

Estadística aplicada 11.67

Trabajo práctico – Informe grupal

Primer cuatrimestre de 2024

Grupo 7

APELLIDO Y NOMBRES	Nº DE LEGAJO	% PARTICIPACIÓN
Calandri, Joaquin	62514	29%
Chaud Olivieri, Agustin	63253	29%
Di Benedetto, Franco	58735	13%
Hitters, Santiago	62160	29%

Plazo de entrega: 07/06/2024

RESUMEN

En el presente informe se estudiaron los datos obtenidos del lanzamiento de muestras de helicópteros de papel del tamaño de hojas A4.

A partir de la información dada en clase y el uso de un software estadístico, se analizaron y compararon dichos datos de forma estadística; también se analizó las relaciones entre variables aleatorias.

INTRODUCCIÓN

El informe se dividió en 5 experimentos, en los que se analizó un fenómeno estadístico puntual en cada uno.

En los primeros tres experimentos se caracterizaron y analizaron estadísticamente una, dos y cuatro muestras respectivamente, donde las muestras del experimento dos y tres eran independientes entre sí.

En el cuarto experimento se cuantificó y exploró las relaciones lineales entre el tiempo de vuelo y algunos factores de variabilidad.

En el quinto experimento se caracterizó y comparó estadísticamente el efecto de un tratamiento en una muestra de individuos heterogénea.

EXPERIMENTO 1

Se armaron 36 helicópteros del croquis A, hojas de tamaño A4, gramaje de 70 g/m². Se cargó el tiempo de vuelo como Y11 en centisegundos, al número de identificador de helicóptero como ID, al orden de lanzamiento como OA, a la verticalidad como Y12 y a los residuos como RES1.

1.1) En la figura 1.1.1 se presentan los estadísticos básicos. Es poco preciso afirmar a qué factor puntual se debió la dispersión, al no ser comparada con otra muestra. Se puede asociar a varios factores que influyen en el tiempo de vuelo, como la adhesión que se genera entre la mano del lanzador y el papel, el ángulo de lanzamiento o el tiempo que se tardó en frenar el cronómetro. La asimetría es positiva resultando en una distribución sesgada hacia la izquierda. Por otro lado, en la figura 1.1.2 se observa que 29 vuelos cayeron verticalmente dentro del área establecida, mientras que 7 no lo hicieron. Este resultado se debe a que no se siguió un método de extrema precisión en los lanzamientos. Los helicópteros con vuelo NOK presentan un mayor varianza, esto resulta lógico ya que al no seguir una verticalidad exacta, la distancia recorrida varía.

Estadísticas

Variable	Conteo total	N	Media	Error estándar de				Mínimo
				la media	Desv.Est.	Varianza	CoefVar	
Y11	36	36	209,139	2,91416	17,4849	305,723	8,36	175
Variable	Q1	Mediana	Q3	Máximo	IQR	Modo	N para moda	Asimetría
Y11	197,25	210	223,25	238	26	211	4	0,13

Figura 1.1.1

Estadísticas

Variable	Y12	Conteo total	N	Media	Error estándar de				Mínimo
					la media	Desv.Est.	Varianza	CoefVar	
Y11	NOK	7	7	215,714	7,24311	19,1635	367,238	8,88	188
	OK	29	29	207,552	3,16251	17,0306	290,042	8,21	175
Variable	Y12	Q1	Mediana	Q3	Máximo	IQR	Modo	N para moda	Asimetría
Y11	NOK	198	211	237	238	39	211	2	-0,16
	OK	195	208	218	238	23	210	3	0,17

Figura 1.1.2

1.2) Del boxplot de la figura 1.2.1 se supone la ausencia de outliers en los tiempos de vuelo. La cola inferior es más larga que la superior, ya que la distancia entre mínimo y Q1 es mayor a la distancia entre el máximo y Q3. La distancia entre Q1 a la media y de Q3 a la media es muy similar.

Se optó por un histograma en lugar de una gráfica de puntos ya que el número de muestras es mayor a 30. El histograma de la figura 1.2.2 muestra una dispersión considerable, presenta una concentración de valores alrededor de la media. Muestra también una distribución normal con un pico anormal en el extremo derecho, da a entender que hubo un problema con los vuelos de mayor tiempo.

La figura 1.2.3 muestra que el 80% de la muestra respeta la verticalidad.

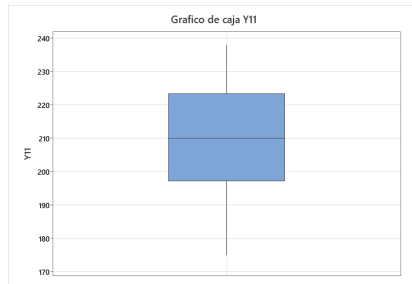


Figura 1.2.1

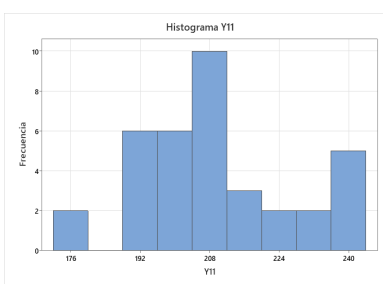


Figura 1.2.2

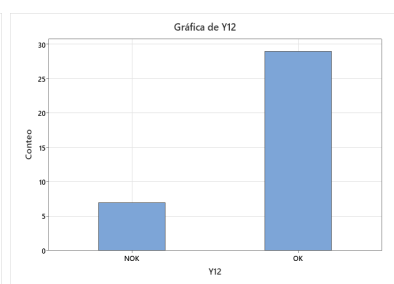


Figura 1.2.3

1.3) Se realizó la prueba de Kolmogorov Smirnov, para verificar la verticalidad. En la prueba se trabajó con un $\alpha = 15\%$. En la figura 1.3.1 se observa que el valor p resultó ser mayor a 15%, entonces la prueba resultó no ser significativa, por ende no se rechazó la hipótesis nula. Concluyendo que siguen una distribución normal.

H0: Los datos siguen una distribución Normal

H1: Los datos no siguen una distribución Normal

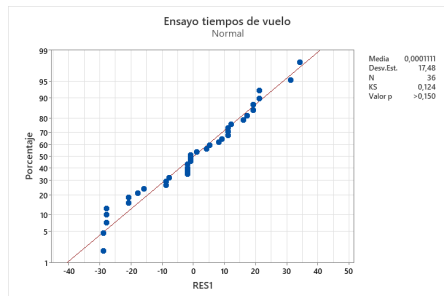


Figura 1.3.1

1.4) En la figura 1.4.1 no se logra ver ninguna tendencia clara, lo que es un indicio de que los residuos son independientes entre sí. Los recaudos tenidos en cuenta fueron respetar un orden de lanzamiento aleatorio, realizar el experimento en un lugar cerrado sin influencia de las condiciones climáticas y medir con una cinta métrica la altura de lanzamiento. Los detalles del lanzamiento se encuentran en el anexo.

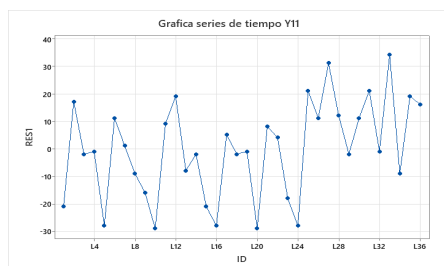


Figura 1.4.1

1.5) De la figura 1.2.1 se puede suponer que no existen valores atípicos en los tiempos de vuelo. Para verificar la suposición se realizó la prueba de Tukey:

La distancia entre el mínimo y Q1 = 22,25 cs es menor a el IQR multiplicado por 1,5 = 39 cs

La distancia entre el máximo y Q3 = 14,75 cs es menor a el IQR multiplicado por 1,5 = 39 cs

Se puede afirmar que no existen outliers leves ni severos en la muestra.

