

ANALISIS EXPLORATORIO DEL CASO: LABORATORIO 1

El objetivo fundamental de este trabajo, es mostrar al alumno(a) como se resuelve un problema completo, esto implica necesariamente aplicar y llevar a cabo un plan de análisis siguiendo las cinco etapas de "Una Investigación Estadística".

ETAPA-1: DESCRIPCIÓN Y PLANTEAMIENTO DEL PROBLEMA

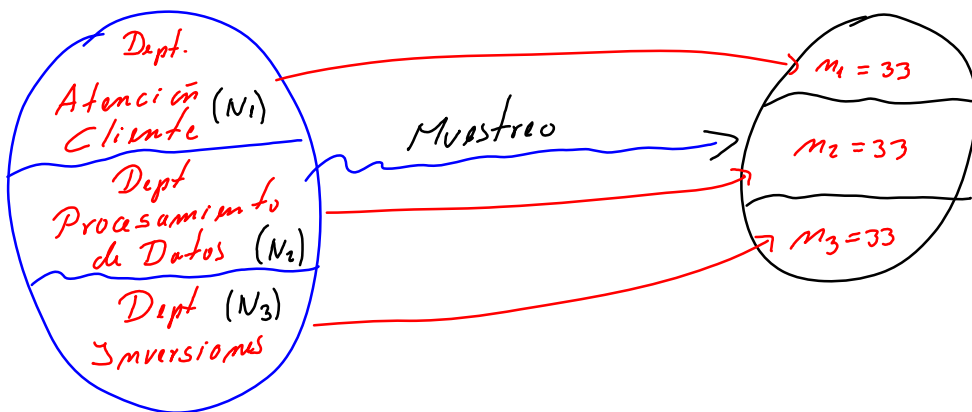
Aunque aparentemente puede parecer una labor sin importancia, describir claramente el problema es "FUNDAMENTAL", prueba de ello, es que la mayor dificultad que han tenido para resolver el problema, es no saber que se quiere, incluso llegando en algunos casos al extremo de no haber leído el problema:

Descripción del Problema: Se evalúa la variable X : mejora del rendimiento en trabajadores oficinistas. Estos trabajadores pertenecen a uno de tres departamentos (Atención al cliente, Procesamiento de Datos e Inversiones), quienes son evaluados en tres planes de trabajo distintos (Horario Flexible, 4 días Laborales y Horario Normal con Ambiente Mejorado)

NOTE, que el director de recursos humanos a priori piensa que existen condiciones de trabajo que podrían mejorar la productividad, lo que no sabe, es cuales son y a que departamentos favorece. Además esta implícito en el problema lo siguiente:

$$\text{Población } (N) \Rightarrow N = N_1 + N_2 + N_3$$

$$N = n_1 + n_2 + n_3 = 33 + 33 + 33 = 99$$



Por otro lado, es evidente que de los 33 trabajadores de cada departamento; 11 son evaluados con el plan de trabajo Horario Flexible; 11 con el plan: 4 días laborales y los restantes 11, son evaluados con el plan Horario Normal con ambiente mejorado.

Entonces, la pregunta fundamental es:

¿Es la mejora del rendimiento la misma para los tres departamentos, independientemente de los tres planes de trabajo que se evalúan para cada departamento.?

Es evidente, que la pregunta tiene dos respuesta posibles: "SI" o "NO". Si la respuesta es "SI", esto significa que la mejora del rendimiento es la misma en los tres departamentos independientemente del plan de trabajo. En este caso, estamos asumiendo lo siguiente:

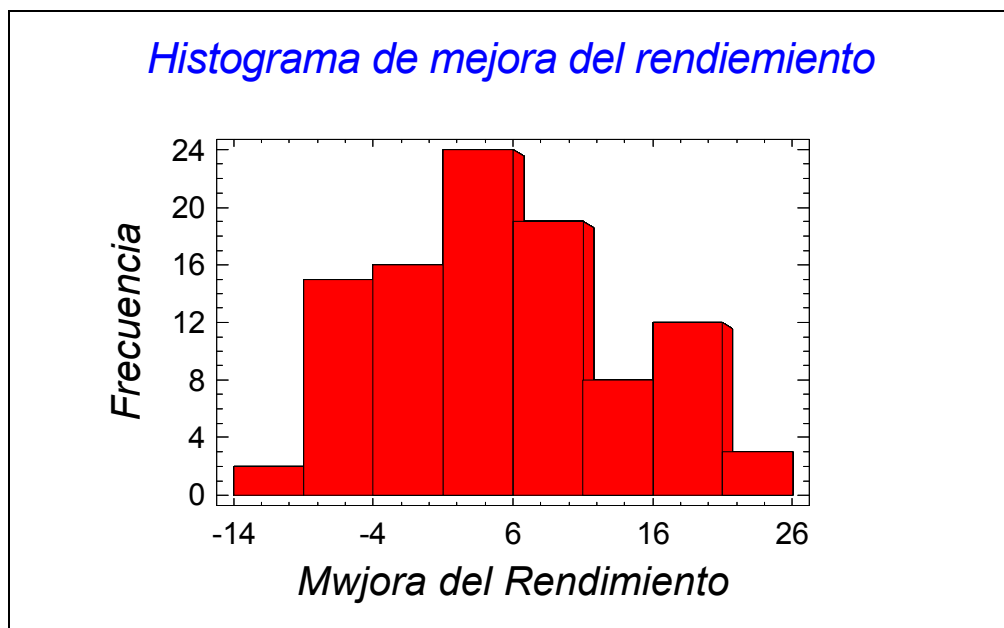
1. Cuando decimos "que la mejora del rendimiento es la misma", no estamos diciendo que los 33 trabajadores de un departamento tengan exactamente la misma mejora de rendimiento, sino que implícitamente nos estamos refiriendo al concepto de "Respuesta" \Leftrightarrow "Tendencia central", es decir, que pensamos que, en el caso de que asumamos que la medida de tendencia central adecuada para este caso es la media aritmética, que el promedio de los trabajadores para los tres planes de trabajo para un determinado departamento es igual, comportamiento que se asume también se replica en los dos departamentos restantes.
2. En este curso estudiamos tres medidas de tendencia central, las cuales están relacionadas con la naturaleza de los datos y la naturaleza del fin. La moda, es una medida de tendencia central que sólo es adecuada si los datos son cualitativos-Nominales, dado que la mediana es posible de obtener en datos de naturaleza ordinal por que estos tienen intrínsecamente la característica de orden, en cambio la media aritmética sólo es posible calcular en datos de naturaleza cuantitativa (Intervalar y razón). Sin embargo, la mediana también es posible de determinar cuando los datos son cuantitativos (dado que este tipo de datos por naturaleza también implica orden + magnitud), por tanto cabe la pregunta, ¿Cuándo es adecuado usar la Mediana y cuando la media Aritmética?, la respuesta es simple, el objetivo de una medida de tendencia central es dar cuenta del centro de la distribución de los datos, lo cual ocurre en el caso de la mediana siempre, sin embargo, para la media aritmética esto sólo ocurre si los datos simétricos o moderadamente asimétricos, en el caso contrario, la mediana siempre proporciona por su definición, el valor que divide la distribución de los datos en dos partes que contienen la misma cantidad de datos.
3. Para probar esto, deberíamos primero determinar si es más adecuado utilizar la media aritmética o la mediana como medida de tendencia central. Supongamos que hemos decidido que la media aritmética es la medida de tendencia central adecuada, entonces, esperamos que todas las medias sean muy parecidas o iguales entre sí.

Para probar si la respuesta media de los planes de trabajo es la misma en los tres departamentos, debemos proceder a realizar un PLAN DE ANALISIS DE DATOS.

