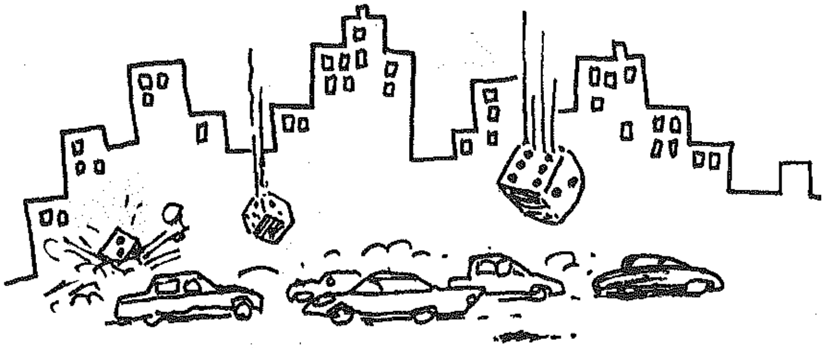
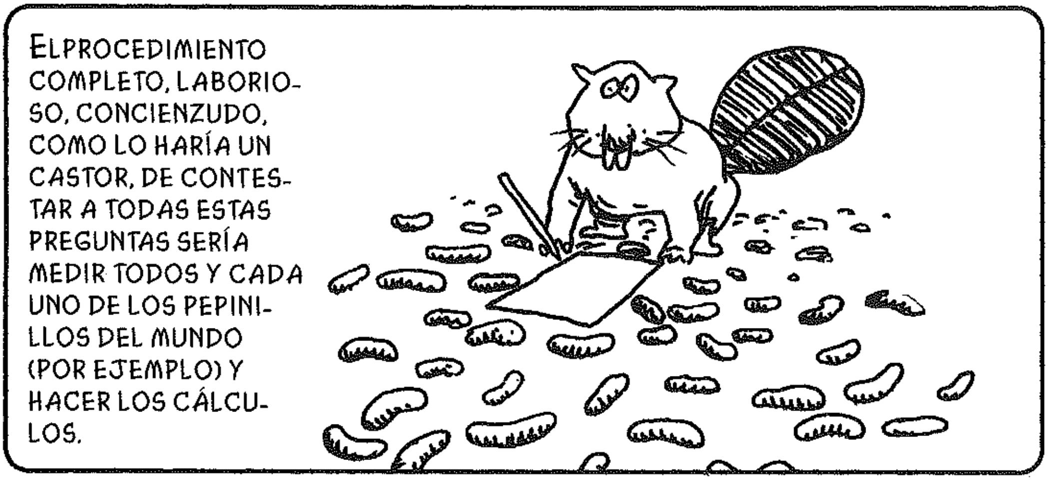
♦ Capítulo 6 ♦
MUESTREO

A ESTAS ALTURAS, TRAS UNA DIETA REGULAR DE MONEDAS, DATOS E IDEAS ABSTRACTAS, A LO MEJOR TE PREGUNTAS QUÉ TIENE QUE VER TODO ESTE MATERIAL ESTADÍSTICO QUE HEMOS DESARROLLADO CON EL **MUNDO REAL**. BUENO, POR FIN LO VAS A DESCUBRIR...

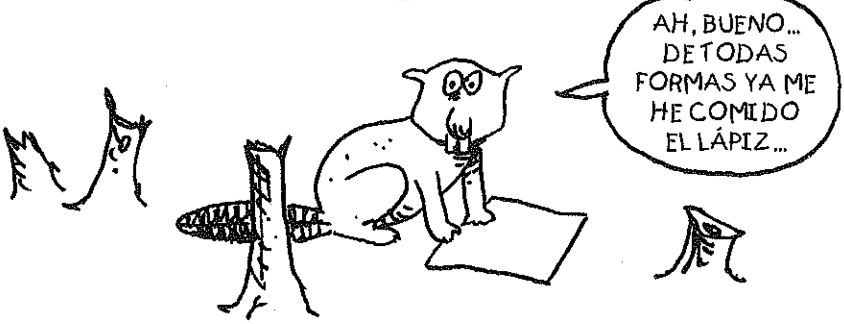


EN ESTE CAPÍTULO EMPEZAMOS A VER LA TAREA **REAL** DE LA ESTADÍSTICA, QUE AL FIN Y AL CABO ES AHORRARNOS **TIEMPO** Y **DINERO**. LA GENTE ODIA PERDER EL TIEMPO EN **TRABAJOS INNECESARIOS**, Y SI HAY ALGO QUE LA ESTADÍSTICA PUEDE HACER ES DECIRNOS EXACTAMENTE CUÁNTA HOLGAZANERÍA NOS PODEMOS PERMITIR.

EL PROBLEMA QUE TIENE EL MUNDO REAL ES QUE LOS CONJUNTOS DE COSAS SON TAN GRANDES QUE RESULTA MUY DIFÍCIL CONSEGUIR LA INFORMACIÓN QUE QUEREMOS:

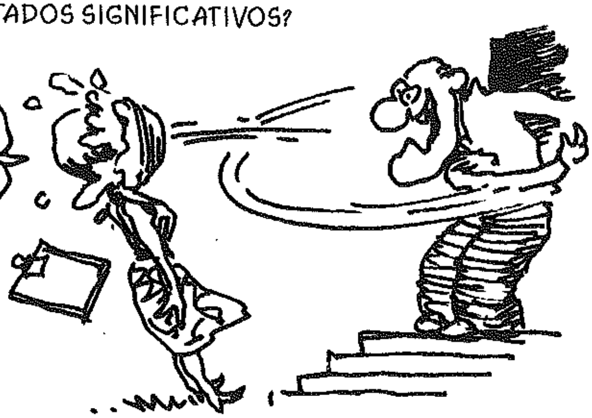


PERO NOSOTROS NO SOMOS CASTORES, ¡SOMOS ESTADÍSTICOS! BUSCAMOS LA FORMA MÁS **SENCILLA**...



NUESTRO MÉTODO ES TOMAR UNA **MUESTRA**... UN SUBCONJUNTO RELATIVAMENTE PEQUEÑO DE LA POBLACIÓN TOTAL, IGUAL QUE CUANDO SE HACE UN SONDEO DE OPINIÓN DURANTE UNAS ELECCIONES.

UNA PREGUNTA OBVIA ES: ¿CUÁNTOS ELEMENTOS DEBE TENER LA MUESTRA PARA OBTENER RESULTADOS SIGNIFICATIVOS?

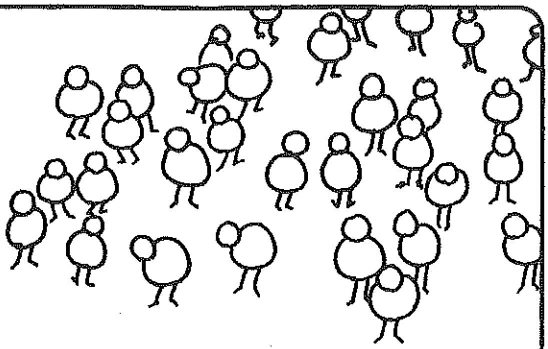


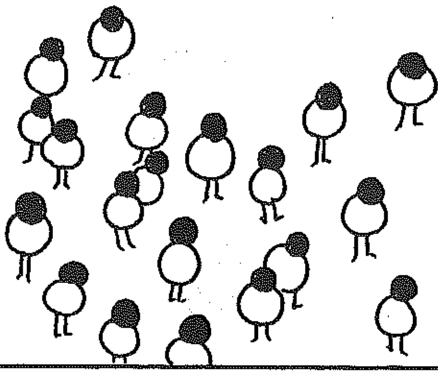
Y LA RESPUESTA, QUE DEBE QUEDARTE GRABADA EN EL CEREBRO PARA SIEMPRE JAMÁS, ES: SI n ES EL NÚMERO DE ELEMENTOS DE LA MUESTRA, ENTONCES TODO ESTÁ GOBERNADO POR

$$\frac{1}{\sqrt{n}}.$$

DISEÑO DEL MUESTREO

ANTES DE EMPEZAR CON LOS NÚMEROS, DEBERÍAMOS SEÑALAR QUE LA **CALIDAD** DE LA MUESTRA ES TAN IMPORTANTE COMO SU **TAMAÑO**. ¿CÓMO PODEMOS ESTAR SEGUROS DE QUE ESCOGEMOS UNA MUESTRA REPRESENTATIVA?



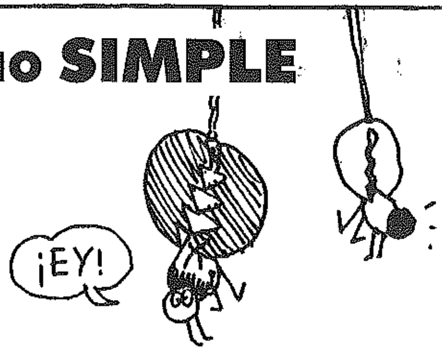


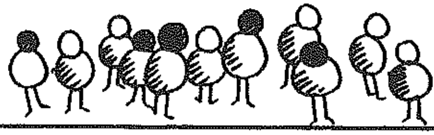
EL MISMO **PROCESO DE SELECCIÓN** ES DE VITAL IMPORTANCIA. POR EJEMPLO, UNA ENCUESTA DE VOTANTES QUE EXCLUYA SISTEMÁTICAMENTE A LOS NEGROS NO TENDRÍA NINGÚN VALOR, Y HAY MILES DE FORMAS MÁS DE ESTROPEAR, O SESGAR, UNA MUESTRA.

