmente podemos ver que la media para Basal se encuentra un poco por arriba del Q2, y para la glucosa luego de los 60 minutos su media se encuentra por debajo de Q2.

El Boxplot para el Nivel de Glucosa Basal es el siguiente:

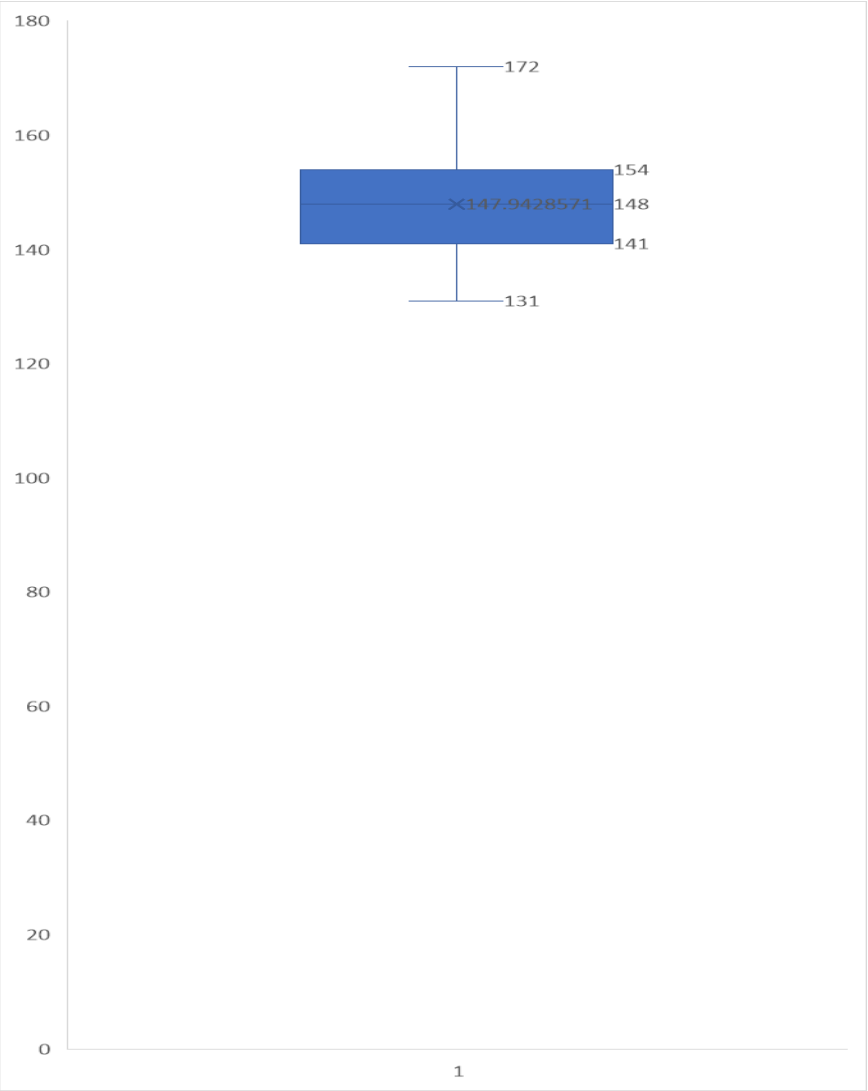


Estos son los valores atípicos para el Nivel de Glucosa Basal:

Basal valores atípicos	
Por debajo	75
	74
	76
	65

	77
	75
	76
	75
Por Arriba	
	97
	103
	92
	104
	101
	99
	94
	97

El Boxplot para el Nivel de Glucosa luego de 60 minutos:



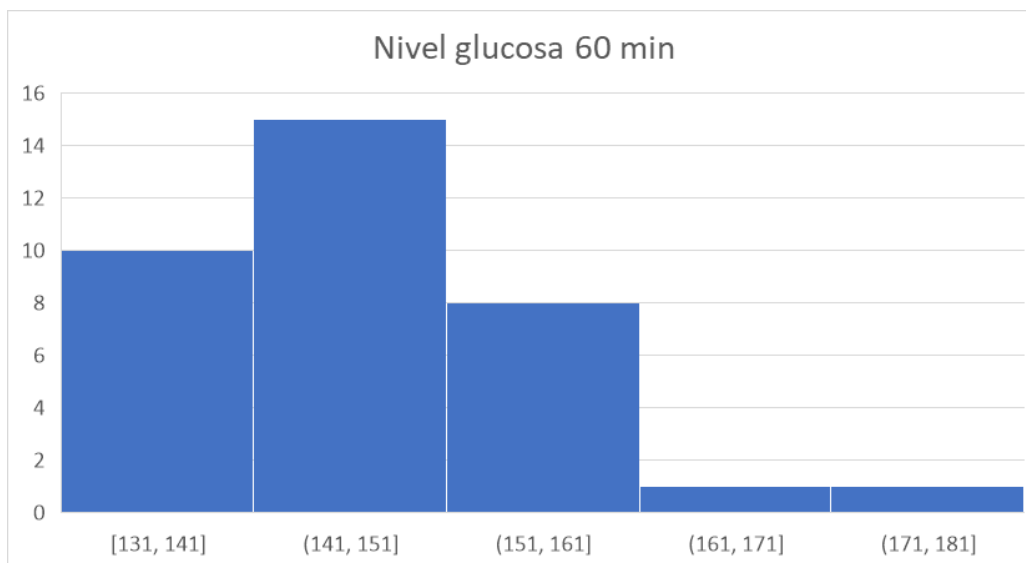
Estos son los valores atípicos para el Nivel de Glucosa luego de 60 minutos:

60 minutos valores atipicos	
Por debajo	136
	138
	131
	140
	134
	133
	138
	139
Por arriba	157
	172
	156
	161
	170
	158
	155
	159

d) Estudiar la normalidad de los datos de cada uno de los grupos de control estudiados para el nivel de glucosa pasados 60 minutos.

Jóvenes

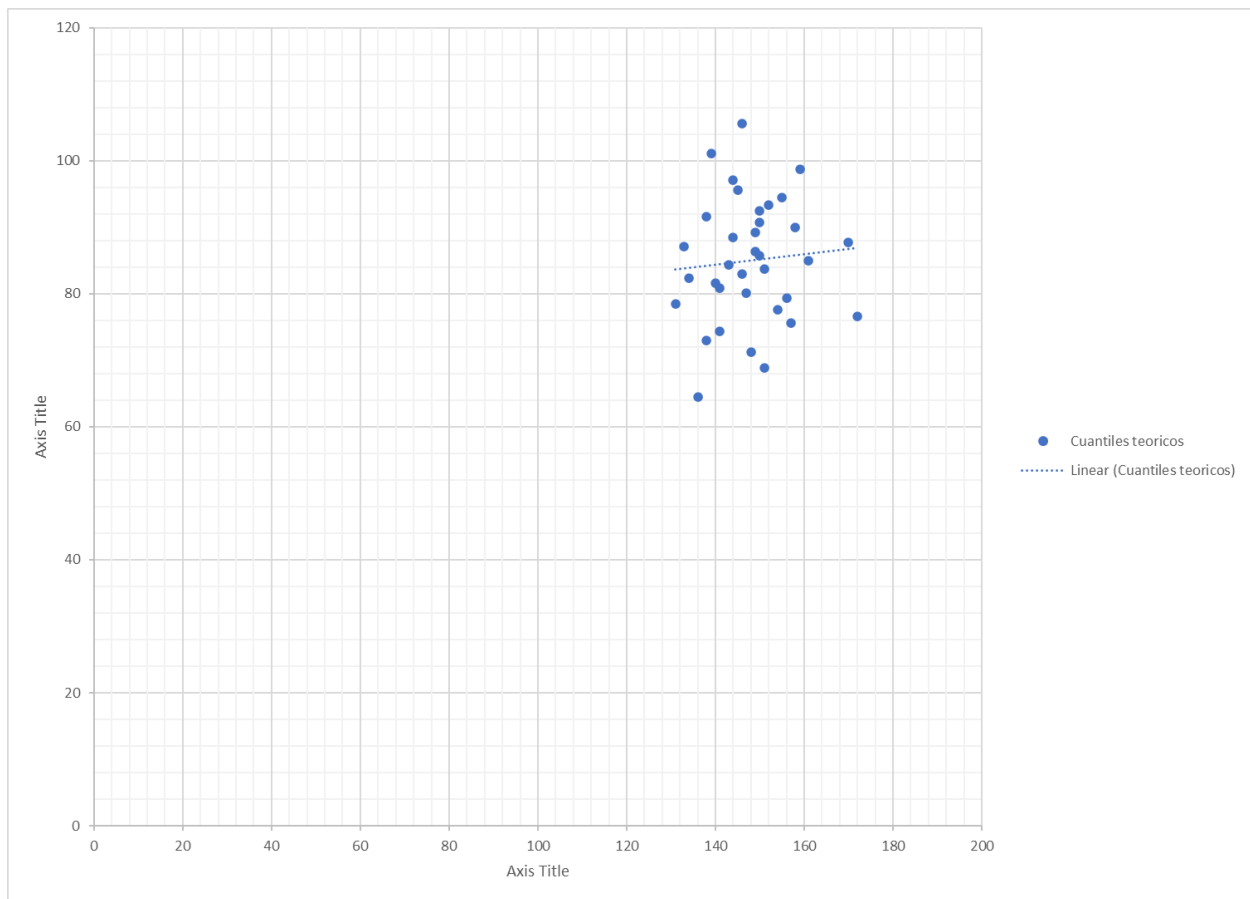
Según el histograma se tiene normalidad al confirmar una forma de campana, aunque se mantiene más elevada del lado izquierdo.