QUE SABEMOS

Departamentos	1 : Reclamaciones
	2 : Procesamiento de Datos
	3 : Inversiones
Plan de trabajo	1 : Horario Flexible
	2 : 4 días laborales
	3 : Jornada normal, con ambiente mejorado

1. SITUACION GENERAL

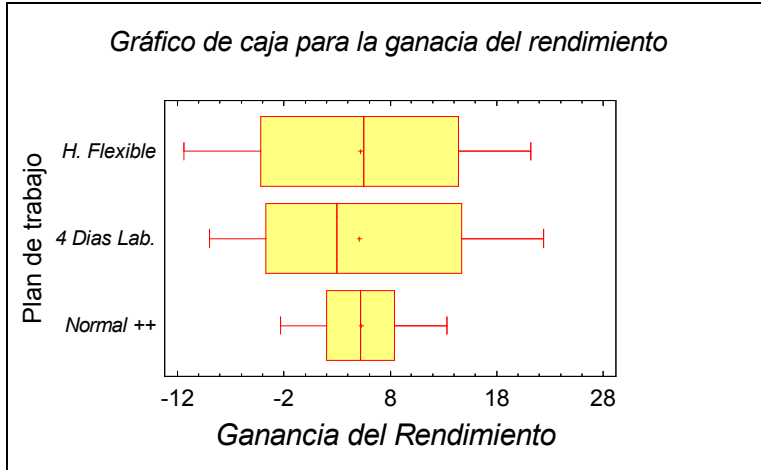


Summary Statistics for *Mejora del Rendimiento*

Count = 99
Average = 5,1798
Median = 5,2
Mode = 17,5
Variance = 67,3869
Standard deviation = 8,20895
Minimum = -11,4
Maximum = 22,4

Es evidente que esta descripción general además de proporcionar una idea de tendencia a la simetría de la distribución de la mejora del rendimiento, no aporta nada más, en consecuencia es necesario realizar un análisis más detallado de la variable en estudio (MEJORA DEL RENDIMIENTO). Por tanto parece razonable evaluar la "mejora del rendimiento" primero por departamento y luego por Plan de trabajo.