1.6) Se calculó el intervalo de confianza para el tiempo medio de vuelo usando la t de Student, ya que se desconoce la varianza poblacional. Intervalo: $\bar{x} \pm E$ $E: t_{v; 1-\alpha/2} \cdot s/\sqrt{n}$

Estadísticas descriptivas

		Error estándar de la media	IC de 90% para μ	
N	Media Desv.Est.			
36	209,14	17,48	2,91	(204,22; 214,06)

μ : media de población de Y11

Figura 1.6.1

En la figura 1.6.1 se ve el intervalo de confianza $p(204,22 \text{ cs} \leq \mu \leq 214,06 \text{ cs}) = 0,9$

Se calculó el intervalo de confianza para la desviación estándar usando chi cuadrado porque se verificó en el punto 1.3 que cumple el supuesto de normalidad.

Método

Evento Y12 = OK
p Proporción donde Y12 = OK
Método Método exacto de Blaker ajustado

Estadísticas descriptivas

			IC de 90% para σ usando Bonett	IC de 90% para σ usando Chi-cuadrada
N	Desv.Est.	Varianza		
36	17,5	306	(15,2; 21,1)	(14,7; 21,8)

Figura 1.6.2

Figura 1.6.3

En la figura 1.6.3 se ve el intervalo de confianza $p(14,7 \text{ cs} \leq \sigma \leq 21,8 \text{ cs}) = 0,9$

Se calculó el intervalo de confianza para el tiempo medio de vuelo de la proporción OK:

Estadísticas descriptivas

N	Evento	Muestra p	IC de 90% para p
36	29	0,805556	(0,669149; 0,893383)

Figura 1.6.4

Como se ve en la figura 1.6.4 se puede afirmar con un 90% de probabilidad que la proporción de vuelos OK será mayor a 0,6691 y menor 0,8933.

1.7) En la figura 1.7.2 se ve el planteamiento de las H0 y H1; como el valor p resultó ser muy cercano a cero se rechaza la hipótesis nula. Además el valor 185 no está contenido en el intervalo de confianza, por lo que no es razonable suponer que el tiempo de vuelo medio poblacional es 185 cs.

Estadísticas descriptivas

N	Media	Desv.Est.	Error estándar de la media	IC de 90% para μ
36	209,14	17,48	2,91	(204,22; 214,06)

μ : media de población de Y11

Figura 1.7.1

Prueba

Hipótesis nula $H_0: \mu = 185$

Hipótesis alterna $H_1: \mu \neq 185$

Valor T	Valor p
8,28	0,000

Figura 1.7.2

1.8) Se realizó la prueba de hipótesis y se obtuvo un p menor a α en los 2 métodos. Como la prueba resultó ser significativa entonces se rechazó la hipótesis nula, por ende no se puede concluir que el desvío poblacional sea de 25 cs.

Estadísticas descriptivas

N	Desv.Est.	Varianza	IC de 90% para σ usando Bonett	IC de 90% para σ usando Chi-cuadrada
36	17,5	306	(15,2; 21,1)	(14,7; 21,8)

Figura 1.8.1

Prueba

Hipótesis nula $H_0: \sigma = 25$

Hipótesis alterna $H_1: \sigma \neq 25$

Método	Estadística de prueba	GL	Valor p
Bonett	—	—	0,003
Chi-cuadrada	17,12	35	0,010

Figura 1.8.2

1.9) No se rechazó la hipótesis nula planteada ya que, como se ve en la figura 1.9.2, se obtuvo un valor p mayor a α . Se puede concluir que la muestra proviene de una población con una probabilidad OK igual a 85%.

Estadísticas descriptivas

N	Evento	Muestra p	IC de 90% para p
36	29	0,805556	(0,669149; 0,893383)

Figura 1.9.1

Prueba

Hipótesis nula $H_0: p = 0,85$

Hipótesis alterna $H_1: p \neq 0,85$

Valor p
0,481

Figura 1.9.2

1.10) En los puntos 1.7, 1.8, 1.9 se realizaron pruebas para comprobar si valores puntuales pueden ser tomados como poblacionales. Mientras que en el 1.6 se construyeron intervalos de confianza para la media, el desvío estándar y la proporción de los OK. Los 4 puntos anteriores se calcularon con un $\alpha = 10\%$

EXPERIMENTO 2

Se armaron 18 helicópteros a partir del croquis A dado en clase, en hojas de cartulina y un gramaje de 120 g/m². Se seleccionaron al azar 18 helicópteros del experimento anterior, repitió el lanzamiento del experimento 1 para ambas muestras. Se cargaron los datos siguiendo las instrucciones dadas.

Una vez se registraron los datos, se comenzó con el análisis tomando de guía los puntos a desarrollar proporcionados por la cátedra.

Se cargó el tiempo de vuelo como Y21 en centisegundos, al número de identificador de helicóptero como ID, al orden de lanzamiento como OA, a la verticalidad como Y22, tipo de papel como TIPO a los residuos como RES1.

2.1) La media del liviano resultó ser menor a la media del pesado.

Estadísticas

Variable	TIPO	Conteo total	N	Media	Error estándar de		Varianza	CoefVar	Mínimo
					la media	Desv.Est.			
Y21	L	18	18	199,833	3,89255	16,5147	272,735	8,26	173
	P	18	18	174,778	3,90371	16,5620	274,301	9,48	145

Variable	TIPO	Q1	Mediana	Q3	Máximo	IQR	Modo	N para moda	Asimetría
Y21	L	184,75	198,5	215,75	224	31	198; 224	2	0,06
	P	165	172,5	184	217	19	165; 173; 184	2	0,96

Figura 2.1.1

No se realizó un histograma porque la cantidad de muestras es menor a 30, se optó por una gráfica de puntos en su lugar. En la figura 2.1.2 no se puede identificar una asimetría clara. Si se comparan ambas muestras los helicópteros pesados presentan una mayor dispersión y una mayor acumulación de muestras cerca de su mediana.

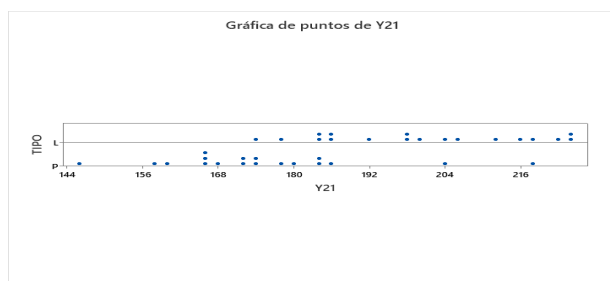


Figura 2.1.2

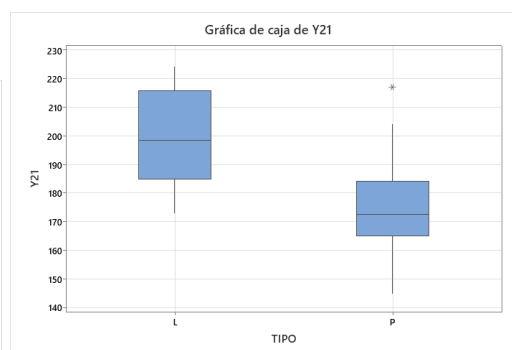


Figura 2.1.3

Las colas inferiores y superiores del boxplot de la figura 2.1.3 referente a los helicópteros pesados son mayores a las del liviano. El rango intercuartílico de los helicópteros livianos es mayor a el de los pesados. Se puede ver que hay un outlier en los helicópteros pesados, se usó la prueba de Tukey y este resultó ser un outlier leve ya que es mayor a $(Q3 + 1,5 * IQR)$ pero menor a $(Q3 + 3 * IQR)$.

