

Estadística
Básica
con
R y R-Commander
(Versión Febrero 2008)

Autores:

A. J. Arriaza Gómez
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
S. Pérez Plaza
A. Sánchez Navas



Universidad
de Cádiz

Servicio de Publicaciones

Copyright ©2008 Universidad de Cádiz. Se concede permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre de GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Free Software Foundation. Una traducción de la licencia está incluida en la sección titulada "Licencia de Documentación Libre de GNU".

Copyright ©2008 Universidad de Cádiz. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation. A copy of the license is included in the section entitled "GNU Free Documentation License".

Edita: Servicio de Publicaciones de la Universidad de Cádiz
C/ Dr. Marañón, 3
11002 Cádiz

<http://www.uca.es/publicaciones>

ISBN:

Depósito legal:

Estadística Básica con R y R-commander
(Versión Febrero 2008)
Autores: A. J. Arriaza Gómez, F. Fernández Palacín,
M. A. López Sánchez, M. Muñoz Márquez, S. Pérez Plaza,
A. Sánchez Navas
©2008 Servicio de Publicaciones de la Universidad de Cádiz
<http://knuth.uca.es/ebrcmdr>

Capítulo 5

Inferencia clásica en poblaciones Normales

1. Conceptos fundamentales

Hasta ahora los objetivos planteados se han limitado a explorar un conjunto de datos describiendo sus características principales o las relaciones entre distintos caracteres. La intención de este capítulo es hacer una primera incursión en lo que se conoce como *análisis inferencial*, en el que a partir del estudio de una muestra pequeña y representativa de miembros de un gran colectivo, se extraen conclusiones que afectan a todos los elementos del mismo. Interesa, por ejemplo, conocer aproximadamente las principales características del colectivo, como pueden ser la media, la desviación típica, su estructura probabilística,...

El enfoque que se le va a dar a este tema se conoce como *clásico*. En él, las características poblacionales a estudiar se consideran parámetros (constantes desconocidas), mientras que los elementos de la muestra se consideran variables aleatorias. La alternativa a este enfoque vendría dada por la *teoría bayesiana*, en el que los parámetros son variables aleatorias, mientras que los datos que se poseen de la población son considerados constantes.

Desde un punto de vista intuitivo, parece razonable que si efectivamente la muestra representa bien al colectivo, los parámetros muestrales

sean muy parecidos a los poblacionales y aunque ciertamente este enfoque de *estimación puntual* es básicamente correcto, adolece de ciertas carencias que lo convierten sólo en una parte del proceso inferencial.

Interesa dar una mayor consistencia al análisis inferencial y ello se consigue desde dos puntos de vista, que en muchas ocasiones son complementarios: la construcción de *intervalos de confianza* y la realización de *contrastos de hipótesis*. Tanto uno como otro tienen en cuenta el margen de error derivado de cierta pérdida de información, que se produce al intentar explicar el comportamiento de una población a partir del conocimiento de una parte muy pequeña de sus miembros. Para ilustrar lo dicho se introduce el siguiente ejemplo:

Ejemplo 5.1

Una máquina está preparada para fabricar piezas de 7 cms de longitud. En una inspección se toman 1000 piezas fabricadas por dicha máquina, comprobándose que la media de éstas es de 7,0037 cms. Si se tomaran decisiones sólo a partir de esta estimación puntual habría que concluir que la máquina se ha desajustado y actuar en consecuencia. Pero se está desaprovechando información importante, como si la varianza de los datos es alta o pequeña, o si, como parece, la distribución de las longitudes es normal. La utilización de dicha información va a permitir construir un intervalo de confianza para la media de la población o confirmar directamente si ésta se puede considerar igual a 7 cms. En todo caso se estará asumiendo un margen de error derivado del proceso de extracción aleatorio de la muestra, ya que si se eligieran otras 1000 piezas la media sería distinta a la anterior.

En el caso de los intervalos de confianza, el objetivo es dar una cierta “garantía” de la presencia del parámetro dentro de un intervalo construido a partir de la muestra, mientras que para el caso de los contrastes, la pretensión es dar respuesta a si el valor del parámetro se encuentra, a la luz de la evidencia muestral, dentro de un conjunto de valores especificados en lo que se conoce como *hipótesis nula* (H_0) o, por el contrario, se haya dentro de su alternativo especificado por la *hipótesis alternativa* (H_1).