PARA NO PROLONGAR EL MISTERIO, LA FORMA DE OBTENER RESULTADOS ESTADÍSTICOS FIABLES ES ESCOGER LA MUESTRA **al azar.**

EL **MUESTREO ALEATORIO SIMPLE**

SUPONGAMOS QUE TENEMOS UNA GRAN POBLACIÓN DE OBJETOS Y UN PROCEDIMIENTO PARA ESCOGER *n* DE ELLOS. SI ESE PROCEDIMIENTO ASEGURA QUE **TODAS LAS MUESTRAS POSIBLES** DE *n* OBJETOS **TIENEN LA MISMA PROBABILIDAD**, ENTONCES ESE PROCEDIMIENTO RECIBE EL NOMBRE DE **muestreo aleatorio simple.**





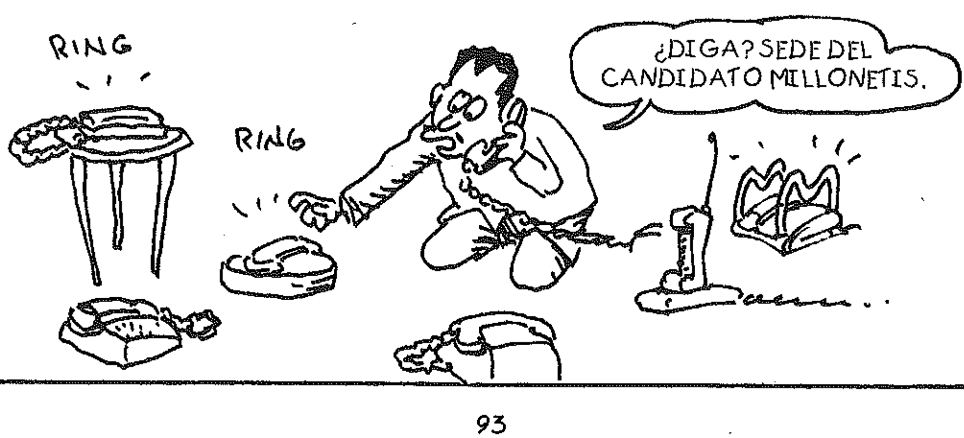
EL MUESTREO ALEATORIO SIMPLE PRESENTA DOS PROPIEDADES QUE LO CONVIERTEN EN UN ESTÁNDAR FRENTE AL QUE MEDIMOS TODOS LOS OTROS MÉTODOS:

1) REPRESENTATIVA: CADA UNIDAD TIENE LAS MISMAS POSIBILIDADES DE SER ESCOGIDA.*

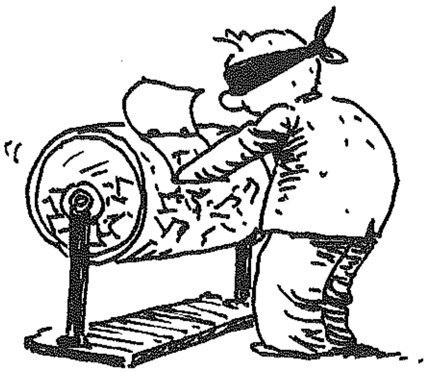
2) INDEPENDENCIA: LA SELECCIÓN DE UNA UNIDAD NO INFLUYE EN LA SELECCIÓN DE OTRAS UNIDADES.

* UN CONCEPTO ESTADÍSTICO MÁS FORMAL ÉS LA AUSENCIA DE SESGO. [N.T.]

POR DESGRACIA, EN EL MUNDO REAL ES MUY DIFÍCIL ENCONTRAR MUESTRAS COMPLETAMENTE INDEPENDIENTES Y REPRESENTATIVAS. POR EJEMPLO, HACER UNA ENCUESTA A LOS VOTANTES MARCANDO NÚMEROS DE TELÉFONO AL AZAR ES UN MÉTODO NO REPRESENTATIVO: NO TIENE EN CUENTA A LOS VOTANTES QUE NO DISPONEN DE TELÉFONO Y CUENTA VARIAS VECES A LOS QUE TIENEN VARIOS NÚMEROS.



TEÓRICAMENTE, ES POSIBLE OBTENER UNA MUESTRA AL AZAR CONSTRUYENDO UN **MARCO DE MUESTREO**: UNA LISTA CON TODAS LAS UNIDADES DE LA POBLACIÓN. UTILIZANDO UN GENERADOR DE NÚMERO ALEATORIO, ESCOGEMOS *n* OBJETOS AL AZAR.



DE IGUAL FORMA, PODEMOS ESCRIBIR TODOS LOS NOMBRES EN TARJETAS Y EXTRAER *n* DE ELLOS DE UN BOMBO.

SIN EMBARGO, NO SIEMPRE ES TAN SENCILLO. EL MARCO PUEDE RESULTAR PROHIBITIVO, CARO, POLÉMICO E, INCLUSO, IMPOSIBLE DE ESTABLECER. POR EJEMPLO, UN ESTUDIO SOBRE LA CALIDAD DEL AGUA DE LA AGENCIA PARA LA PROTECCIÓN DEL MEDIO AMBIENTE DE ESTADOS UNIDOS NECESITABA UN MARCO DE MUESTRA DE LOS LAGOS DEL PAÍS, ASÍ QUE ALGUIEN TENÍA QUE DECIDIR:

¿EXISTE ALGÚN OTRO MÉTODO MÁS **EFICAZ** Y **RENTABLE** QUE UNA MUESTRA ALEATORIA SIMPLE? SÍ, SI ES QUE YA SABES ALGO DE LA POBLACIÓN. POR EJEMPLO...

EL MUESTREO

estratificado

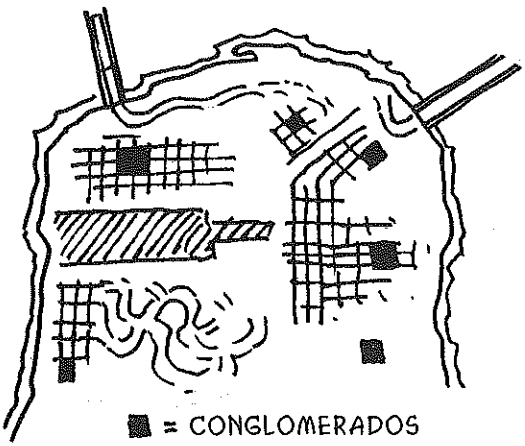
DIVIDE LAS UNIDADES DE POBLACIÓN EN GRUPOS HOMOGÉNEOS (**ESTRATOS**) Y LUEGO LLEVA A CABO MUESTREO ALEATORIO SIMPLE DE CADA GRUPO.

POR EJEMPLO, LA POBLACIÓN DE TODAS LAS CONSERVAS EN VINAGRE SE PUEDE ESTRATIFICAR POR EL **TIPO DE CONSERVA**. DENTRO DE CADA TIPO, O ESTRATO, EL TAMAÑO SERÁ MENOS VARIABLE.

EL MUESTREO POR

conglomerados

AGRUPA LA POBLACIÓN EN PEQUEÑOS CONGLOMERADOS, REALIZA UNA MUESTRA ALEATORIA SIMPLE DE ELLOS Y TIENE EN CUENTA ABSOLUTAMENTE TODO DENTRO DE CADA CONGLOMERADO MUESTREADO. ESTO PUEDE RESULTAR RENTABLE SI LOS COSTES DE TRANSPORTE ENTRE LAS UNIDADES DE MUESTRA ALEATORIA SON ELEVADOS.

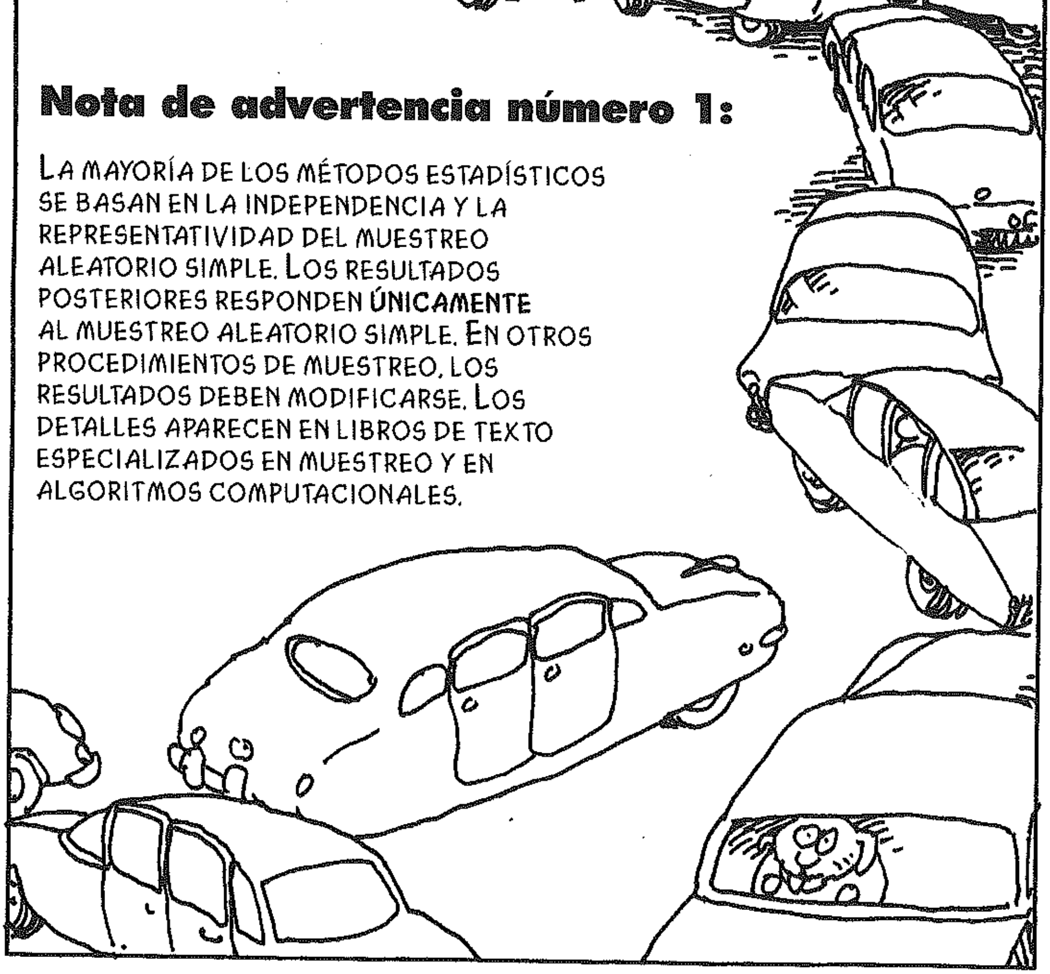


UN BUEN EJEMPLO ES EL DE UNA ENCUESTA SOBRE LA VIVIENDA, QUE DIVIDE LA CIUDAD EN BLOQUES Y ESTUDIA CADA UNIDAD DE VIVIENDAS DE CADA BLOQUE DE LA MUESTRA.