2. POR FORMA DE TRABAJO

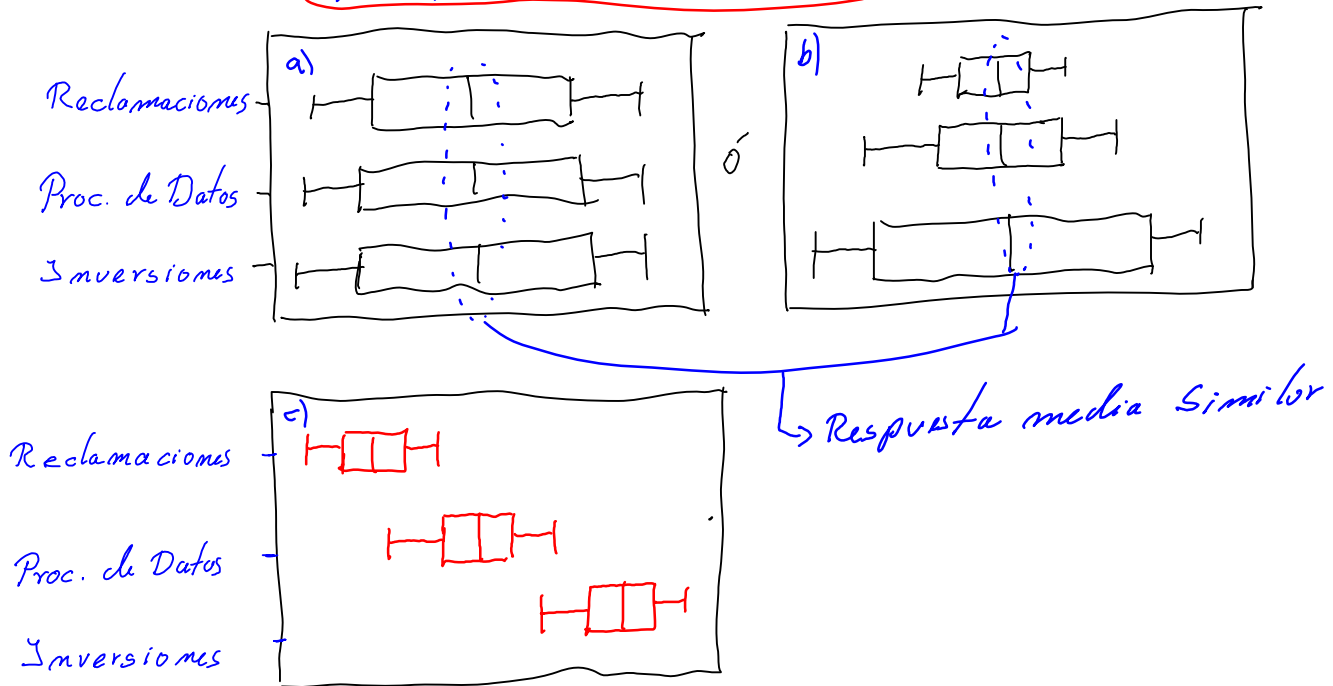


Code	Count	Average	Median	Variance	Deviation Std
1	33	5,18788	5,5	97,3267	9,86543
2	33	5,08485	3,0	92,1401	9,59896
3	33	5,26667	5,2	16,8885	4,10957
Total	99	5,1798	5,2	67,3869	8,20895

Si observamos el gráfico de caja múltiple como el resumen de medidas descriptivas asociado, notaremos que la respuesta en términos de tendencia central es muy similar para los tres planes de trabajo, esto tanto si comparamos sus medianas como sus medias aritméticas. Por otro lado, si observamos la dispersión, tanto en el gráfico de caja como en el resumen de medidas descriptivas, podemos observar la importante diferencia que existe de esta medida de un plan de trabajo a otro. En concreto, las cajas asociadas a los planes de trabajo horario flexible y 4 días laborales, son similares en tamaño, pero significativamente más grande que la caja asociada al plan de trabajo "horario normal ++", en consecuencia la dispersión de los planes de trabajo "horario flexible" y "4 días laborales" es significativamente mayor que los trabajadores que se sometieron al plan de trabajo "horario normal con ambiente mejorado", este fenómeno puede tener una de las siguientes causas:

- i. Para el caso de Horario Flexible, como es obvio se probó sobre 33 trabajadores, de los cuales 11 son del departamento de Reclamaciones, 11 de procesamiento de datos y 11 de inversiones, respecto de los cuales, no sabemos si todos tienen el mismo comportamiento o no, me explico:

Horario Flexible



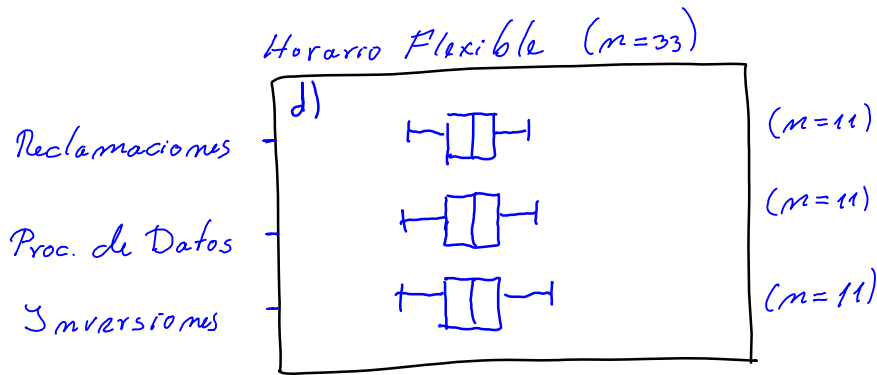
Los casos a, b y c, explican igualmente tanto la dispersión como la respuesta media de los trabajadores que fueron sometidos al plan de trabajo "Horario Flexible", sin embargo, cual de las situaciones exactamente es la que ocurre, sólo podremos conocerla realizando un análisis con un mayor nivel de detalle.