De la figura 2.1.4 se calculó que el 88,8% de los helicópteros livianos fueron OK mientras que solo el 66,6% de los pesados fueron OK. Muestra una tendencia a que los livianos resultan mantener mejor la verticalidad.

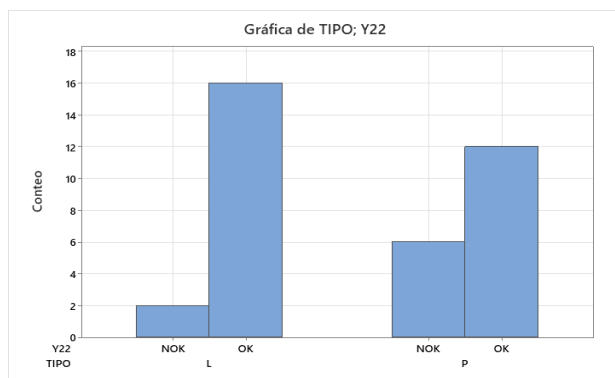


Figura 2.1.4

2.2) Se realizó la prueba de Kolmogorov Smirnov, para verificar la verticalidad. Se trabajó con un $\alpha = 15\%$ y el valor p resultó ser mayor a eso, se puede afirmar que se cumple el supuesto de normalidad ya que se sostiene la hipótesis nula.

H0: Los datos siguen una distribución Normal

H1: Los datos no siguen una distribución Normal

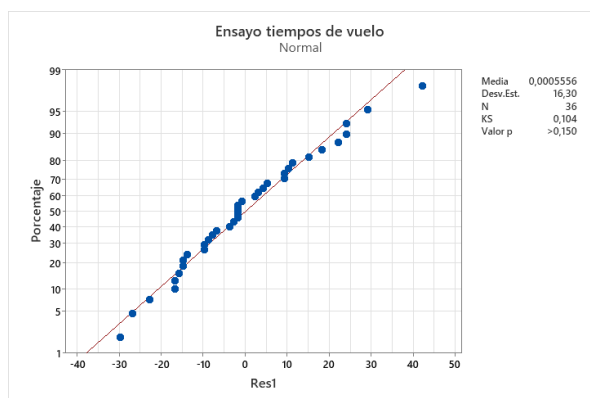


Figura 2.2.1

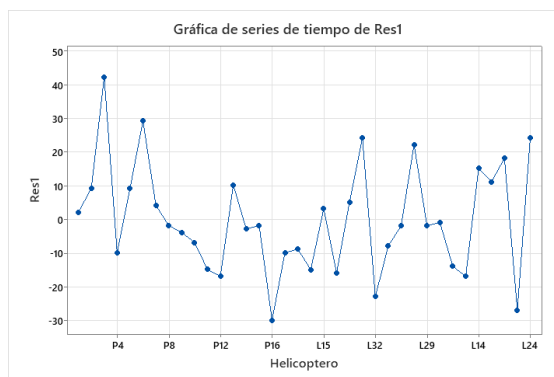


Figura 2.3.1

2.3) En la figura 2.3.1 no se ve ninguna tendencia clara, lo que es un indicio de que los residuos son independientes entre sí. Los recaudos que se tuvieron en cuenta fueron los mismos que en el experimento uno, estos se encuentran en el punto 1.4.

2.4) En la figura 2.4.2 se ven planteados la H0 y la H1. A través del método de Fisher, utilizado porque los datos siguen una distribución normal, se calculó el valor-p. Como dicho valor resultó ser muy alto (cercano a 1) no se rechazó la H0, por lo que se cumple el supuesto de igualdad de varianzas para ambas muestras.

Método

σ_1 : desviación estándar de Y21 cuando TIPO = L

σ_2 : desviación estándar de Y21 cuando TIPO = P

Relación: σ_1/σ_2

Se utilizó el método F. Este método es exacto sólo para datos normales.

Prueba

Hipótesis nula $H_0: \sigma_1 / \sigma_2 = 1$

Hipótesis alterna $H_1: \sigma_1 / \sigma_2 \neq 1$

Nivel de significancia $\alpha = 0,1$

Estadística				
Método	de prueba	GL1	GL2	Valor p
F		0,99	17 17	0,991

Figura 2.4.1

Figura 2.4.2

2.5) En la figura 2.5.3 se ve que el intervalo para las diferencia de los tiempos medios de vuelo no incluye al cero, por lo que se puede afirmar con un 90% de confianza que los tiempos difieren. En la figura 2.5.4 se ve que el intervalo de confianza para el cociente de desviaciones estándar incluye al 1, por lo que puede afirmarse con 90% de confianza que las desviaciones estándar de ambas muestras son iguales. La figura 2.5.6 muestra evidencia de que las proporciones de vuelos OK difieren si el helicóptero es liviano o pesado, ya que el intervalo de confianza no incluye al cero.

Estadísticos descriptivos: Y21

Método

μ_1 : media de población de Y21 cuando TIPO = L

μ_2 : media de población de Y21 cuando TIPO = P

Diferencia: $\mu_1 - \mu_2$

Se presupuso igualdad de varianzas para este análisis.

TIPO	N	Media	Desv.Est.	Error estándar de la media
L	18	199,8	16,5	3,9
P	18	174,8	16,6	3,9

Figura 2.5.1

Figura 2.5.2

Estimación de la diferencia

Diferencia agrupada	Desv.Est.	IC de 90% para la diferencia
25,06	16,54	(15,73; 34,38)

Figura 2.5.3

Relación de desviaciones estándar

Relación estimada	IC de 90% para la relación usando F
0,997143	(0,662; 1,503)

Figura 2.5.4

Estadísticos descriptivos: Y22

TIPO	N	Evento	Muestra	p
L	18	16	0,888889	
P	18	12	0,666667	

Figura 2.5.5

Estimación de la diferencia

Diferencia	IC de 90% para la diferencia
0,222222	(0,002570; 0,441874)
<i>IC basado en la aproximación a la normal</i>	

Figura 2.5.6

2.6) Las hipótesis nulas se ven presentes en las figuras 2.6.1, 2.6.2 y 2.6.3. Las 3 hipótesis nulas anteriores asumen la igualdad entre las muestras. Los análisis y comentarios se encuentran en los 3 puntos anteriores.

Prueba

Hipótesis nula $H_0: p_1 - p_2 = 0$
Hipótesis alterna $H_1: p_1 - p_2 \neq 0$

Método	Valor Z	Valor p
Aproximación normal	1,66	0,096
Exacta de Fisher		0,228

La aproximación normal puede ser inexacta para muestras pequeñas.

Figura 2.6.1

Prueba

Hipótesis nula $H_0: \mu_1 - \mu_2 = 0$
Hipótesis alterna $H_1: \mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
4,54	34	0,000

Figura 2.6.2

Prueba

Hipótesis nula $H_0: \sigma_1 / \sigma_2 = 1$
Hipótesis alterna $H_1: \sigma_1 / \sigma_2 \neq 1$
Nivel de significancia $\alpha = 0,1$

Estadística				
Método de prueba	GL1	GL2	Valor p	
F	0,99	17	17	0,991

Figura 2.6.3

2.7) En los 2 puntos anteriores se concluyó que los tiempos medios de vuelo de ambas muestras difieren, que las desviaciones estándar de ambas muestras son iguales y que las proporciones de vuelos ok difieren según el tipo de helicóptero. Se concluyó lo mismo con métodos diferentes, uno es un intervalo de confianza y el otro una prueba de hipótesis.

