DISEÑO DE EXPERIMENTOS

INTRODUCCION

Licenciatura en Estadística
Profesores:
Dr. José Alberto Pagura
Lic. Julia Inés Fernández

INTRODUCCION

Experimentación

- En muchas áreas del conocimiento es posible emplear la experimentación para encontrar o detectar relaciones de tipo causal entre variables de un sistema o de un proceso.

- Esta clase de estudios presupone la posibilidad de controlar un conjunto de factores que se piensa, son causales de los resultados.

- **Experimento**: prueba o serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema para observar e identificar las razones de los cambios que pudieran observarse en la respuesta de salida.

Ejemplos

- Se desea estudiar si el rendimiento de una determinada variedad de maíz es afectada por diferentes abonos y tipos de suelo.

- Se piensa que diferentes algoritmos computacionales para la resolución de un determinado tipo de problema, dan el mismo resultado pero en diferentes tiempos de proceso. Se desea determinar cual de ellos es más "rápido".

- Se propone el uso de un nuevo programa de entrenamiento para la realización de determinada tarea asumiendo que es mejor que el que se emplea actualmente.

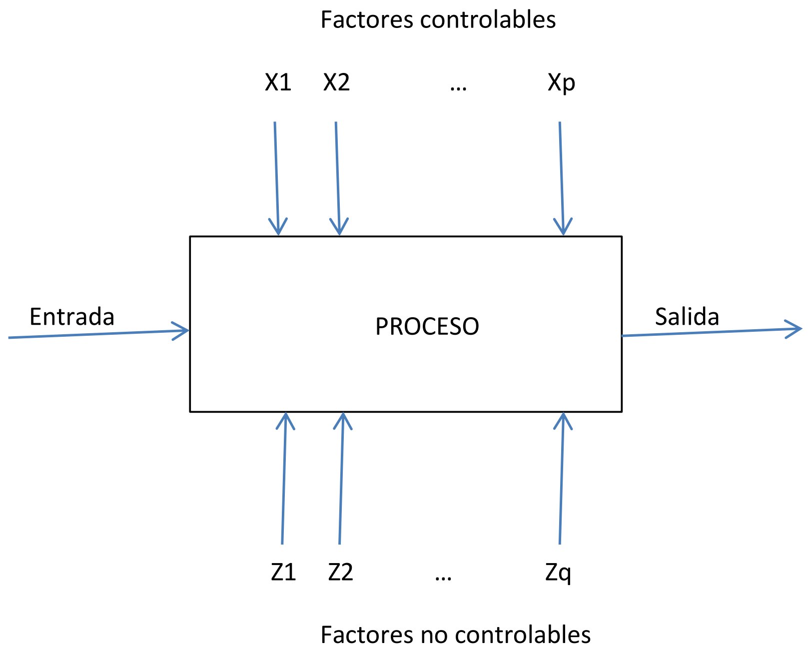
Ejemplos

- Las variables consideradas causales, ¿son las únicas con efecto sobre la variable que se considera afectada?

- Para resolver esta clase de problemas debe determinarse si las variables que se consideran influyentes efectivamente lo son, y esto en presencia de otras variables que pueden estar afectando el comportamiento del sistema y que no son tenidas en cuenta en el estudio