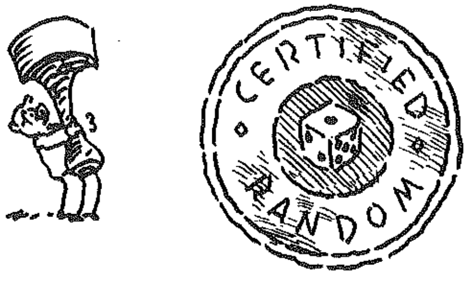
EL MUESTREO **sistemático** EMPIEZA CON UNA UNIDAD ESCOGIDA AL AZAR Y LUEGO SELECCIONA CADA UNIDAD QUE SE ENCUENTRE A *k* UNIDADES DE AQUELLA. POR EJEMPLO, UN **ESTUDIO DEL TRÁFICO EN AUTOPISTAS** PODRÍA ESTUDIAR UNO DE CADA CIEN COCHES QUE PASARA POR EL PEAJE. ESTE PLAN ES FÁCIL DE APLICAR Y PUEDE SER MÁS EFICAZ SI LOS PATRONES DEL TRÁFICO VARÍAN CON EL PASO DE LAS HORAS.

Nota de advertencia número 1:

LA MAYORÍA DE LOS MÉTODOS ESTADÍSTICOS SE BASAN EN LA INDEPENDENCIA Y LA REPRESENTATIVIDAD DEL MUESTREO ALEATORIO SIMPLE. LOS RESULTADOS POSTERIORES RESPONDEN **ÚNICAMENTE** AL MUESTREO ALEATORIO SIMPLE. EN OTROS PROCEDIMIENTOS DE MUESTREO, LOS RESULTADOS DEBEN MODIFICARSE. LOS DETALLES APARECEN EN LIBROS DE TEXTO ESPECIALIZADOS EN MUESTREO Y EN ALGORITMOS COMPUTACIONALES.



Nota de advertencia número 2:



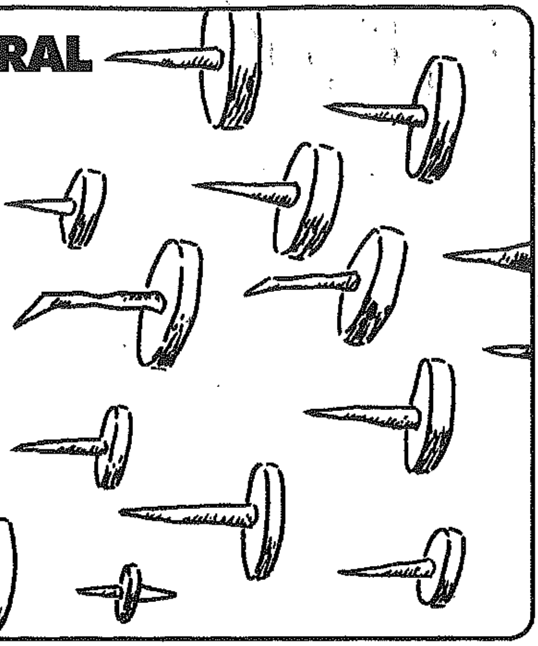
NO EXISTE ANÁLISIS ESTADÍSTICO FIABLE SIN UN DISEÑO ALEATORIZADO, NO IMPORTA CUÁNTO SE MODIFIQUE DESPUÉS. LA BELLEZA DEL MUESTREO ALEATORIO RESIDE EN QUE «GARANTIZA ESTADÍSTICAMENTE» LA EXACTITUD DEL ESTUDIO.

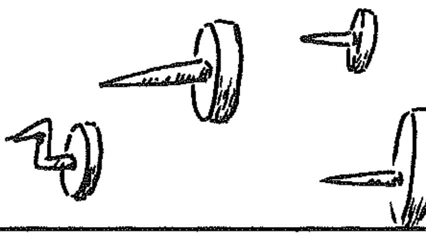
UNO DE LOS MÉTODOS MÁS COMUNES TIENDE A MENUDO A LA PARCIALIDAD: SE TRATA DEL MUESTREO **oportunista.** ESTE MÉTODO EVITA TODA LA PROBLEMÁTICA DE DISEÑAR UN PROCEDIMIENTO Y SE LIMITA A TOMAR LAS PRIMERAS *n* UNIDADES DE LA POBLACIÓN QUE SE PRESENTEN.

UN CLÁSICO EJEMPLO ES EL LIBRO DE SHERE HITE *MUJERES Y AMOR.* SE ENVIARON 100.000 CUESTIONARIOS A ORGANIZACIONES DE MUJERES (UN **MUESTREO OPORTUNISTA**), Y SÓLO UN **4,5%** SE RELLENARON Y ENTREGARON (**RESPUESTA PARCIAL**). ASÍ QUE SUS «RESULTADOS» ESTABAN BASADOS EN UNA MUESTRA DE MUJERES QUE, POR UNA RAZÓN U OTRA, TENÍAN UNA GRAN MOTIVACIÓN PARA CONTESTAR LAS PREGUNTAS DE LA ENCUESTA.

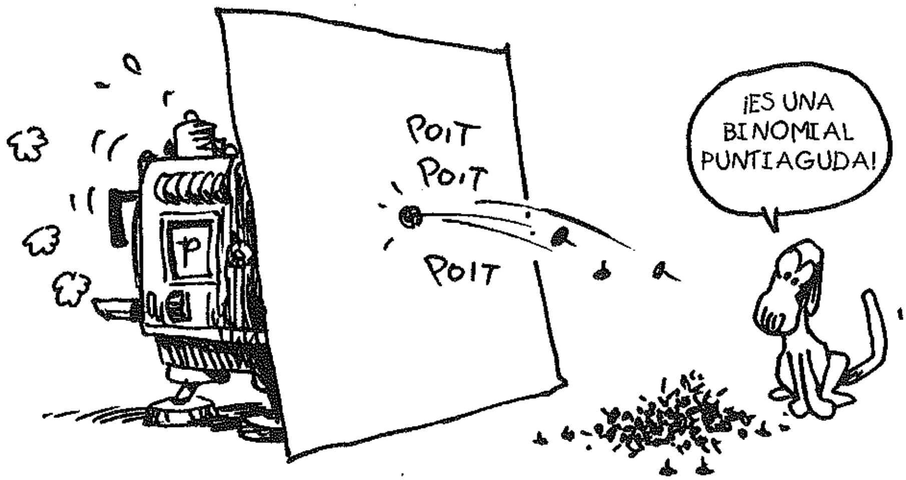
El TAMAÑO MUESTRAL y el error típico

Y AHORA VAMOS A DAR EN EL **CLAVO**... PERO CON CLAVOS DE VERDAD. SUPONGAMOS QUE LA **FÁBRICA DE CLAVOS BERNOULLI** PRODUCE CLAVOS A MILES Y ALGUNOS, CLARO, RESULTAN **DEFECTUOSOS**.





EL ASTUTO LECTOR SE DARÁ CUENTA EN SEGUIDA DE QUE SE TRATA DE UN **SISTEMA DE BERNOULLI**: CADA NUEVO CLAVO ES EL RESULTADO DE UNA PRUEBA DE BERNOULLI CON PROBABILIDAD p DE ÉXITO (EN ESTE CASO, NO SER DEFECTUOSO) Y PROBABILIDAD $1 - p$ DE FRACASO (SER DEFECTUOSO).



PENSAMOS EN ESTA SITUACIÓN COMO SI HUBIESE UNA **«MÁQUINA DE BERNOULLI» REAL AUNQUE ESCONDIDA** CUYA PROBABILIDAD p RIGE LOS RESULTADOS QUE OBSERVAMOS EN EL LLAMADO «MUNDO REAL».

COMO LA MÁQUINA DE BERNOULLI ES INVISIBLE, NO SABEMOS CUÁL ES LA PROBABILIDAD p, PERO NOS GUSTARÍA DESCUBRIRLO. ASÍ QUE TOMAMOS UNA **MUESTRA ALEATORIA** DE N CLAVOS Y VEMOS QUE, DE TODOS ELLOS, x NO TIENEN NINGÚN DEFECTO.

BIEN, LA PROPORCIÓN DE ÉXITOS NO DEBERÍA DIFERIR MUCHO DE p... ASÍ QUE LA LLAMAMOS $\hat{p}$, «PE CON CIRCUNFLEJO».

$$\hat{p} = \frac{x}{n}$$

$\hat{p}$ ES EL NÚMERO DE ÉXITOS x DE LA MUESTRA, DIVIDIDO ENTRE EL TAMAÑO n DE ÉSTA. POR EJEMPLO, SI p FUERA 0,85, Y HUBIÉSEMOS TOMADO $n = 1.000$ TORNILLOS COMO MUESTRA, QUIZÁ ALREDEDOR DE $x = 832$ ESTARÍAN BIEN Y ENTONCES $\hat{p} = 0,832$.

