



# TALLER: USO DE MINITAB PARA PRINCIPIOS DE INVESTIGACIÓN

**Prof. José Neville Díaz Caraballo, MS**  
**Departamento de Matemáticas**  
**UPR Aguadilla**

# INTRODUCCIÓN A MINITAB



Minitab - IGS-Avaluo-Precalculo.MPJ

File Edit Data Calc Stat Graph Editor Tools Window Help

Session

```
L93      669.0
LA1      669.0
M21      669.0
M76      642.0
M91      657.0
M92      725.0
M93      647.0
```

**Descriptive Statistics: aprom**

Variable	seccion	Mean	StDev	CoefVar	Q1	Median	Q3
aprom	F56	585.5	47.4	8.09	*	585.5	*
	L01	602.9	55.0	9.13	566.0	583.5	647.5
	L11	611.6	54.7	8.94	574.8	612.5	649.5

Worksheet 5 \*\*\*

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
	IES	ICE	Pre-Prueba	Post-Prueba	WebAssign	Prom								
1	0.00	200	2	12	793	74								
2	1.00	350	3	12	243	4								
3	2.00	500	3	10	238	11								
4	3.00	650	2	12	655	48								
5	4.00	800	5	11	345	17								
6	0.80	320	5	11	102	7								
7	1.20	380	4	12	732	67								
8	1.60	440	0	10	621	69								
9	2.20	520	7	12	770	46								
10	2.70	605	3	12	691	80								

Current Worksheet: Worksheet 5



# MINITAB

- Pantallas
- Hojas de Cálculo
- Proyectos
- Menú



# CONJUNTO DE DATOS JFKLAX

MATH.UPRAG.EDU/JFKLAX.MTW

- Este conjunto de datos tiene los tiempos de atrasos en salidas y llegadas de los aeropuertos JFK y Los Ángeles.



# ESTADÍSTICA DESCRIPTIVA

## TENDENCIA CENTRAL

- Media
- Mediana
- Media Podada
- Media Ponderada



# ESTADÍSTICA DESCRIPTIVA

## VARIABILIDAD

- “The median is not the message”
- No se comparan las desviaciones estándar

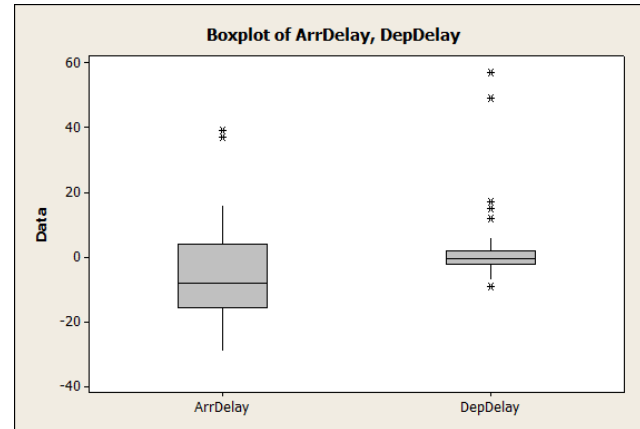
$$s_x = \sqrt{\sum \frac{(x - \bar{x})^2}{n - 1}}$$

- Coeficiente de variación

$$C.V = \frac{s_x}{\bar{x}}$$



- “Boxplot” importancia



- Correlación

$$r = \frac{1}{n - 1} \sum \frac{(x - \bar{x})}{s_x} \cdot \frac{(y - \bar{y})}{s_y}$$

- La correlación no implica causalidad es una frase en la ciencia y la estadística que pone de relieve que una correlación entre dos variables no implica necesariamente que uno causa el otro.



# RECOMENDACIONES PARA UN BUEN DISEÑO DE EXPERIMENTO

- Aleatoriedad : El elemento crucial
- Aleatoriedad en el orden de los tratamientos
- Aleatoriedad del tipo de tratamiento
- Grupo control
- Efecto placebo
- “Blinding”
- “Double Blind”





# DISEÑO DE EXPERIMENTOS

- Los tres elementos básicos del diseño de experimentos son:
- Aleatoriedad: Reducir el efecto de lo que no podemos controlar. Podemos obtener un estimado del error y luego concluir si las diferencias observadas y experimentales eran significativas.
- Repetición: (Una réplica de un experimento).  
Esto nos permite obtener un estimado del error.
- Bloquear: El efecto de un elemento ruidoso pero controlable se cuantifica aparte.