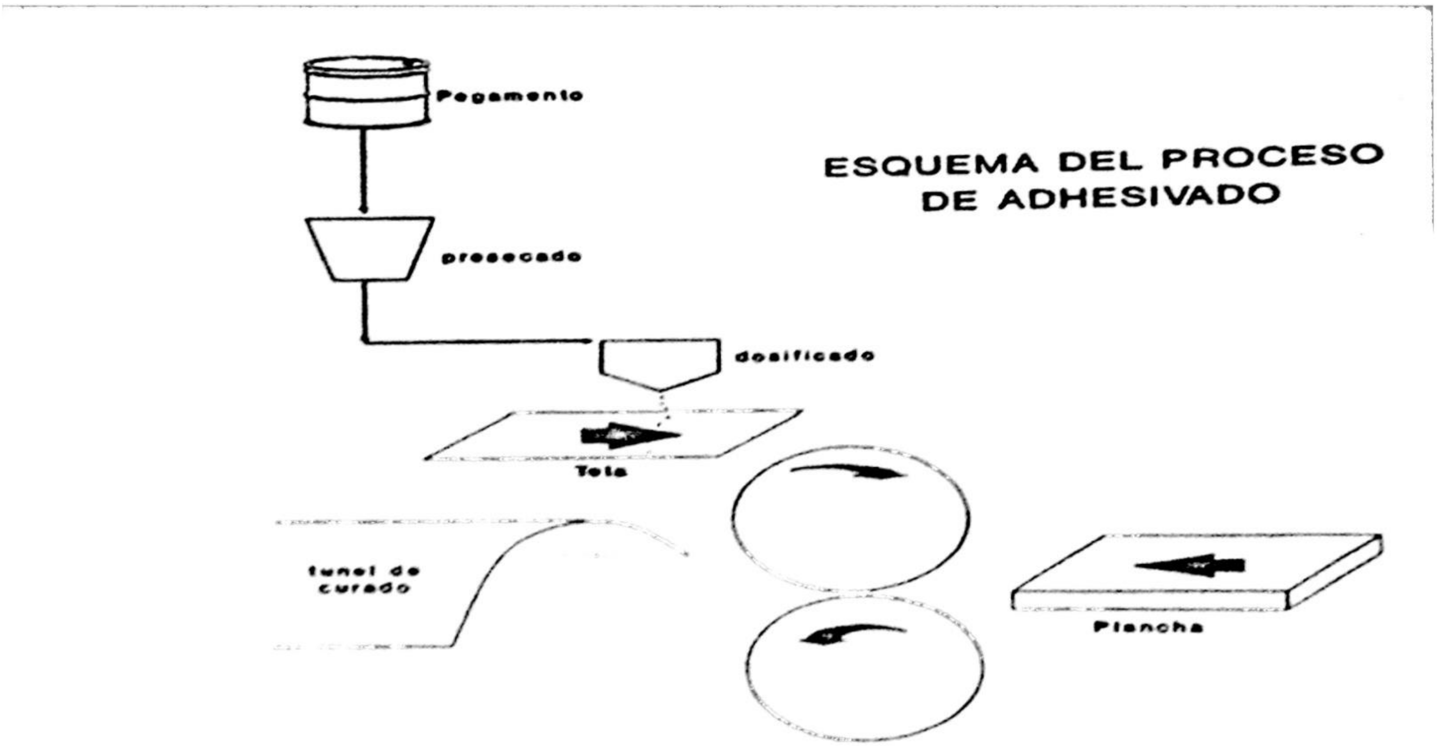
El proceso o sistema a estudiar

- Se desea estudiar el comportamiento de una variable “Y” que podría depender de los valores de variables X_1, X_2,…, X_p, controlables, y de Z_1, Z_2,…, Z_q no controlables.



Ejemplos

- Se desea mejorar la fuerza de adhesión obtenida en un ***proceso de adhesivado*** de planchas de poliuretano, (Romero, 2004)



Fuentes comunes de dificultad en un estudio experimental

Error experimental

- Siempre hay variaciones en las mediciones resultantes, aunque se hagan bajo iguales condiciones
- Tales variaciones son producidas por factores distorsionantes (conocidos y desconocidos) y por errores de medición (seguramente una pequeña parte se podrá atribuir al error de medición).
- La variabilidad de una respuesta, no atribuible a influencias conocidas se denomina error experimental.
- Los efectos perniciosos del error experimental pueden reducirse mediante un buen diseño. Si no se logra su reducción, es posible que efectos importantes queden ocultos o bien concluir que efectos no relevantes, lo son.

Error experimental

- La variación natural entre unidades experimentales
- La variabilidad en la medición de la respuesta
- La imposibilidad de reproducir las condiciones del tratamiento, con exactitud, de una unidad a otra
- La interacción de los tratamientos con las unidades experimentales
- Cualquier otro factor externo que influya en las características medidas

Confusión por factores no medidos

- Cualquier fenómeno que se estudie estará influenciado por un cantidad de factores, muchos de los cuales pueden pasar desapercibidos para el investigador
- Un riesgo latente es el descubrir relaciones entre variables, que realmente estén causadas por una tercera que influye sobre ambas
- Se pueden diseñar experimentos que con pocas pruebas sean capaces de evaluar la influencia de una gran cantidad de posibles factores influyentes

Complejidad de los efectos

- Supongamos estudiar los efectos del café y el alcohol sobre los tiempos de respuesta de conductores (en simuladores)
- Se encuentra que: si no se había tomado café, una copa de licor aumenta el tiempo medio de respuesta en 0.45" y si no se tomó alcohol, una taza de café lo reduce en 0.20"
- Si se quiere evaluar el efecto de varias tazas de café y tragos de licor, sería fácil si ellos son lineales y aditivos
- Pero es probable que no tengan efecto lineal y que interactúen
- Hay diseños "simples" que permiten con pocas pruebas estimar todos esos efectos

