UNIVERSIDAD DE ALMERÍA

Grado en Matemáticas

ESCUELA POLITÉCNICA SUPERIOR Y FACULTAD DE CIENCIAS EXPERIMENTALES

División de Ciencias Experimentales

Trabajo Fin de Grado

Diseño unifactorial con covariable

Curso académico: 2013/2014

Convocatoria: Junio

Autora: María Dolores Fernández de Henestrosa González

Tutor: Ignacio Jesús Martínez López

Índice general

Objetivo del trabajo

El análisis de la covarianza es un procedimiento de control estadístico que permite eliminar el ruido que generan variables no controladas sobre la variable respuesta en un diseño estadístico de experimentos. El análisis de la covarianza es una técnica que en un gran numero de ocasiones, es útil para mejorar la precisión del experimento.

A la variable no controlada, pero que se puede observar junto a la variable respuesta, se le denomina variable concomitante o covariable. Es muy importante destacar el carácter cuantitativo de la covariable, a diferencia del carácter cualitativo del factor habitualmente considerado en el diseño experimental, además debe existir una relación lineal entre la variable respuesta y la covariable. El procedimiento estudiado se basa en analizar la posibilidad de interacción entre las variables cualitativas y cuantitativas en un modelo lineal. Su análisis requiere del estudio matemático de un modelo, combinación del diseño de experimentos y del análisis de regresión de forma conjunta.

En este trabajo se estudia el diseño unifactorial con covariable, es decir, el estudio de la posible influencia de una variable cualitativa (factor) sobre una variable respuesta observada de forma simultánea a una variable cuantitativa (covariable). Para ello se presenta el modelo lineal y los contrastes a resolver para concluir las posibles influencias. Se estiman los parámetros presentes en este modelo y se construye los estadísticos que permiten resolver los contrastes planteados.

El trabajo está estructurado de la siguiente forma: en el primer capítulo se presentan los conceptos fundamentales del diseño experimental, así como una breve evolución histórica. Se incluyen los principales resultados del diseño unifactorial completamente aleatorizado y del modelo lineal de regresión, por estar relacionado con el modelo a estudiar. El siguiente capítulo aborda el problema de interés en el trabajo, realizando los siguientes pasos: se presenta el modelo lineal y los contrastes a resolver, se estiman los parámetros y se construye los estadísticos que permite resolver dichos contrastes. Por ultimo, se presenta un caso real modelizado mediante el diseño unifactorial con covariable. Para su resolución se ha utilizado Statgraphics Centurion XVI

Capítulo 1

Diseño estadístico de experimentos

Nadie ha tenido tanto impacto en los principios estadísticos del diseño de experimentos como Ronald A. Fisher (1890-1962). En octubre de 1919, Fisher fue contratado en Rothamsted Experimental Station, Inglaterra. Su trabajo allí consistía en aplicar un exhaustivo análisis estadístico a los datos de investigaciones agrícolas.

Fue durante su ejercicio en Rothamsted, donde desarrolló y consolidó los principios básicos del diseño y análisis de datos experimentales, que hasta la fecha son necesarios para llegar a resultados validos. De 1919 A 1925 estudió y analizó experimentos relativos al trigo que se habían realizado desde 1843. De sus investigaciones estadísticas, de éstos y otros experimentos, Fisher desarrolló el análisis de la varianza y unificó sus ideas básicas sobre los principios del diseño de experimentos.

En 1926, Fisher publica en Journal of the Ministry of Agriculture of Great Britain, el primer resumen completo de sus ideas en el artículo "The Arragement of Fiel Experments". En este importante trabajo describe las tres componentes fundamentales de los diseños experimentales en el área de pruebas agrícolas: control local de las condiciones de campo para reducir el error experimental, replicación como un medio para estimar la varianza del error experimental y aleatorización para obtener una estimación válida de esa varianza.

Los desarrollos posteriores en diseños de experimentos fueron encabezados por George E. P. Box, quien trabajó como estadístico durante ocho años en la industria química en Inglaterra y desarrollo la metodología de superficies de respuesta, la cual incluye nuevas familias de diseños y una estrategia para la experimentación secuencial. En 1950,William G. Cochran, junto con Gertrude Mary Cox, desarrollan la descomposición en forma de cuadrados que hoy día realizamos. Cochran, en el trabajo "Analysis of Covariance: Its Nature and Uses"publicado en 1957 en Biometrics, introduce el concepto de covariable asociado a un diseño experimental y desarrolla lo que hoy conocemos como el análisis de la covarianza.

Entre 1950 y 1980 el diseño de experimentos se convierte en una herramienta que se aplica en el área de investigación y desarrollo. En los ochenta se da un gran impulso a la aplicación del diseño de experimentos, debido a éxito en calidad de la industria japonesa. El movimiento por la calidad, encabezado por Deming e Ishikawa, promovió el uso de la estadística en calidad,

donde el diseño de experimentos demostró su utilidad tanto para resolver problemas como para diseñar. En Japón destaca el trabajo de Genichi Taguchi, cuyos conceptos sobre diseño robusto tuvieron un impacto significativo en la academia del mundo occidental.

El objetivo básico de un experimento estadístico es estudiar el efecto que tiene un conjunto de variables cualitativas, denominadas *factores* sobre una variable de interés llamada *variable respuesta*. La aplicación de las técnicas del diseño experimental en las fases iniciales del desarrollo de un proceso puede suponer mejoras en el rendimiento del proceso, reducir la variabilidad y el tiempo de desarrollo, y lo más importante, reducir los costes globales.

Las etapas que todo diseño experimental debe seguir son las siguientes: primero identificar y enunciar el problema, después elegir los factores y los niveles, especificando el modelo y seleccionando los elementos (unidades experimentales) sobre los que se va a llevar a cabo la experimentación. Una vez hecho esto tenemos que seleccionar la variable respuesta y elegir del diseño experimental. Por último realizamos el experimento y un análisis estadístico de los datos, llegando a obtener las conclusiones y pudiendo dar recomendaciones sobre como actuar en el proceso.

Existen tres principios básicos a tener en cuenta en cualquier diseño de experimentos, que coinciden con los principios que enunció Fisher. La aleatorización, que consiste en asignar al azar todos los efectos de variables perturbadoras a las unidades experimentales. La homogeneidad estadística de las comparaciones, pudiendo conseguirse mediante la introducción de bloques en el diseño o mediante la factorización y por último la replicación del experimento, que consiste en repetir la experimentación bajo las mismas condiciones.

A continuación se muestran los conceptos básicos del diseño unifactorial completamente aleatorizado y del modelo de regresión lineal simple, modelos lineales a partir de los cuales se introduce el estudio del análisis de la covariable.

1.1. Diseño unifactorial completamente aleatorizado

El diseño unifactorial completamente aletatoriazado aparece a partir del análisis de la varianza. Es el diseño más sencillo que permite estudiar la influencia de un factor cualitativo sobre una variable respuesta observada, agrupada en k-grupos o tratamientos.

Sean $Y_1, Y_2, ..., Y_k$ variables aleatorias independientes, con $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, donde μ_i y σ^2 son parámetros desconocidos. Nótese que la varianza es constante en los k poblaciones consideradas. Supongamos que para cada población tomamos una muestra aleatoria simple, de tamaño n_i notada $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$. En esta situación podemos introducir el modelo lineal del diseño unifactorial completamente aleatorizado como

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i,$$

donde $\alpha_i = \mu_i - \mu$,es el efecto debido al tratamiento i-ésino, que puede considerarse como

desviación de la media de la población μ_i respecto a la media global μ, con

$$\mu = \frac{\sum_{i=1}^{k} n_i \mu_i}{n}$$

Los efectos debidos a los tratamientos verifican la siguiente condición:

$$\sum_{i=1}^{k} n_i \alpha_i = 0$$

Además ϵ_{ij} es el habitual error aleatorio, modelizado mediante una distribución normal

$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \qquad i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i.$$

El estudio de la influencia del factor sobre la variable respuesta se establece mediante la resolución del siguiente contraste:

$$\begin{cases} H_0 : \alpha_i = 0 \quad \forall i = 1, \ldots, k \\ H_1 : \text{algún } \alpha_i \neq 0 \end{cases}$$

La hipótesis nula establece que los efectos del factor son nulos, equivalentemente a contrastar que todas las medias μ_i se pueden suponer análogas, frente a la hipótesis alternativa, en la que algún efecto debido no debe+ considerarse nulo.