NOS PREGUNTAMOS: ¿ES BUENA ESTA ESTIMACIÓN?

Y CONTESTAMOS CON OTRO INTERROGANTE: ¿QUÉ SIGNIFICA LA PRIMERA PREGUNTA?

NO PODEMOS SABER LA DIFERENCIA EXACTA ENTRE $\hat{p}$ Y p, PORQUE NO CONOCEMOS EL VALOR p. LA **AUTÉNTICA** PREGUNTA ES LA SIGUIENTE: SI TOMÁRAMOS **MUCHAS MUESTRAS DE 1.000 CLAVOS** Y OBSERVÁRAMOS EL NÚMERO $\hat{p}$ DE CADA MUESTRA, ¿CUÁL SERÍA LA DISTRIBUCIÓN DE ESOS VALORES DE $\hat{p}$ ALREDEDOR DE p?

DE HECHO, ESTOS VALORES DE $\hat{p}$ CADA VEZ SE PARECEN MÁS A UNA **VARIABLE ALEATORIA**: LA SELECCIÓN DE UNA **MUESTRA DE n UNIDADES ES UN EXPERIMENTO ALEATORIO**, ¡Y LA OBSERVACIÓN $\hat{p}$ ES UN RESULTADO NUMÉRICO!

PARA SER EXACTOS, SI X ES EL NÚMERO DE ÉXITOS DE LA MUESTRA, ENTONCES X NO ES MÁS QUE NUESTRA VIEJA AMIGA LA VARIABLE ALEATORIA BINOMIAL (n PRUEBAS, PROBABILIDAD p)... Y DEFINIMOS LA **PROPORCIÓN OBSERVADA** COMO LA VARIABLE ALEATORIA

$$\hat{P} = \frac{X}{n}$$

COMO LO SABEMOS TODO SOBRE X, PODEMOS DEDUCIR SIN PROBLEMAS UNOS CUANTOS HECHOS SOBRE $\hat{P}$:

1) LA MEDIA DE $\hat{P}$ ES $E[\hat{P}] = p$

2) LA DESVIACIÓN TÍPICA DE $\hat{P}$ ES

$$\sigma(\hat{P}) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

3) PARA UNA n MUY GRANDE, $\hat{P}$ ES APROXIMADAMENTE NORMAL.

¡Y ESO ES TODO! LOS VALORES OBSERVADOS DE $\hat{P}$ SE CENTRARÁN EN p (EVIDENTEMENTE), Y SU DESVIACIÓN TÍPICA, O DISPERSIÓN, SERÁ PROPORCIONAL AL NÚMERO MÁGICO QUE HABÍAMOS MENCIONADO AL PRINCIPIO DEL CAPÍTULO:

$$\frac{1}{\sqrt{n}}$$

Y COMO $\hat{P}$ ES BASTANTE NORMAL, PODEMOS USAR LA REGLA EMPÍRICA PARA CONCLUIR QUE APROXIMADAMENTE UN 68% DE TODAS LAS ESTIMACIONES QUEDARÁN A MENOS DE UNA DESVIACIÓN TÍPICA DEL VALOR REAL p.

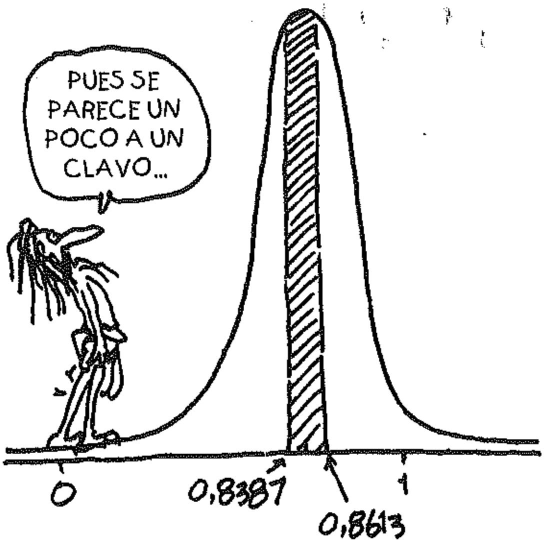
Y VOLVIENDO A LOS CLAVOS, CON $n = 1.000$ Y $p = 0,85$, LA DESVIACIÓN TÍPICA DE $\hat{p}$ ES

$$\sigma(\hat{P}) = \sqrt{\frac{(0,85)(0,15)}{1.000}}$$

$$= 0,0113$$

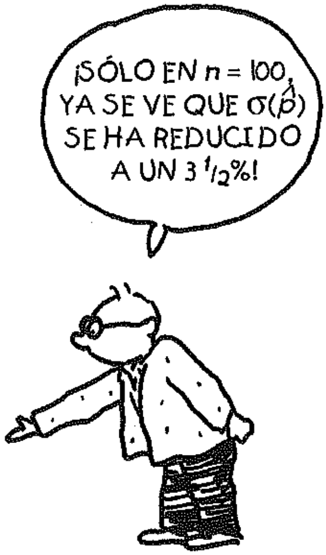
ASÍ QUE ESPERAMOS QUE ALREDEDOR DE UN 68% DE NUESTRAS ESTIMACIONES QUEDEN DENTRO DEL PEQUEÑO INTERVALO

$$0,8387 \leq \hat{p} \leq 0,8613$$



LA DESVIACIÓN TÍPICA DE $\hat{P}$ ES UNA MEDIDA DEL **error muestral.**

COMO YA HEMOS VISTO, PARA LA BINOMIAL ESTE ERROR MUESTRAL ES INVERSAMENTE PROPORCIONAL A $\sqrt{n}$. SI SE AUMENTA EL TAMAÑO MUESTRAL EN UN FACTOR 4, LA DISPERSIÓN $\sigma(\hat{P})$ SE REDUCE EN UN FACTOR 2.



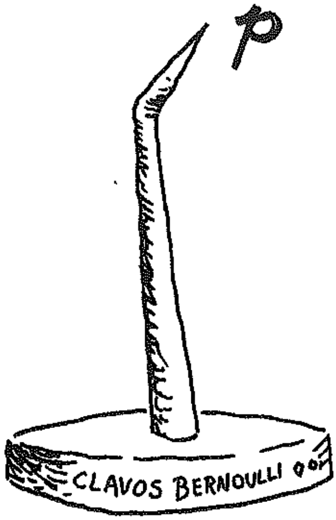
TAMAÑOS MUESTRALES DE CLAVOS, $p = 0,85$

n	1	4	16	25	100	10.000
$\sqrt{n}$	1	2	4	5	10	100
$\sigma(\hat{p})$	0,357	0,1785	0,089	0,071	0,0357	0,0036

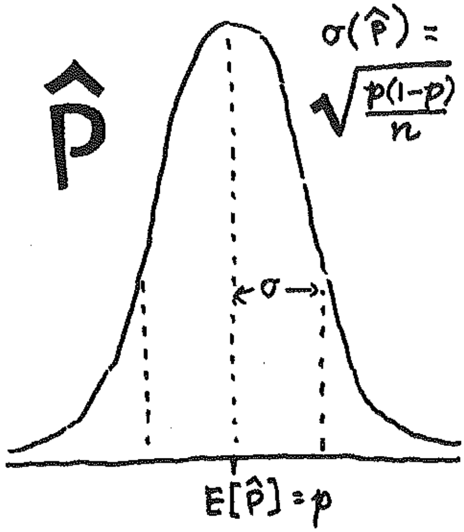
NOTA LINGÜÍSTICA: UNA **ESTIMACIÓN** ES UNA SOLA MEDIDA U OBSERVACIÓN. UN **ESTIMADOR** ES UNA REGLA PARA OBTENER ESTIMACIONES. EN ESTE CASO EL ESTIMADOR ES LA VARIABLE ALEATORIA $\hat{P} = \frac{X}{n}$.

CASI TODA LA ESTADÍSTICA IMPLICA UN PROCESO DE CUATRO ETAPAS POR EL QUE ACABAMOS DE PASAR:

DEFINIR LA POBLACIÓN CON UN PARÁMETRO DESCONOCIDO.



ENCONTRAR UN ESTIMADOR, SU DISTRIBUCIÓN MUESTRAL TEÓRICA Y SU DESVIACIÓN TÍPICA.

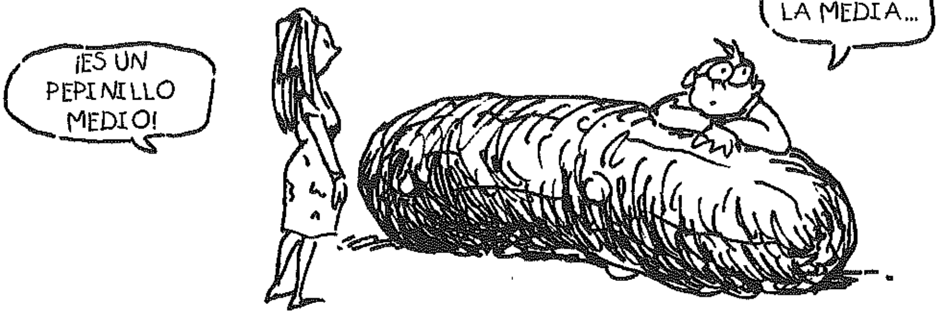


EXTRAER UNA MUESTRA ALEATORIA Y ENCONTRAR LA ESTIMACIÓN.

HACER UN INFORME CON LOS RESULTADOS Y SU ERROR MUESTRAL O ESTADÍSTICO.