Observe, que los casos a y b generan igual dispersión e igual respuesta media, sin embargo, aún cuando el caso "c" genera la misma dispersión general, la respuesta media es distinta, por tanto, esta gran dispersión puede explicar dos comportamientos distintos:

- Los tres planes de trabajo para este departamento tienen gran dispersión, explicado por a o b.
- Los tres planes de trabajo no tienen gran dispersión, pero si respuestas medias bien distintas, proporcionando con este comportamiento una explicación igualmente adecuada a la significativa dispersión de la ganancia de rendimiento de los trabajadores sometidos a este plan de trabajo.

Para el plan de trabajo "4 días laborales", la explicación anterior también es válida.

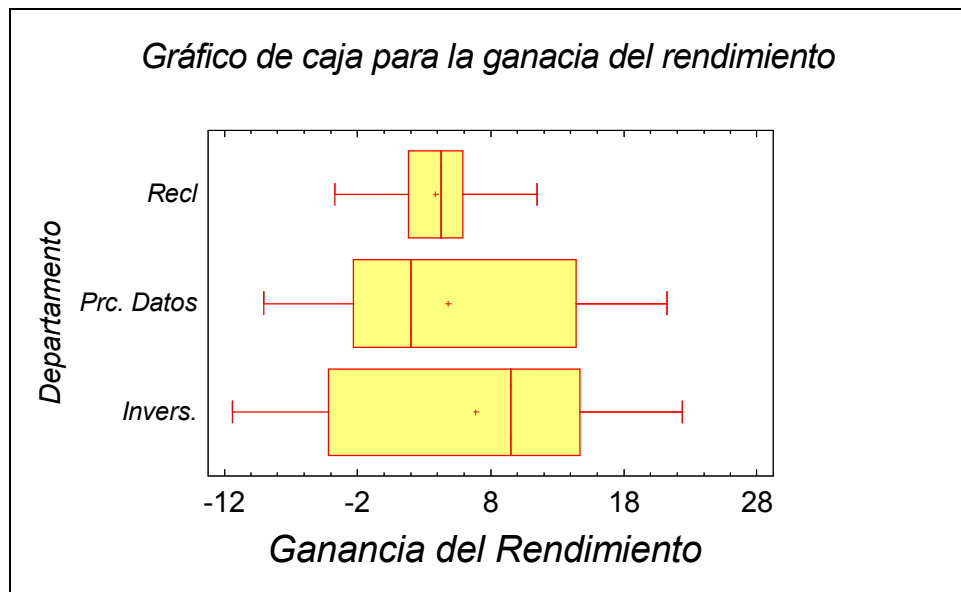
- i. Para el caso del plan de trabajo "horario normal con ambiente mejorado", la explicación no es exactamente la misma, aún cuando existe una posibilidad que también pueda ser válida, en este caso, resulta más probable el siguiente comportamiento:



Este comportamiento, en caso de corresponder a la verdadera situación para el plan de trabajo "horario normal ++", muestra que para los tres departamentos tiene igual comportamiento tanto respecto de la respuesta media como de la dispersión.

En resumen, si realizamos un análisis descriptivo de subgrupos, que sólo considera la situación por departamentos o por plan de trabajo, estamos implícitamente asumiendo que los comportamientos que se observan, son transversales para el otro criterio de clasificación, juicio para el cual no tenemos certeza alguna, y en el peor de los casos, podrías cometer un grave error si en realidad el verdadero comportamiento es el que muestra el dibujo c. En estadística, esta situación se denomina confusión de efectos, y pone de manifiesto la necesidad de analizar los datos considerando simultáneamente todos los criterios de clasificación que se tengan disponibles.

3. POR DEPARTAMENTO



Summary Statistics

Code	Count	Average	Median	Variance	Standard Deviation	Sum
1	33	3,88485	4,3	11,6388	3,41157	128,2
2	33	4,78788	2,0	91,3667	9,55859	158,0
3	33	6,86667	9,5	98,5448	9,92697	226,6
Total	99	5,1798	5,2	67,3869	8,20895	512,8

Para este caso, las explicaciones son análogas a las dadas para el análisis por plan de trabajo.