EXPERIMENTO 3

3.1) Se desea ver si hay diferencias entre las medias de los tiempos de vuelo de las cuatro muestras a analizar. La hipótesis global es: todas las μ_i son iguales. La hipótesis alternativa es: al menos una μ_i es distinta del resto. Donde la condición de rechazo es $p\text{-value} < \alpha$. Si es rechazada H_0 , se tiene evidencia de que alguna de las medias es distinta a las demás, de no rechazar se probaría la hipótesis global. Los papeles utilizados fueron de los siguientes gramajes: L = 70 g/m², P = 120 g/m², A = 150 g/m² y B = 300g/m².

3.2) En la figura 3.2.1 se ve la dispersión y la asimetría de cada muestra. La menor dispersión la tiene la muestra A. Todas las muestras parecen tener asimetría positiva. Los helicópteros hechos con papel B tienen los menores tiempos de vuelo, mientras que los tiempos de vuelo de los otros papeles son similares. Esto puede suponer que su media va a diferir de forma significativa, al resto de las medias. En la figura 3.2.2, se logra detectar un outlier en la muestra P. Este se confirma que es leve a través de la prueba de Tukey. Se ve un solapamiento importante entre las cajas A y L. En la figura 3.2.7 en la tabla con título "vs orden" no se reconoce ningún patrón, esto da a entender que los residuos son independientes entre sí.

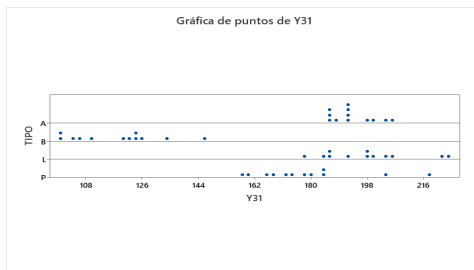


Figura 3.2.1

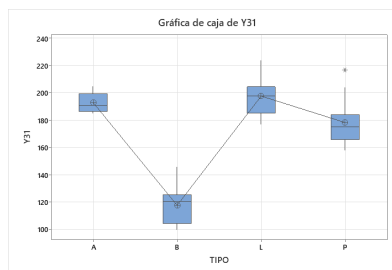


Figura 3.2.2

Resumen del modelo

S	R-cuadrado	R-cuadrado(ajustado)	R-cuadrado (pred)
13,8212	85,37%	84,37%	82,59%

Figura 3.2.3

Información del factor

Factor	Niveles	Valores
TIPO	4	A; B; L; P

Figura 3.2.4

Medias

TIPO	N	Media	Desv.Est.	IC de 90%
A	12	193,00	7,10	(186,30; 199,70)
B	12	117,75	14,15	(111,05; 124,45)
L	12	197,75	14,50	(191,05; 204,45)
P	12	178,33	17,42	(171,63; 185,04)

Desv.Est. agrupada = 13,8212

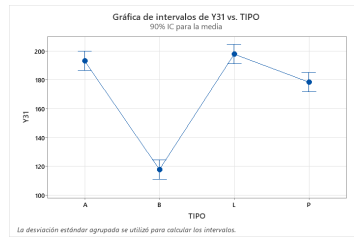


Figura 3.2.5

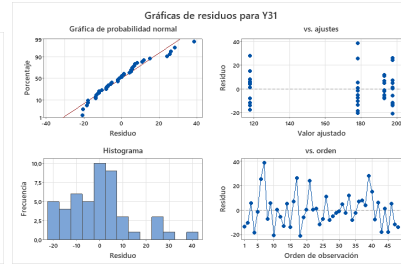


Figura 3.2.6

Figura 3.2.7

3.3) El 1-way ANOVA se usa para analizar las medias de más de 2 grupos para ver si alguno de los grupos difiere significativamente de los demás en variables de interés, en este caso la media. Para realizar el 1 way anova se requiere que la variable (tiempo de vuelo) cumpla los supuestos de normalidad para todas las poblaciones (tipos de helicópteros) y el de homocedasticidad. La variabilidad total se desdobra en dos: la variabilidad entre las diferencias dentro de los grupos y la variabilidad entre las diferencias entre los grupos. En la figura 3.3.1 se ve la tabla ANOVA.

En la columna GL (grados de libertad) se encuentran las siguientes intersecciones:

En la intersección GL-TIPO se cuenta la cantidad de tipos de papeles usados menos uno.

En la intersección GL-Error se cuenta la cantidad de observaciones menos la cantidad de grupos.

En la intersección GL-Total se encuentra la suma de las dos intersecciones anteriores.

En la columna SC Ajust se encuentra la suma de cuadrados totales ajustados. En la columna MC Ajust se encuentra los cuadrados medios obtenidos. En la columna Valor F se encuentra el valor de los cuadrados medios de X dividido el cuadrado de los valores medios del error. En la columna Valor p se encuentra un valor que simboliza la probabilidad de obtener los resultados observados en el análisis, asumiendo que la hipótesis nula es cierta.

El modelo teórico para cada observación en un ANOVA de un solo factor es:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Donde cada valor observado y_{ij} es una variable aleatoria. μ es la media global de todas las observaciones. τ_i es un parámetro propio del i-ésimo nivel del factor, $\tau_i = \mu_i - \mu$. ε_{ij} es la componente aleatoria que incluye todas las demás fuentes de variabilidad presentes en la j-ésima condición, y no puede explicarse con el valor i seleccionado. i va desde 1 hasta a donde a es simboliza la cantidad de modelos ajustados al final.

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
TIPO	3	49043	16347,6	85,58	0,000
Error	44	8405	191,0		
Total	47	57448			

Figura 3.3.1

Método

Hipótesis nula	Todas las medias son iguales
Hipótesis alterna	No todas las medias son iguales
Nivel de significancia α	= 0,1

Se presupuso igualdad de varianzas para el análisis.

Figura 3.3.2

3.4) En la figura 3.3.2 se plantea la hipótesis nula. Como se ve en la figura 3.3.1 se calculó que el valor-p resultó ser cercano a cero, por ende se rechaza dicha H_0 y se concluye con una confianza del 90% que no todas las medias son iguales. Como el p-value es menor a α entonces se rechaza la hipótesis global, la prueba es significativa. Como conclusión no todas las medias son iguales.

3.5) El error experimental del modelo se obtiene del análisis de la varianza, en la intersección entre el error y el MC Ajust. En la figura 3.3.1 se observa que el resultado es 191,0 cs^2 .

3.6) Los supuestos estadísticos para dar buenos estos resultados no se cumplen razonablemente bien. El supuesto de normalidad se cumple, se verifica a través de la prueba de Kolmogorov Smirnov. En la figura 3.6.1 vemos que el valor p nos dió mayor a 15%, entonces la prueba es no significativa, no se rechaza la hipótesis nula de que los datos siguen una distribución normal. Hay indicios de que se cumple el supuesto de independencia en la figura 3.2.7. en el gráfico de residuos. Las varianzas entre los grupos no son iguales, esto se verifica con el método de Bartlett en la figura 3.6.3. El valor p es menor al α , entonces la prueba es significativa, se rechaza la hipótesis nula.

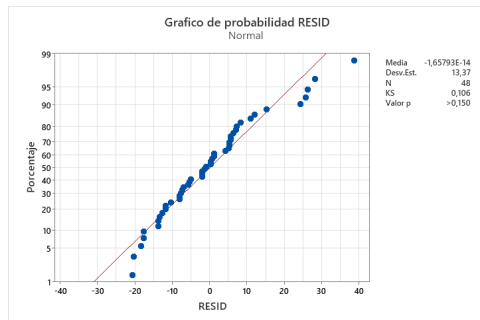


Figura 3.6.1

Método

Hipótesis nula: Todas las varianzas son iguales
Hipótesis alterna: Por lo menos una varianza es diferente
Nivel de significancia $\alpha = 0,1$

Se utiliza el método de Bartlett. Este método es exacto sólo para datos normales.