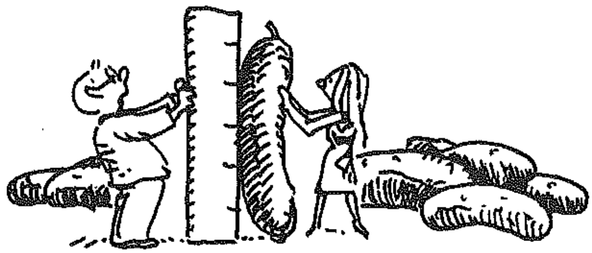
Distribución muestral de la MEDIA

Y AHORA PASAMOS DE LOS CLAVOS A LOS PEPINILLOS EN VINAGRE...



A LOS FABRICANTES DE BOTES LES GUSTARÍA SABER EL **TAMAÑO MEDIO** DE UN PEPINILLO SIN TENER QUE EXAMINAR TODOS LOS PEPINOS DEL CONTINENTE. SELECCIONAN *n* PEPINILLOS AL AZAR Y LOS MIDEN, $x_1, x_2, \ldots, x_n$.

AHORA QUIZÁ YA TE HAYAS ACOSTUMBRADO A QUE CADA X_i ES UNA **VARIABLE ALEATORIA**: EL RESULTADO NUMÉRICO DE UN EXPERIMENTO ALEATORIO.



SI μ ES EL TAMAÑO MEDIO (DESCONOCIDO) DE UN PEPINILLO, Y σ ES LA DESVIACIÓN TÍPICA DE LA **DISTRIBUCIÓN DEL TAMAÑO DEL PEPINILLO**, ENTONCES

$$E[X_i] = \mu$$
$$\sigma(X_i) = \sigma$$

PARA CADA i (YA QUE X_i PODRÍA HABER SIDO EL TAMAÑO DE CUALQUIER PEPINILLO).

A CONTINUACIÓN OBSERVAMOS LA MEDIA MUESTRAL: EL TAMAÑO MEDIO DE LOS PEPINILLOS ESCOGIDOS. ES UNA NUEVA VARIABLE ALEATORIA QUE VIENE DADA POR:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

IGUAL QUE ANTES, NOS GUSTARÍA SABER LO «CERCA» QUE SE ENCUENTRA DE μ, ES DECIR, SI REALIZÁRAMOS ESTE MUESTREO REPETIDAS VECES, ¿CUÁL SERÍA LA DISTRIBUCIÓN DE $\bar{X}$? COMO TENEMOS DATOS DE $x_1, x_2, \dots$ Y x_n, TAMBIÉN SABEMOS QUE

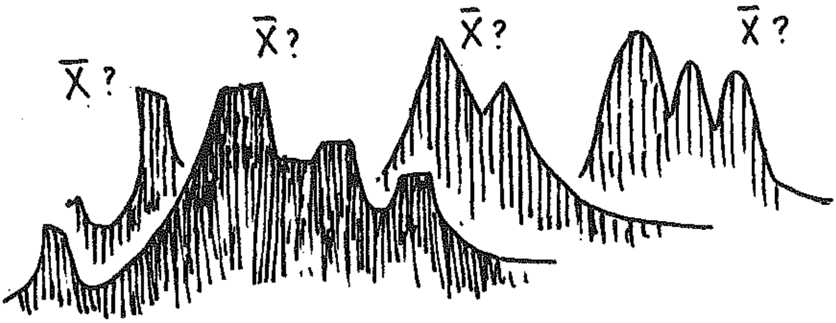
$$E[\bar{X}] = \mu$$

$$\sigma(\bar{X}) = \sigma/\sqrt{n}$$

¡DE NUEVO NOS ENCONTRAMOS CON ESE DENOMINADOR MÁGICO! LA DISPERSIÓN DE LAS MEDIAS MUESTRALES QUE HEMOS OBSERVADO ES PROPORCIONAL A

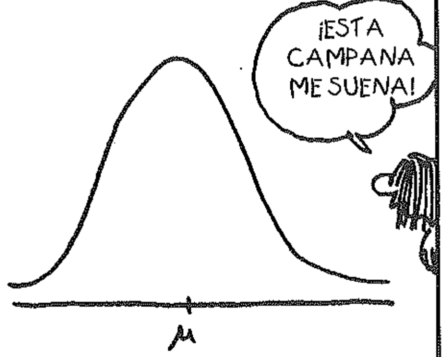
$$\frac{1}{\sqrt{n}}.$$

SIN EMBARGO, DESCONOCEMOS LA FORMA DE LA DISTRIBUCIÓN DE $\bar{X}$. LA **DISTRIBUCIÓN DE PROBABILIDAD** MUESTRAL DE $\hat{p}$ ERA CASI NORMAL PORQUE ESTABA BASADA EN UNA VARIABLE ALEATORIA BINOMIAL. PERO, ¿¿QUÉ PASA CON $\bar{X}$, EL ESTIMADOR **DE LA MEDIA** MUESTRAL??



¡RESULTA QUE $\bar{X}$ TAMBIÉN ES APROXIMADAMENTE NORMAL! ESTE FAMOSO RESULTADO SE LLAMA TAMBIÉN

TEOREMA CENTRAL DEL LÍMITE

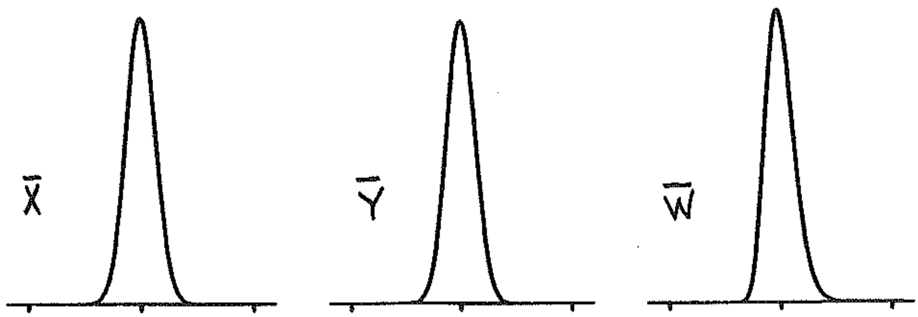


Y DICE ASÍ: SI TOMAMOS MUESTRAS ALEATORIAS DE TAMAÑO n DE UNA POBLACIÓN DE MEDIA μ Y DESVIACIÓN TÍPICA σ, ENTONCES, A MEDIDA QUE n SE HACE MAYOR, $\bar{X}$ SE ACERCA A LA **DISTRIBUCIÓN NORMAL** CON MEDIA μ Y DESVIACIÓN TÍPICA $\sigma/\sqrt{n}$. ENTONCES,

$$Pr(a \leq \bar{X} \leq b) \approx Pr\left(\frac{a-\mu}{\sigma/\sqrt{n}} \leq Z \leq \frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

¿QUÉ TIENE ESTO DE EXTRAORDINARIO? NOS DICE QUE, SIN IMPORTAR LA FORMA QUE TENGA LA DISTRIBUCIÓN ORIGINAL (EN ESTE CASO, DEL TAMAÑO DE LOS PEPINILLOS), LA DISTRIBUCIÓN DE LA **MEDIA MUESTRAL** CONVERGE A UNA **NORMAL**. PARA ENCONTRAR LA DISTRIBUCIÓN DE $\bar{X}$, **TAN SÓLO** NECESITAMOS SABER LA MEDIA Y LA DESVIACIÓN TÍPICA POBLACIONALES.

ESTAS TRES DENSIDADES DE PROBABILIDAD DE AQUÍ ARRIBA TIENEN LA MISMA MEDIA Y DESVIACIÓN TÍPICA. A PESAR DE QUE TIENEN FORMAS DIFERENTES, CUANDO $n = 10$, LAS DISTRIBUCIONES MUESTRALES DE LA MEDIA, $\bar{X}$, SON CASI IDÉNTICAS.



La distribución t

POR MUY ASOMBROSO QUE SEA EL TEOREMA CENTRAL DEL LÍMITE, PRESENTA COMO MÍNIMO DOS PROBLEMAS:

UNO: DEPENDE DE UN TAMAÑO MUESTRAL MUY GRANDE.

DOS: PARA UTILIZARLO, NECESITAMOS CONOCER σ, LA DESVIACIÓN TÍPICA.

PERO, A MENUDO, LAS MUESTRAS SON PEQUEÑAS, Y NORMALMENTE SE DESCONOCE σ. SIN DUDA, EN EL CASO DE LOS PEPINILLOS NO TENEMOS **LA MENOR IDEA** DE CUÁNTO DISTA DE LA MEDIA EL TAMAÑO DE CADA UNO.

LO QUE PODEMOS HACER EN ESTE CASO ES **ESTIMAR** σ UTILIZANDO LA **DESVIACIÓN TÍPICA DE LA MUESTRA**, QUE, COMO RECORDARÁS, VIENE DADA POR LA FÓRMULA

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

ENTONCES, EN EL LUGAR DE LA VARIABLE ALEATORIA

$$z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

SUSTITUIMOS σ POR s, Y DEFINIMOS UNA **NUEVA VARIABLE ALEATORIA** t

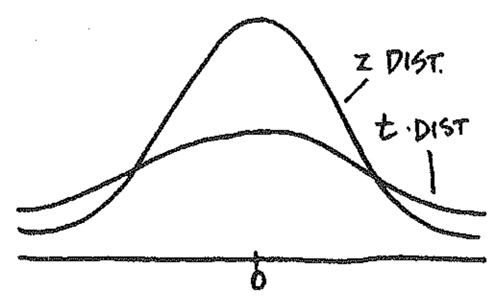
$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$$

PUEDES PENSAR EN LA VARIABLE ALEATORIA t COMO EN **LO MEJOR QUE SE PUEDE HACER DADAS LAS CIRCUNSTANCIAS.** SU DISTRIBUCIÓN RECIBE EL NOMBRE DE **t DE STUDENT**, PORQUE SU INVENTOR, **WILLIAM GOSSET**, LA PUBLICÓ CON EL SEUDÓNIMO DE «STUDENT».