Se llama *nivel de confianza*, $1 - \alpha$, de un intervalo a la probabilidad (a priori) de que el intervalo contenga el valor del parámetro a estimar. La interpretación habitual del nivel de confianza es la probabilidad de que el intervalo de confianza, ya obtenido, contenga el valor del parámetro. Esta interpretación es incorrecta pues una vez obtenido el intervalo el valor del parámetro está o no está y no tiene sentido hablar de la probabilidad de que esto ocurra. $1 - \alpha$ debe interpretarse como la proporción teórica de intervalos (ya construidos) que contiene al valor del parámetro.

Para el caso de los contrastes, α es la probabilidad de rechazar la hipótesis nula cuando ésta es cierta y se conoce también como probabilidad de *error de tipo I*, $1 - \alpha$ también se llama aquí nivel de confianza. En el caso de los contrastes, existe un error asociado al α que se conoce como β y que indica la probabilidad de no rechazar la hipótesis nula cuando es falsa, conocido también como probabilidad de *error de tipo II*, $1 - \beta$ se conoce como *potencia del test*. Ambos errores son contrapuestos y fijado un tamaño muestral cuando uno de los dos crece el otro decrece. El cuadro que sigue recoge las distintas situaciones que pueden darse a la hora de realizar un contraste en término de los errores y aciertos.

		Decisión estadística	
		No rechazar H_0	Rechazar H_0
Estado Real de la cuestión	H_0 cierta	Correcta	Error tipo I
	H_0 falsa	Error tipo II	Correcta

En el peor de los casos, a la hora de realizar un estudio inferencial se cuenta con la información muestral, mientras que en las ocasiones más favorables, se tiene un conocimiento bastante aproximado de la estructura de probabilidad de la población analizada. Cuando se hace uso de la distribución de probabilidad de la población estudiada se dice que la inferencia realizada es *paramétrica*, mientras que si sólo se hace uso de la muestra, la inferencia es *no paramétrica*. El objetivo en los contrastes paramétricos es intentar obtener información sobre los parámetros desconocidos de la distribución de la población bajo estudio. En el caso de los contrastes no paramétricos, su objetivo es intentar determinar alguna característica de la población o de la muestra bajo estudio.

Puesto que los contrastes paramétricos utilizan más información que los no paramétricos, ofrecen mejores resultados. Por ello, siempre que sea posible se debe recurrir a los primeros.

Dependiendo de la estructura de sus hipótesis, se distingue entre los siguientes tipos de contrastes:

1. *Contrastes bilaterales*: en ellos se propone un valor puntual para el parámetro bajo estudio, de forma que se rechazará bien porque la evidencia muestral lleve a decidir que el valor es mayor que el propuesto o bien que es menor. Formalmente:

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

2. *Contrastes unilaterales*: en ellos se propone que el valor del parámetro se encuentre por debajo (o por encima) de un cierto valor. Las dos situaciones se plantearían de la siguiente forma:

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$

Se puede observar que en todos los casos el signo igual está incluido en la hipótesis nula, el motivo de ello se encuentra en el procedimiento que se va a utilizar para realizar el contraste.

Las distribuciones asociadas al proceso de muestreo son la **normal** y la **t de student** para el estudio de medias, la **Chi-cuadrado** para la varianza y la **F de Snedecor** para la comparación de varianzas; todas ellas estudiadas en el anterior capítulo. En general, interesa analizar el comportamiento de la media, aunque el mismo va a depender del conocimiento o no que se tenga de su varianza o si, para el caso de dos poblaciones sus varianzas coinciden. No hay que olvidar que la varianza determina la escala de la variable y siempre es más fácil comparar aquellas poblaciones con el mismo factor de escala.



Figura 5.1: Ventana de diálogo para el test t

Es muy importante entender que en el *contraste de hipótesis* los roles que juegan las hipótesis nula y alternativa no son equiparables y mucho menos intercambiables. En todo caso, hay que ver este enfoque como una regla de confirmación sobre una cuestión que el investigador cree razonablemente que es cierta, siendo la función del contraste la de validarla o, por el contrario, si la evidencia muestral en contra es muy fuerte, la de rechazarla.