Figura 3.6.2

Pruebas

	Estadística	
	Método	de prueba Valor p
Bartlett	7,72	0,052

Figura 3.6.3

3.7) Se rechaza la hipótesis global, se deben identificar qué medias difieren del resto, se usó el proceso de comparaciones múltiples con la prueba de Tukey. Como se ve en la figura 3.7.1 las muestras L y A forman parte del grupo A, P forma parte del grupo B y B del grupo C. Esto resulta extraño porque el gramaje de P es menor al de A y mayor al de L, estaría acotado por estos. Se debería realizar un nuevo muestreo de P, para verificar dicho comportamiento. En la figura 3.7.2 se puede ver mejor la agrupación de las muestras, donde la media de A está contenida en el intervalo de confianza de L y la media de L en el intervalo de confianza de A. Las medias de B y P no se encuentran contenidas en el intervalo de confianza de ningún otro grupo.

Agrupar información utilizando el método de Tukey y una confianza de 90%

TIPO	N	Media	Agrupación
L	12	197,75	A
A	12	193,00	A
P	12	178,33	B
B	12	117,75	C

Las medias que no comparten una letra son significativamente diferentes.

Figura 3.7.1

Medias

TIPO	N	Media	Desv.Est.	IC de 90%
A	12	193,00	7,10	(186,30; 199,70)
B	12	117,75	14,15	(111,05; 124,45)
L	12	197,75	14,50	(191,05; 204,45)
P	12	178,33	17,42	(171,63; 185,04)

Desv.Est. agrupada = 13,8212

Figura 3.7.2

EXPERIMENTO 4

4.1) La figura 4.1.1 refleja una correlación lineal positiva entre las variables Y41 y X1, lo que se refleja en ambos límites del intervalo de confianza superiores a cero. Esto se interpreta como que a medida que aumenta la longitud del ala aumentará también el tiempo de vuelo.

Aun así se realizó el análisis de correlación de Pearson para el coeficiente de correlación para validar el análisis gráfico efectuado. Se plantearon las siguientes hipótesis:

Ho) $\rho = 0$ "No hay correlación entre las variables"

Ha) $\rho \neq 0$ "Hay correlación entre las variables"

Como la figura 4.1.3 muestra que el p-value es menor a 10%, por lo que se concluye que hay correlación entre las variables. La figura 4.1.2 muestra el coeficiente de correlación.

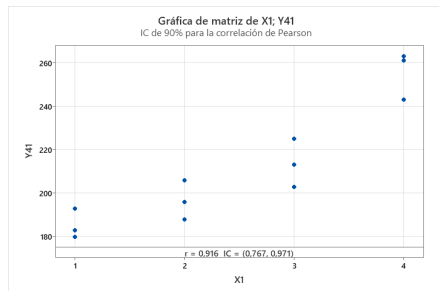


Figura 4.1.1

Correlaciones

X1
Y41 0,916

Figura 4.1.2

Correlaciones en parejas de Pearson

Muestra 1	Muestra 2	N	Correlación	IC de 90% para	p Valor p
Y41	X1	12	0,916	(0,767; 0,971)	0,000

Figura 4.1.3

Como se ve en la figura 4.1.4 no se ve reflejada ninguna correlación lineal clara entre las variables Y41 y X2, lo que se refleja en el límite inferior del intervalo menor a cero y el límite superior mayor a cero. Esto se interpreta como que la cantidad de clips no tendrá influencia en el tiempo de vuelo. Al igual que para la otra variable, se realizó el análisis de correlación de Pearson para el coeficiente de correlación para validar el análisis gráfico efectuado. Se plantearon las hipótesis de la misma forma, por lo que la figura 4.1.5 muestra un valor-p que hace suponer como válida la hipótesis nula "No hay correlación entre las variables".

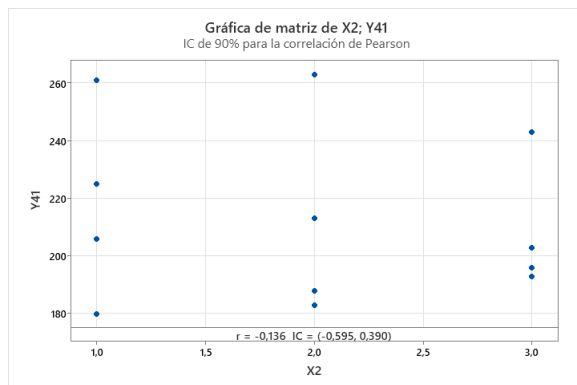


Figura 4.1.4

Correlaciones en parejas de Pearson

Muestra 1	Muestra 2	N	Correlación	IC de 90% para	p Valor p
Y41	X2	12	-0,136	(-0,595; 0,390)	0,674

Figura 4.1.5

4.2) El modelo de RLS de 1º orden para X_1 es: $Y_i = \beta_0 + \beta_1 \cdot X_1 + \varepsilon_i$. En dicho modelo la variable explicada Y_i es el tiempo de vuelo del helicóptero en cs, la variable explicativa X_1 es la longitud del ala del cm, β_1 es el parámetro poblacional desconocido y ε_i el término de perturbación aleatoria.

Por otro lado, el modelo de RLS de 1º orden para X_2 es: $Y_i = \beta_0 + \beta_1 \cdot X_2 + \varepsilon_i$. En dicho modelo la variable explicada Y_i es el tiempo de vuelo del helicóptero en cs, la variable explicativa X_2 es la cantidad de clips, β_1 es el parámetro poblacional desconocido y ε_i el término de perturbación aleatoria.

4.3) La figura 4.3.1 refleja que el valor-p de la variable X_1 es menor a 10%, por lo que se concluye que β_1 es estadísticamente significativo con una probabilidad de equivocarse del 10%. Además figura 4.3.3 refleja que el R-cuadrado es 83,85% por lo que el modelo es válido.

Por otro lado, la figura 4.3.4 refleja que el valor-p de la variable X_2 es mayor a 10%, por lo que se concluye que β_1 NO es estadísticamente significativo con una probabilidad de equivocarse del 10%. Además la figura 4.3.3 refleja que el R-cuadrado es 1,84%, lo que implica la inexistencia de un modelo de primer orden.

Análisis de Varianza

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Regresión	1	7797,6		7797,60		51,91	0,000
X_1	1	7797,6		7797,60		51,91	0,000
Error	10	1502,1		150,21			
Falta de ajuste	2	761,4		380,70		4,11	0,059
Error puro	8	740,7		92,58			
Total	11	9299,7					

Figura 4.3.1

Coefficientes

Término	Coef	EE	del coef.	Valor T	Valor p	FIV
Constante	155,83		8,67	17,98	0,000	
X_1	22,80		3,16	7,21	0,000	1,00

Figura 4.3.2

Resumen del modelo

S	R-cuadrado	R-cuadrado(ajustado)	R-cuadrado (pred)
12,2559	83,85%	82,23%	76,58%

Figura 4.3.3

Análisis de Varianza

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Regresión	1	171,13		171,13		0,19	0,674
X_2	1	171,12		171,12		0,19	0,674
Error	10	9128,54		912,85			
Falta de ajuste	1	7,04		7,04		0,01	0,935
Error puro	9	9121,50		1013,50			
Total	11	9299,67					

Figura 4.3.4

Coefficientes

Término	Coef	EE	del coef.	Valor T	Valor p	FIV
Constante	222,1		23,1	9,62	0,000	
X_2	-4,6		10,7	-0,43	0,674	1,00

Figura 4.3.5

Resumen del modelo

S	R-cuadrado	R-cuadrado(ajustado)	R-cuadrado (pred)
30,2135	1,84%	0,00%	0,00%

Figura 4.3.6

4.4) En la figura 4.4.1 se pueden observar el modelo cuadrático muestral y sus valores de calidad de ajuste. El Y_{41} representa el tiempo de vuelo en cs y el X_1 representa los niveles de longitud de las alas. Los valores que se deben considerar para este estudio son el coeficiente de determinación R^2 y la desviación estándar del error expresada como S. El R^2 en este caso tiene un valor de 83,8%, entonces se entiende que el modelo explica el 83,8% de la variabilidad del tiempo de vuelo de los helicópteros.