# DIAMETER OF SAND GRANULES VS. SLOPE ON BEACH

- Calcular la correlación entre X & Y
- Crear un diagrama de puntos.
- Crear un modelo de regresión lineal.
- Verificar que el modelo es bueno.



## REGRESIÓN LINEAL SIMPLE JFKLAX

- Calcular la correlación entre ArrDelay y DepDelay
- Crear un diagrama de puntos.
- Crear un modelo de regresión lineal.
- Verificar que el modelo es bueno.



## PRUEBA DE T PAREADA

- Ejemplo de prueba en bloques o medidas repetidas
- “Baseline” ¿Cuánto deseas demostrar que mejora?
- ¿Por qué las pruebas pareadas son mejores?
- $\text{Cov}(X, Y) = \text{Var}(X) - \text{Var}(Y) - \text{Cov}(X, Y)$
- La Prueba pareada es más poderosa que la de muestras independientes



- Comparar Pre Prueba BoxPlot vs Post Prueba Boxplot
- Recordar tratamiento y experimento completamente aleatorio
- Visitar <http://math.uprag.edu/Taller.MPJ>



## OBJETIVOS

- Conocer las presunciones del ANOVA
- Verificar que los datos cumplan con las presunciones
- Analizar e interpretar los resultados
- Conocer e interpretar la interacción entre variables



- ANOVA fue desarrollado por el estadístico inglés Ronald A. Fisher (1890 – 1962)
- A pesar de su uso constante muchos fallan al no verificar las presunciones de ANOVA.



## USO CORRECTO DEL ANOVA

- ANOVA = Análisis de variación

El término *análisis de varianza* es un nombre poco apropiado. En ANOVA usamos las cantidades de varianza para estudiar la igualdad o no igualdad de las medias de la poblacionales, pensando que cada nivel del factor representan datos de una población distinta





## PRESUNCIONES DEL “ONE WAY ANOVA”

- Las muestras provienen de una población aproximadamente normal. (La prueba es robusta)
- Las poblaciones tienen la misma variabilidad. (La prueba es robusta, tamaño de la muestra igual o aproximadamente igual)
- Los datos son obtenidos usando un muestreo aleatorio simple.
- Las muestras son independientes.
- Solo hay un factor.



MATH.UPRAG.EDU/TALLERSANDRA.MPJ

- Fortaleza textil vs porcentaje algodón

Factor Algodón

5%

10%

15%

20%

25%



## IMPORTANTE

- P-Value = es el nivel significancia observada
- Normality Plot
- La presunción de independencia es crítica. Si el factor de interés son asignadas al azar, al igual que las unidades experimentales, la presunción de independencia es usualmente satisfecha.



- Prueba de igualdad de varianza
- En un estudio de investigación, el error de tipo I también denominado error de tipo alfa ( $\alpha$ ) o falso positivo, es el error que se comete cuando el investigador rechaza la hipótesis nula ( $H_0$ ) siendo esta verdadera en la población.



- En un estudio de investigación, el error de tipo II, también llamado error de tipo beta ( $\beta$ ) ( $\beta$  es la probabilidad de que exista este error) o falso negativo, se comete cuando el investigador no rechaza la hipótesis nula siendo esta falsa en la población.



## HIPÓTESIS NULA Y ALTERNA

$$H_0 : \mu_0 = \mu_1 \dots = \mu_n$$

$$H_a : \text{Al menos una } \mu_i \neq \mu_j$$

¿Por qué no hacer varias pruebas  $t$ ?



- No es lo mejor porque nos va a aumentar el error Tipo I, Por ejemplo, suponga que deseamos probar la igualdad de 5 promedios usando pruebas  $t$ .
- Entonces existen 10 posibles combinaciones y si la probabilidad de aceptar correctamente la hipótesis nula para cada par es  $1-\alpha=.95$
- La probabilidad de aceptar correctamente la hipótesis nula para las 10 pruebas es  $.95^{10} = .60$  si las pruebas son independientes. Esto aumenta sustancialmente el error tipo I.



- Las comparaciones de interés deben ser planificadas antes. No podemos ver qué sugieren los datos para luego determinar que comparación me conviene realizar.
- Tampoco podemos seleccionar un  $\alpha$  conveniente de .10. Debemos establecer las reglas del juego con anticipación, de lo contrario estaríamos cometiendo fraude científico.