Recursos estadísticos

1. Métodos eficientes de ***diseño de experimentos*** que permitan obtener respuestas que sean lo menos ambiguas y lo menos afectadas posible por lo errores experimentales

2. Métodos de ***análisis de datos***, que sean sensibles para indicar lo que se puede deducir de la hipótesis en vigor y den pie a nuevas ideas a considerar

Importancia de tales recursos

- De los dos recursos enunciados, es crucial el diseño del experimento.
- Si está mal elegido, de modo que los datos resultantes no contengan mucha información, poco se podrá extraer, por muy detallado o sofisticado que sea el análisis
- Si el diseño se escoge adecuadamente, habrá disponible una gran cantidad de INFORMACIÓN y análisis simples podrían ser suficientes

El papel del diseño experimental

- El diseño experimental consiste en definir cómo y bajo que condiciones, habrán de reunirse los datos que se contrastarán con las hipótesis
- Cada diseño contiene un **grupo de experimentos o ensayos a realizar**
- La elección del diseño siempre depende de las hipótesis en vigor
- El diseño elegido debe investigar las zonas oscuras que tiene el conocimiento actual del problema y cuya iluminación se considere un avance importante

Definición

- Diseñar un experimento significa manejar el material experimental y realizar la adjudicación de los tratamientos, de modo tal que los resultados obtenidos permitan que las hipótesis de interés sean juzgadas con la mayor potencia posible y los parámetros que describen aspectos relevantes de la investigación se estimen con la mayor precisión, dentro de los límites de un presupuesto limitado.

  William Cochran

Fisher y los principios de la experimentación

- Ronald Fisher, entre 1919-1933 desempeñó tareas en la Rothamsted Experimental Station. En ese período desarrolló y consolidó los principios básicos del diseño de experimentos y el análisis de los datos obtenidos con esas estrategias, prácticas necesarias para llegar a resultados válidos y enunció tres principios que deben tenerse en cuenta en la realización de cualquier estudio experimental:

ALEATORIZACION

REPLICACION

CONTROL LOCAL

Principios de la experimentación

- **Aleatorización.** En un experimento la adjudicación del material experimental y el orden en que se realizan las pruebas, se hace en forma aleatoria. Esto garantiza lograr observaciones independientes y ayuda a cancelar los efectos de factores extraños que pueden estar presentes.
- **Replicación.** Permite estimar el error experimental. Esto hace posible determinar si las diferencias observadas son estadísticamente significativas. Además si se consigue un error "pequeño" se pueden calcular estimaciones precisas del efecto de los factores del experimento.
- **Control local.** Se utiliza para incrementar la precisión del experimento. Se procura que los tratamientos se apliquen sobre material experimental homogéneo.

Terminología básica

- **EXPERIMENTO:** conjunto de pruebas cuyo objetivo es obtener información que permita avanzar en la comprensión del comportamiento del sistema. Por ejemplo, en diseños industriales probablemente se busque condiciones que conducen a la mejora de productos o de procesos.
- **RESPUESTAS:** aquellas características de interés objeto del estudio (tiempo necesario hasta lograr una reacción, dureza de una superficie, consumo de energía de un proceso) . En un experimento puede haber más de una respuesta de interés.

Terminología básica

- **FACTORES CONTROLADOS**: aquellas características del producto o proceso para los cuales se ensayarán distintas variantes o alternativas con el fin de estudiar como influyen sobre las respuestas de interés. Pueden ser de tipo *cualitativo* (proveedor, variedad del material de un electrodo, agregado o no de un aditivo, etc.) o *cuantitativo* (temperatura, velocidad de giro, dosis de aditivo, viscosidad, etc.)

Terminología básica

- **Variantes o Niveles**: nombre de las alternativas que se prueban o ensayan de los factores cualitativos o cuantitativos, respectivamente.
- Si los niveles o variantes del factor son los únicos de interés, se dirá que los efectos a evaluar son fijos
- Si los niveles o variantes a ensayar constituyen una muestra de los que son posibles para el factor, se dirá que los efectos son aleatorios.

Terminología básica

- **TRATAMIENTO**: combinación de variantes y/o niveles de cada factor, que se utiliza en una determinada prueba.
- **REPLICAS**: cantidad de ensayos que se realizan con el mismo tratamiento. Generalmente se procura que los experimentos sean *balanceados*, es decir, que se hagan igual cantidad de réplicas de cada tratamiento.

Pasos para elaborar el diseño de un experimento

- La actividad de diseñar un experimento será realizada preferiblemente por un equipo compuesto por personas de las diferentes áreas afectadas por el tema. Las etapas a cumplir serán:
- Definir claramente los objetivos.
- Definir en forma operativa las características sobre las que se quiere investigar el posible efecto de ciertos factores.
- Seleccionar los factores a incluir en el experimento.