En la figura 4.4.2 ocurre que el Rcuadrado tiene valor de 1,8% por ende el modelo explica sólo el 1,8% de la variabilidad del tiempo de vuelo, por lo que no es un buen modelo explicativo.

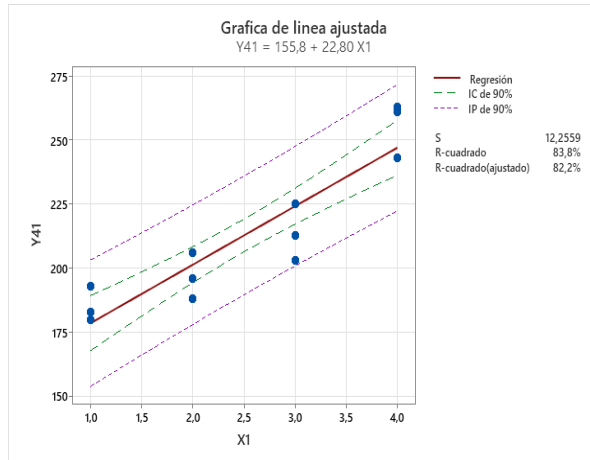


Figura 4.4.1

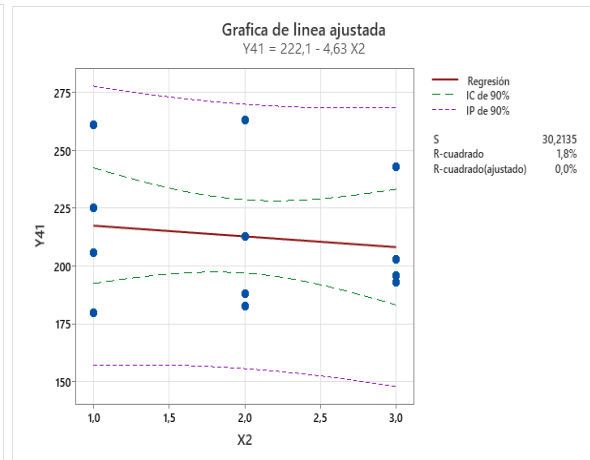


Figura 4.4.2

4.5) Para validar el modelo se deben comprobar los supuestos de mínimos cuadrados se realizó un análisis de residuos. Tanto en la figura 4.5.1 como en la figura 4.5.2 el valor de p es mayor a 15%, por ende se acepta como válida la H_0 : los residuos siguen una distribución normal.

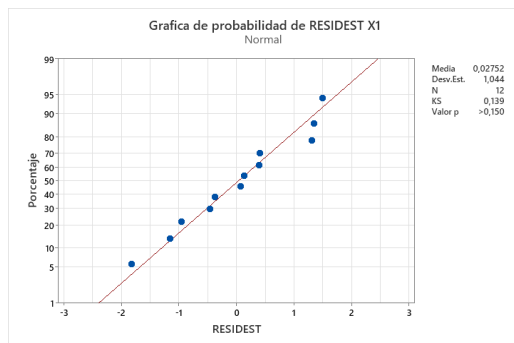


Figura 4.5.1

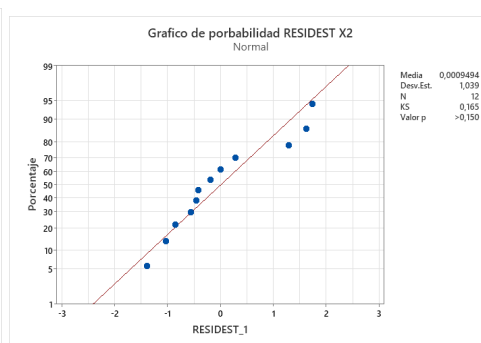


Figura 4.5.2

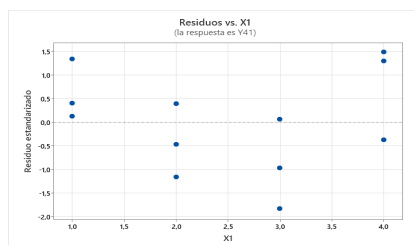


Figura 4.5.3

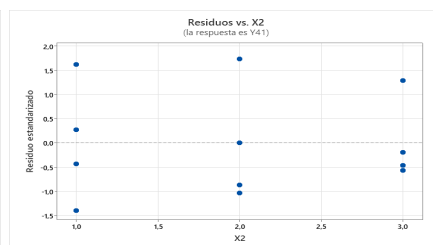


Figura 4.5.4

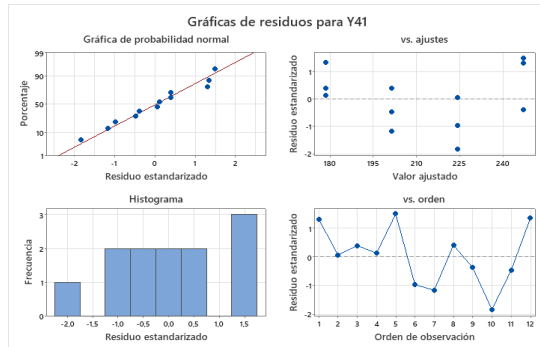


Figura 4.5.5

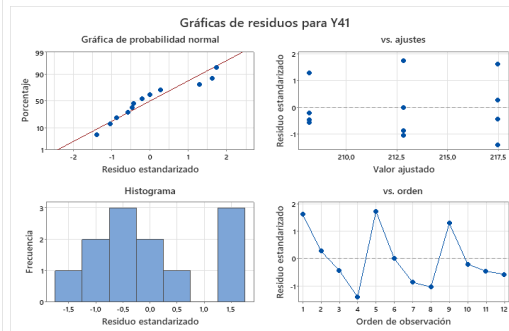


Figura 4.5.6

En las figuras 4.5.5 y 4.5.6 en el gráfico de residuos en función de los valores ajustados, se puede ver que los residuos obtenidos se encuentran centrados en el valor nulo. De esta manera, no se tiene evidencia gráfica que implique la refutación del supuesto. En este caso se está validando el supuesto de esperanza residual nula.

En las figuras 4.5.5 y 4.5.6 en el gráfico de residuos en función del orden no se nota ningún patrón marcado de los residuos. Esta observación brinda indicios de que los residuos se encuentran incorrelacionados y la varianza de los mismos se mantiene constante. Así se valida el supuesto de homocedasticidad y el de independencia.

4.6) El modelo de RLS de 2º orden para $X_1 : Y_i = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_1^2 + \varepsilon_i$

La variable explicada (Yi) es el tiempo de vuelo del helicóptero en cs, la variable explicativa X_1 es la longitud del ala del helicóptero en cm, x_i^2 representa el cuadrado de la longitud de las alas en cm^2 , β_i son los parámetros poblacionales desconocidos y ε_i el término de perturbación aleatoria.

El modelo de RLS de 2º orden para $X_2 : Y_i = \beta_0 + \beta_1 \cdot X_2 + \beta_2 \cdot X_2^2 + \varepsilon_i$

La variable explicada es la misma pero la variable explicativa X_2 es la cantidad de clips y x_i^2 representa la cantidad de clips elevada al cuadrado. β_i es el parámetro poblacional desconocido y ε_i el término de perturbación aleatoria.

