

Diseño de Experimentos

Ejemplo: Datos de Duraznos

M. Sc. Marcelo E. ALBERTO

Tabla de Datos

- Se trabaja con la tabla de datos de duraznos utilizada en el primer trabajo del curso
- La tabla contiene datos de las variables
 - Peso (en g)
 - Diámetro Sutural (en mm)
 - Diámetro Contrasantal (en mm)
- Los duraznos provienen de tres zonas diferentes

Objetivos

1. Se dese saber si el **peso** de los duraznos varía significativamente de una zona a otra
2. Se dese saber si el **diámetro sutural** de los duraznos varía significativamente de una zona a otra
3. Se dese saber si el **diámetro contrasutural** de los duraznos varía significativamente de una zona a otra

Análisis de los Datos

- Se piensa el problema como un diseño experimental en el cual los “tratamientos” son las tres zonas consideradas.

3 Zonas → 1 Factor con 3 Niveles

- En rigor, no se trata de un experimento sino de un estudio *observacional* ya que el factor no es administrado por los investigadores.

Pero el modelado matemático es el mismo!!!

Objetivo 1: Peso

- La variable respuesta es el peso del fruto
- Podemos plantear la siguiente hipótesis

El peso de los duraznos, en promedio, varía entre las zonas

(hipótesis de trabajo)

- Asumiendo que el peso es una v.a. con promedios μ_1 , μ_2 , μ_3 para cada zona (desconocidos), podemos plantear las 2 hipótesis estadísticas:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

H_1 : al menos dos promedios difieren

Objetivo 1: Peso

- La técnica estadística para poner a prueba estas hipótesis se denomina *Análisis de la Varianza* y es conocida por sus siglas **ADEVA**, **ANAVA** o **ANOVA**.
- Recordemos que las pruebas estadísticas siempre se hacen sobre la hipótesis nula.
- Notemos que, **en este caso**, la hipótesis nula es contraria a la hipótesis de trabajo: *El peso de los duraznos, en promedio, varía entre las zonas* (la hipótesis nula afirma lo contrario).
- Realizaremos el ANOVA en Statgraphics y en R.