Otras etapas

- Seleccionar las variantes o niveles a ensayar por cada factor.
- Definir en que consistirá cada prueba.
- Decidir el número de pruebas a realizar y el tratamiento a aplicar en cada una de ellas.
- Organizar el trabajo experimental, asignando las responsabilidades y precisando las necesidades de tiempos y de medios.

Análisis de los resultados

- Una vez realizado el experimento y obtenido los resultados de las pruebas, el equipo deberá proceder al análisis de los resultados y obtención de las conclusiones que se derivan del mismo.
- El análisis, en general, implica obtener respuestas a las siguientes preguntas:
- ¿Qué factores tienen un efecto significativo sobre la la respuesta?
- ¿Cuál es la naturaleza de la relación entre los efectos significativos encontrados y la variable respuesta? ¿Concuerdan estas conclusiones con los conocimiento técnicos previos?

Análisis de los resultados

- ¿Cuáles serían los niveles o variantes óptimos para los diferentes factores en función de sus efectos sobre las medias de las respuestas?
- ¿Qué respuesta media cabe predecir trabajando en las condiciones óptimas encontradas?
- ¿Hay efectos significativos sobre la variancia de la respuesta? ¿Cuál es su naturaleza? ¿Cuál es la variancia previsible de la respuesta en función de las condiciones operativas utilizadas?

Análisis de los resultados

- Finalmente: ¿Cuál es la condición de “trabajo” óptima desde el punto de vista técnico y económico?
- ¿Qué media y que variancia caben predecir para las respuestas trabajando en dichas condiciones?
- ¿Es aconsejable algún experimento complementario para aclarar cuestiones que no han quedado claras o para profundizar en el conocimiento de efectos parcialmente importantes?

Análisis adicionales

- Siempre se debe investigar, al analizar los resultados, la posible existencia de resultados anómalos, debidos por ejemplo a la interferencia de algún hecho inusual, con el fin de corregirlos o de repetir en su caso las pruebas correspondientes

DISEÑOS EXPERIMENTALES CON UN SOLO FACTOR A EFECTOS FIJOS

Diseño completamente aleatorizado

- El caso más simple de un experimento es cuando hay un solo factor del cual se quiere estudiar su efecto sobre alguna variable respuesta
- En ese caso el diseño depende del material experimental disponible. Si hay material homogéneo suficiente, se pueden asignar aleatoriamente las variantes (o niveles) del factor a distintos grupos del material
- Este diseño se denomina COMPLETAMENTE ALEATORIZADO

Ejemplo 1

- Un horticultor aficionado llevó a cabo un experimento cuyo objetivo es descubrir si un cambio en la mezcla de fertilizantes traería como consecuencia un aumento de la cosecha de tomates.
- Dispone de 11 plantas de tomates todas ellas en el mismo surco; a 5 les suministró la mezcla estándar (fertilizante A) y a 6 las trató con una mezcla supuestamente mejorada (fertilizante B)
- Las mezclas se administraron aleatoriamente de la siguiente forma: se tomaron 11 naipes: 5 rojos y 6 negros, habiendo decidido asignar el fertilizante A si el naipe es rojo y el B si es negro.

Ejemplo 1 (continuación)

- Se mezclaron los 11 naipes y se obtuvo la sucesión: R, R, N, N, R, N, N, N, R, R, N

Planta	1	2	3	4	5	6	7	8	9	10	11
Fertilizante	A	A	B	B	A	B	B	B	A	A	B
Peso de Tomates (libras)	29,9	11,4	26,6	23,7	25,3	28,5	14,2	17,9	16,5	21,1	24,3