Para la estimación de los parámetros del modelo, son habituales los siguientes estadísticos muestrales:

- Suma y media de las observaciones del tratamiento i-ésimo, $Y_{i.}$ y $\bar{Y}_i$, respectivamente, son

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij} \qquad \bar{Y}_i = \frac{Y_{i.}}{n_i} = \frac{1}{n} \sum_{j=1}^{n_i} Y_{ij}$$

- Suma y media global de todas las observaciones, $Y_{..}$ y $\bar{Y}$respectivamente, son:

$$Y_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{k} Y_{i.} \qquad \bar{Y} = \frac{Y_{..}}{n} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n} \sum_{i=1}^{k} Y_{i.}$$

Los parámetros presentes en el modelo que se han de estimar son la media de todas las observaciones, μ, la media de la i-ésima población, μ_i, los efectos debidos a los tratamientos, α_i y la varianza del error, σ^2.
El método de mínimos cuadrados, consiste en minimizar la suma de los cuadrados de los residuos, es decir, minimizar la siguiente expresión

$$R = \sum_{i=1}^{k} \sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_{ij})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \mu - \alpha_i)^2$$

Obteniéndose los estimadores mínimo cuadráticos con las siguientes distribuciones:

$$\begin{aligned}\hat{\mu}_i &= \bar{Y} \sim \mathcal{N}(\mu_i, \sigma^2)\\ \hat{\mu} &= \bar{Y} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)\\ \hat{\alpha}_i &= \bar{Y}_i - \bar{Y} \sim \mathcal{N}\left(\alpha_i, \left(\frac{1}{n_i} - \frac{1}{n}\right)\sigma^2\right)\end{aligned}$$

Aplicando el método de máxima verosimilitud, se ha de maximizar la función de verosimilitud:

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(Y_{ij}-\mu_i)^2}{2\sigma^2}\right\}$$

se obtienen estimadores análogos para μ, μ_i y α_i y para la varianza σ^2 el estimador máximo verosímil es:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2.$$

A diferencia de los demás estimadores, este último estimador no es insesgado. El estimador máximo verosímil σ^2 puede expresarse en función de los residuos como

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}e_{ij}^2 \qquad i = 1,\ldots,k \quad j = 1,\ldots,n_i$$

Los n residuos del modelo no son independientes, ya que están sometidos a las restricciones

$$\sum_{j=1}^{ni} e_{ij} = \sum_{j=1}^{ni}(Y_{ij} - n_i\bar{Y}_i) = 0$$

lo que implica k ecuaciones de restricción y, por tanto, $n-k$ grados de libertad. A partir de los residuos definimos la varianza residual como el estimador insesgado de la varianza

$$S_R^2 = \sum_{i=1}^{k}\sum_{j=1}^{ni}\frac{e_{ij}^2}{n-k} = \sum_{i=1}^{k}\sum_{j=1}^{ni}\frac{(Y_{ij}-\bar{Y}_i)^2}{n-k}$$

La suma de las desviaciones cuadráticas de los valores observados respecto a la media total admite la siguiente descomposición:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{Y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2$$

Equivalente a

$$SCT = SCTR + SCE$$

siendo:

- SCT, Suma de cuadrados total, supone la variabilidad total existente en la observación, tiene asociados $n-1$ grados de libertad y se puede expresar como:

$$SCT = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2 = (n-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{n}$$

- $SCTR$, Suma de cuadrados entre tratamientos, es la variabilidad explicada por los tratamientos, tiene $k-1$ grados de libertad y puede expresarse como:

$$SCTR = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\alpha}_i^2 = \sum_{i=1}^{k}\frac{Y_i^2}{n_i} - \frac{Y_{\cdot\cdot}^2}{n}$$

- SCE, Suma de cuadrados de error, es la parte de la variabilidad no explicada por los tratamientos, tiene $n-k$ grados de libertad y se expresa como

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} e_{ij}^2 = SCT - SCTR$$

Las sumas de cuadrados anteriores divididas por sus grados de libertad permiten definir las medias cuadráticas, que son estimadores insesgados de σ^2 bajo la hipótesis nula H_0.

$$MCTR = \frac{SCTR}{k-1}$$

$$MCE = \frac{SCE}{n-k}$$

con distribuciones:

$$\frac{(k-1)MCTR}{\sigma^2} \sim \chi^2_{k-1}$$

$$\frac{(n-k)MCE}{\sigma^2} \sim \chi^2_{n-k}$$

Por lo que se puede definir el estadístico

$$F = \frac{MCTR}{MCE} \sim F_{k-1,n-k}$$

rechazando la hipótesis nula si $F > F_{k-1,n-k,1-\alpha}$.

Toda la información anterior se puede mostrar de forma resumida en la siguiente tabla ANOVA:

Fuente de variación	*S.C.*	*g.l.*	*M.C.*	Estadístico
Tratamientos	$SCTR$	$k-1$	$MCTR$	F
Error	SCE	$n-k$	MCE	
Total	SCT	$n-1$		

1.2. Modelo de Regresión Lineal Simple

El análisis de regresión estudia las posibles relaciones de dependencia entre variables, con la finalidad de obtener modelos que expliquen diferentes fenómenos.También se pretende estimar y predecir valores de interés para el investigador. El modelo de regresión lineal simple es:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, \ldots, n$$

donde el error aleatorio ϵ_i se modeliza mediante $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ y verificando que los errores son incorrelados, es decir, $Cov(e_i, e_j) = 0$.

Para estimar los parámetro definimos los siguientes estadísticos muestrales:

- $S_{xy} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(X_i - \bar{Y})}{n} = \sum_{i=1}^{n} \frac{X_i Y_i}{n} - \bar{X}\bar{Y}$
- $S_{xx}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} = \sum_{i=1}^{n} \frac{X_i^2}{n} - \bar{X}^2$
- $S_{yy}^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n} = \sum_{i=1}^{n} \frac{Y_i^2}{n} - \bar{Y}^2$

Aplicando el método de mínimos cuadrados, los estimadores de los parámetros β_0 y β_1 han de minimizar la función

$$R = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Obteniendose los estimadores mínimo cuadráticos con las siguientes distribuciones:

$$\begin{aligned} \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}^2} \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{n S_{xx}^2}\right) \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} = \bar{Y} - \frac{S_{xy}}{S_{xx}^2} \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2}{n}\left(1 + \frac{\bar{X}^2}{S_{xx}^2}\right)\right) \end{aligned}$$

Aplicando el método de máxima verosimilitud obtenemos los mismos estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ que con el método de mínimos cuadrados y el estimador de σ^2 es:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$

donde $e_i = Y_i - \hat{Y}_i$ es el residuo del modelo de regresión. Estos residuos verifican:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} e_i \hat{Y}_i = \sum_{i=1}^{n} e_i X_i = 0$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

Por lo que los grados de libertad asociados a los errores han de ser $n-2$ siendo el estimador insesgado de σ^2 la varianza residual

$$S_R^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

Con el objeto de ver la significación del modelo de regresión, planteado el contraste

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

es posible aplicar la metodología del ANOVA al análisis de regresión, para la resolución del contraste, mediante la siguiente descomposición de la variabilidad

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Equivalente a

$$SCT = SCR + SCE$$

siendo:

- SCT, Suma de cuadrados total, representa la variabilidad total existente en las observaciones, tiene $n-1$ grados de liberta y se puede expresar como

$$SCT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = nS_{yy}^2$$

- SCR, Suma de cuadrados de regresión, es la variabilidad explicada por la regresión, tiene un grado de libertad y se expresa como

$$SCR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = n\hat{\beta}_1 S_{xy} = n\hat{\beta}_1^2 S_{xx}$$

- SCE, Suma de cuadrados de error, es la variabilidad no explicada por la regresión, tiene $n-2$ grados de libertad y se expresa como

$$SCE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 = nS_{yy}^2 - n\hat{\beta}_1 S_{xy}$$

Las sumas de cuadrados anteriores divididas por sus grados de libertad permiten definir las medias cuadráticas, de igual modo que en el procedimiento del ANOVA:

$$MCR = \frac{SCR}{1}$$

$$MCE = \frac{SCE}{n-2}$$

Con distribuciones

$$\frac{MCR}{\sigma^2} \sim \chi^2_1$$

$$\frac{(n-2)MCE}{\sigma^2} \sim \chi^2_{n-2}$$

Por lo que se puede definir el estadístico

$$F = \frac{MCR}{MCE} \sim F_{1,n-2}$$

rechazando la hipótesis nula si $F > F_{1,n-2,1-\alpha}$.

La resolución del contraste se muestra en la siguiente tabla ANOVA:

Fuente de variación	*S.C.*	*g.l.*	*M.C.*	Estadístico
Regresión	SCR	1	MCR	F
Error	SCE	$n-2$	MCE	
Total	SCT	$n-1$		

Capítulo 2

Diseño unifactorial con covariable

2.1. El modelo lineal

El diseño unifactorial con covariable corresponde a un diseño unifactorial donde, junto a cada observación Y_{ij} de la variable respuesta, se dispone de la información adicional aportada por la variable X_{ij}, relacionada linealmente con ella. Esta variable X_{ij} de información concomitante recibe el nombre de covariable. El modelo lineal del diseño unifactorial con covariable, relaciona a la variable respuesta con un factor de naturaleza cualitativa y una covariable de naturaleza cuantitativa y puede expresarse como:

$$Y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \bar{X}) + \epsilon_{ij} \qquad i = 1, \quad k; \quad j = 1, ..., p,$$

donde Y_{ij} ha sido observada de forma simultanea con X_{ij}, ambas están relacionadas entre si de forma lineal, α_i es el efecto debido al tratamiento i-ésimo del factor en estudio, $\bar{X}$ es la media de la información adicional apuntada por las observaciones de la covaribale

$$\bar{X} = \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{X_{ij}}{kp}$$

verificando las siguientes propiedades:

$$\sum_{i=1}^{k} \alpha_i = 0 \qquad \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}) = 0.$$

Además β es el coeficiente de regresión lineal entre Y_{ij} y X_{ij} y ϵ_{ij} es el habitual error aleatorio de un modelo lineal del diseño experimenta, modelizado mediante una distribución normal con esperanza nula y varianza $\sigma^2, \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. El modelo anterior puede ser reescrito, eliminando de la observación Y_{ij} la información aportada por la covarianza como

$$Z_{ij} = Y_{ij} - \beta(X_{ij} - X) = \mu + \alpha_i + \epsilon_{ij}$$

con $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. En este caso, el modelo lineal corresponde al modelo lineal de un diseño unifactorial completamente aleatorizado.
El objetivo de un diseño unifactorial con covariable es estudiar la influencia del factor sobre la variable respuesta mediante la resolución del contraste:

$$\begin{cases} H_0 : \alpha_i = 0 \quad \forall i = 1, \ldots, k \\ H_1 : \text{algún } \alpha_i \neq 0 \end{cases}$$

