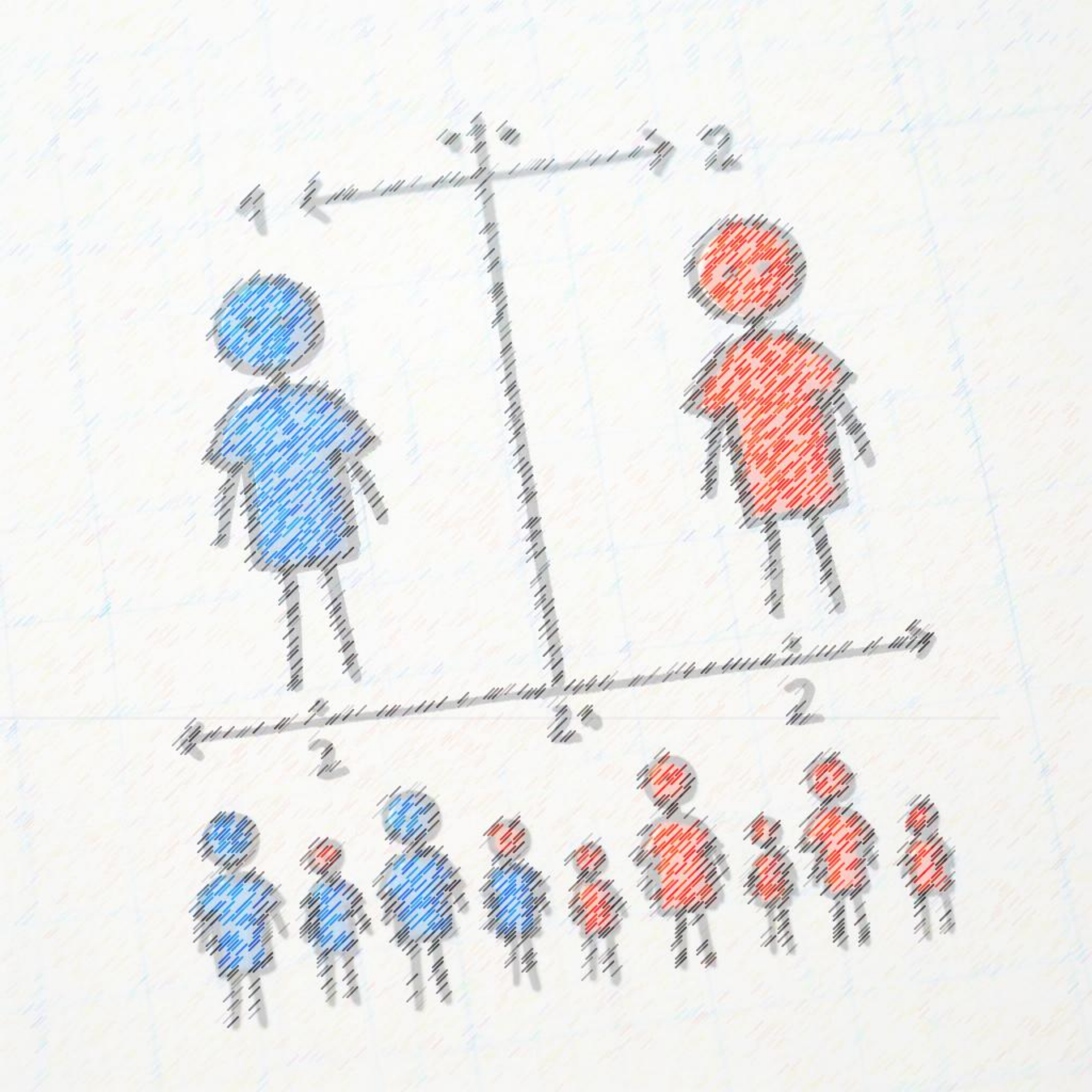
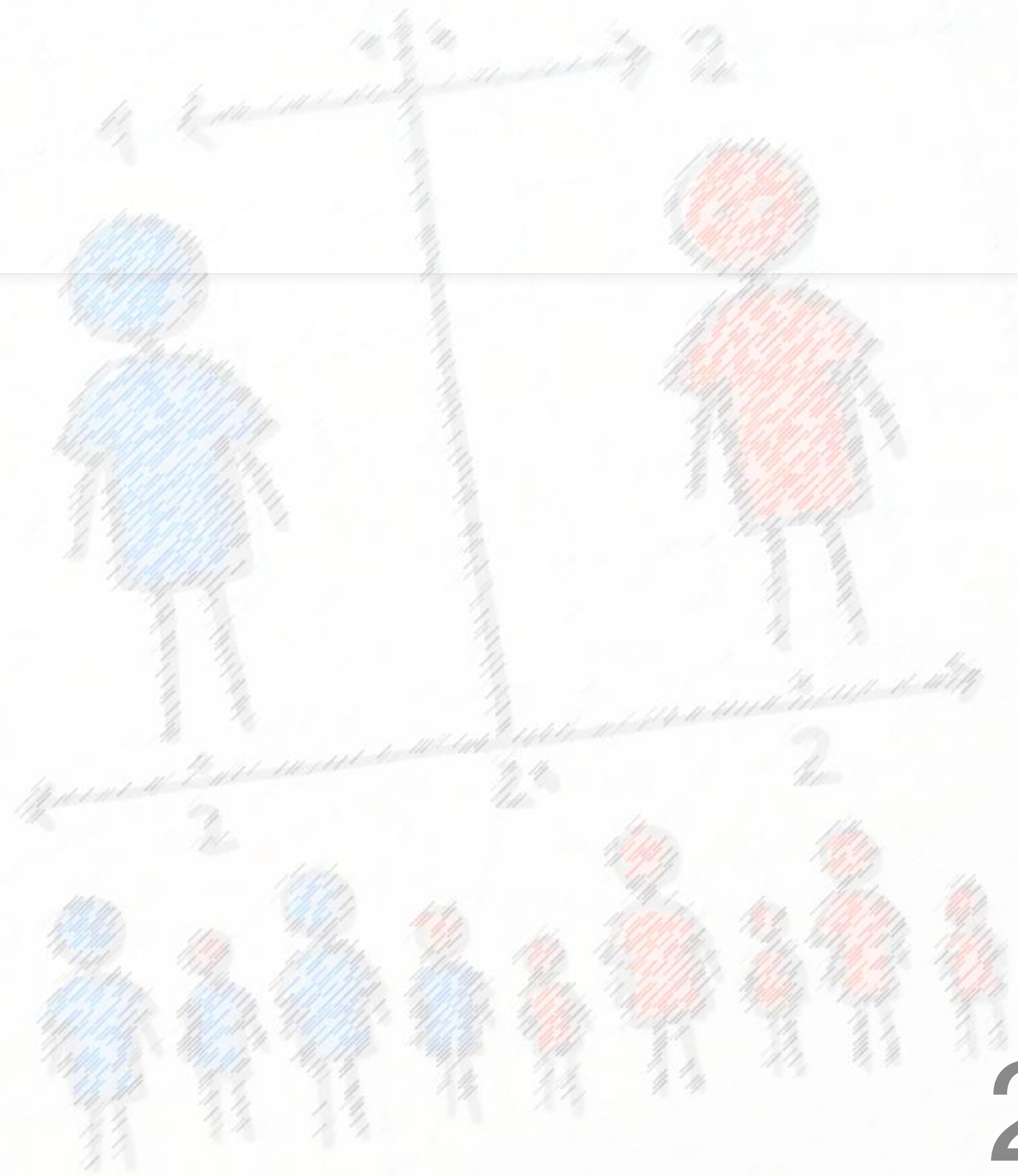
Experimentos comparativos simples

Basado en *Diseño y Análisis de Experimentos*, de Douglas C. Montgomery

Ignacio Díaz Oreiro
CI0131 Diseño de Experimentos



Agradecimiento a la profesora Dra. Kryscia Ramírez Benavides, por facilitar material utilizado en esta presentación.



Experimentos comparativos simples

- Se denomina experimentos comparativos simples a aquellos experimentos en que se comparan dos condiciones.
- A estas condiciones se les podría llamar tratamientos o dos niveles de un factor.
- Por ejemplo, determinar si dos versiones o configuraciones diferentes de un producto generan resultados equivalentes respecto de una variable de respuesta.

Experimentos comparativos simples

Ejemplo

- Se desea comparar la eficacia de un algoritmo de compresión de datos sin pérdida, respecto de una variante propuesta (algoritmo modificado).
- Se quiere determinar cuál logra una mayor tasa de compresión manteniendo la calidad y la integridad de los datos.
- La variable de respuesta será la relación de compresión lograda, expresada como una tasa de compresión, medida como la razón entre el tamaño del archivo original y el tamaño del archivo comprimido.

Experimentos comparativos simples

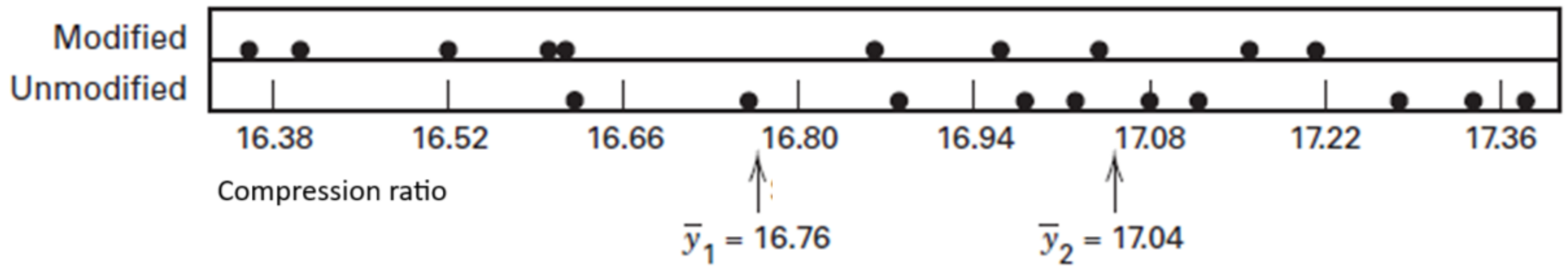
Ejemplo

- Los valores esperados se encuentran en un rango entre 16.0 y 18.0 (por ejemplo, esto indica que el archivo original es entre 16.0 y 18.0 veces más grande que el archivo comprimido).
- Se realizarán 10 observaciones de cada nivel (algoritmo original y algoritmo modificado).
- Esto dado que el proceso es costoso (procesamiento y tiempo) debido al conjunto de datos que se utilizará, el tamaño de los archivos a comprimir y la evaluación de la integridad de los archivos después de realizarles el proceso de descompresión.

Experimentos comparativos simples

Ejemplo

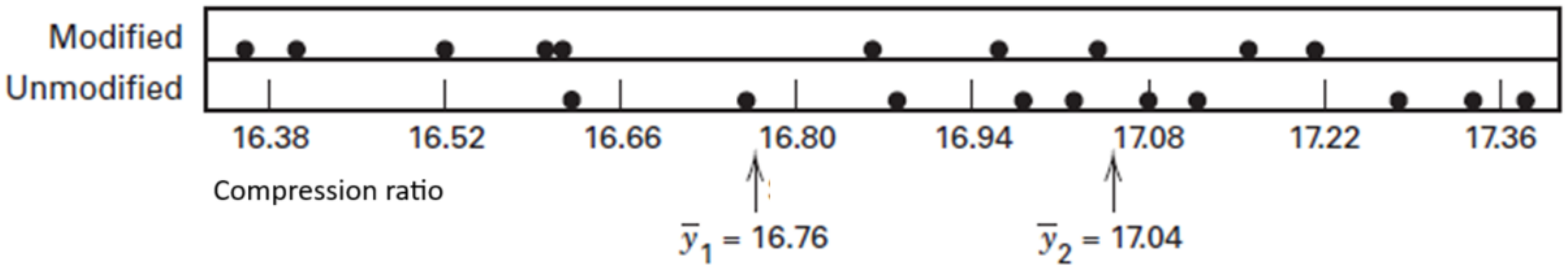
- Se obtienen las siguientes observaciones:
- Como parte del análisis exploratorio de datos se realiza un diagrama de puntos.



j	Modified Algorithm y_{1j}	Unmodified Algorithm y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Experimentos comparativos simples

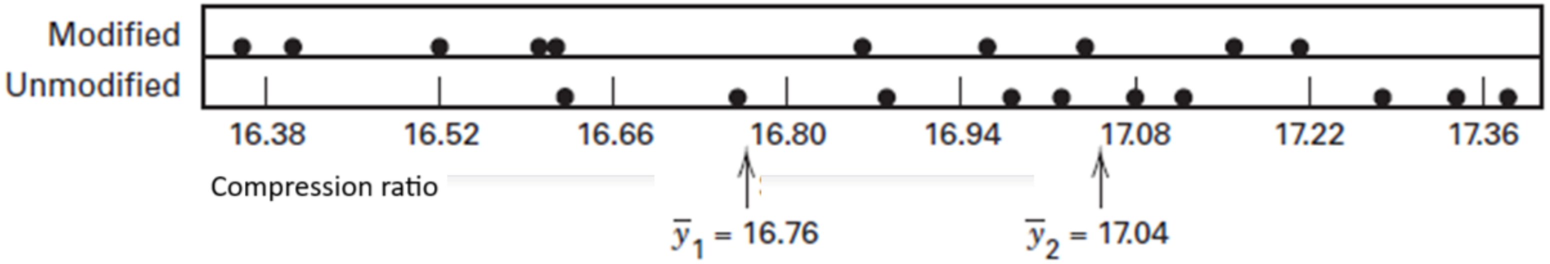
Ejemplo



j	Modified Algorithm y_{1j}	Unmodified Algorithm y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

- El análisis visual parece indicar que el algoritmo sin modificar obtiene una tasa mayor que el modificado, en lo que parece ser una cantidad no trivial .