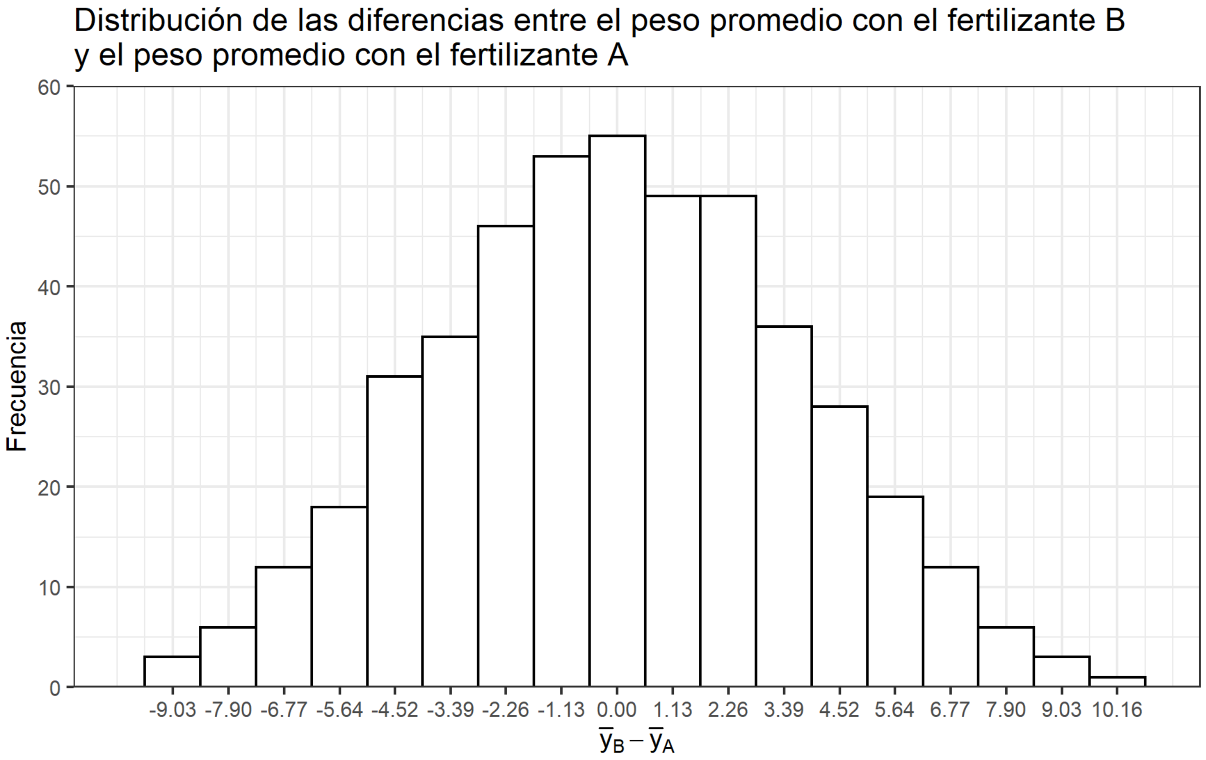
Ejemplo 1. Test de significación basado en la distribución de aleatorización

- Fisher argumentó que con tal experimento aleatorizado es posible plantear un contraste de significación sin realizar ninguna hipótesis sobre la distribución de la variable respuesta.
- La hipótesis nula es que la mezcla modificada no tiene efecto en los resultados y por lo tanto, no tiene efecto sobre la media.
- Si la hipótesis nula es cierta, los fertilizantes A y B son solo etiquetas. Por ejemplo, la planta 1 hubiese dado 29,9 libras tanto si hubiese sido marcada con A como con B.

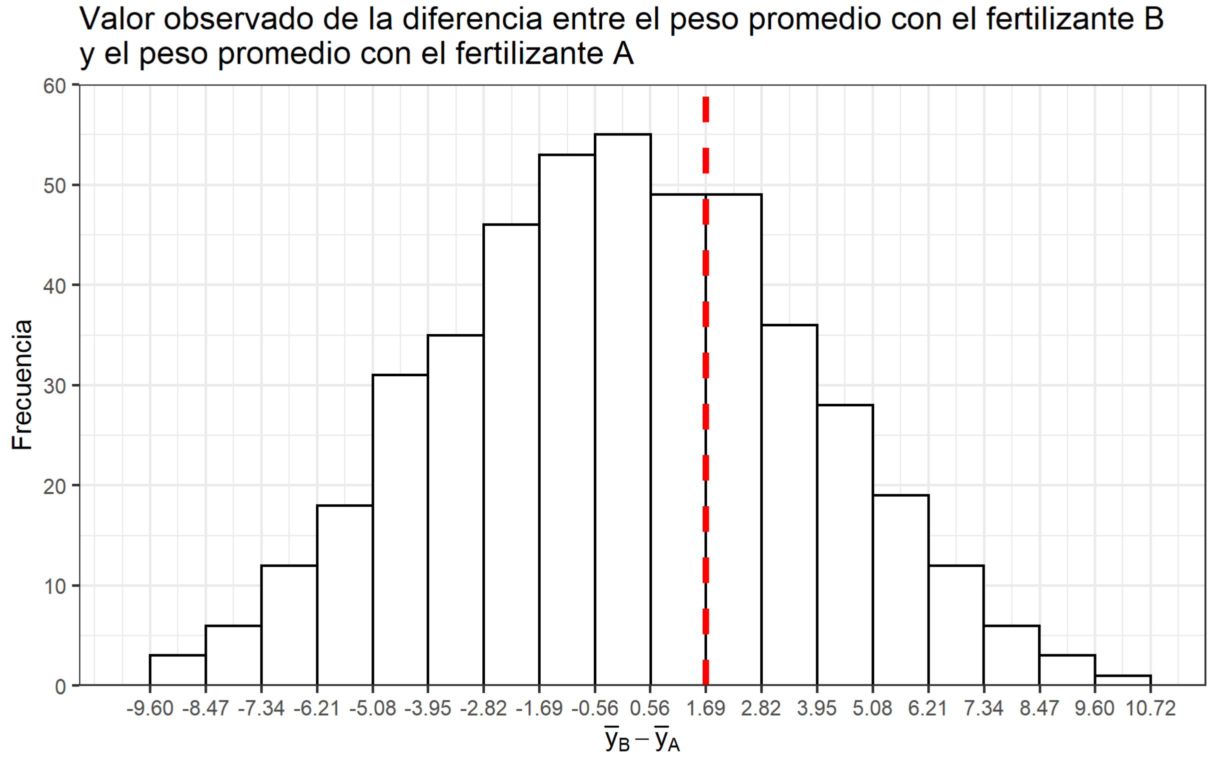
Ejemplo 1. Test de significación basado en la distribución de aleatorización