Utilizando la notación habitual del diseño de experimentos referida a los estadísticos muestrales se introducen las sumas y medias por tratamientos para las variable X e Y con la siguientes expresiones

$$Y_{i.} = \sum_{j=1}^{p} Y_{ij} \qquad \bar{Y}_i = \frac{Y_{i.}}{p} = \frac{1}{p}\sum_{j=1}^{p} Y_{ij}$$

$$X_{i.} = \sum_{j=1}^{p} X_{ij} \qquad \bar{X}_i = \frac{X_{i.}}{p} = \frac{1}{p}\sum_{j=1}^{p} X_{ij}$$

y de forma global tenemos los estadísticos conocidos $Y_{..}$, $\bar{Y}$, $X_{..}$ y $\bar{X}..$ en función de estos estadísticos e l modelo lineal se puede expresar como:

$$\bar{Y}_i = \mu + \alpha_i + \beta(\bar{X}_i - \bar{X})$$

Por último, es de interés mostrar la expresión de diferencia entre los efectos debidos a dos tratamientos cualesquier α_i y α_j expresada como:

$$\alpha_i - \alpha_j = \bar{Y}_i - \bar{Y}_j - \beta(\bar{X}_i - \bar{X}) = (\bar{Y}_i - \beta\bar{X}_i) - (\bar{Y}_j - \beta\bar{X}_j) = Y_{\bar{X}_i} - Y_{\bar{X}_j}$$

donde $Y_{\bar{X}}$ es el valor del ajuste de regresión para $X = \bar{X}$.

2.2. Estimación de parámetros

Siguiendo la metodología de la resolución del diseño experimental, para la estimación de parámetros es preciso la definición de los siguientes estadísticos muestrales:

- $SS_{xx} = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X})^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} X_{ij}^2 - \frac{X_{..}^2}{kp} = nS_{xx}^2$
- $SS_{yy} = \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} Y_{ij}^2 - \frac{Y_{..}^2}{kp} = nS_{yy}^2$
- $SS_{xy} = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X})(Y_{ij} - \bar{Y}) = \sum_{i=1}^{k}\sum_{j=1}^{p} X_{ij}Y_{ij}\frac{X_{..}Y_{..}}{kp} = nS_{xy}^2$
- $T_{xx} = \sum_{i=1}^{k}(X_{i.} - \bar{X})^2 = \sum_{i=1}^{k}\frac{X_{i.}^2}{p} - \frac{X_{..}^2}{kp}$
- $T_{yy} = \sum_{i=1}^{k}(Y_{i.} - \bar{Y})^2 = \sum_{i=1}^{k}\frac{Y_{i.}^2}{p} - \frac{Y_{..}^2}{kp}$
- $T_{xy} = \sum_{i=1}^{k}(X_{i.} - \bar{X})(Y_{i.} - \bar{y})\sum_{i=1}^{k}\frac{X_{i.}Y_{i.}}{p} - \frac{X_{..}Y_{..}}{kp}$

- $E_{xx} = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X}_i)^2 = SS_{xx} - T_{xx}$
- $E_{yy} = \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \bar{Y}_i)^2 = SS_{yy} - T_{yy}$
- $E_{xy} = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i) = SS_{xy} - T_{xy}$

Los estadísticos muestrales anteriores, asociados a las correspondientes covarianzas entre Y y X, pueden expresares de forma análoga a una tabla ANOVA, teniendo en cuenta que las expresiones con subíndices iguales corresponden a sumas de cuadrados mientras que los subíndices distintos corresponden a los factores cruzados, y por tanto pueden ser positivos o negativos. En la tabla de tipo ANOVA, se verifica la suma en la diferentes columnas.

Fuente	X	XY	Y
Tratamientos	T_{xx}	T_{xy}	T_{yy}
Error	E_{xx}	E_{xy}	E_{yy}
Total	SS_{xx}	SS_{xy}	SS_{yy}

La estimación de los parámetros mediante el método de mínimos cuadrados minimiza la suma de los cuadrados de los residuos, equivalente a minimizar la expresión

$$L = \sum_{i=1}^{k}\sum_{j=1}^{p} e_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \mu_i - \beta(X_{ij} - \bar{X}))^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X}))^2$$

Las ecuaciones normales que se obtienen son

$$\begin{cases} \dfrac{\partial L}{\partial \mu} = 0 \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \\ \dfrac{\partial L}{\partial \beta} = 0 \end{cases}$$

Desarrollando la primera ecuación normal obtenemos:

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= \frac{\partial \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X}))^2}{\partial \mu} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{p} -2(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X})) = 0 \end{aligned}$$

$$\iff -\sum_{i=1}^{k}\sum_{j=1}^{p} Y_{ij} \; + \; n\mu + \; \sum_{i=1}^{k} p\,\alpha_i \; + \; \sum_{i=1}^{k}\sum_{j=1}^{p} \beta(X_{ij} - \bar{X}) = 0$$
$$\iff n\hat{\mu} = \; Y_{\cdot\cdot} \iff \hat{\mu} = \bar{Y}$$

Si desarrollamos la segunda ecuación obtenemos:

$$\begin{aligned}
\frac{\partial L}{\partial \alpha_i} &= \frac{\partial \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X}))^2}{\partial \alpha_i} \\
&= \sum_{j=1}^{p} -2(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X})) = 0 \\
&\iff -\sum_{j=1}^{p} Y_{ij} + p\hat{\mu} + p\hat{\alpha}_i + \hat{\beta}\sum_{j=1}^{p}(X_{ij} - \bar{X}) = 0 \\
&\iff p\hat{\mu} + p\hat{\alpha}_i + \hat{\beta}\sum_{j=1}^{p}(X_{ij} - \bar{X}) \; = Y_{i\cdot} \quad i = 1, 2..., k
\end{aligned}$$

Desarrollando la tercera ecuación obtenemos:

$$\begin{aligned}
\frac{\partial L}{\partial \beta} &= \frac{\partial \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X}))^2}{\partial \beta} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{p} -2(X_{ij} - \bar{X})(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X})) = 0 \\
&\iff -\sum_{i=1}^{k}\sum_{j=1}^{p} Y_{ij}(X_{ij} - \bar{X}) + \sum_{i=1}^{k}\sum_{j=1}^{p} \hat{\mu}(X_{ij} - \bar{X}) + \\
&\quad + \sum_{i=1}^{k}\sum_{j=1}^{p} \hat{\alpha}_i(X_{ij} - \bar{X}) + \sum_{i=1}^{k}\sum_{j=1}^{p} \hat{\beta}(X_{ij} - \bar{X})^2 = 0 \\
&\iff \sum_{i=1}^{k}\sum_{j=1}^{p}(-Y_{ij}(X_{ij} - \bar{X}) + \hat{\mu}(X_{ij} - \bar{X})) + \sum_{i=1}^{k}\sum_{j=1}^{p} \hat{\alpha}_i(X_{ij} - \bar{X}) + \\
&\quad + \sum_{i=1}^{k}\sum_{j=1}^{p} \hat{\beta}(X_{ij} - \bar{X})^2 = 0 \\
&\iff \sum_{i=1}^{k}\sum_{j=1}^{p}(\bar{Y} - Y_{ij})(X_{ij} - \bar{X}) + \sum_{i=1}^{k} \hat{\alpha}_i \sum_{j=1}^{p}(X_{ij} - \bar{X}) + \hat{\beta}\sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X})^2 = 0 \\
&\iff \sum_{i=1}^{k} \hat{\alpha}_i \sum_{j=1}^{p}(X_{ij} - \bar{X}) + \hat{\beta}\sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X})^2 - \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij} - \bar{Y})(X_{ij} - \bar{X}) = 0 \\
&\iff \sum_{i=1}^{k} \hat{\alpha}_i \sum_{j=1}^{p}(X_{ij} - \bar{X}) + \hat{\beta} SS_{xx} - SS_{xy} = 0
\end{aligned}$$

En resumen, las tres ecuaciones normales pueden expresarse como:

$$\begin{cases} \hat{\mu} = \bar{Y} \\ \\ Y_{i.} = p\hat{\mu} + p\hat{\alpha}_i \ + \hat{\beta} \ \sum_{j=1}^{p}(X_{ij} - \bar{X}) \quad i = 1, 2..., k \\ \\ SS_{xy} = \sum_{i=1}^{k} \hat{\alpha}_i \sum_{j=1}^{p}(X_{ij} - \bar{X}) + \hat{\beta} SS_{xx} \end{cases}$$

Estimamos los parámetros a partir de las ecuaciones normales obtenidas.
A partir de la segunda ecuación normal se obtiene el estimador del efecto debido:

$$p\ \hat{\alpha}_i = \ Y_{i.} - p\ \bar{Y} - \beta\ \sum_{j=1}^{p}(X_{ij} - \bar{X}) \Longleftrightarrow \hat{\alpha}_i = \ \bar{Y}_i - \ \bar{Y} - \hat{\beta}\ (\bar{X}_i - \bar{X}).$$