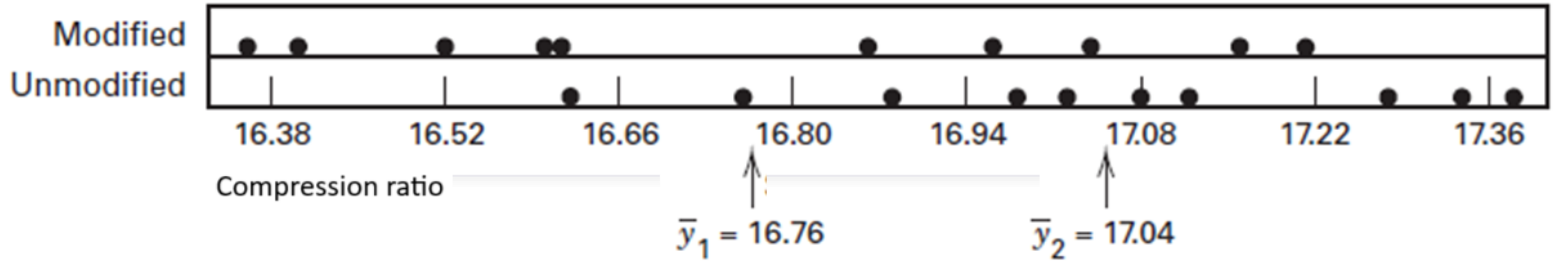
Ejemplo



- Sin embargo, no es evidente que esta diferencia sea de la magnitud suficiente para implicar que las dos implementaciones son en realidad diferentes.
- Quizás esta diferencia observada en las tasas promedio sea el resultado de fluctuaciones del muestreo y las dos implementaciones sean idénticas en realidad.
- Podrías suceder que otras dos muestras produzcan el resultado contrario, con la tasa del algoritmo modificado superando la del algoritmo original sin modificar.

Experimentos comparativos simples

Ejemplo



- Puede usarse una técnica de inferencia estadística llamada prueba de hipótesis (o prueba de significación) para auxiliar al experimentador en la comparación de estas dos implementaciones.
- La prueba de hipótesis permite que la comparación de las dos implementaciones se haga en términos objetivos, con el conocimiento de los riesgos asociados si se llega a una conclusión equivocada.

Conceptos estadísticos básicos

- A cada una de las observaciones del experimento del algoritmo de compresión citado anteriormente se le llamaría una corrida.
- Observe que las corridas individuales difieren, por lo que existen fluctuaciones, o ruido, en los resultados.
- Es común llamar a este ruido el error experimental o simplemente el error. Se trata de un error estadístico, lo cual significa que se origina por la variación que no está bajo control y que generalmente es inevitable.

Conceptos estadísticos básicos

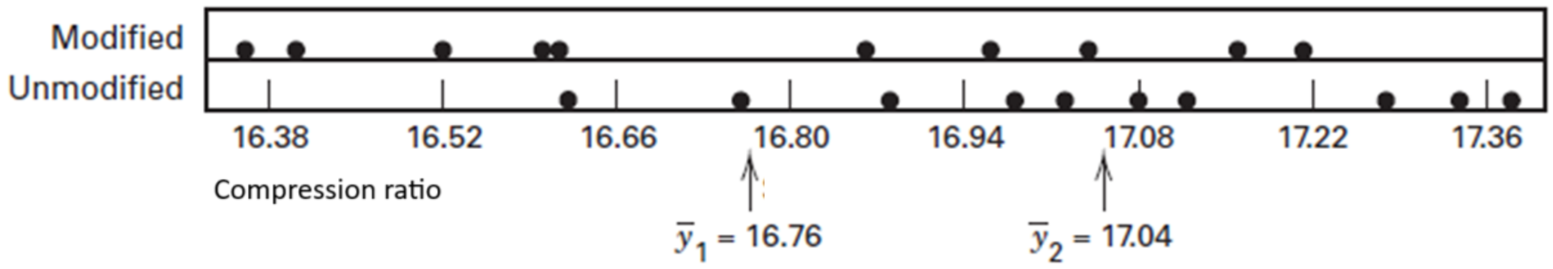
- La presencia del error o ruido implica que la variable de respuesta, la tasa de compresión, es una variable aleatoria.
- Una variable aleatoria puede ser discreta o continua.
- Si el conjunto de todos los valores posibles de la variable aleatoria es finito o contablemente infinito, entonces la variable aleatoria es discreta, mientras que, si el conjunto de todos los valores posibles de la variable aleatoria es un intervalo, entonces la variable aleatoria es continua.

Descripción gráfica de la variabilidad

- Es frecuente usar métodos gráficos simples como ayuda para analizar los datos de un experimento.
- El diagrama de puntos es un recurso muy útil para representar un cuerpo reducido de datos (digamos hasta unas 20 observaciones).
- El diagrama de puntos le permite al experimentador ver de inmediato la localización o tendencia central de las observaciones y su dispersión.

Conceptos estadísticos básicos

Descripción gráfica de la variabilidad

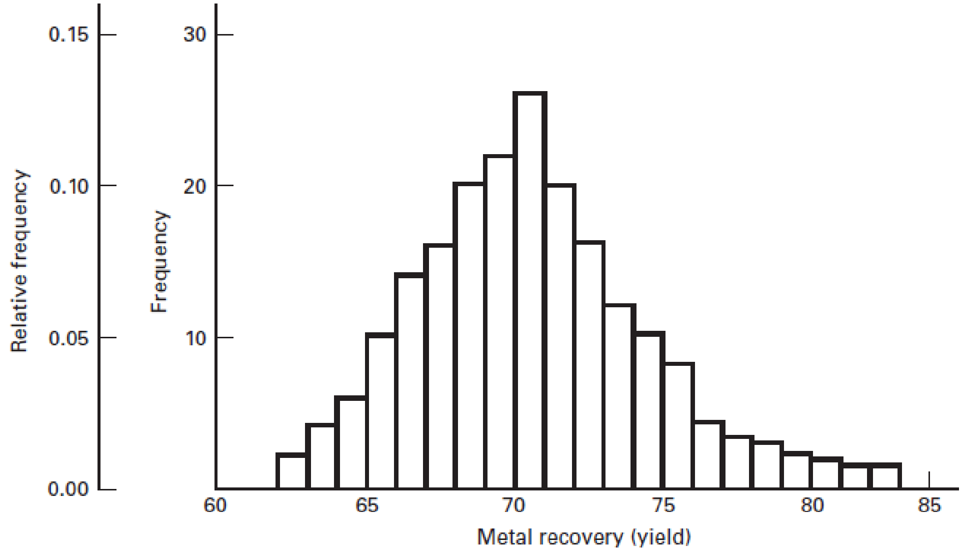


- En el experimento de la tasa de compresión del algoritmo de compresión de datos, el diagrama de puntos revela que probablemente las dos implementaciones difieran en la tasa promedio, pero que ambas presentan aproximadamente la misma variación en dicha tasa.

Conceptos estadísticos básicos

Descripción gráfica de la variabilidad

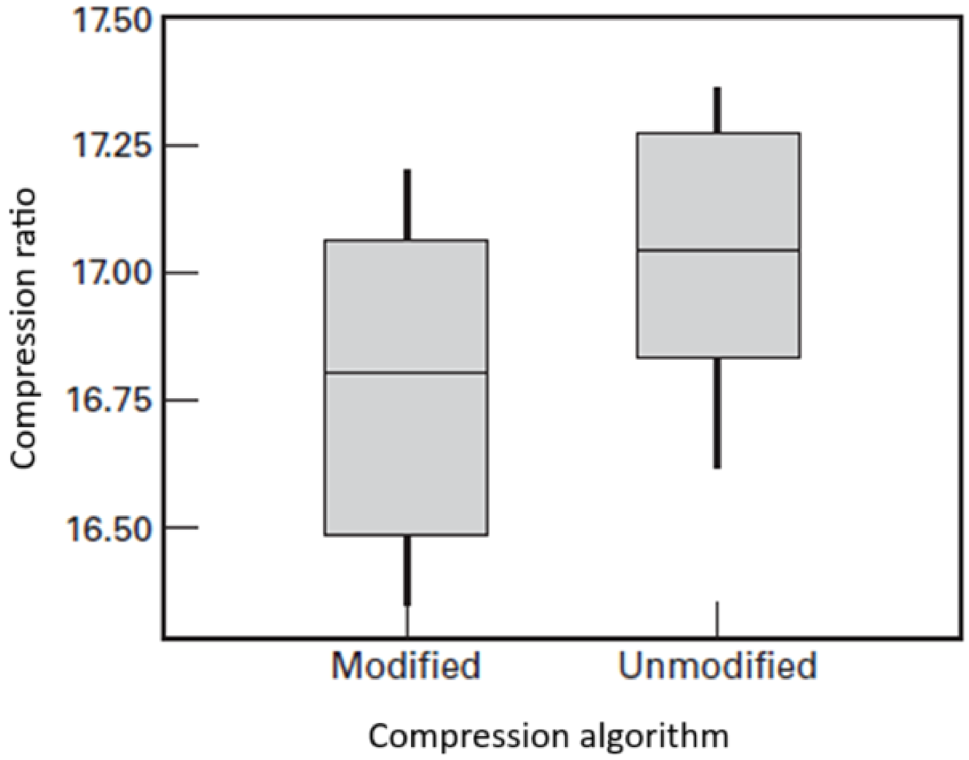
- Cuando los datos son muy numerosos, será preferible usar un histograma.
- Por ejemplo, un histograma de 200 observaciones de la recuperación de metal en un proceso de fundición.
- El histograma muestra la tendencia central, la dispersión y la forma general de la distribución de los datos.



Conceptos estadísticos básicos

Descripción gráfica de la variabilidad

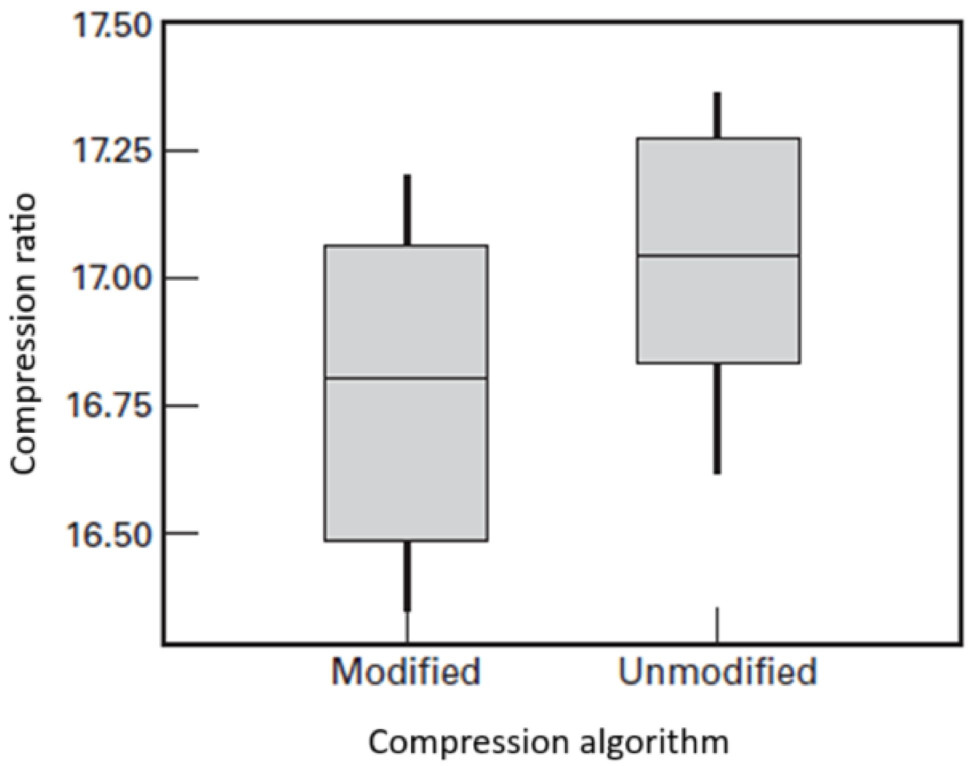
- El diagrama de caja es una manera muy útil de representar gráficamente los datos.
- Muestra el mínimo, el máximo, los cuartiles inferior y superior (el percentil 25 y 75, respectivamente) y la mediana (el percentil 50).
- Se trazan dos líneas (o bigotes) que se extienden de los extremos de la caja hasta los valores mínimo y máximo.