En este capítulo se estudiarán problemas que involucran a una o dos poblaciones, mientras que en el capítulo 7 se generalizarán los resultados a más de dos poblaciones. Se aceptará, a expensas de poder comprobarlo en el próximo capítulo, que las poblaciones siguen distribuciones normales; caso de que esto no fuera cierto, habría que replantear el análisis desde una perspectiva no paramétrica. Además, se supondrá que las muestras extraídas son aleatorias y que no existen valores anómalos. Igual que para la normalidad, en el próximo capítulo se comprobarán estos supuestos.

2. Inferencias sobre una población

En esta sección se abordará el estudio de la media de una población, de la que se dispone de una muestra aleatoria simple de tamaño n . Aunque en el caso, poco frecuente, de que se conozca la varianza de la población se podría utilizar la distribución Normal, y que cuando el tamaño de la muestra sea grande ($n \geq 50$) la distribución t de student se puede reemplazar por la $N(0, 1)$, en general se empleará la propia t de student.

Ejemplo 5.2

Se considera que el fichero de datos `peso_altura.dat` es una muestra aleatoria simple de la población adulta de un municipio andaluz. Dicha muestra se utilizará para estudiar los valores medios del *peso* y la *altura* de la población.

- Las características muestrales se obtienen como siempre en Estadísticos→Resúmenes→Resúmenes numéricos..., seleccionando las correspondientes variables e indicando que se haga en función del sexo:

```
> numSummary(Datos[,c('ALTURA', 'PESO')],
groups=Datos$SEXO, statistics=c('mean', 'sd',
'quantiles'))
Variable: ALTURA
```

	mean	sd	0%	25%	50%	75%	100%	n
Mujer	171.0000	5.676462	159	167.00	170.5	175	182	46
Varón	177.1296	6.901043	167	171.25	178.0	182	194	54

```
Variable: PESO
```

	mean	sd	0%	25%	50%	75%	100%	n
Mujer	66.95652	4.340796	59	63.00	68.0	70	75	46
Varón	86.24074	10.504150	64	77.25	86.5	93	109	54

- Intervalos de confianza.** A continuación se obtendrán los intervalos de confianza del 95% para la altura de los hombres. Para ello se filtra la base de datos por la variable *sexo*. A continuación se marca Estadísticos→Medias→Test t para una muestra, seleccionando en la ventana de diálogo la variable que interesa, en este caso la *altura*, y comprobando que el nivel de confianza está fijado en el 0,95 (fig 5.1). Las instrucciones que se generan son:

```
> t.test(Hombres$ALTURA, alternative='two.sided', mu=0.0,
conf.level=.95)
One Sample t-test
data: Hombres$ALTURA
t = 188.6138, df = 53, p-value <2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
175.2460 179.0133
sample estimates:
mean of x
177.1296
```


De la salida interesa la parte que hace referencia al intervalo de confianza, la media de altura de la población de hombres se encuentra dentro del intervalo (175,24; 179,01) con una confianza, que no una probabilidad, del 95 %.

- Contraste bilateral.** Como se puede observar en las instrucciones de **R** generadas por **Rcmdr**, además de la variable y el nivel de confianza, el procedimiento `t.test` incluye dos opciones más. La primera de ellas es `alternative` y admite tres posibilidades: contraste bilateral `two.sided`, contraste unilateral $H_1 : \mu < \mu_0$ `less` y contraste unilateral $H_1 : \mu > \mu_0$ `greater`. La segunda opción permite fijar un valor para la hipótesis nula `mu=0.0`. Para realizar los distintos contrastes se va a retocar la línea de instrucciones. En primer lugar se desea realizar el contraste:

$$H_0 : \mu = 175$$

$$H_1 : \mu \neq 175$$

con un nivel de significación $\alpha = 0,01$. Editando la línea de instrucciones y ejecutando se tiene:

```
> t.test(Hombres$ALTURA, alternative='two.sided', mu=175.0,
conf.level=.99)
One Sample t-test
data: Hombres$ALTURA
t = 2.2677, df = 53, p-value = 0.02745
alternative hypothesis: true mean is not equal to 175
99 percent confidence interval:
174.6205 179.6388
sample estimates:
mean of x
177.1296
```

Se puede observar que, respecto a la salida anterior al aumentar el nivel de confianza ha aumentado la amplitud del intervalo y que el resto es prácticamente igual. Respecto al contraste se concluye que puesto que el `p-value`= 0,027, es mayor que el nivel de significación, $\alpha = 0,01$, no hay evidencias para rechazar la hipótesis nula. Se puede ver que en este caso el valor que H_0 propone para la media se encuentra dentro del intervalo de confianza. Esto no ocurría en la salida anterior donde se había fijado el nivel de confianza en 0,95, pues en ese caso 175 estaba fuera del intervalo.