Sustituyendo el valor de $\hat{\alpha}_i$ en la tercera ecuación se obtiene

$$\sum_{i=1}^{k}\ (\bar{Y}_i - \ \bar{Y}) \sum_{j=1}^{p}(X_{ij} - \bar{X}) - \hat{\beta}\ \sum_{i=1}^{k}\ (\bar{X}_i - \bar{X}) \sum_{j=1}^{p}(X_{ij} - \bar{X}) + \hat{\beta} SS_{xx} = SS_{xy}$$

Ahora bien, teniendo en cuenta que :

$$\sum_{i=1}^{k}\ (\bar{Y}_i - \ \bar{Y}) \sum_{j=1}^{p}(X_{ij} - \bar{X}) = \sum_{i=1}^{k}(Y_i - \bar{Y})(X_i - \bar{X}) = T_{xy}$$

Y que:

$$\sum_{i=1}^{k}\ (\bar{X}_i - \bar{X}) \sum_{j=1}^{p}(X_{ij} - \bar{X}) = \sum_{i=1}^{k}(X_i - \bar{X})(X_i - \bar{X}) = T_{xx}$$

tenemos, sustituyendo, que

$$T_{xy} - \hat{\beta} T_{xx} + \hat{\beta} SS_{xx} = SS_{xy}$$

Y despejando llegamos a su estimación:

$$\hat{\beta} = \frac{SS_{xy} - T_{xy}}{SS_{xx} - T_{xx}} = \frac{E_{xy}}{E_{xx}}$$

Para la estimación de σ^2 es preciso aplicar el método de máxima verosimilitud puesto que corresponde a la varianza del error. Para los parámetros restantes, su estimador máximo verosímil coincide con el estimador mínimo cuadrático.

Bajo la suposición de normalidad hecha para los errores aleatorios, la estimación de máximo verosimilitud de σ^2 consiste en maximizar la función de verosimilitud cuya expresión es:

$$\begin{aligned} L \ &= \ \prod_{i=1}^{k}\prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e_{ij}}{2\sigma^2}\right) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^{k}\sum_{j=1}^{p}\frac{e_{ij}}{2\sigma^2}\right) \\ &= \ (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^{k}\sum_{j=1}^{p}\frac{(Y_{ij} - \mu - \alpha_i - \beta(X_{ij} - \bar{X}))^2}{2\sigma^2}\right) \end{aligned}$$

Equivalente a maximizar

$$\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{(Y_{ij}-\mu-\alpha_i-\beta(X_{ij}-\bar{X})^2}{2\sigma^2}$$

Maximizando $\log L$

$$\begin{aligned}\frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{(Y_{ij}-\mu-\alpha_i-\beta(X_{ij}-\bar{X})^2}{2(\sigma^2)^2} = 0\\ &\iff -\frac{n}{2} + \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{(Y_{ij}-\mu-\alpha_i-\beta(X_{ij}-\bar{X}))^2}{2\sigma^2} = 0\\ &\iff \sum_{j=1}^{p}\frac{(Y_{ij}-\mu-\alpha_i-\beta(X_{ij}-\bar{X}))^2}{\sigma^2} = n\end{aligned}$$

Despejando, el estimador máximo verosímil de σ^2 es:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}(X_{ij}-\bar{X}))^2}{n}\\ &= \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij}-\bar{Y}-\bar{Y}_i+\bar{Y}+\hat{\beta}(\bar{X}_i-\bar{X})-\hat{\beta}(X_{ij}-\bar{X})\right)^2\\ &= \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij}-\bar{Y}_i-\bar{Y}_i+\hat{\beta}(\bar{X}_i-\bar{X}-X_{ij}+\bar{X})\right)^2\\ &= \sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij}-\bar{Y}_i)^2+\hat{\beta}^2\sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij}-\bar{X}_i)^2-2\sum_{i=1}^{k}\sum_{j=1}^{p}(Y_{ij}-\bar{Y}_i)\hat{\beta}(X_{ij}-\bar{X}_i)\\ &= E_{yy}+\frac{E_{xy}^2E_{xx}}{E_{xx}^2}-2\frac{E_{xy}E_{xy}}{E_{xx}} = E_{yy}+\frac{E_{xy}^2}{E_{xx}}-2\frac{E_{xy}^2}{E_{xx}} = E_{yy}-\frac{E_{xy}^2}{E_{xx}}\end{aligned}$$

Equivalente a $\sigma^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{e_{ij}^2}{n}$

Los estimadores obtenidos para los parámetros $\hat{\mu}$, $\hat{\alpha}_i$ y $\hat{\beta}$ son insesgados. Así la esperanza de μ es

$$\begin{aligned}E[\hat{\mu}] &= \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{1}{kp}E[Y_{ij}] = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{1}{kp}(\mu+\alpha_i+\beta(X_{ij}-\bar{X})\\ &= \mu+\sum_{i=1}^{k}(p\alpha_i)+\beta\sum_{i=1}^{k}\sum_{j=1}^{k}(X_{ij}-\bar{X}) = \mu\end{aligned}$$

Para el estimador $\hat{\beta}$ su esperanza es:

$$E[\hat{\beta}] = E\left[\frac{E_{xy}}{E_{xx}}\right] = E\left[\frac{\sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij}-\bar{X}_i)(Y_{ij}-\bar{Y}_i)}{E_{xx}}\right]$$

$$
\begin{aligned}
&= \frac{1}{E_{xx}} \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}_i)(E[Y_{ij}] - E[\bar{Y}_i]) \\
&= \frac{1}{E_{xx}} \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}_i)(\mu + \alpha_i + \beta(X_{ij} - \bar{X}) - \mu - \alpha_i - \beta(\bar{X}_i - \bar{X})) \\
&= \beta \frac{1}{E_{xx}} \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i) = \beta \frac{E_{xx}}{E_{xx}} = \beta
\end{aligned}
$$

Equivalentemente para $\hat{\alpha}_i$, utilizando que $E[\hat{\beta}] = \beta$, tenemos:

$$
\begin{aligned}
E[\hat{\alpha}_i] &= E[\bar{Y}_{i.} - \bar{Y} - \hat{\beta}(\bar{X}_{i.} - \bar{X})] = E[\sum_{j=1}^{p} \frac{Y_{ij}}{p} - \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{Y_{ij}}{n} - \hat{\beta}(\bar{X}_{i.} - \bar{X})] \\
&= \frac{1}{p} \sum_{j=1}^{p} E[Y_{ij}] - \mu - \beta(\bar{X}_{i.} - \bar{X}) \\
&= \frac{1}{p} \sum_{j=1}^{p} E\left[\mu + \alpha_i + \beta(X_{ij} - \bar{X})\right] - \mu - \beta(\bar{X}_i - \bar{X}) \\
&= \frac{1}{p} \left(p\mu + p\alpha_i + \beta(\bar{X}_i - p\bar{X})\right) - \mu - \beta(\bar{X}_i - \bar{X}) \\
&= \mu + \alpha_i + \beta(\bar{X}_i - \bar{X}) - \mu - \beta(\bar{X}_{i.} - \bar{X}) = \alpha_i
\end{aligned}
$$

Sin embargo el estimador máximo verosímil $\hat{\sigma}^2$ no es insesgado puesto que bajo normalidad $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$ por lo que se precisa definir la varianza residual como el estimador insesgado de σ^2. Teniendo en cuenta que los grados de libertad del error son $k(p-1)-1$,correspondientes a los $(p-1)$ grados de libertad de cada muestra por k muestras y restando una unidad por el regresor correspondiente a la covariable, definimos la varianza residual como:

$$S_R^2 = \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{e_{ij}}{k(p-1)-1}$$

2.3. Descomposición de la variabilidad

Para aplicar la metodología del diseño experimental es necesario la descomposición de la variabilidad existente. Esta descomposición es

$$SCT = SCR + SCTR + SCE$$

La variabilidad total es

$$SCT = \sum_{i=1}^{k} \sum_{j=1}^{p} (Y_{ij} - \bar{Y})^2 = nS_{yy}^2 = SS_{yy}$$

y tiene $kp-1$ grados de libertad.
Si tenemos en cuenta la variabilidad explicada por la covariable esta corresponde a la regresion explicada del modelo de regresion :

$$SCR = kp\hat{\beta}^2 S_{xx}$$

Por tanto la variabilidad restante a explicar mediante los tratamientos el modelo unifactorial con covariable, podría expresarse como:

$$SCT - SCR = SCTR + SCE$$

Donde SCE es la suma de los residuos al cuadrado calculados como

$$SCE = E_{yy} - \frac{E_{xy}^2}{E_{xx}}$$

Pudiendo obtener la variabilidad debido a los tratamientos como

$$SCTR = SCT - SCR - SCE$$

Notar que SCR es la correspondiente al modelo de regresión lineal simple con Y como variable respuesta y X como regresor, mientras que $SCTR$ no coincide con la suma de cuadrados de tratamientos del diseño unifactorial ya que está influida por la variabilidad eliminada por la covariable, SCR, por lo que se dice corregida por la regresión o por la covariable.
Las sumas de cuadrado anteriores divididas por sus grados de libertad permiten definir las medias cuadráticas, de igual modo que en el procedimiento del ANOVA:

$$MCR = \frac{SCR}{1}$$
$$MCTR = \frac{SCTR}{k-1}$$
$$MCE = \frac{SCE}{k(p-1)-1}$$

Estas medias cuadráticas bajo normalidad cumplen que:

$$\frac{(n-k-1)MCE}{\sigma^2} \sim \chi^2_{n-k-1}$$
$$\frac{(k-1)MCTR}{\sigma^2} \sim \chi^2_{k-1}$$