Conceptos estadísticos básicos

Descripción gráfica de la variabilidad

- Los diagramas de caja de las dos muestras de la tasa de compresión revelan con claridad la diferencia en la compresión promedio entre las dos implementaciones.
- Indican también que ambas implementaciones producen tasas de compresión razonablemente simétricas con una variabilidad o dispersión similar.



Inferencias acerca de las diferencias en las medias
Diseños aleatorios

- Supongamos que se cuenta con un diseño experimental completamente aleatorizado.
- En este diseño, los datos se consideran como si fueran una muestra aleatoria de una distribución normal.
- ¿Cómo pueden analizarse los datos de un experimento comparativo simple utilizando procedimientos de prueba de hipótesis para comparar las medias de dos tratamientos?

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

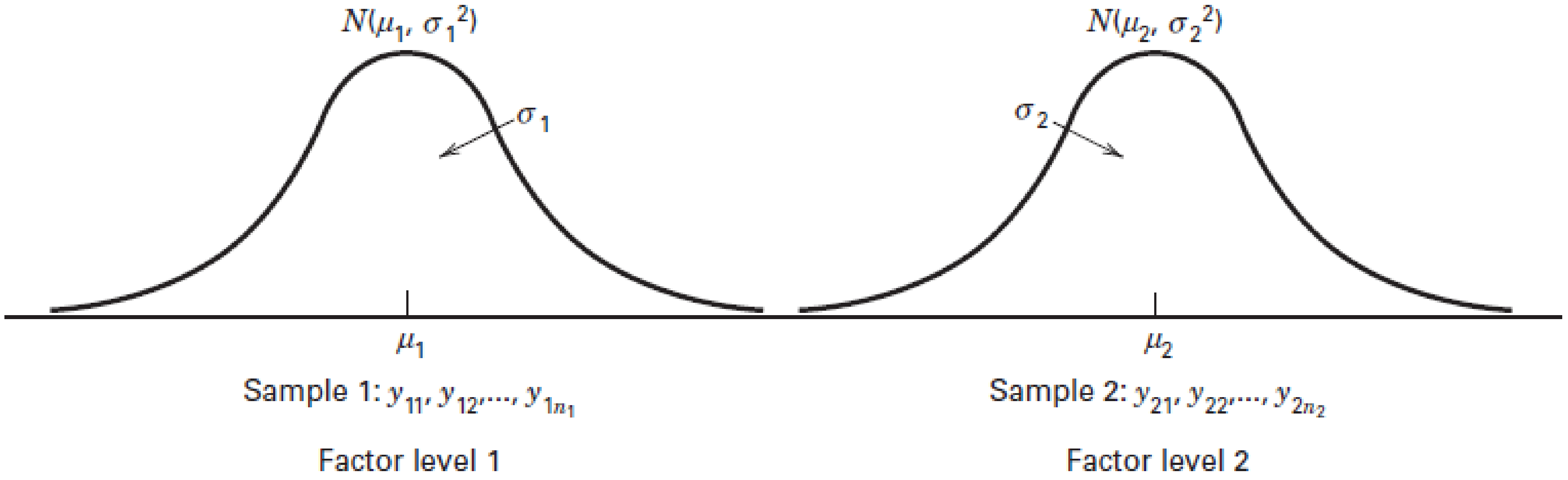
- En el experimento de la compresión de datos se quiere comparar la tasa de compresión de dos implementaciones: una con el algoritmo original y otra con el algoritmo modificado.
- Estas dos implementaciones son dos niveles del factor “implementación".
- Sea que y_{11}, y_{12}, ..., y_{1n} representan las n_1 observaciones del primer nivel del factor y que y_{21}, y_{22}, ..., y_{2n} representan las n_2 observaciones del segundo nivel del factor.

	Modified Algorithm	Unmodified Algorithm
j	y_{1j}	y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Supongamos que las muestras se sacan al azar de dos poblaciones normales independientes:



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Con frecuencia los resultados de un experimento se describen con un modelo estadístico que describe los datos. Por ejemplo, para el experimento anterior:

$$y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, n_i \end{cases}$$

- Donde y_{ij} es la observación j-ésima del nivel i del factor, μ_i es la media de la respuesta para el nivel i-ésimo del factor, y ε_{ij} es una variable aleatoria normal asociada con la observación ij-ésima.
- Se supone que las ε_{ij} son NID(0, σ^2_i), i = 1, 2 y se acostumbra a referenciarlas como el componente del error aleatorio del modelo.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Una hipótesis estadística es un enunciado o afirmación ya sea acerca de los parámetros de una distribución de probabilidad o de los parámetros de un modelo.
- La hipótesis refleja alguna conjetura acerca de la situación del problema.
- Por ejemplo, en el experimento de la compresión, puede pensarse que las tasas de compresión promedio de las dos implementaciones del algoritmo son iguales.
- Esto puede enunciarse formalmente como:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

- Donde μ_1 es la tasa de compresión promedio del algoritmo modificado y μ_2 es la tasa de compresión promedio del algoritmo original, sin modificar.
- Al enunciado $H_0 : \mu_1 = \mu_2$ se le llama la hipótesis nula y a $H_1 : \mu_1 \neq \mu_2$ se le llama la hipótesis alternativa.
- A la hipótesis alternativa que se especifica aquí se le llama hipótesis alternativa de dos colas porque sería verdadera si $\mu_1 < \mu_2$ o $\mu_1 > \mu_2$.

Prueba de hipótesis

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- Para probar una hipótesis se proyecta un procedimiento para tomar una muestra aleatoria, calcular un estadístico de prueba apropiado para después rechazar o no estar en posición de rechazar la hipótesis nula H_0.
- Parte de este procedimiento consiste en especificar el conjunto de valores del estadístico de prueba que llevan al rechazo de H_0.
- A este conjunto de valores se le llama la región crítica o región de rechazo de la prueba.

Prueba de hipótesis

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

- Pueden cometerse dos tipos de errores cuando se prueban hipótesis.
- Si la hipótesis nula se rechaza cuando es verdadera, ha ocurrido un error tipo I.
- Si la hipótesis nula no se rechaza cuando es falsa, se ha cometido un error tipo II.
- Las probabilidades de estos dos errores se expresan con símbolos especiales:

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$
$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 | H_0 \text{ is false})$$

Prueba de hipótesis

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$
$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 | H_0 \text{ is false})$$

- En ocasiones es más conveniente trabajar con la potencia de la prueba, donde:

$$\text{Power} = 1 - \beta = P(\text{reject } H_0 | H_0 \text{ is false})$$

- El procedimiento general en la prueba de hipótesis es especificar un valor de la probabilidad $\boldsymbol{\alpha}$ del error tipo I, llamada el nivel de significación de la prueba, y después diseñar el procedimiento de prueba de tal modo que la probabilidad **β** del error tipo II tenga un valor convenientemente pequeño.

Prueba de hipótesis

La prueba *t* de dos muestras.

- Considere que puede suponerse que las varianzas de las tasas de compresión fueron idénticas para ambas implementaciones del algoritmo.
- El estadístico de prueba para comparar las medias de dos tratamientos en el diseño completamente aleatorizado es la prueba *t* de dos muestras.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

El estadístico de prueba que deberá usarse (prueba *t* de dos muestras) es:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- $\bar{y}_1, \bar{y}_2$ son las medias muestrales
- n_1, n_2 son los tamaños de las muestras
- S_p^2 es una estimación de la varianza común $\sigma_1^2 = \sigma_2^2 = \sigma^2$ calculada a partir de:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- S_1^2, S_2^2 son las dos varianzas muestrales individuales.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- A este rasgo en el denominador se le denomina a menudo error estándar de la diferencia de medias en el numerador.
- Se abrevia como: se ($\bar{y}_1$ - $\bar{y}_2$)

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\boxed{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Para determinar si debe rechazarse $H_0 : \mu_1 = \mu_2$ se compara t_0 con la distribución t con $n_1 + n_2 - 2$ grados de libertad.
- Si $| t_0 | > t_{\alpha/2, n1+n2-2}$ (donde $t_{\alpha/2, n1+n2-2}$ es el punto porcentual $\alpha/2$ superior de la distribución t con $n_1 + n_2 - 2$ grados de libertad) entonces se rechazaría H_0 y se concluiría que las fuerzas promedio de las dos implementaciones del algoritmo de compresión difieren.

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Si H_0 es verdadera, se esperaría que 100(1 – α) por ciento de los valores de t_0 estén entre $-t_{\alpha/2,\, n1 + n2 - 2}$ y $t_{\alpha/2,\, n1 + n2 - 2}$
- Una muestra que produjera un valor de t_0 que estuviera fuera de estos límites sería inusual si la hipótesis nula fuera verdadera y es evidencia de que H_0 deberá rechazarse.

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- En algunos problemas quizá quiera rechazarse H_0 únicamente si una de las medias es mayor que la otra.
- Por lo tanto, se especificaría una hipótesis alternativa de una cola $H_1: \mu_1 > \mu_2$ y H_0 solo se rechazaría si $t_0 > t_{\alpha/2\,,\, n1 + n2 - 2}$
- Si se desea rechazar H_0 solo si $\mu_1 < \mu_2$ *, entonces la hipótesis alternativa es* $H_1: \mu_1 < \mu_2$ y H_0 se rechazaría si $t_0 < -t_{\alpha/2\,,\, n1 + n2 - 2}$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

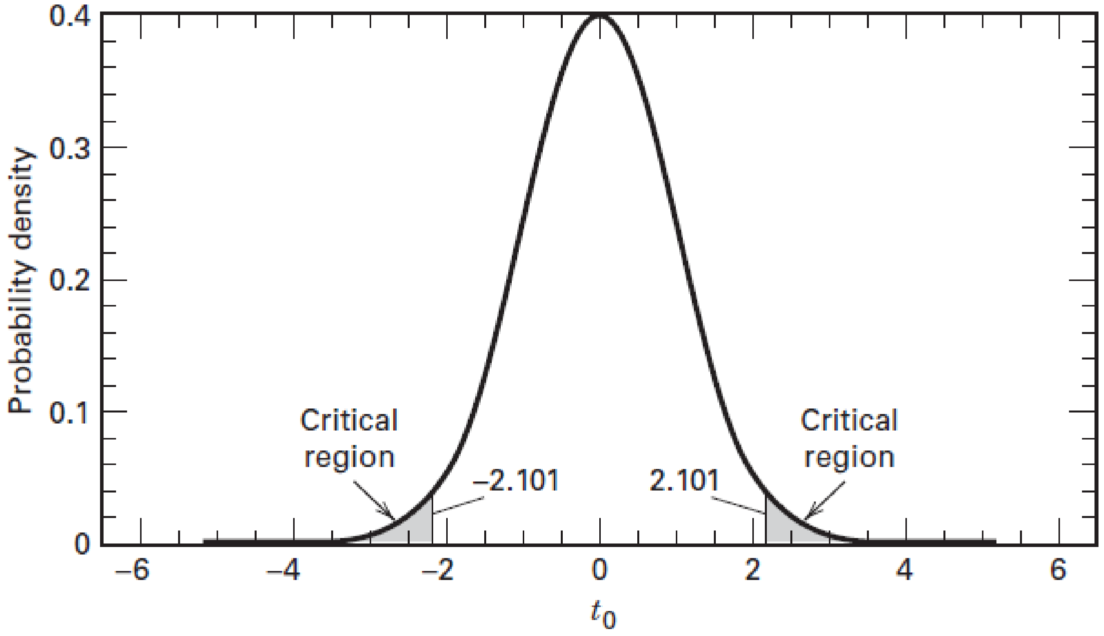
- En el ejemplo de la compresión de datos se tendría:

Modified Algorithm	Unmodified Algorithm
$\bar{y}_1 = 16.76$	$\bar{y}_2 = 17.04$
$S_1^2 = 0.100$	$S_2^2 = 0.061$
$S_1 = 0.316$	$S_2 = 0.248$
$n_1 = 10$	$n_2 = 10$

- Puesto que S_1 y S_2 son razonablemente similares, no es improcedente concluir que las desviaciones estándar (o las varianzas) poblacionales son iguales.
- Por lo tanto, puede usarse la prueba t de dos muestras para probar las hipótesis $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

Prueba de hipótesis

- Los grados de libertad son 18 ($n_1 + n_2 - 2 = 10 + 10 - 2$) y si se elige $\alpha = 0.05$, entonces $H_0 : \mu_1 = \mu_2$ se rechazaría si el valor teórico del estadístico de prueba $t_0 > t_{0.025,\,18} = 2.101$, o si $t_0 < -t_{0.025,\,18} = -2.101$
- Estos límites de la región crítica se ilustran en la distribución t con 18 grados de libertad.