Para confirmar se utilizó el qq plot y se obtuvo los siguientes cuantiles teóricos:

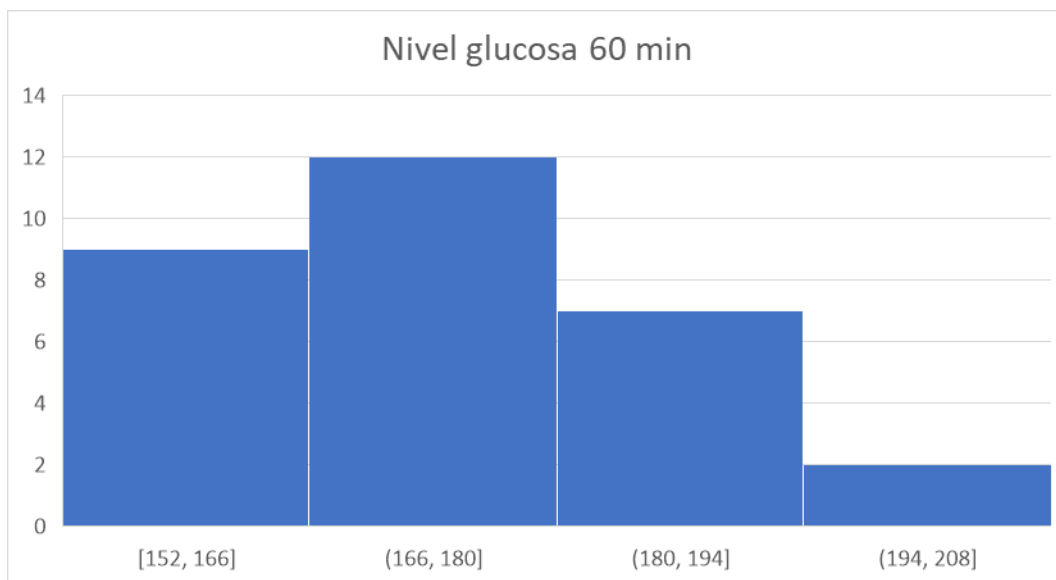
Nivel glucosa 60 min	Cuantiles teóricos
136	64.49418074
151	68.91083817
148	71.28582289
138	73.00861871
141	74.40032162
157	75.59054492
172	76.64541962
154	77.60361961
131	78.48992928
156	79.3213342
147	80.11010421
141	80.86549901
140	81.59477975
134	82.30384482
146	82.99765047
151	83.68050348
143	84.35627573
161	85.02857143
150	85.70086712
149	86.37663938
133	87.05949239
170	87.75329804
144	88.4623631
149	89.19164385
158	89.94703865
150	90.73580866
138	91.56721358
150	92.45352325
152	93.41172324
155	94.46659794
145	95.65682124
144	97.04852414
159	98.77131997
139	101.1463047
146	105.5629621

Al observar la grafica obtenida por los datos, se observa una línea recta positiva que indica que los datos tienen normalidad.