## MATH.UPRAG.EDU/TALLERSANDRA.MPJ

- Son el promedio del volumen de 12 latas de refrescos iguales para 4 tipos de refrescos distintos. (Coca Cola Regular, Coca Cola Dieta, Pepsi Regular y Pepsi Dieta).
- $H_0 : \mu_0 = \mu_1 \dots = \mu_n$
- $H_a : \text{Al menos una } \mu_i \neq \mu_j$

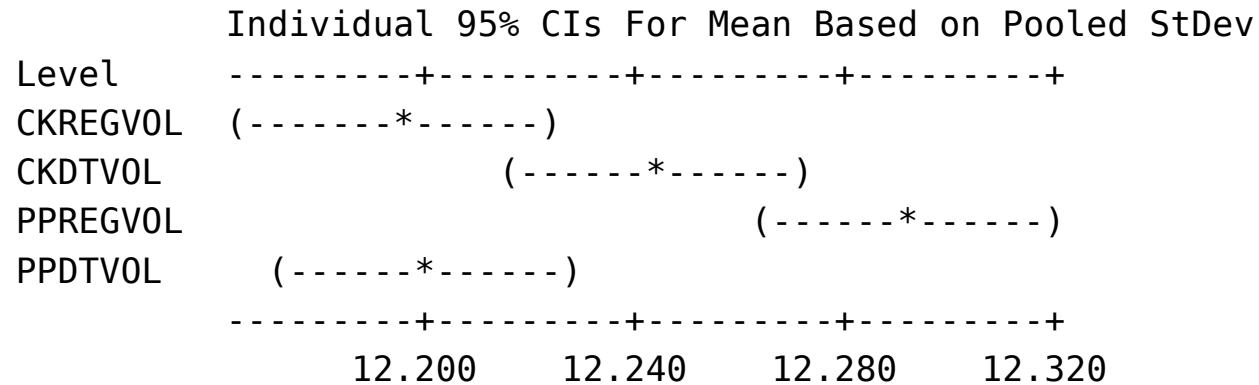


**One-way ANOVA: CKREGVOL, CKDTVOL, PPREGVOL, PPDTVOL**

Source	DF	SS	MS	F	P
Factor	3	0.22132	0.07377	9.43	0.000
Error	140	1.09528	0.00782		
Total	143	1.31660			

S = 0.08845    R-Sq = 16.81%    R-Sq(adj) = 15.03%

Level	N	Mean	StDev
CKREGVOL	36	12.1944	0.1145
CKDTVOL	36	12.2444	0.0809
PPREGVOL	36	12.2917	0.0906
PPDTVOL	36	12.2000	0.0586



Pooled StDev = 0.0885



## VERIFICAR NORMALIDAD

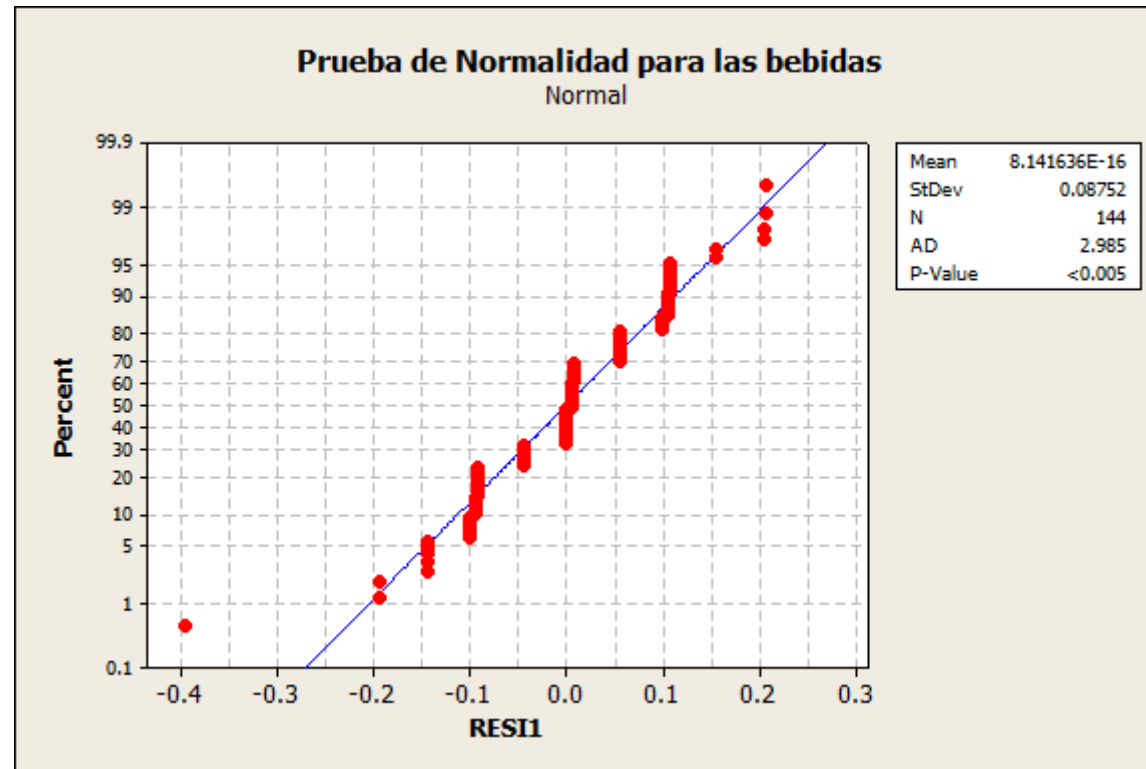
- “Normality Plot” de residuales
- Análisis de residuales
- Prueba de igualdad de varianza (ANOVA test for equal variance)
- Bartletts presume normalidad, Levenes no presume normalidad