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Usando las ecuaciones para prueba *t* de dos muestras:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

$$S_p = 0.284$$

- Y el estadístico de prueba t_0 es:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{-0.28}{0.127} = -2.20$$

- Puesto que $t_0 = -2.20 < -t_{0.025,\,18} = -2.101$ se rechaza H_0 y se concluiría que las tasas de compresión promedio de las dos implementaciones del algoritmo son diferentes.

Prueba de hipótesis

El uso de *valores P* en la prueba de hipótesis.

- Una manera de reportar los resultados de una prueba es estableciendo que la hipótesis nula fue rechazada o no para un valor de α (nivel de significación) específico.
- Por ejemplo, en el experimento de compresión de datos puede decirse que $H_0: \mu_1 = \mu_2$ se rechazó con el nivel de significación 0.05.
- Esto puede no ser adecuado porque no precisa si el valor del estadístico de prueba apenas rebasó la región de rechazo o si se adentró bastante en la misma.

Prueba de hipótesis

- Además, al darse los resultados de esta manera se impone, a otros usuarios de la información, el nivel de significación predefinido.
- Este enfoque puede ser insatisfactorio porque algunos investigadores o responsables de la toma de decisiones podrían sentirse incómodos con los riesgos que implica el valor $\alpha = 0.05$.
- Para evitar estas dificultades, en la práctica se ha adoptado extensivamente el enfoque del *valor P*.

Prueba de hipótesis

- El valor P es la probabilidad de que el estadístico de prueba asuma un valor que sea al menos tan extremo como el valor observado del estadístico cuando la hipótesis nula H_0 es verdadera.
- Por lo tanto, un valor P transmite mucha información acerca del peso de la evidencia en contra de H_0 y, por consiguiente, se puede llegar a una conclusión con cualquier nivel de significación especificado.
- En términos más formales, el valor P se define como el nivel de significación menor que llevaría a rechazar la hipótesis nula.

Prueba de hipótesis

- No siempre es sencillo calcular el valor P exacto de una prueba.
- Sin embargo, los programas de computación para realizar análisis estadísticos reportan valores P.
- En el ejemplo del algoritmo de compresión, para el valor t_0 = -2.20 se obtiene el valor P = 0.0411
- Por lo tanto, la hipótesis nula $H_0 : \mu_1 = \mu_2$ se rechazaría con cualquier nivel de significación $\alpha > 0.0411$.

Prueba de hipótesis

Computer Output for the Two-Sample *t*-Test

```
Minitab
Two-sample T for Modified vs Unmodified

              N     Mean    Std. Dev.   SE Mean
Modified     10   16.764        0.316      0.10
Unmodified   10   17.042        0.248     0.078

Difference = mu (Modified) − mu (Unmodified)
Estimate for difference: −0.278000
95% CI for difference: (−0.545073, −0.010927)
T-Test of difference = 0 (vs not = ): T-Value = −2.19
P-Value = 0.042 DF = 18
Both use Pooled Std. Dev. = 0.2843
```

Prueba de hipótesis

Verificación de supuestos de la prueba *t*.

Para utilizar el procedimiento de la prueba *t* se establecen los supuestos de que:

- ambas muestras se toman de poblaciones independientes que pueden describirse con una distribución normal,
- las desviaciones estándar o las varianzas de ambas poblaciones son iguales,
- y que las observaciones son variables aleatorias independientes.

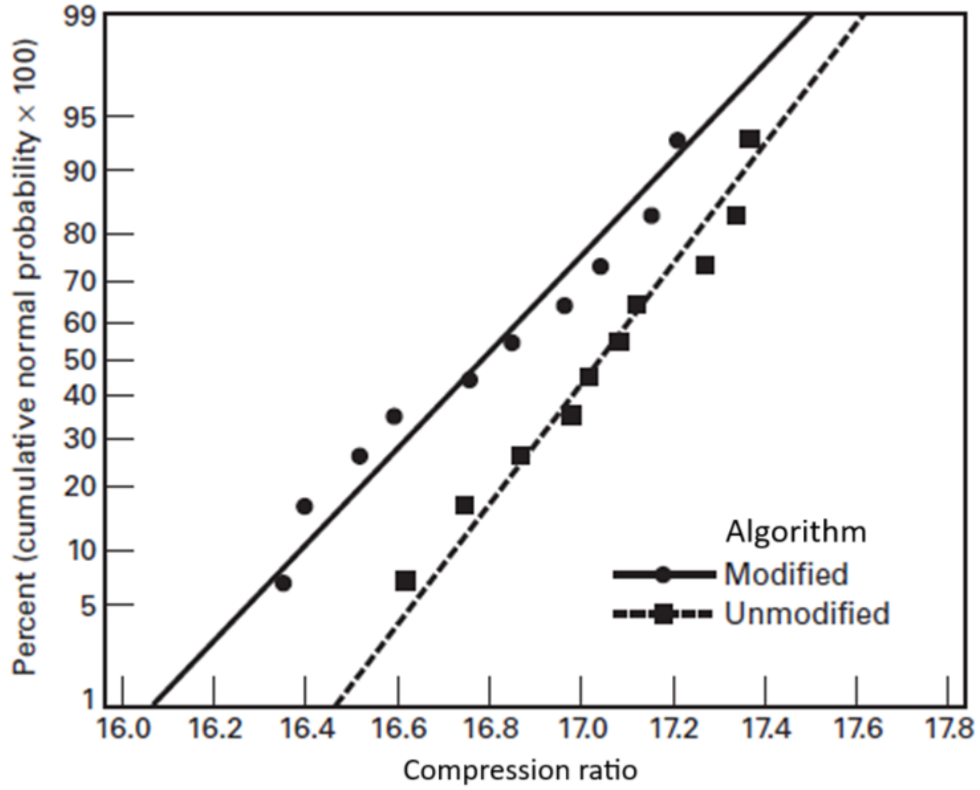
Prueba de hipótesis

Verificación de supuestos de la prueba *t*.

- El supuesto de independencia es crítico, pero si el orden de las corridas está aleatorizado (y, de ser apropiado, se seleccionan al azar otras unidades y materiales experimentales), este supuesto por lo general se satisfará.
- Los supuestos de la igualdad de varianzas y la normalidad se pueden verificar utilizando una gráfica de probabilidad normal.

Prueba de hipótesis

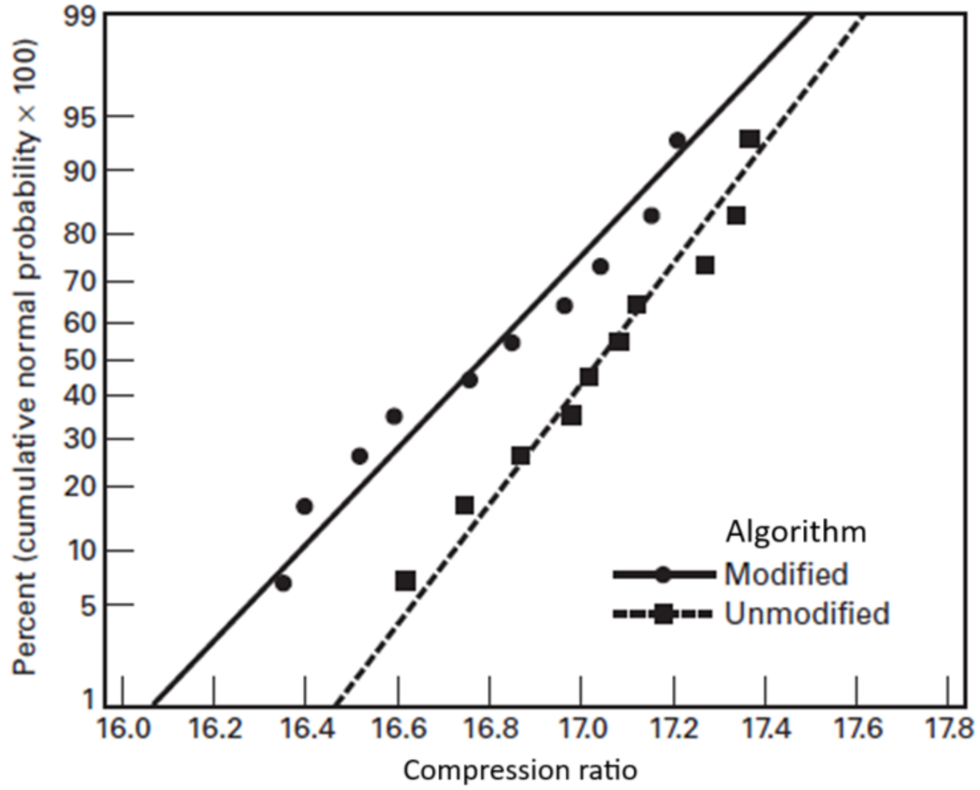
- Para construir una gráfica de probabilidad normal, primero se ordenan de menor a mayor las observaciones de la muestra.
- Luego se grafican las observaciones ordenadas contra sus respectivas frecuencias acumuladas observadas.



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

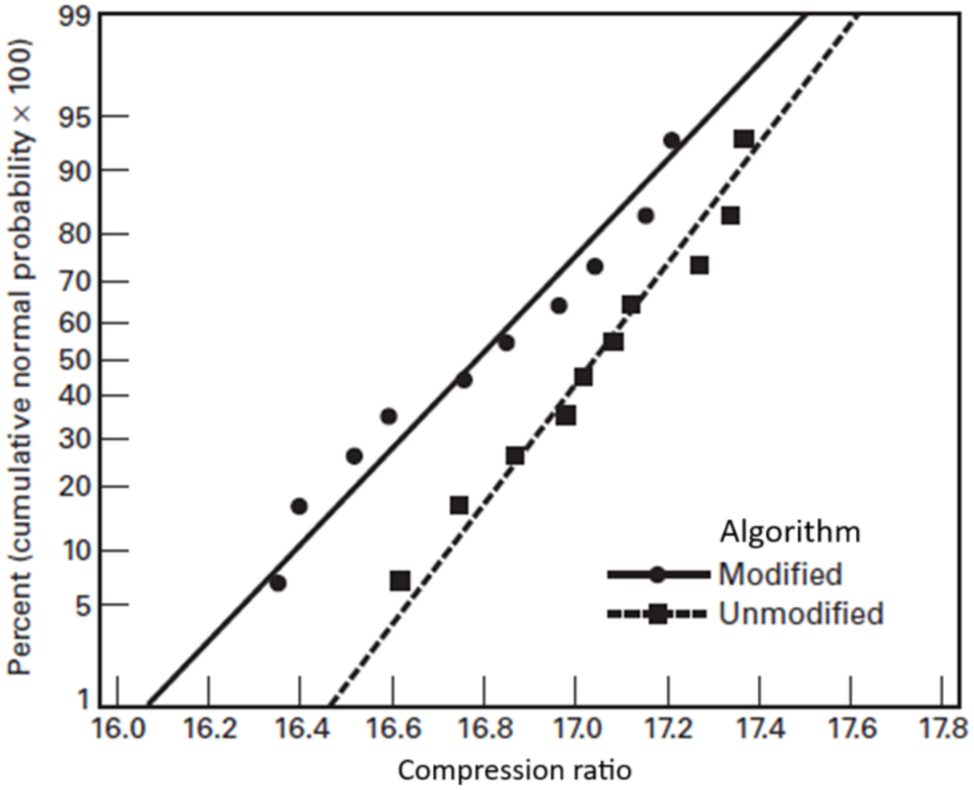
Prueba de hipótesis

- La escala indica que, si la distribución normal describe adecuadamente los datos, los puntos estarán aproximadamente sobre una línea recta.
- Si los puntos se desvían significativamente de una recta, el modelo no es apropiado.
- Determinar si los datos graficados pertenecen o no a una recta puede ser subjetivo.