Objetivo 1: Peso

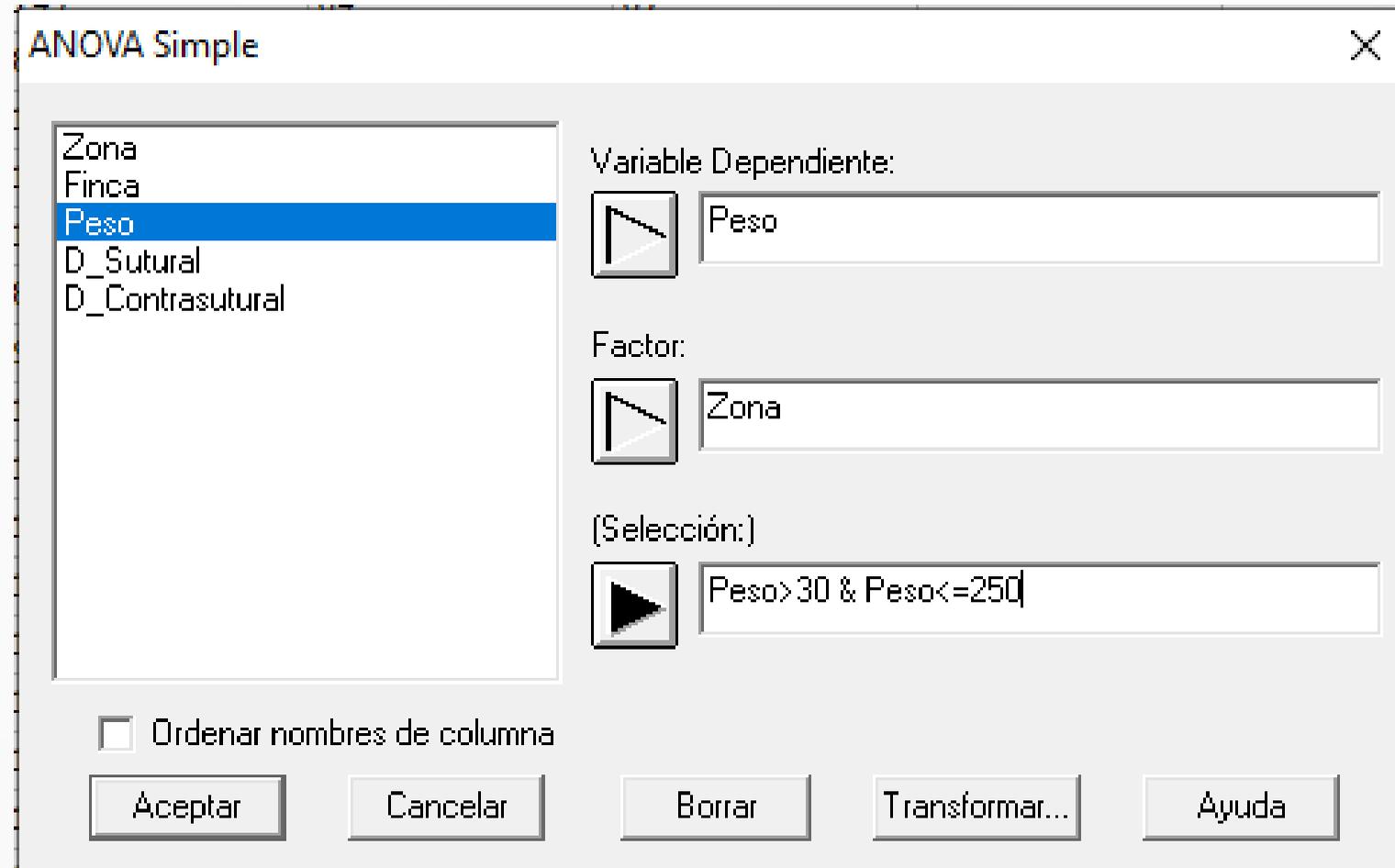
ANOVA en Statgraphics

The screenshot shows the Statgraphics Centurion interface. The 'Comparar' menu is open, showing 'Análisis de Varianza' selected. A sub-menu is also open, with 'ANOVA Simple...' highlighted. The background shows a data table with columns 'Zona' and 'Finca'.