Ho: HAY NORMALIDAD

Ha: NO HAY NORMALIDAD

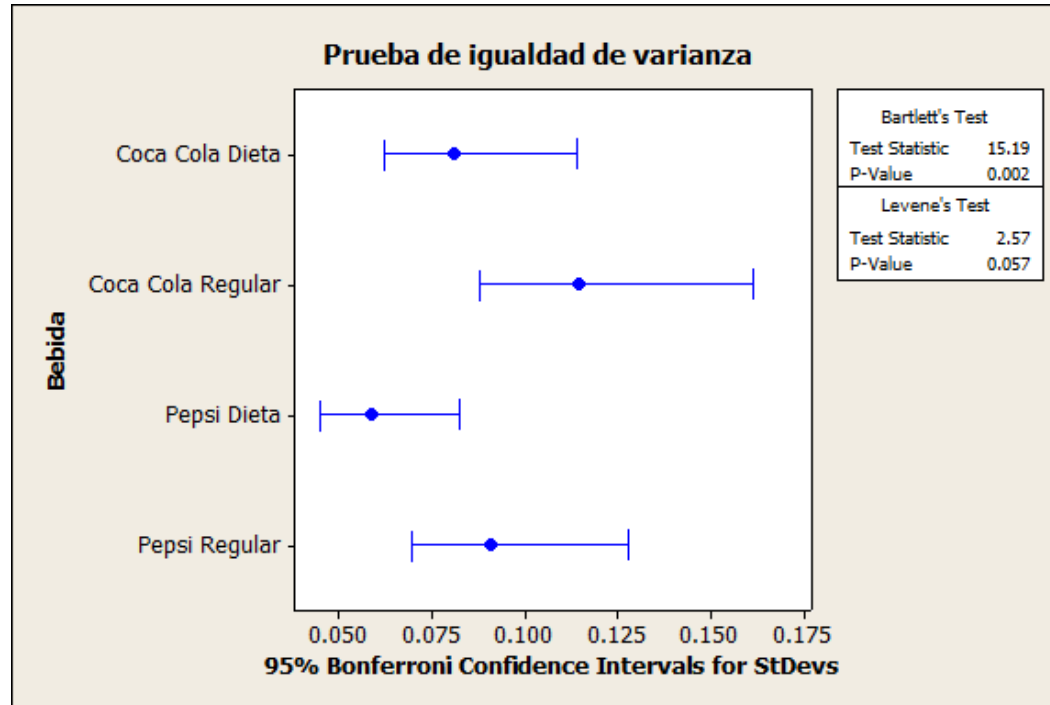
- Verificar normalidad



- P-value  $<.05$  se rechaza Ho, por lo tanto los datos no provienen de una distribución normal



- Verificar igualdad de varianza



- Ambas pruebas rechazan la  $H_0$ , es decir, al menos un grupo tienen una variabilidad significativamente distinta.