Prueba de hipótesis

- El supuesto de la igualdad de varianzas puede verificarse comparando las pendientes de las dos rectas.
- En este caso, el supuesto de la igualdad de las varianzas es razonable.
- Si se viola este supuesto, deberá usarse una versión modificada de la prueba *t* .



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Prueba de hipótesis

- Cuando ocurren violaciones importantes de los supuestos, se afectará el desempeño de la prueba *t* , tanto el nivel de significación de la prueba como la capacidad para detectar diferencias entre las medias.
- En general, violaciones pequeñas o moderadas no son motivo de preocupación particular, pero no deberá ignorarse cualquier falla del supuesto de independencia, así como los indicios claros de que no se satisface el supuesto de normalidad.
- Un recurso para resolver este problema son las transformaciones.
- También es posible utilizar procedimientos no paramétricos para la prueba de hipótesis cuando las observaciones provienen de poblaciones no normales.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Intervalos de confianza

- Aun cuando la prueba de hipótesis es un procedimiento útil, en ocasiones no cuenta la historia completa.
- Muchas veces es preferible proporcionar un intervalo dentro del cual cabría esperar que estaría incluido el valor del parámetro o los parámetros en cuestión.
- En muchos experimentos se sabe de antemano que las medias μ_1 y μ_2 difieren, por lo que la prueba de la hipótesis $H_0 : \mu_1 = \mu_2$ es de escaso interés.
- Por lo general el experimentador estaría más interesado en un intervalo de confianza para la diferencia en las medias $\mu_1 - \mu_2$.

Intervalos de confianza

- Para definir un intervalo de confianza, suponga que θ es un parámetro desconocido. Para obtener una estimación del intervalo de θ, es necesario encontrar dos estadísticos *L* y *U* tales que la declaración de probabilidad

$$P(L \leqslant \theta \leqslant U) = 1 - \alpha$$

sea verdadera. Al intervalo

$$L \leqslant \theta \leqslant U$$

se le llama intervalo de confianza de 100(1 - α) por ciento para el parámetro θ .

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Intervalos de confianza

$$P(L \leqslant \theta \leqslant U) = 1 - \alpha$$

- La interpretación de este intervalo es que si, en muestreos aleatorios repetidos se construye un gran número de estos intervalos, $100(1 - \alpha)$ por ciento de ellos contendrán el verdadero valor de θ.
- A los estadísticos L y U se les llama los límites de confianza inferior y superior, respectivamente, y a $(1 - \alpha)$ se le llama el coeficiente de confianza.
- Si $\alpha = 0.05$, a $L \leq \theta \leq U$ se le llama intervalo de confianza de 95% para θ.

Intervalos de confianza

$$P(L \leqslant \theta \leqslant U) = 1 - \alpha$$

- Los intervalos de confianza tienen una interpretación de frecuencia;
- Es decir, no se sabe si la declaración es verdadera para esta muestra específica, pero sí se sabe que el método usado para generar el intervalo produce declaraciones correctas en 100 (1 - α) por ciento de las veces.

Intervalos de confianza

- Suponga que quiere encontrarse un intervalo de confianza de 100(1 - α) por ciento para la verdadera diferencia de las medias $\mu_1 - \mu_2$ en el problema del algoritmo de compresión. El intervalo puede deducirse de la siguiente manera.
- El estadístico

$$\frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Se distribuye como $t_{n1 + n2 - 2}$, por lo tanto:

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Intervalos de confianza

$$\frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$P\left(-t_{\alpha/2,n_1+n_2-2} \leq \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{\alpha/2,n_1+n_2-2}\right) = 1 - \alpha$$

O, lo que es lo mismo

$$P\left(\bar{y}_1 - \bar{y}_2 - t_{\alpha/2,n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + t_{\alpha/2,n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

Que, de la definición del intervalo $P(L \leqslant \theta \leqslant U) = 1 - \alpha$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Intervalos de confianza

$$P(L \leqslant \theta \leqslant U) = 1 - \alpha$$

Se observa que

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{y}_1 - \bar{y}_2 + t_{\alpha/2, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Es un intervalo de confianza de 100(1 - a) por ciento para $\mu_1 - \mu_2$.

Si se toma esta ecuación y se aplica en el ejemplo del algoritmo de compresión:

$$16.76 - 17.04 - (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}} \leq \mu_1 - \mu_2 \leq 16.76 - 17.04 + (2.101)0.284\sqrt{\tfrac{1}{10} + \tfrac{1}{10}}$$

$$-0.28 - 0.27 \leq \mu_1 - \mu_2 \leq -0.28 + 0.27$$

$$-0.55 \leq \mu_1 - \mu_2 \leq -0.01$$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Intervalos de confianza

$$-0.28 - 0.27 \leq \mu_1 - \mu_2 \leq -0.28 + 0.27$$

$$-0.55 \leq \mu_1 - \mu_2 \leq -0.01$$

- Por lo tanto, el intervalo de confianza de 95% estimado para la diferencia en las medias se extiende de -0.55 a -0.01 .
- Expresado en otros términos, el intervalo de confianza es $\mu_1 - \mu_2 = -0.28 \pm 0.27$, o la diferencia en las tasas promedio es -0.28 , y la precisión de esta estimación es de ±0.27 .
- Observe que como $\mu_1 - \mu_2 = 0$ no está incluida en este intervalo, los datos no apoyan la hipótesis de que $\mu_1 = \mu_2$ con el nivel de significación de 5% .
- Es probable que la compresión media de la implementación sin modificar exceda la compresión media de la implementación modificada del algoritmo.

Elección del tamaño de la muestra

- La selección de un tamaño de muestra apropiado es una de las partes más importantes de cualquier problema de diseño experimental.
- Una forma de hacer esta selección es considerar el impacto del tamaño de la muestra en la estimación de la diferencia de dos medias.
- Sabemos que el intervalo de confianza de 100 (1 - α)% sobre la diferencia de dos medias es una medida de la precisión de la estimación de la diferencia de las dos medias.
- El largo de este intervalo está determinado por $t_{\alpha/2,\, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Elección del tamaño de la muestra

- Consideremos el caso donde los tamaños de muestra de las dos poblaciones son iguales, de modo que $n_1 = n_2 = n$.

$$t_{\alpha/2,\, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

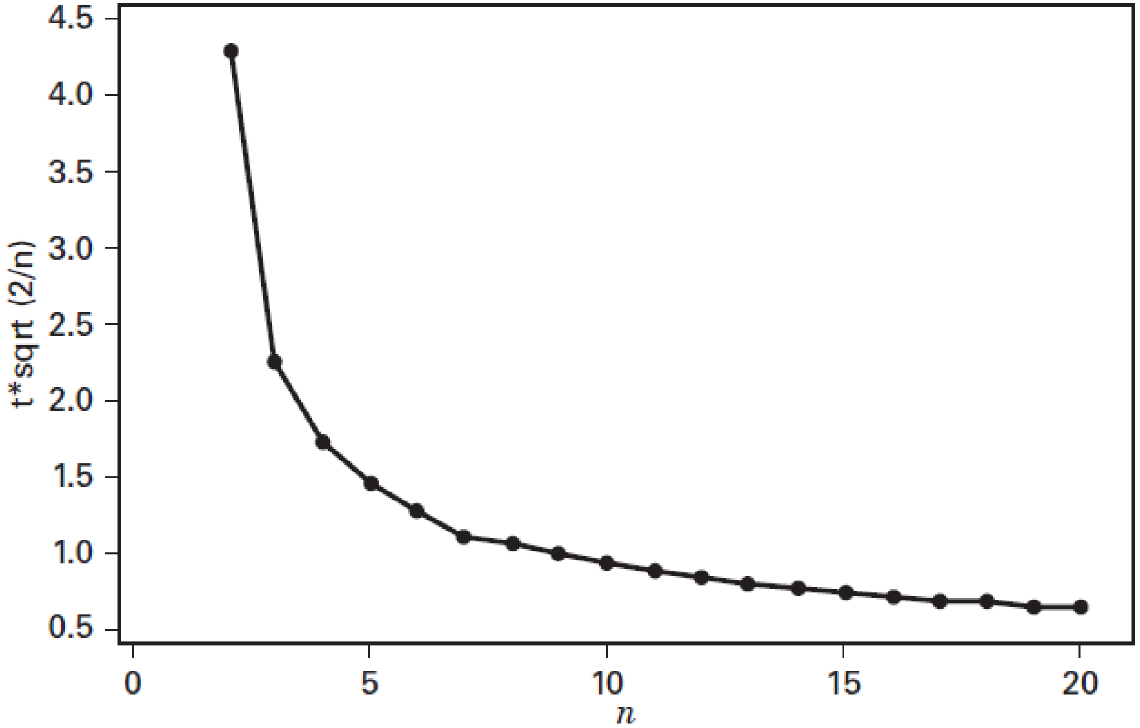
- Entonces la longitud del IC está determinada por:

$$t_{\alpha/2,\, 2n-2}\, S_p \sqrt{\frac{2}{n}}$$

- En consecuencia, la precisión con la que se estima la diferencia entre las dos medias depende de dos cantidades:
  - S_p , sobre la que no tenemos control,
  - $t_{\alpha/2\,,\, 2n-2}\sqrt{(2/n)}$ que podemos controlar eligiendo el tamaño de muestra n.

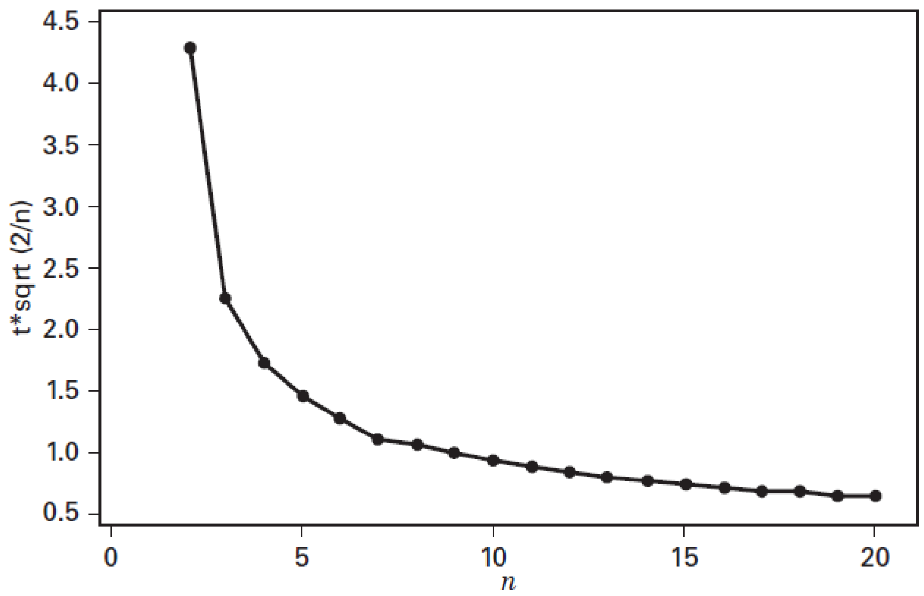
Elección del tamaño de la muestra

- Se pude graficar $t_{\alpha/2\,,\,2n-2}\sqrt{(2/n)}$ versus n para $\alpha = 0.05$.
- Observe que la curva desciende rápidamente a medida que n aumenta hasta alrededor de n = 10 y menos rápidamente más allá de eso.



Plot of $t_{\alpha/2,\,2n-2}\sqrt{2/n}$ versus sample size in each population n for $\alpha = 0.05$

Elección del tamaño de la muestra



Dado que *Sp* es relativamente constante y $t_{\alpha/2\,,\,2n-2}\sqrt{(2/N)}$ no va a cambiar mucho para tamaños de muestra más allá de n = 10 o 12, podemos concluir que, elegir un tamaño de muestra de n = 10 o 12 de cada población en un IC del 95% de dos muestras, dará como resultado un IC con la mejor precisión de estimación para la diferencia en las dos medias que es posible, dada la cantidad de variabilidad inherente presente en las dos poblaciones.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Elección del tamaño de la muestra

- También podemos usar un marco de prueba de hipótesis para determinar el tamaño de la muestra. La elección del tamaño de la muestra y la probabilidad β del error tipo II guardan una estrecha relación.

- Suponga que se están probando las hipótesis:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

- Y suponga que las medias no son iguales, por lo que hay un δ tal que $\delta = \mu_1 - \mu_2$.