- **Contraste unilateral.** Se plantea ahora la realización del contraste:

$$H_0 : \mu \geq 180$$

$$H_1 : \mu < 180$$

con un nivel de significación $\alpha = 0,1$. Se edita de nuevo la línea de instrucciones y se ejecuta:

```
> t.test(Hombres$ALTURA, alternative='less', mu=180.0,
conf.level=.90)
One Sample t-test
data: Hombres$ALTURA
t = -3.0565, df = 53, p-value = 0.001752
alternative hypothesis: true mean is less than 180
90 percent confidence interval:
-Inf 178.3483
sample estimates:
mean of x
177.1296
```

En este caso el p -valor=0,0017 es mucho menor que el nivel de significación y por tanto se rechaza la hipótesis nula. Igualmente se puede comprobar que 180 no pertenece al intervalo de confianza.

3. Inferencias sobre dos poblaciones

Para el caso de comparar las medias de dos poblaciones, además de comprobar las hipótesis sobre normalidad y aleatoriedad, que como ya se ha comentado se verán en el próximo capítulo, se plantean distintas situaciones. En primer lugar habrá que determinar si se tienen muestras independientes o pareadas (relacionadas). La diferencia entre uno y otro caso es que en el segundo, se dan dos mediciones de la misma o similar característica para cada individuo o para dos individuos de idénticas, respecto de los restantes, características relevantes de la muestra.

Si se miden el peso de 50 alevines de truchas antes y después de una cierta dieta alimenticia, ambas observaciones están relacionadas. La aplicación de dos pomadas en diferentes zonas de la piel de un individuo y la observación de ambas respuestas conduce a observaciones pareadas. A veces la dependencia no resulta tan evidente. La longitud de

la cola de trabajo de dos impresoras pueden parecer dos observaciones independientes, sin embargo, si ambas impresoras presentan idénticas características tanto en prestaciones como en accesibilidad, la elección del usuario dependerá de las longitudes de las colas existentes, introduciendo dependencia entre ambas longitudes.

Otra cuestión a tener en cuenta, para el caso de muestras independientes, es si las varianzas de las poblaciones se pueden considerar iguales o no.

3.1. Muestras independientes

Ejemplo 5.3

Para el caso de muestras independientes se usará el fichero `parque_eolico.dat`, que contiene datos de la velocidad del viento, registrados durante 730 horas de forma simultánea, en dos localizaciones alternativas (*Parque1* y *Parque2*). Se tratará de establecer la localización más aconsejable para la instalación de un parque de producción de energía eólica.

Hay que tener en cuenta, al importar este conjunto de datos, que el carácter decimal viene dado en este fichero mediante una coma. Por otra parte, la estructura de la base de datos es de dos columnas, conteniendo cada una de ellas las mediciones en cada localización. Aunque **R** puede trabajar con esta estructura de datos, resulta más manejable para **Rcmdr** si es transformada en dos variables, una continua que contenga las mediciones de viento y otra factor que indique la localización. Esto se realiza desde el menú Datos→Conjunto de datos activo→Apilar variables del conjunto de datos activo... En la ventana de diálogo (fig. 5.2) se pide el nombre de la nueva base de datos que se ha venido a llamar `eolico_apilado`, el nombre de la variable apilada, `velocidad`, y el nombre de la nueva variable factor, `parque`, cuyas clases se han denominado `Parque1` y `Parque2`.

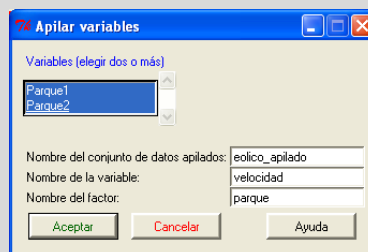


Fig. 5.2: Ventana para apilar `parque_eolico.dat`

Como se ha dicho es conveniente saber si las varianzas se pueden considerar iguales o no a la hora de comparar las dos poblaciones. Una primera idea sobre la igualdad de varianzas es mediante la representación simultánea de los diagramas de caja de las muestras. Desde Gráficas→Diagrama de caja..., se selecciona la variable *velocidad* y el grupo *parque*, obteniéndose la figura 5.3.