2.4. Resolución de los contrastes

Los contrastes a resolver con este modelo son:

- Para comprobar si el factor es significativo, la hipotesisi nula es

  $$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0$$

  Para resolver este contraste, mediante la descomposición de la variabilidad anterior, definimos el estadístico F como

  $$F = \frac{MCTR}{MCE} \sim F_{k-1,kp-k-1}$$

  rechazando la hipótesis nula si $F > F_{k-1,kp-k-1,1-\alpha}$ con un nivel de confianza $1-\alpha$

- Para comprobar la linealidad exigida para la covariable mediante

$$H_0 : \beta = 0$$

el estadístico de contraste, aplicando la descomposición de la variabilidad obtenida, es

$$F_R = \frac{MCR}{MCE} \sim F_{1,k(p-1)-1}$$

rechazando la hipótesis nula si $F_R > F_{1,k(p-1)-1,1-\alpha}$ con un nivel de confianza $1 - \alpha$

Los resultados pueden reunirse en la siguiente tabla ANOVA

Fuente variación	S.C.	g.l.	M.C.	Estadístico
Regresión	SCR	1	MCR	F_R
Tratamientos	$SCTR$	$k-1$	$MCTR$	F
Error	SCE	$k(p-1)-1$	MCE	
Total	SCT	$kp-1$		

Capítulo 3

Ejemplo

Consideramos el estudio realizado para determinar si existe una diferencia en la resistencia de una fibra monofilamento producida por tres maquinas diferentes. Cabe la posibilidad que las muestras elegidas no sean homogéneas, en particular que el grosor no sea análogo y repercuta en la resistencia. Se decide medir el grosor e intentar incorporarlo como una covariable y se muestran en la siguiente tabla.

Máquina 1		Máquina 2		Máquina 3	
Resistencia	Grosor	Resistencia	Grosor	Resistencia	Grosor
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15

La resistencia de la fibra también puede verse afectada por su grosor, por tanto el efecto del grosor puede ser considerada como covariable, si tiene relación lineal ya que ha sido medida de forma simultanea con la resistencia.

Si realizamos un ANOVA clásico para diagnosticar la influencia de la máquina en la resistencia se obtiene la siguiente tabla

Fuente Variación	*S.C.*	*g.l.*	*M.C.*	F
Tratamiento	140	2	70,2	4,09
Error	206,0	12	17,16	
Total	346,4	14		

con p-valor igual 0.0442. Como este p-valor es menor que 0.05, este factor tendrá un efecto significativo sobre la variable respuesta Resistencia al 95 % de confianza. Por tanto al hacer el ANOVA se rechaza la hipótesis nula, es decir, que la resistencia depende de la máquina con la que se ha produce el monofilamento con un 95 % de certeza.
Si se considera la información aportada por el grosor, el modelo corresponde a un diseño unifactorial con covariable cuya tabla ANOVA pasa a ser

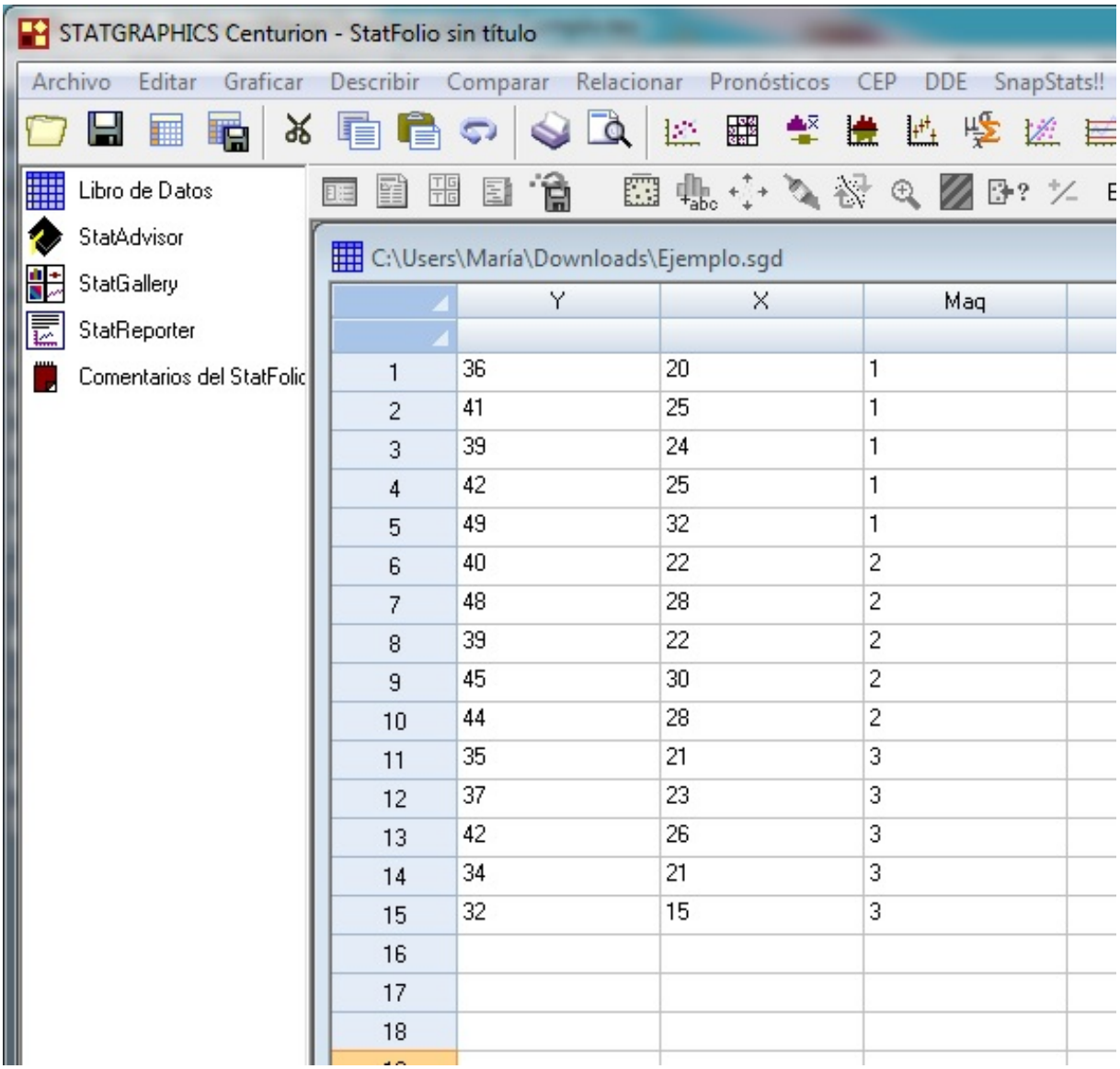
Fuente Variación	*S.C.*	*g.l.*	*M.C.*	Estadístico
Regresión	305,13	1	305,13	
Tratamientos	13,28	2	6,64	2,61
Error	27,98	11	2,54	
Total	346,4	14		

En este caso el p-valor de es 0.1181, que por ser mayor de 0.05, el factor Máquina no tendrá un efecto significativo sobre la variable respuesta Resistencia al 95 % de confianza. Se observa entonces como al hacer el análisis de la covarianza se acepta la hipótesis nula, es decir, la resistencia del monofilamento no depende de la máquina con la que se produce con un 95 % de certeza.
Por tanto, podemos concluir que las máquinas, según lo observado muestran una resistencia análoga y las diferencias mostradas por el diseño unifactorial inicial se debe a las diferencias de grosor de las muestras observadas. Sin embargo, en el segundo diseño, al eliminar la influencia del grosor por la introducción de la covariable el diseño muestra que las máquinas son análogas.
En la tabla ANOVA del diseño con covariable también se comprueba la linealidad exigida para la covariable en la fuente de variación asociada a la regresión. Este analisis es análogo al estudio de regresión de la Resistencia, variable respuesta, frente al grosor, covariable, como se muestra en la tabla ANOVA de un modelo de regresión lineal simple

Fuente Variación	*S.C.*	*g.l.*	*M.C.*	F
Regresion	305,13	1	305,13	46,12
Error	41,26	13	3,1746	
Total	346,4	14		

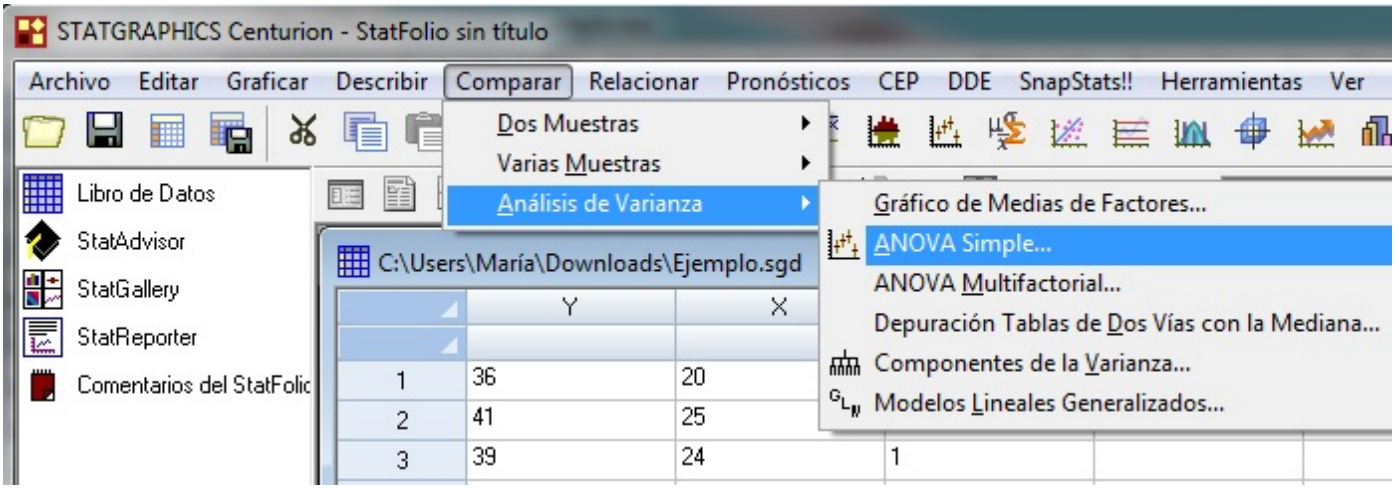
Puesto que el p-valor es 0.00 menor que 0.05, se acepta la condición de linealidad exigida a la covariable sobre la variable respuesta.