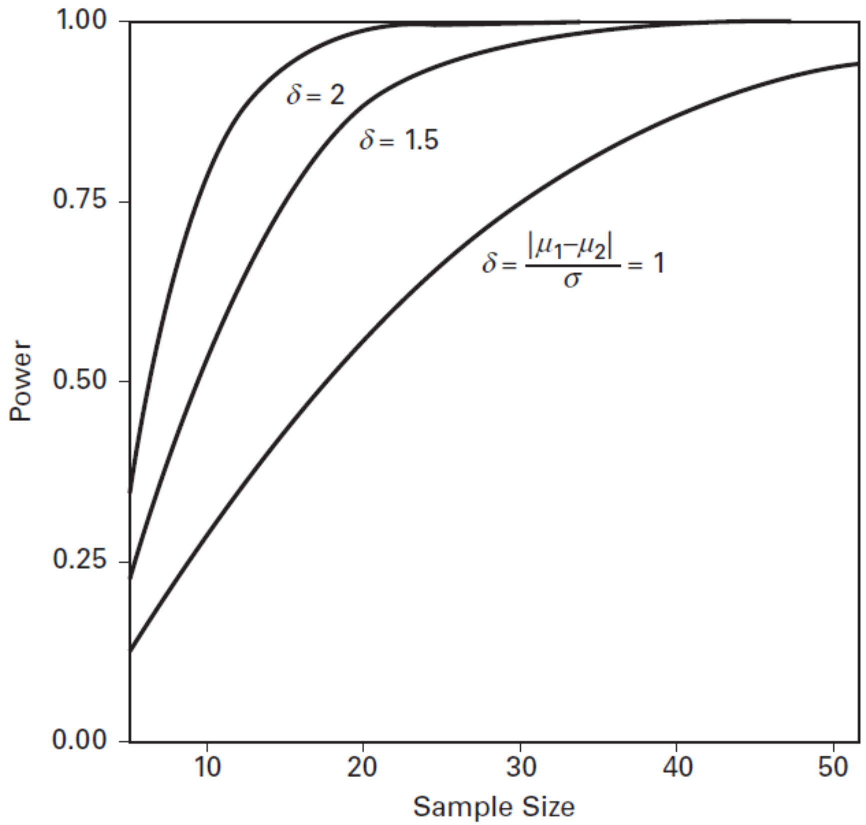
- Puesto que $H_0 : \mu_1 = \mu_2$ no es verdadera, la preocupación principal es cometer la equivocación de no rechazar H_0 .

Elección del tamaño de la muestra

- La probabilidad del error tipo II depende de la verdadera diferencia en las medias δ .
- A una gráfica de β contra δ para un tamaño particular de la muestra se le llama la curva de característica de operación, o curva OC, de la prueba.
- El error β también es una función del tamaño de la muestra. En general, para un valor dado de δ, el error β se reduce cuando el tamaño de la muestra se incrementa.
- Es decir, es más fácil detectar una diferencia especificada en las medias para tamaños grandes de la muestra que para los tamaños pequeños.

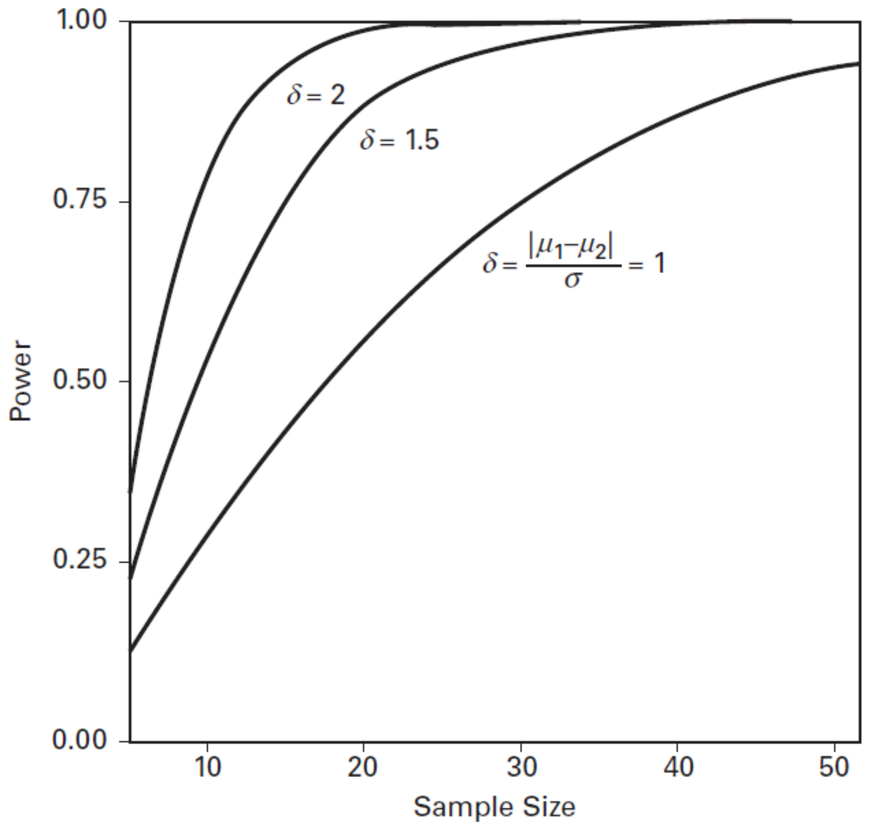
Elección del tamaño de la muestra

- Una alternativa a la curva OC es una curva de potencia, que normalmente traza la potencia (1 - β) frente al tamaño de la muestra para una diferencia específica en las medias.
- Si las dos varianzas poblacionales σ_1^2 y σ_2^2 son desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) y para un nivel de significación α de 0.05 se tiene la siguiente curva de potencia:



Elección del tamaño de la muestra

- Esta curva también supone que los tamaños de muestra de las dos poblaciones son iguales y que el tamaño de muestra de la escala horizontal (n) es el tamaño de muestra total, por lo que el tamaño de muestra en cada población es n/2.
- Observe también que la diferencia de medias se expresa como una razón de la desviación estándar común $$\delta = \frac{|\mu_1 - \mu_2|}{\sigma}$$

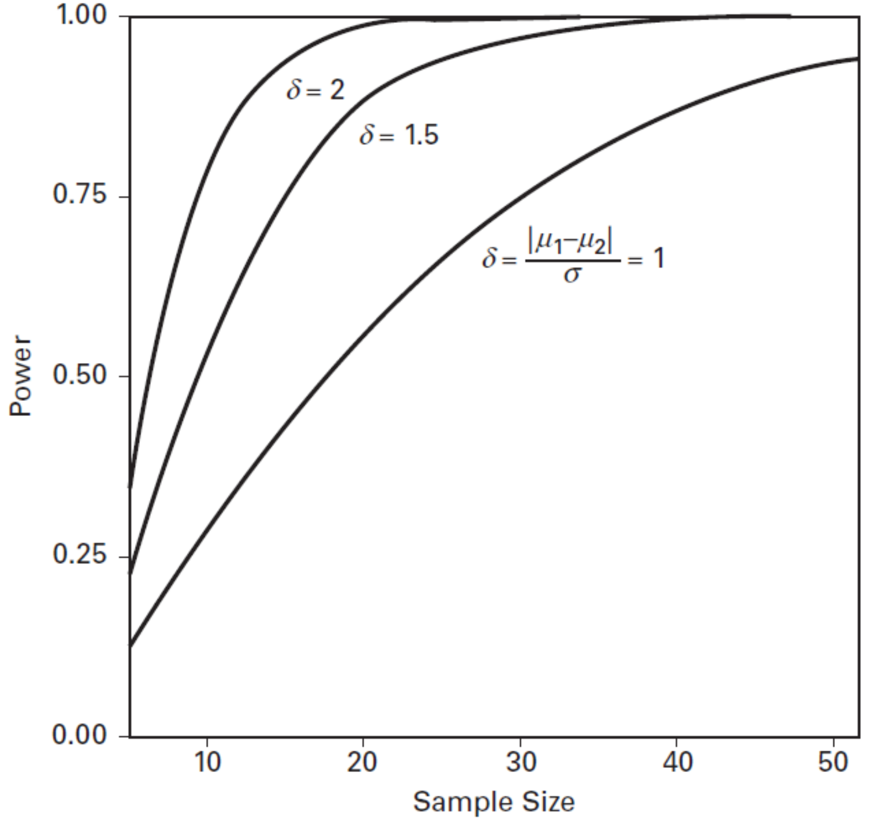


Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Elección del tamaño de la muestra

Al examinar estas curvas observamos lo siguiente:

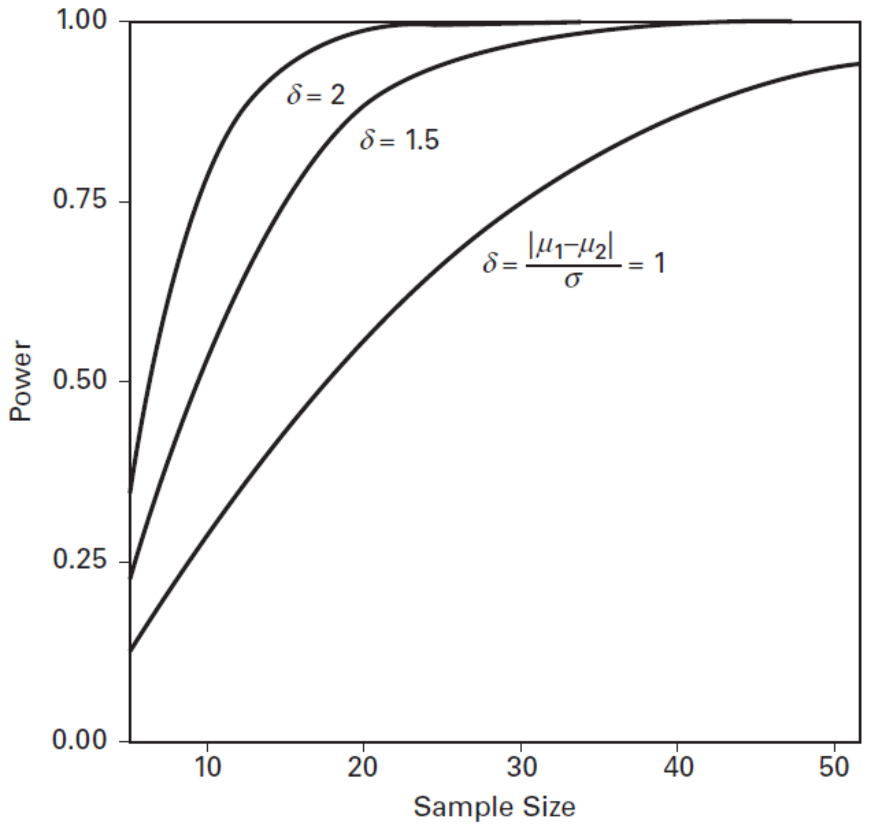
1) A mayor diferencia de medias, mayor potencia (menor probabilidad de error tipo II). Es decir, para un tamaño de muestra y un nivel de significación específicos, la prueba detectará diferencias grandes en las medias más fácilmente que pequeñas diferencias.



Elección del tamaño de la muestra

2) A medida que aumenta el tamaño de la muestra, aumenta la potencia de la prueba (la probabilidad de error de tipo II disminuye) para una diferencia dada en las medias y el nivel de significación.

Es decir, para detectar una diferencia específica en las medias, podemos hacer que la prueba sea más poderosa aumentando el tamaño de la muestra.