Adultos

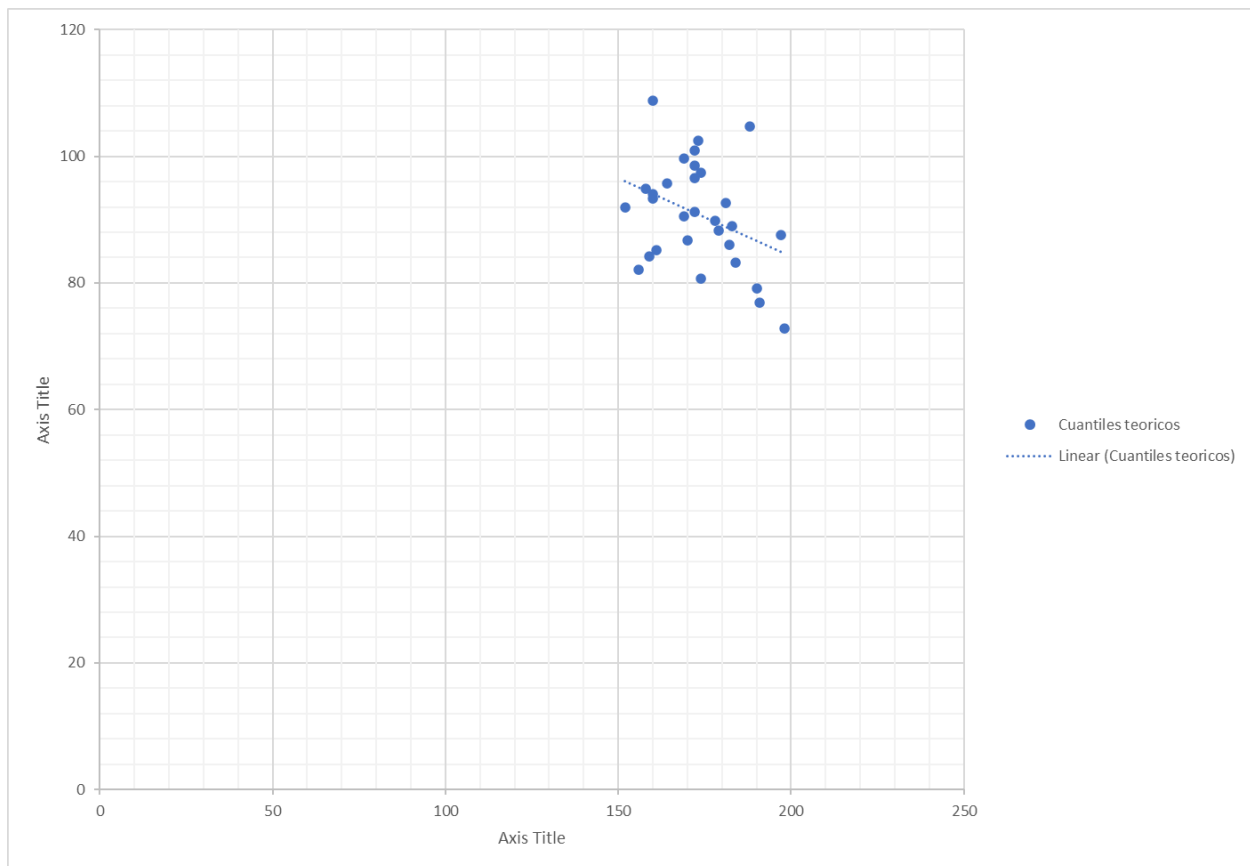
Según el histograma se tiene normalidad al obtener una campana, con el lado izquierdo más elevado. Podemos observar que ambos Jóvenes y adultos tienen una campana similar.



También se utilizó el qq plot para confirmar que los datos son normales; se obtuvieron los siguientes cuantiles teóricos:

Nivel glucosa 60 min	Cuantiles teóricos
198	72.82751089
191	76.9158931
190	79.13153945
174	80.74913492
156	82.06386016
184	83.19511112
159	84.20397771
161	85.12633872
182	85.98533867
170	86.79699749
197	87.5730587
179	88.32257428
183	89.052856
178	89.77008718
169	90.47974561
172	91.18692106
152	91.89657948
181	92.61381067
160	93.34409238
160	94.09360797
158	94.86966918
164	95.681328
172	96.54032794
174	97.46268896
172	98.47155554
169	99.6028065
172	100.9175317
173	102.5351272
188	104.7507736
160	108.8391558

Al observar la línea de los datos, podemos ver que es recta y es negativa. En otras palabras, aunque muestra la normalidad de los datos, también nos indica que el comportamiento teórico es opuesto a lo observado.