- La hipótesis alternativa establece que el fertilizante modificado da una media mayor.
- Las etiquetas pueden ser colocadas de 462 formas posibles (permutaciones de 11 con 5 y 6 repetidos).
- Se construye un conjunto de referencia calculando las 462 diferencias de medias: distribución de aleatorización si la hipótesis nula es cierta.

Ejemplo 1. Distribución de aleatorización de la diferencia de medias



Ejemplo 1. Distribución de aleatorización y diferencia observada en el experimento



Ejemplo 1. Test de significación basado en la distribución de aleatorización

- Luego, se verifica si el valor encontrado, es excepcional o puede considerarse proveniente de ese conjunto de referencia
- En el presente experimento, $\overline{y}_B = 22,53$ e $\overline{y}_A = 20,84$ El valor de la diferencia de las medias es $1,69$
- En la distribución de aleatorización se encontraron 154 diferencias mayores que ese valor (33%).
- En consecuencia: NO HAY RAZON PARA DUDAR DE LA HIPOTESIS NULA.

La distribución t como aproximación a la distribución de aleatorización(1)

Bajo la hipótesis de muestreo aleatorio de distribuciones normales, la estadística:

$$t_0 = \frac{(\bar{y}_B - \bar{y}_A) - (\mu_B - \mu_A)}{s\sqrt{1/n_A + 1/n_B}}$$

Se distribuye, según una t de Student con $(n_A + n_B - 2)$ grados de libertad.

Dado que bajo la Ho la diferencia de medias poblacionales es 0, resulta: t_0=1,69/3,82=0,44, y $P(t \geq t_0)$=0,34, por lo que no se rechaza Ho .

La distribución t como aproximación a la distribución de aleatorización(2)

- Los "valores p" son muy parecidos si la prueba se hace con la distribución de aleatorización o con la distribución t.
- Se puede demostrar que generalmente la distribución de aleatorización se aproxima razonablemente bien mediante una distribución t a escala apropiada.
- Entonces, si se ha aleatorizado, pueden emplearse los test t como aproximaciones de los test de aleatorización exactos y no se necesitan las hipótesis de muestreo aleatorio ni de normalidad exacta.
- Si se construye una gráfico con la distribución de aleatorización y la t escalada, se podrá apreciar el parecido.

Ejercicio propuesto

Si los datos obtenidos en un experimento con 8 plantas de tomate fueron:

Fertilizante	B	A	B	A	A	A	B	B
Libras de tomate	32	30	31	29	30	29	31	30

Construir la distribución de aleatorización de referencia y la distribución t a escala que se aproxima. ¿Cuál es el valor p en cada caso?

Ejemplo 2

- Un ingeniero está interesado en maximizar la resistencia a la tensión de una nueva fibra sintética que se empleará en la confección de camisas.

- La resistencia es influida por el % de algodón presente en la fibra. Se deciden probar muestras con 5 niveles de % de algodón (se ensayarán 5 probetas con cada % de algodón)

Resultados del experimento

Porcentaje de algodón	orden del ensayo	resistencia	orden del ensayo	resistencia	orden del ensayo	resistencia	orden del ensayo	resistencia	orden del ensayo	resistencia
15%	6	7	12	7	15	15	19	11	25	9
20%	1	12	3	17	8	12	11	18	14	18
25%	7	14	9	18	13	18	18	19	20	19
30%	2	19	5	25	10	22	22	19	24	23
35%	4	7	16	10	17	11	21	15	23	11

Montgomery, D.C. (2004) Diseño y análisis de experimentos. Editorial Limusa S.A.