A continuación se muestra cual es el procedimiento con Statgraphics Centurion para realizar el ANOVA y el diseño unifactorial con covariable. Para ambos casos, lo primero es introducir los datos. Llamamos Y a la resistencia, X a la variable grosor y Maq al numero de máquina.

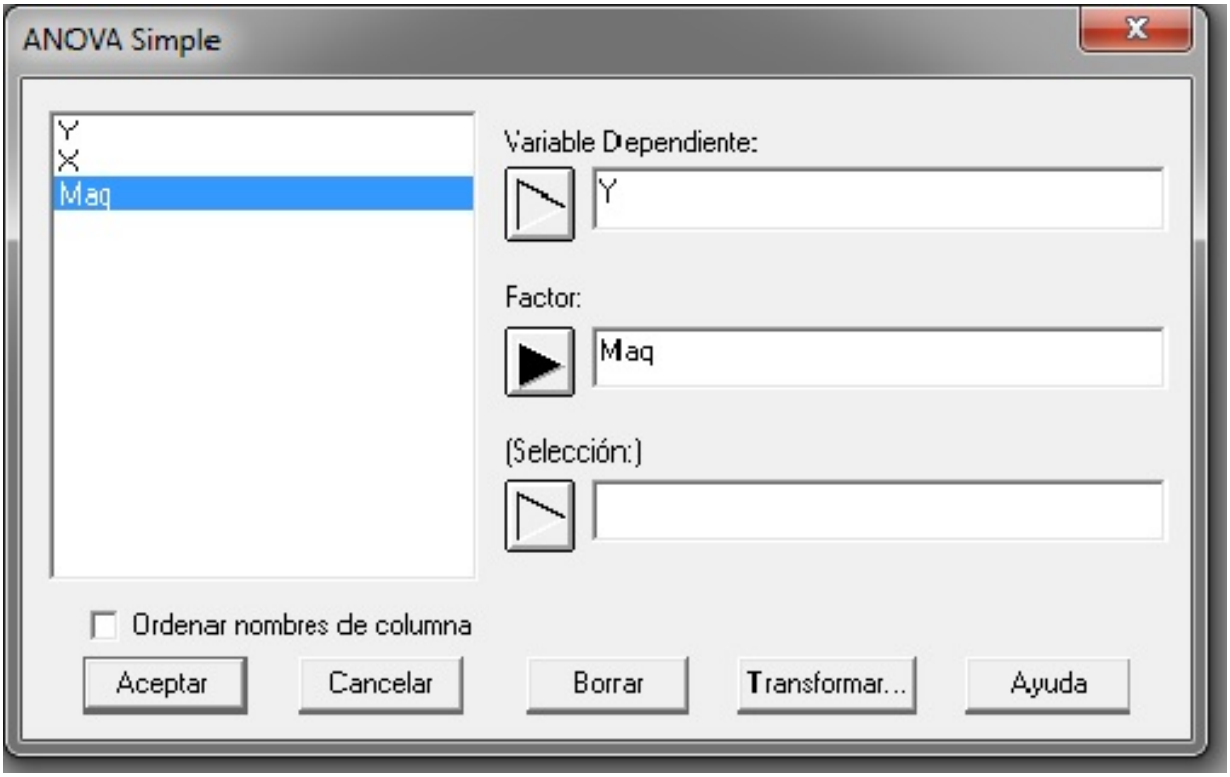


	Y	X	Maq
1	36	20	1
2	41	25	1
3	39	24	1
4	42	25	1
5	49	32	1
6	40	22	2
7	48	28	2
8	39	22	2
9	45	30	2
10	44	28	2
11	35	21	3
12	37	23	3
13	42	26	3
14	34	21	3
15	32	15	3
16			
17			
18			

A continuación, para hacer el ANOVA, se selecciona en la barra de herramientas la pestaña Comparar. Al picar sobre la pestaña comparar tenemos varias opciones, se ha de seleccionar la pestaña desplegable Análisis de Varianza donde se abrirán una serie de opciones de las cuales seleccionamos ANOVA Simple .

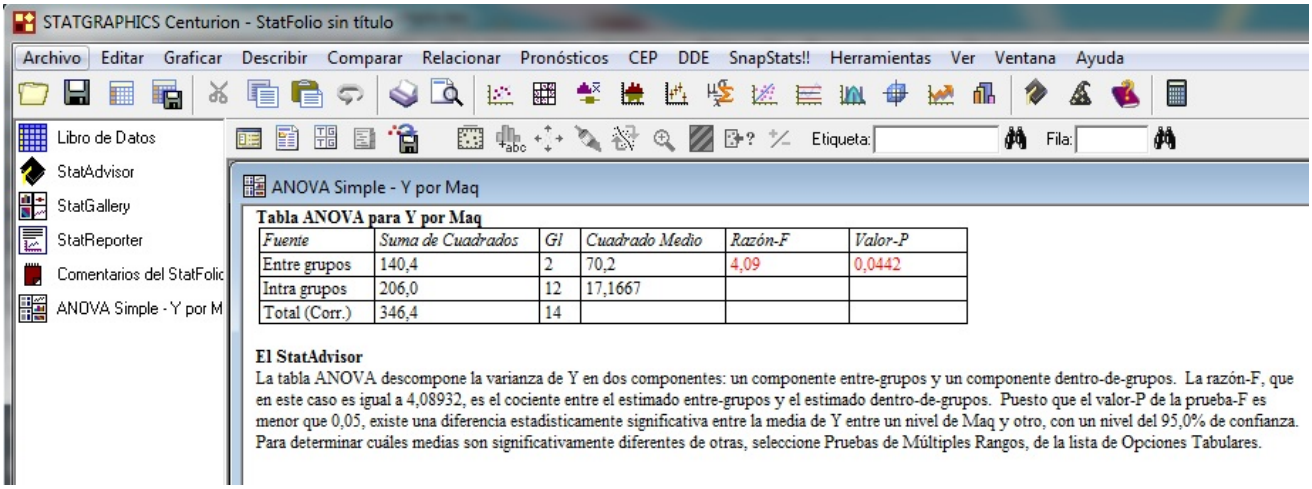


Al picar en ANOVA Simple, obtenemos un cuadro de diálogo llamado ANOVA Simple en cual aparecen todas las variables a la izquierda y a la derecha las siguientes casillas: Variable Dependiente y Factor. En este cuadro primero debemos introducir la variable dependiente, Y(Resistencia), y en segundo lugar el factor, Maq.

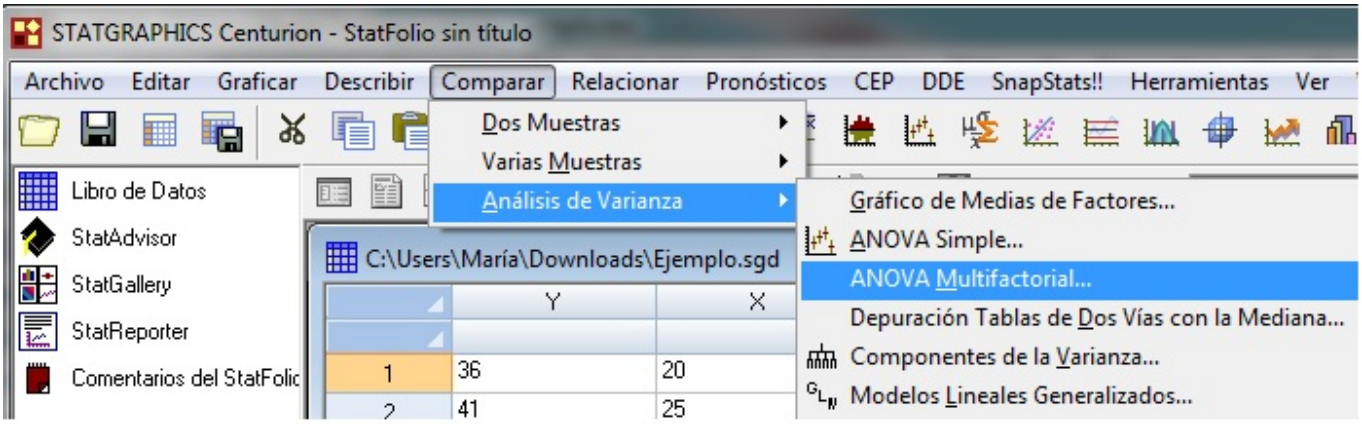


Una vez que picamos en el botón aceptar del diálogo se abre otro diálogo llamado Tablas y Gráficos en el que tenemos que seleccionar que queremos que nos muestre el programa por pantalla