# KRUSKAL - WALLIS

- **Kruskal-Wallis Test: Volumen versus Bebida**

Kruskal-Wallis Test on Volumen

Bebida	N	Median	Ave Rank	Z
Coca Cola Dieta	36	12.30	79.0	1.08
Coca Cola Regular	36	12.20	59.6	-2.14
Pepsi Dieta	36	12.20	56.3	-2.70
Pepsi Regular	36	12.30	95.2	3.76
Overall	144		72.5	

H = 20.39    DF = 3    P = 0.000

H = 23.47    DF = 3    P = 0.000    (adjusted for ties)



## EJERCICIO DE PRÁCTICA

- Se desea verificar si el promedio excavación en cuatro lugares arqueológicos son distintos.

$$H_0: \mu_0 = \mu_1 \dots = \mu_n$$

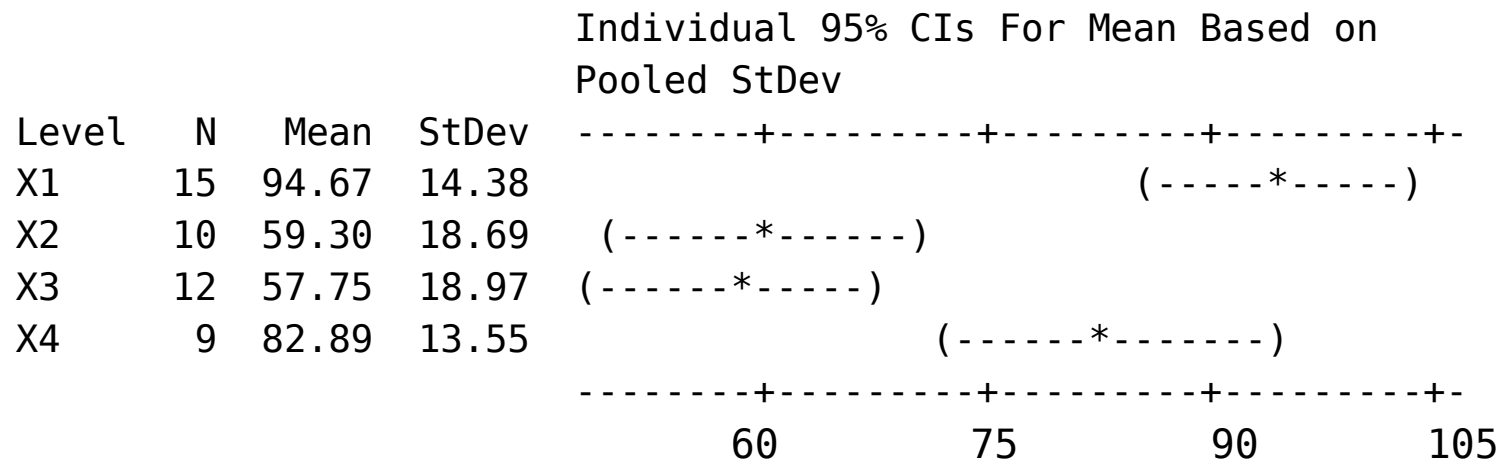
$$H_a: \mu_i \neq \mu_j$$



## One-way ANOVA: X1, X2, X3, X4

Source	DF	SS	MS	F	P
Factor	3	12397	4132	15.14	0.000
Error	42	11465	273		
Total	45	23862			

S = 16.52    R-Sq = 51.95%    R-Sq(adj) = 48.52%



Pooled StDev = 16.52





## TWO WAY ANOVA

- Las presunciones son las misma del One Way ANOVA porque es una generalización del mismo.
- Lo que añade es que podemos verificar el efecto de varias variables independientes. Además, podemos verificar si existe interacción.



## AFILIACIÓN POLÍTICA

- Respuesta: Porcentaje de votantes en una elección.
- Factor 1: Condados en Montana  
Factor 2: Afiliación Política

Referencia: County and City Data Book U.S.  
Dept. of Commerce



- Verificar normalidad
- Verificar igualdad de varianza
- Analizar los resultados
- En este ejemplo, al tener una sola observación por celda no podemos verificar la interacción.



## TRAMPA PARA LAS POLILLAS

- Respuesta: número de polillas encontradas en la trampa después de 48 horas

Factor 1: Localización de la trampa en el árbol.