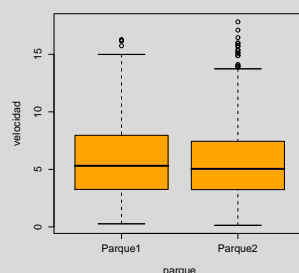


Fig. 5.3: Velocidad según tipo de parque

La comparación de los diagramas sugiere la igualdad de varianzas. El test *F* permite contrastar dicha hipótesis, desde Estadísticos→Varianzas→Test *F* para dos varianzas... seleccionando en este caso como factor la variable *parque* y como explicada la variable *velocidad*.

```
> tapply(eolico_apilado$velocidad, eolico_apilado$parque, var,
na.rm=TRUE)
Parque1 Parque2
10.50574 10.59477
> var.test(velocidad ~ parque, alternative='two.sided',
conf.level=.95, data=eolico_apilado)
F test to compare two variances
data: velocidad by parque
F = 0.9916, num df = 729, denom df = 729, p-value = 0.9093
alternative hypothesis: true ratio of variances is not equal to
1
95 percent confidence interval:
0.8574994 1.1466647
sample estimates:
ratio of variances
0.9915968
```

Como $p\text{-valor} = 0,9093 > 0,05$ no hay motivos para rechazar la igualdad de varianzas. Siendo así, como se supone que los datos están distribuidos normalmente y las varianzas son iguales, los dos parques eólicos serán igualmente productivos cuando la diferencia de sus medias no se separe significativamente de 0. Para realizar este contraste se selecciona Estadísticos→Medias→Test *t* para muestras independientes... y en la ventana de diálogo emergente se selecciona como grupo la variable *parque* y como variable explicada la *velocidad*, marcando la opción *bilateral* con el 95% de nivel de confianza y suponiendo las varianzas iguales.

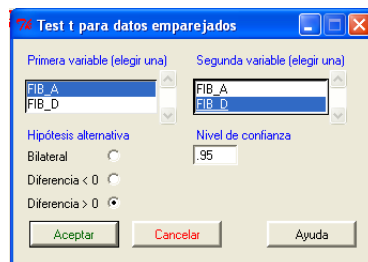


Figura 5.4: Contraste unilateral de fenofibrato

```
> t.test(velocidad~parque, alternative='two.sided',
conf.level=.95, var.equal=TRUE, data=eolico_apilado)
Two Sample t-test
data: velocidad by parque
t = 0.9937, df = 1458, p-value = 0.3205
alternative hypothesis: true difference in means is not equal to
0
95 percent confidence interval:
-0.1645533 0.5024437
sample estimates:
mean in group Parque1 mean in group Parque2
5.801795 5.632849
```

Al ser el p -valor = 0,32 > 0,05 no se rechaza que la diferencia de las medias sea cercana a cero.

3.2. Muestras pareadas

Ejemplo 5.4

Para el caso de muestras pareadas se tomará el conjunto de datos *fenofibrato.dat* en el que se quiere analizar si el tratamiento durante un año con fenofibrato reduce el fibrinógeno, contando para ello con una muestra de 32 individuos. Se efectúa el Test t en Estadísticos→Medias→Test t para datos relacionados..., realizando un contraste unilateral (figura 5.4).

```

> t.test(Datos$FIB_A, Datos$FIB_D, alternative='greater',
conf.level=.95,paired=TRUE)
Paired t-test
data: Datos$FIB_A and Datos$FIB_D
t = 7.5391, df = 31, p-value = 8.48e-09
alternative hypothesis: true difference in means is greater than
0
95 percent confidence interval:
57.8178 Inf
sample estimates:
mean of the differences
74.59375