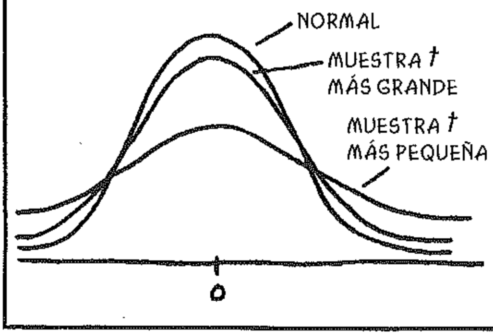
(GOSSET ERA EMPLEADO DE LA FÁBRICA DE CERVEZA **GUINNESS**, Y LE OBLIGARON, POR ALGÚN MOTIVO, A UTILIZAR SEUDÓNIMO.)

CON LA PRESUNCIÓN DE QUE LA **DISTRIBUCIÓN POBLACIONAL ORIGINAL ERA NORMAL**, O CASI NORMAL, «STUDENT» PUDO LLEGAR A UNA CONCLUSIÓN:

t TIENE MÁS DISPERSIÓN QUE z, ES MÁS «PLANA» QUE LA NORMAL. ESO ES PORQUE EL USO DE s INTRODUCE MAYOR INCERTIDUMBRE Y HACE QUE t SEA MENOS PRONUNCIADA QUE z.

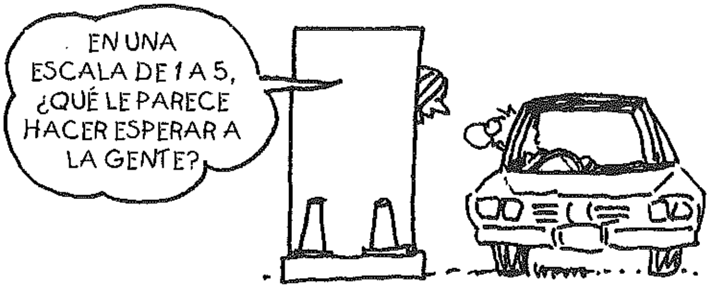


LA CANTIDAD DE DISPERSIÓN DEPENDE DEL **TAMAÑO MUESTRAL**. CUANTO MAYOR SEA LA MUESTRA, MÁS SEGUROS PODEMOS ESTAR DE QUE s SE ACERCA A σ, Y t SE ACERCA MÁS A z, LA NORMAL.



GOSSET CONSIGUIÓ CALCULAR TABLAS DE t PARA VARIOS TAMAÑOS MUESTRALES. VEREMOS CÓMO USARLAS EN EL PRÓXIMO CAPÍTULO.

EN ESTE CAPÍTULO HEMOS TRATADO UN PROBLEMA CLAVE DE LA **ESTADÍSTICA DEL MUNDO REAL**: CÓMO SELECCIONAR UNA **MUESTRA** DE UNA POBLACIÓN GRANDE PARA QUE EL ANÁLISIS ESTADÍSTICO SEA VÁLIDO. ADEMÁS DEL «ESTÁNDAR DORADO» DE LA MUESTRA ALEATORIA SIMPLE, TAMBIÉN HEMOS DESCRITO OTROS ESQUEMAS MUESTRALES QUE SE UTILIZAN POR SU EFICACIA, PRECIO Y ASPECTO PRÁCTICO.



A CONTINUACIÓN, DANDO POR SUPUESTO EL MUESTREO ALEATORIO SIMPLE, HEMOS VISTO LA **DISTRIBUCIÓN** DE VARIOS ESTADÍSTICOS MUESTRALES. ES DECIR, HEMOS CONTEMPLADO LA MUESTRA COMO **EXPERIMENTO ALEATORIO** Y ASÍ SUS ESTADÍSTICOS SE HAN CONVERTIDO EN **VARIABLES ALEATORIAS**.

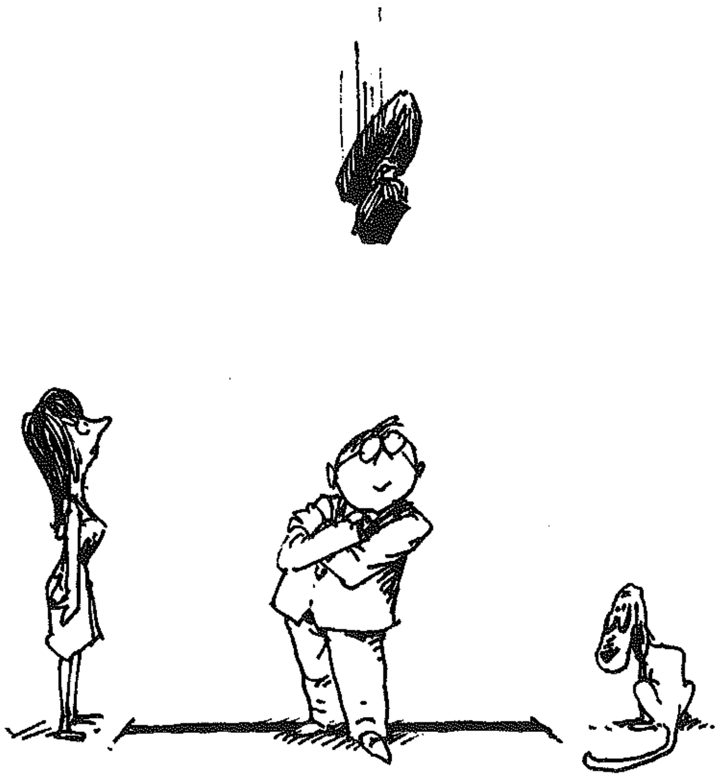
HEMOS DESCUBIERTO QUE LAS **PROPORCIONES** MUESTRALES $\hat{p}$ TENÍAN UNA DISTRIBUCIÓN MÁS O MENOS NORMAL, MIENTRAS QUE LAS DE LA **MEDIA** MUESTRAL $\bar{X}$ DEPENDÍAN DEL TAMAÑO DE LA MUESTRA. EN LAS MUESTRAS DE MAYOR TAMAÑO, LA DISTRIBUCIÓN ERA APROXIMADAMENTE NORMAL, MIENTRAS QUE EN LAS DE MENOR TAMAÑO, UTILIZAMOS LA DISTRIBUCIÓN t DE STUDENT.

* CON TÉ, ¡POR SUPUESTO!

EN LOS DOS CAPÍTULOS SIGUIENTES, VEREMOS CÓMO UTILIZAR ESTAS DISTRIBUCIONES PARA HACER **INFERENCIAS ESTADÍSTICAS**: DADA UNA SOLA OBSERVACIÓN, COMO UN SONDEO DE OPINIÓN, ¿CÓMO USAMOS NUESTRO CONOCIMIENTO DE $\hat{p}$ Y $\bar{X}$ PARA EVALUARLO?

♦ Capítulo 7 ♦

INTERVALOS DE CONFIANZA



EN EL CAPÍTULO ANTERIOR ESTUDIAMOS EL MUESTREO. COMENZANDO CON UNA POBLACIÓN GRANDE, IMAGINAMOS TOMAR MUCHAS MUESTRAS Y DEDUJIMOS LA DISTRIBUCIÓN DE ALGUNOS ESTIMADORES MUESTRALES.

EN ESTE CAPÍTULO, HAREMOS LO CONTRARIO. CON **UNA** MUESTRA, NOS PLANTEAMOS LA SIGUIENTE PREGUNTA: ¿QUÉ SISTEMA ALEATORIO HA GENERADO SUS ESTADÍSTICOS?

Esto representa un cambio en nuestra forma de pensar: del razonamiento deductivo a la **inducción**.

En el **razonamiento deductivo** vamos de una hipótesis a una conclusión: «Si Lord Fastback cometiera un asesinato, limpiaría las huellas dactilares de la pistola.»

El **razonamiento inductivo**, por el contrario, discurre **hacia atrás**, desde un conjunto de observaciones a una hipótesis razonable:

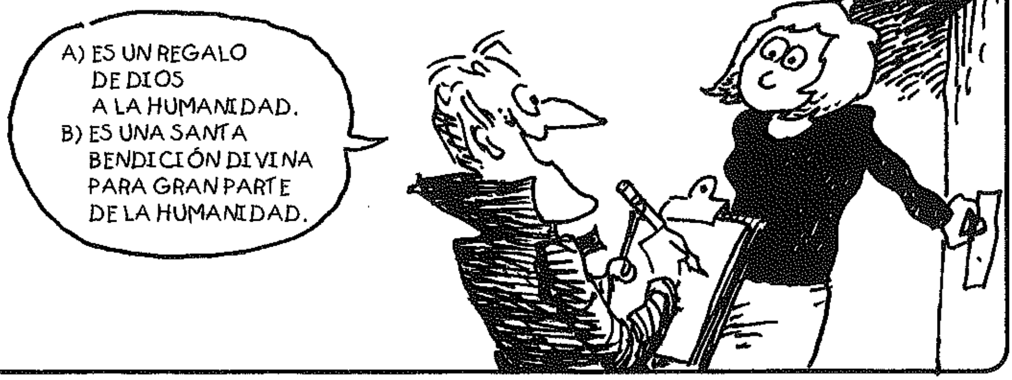
La ciencia, también la estadística, es de algún modo un trabajo detectivesco. Empezamos con un conjunto de observaciones, y nos preguntamos qué se puede decir de los sistemas que las generaron.

LA ESTIMACIÓN

CON INTERVALOS DE CONFIANZA

ES UNA DE LAS FORMAS MÁS EFECTIVAS DE INFERENCIA ESTADÍSTICA, Y SE PUEDE VER A DIARIO ANTES DE UNAS **ELECCIONES**...