Factor 2: Tipos de atracción en la trapa (olor, azúcar, químico)



- Verificar normalidad
- Verificar igualdad de varianza
- Analizar los resultados
- En este ejemplo, al tener repeticiones en las observaciones por celda, podemos verificar la interacción.



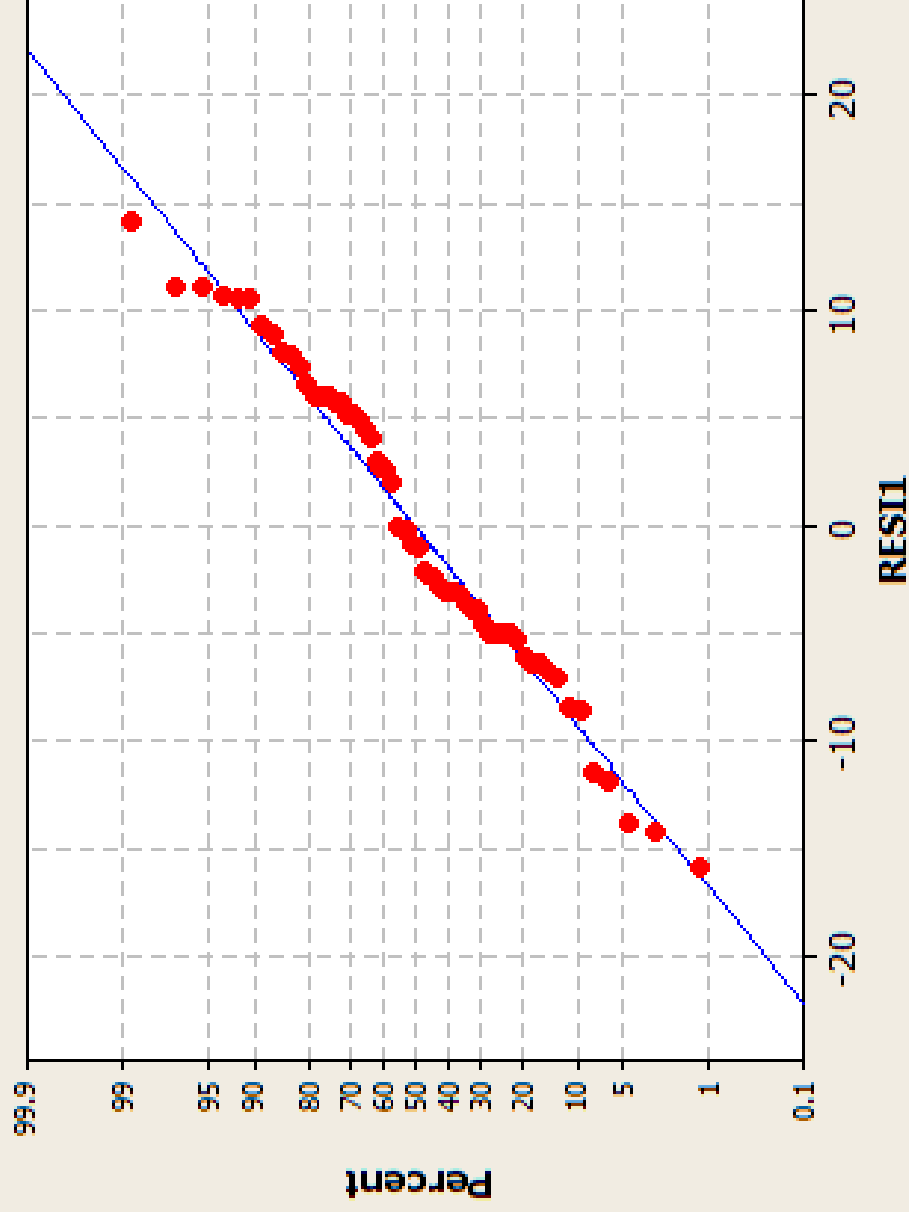
## Two-way ANOVA: Quantity versus Location, Type

Source	DF	SS	MS	F	P
Location	3	1981.38	660.461	10.45	0.000
Type	2	113.03	56.517	0.89	0.416
Interaction	6	114.97	19.161	0.30	0.932
Error	48	3033.60	63.200		
Total	59	5242.98			