Con abse a todo lo anterior, es fundamental realizar análisis de la ganancia de rendimiento pero observada simultaneamente bajo los dos criterios de clasificación, Departamaneto y Plan de trabajo.

iv) POR DIVISION Y FORMA DE TRABAJO

Para explorar la variable aumento del rendimiento, observada bajo la acción combinada de división y forma de trabajo, se debe realizar lo siguiente:

1.- Crearemos la variable **DYT**, que contendrá las 9 categorías distintas producto de la combinación de las categorías de las variables Departamento y Plan de Trabajo. Esto se explica en la siguiente tabla.

Plan Depto	Plan de Trabajo		
	H P/rx. (1)	U drs (2)	Normal (3)
Reclamaciones 1	1	2	3
Proc de Datos 2	4	5	6
Inversiones 3	7	8	9

Ahora la cuestión es, ¿Cómo utilizamos esta información para generar una nueva variable cualitativa en STATGRAPHICS la cual contenga 9 categorías, fruto de la combinación de las variables categoricas Departamento y Plan de trabajo. En realidad, esta labor es sencilla, solo debemos conocer la logica de STATGRAPHICS que es como sigue:

$$(Expresión\ Lógica) = \begin{cases} V \rightarrow \text{Responde "1"} \\ F \rightarrow \text{Responde "0"} \end{cases}$$

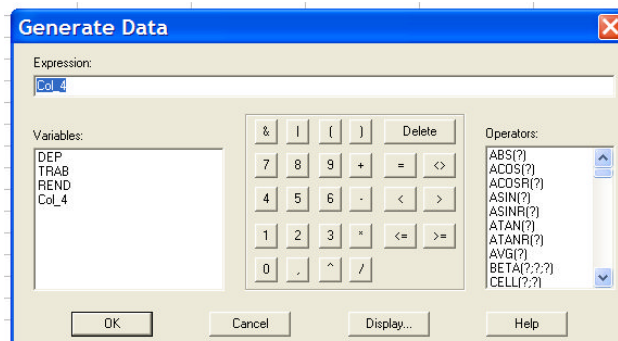
Además se debe considerar, que por razones de simplicidad aún cuando es posible crear la nueva variable de una sola vez, dicho proceso lo realizaremos creando tantas variables auxiliares como categorías tenga la nueva variable que se requiere crear, en este caso, 9 variables auxiliares, las cuales por cuestiones de parsimonia se propone crear siguiendo la siguiente pauta:

Departamento (Dept)	F. Trabajo (Trab)	Pregunta	A_i	Statgraphics
1	1	(Dept = 1) y (Trab = 1)	$A_1 = 1$	$1 * ((Dept = 1) \& (Trab = 1))$
1	2	(Dept = 1) y (Trab = 2)	$A_2 = 2$	$2 * ((Dept = 1) \& (Trab = 2))$
1	3	(Dept = 1) y (Trab = 3)	$A_3 = 3$	$3 * ((Dept = 1) \& (Trab = 3))$
2	1	(Dept = 2) y (Trab = 1)	$A_4 = 4$	$4 * ((Dept = 2) \& (Trab = 1))$
2	2	(Dept = 2) y (Trab = 2)	$A_5 = 5$	$5 * ((Dept = 2) \& (Trab = 2))$
2	3	(Dept = 2) y (Trab = 3)	$A_6 = 6$	$6 * ((Dept = 2) \& (Trab = 3))$
3	1	(Dept = 3) y (Trab = 1)	$A_7 = 7$	$7 * ((Dept = 3) \& (Trab = 1))$
3	2	(Dept = 3) y (Trab = 2)	$A_8 = 8$	$8 * ((Dept = 3) \& (Trab = 2))$
3	3	(Dept = 3) y (Trab = 3)	$A_9 = 9$	$9 * ((Dept = 3) \& (Trab = 3))$

De esta forma, cada una de las variables auxiliares A_i ($i=1, \dots, 9$) tomara solo dos valores posibles "0" o "1", con $i=1, 2, \dots, 9$. Por otra parte, las nuevas variables A_i son disjuntas, es decir, que si se suman todas las variables auxiliares, la nueva variable solo tomara valores en $i=1, 2, \dots, 9$.