EN UNAS ELECCIONES RECIENTES, EN ALGÚN LUGAR, EL SENADOR **ASTUTO** ENCARGA UN SONDEO DE OPINIÓN A LA COMPAÑÍA **GRANDES INVESTIGACIONES HOLMES**. EL ENCUESTADOR HOLMES TOMA UNA MUESTRA ALEATORIA SIMPLE DE 1.000 VOTANTES Y LES PREGUNTA QUÉ OPINIÓN LES MERECE ASTUTO.



DESPUÉS DE CENSURAR LOS COMENTARIOS DE UNAS CUANTAS OBSERVACIONES EXTREMAS GRUÑONAS, HOLMES CREE QUE **550** VOTANTES ESTÁN A FAVOR DE SU CLIENTE, EL SENADOR ASTUTO.

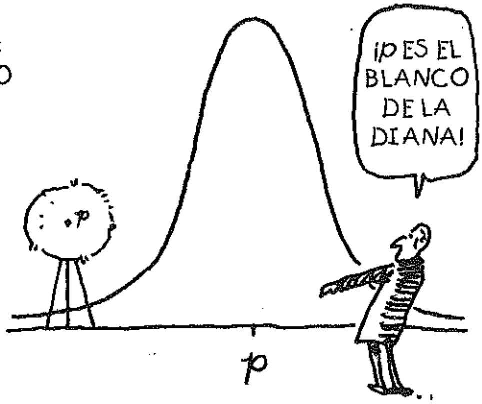
ESTA ES LA ÚNICA OBSERVACIÓN.

CUANDO ASTUTO SE CALMA, HOLMES LE EXPLICA LO QUE QUIERE DECIR CON **UN 95% DE SEGURIDAD**: SABE QUE SU PROCEDIMIENTO DE ESTIMACIÓN TIENE UNA **PROBABILIDAD DE UN 95%** DE PRODUCIR UN INTERVALO EN EL QUE SE ENCUENTRE p, ES DECIR, EN SU AÑOS COMO ENCUESTADOR, p ESTABA DENTRO DEL INTERVALO DE CONFIANZA QUE RODEA EL VALOR OBSERVADO, $\hat{p}$, EN UN **95%** DE LAS OCASIONES.


EL SENADOR ASTUTO AÚN ESTÁ ALGO CONFUSO, ASÍ QUE HOLMES LE DA UNA CLASE DE **tiro al arco**.
¡DISPARA! LO QUE SEA QUE ME QUITE DE LA CABEZA LA ESTADÍSTICA DE LAS NARICES.
VAMOS A CONSIDERAR A UNA ARQUERA QUE DISPARA A UNA DIANA. SUPONGAMOS QUE DA EN EL BLANCO DE 10 CENTÍMETROS UN 95% DE LAS VECES QUE DISPARA. ES DECIR, SÓLO UNA FLECHA DE CADA 20 NO DA EN EL BLANCO.
Y DETRÁS DE LA DIANA ENCONTRAMOS A UN VALEROSO DETECTIVE, QUE NO VE EL BLANCO. LA ARQUERA DISPARA UNA SOLA FLECHA.
EL DETECTIVE CONOCE EL NIVEL DE HABILIDAD DE LA ARQUERA Y DIBUJA UN CÍRCULO CON RADIO DE 10 CENTÍMETROS ALREDEDOR DE LA FLECHA. ¡TIENE UNA **SEGURIDAD DE UN 95%** DE QUE EN ESE CÍRCULO SE ENCUENTRA EL CENTRO DE LA DIANA!
10 CM
HA RAZONADO QUE SI DIBUJABA CÍRCULOS DE 10 CENTÍMETROS DE RADIO ALREDEDOR DE **MUCHAS FLECHAS**, EL BLANCO SE ENCONTRARÍA DENTRO DE ESOS CÍRCULOS UN 95% DE LAS VECES.
CENTRO AUTÉNTICO
X = FLECHA
(LOS PROBABILISTAS UTILIZAN EL TÉRMINO **ESTOCÁSTICO** PARA DESCRIBIR LOS MODELOS ALEATORIOS. PROVIENE DEL GRIEGO **STOCHAZESTHAI**, QUE SIGNIFICA APUNTAR A UN OBJETIVO, O ADIVINAR, DE **STOCHOS**, OBJETIVO O DIANA.)

AHORA, HOLMES TRADUCE LA LECCIÓN DE TIRO AL ARCO AL LENGUAJE QUE DESARROLLAMOS EN EL CAPÍTULO ANTERIOR.

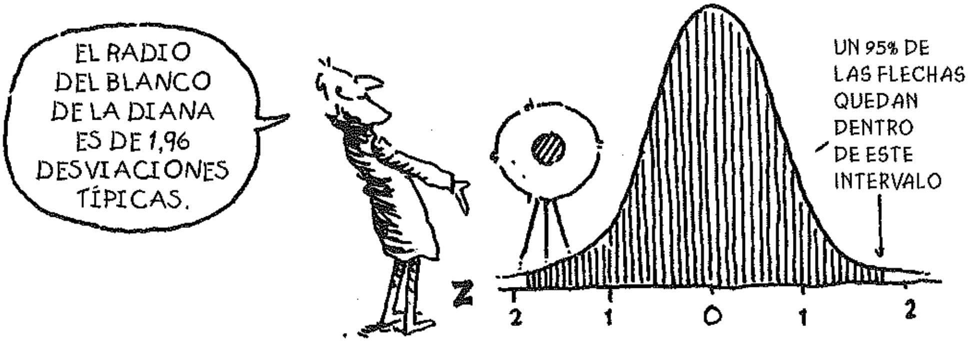
Primer paso: DISPARAR MUCHAS FLECHAS. CON UN CÁLCULO DE PROBABILIDAD SE DESCUBRE EL TAMAÑO DEL BLANCO DE LA «DIANA». LAS ESTIMACIONES DE $\hat{p}$ SON NUESTRAS FLECHAS. YA HEMOS VISTO QUE LA DISTRIBUCIÓN MUESTRAL DE $\hat{p}$ ES CASI **NORMAL**, CON MEDIA p Y DESVIACIÓN TÍPICA



$$\sigma(\hat{p}) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

COMO LA CURVA ES NORMAL, UTILIZAMOS LA TRANSFORMACIÓN z Y UNA TABLA ESTÁNDAR PARA ENCONTRAR LA AMPLITUD DEL INTERVALO EN EL QUE ESTÁN UN 95% DE LAS «FLECHAS». (DENTRO DE UNAS PÁGINAS, VEREMOS CÓMO HACERLO CON EXACTITUD.) LOS CÁLCULOS NOS DICEN QUE LA AMPLITUD ES DE **1,96** DESVIACIONES TÍPICAS:

$$0{,}95 = Pr\,(-1{,}96 \leq Z \leq 1{,}96)$$



AHORA TOCA UN POCO DE ÁLGEBRA. SEGÚN LA DEFINICIÓN DE LA TRANSFORMACIÓN z,

$$0{,}95 \simeq \Pr\left(-1{,}96 \leq \frac{\hat{p}-p}{\sigma(p)} \leq 1{,}96\right)$$

QUE SE CONVIERTE EN

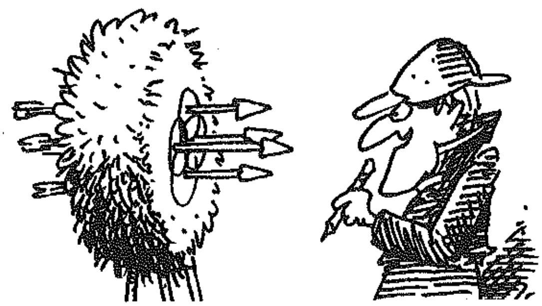
$$0{,}95 \simeq \Pr\left(p-1{,}96\sigma(p) \leq \hat{p} \leq p+1{,}96\sigma(p)\right)$$

QUE ES SÓLO OTRA FORMA DE DECIR QUE EL 95% DE LAS «FLECHAS» $\hat{p}$ QUEDAN ENTRE $p - 1{,}96\sigma(p)$ Y $p + 1{,}96\sigma(p)$.

DESDE NUESTRA POSICIÓN VEMOS LA DIANA POR DETRÁS. OTRA VUELTA DE TUERCA DEL ÁLGEBRA LO CONVIERTE EN

$$0{,}95 \simeq \Pr\left(\hat{p}-1{,}96\sigma(p) \leq p \leq \hat{p}+1{,}96\sigma(p)\right)$$

AQUÍ ESTAMOS DIBUJANDO CÍRCULOS ALREDEDOR DE MUCHAS FLECHAS (ES DECIR, ESTABLECIENDO INTERVALOS ALREDEDOR DE $\hat{p}$) Y DECIMOS QUE UN 95% DE ELLOS INCLUYE p.



PERO NOS ENCONTRAMOS CON UN PROBLEMILLA... **NO SABEMOS CUÁNTO MIDE EN REALIDAD EL BLANCO DE LA DIANA**, PORQUE NO CONOCEMOS EL VALOR DE p, Y LA AMPLITUD DEBE SER UN MÚLTIPLO DE $\sigma(p)$.

ASÍ QUE NOS LO AMAÑAMOS UN POCO Y UTILIZAMOS EL **ERROR ESTÁNDAR O TÍPICO** (SE)* DE $\hat{P}$.

$$SE(\hat{P}) = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

EN SU LUGAR... SE ACERCA BASTANTE... ES TODO LO QUE PODEMOS HACER... ¡Y HASTA TIENE UNA JUSTIFICACIÓN TEÓRICA!

* STANDARD ERROR = SE. [N.T.]