Análisis estadístico de los datos

- El análisis de la información resultante se hace mediante el Análisis de la Variancia (ANOVA), técnica que fue derivada alrededor de 1930 por Ronald A. Fisher.

Análisis estadístico de los datos

- Denominemos con:
- a: número de tratamientos o poblaciones
- n_i: número de réplicas o de unidades en cada tratamiento (n si el experimento es balanceado.
- N es el total de observaciones
- y_{ij}: valor de la variable respuesta que se estudia, en la réplica j del tratamiento i

Modelo estadístico y prueba de hipótesis

- Modelo de medias

$$y_{ij} = \mu_i + \varepsilon_{ij}$$
$$i = 1,2,\ldots,a$$
$$j = 1,2,\ldots,n$$
$$\varepsilon_{ij} \sim N(0,\sigma^2), \text{ independientes}$$

- Hipótesis a probar

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_a$$
$$H_1: \mu_i \neq \mu_j \text{ para al menos un par } (i,j)$$

Modelo estadístico y prueba de hipótesis

- Modelo de efectos

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$
$$i = 1,2, \ldots, a$$
$$j = 1,2, \ldots, n$$
$$\varepsilon_{ij} \sim N(0, \sigma^2), \text{ independientes}$$

- Hipótesis a probar

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1: \tau_i \neq 0 \text{ para al menos un } i$$

¿Cómo se prueba el acuerdo entre datos y hipótesis?

- La forma de resolver el problema se basa en la consideración de la variabilidad entre las diferentes condiciones (tratamientos o poblaciones) con respecto a la existente dentro de los tratamientos (¿a qué corresponde?)
- Esta es la idea clave para probar cualquier hipótesis en un ANOVA
- La variabilidad total se puede particionar en:
- $SC_{total} = SC_{factor} + SC_{error}$

Análisis estadístico

- Llamando $y_{i.} = \sum_{j=1}^{n} y_{ij}$, $\bar{y}_{i.} = \dfrac{y_{i.}}{n}$, $y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$, $\bar{y}_{..} = \dfrac{y_{..}}{N}$
- La suma de cuadrados que recoge la variabilidad total es:

$$SC_{total} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{..}\right)^2$$

- Que puede expresarse como:

$$SC_{total} = \sum_{i=1}^{a} n_i \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i.}\right)^2$$

- Resultando una componente que refleja la variabilidad entre tratamientos SC_{factor} y otra, dentro de los tratamientos SC_{error}
- Si no hay efecto del factor, el primer sumando tendrá un valor pequeño .

Sumas de cuadrados, cuadrados medios y sus esperanzas

Considerando el modelo de efectos, se encuentran las esperanzas de los cuadrados medios

$$SC_{factor} = n\sum_{i=1}^{a}\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 \quad ; \quad CM_{factor} = \frac{SC_{factor}}{a-1}$$

$$E\left(CM_{factor}\right) = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$SC_{error} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i.}\right)^2 \quad ; \quad CM_{error} = \frac{SC_{error}}{N-a}$$

$$E\left(CM_{error}\right) = \sigma^2$$

Teorema de Cochran

- Sea

$$Z_i \sim NID(0;1) \text{ para } i = 1,2,\ldots,v$$

$$\sum_{i=1}^{v} Z_i^2 = Q_1 + Q_2 + \ldots + Q_s \text{ donde } s \leq v$$

$$Q_i \text{ tiene } v_i \text{ grados de libertad}$$

- Entonces $Q_1, Q_2, \ldots, Q_s$ son variables aleatorias chi cuadrado independientes con $v = v_1, v_2, \ldots, v_s$ grados de libertad, si solo si :

$$v = v_1 + v_2 + \ldots + v_s$$

Comprobación de la H_0

- Si H_0 es verdadera, CM_{factor} tendrá un valor muy pequeño, ya que los valores de "resistencia media" para cada % de algodón serán muy parecidos.
- Se propone probar la H_0 por medio del cociente: $F = CM_{factor} / CM_{residual}$. Esta estadística sigue, si H_0 es cierta una distribución $\mathcal{F}_{I-1,n-I-1}$.
- Valores grandes de F indican el rechazo de H_0.
- Se calcula el p-value (probabilidad de obtener una F tan grande como la observada o mayor) y si es menor que 0.05, se rechaza H_0

Rechazo H_0 ¿cuáles medias son diferentes?