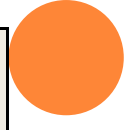


# Prueba de Normalidad para las bebidas

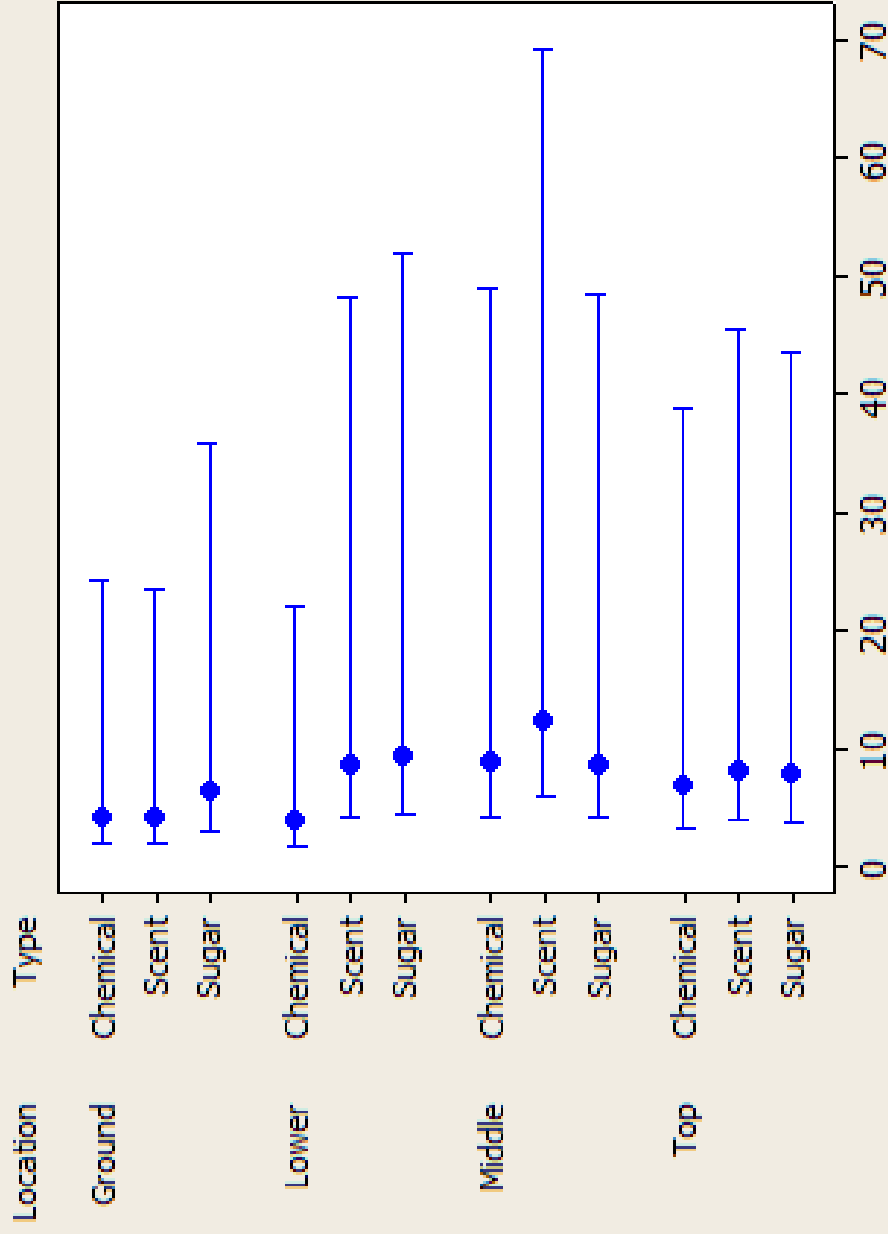
Normal



Mean	3.552714E-16
StDev	7.171
N	60
AD	0.589
P-Value	0.119



## Prueba de igualdad de varianza



Bartlett's Test	
Test Statistic	8.96
P-Value	0.626
Levene's Test	
Test Statistic	0.64
P-Value	0.788

95% Bonferroni Confidence Intervals for StDevs





# PRÓTESIS

	Tipo de Prótesis		
Edad	A	B	C
Menos de 20	7	9	10
20-29	8	9	10
30-39	9	9	12
40-49	10	9	12
50 +	11	12	14



# ¿PREGUNTAS?

Es verdad muy cierta, que si no está en nuestro poder el discernir las mejores opiniones, debemos seguir las más probables.

René Descartes