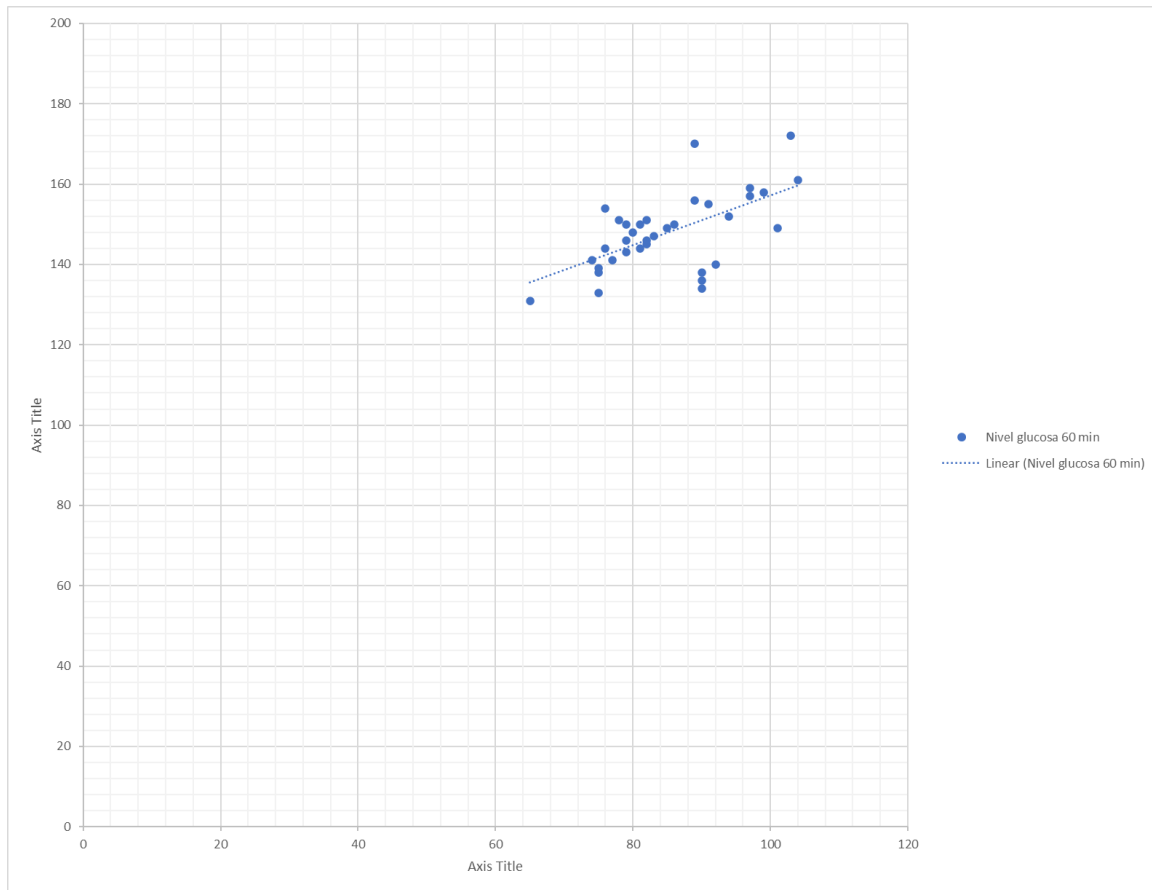
Ejercicio 2

a) Estudiar la relación lineal existente entre estas dos variables de estudio gráficamente y mediante algún valor estadístico de forma razonada.

Al obtener la correlación podemos observar que se tiene un alto nivel de correlación entre los datos y que el Nivel de Glucosa luego de 60 minutos es dependiente de la Glucosa Basal ya que de manera inversa no se nota una correlación

	<i>Nivel glucosa basal</i>	<i>Nivel glucosa 60 min</i>
Nivel glucosa basal	1	
Nivel glucosa 60 min	0.610539498	1

Al realizar un diagrama de dispersión podemos, podemos ver como los datos se encuentran cerca de la línea, mostrando su alto nivel de correlación.



b) Obtener un modelo lineal que explica el nivel de glucosa en sangre a los 60 minutos en función del nivel basal del paciente y realizar la estimación para un paciente cuyo nivel basal es 83 mg/Dl

Para obtener el modelo lineal se obtuvo la regresión, Anova y los coeficientes.