	Zona	Finca				Col_6	Col
1	1	301					
2	1	1502	87	57	53		
3	1	301	139	65	64		

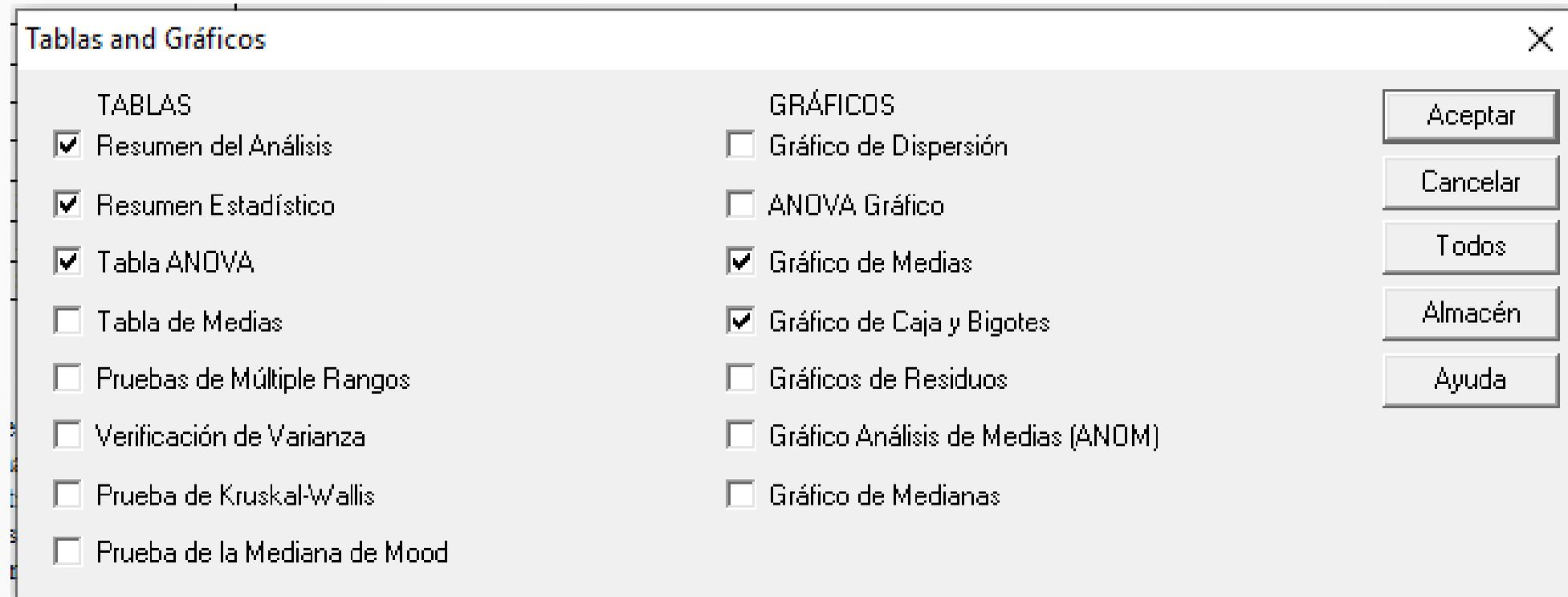
Objetivo 1: Peso

ANOVA en Statgraphics



Objetivo 1: Peso

ANOVA en Statgraphics



Objetivo 1: Peso

ANOVA en Statgraphics

ANOVA Simple - Peso por Zona (Peso>30 & Peso<=250)

ANOVA Simple - Peso por Zona (Peso>30 & Peso<=250)

Variable dependiente: Peso
Factor: Zona
Selección de la Variable: Peso>30 & Peso<=250

Número de observaciones: 1114
Número de niveles: 3

ANOVA Simple - Peso por Zona (Peso>30 & Peso<=250)

Resumen Estadístico para Peso

<i>Zona</i>	<i>Recuento</i>	<i>Promedio</i>	<i>Desviación Estándar</i>	<i>Mínimo</i>	<i>Máximo</i>
1	605	130,402	31,8043	42,0	232,0
2	129	147,829	37,5876	80,0	235,0
3	380	139,745	39,2811	50,0	240,0
Total	1114	135,607	35,7032	42,0	240,0

Objetivo 1: Peso

ANOVA en Statgraphics

ANOVA Simple - Peso por Zona (Peso>30 & Peso<=250)

Tabla ANOVA para Peso por Zona

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	42169,9	2	21085,0	17,02	0,0000
Intra grupos	1,37659E6	1111	1239,06		
Total (Corr.)	1,41876E6	1113			

Probabilidad de
Obtener un valor
Igual o más extremo
Siendo H_0 Verdadera

$$\bullet \text{ valor} - p < \alpha$$



Rechazar H_0

$\alpha :=$ nivel de significación

$$P[\text{Rechazar } H_0 | H_0 \text{ es verdadera}]$$

Objetivo 1: Peso

ANOVA en Statgraphics

Si rechazamos H_0 entonces aceptamos H_1 , por lo tanto...