Las variables explicativas son las mismas que para el modelo de regresión simple de primer orden.

Para que el modelo sea factible se deben cumplir los siguientes 4 supuestos: supuesto de normalidad, supuesto de no autocorrelación lineal, supuesto de homocedástica y el supuesto de que la media de los residuos es 0.

En la figura 4.6.3 se observa que el error es de 88,53 cs², el cual refleja la variabilidad que no puede explicar el modelo cuadrático.

Se observa en la figura 4.6.1 que el R cuadrado del modelo de segundo orden es mayor al de 1er orden de la figura 4.4.1. Además la S es menor a la S de 1er orden, por lo que el modelo cuadrático es mejor.

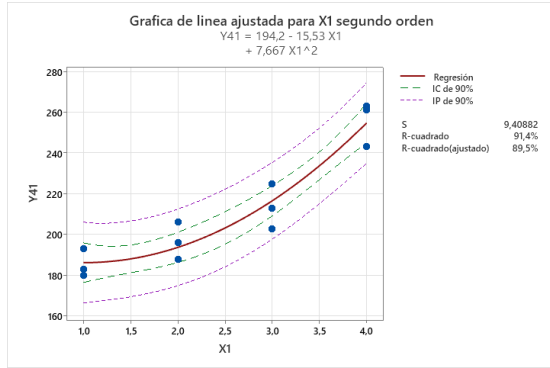


Figura 4.6.1

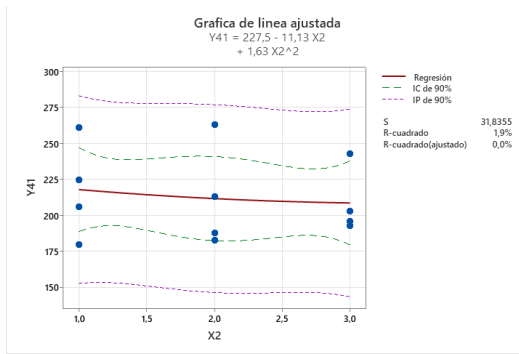


Figura 4.6.2

Análisis de Varianza

Fuente	GL	SC	MC	F	P
Regresión	2	8502,93	4251,47	48,03	0,000
Error	9	796,73	88,53		
Total	11	9299,67			

Figura 4.6.3

Análisis de varianza secuencial

Fuente	GL	SC	F	P
Lineal	1	7797,60	51,91	0,000
Cuadrático	1	705,33	7,97	0,020

Figura 4.6.4

4.7) Ecuación del modelo de RLM de 1er orden aditivo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
Siendo X1=nivel de ala y X2=nro de clips.

Ecuación de regresión

$$Y41 = 165,1 + 22,80 X1 - 4,63 X2$$

4.8) Con el fin de analizar la correlación de las variables X1 y X2 con la variable Y, se diagramó la matriz de correlaciones presente en la figura 4.8.1. Se puede observar que el tamaño de las alas (variable X1) está correlativamente ligadas al tiempo el cual tarda el objeto en caer (var. Y1). Para constatar esta correlación se observó el $r=0.916$, lo indica un grado alto de relación correlativa entre las variables X1 e Y1. El valor de $p=0$ es menor al valor de significación, por lo que se puede rechazar la hipótesis nula y por ende el modelo es significativo.

Para el análisis de la correlación entre las variables X2 e Y, no se puede constatar gráficamente la correlatividad de dichas variables. A su vez se observó que el $r=-0,136$ el cual cuanto más se aproxima a cero, más débil es la relación lineal. El valor de $p=0,674$ es inferior al de alpha, por lo que no se rechaza la hipótesis nula. Por ende no se puede afirmar que el modelo es significativo.

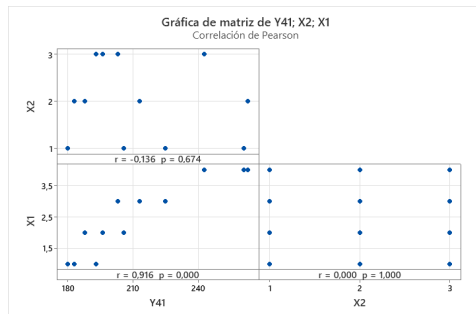


Figura 4.8.1

4.9) La figura 4.9.3 muestra que aunque X1 sea significativo, el coeficiente X2 no lo es ya que su valor-p es mayor a 10%.

Analizando el p-value obtenido a partir de la tabla ANOVA de regresión observamos que el p-value es menor al 0,1 por ende la hipótesis global significativa, el MPO propuesto funciona.

Analizando los p-valores obtenidos de la tabla ANOVA se puede constatar que de acuerdo a la prueba de significatividad para los parámetros, para la variable X1 el p-value es menor a 0,1 y para la variable X2 es mayor, por ende la variable X1 es significativa y la variable X2 no.

Por ende, se puede concluir que β_1 es estadísticamente significativo con una probabilidad de equivocarnos del 10%. En base al resultado de este ensayo, se debe retener el parámetro β_0 para el armado del modelo.

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	7968,7	3984,4	26,94	0,000
X1	1	7797,6	7797,6	52,73	0,000
X2	1	171,1	171,1	1,16	0,310
Error	9	1330,9	147,9		
Total	11	9299,7			

Figura 4.9.1

Coefficientes

Término	Coef	EE del coef.	Valor T	Valor p	FIV
Constante	165,1	12,2	13,58	0,000	
X1	22,80	3,14	7,26	0,000	1,00
X2	-4,63	4,30	-1,08	0,310	1,00

Figura 4.9.2

4.10) En los ajustes resultantes de la tabla se obtuvo un valor de S de 12,1607 y R-cuadrado(ajustado) de 82,51%, valores altamente significativos. Pero aun así siguen siendo menores que los valores del modelo de RLSD.

Resumen del modelo

S	R-cuadrado	R-cuadrado(ajustado)	R-cuadrado (pred)
12,1607	85,69%	82,51%	73,71%

Figura 4.10.1

4.11) Como se ve en la figura 4.11.2, el valor del p-value es mayor al $\alpha=15\%$ considerado en la prueba de normalidad por ende no se rechaza H_0 y los residuos siguen una distribución normal. Además, como ocurrió en el punto 4.5) en el gráfico de residuos en función de los valores ajustados, puede verse que los residuos obtenidos se encuentran centrados en el valor nulo. De esta manera, no se tiene evidencia gráfica que implique la refutación del supuesto. En este caso se valida el supuesto de esperanza residual nula.

En la figura 4.11.1 en el gráfico de residuos en función del orden no se nota ningún patrón marcado de los residuos. Esta observación brinda indicios de que los residuos se encuentran incorrelacionados y la varianza de los mismos se mantiene constante. Así se valida el supuesto de homocedasticidad y el de independencia.

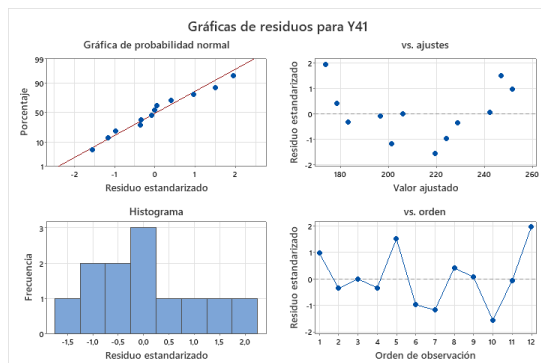


Figura 4.11.1

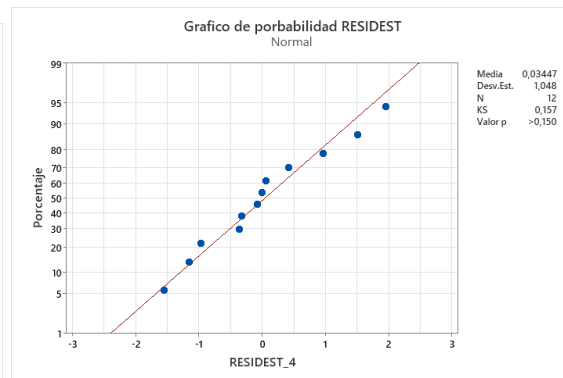


Figura 4.11.2

4.12) El modelo mejor ajustado para los tiempos de vuelo en función de las variables independientes es el modelo de RLSO cuyos valores ajustados de R fueron los más significativos y sus cumplieron con la condición de ser menor al α . Además el error representado por S también fue el menor.