<i>Regression Statistics</i>	
Multiple R	0.610539498
R Square	0.372758479
Adjusted R Square	0.35375116
Standard Error	7.67597004
Observations	35

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	1155.508684	1155.508684	19.61131303	9.8273E-05
Residual	33	1944.37703	58.92051606		
Total	34	3099.885714			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	95.09281266	12.00448777	7.921438589	3.90479E-09	70.66949865	119.5161267	70.66949865	119.5161267
Nivel glucosa basal	0.621556303	0.140354758	4.428466217	9.8273E-05	0.336002401	0.907110204	0.336002401	0.907110204

Con esto se creó el siguiente modelo lineal:

Modelo	Coeficiente		Nivel de Glucosa		Nivel Basal
lineal	95.09281266	+	0.621556303	*	Paciente

Después de estimar con el nivel Basal de 83 se espera que su nivel de glucosa luego de 60 minutos sea de 146.68 mg/DL.

c) ¿Qué tanto por ciento del nivel de glucosa en sangre pasados 60 minutos queda no queda explicado por el anterior modelo?

Para obtener cuanto no queda explicado se debe restar $1-R^2$, utilizando este modelo el porcentaje no explicado es de 62.72%.

d) Si aumentásemos el nivel basal de un paciente en 5 mg/Dl ¿Qué variación experimentaría su nivel de glucosa al cabo de 60 minutos?

Se utiliza el modelo lineal anteriormente presentado y al realizar el calculo con los 5mg/Dl se obtiene el resultado de 98.20mg/DL este resultado se le debe restar el coeficiente anteriormente presentado. Esto nos otorga una diferencia o variación de 3.1078

Ejercicio 3

a) Se quiere estudiar si se puede admitir que el nivel medio de glucosa en sangre en el momento de la ingestión en los jóvenes es 88 mg/Dl. Obtener el intervalo de confianza al 95% y al 99% para el nivel medio de glucosa en sangre de los jóvenes y posteriormente contesta a la cuestión planteada con los resultados obtenidos o con un contraste de hipótesis.

Se realiza de nuevo el modelo de estadística descriptiva para el Nivel de Glucosa Basal:

<i>Nivel glucosa basal</i>	
Mean	85.02857143
Standard Error	1.585377358
Median	82
Mode	90
Standard Deviation	9.379218939
Sample Variance	87.9697479
	-
Kurtosis	0.463968504
Skewness	0.336054724
Range	39
Minimum	65
Maximum	104
Sum	2976

Count	35
-------	----

Y se realizan los cálculos del intervalo de confianza con los siguientes resultados:

	Nivel de confianza	Grado de error	Media	Rango Error -	Rango Error +
Intervalo	95%	3.107282524	85.02857143	81.9212889	88.13585395
Intervalo	99%	4.083661457	85.02857143	80.94490997	89.11223289

Con esto podemos concluir que según el intervalo de confianza obtenido la media de 88 se encuentra dentro del error, al utilizar la media obtenida de los datos para jóvenes. Por lo que es correcto asumir que la media es 88.

b) Obtener los intervalos de confianza al 95% para la diferencia de medias en el nivel basal de glucosa entre adultos y jóvenes e interpreta los resultados. ¿Se puede concluir que el nivel basal de glucosa de los jóvenes y los adultos es el mismo con nivel de significación del 5%? Suponiendo que se cumplen las condiciones iniciales teóricas para obtener los intervalos de confianza.