$$\mu_1 \neq \mu_2 \wedge \mu_1 \neq \mu_3 \wedge \mu_2 \neq \mu_3$$

o bien

$$\mu_1 \neq \mu_2 \wedge \mu_1 \neq \mu_3 \wedge \mu_2 = \mu_3$$

o bien

$$\mu_1 = \mu_2 \wedge \mu_1 \neq \mu_3 \wedge \mu_2 \neq \mu_3$$

o bien

⋮

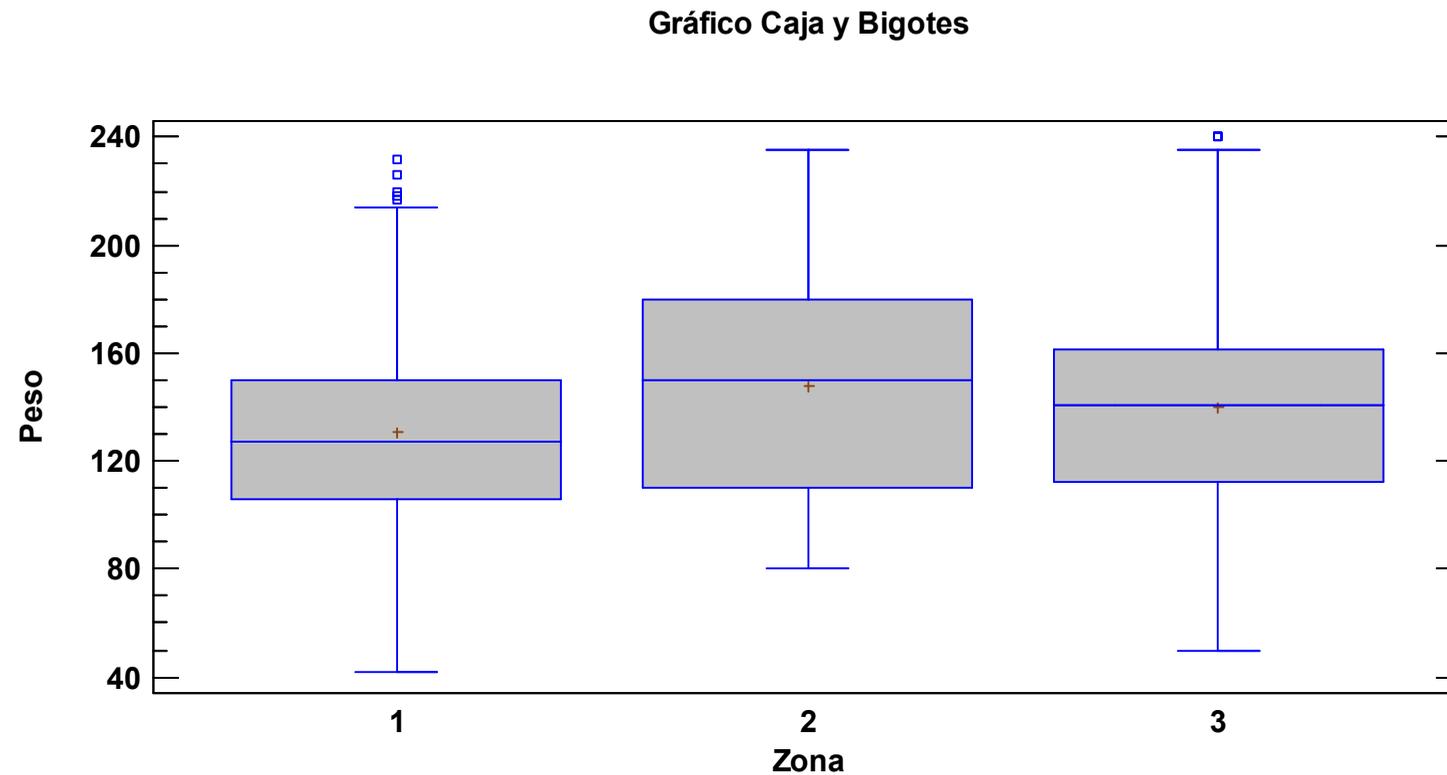
M.Sc. Marcelo E. Alberto

¡OMG!
COMBINATORIA!!!