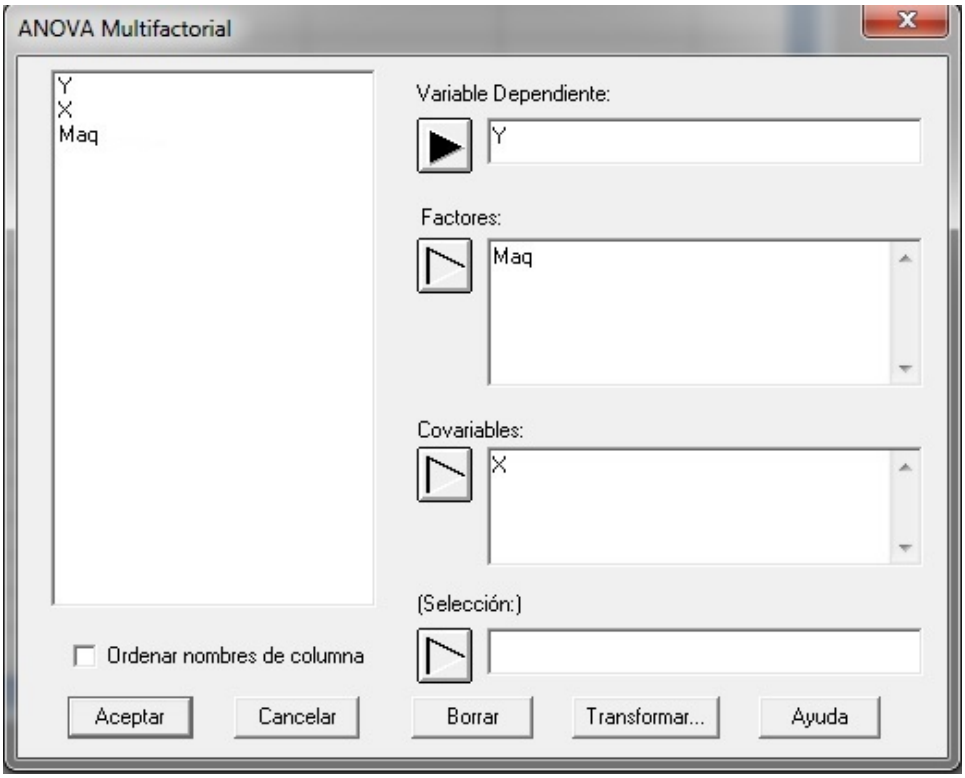
En este caso vamos a seleccionamos Tabla ANOVA, aunque si queremos información adicional podemos seleccionar lo que nos interese en cada caso.



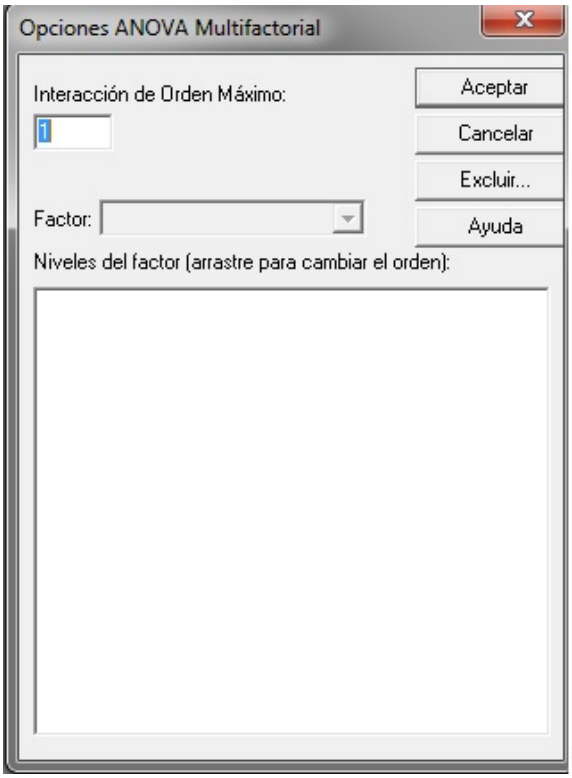
A continuación, vamos a realizar el Análisis de la covarianza. Para ello se selecciona en la barra de herramientas la pestaña comparar. Al picar sobre la pestaña comparar tenemos varias opciones, se ha de seleccionar la pestaña Analisis de Varianza. Una vez seleccionado esta pestaña volvemos a poder elegir entre varias, seleccionamos por ultimo ANOVA Multifactorial.



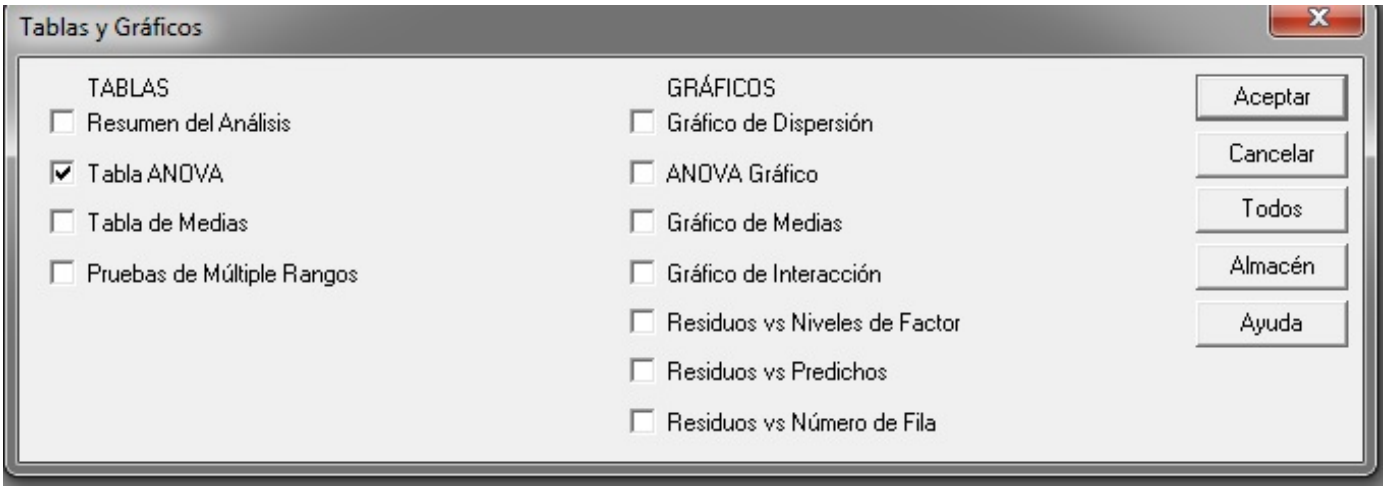
Al picar sobre el botón Anova Multifactorial se abre un dialogo llamado Anova Multifactorial similar al que se obtenía al picar en Anova Simple. A la izquierda aparecen las variables y a la derecha las siguientes casillas: Variable Dependiente, Fator y Covariable. Primero seleccionamos la variable dependiente, Y, después el factor, Maq, y por ultimo la covariable, X



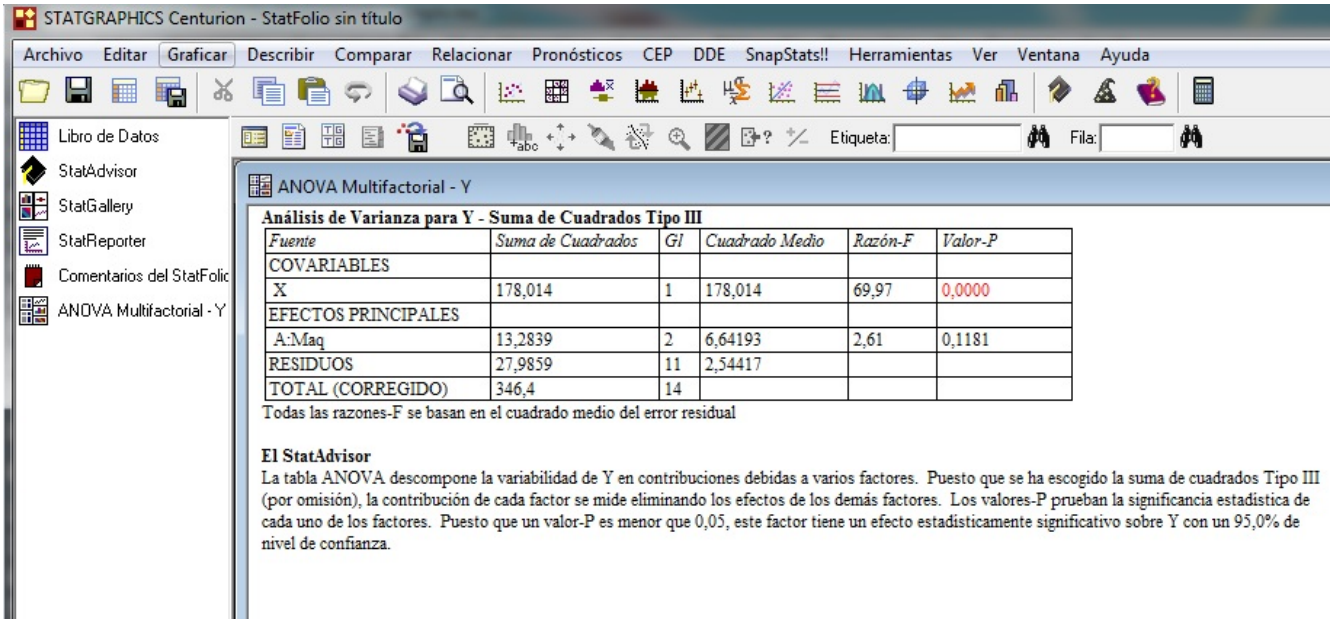
Al picar en aceptar en el cuadro de diálogo Anova Multifactorial, se abre otro dialogo llamado Opciones Anova Multifactorial, en el que dejamos las opciones por defecto y picamos en aceptar