Se obtiene la estadística descriptiva de ambos grupos obteniendo:

<i>Nivel glucosa basal jovenes</i>		<i>Nivel glucosa basal adultos</i>	
Mean	85.02857143	Mean	90.83333333
Standard Error	1.585377358	Standard Error	1.544797225
Median	82	Median	90.5
Mode	90	Mode	88
Standard Deviation	9.379218939	Standard Deviation	8.461202871
Sample Variance	87.9697479	Sample Variance	71.59195402
Kurtosis	-0.463968504	Kurtosis	-1.025807155
Skewness	0.336054724	Skewness	-0.08006821
Range	39	Range	29
Minimum	65	Minimum	77
Maximum	104	Maximum	106
Sum	2976	Sum	2725
Count	35	Count	30

Con esto podemos obtener los intervalos de confianza y sumarlos:

Intervalo Jóvenes	95%	3.107282524
Intervalo adultos	95%	3.027746925
		6.135029449

Con esto podemos descubrir si tienen el mismo Nivel Basal de Glucosa al buscar que lleguen a 0. Estos son los resultados:

	Adultos - Jovenes	Jovenes - Adultos
Sumar	10.35469506	-1.254828752
Restar	1.254828752	-10.35469506

Podemos observar que el resultado nunca llega a 0 por lo que concluimos que no tienen el mismo Nivel Basal de Glucosa. Esto puede deberse a las diferencias de edad.

c) Se quiere estudiar la proporción de la población con un nivel basal de glucosa superior a 95 mg/Dl (prediabetes). A partir de la muestra del fichero (tomando todos los datos) obtener un intervalo de confianza al 98% y contrastar la hipótesis que la proporción de la población con glucosa superior a 95 mg/Dl es 0,15 con nivel de significación del 5%.

Para responder la primera parte de la pregunta se toman los siguientes datos:

Grupo de control	Nivel glucosa basal	Nivel glucosa 60 min
1	97	157
1	103	172
1	104	161
1	101	149
1	99	158
1	97	159
2	99	198
2	96	191
2	97	159
2	102	161
2	104	182
2	96	181
2	98	160
2	102	164
2	106	172
2	99	172
2	98	173

Y con el siguiente modelo de estadística descriptiva:

<i>Nivel glucosa basal</i>	
Mean	99.88235294
Standard Error	0.756745902
Median	99
Mode	97
Standard Deviation	3.120143285
Sample Variance	9.735294118

	-
Kurtosis	0.966356902
Skewness	0.501371351
Range	10
Minimum	96
Maximum	106
Sum	1698
Count	17

Para obtener el intervalo de confianza de esta población primero se identificaron que era 17 datos de los 65 originales por lo que representan un 26.15% de la población. Con esto se obtiene el error de cuánta gente representan estos datos.

Intervalo	98%	0.12680889
Poblacion	17	
Poblacio total	65	
	26.15%	

Suma	Resta
38.8347%	13.4730%

Se puede observar que el rango que representa es entre el 38.83% de la población al 13.47% de la población. Un rango bastante elevado.

Para responder la segunda parte se utiliza el intervalo de 95% con el % de población previamente obtenida obteniendo los siguientes resultados:

Intervalo	95%	0.106837357
-----------	-----	-------------

Suma	Resta
36.84%	15.47%

Al comparar las hipótesis podemos concluir que la población con glucosa superior a 95 es 0.15, ya que ese nivel de confianza se encuentra dentro del rango 0.1547 a 0.3684.

d) ¿Se detecta una variación significativa del nivel de glucosa en sangre en el grupo de los adultos después de la toma?

Para realizar esta variación se debe obtener la estadística descriptiva del Nivel Basal y el Nivel de Glucosa 60 minutos.

<i>Nivel glucosa basal</i>		<i>Nivel glucosa 60 min</i>	
Mean	90.83333333	Mean	173.2666667
Standard Error	1.544797225	Standard Error	2.247314276
Median	90.5	Median	172
Mode	88	Mode	172
Standard Deviation	8.461202871	Standard Deviation	12.30904723
Sample Variance	71.59195402	Sample Variance	151.5126437
Kurtosis	-1.025807155	Kurtosis	-0.635649125
Skewness	-0.08006821	Skewness	0.286622679
Range	29	Range	46
Minimum	77	Minimum	152
Maximum	106	Maximum	198
Sum	2725	Sum	5198
Count	30	Count	30

Luego se busca el intervalo de confianza para ambos resultados:

Intervalo Basal	95%	3.027746925
Intervalo 60 min	95%	4.404655043
		7.432401968

Con esto ya podemos descubrir cuanta variación se tiene con los siguientes resultados:

	60 min - Basal	Basal - 60 min
Sumar	84.95263287	-79.91403379
Restar	79.91403379	-84.95263287

Ya con esto se puede confirmar que se tiene un nivel de variación bastante significativo luego de que se tomen las pastillas.