Objetivo 1: Peso

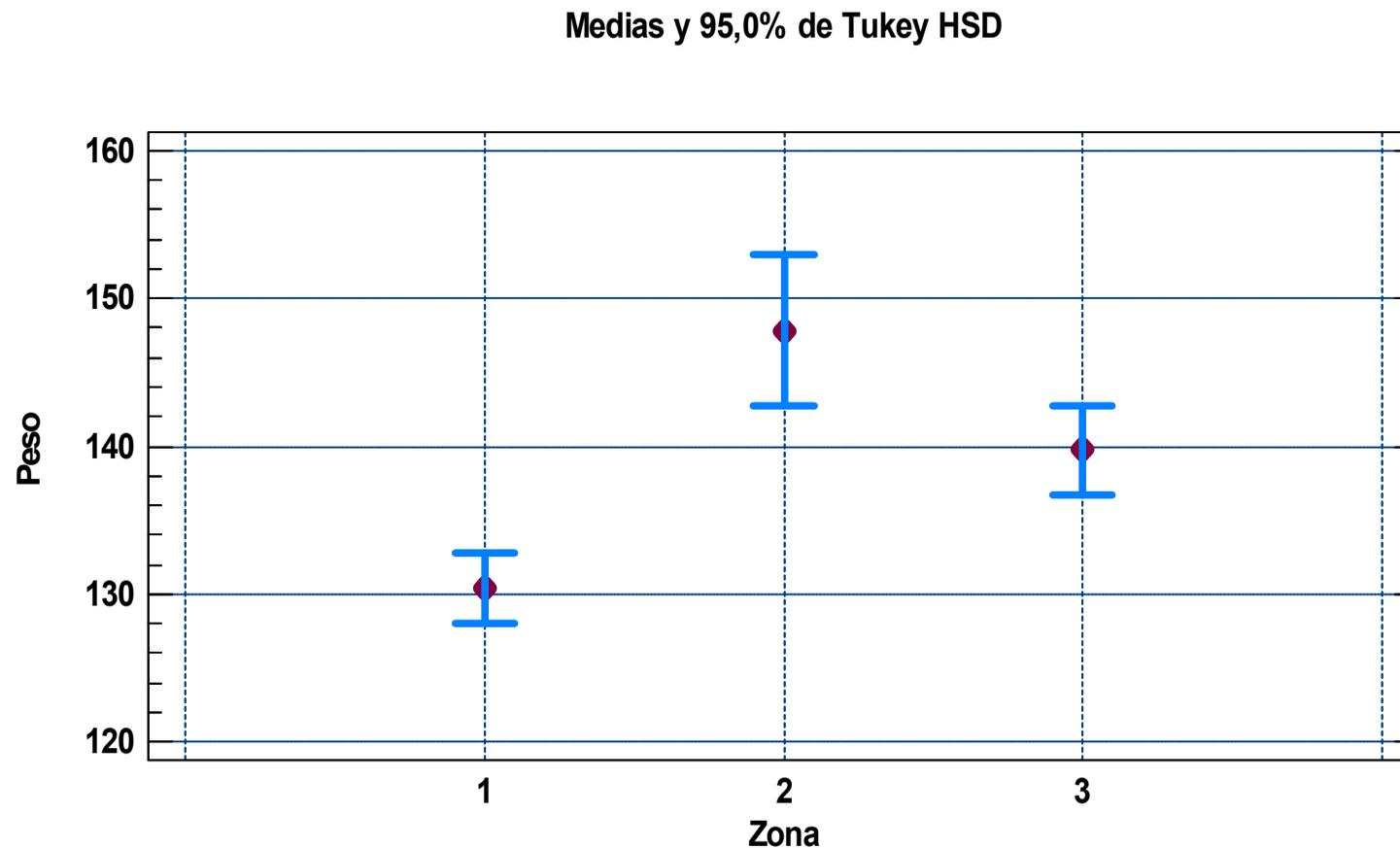
ANOVA en Statgraphics



Objetivo 1: Peso

ANOVA en Statgraphics

Son diferentes dos promedios cuyos Intervalos de Confianza NO se solapan



Objetivo 1: Peso

ANOVA en *R*

- Traemos los datos a R

```
read.table("durazno406.txt", header=TRUE) → durazno
```

- Declaramos la tabla “durazno” en la memoria operativa

```
attach(durazno)
```

- Ordenamos el ANOVA

```
aov(Peso~Zona) → anova.durazno
```

- Veamos los resultados

```
summary(anova.durazno)
```

Objetivo 1: Peso

ANOVA en *R*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Zona	2	42170	21085	17.02	5.26e-08 ***
Residuals	1111	1376594	1239		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Objetivo 1: Peso

CONCLUSIONES

- La evidencia muestral sugiere que los duraznos de la Zona 1 tienen un promedio significativamente diferente del resto.
- No existe evidencia para concluir que los duraznos de las Zonas 2 y 3 difieran significativamente

ZONA	PROMEDIO	GRUPO
1	130	A
2	148	B
3	140	B

Ejercicios

- Verificar los objetivos 2 y 3
- Con los datos de Meteorología: utilizando ANOVA pruebe que la Temperatura Horaria varía en promedio entre los meses del año

Mmm...
No era sólo una charla





Muchas gracias !!!