```

Al ser el p -valor $< 0,001$ se rechaza la hipótesis nula, con lo que se acepta que la diferencia, entre los niveles iniciales y finales, es positiva. Con ello se puede deducir que el tratamiento anual con fenofibrato reduce los niveles de fibrinógeno en el organismo y existen así evidencias acerca de su efectividad. Si se deseara confirmar que el tratamiento produce un descenso de más de 50 puntos en el nivel de fenofibrato, se debería tocar ligeramente la instrucción **R** incluyendo ese dato:

```

> t.test(Datos$FIB_A, Datos$FIB_D, alternative='greater',
conf.level=.95, paired=TRUE, mu=50)
Paired t-test
data: Datos$FIB_A and Datos$FIB_D
t = 2.4857, df = 31, p-value = 0.009265
alternative hypothesis: true difference in means is greater than
50
95 percent confidence interval:
57.8178 Inf
sample estimates:
mean of the differences
74.59375

```

De nuevo dado que $p < 0,001$ se rechaza la hipótesis de que $\mu_A \leq \mu_D + 50$ y se concluye que el medicamento produce una disminución de más de 50 puntos en el nivel de fenofibrato.

4. Ejercicios

5.1 Utilizando el fichero de datos `peso_altura.dat` realice los siguientes ejercicios:

a) Obtenga el intervalo de confianza del 90 % para la altura de las mujeres.

b) Obtenga los intervalos del 95 % para el peso de hombres y mujeres.

c) Para un nivel de confianza del 99 % contraste si la media de la altura de las mujeres es mayor o igual a 173 cms y la de los hombres menor o igual a 175 cms. ¿Puede indicar la razón de este aparente contrasentido?

5.2 Para estudiar la diferencia de estaturas medias, medidas en centímetros, de estudiantes varones en las facultades de ciencias de Cádiz y Málaga, se toma una muestra aleatoria de 15 estudiantes en cada facultad, obteniéndose:

Cádiz	182	170	175	167	171	174	181	169
	174	174	170	176	168	178	180	
Málaga	181	173	177	170	170	175	169	169
	171	173	177	182	179	165	174	

Obtenga el intervalo de confianza al 99 % para la diferencia de estaturas medias entre ambos colectivos de estudiantes. Se supone que las estaturas siguen una distribución normal.

5.3 Se está realizando un estudio sobre la evolución del nivel de colesterol de las personas, para lo cual se seleccionan 10 individuos al azar y se les somete a una nueva dieta alimenticia durante seis meses, tras la cual se les volvió a medir el nivel de colesterol en mg/dl. Suponiendo normalidad, obtenga un intervalo de confianza al 90 % para la diferencia de medias.

Antes	200	156	178	241	240	256	245	220	235	200
Después	190	145	160	240	240	255	230	200	210	195

5.4 Una fábrica produce barras de hierro cuya longitud sigue una distribución Normal. A partir de la muestra:

100,9 101,2 100,2 100,4 99,8
 100,1 101,5 100,4 101,7 99,5.

a) Encuentre un intervalo de confianza para la longitud media.

b) Tras revisar la maquinaria, se obtuvo una nueva muestra:

99,7 100,7 97,8 98,8 101,4
 100,3 98,7 101,1 99,4 99,5.

Estudie si se produjo algún cambio en la longitud media de la barras.

5.5 Una empresa de transporte de mercancías tiene dos oficinas en una determinada ciudad. Al objeto de asignar un nuevo trabajador a una de las dos oficinas, la dirección de la empresa decide analizar la productividad de cada una de ellas, contabilizándose las facturaciones en los últimos doce meses (miles de euros).

Ofic. 1	13,7	12,1	12,3	8,9	9,7	10,1	12,7	11,0	13,2	9,7	10,1	9,9
Ofic. 2	9,8	9,9	10,0	10,3	9,5	9,3	11,1	13,9	9,8	9,5	7,3	7,9

Suponiendo la normalidad de ambas poblaciones, ¿existen diferencias de facturación entre las dos oficinas?

5.6 Una empresa le propone al director de una fábrica un nuevo método que, supuestamente, reduce el tiempo empleado en el montaje de uno de sus productos. Con el propósito de comparar tal método con el empleado habitualmente, seleccionó aleatoriamente a siete de sus empleados para que llevarsen a cabo el montaje con los dos sistemas y anotó los tiempos empleados en el montaje, obteniendo los siguientes resultados:

Trabajador	1	2	3	4	5	6	7
Método habitual	38	32	41	35	42	32	45
Método nuevo	30	32	34	37	35	26	38

Supuesto que el tiempo de montaje sigue una distribución Normal, ¿se puede afirmar que efectivamente el nuevo método reduce el tiempo en más de dos minutos?