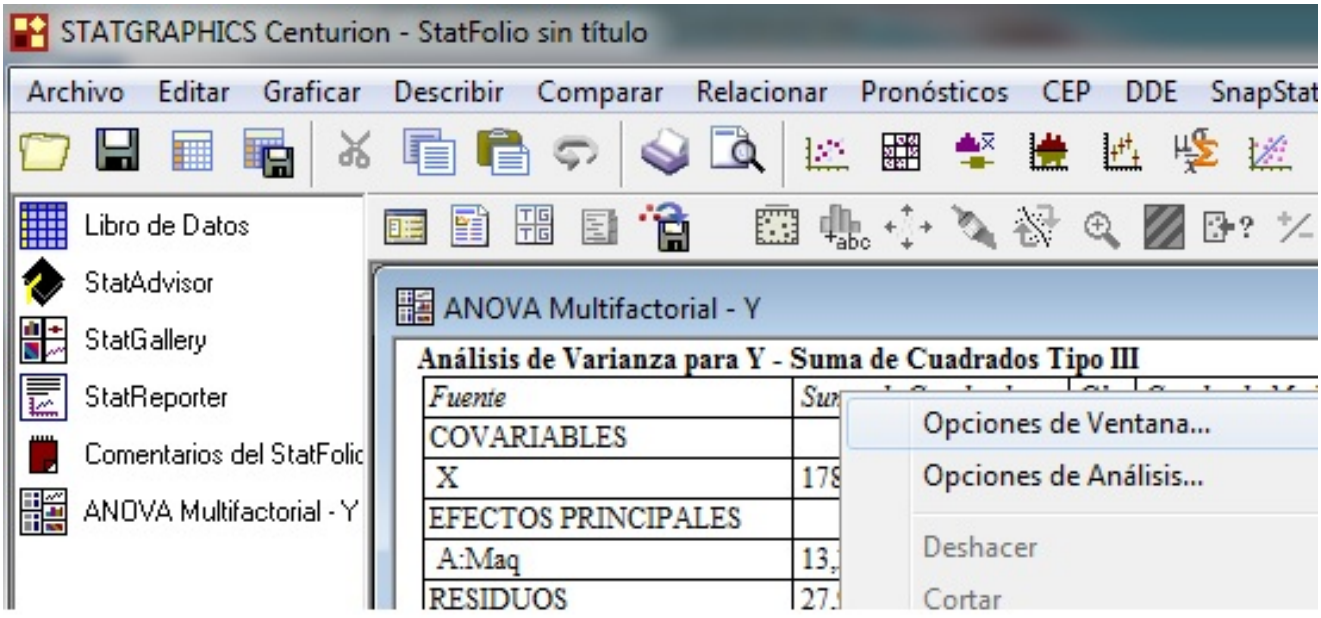
Por último se abre otro cuadro de dialogo, Tablas y Gráficos, donde seleccionamos lo que queremos que el programa nos muestre por pantalla. En este caso solo seleccionamos Tabla Anova pero si queremos información adicional podemos seleccionar cualquier otro botón.



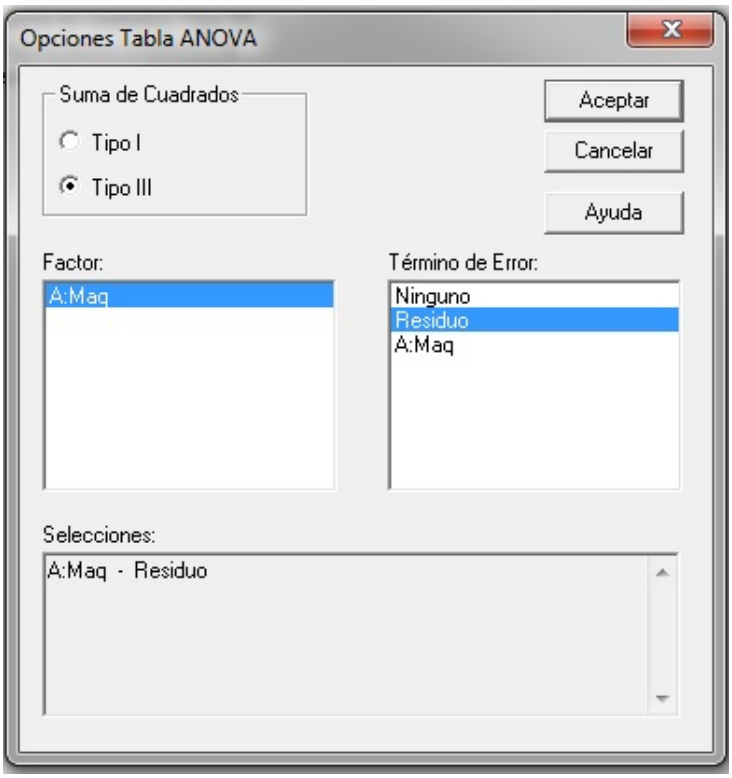
Al aceptar el cuadro de diálogo Tablas y Gráficos obtenemos



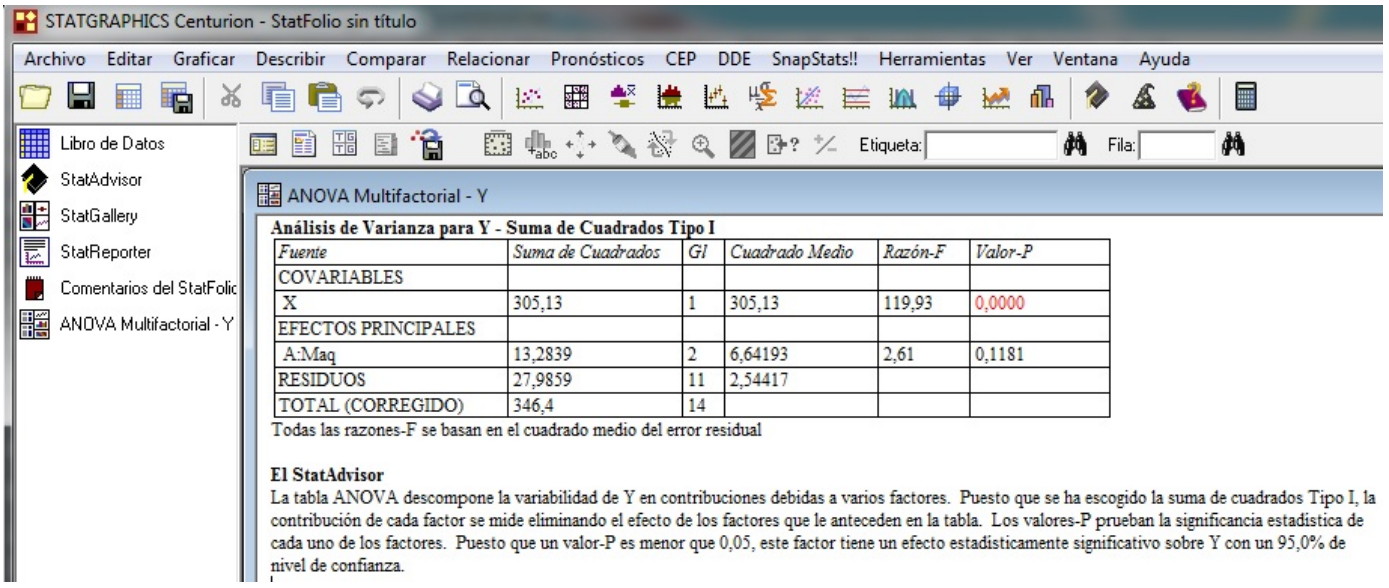
Por último picamos sobre la tabla Análisis de la varianza para Y - Suma de Cuadrados Tipo III con el boton derecho del ratón y seleccionamos la opción Opciones de ventana



Al picar en Opciones de Ventana se abre el siguiente dialogo



Donde seleccionamos Tipo I ya que en Statgraphics el TIPO I corrige las fuentes de variación según las anteriores del modelo, es decir, en este caso, corrige los tratamientos (Máquinas) según la covariable (grosor). Obteniendo finalmente la tabla ANOVA del diseño unifactorial con covariable para nuestro ejemplo



Bibliografía

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13, 261-281.*

Heyer, H. Theory of statistical experiments. *Spromher-Verlaq, 1982.*

Hicks, C. R. Fundamental Concepts in the Design of Experiments. *OUP USA, 1999.*

Hinkelmann, K. Kempthorne, 0. Design and analysis of experiments. *Willey, 1993.*

Kuehl, R. O. Diseño de experimentos : principios estadísticos de diseño y análisis de investigación. *Thomson Learning, 2001.*

Montgomery, D.C. Diseño y Análisis de Experimentos. *Limusa-Wiley, 2005.*

Peña, D. Regresión y diseño de experimentos. *Alianza Editorial, 2010.*

Yandell, B.S. Practical Data Analysis For Designed Experiments. *Chapman & Hall, 1997.*