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

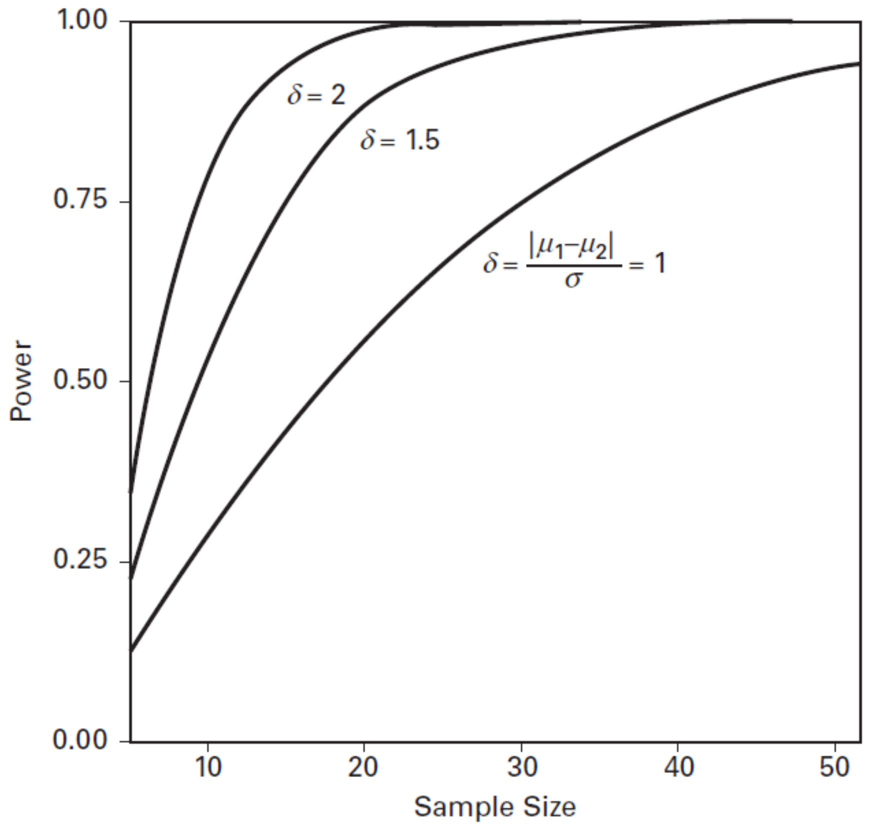
Elección del tamaño de la muestra

- Por ejemplo, para el problema del algoritmo de compresión:
- Suponga que una diferencia en la tasa media de 0,5 tiene un impacto práctico en el uso del algoritmo, por lo que, si la diferencia en las medias es al menos así de grande, nos gustaría detectarla con una alta probabilidad.
- Por lo tanto, debido a que $\mu_1 - \mu_2 = 0.5$ es la diferencia crítica en las medias que deseamos detectar, encontramos que el parámetro de la curva de potencia sería $\delta = 0.5 / \sigma$.
- Desafortunadamente, implica saber la desviación estándar σ (desconocida).

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Elección del tamaño de la muestra

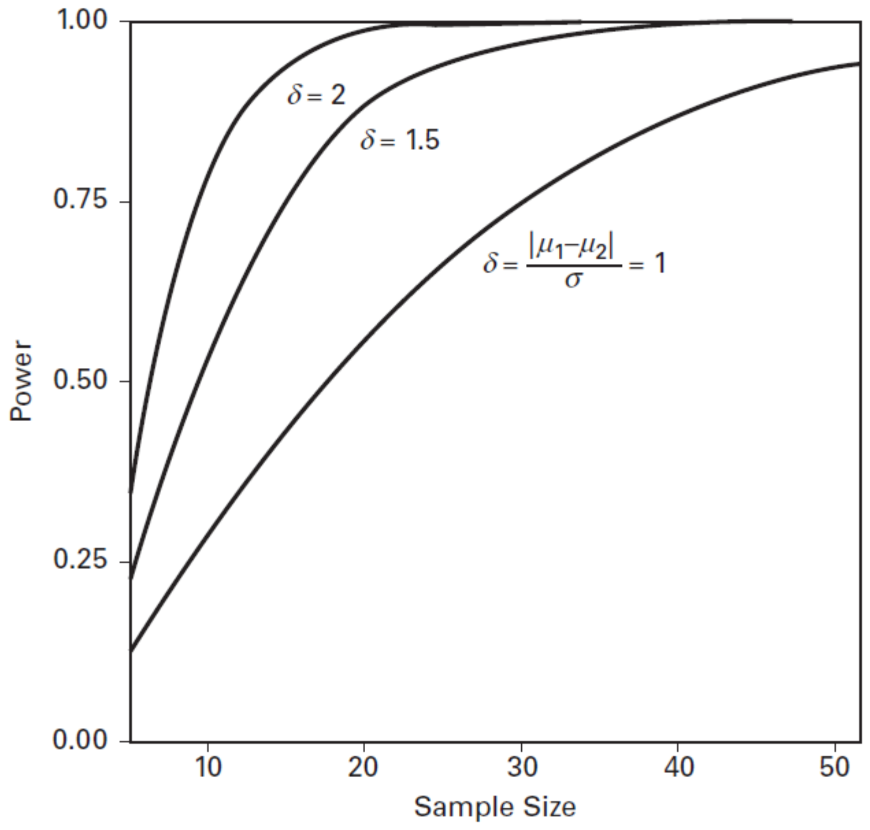
- Sin embargo, supongamos que, basándonos en experiencias pasadas u otras investigaciones, pensamos que es muy poco probable que la desviación estándar supere los 0.25 .
- Luego, sustituyendo $\sigma = 0.25$ en la expresión $\delta = 0.5 / \sigma$ da como resultado $\delta = 2$.



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

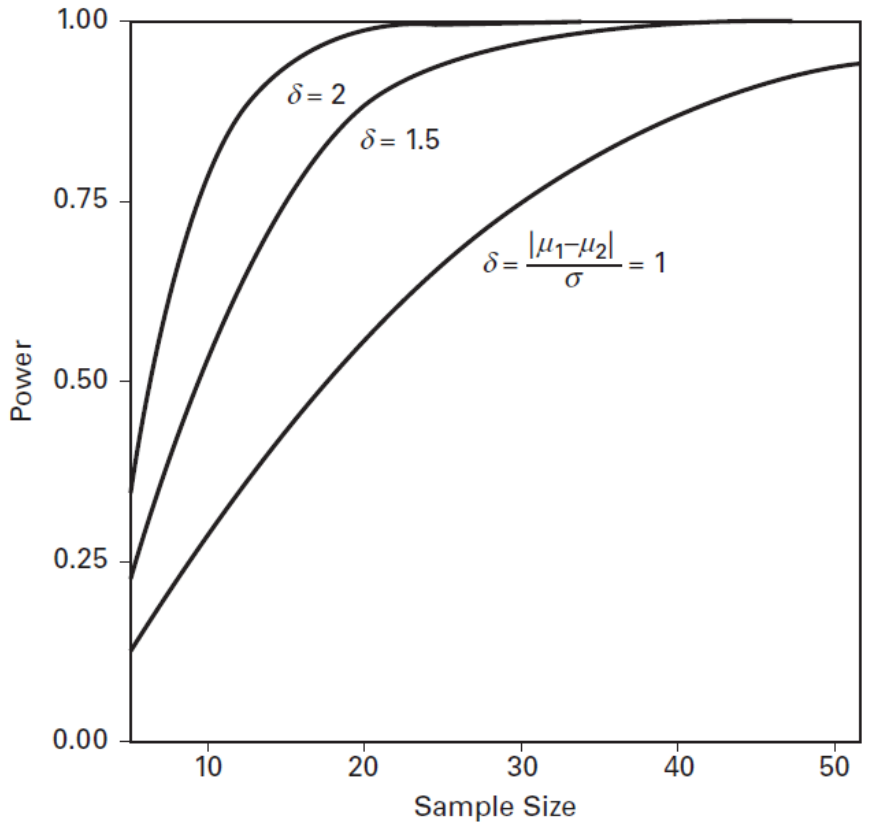
Elección del tamaño de la muestra

- Si deseamos rechazar la hipótesis nula cuando la diferencia de medias $\mu_1 - \mu_2 = 0.5$ con probabilidad de al menos 0.95 (potencia = 0.95) con $\alpha = 0.05$, entonces, al consultar la figura, encontramos que el tamaño de muestra requerido es 16, aproximadamente.
- Este es el tamaño total de n, por lo que cada muestra de cada población debe ser 8.



Elección del tamaño de la muestra

- En el ejemplo, el experimentador en realidad usó un tamaño de muestra de 10.
- El experimentador pudo haber elegido aumentar ligeramente el tamaño de la muestra para protegerse contra la posibilidad de que la estimación anterior de la desviación estándar común σ fuera demasiado conservadora y probablemente fuera un poco más grande que 0,25.



Inferencias acerca de las diferencias en las medias. Diseños aleatorios

El caso cuando $\sigma_1^2 \neq \sigma_2^2$

- Si se está probando: $$H_0: \mu_1 = \mu_2$$ $$H_1: \mu_1 \neq \mu_2$$
- Y no hay bases para suponer que las varianzas σ_1^2 y σ_2^2 son iguales (pero sí se cuenta con normalidad), entonces es necesario hacer ligeras modificaciones en la prueba *t* de dos muestras.
- En este caso el estadístico de prueba es: $$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

El caso cuando $\sigma_1^2 \neq \sigma_2^2$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

- Este estadístico no se distribuye exactamente como *t*. No obstante, *t* es una buena aproximación de la distribución de t_0 si para los grados de libertad se usa:

$$v = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$$

- A partir de este estadístico se podrá encontrar un intervalo de confianza para la diferencia en las medias en el caso de varianzas desiguales.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

El caso cuando $\boldsymbol{\sigma}_1^2$ y $\boldsymbol{\sigma}_2^2$ son conocidas

- Si las varianzas poblacionales son conocidas, entonces las hipótesis: $$H_0: \mu_1 = \mu_2$$ $$H_1: \mu_1 \neq \mu_2$$
- Pueden probarse con el estadístico: $$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$
- Si ambas poblaciones son normales, o si los tamaños de las muestras son lo suficientemente grandes para aplicar el teorema del límite central, la distribución de Z_0 es N(0, 1) si la hipótesis nula es verdadera.

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

El caso cuando $\boldsymbol{\sigma}_1^2$ y $\boldsymbol{\sigma}_2^2$ son conocidas

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

- Por lo tanto, la región crítica se encontraría utilizando la distribución normal en lugar de la distribución *t*.
- Específicamente, H_0 se rechazaría si $|Z_0| > Z_{\alpha/2}$ donde $Z_{\alpha/2}$ es el punto porcentual $\alpha/2$ superior de la distribución normal estándar.
- A diferencia de la prueba *t* anterior, en la prueba de las medias con varianzas conocidas no se requiere que el muestreo se haga de poblaciones normales.
- Puede aplicarse el teorema del límite central para justificar una distribución normal aproximada para la diferencia en las medias muestrales $\bar{y}_1 - \bar{y}_2$.

Comparación de una sola media con un valor especifico

- Algunos experimentos incluyen la comparación de la media μ de una sola población con un valor especificado, por ejemplo μ_0.
- Las hipótesis son $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$
- Si la población es normal con varianza conocida, o si la población no es normal pero el tamaño de la muestra es lo suficientemente grande para aplicar el teorema del límite central, entonces la hipótesis puede probarse utilizando una aplicación directa de la distribución normal.

Comparación de una sola media con un valor especifico

- El estadístico de prueba (prueba Z de una muestra) se define: $$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$
- Si $H_0 : \mu = \mu_0$ es cierta, la distribución de Z_0 es N(0, 1). Por lo tanto H_0 se rechazaría si $|Z_0| > Z_{\alpha/2}$.
- El intervalo de confianza de 100(1 - α) por ciento para la verdadera media poblacional es: $$\bar{y} - Z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{y} + Z_{\alpha/2}\sigma/\sqrt{n}$$

Comparación de una sola media con un valor especifico

- El valor de la media μ_0 especificado en la hipótesis nula puede ser resultado de:
  - Evidencia, conocimientos o experimentación previos.
  - Alguna teoría o modelo que describe la situación bajo estudio.
  - Especificaciones contractuales.

Comparación de una sola media con un valor especifico

- Ejemplo: Un proveedor ofrece un nuevo tipo de servidor a un centro de datos que busca mejorar la capacidad de procesamiento de sus servicios.
- El director del centro de datos desea saber si la capacidad de procesamiento promedio del servidor, medida en millones de instrucciones por segundo (MIPS), excede los 200 MIPS.
- Si es así, el centro de datos considerará adoptar este nuevo tipo de servidor.

Comparación de una sola media con un valor especifico

- La experiencia pasada con servidores similares indica que un valor razonable para la varianza de la capacidad de procesamiento es 100(MIPS)^2 .
- Las hipótesis son:

$H_0: \mu = 200$ (la capacidad de procesamiento promedio del servidor es 200 MIPS)

$H_1: \mu > 200$ (la capacidad de procesamiento promedio del servidor excede los 200 MIPS)

Comparación de una sola media con un valor especifico

- Las hipótesis son: $H_0: \mu = 200$, $H_1: \mu > 200$
- Observe que se trata de una hipótesis alternativa de una cola.
- Por lo tanto, el tipo de servidor se aceptaría sólo si la hipótesis nula $H_0: \mu = 200$ pudiera rechazarse, es decir, si $Z_0 > Z_\alpha$.
- Se seleccionan cuatro servidores aleatoriamente para realizar el experimento, y el valor promedio observado es $\bar{y} = 214$ MIPS.
- El valor del estadístico de prueba es:

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{214 - 200}{10/\sqrt{4}} = 2.80$$

Comparación de una sola media con un valor especifico

$$Z_0 = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{214 - 200}{10/\sqrt{4}} = 2.80$$

- Las hipótesis son: $H_0: \mu = 200$, $H_1: \mu > 200$
- Si se especifica un error tipo 1 de $\alpha = 0.05$, se identifica que $Z_\alpha = Z_{0.05} = 1.645$ (mediante tablas de distribución o programas de software).
- Por lo tanto, H_0 se rechaza y se concluye que la capacidad de procesamiento excede los 200 MIPS.