2. La creación de las variables A_i en el programa STATGRAPHICS, se realiza de la siguiente forma:

- Asegurese de tener cargado en el visr de datos el archivo de mejora del rendimeiento.
- Marque una columna vacia sobre su nombre para que se marque completa..
- Utilizando la tecla de propiedades del mouse, normalmente la derecha, marque la opción que dice: GENERAR DATOS.
- En generar datos aparecera el siguiente cudro de dialogo



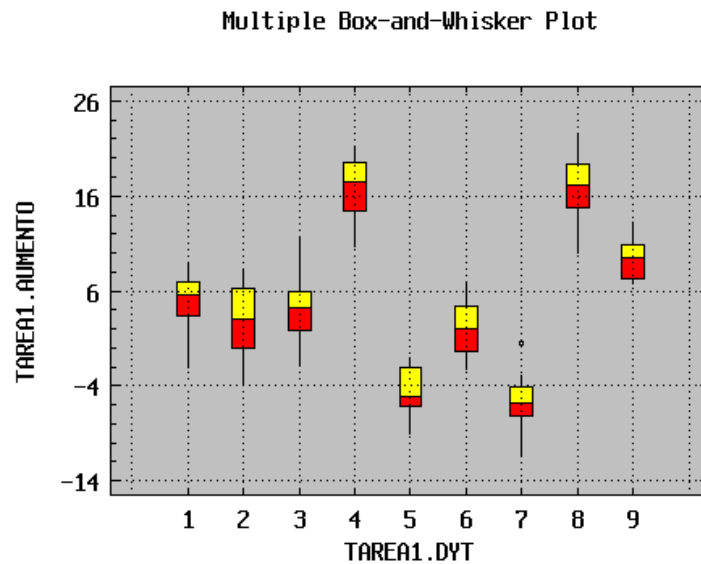
Escribiendo en la linea marcada como EXPRESION, la transformación que desea generar, en nuestro caso: $1 * ((Dept = 1) \& (Trab = 1))$, si queremos generar A_1 . Las restantes A_i se generan de forma analoga.

No olvide poner nombre a las variables.

3.- Por último, se creara la variable **DYT**, que contendrá todas las combinaciones posibles de **DIVISION** y **TRABAJO**, lo que se realizara como sigue:

$$\text{DYT} = A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9$$

5. Finalmente podemos utilizar la nueva variable DYT para construir el gráfico de caja y un resumen de medidas descriptivas según segmento.



Summary Statistics

Code	Count	Average	Median	Standard Variance	Deviation
1	11	4,32727	5,5	11,0242	3,32027
2	11	3,12727	3,0	13,4142	3,66254
3	11	4,2	4,3	11,85	3,44238
4	11	16,8909	17,5	10,4129	3,2269
5	11	-4,74545	-5,1	5,65873	2,37881
6	11	2,21818	2,0	7,93364	2,81667
7	11	-5,65455	-5,9	9,22273	3,03689
8	11	16,8727	17,2	12,4122	3,52309
9	11	9,38182	9,5	4,15764	2,03903
Total	99	5,1798	5,2	67,3869	8,20895

Depto de Reclamaciones
 Depto de Proc. de Datos
 Depto de Inversiones

Conjeturas

1. Departamento de reclamos no presenta diferencias en la respuesta, respecto del aumento en la productividad.
2. Departamento de procesamiento de datos.
 - El mayor aumento de productividad se genera con el plan de trabajo de horario flexible.
 - La menor productividad se produce en un plan de trabajo de semana laboral de 4 días.
3. Departamento de inversiones.
 - La mayor productividad se genera con el plan de trabajo de semana laboral de 4 días.
 - La menor productividad esta dada por la forma de trabajo de horario flexible.
4. El aumento del rendimiento del plan de trabajo, horario normal con ambiente mejorado, tiene una respuesta independiente de la división.

Espero que estas meditaciones un poco apuradas les sean de utilidad.

Jorge Cortés