- Si el test F resulta significativo, habría que estudiar entre cuales de los tratamientos existen diferencias, ya que el rechazo de la H_0 solo indica que al menos una de las medias es diferente de las otras, no surgiendo de esta prueba, cuales medias son diferentes
- Se deberá acompañar al test con el cálculo de intervalos de confianza... ¿cuáles?

Comparaciones Múltiples

- Existen varios métodos para realizar comparaciones simultáneas entre medias. Resulta de interés considerar los intervalos para diferencias de medias de a dos, de Fisher y de Tuckey o intervalos LSD (hay más alternativas
- Como regla práctica, intervalos que no cubran al valor 0 estarán indicando diferencias significativas entre las medias involucradas
- Otra opción es calcular CONTRASTES y evaluar su significatividad con un test

Contrastes-caso **balanceado**

Un contraste se define como: $\Gamma = \sum_{i=1}^{a} c_i \mu_i$

donde $\sum_{i=1}^{a} c_i = 0$

Se desea probar $H_0) \sum_{i=1}^{a} c_i \mu_i = 0$, $H_1) \sum_{i=1}^{a} c_i \mu_i \neq 0$

La prueba de hipótesis se realizará considerando:

$C = \sum_{i=1}^{a} c_i \bar{y}_{i.}$; $V(C) = \sum_{i=1}^{a} c_i^2 \frac{\sigma^2}{n}$

Entonces, si H_0 es verdadera: $C/\sqrt{V(C)} \sim N(0,1)$

Contrastes-caso **balanceado**

para el caso de varianza σ^2 desconocida

$$\frac{\sqrt{n}\sum_{i=1}^{a} c_i \bar{y}_i}{\sqrt{CM_{error}\sum_{i=1}^{a} c_i^2}} \sim t_{N-a}$$

Puede también considerarse que: $\dfrac{n\left(\sum_{i=1}^{a} c_i \bar{y}_i\right)^2}{CM_{error}\sum_{i=1}^{a} c_i^2} \sim F_{1,N-a}$

y: $SC_c = \dfrac{n\left(\sum_{i=1}^{a} c_i \bar{y}_i\right)^2}{\sum_{i=1}^{a} c_i^2}$ con un grado de libertad

Contrastes-caso **no balanceado**

El contraste se define como $$C = \sum_{i=1}^{a} c_i n_i \bar{y}_{i.}$$

Donde $$\sum_{i=1}^{a} n_i c_i = 0$$

Y la suma de cuadrados del contraste es:

$$SC_c = \frac{\left(\sum_{i=1}^{a} c_i n_i \bar{y}_i \right)^2}{\sum_{i=1}^{a} n_i c_i^2}$$

Contrastes ortogonales

- Dos contrastes C y D con coeficientes c_i y d_i son ortogonales si $\sum_{i=1}^{a} c_i d_i = 0$ en el caso balanceado,
- y $\sum_{i=1}^{a} n_i c_i d_i = 0$ en el caso no balanceado.
- La suma de cuadrados del factor, con *a-1* grados de libertad, puede descomponerse en *a-1* contrastes ortogonales.

Contrastes ortogonales

- Una aplicación particular es el estudio de la naturaleza de la relación entre un factor cuantitativo y la respuesta.
- Si el efecto del factor es significativo, puede descomponerse la suma de cuadrados del factor, en contrastes asociados con componente lineal, cuadrática, etc. utilizando ciertos coeficientes

Coeficientes de los contrastes lineal, cuadrático y cúbico para factores con hasta 5 niveles

3 niveles			4 niveles				5 niveles					
$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	
-1	0	1	-3	-1	1	3	-2	-1	0	1	2	Comp. lineal
1	-2	1	1	-1	-1	1	2	-1	-2	-1	2	Comp. cuadrática
----	----	----	-1	3	-3	1	-1	2	0	-2	1	Comp. cúbica

Verificación de supuestos

- El análisis de la variancia a efectos fijos (como el que se estudia aquí) plantea para los datos el siguiente modelo:
- $y_{ij}=\mu+\tau_i+\varepsilon_{ij}$
- $\mu_i=\mu+\tau_i$ es la media esperada para el tratamiento i
- ε_{ij} residuos normales con media 0 y variancia σ^2 constante (homocedasticidad) no correlacionados entre ellos

Verificación de supuestos-Análisis de residuos

- Los residuos deben seguir una distribución normal. Esto se podrá evaluar graficando los mismos sobre papel normal
- Un gráfico de los residuos frente al orden de las observaciones deberá presentar un comportamiento errático
- Un gráfico de los residuos frente a los valores ajustados permitirá evaluar el supuesto de homocedasticidad