EXPERIMENTO 5

5.1) En este experimento, a diferencia del de 2MI se trabaja con muestras de a pares y se quiere demostrar como la diferencia de las alas de los helicópteros pueden hacer que varíe el análisis de la muestra. Se doblan las alas de uno de los helis del par y se miden los tiempos de vuelo de cada uno.

5.2) La variable aleatoria de este experimento es la diferencia entre los tiempos de vuelo de cada par. Los supuestos son que sigue una distribución normal y que son valores independientes.

5.3) En la figura 5.3.1 a partir del cálculo de residuos se realiza de prueba de normalidad con Kolmogorov.

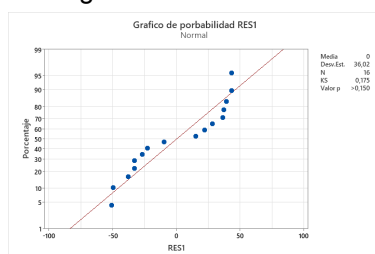


Figura 5.3.1

5.4)

5.5)

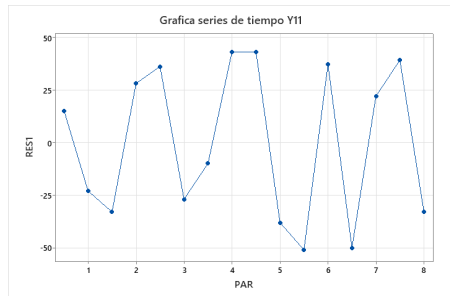


Figura 5.4.1

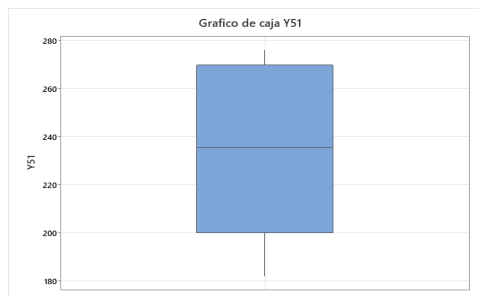


Figura 5.5.1

5.6)

Estadísticas descriptivas

Error estándar de la media IC de 90% para μ				
N	Media	Desv.Est.		
16	232,88	36,02	9,00	(217,09; 248,66)

μ : media de población de Y51

Figura 5.6.1

5.7)

Método

μ_1 : media de población de Y51 cuando ALA = D

μ_2 : media de población de Y51 cuando ALA = E

Diferencia: $\mu_1 - \mu_2$

Se presupuso igualdad de varianzas para este análisis.

Figura 5.7.1

Estadísticos descriptivos: Y51

Error estándar de la media				
ALA	N	Media	Desv.Est.	
D	8	199,9	13,6	4,8
E	8	265,9	10,2	3,6

Figura 5.7.2

Prueba

Hipótesis nula $H_0: \mu_1 - \mu_2 = 0$

Hipótesis alterna $H_1: \mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
-10,95	14	0,000

Figura 5.7.3

Estimación de la diferencia

Desv.Est. IC de 90% para la diferencia agrupada		
Diferencia	agrupada	diferencia
-66,00	12,05	(-76,61; -55,39)

Figura 5.7.4

No se logró completar el experimento 5, pero se dejaron los gráficos planteados.

ANEXO

Para determinar la altura se utilizó una cinta métrica y para visualizarlo con mayor facilidad se marcó con cinta de papel el lugar desde lanzamiento hasta el piso.

En cada lanzamiento el integrante A se colocó de frente para verificar la altura, si estaba alineado respecto al eje del objetivo y si el helicóptero caía dentro del objetivo. El integrante B se ocupó de lanzar los helicópteros. El integrante C se ocupó de verificar si estaba alineado respecto a el otro eje, comandar el lanzamiento y cronometrarlo. El integrante D se ocupó de registrar el tiempo de lanzamiento en centímetros y si cayó dentro o no del objetivo. Como se muestra en la figura 6.1. Se siguió esta distribución para el resto de los experimentos. Una vez registrados los datos se colocaron en el software estadístico. En la figura 6.2 se ve con claridad el momento exacto en el que un helicóptero es lanzado.

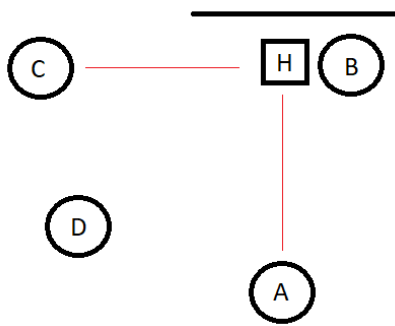


Figura 6.1

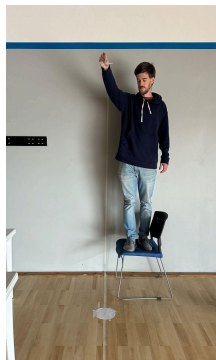


Figura 6.2

Para evitar la máxima cantidad de interferencia de y entre los experimentos se realizaron todos los experimentos el mismo día y dentro de un espacio cerrado.

En la figura 6.3 se ve el croquis A hecho del papel con gramaje 70 g/m².

En la figura 6.4 se ve el croquis A hecho de cartulina.

En la figura 6.5 se ve el croquis B.

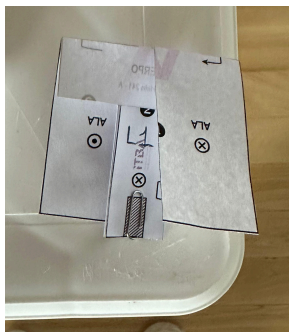


Figura 6.3

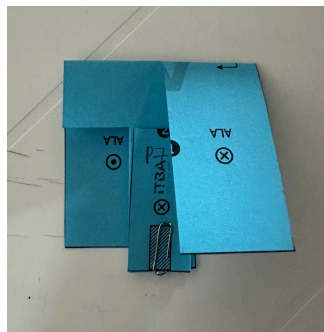


Figura 6.4



Figura 6.5

En la figura 6.6 se observa la disposición de los helicópteros previa a su lanzamiento.



Figura 6.6

Bibliografía

Box, G.E.P., Hunter, J.S. & Hunter, W.G. (2005). *Statistics for Experimenters*. Second Edition. Hoboken: Wiley.

Devore, J. (2019). *Introducción a la probabilidad y estadística para ingeniería y ciencias*. Novena edición. México: Cengage.

Montgomery, D.C, Runger, G.R. and Hubele, N.F. (2011). *Engineering Statistics*. Fifth Edition. Hoboken: John Wiley & Sons.

Montgomery, D.C.; Peck, E.A. & Vining, G.G. (2012). *Introduction to Linear Regression Analysis*. Fifth Edition. Hoboken: John Wiley & Sons.

Walpole, R.E.; Myers, R.H.; Myers, S.L. y Ye, K. (2007). *Probabilidad y estadística para ingeniería y ciencias*. Octava edición. México: Pearson.