Comparación de una sola media con un valor especifico

- Cuando no se conoce la varianza poblacional, es necesario establecer el supuesto adicional de que la población sigue una distribución normal, aunque las desviaciones moderadas de la normalidad no afectarán seriamente los resultados.
- Para probar $H_0: \mu = \mu_0$ en el caso de varianza desconocida se usa la varianza muestral S^2 para estimar σ^2 .
- El estadístico resultante será una prueba *t* de una muestra: $t_0 = \dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$

Inferencias acerca de las diferencias en las medias. Diseños aleatorios

Comparación de una sola media con un valor especifico

- La hipótesis nula $H_0 : \mu = \mu_0$ se rechazaría si $| t_0 | > t_{\alpha/2, n-1}$, donde $t_{\alpha/2, n-1}$ denota el punto porcentual $\alpha/2$ superior de la distribución t con $n_1 - 1$ grados de libertad.

$$t_0 = \frac{\bar{y} - \mu_0}{S/\sqrt{n}}$$

- El intervalo de confianza de $100(1 - \alpha)$ por ciento es en este caso:

$$\bar{y} - t_{\alpha/2,n-1}S/\sqrt{n} \leq \mu \leq \bar{y} + t_{\alpha/2,n-1}S/\sqrt{n}$$

Inferencias acerca de las diferencias en las medias

Diseños de comparaciones pareadas

- En algunos experimentos comparativos simples puede conseguirse un mejoramiento significativo de la precisión haciendo comparaciones de observaciones pareadas (o emparejadas) del material experimental.

El problema de las comparaciones pareadas

- Considere una aplicación que ejecuta un conjunto de tareas computacionales intensivas.
- Se cuenta con dos algoritmos diferentes para realizar la misma tarea, y se desea determinar si uno de ellos tiene un desempeño significativamente mejor.
- Se usará una puntuación de rendimiento normalizada o "benchmark score“, que asigna valores enteros entre 1 y 10, que representan una escala de puntuación derivada de una fórmula que incluye los factores: tiempo de ejecución, uso de memoria y eficiencia general.

El problema de las comparaciones pareadas

- Sería posible realizar un experimento de la siguiente manera:
- Se seleccionan al azar 20 tareas computacionales similares. La mitad de las tareas se asignan aleatoriamente al Algoritmo 1 y la otra mitad al Algoritmo 2.
- La asignación exacta de las tareas se determinaría de manera aleatoria (diseño completamente aleatorizado)
- Luego la puntuación de rendimiento promedio de las dos muestras podría compararse utilizando la prueba *t* de dos muestras.

El problema de las comparaciones pareadas

- Ahora, si las tareas tienen características distintas (por ejemplo, diferentes niveles de complejidad o estructura de datos que afectan el tiempo de ejecución), esto podría introducir variabilidad adicional en las mediciones y dificultar la detección de una diferencia real entre los algoritmos.
- Esta falta de homogeneidad entre las tareas contribuirá a la variabilidad de las mediciones de los valores y el cálculo de la puntuación de rendimiento y tenderá a inflar el error experimental, haciendo más difícil detectar una diferencia real entre los algoritmos.

El problema de las comparaciones pareadas

- Considere un diseño experimental alternativo:
- Cada tarea se ejecuta dos veces, una con cada algoritmo.
- Se asigna aleatoriamente cuál algoritmo se ejecuta primero para cada tarea.
- Esto crea pares de puntuaciones de rendimiento (uno para cada algoritmo) para cada tarea, lo que permite comparar directamente el desempeño de los algoritmos en condiciones similares.

El problema de las comparaciones pareadas

- El experimento, cuando se llevó a cabo de acuerdo con este diseño con 10 tareas, produjo los datos que se muestran en la tabla:

Task	Algorithm 1	Algorithm 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

El problema de las comparaciones pareadas

- Un modelo estadístico que describe los datos de este experimento puede expresarse como:

$$y_{ij} = \mu_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, 10 \end{cases}$$

  - y_{ij} es la observación de la puntuación para el algoritmo *i* en la tarea *j*
  - μ_i es la verdadera puntuación promedio del algoritmo i-ésimo
  - β_j es un efecto sobre la puntuación debido a la tarea j-ésima
  - ε_{ij} es el error experimental aleatorio con media cero y varianza σ_i^2
    - σ_1^2 es la varianza de las puntuaciones hechas con el algoritmo 1
    - σ_2^2 es la varianza de las puntuaciones hechas con el algoritmo 2

El problema de las comparaciones pareadas

$$y_{ij} = \mu_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, 10 \end{cases}$$

- Observe que si calcula la j-ésima diferencia pareada: $d_j = y_{1j} - y_{2j} \qquad j = 1, 2, \ldots, 10$

  el valor esperado de esta diferencia es:

$$\begin{aligned} \mu_d &= E(d_j) \\ &= E(y_{1j} - y_{2j}) \\ &= E(y_{1j}) - E(y_{2j}) \\ &= \mu_1 + \beta_j - (\mu_2 + \beta_j) \\ &= \mu_1 - \mu_2 \end{aligned}$$

- Es decir, pueden hacerse inferencias acerca de la diferencia en las puntuaciones promedio de los dos algoritmos $\mu_1 - \mu_2$ haciendo inferencias acerca de la media de las diferencias μ_d.

El problema de las comparaciones pareadas

$$y_{ij} = \mu_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \ldots, 10 \end{cases}$$

- Observe que el valor aditivo de las β_j de las tareas se cancela cuando las observaciones están pareadas de esta manera.
- Luego, probar $H_0 : \mu_1 = \mu_2$ es equivalente a probar:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

- Esto es una prueba *t* de una muestra.

El problema de las comparaciones pareadas

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

- El estadístico de prueba para esta hipótesis es: $t_0 = \dfrac{\bar{d}}{S_d/\sqrt{n}}$

Donde: $\bar{d} = \dfrac{1}{n}\sum_{j=1}^{n} d_j$ es la media muestral de las diferencias, y

$$S_d = \left[\frac{\sum_{j=1}^{n}(d_j - \bar{d})^2}{n-1}\right]^{1/2} = \left[\frac{\sum_{j=1}^{n} d_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} d_j\right)^2}{n-1}\right]^{1/2}$$

es la desviación estándar muestral de las diferencias.

El problema de las comparaciones pareadas

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

- $H_0 : \mu_d = 0$ se rechazaría si $| t_0 | > t_{\alpha/2,\, n-1}$
- Dado que las observaciones de los niveles del factor están pareadas en cada unidad experimental, a este procedimiento suele llamársele prueba *t* pareada.

El problema de las comparaciones pareadas

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

- De la tabla original se pueden calcular las diferencias:

Task	Algorithm 1	Algorithm 2
1	7	6
2	3	3
3	3	5
4	4	3
5	8	8
6	3	2
7	2	4
8	9	9
9	5	4
10	4	5

$$d_1 = 7 - 6 = 1$$
$$d_2 = 3 - 3 = 0$$
$$d_3 = 3 - 5 = -2$$
$$d_4 = 4 - 3 = 1$$
$$d_5 = 8 - 8 = 0$$
$$d_6 = 3 - 2 = 1$$
$$d_7 = 2 - 4 = -2$$
$$d_8 = 9 - 9 = 0$$
$$d_9 = 5 - 4 = 1$$
$$d_{10} = 4 - 5 = -1$$

El problema de las comparaciones pareadas

- Por lo tanto:

$$\bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j = \frac{1}{10}(-1) = -0.10$$

$$S_d = \left[\frac{\sum_{j=1}^{n} d_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} d_j\right)^2}{n-1}\right]^{1/2} = \left[\frac{13 - \frac{1}{10}(-1)^2}{10-1}\right]^{1/2} = 1.20$$

- Suponga que se había elegido α = 0.05. Entonces, para tomar una decisión se calcularía t_0 y H_0 se rechazaría si $|t_0| > t_{0.025,9} = 2.262$.

$$\begin{aligned}
d_1 &= 7 - 6 &&= 1\\
d_2 &= 3 - 3 &&= 0\\
d_3 &= 3 - 5 &&= -2\\
d_4 &= 4 - 3 &&= 1\\
d_5 &= 8 - 8 &&= 0\\
d_6 &= 3 - 2 &&= 1\\
d_7 &= 2 - 4 &&= -2\\
d_8 &= 9 - 9 &&= 0\\
d_9 &= 5 - 4 &&= 1\\
d_{10} &= 4 - 5 &&= -1
\end{aligned}$$

El problema de las comparaciones pareadas

- H_0 se rechazaría si $| t_0 | > t_{0.025\,,\,9} = 2.262$.
- El valor calculado del estadístico de prueba *t* pareada es:

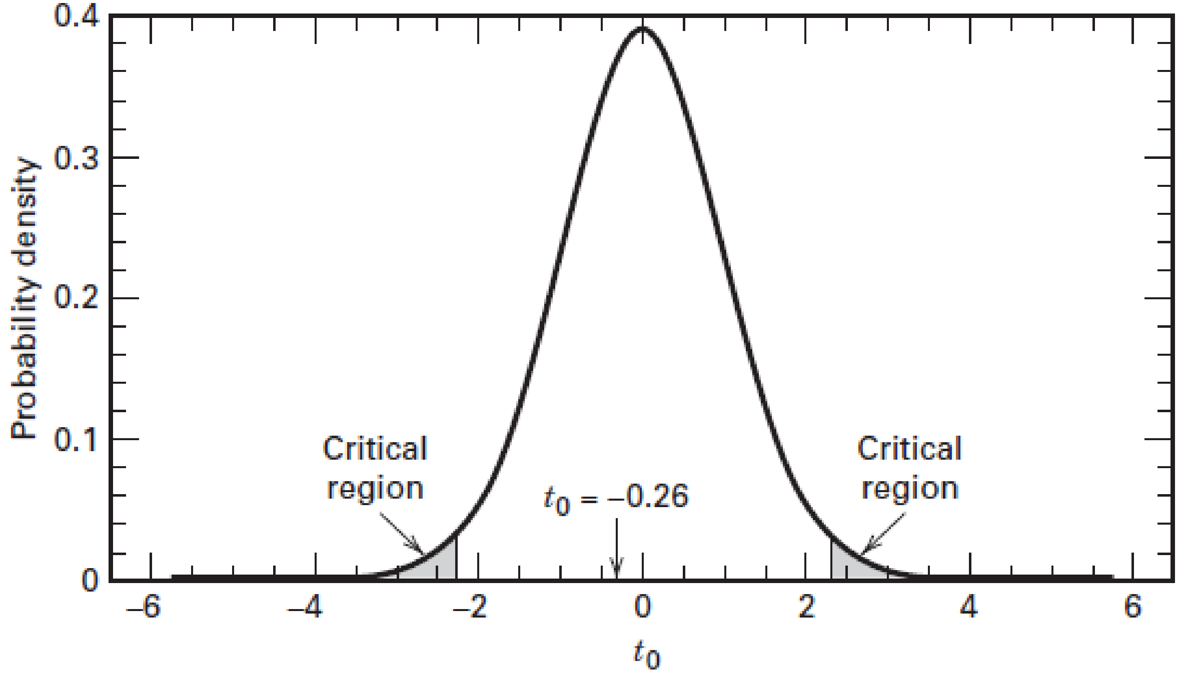
$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{-0.10}{1.20/\sqrt{10}} = -0.26$$

- Como si $| t_0 | = 0.26$ no es mayor que $t_{0.025\,,\,9} = 2.262$ H_0 no puede rechazarse.
- Entonces, no hay diferencia significativa entre el desempeño de ambos algoritmos.

Inferencias acerca de las diferencias en las medias. Diseños de comparaciones pareadas

El problema de las comparaciones pareadas

- Como si $|\, t_0 \,| = 0.26$ no es mayor que $t_{0.025\,,\,9} = 2.262$ H_0 no puede rechazarse.
- Comparaciones pareadas permitieron comparar las puntuaciones de rendimiento de los algoritmos, minimizando el impacto de la variabilidad entre las tareas.



Referencias

- Montgomery, D.C. (2013). Design of Experiments. John Wiley & sons.
- Wickman, H., Grolemund, G. (2017). *R for Data Science. Visualize, model, transform, tidy, and import data.* O’Reilly.
- Lawson, J. (2014). *Design and Analysis of Experiments with R* (Vol. 115). CRC press.