

PRÁCTICA 7. ANÁLISIS DE LA VARIANZA CON UN FACTOR A DOS NIVELES. COMPARACIÓN DE DOS POBLACIONES NORMALES

Objetivo

El objeto de la presente sesión de práctica informática es complementar y afianzar los conceptos relativos a las técnicas de inferencia relativos a la comparación de dos poblaciones normales vistos en clase (apartados 2 de la UD 5.2 y primeros apartados de la UD5.3 ANOVA con un factor a dos niveles). Se pretende que el alumno se familiarice con las opciones que el programa Statgraphics ofrece al respecto.

1. Ejemplo

Un programador desea comparar dos sistemas de depuración automática de programas. Para ello, registra el tiempo que tarda el sistema 1 con 10 programas seleccionados al azar:

Sistema 1: 63, 61, 57, 61, 58, 60, 64, 60, 59 y 58

En otra muestra de 10 programas registra el tiempo que tarda el sistema 2:

Sistema 2: 62, 59, 61, 60, 58, 63, 61, 58, 63 y 60

¿Podemos decir que hay diferencias significativas entre los tiempos medios de depuración proporcionados por uno y otro sistema?

¿Podemos aceptar que hay diferencias significativas entre las varianzas de los tiempos de depuración proporcionados por uno y otro sistema?

Para responder a este tipo de preguntas se deben utilizar las técnicas de *comparación de poblaciones*.

Introduce los datos en el editor de Statgraphics, creando dos columnas: TIEMPO con los 20 valores del tiempo y SISTEMA con los 20 códigos de sistema (10 unos y 10 doses).

Para comparar las dos medias se puede utilizar la opción del ANOVA de Statgraphics:

Compare > Analysis of Variance> Multifactor ANOVA... (También se puede seleccionar One-Way ANOVA). En el cuadro de diálogo:

Dependent Variable: TIEMPO

Factors: SISTEMA

La tabla de ANOVA resulta:

Analysis of Variance for TIEMPO

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:SISTEMA	0,8	1	0,8	0,19	0,6673
RESIDUAL	75,4	18	4,18889		
TOTAL	76,2	19			

Como $p\text{-value}=0,6673$ es mayor que 0,05, no hay diferencias significativas entre las medias para un riesgo de primera especie del 5% (o nivel de confianza 95%).

La tabla del ANOVA proporciona:

Suma de Cuadrados Total (SCTotal)=76,2 con grados de libertad totales = $n^{\circ}\text{total de datos}-1=20-1=19$

Suma de Cuadrados del factor sistema (SCSistema)=0,8 con grados de libertad= $n^{\circ}\text{ de niveles del factor}-1=2-1=1$


Suma de Cuadrados residual (SCResidual)=75,4 con grados de libertad residuales = $\text{grados de libertad totales}-\text{grados de libertad del factor}=19-1=18$

Los cuadrados medios son sumas de cuadrados/grados de libertad.

La F-ratio= $\text{CMSistema}/\text{CMResidual}=0,19$.

Si no se dispone del p-value para saber si hay diferencia significativa entre las dos medias de sistema se puede comparar la F-ratio con la F de tablas (Tabla F de 1 cola, página 1). La F tablas se busca con 1 grado de libertad para el numerador y 18 grados de libertad para el denominador, y para $\alpha=0,05$ resulta 4,41. Como Fratio es menor que F tablas no hay diferencia significativa entre las dos medias para $\alpha=5\%$.

Para comparar las dos varianzas se utiliza la opción: **Compare > Two-Samples > Two-Sample Comparison**. En el cuadro de diálogo cambiar el campo Input a la opción Data and Code columns. En Data: TIEMPO y en Sample Code: SISTEMA.

Seleccionamos en **Tabular options**  la opción “**Comparison of Standard Deviations**”. En dicha ventana el intervalo de confianza para el ratio de varianzas resulta:

95,0% Confidence Intervals

Ratio of Variances: [0,365657;5,92679]

Por defecto se calcula con un nivel de confianza del 95%. Contiene el valor 1, por lo que no hay diferencia significativa entre las varianzas. Para cambiar el nivel de confianza, en Pane Options modificar alfa.

¿Hay diferencia entre las varianzas con un riesgo de 1ªespecie $\alpha=15\%$?

Modifica el valor de α con Pane Options, y analiza si el nuevo intervalo para el ratio de varianzas contiene el 1.