AHORA LA FÓRMULA ES

$$0{,}95 \simeq \Pr\left(\hat{p} - 1{,}96\,SE(\hat{p}) \leq p \leq \hat{p} + 1{,}96\,SE(\hat{p})\right)$$

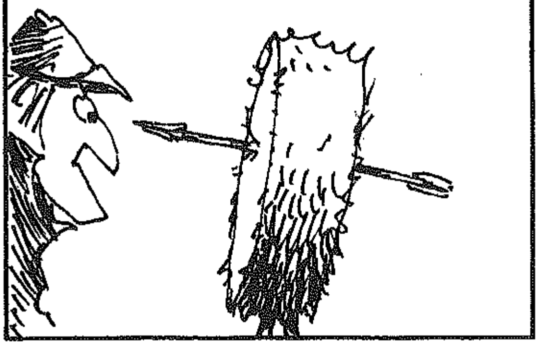
DE NUEVO, ESTA ECUACIÓN DESCRIBE LA PROBABILIDAD DE QUE LA AUTÉNTICA PROPORCIÓN DE LA POBLACIÓN FIJADA QUEDE DENTRO DEL **INTERVALO ALEATORIO**

$$(\hat{p} - 1{,}96\,SE(\hat{p}),\ \hat{p} + 1{,}96\,SE(\hat{p})).$$

SI LLEVÁSEMOS A CABO REPETIDAS MUESTRAS, ESTOS INTERVALOS INCLUIRÍAN p EN UN 95% DE LAS OCASIONES.

YA HEMOS HECHO EL CÁLCULO DE PROBABILIDADES Y HA LLEGADO LA HORA DEL...

Segundo paso: EL TRABAJO DETECTIVESCO. EN UNA ENCUESTA REAL, HOLMES SÓLO LLEVA A CABO UNA MUESTRA ALEATORIA SIMPLE DE 1.000 VOTOS, DESCUBRE QUE $\hat{p} = 0{,}550$, Y QUIERE INFERIR EL VALOR DE p.



ASÍ QUE UTILIZA EL PRIMER PASO PARA CALCULAR

$$SE(\hat{p}) = \sqrt{\frac{(0{,}55)(0{,}45)}{1000}} = 0{,}0157$$

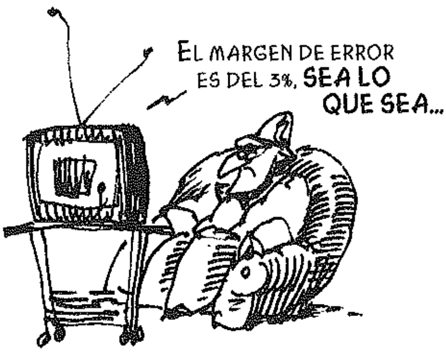
Y LLEGA A LA CONCLUSIÓN DE QUE PODEMOS ESTAR UN 95% SEGUROS DE QUE p SE ENCUENTRA EN EL INTERVALO

$$\hat{p} \pm 1{,}96\,SE(\hat{p})$$

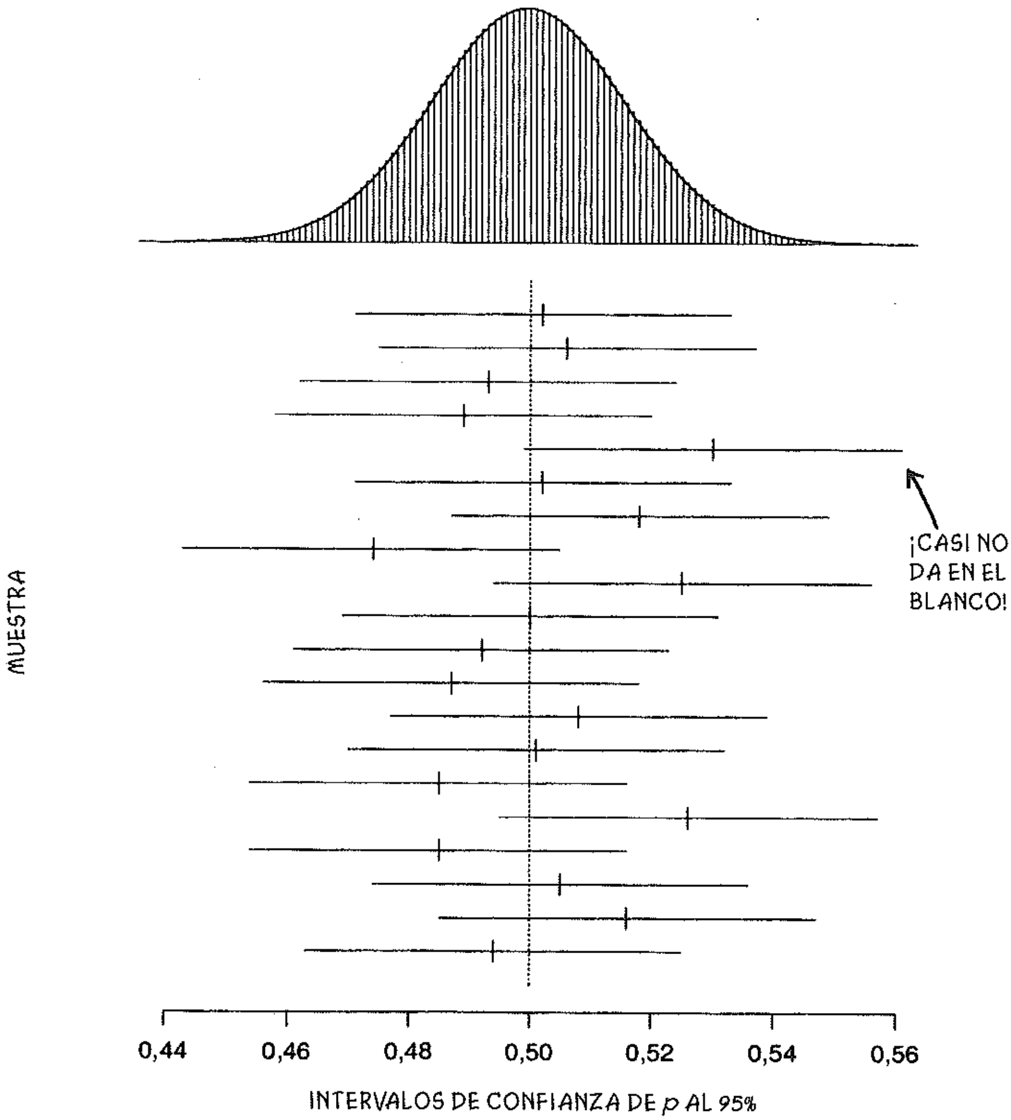
$$= 0{,}550 \pm (1{,}96)(0{,}0157)$$

$$= \mathbf{0{,}550 \pm 0{,}031}$$

ESTO ES LO QUE QUIEREN DECIR LAS ENCUESTAS CUANDO SE REFIEREN A SU «MARGEN DE ERROR». EN ESTE CASO, HOLMES VIO QUE $0{,}519 \leq p \leq 0{,}581$, EN OTRAS PALABRAS, QUE $p = 55\%$ CON UN 3% DE MARGEN DE ERROR. (NORMALMENTE, LAS ENCUESTAS UTILIZAN UN 95% DE CONFIANZA.)



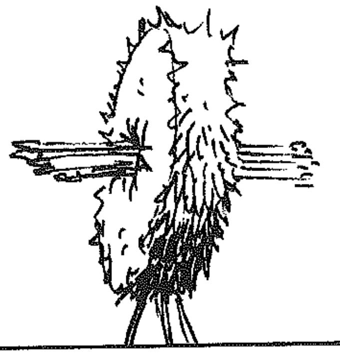
ESTA PÁGINA MUESTRA LOS RESULTADOS DE UNA SIMULACIÓN POR ORDENADOR DE VEINTE MUESTRAS DE TAMAÑO $n = 1.000$. SUPONEMOS QUE EL VALOR REAL DE $p = 0,5$. EN LA PARTE SUPERIOR PUEDES VER LA DISTRIBUCIÓN MUESTRAL DE $\hat{p}$ (NORMAL, CON MEDIA p Y $\sigma = \sqrt{\frac{p(1-p)}{n}}$). EN LA PARTE INFERIOR SE ENCUENTRAN LOS INTERVALOS DE CONFIANZA DE CADA MUESTRA, AL 95%. COMO MEDIA, UNO DE CADA VEINTE (O UN 5%) DE ESTOS INTERVALOS NO INCLUIRÁ EL PUNTO $p = 0,5$.



A PESAR DE QUE UN 95% DE CONFIANZA ESTÁ BASTANTE BIEN PARA LOS SONDEOS DE LA PRENSA, NO ES LO BASTANTE BUENO PARA EL **SENADOR ASTUTO**. ¡ÉL QUIERE UN 99%!

¿CÓMO SE PUEDE AUMENTAR LA CONFIANZA? SI USAMOS LA DIANA DE TIRO AL ARCO, TENEMOS DOS FORMAS DE HACERLO: UNA SERÍA **AUMENTAR EL TAMAÑO DEL CÍRCULO** QUE DIBUJAMOS...

Y OTRA FORMA SERÍA EMPEZAR POR **MEJORAR LA PUNTERÍA** DE LA ARQUERA PARA QUE LAS FLECHAS QUEDARAN MÁS CERCA DEL BLANCO DE LA DIANA.

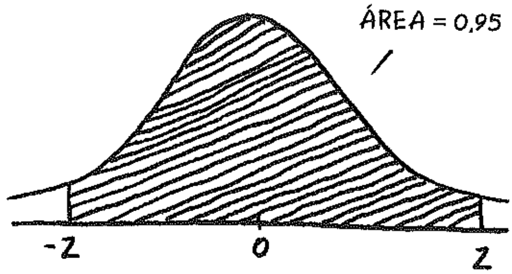


EL PRIMER MÉTODO EQUIVALE A **AMPLIAR EL INTERVALO DE CONFIANZA**. CUANTO MAYOR SEA EL MARGEN DE ERROR, MÁS SEGUROS PODEMOS ESTAR DE QUE EL VALOR REAL DE p SE ENCUENTRA EN EL INTERVALO.

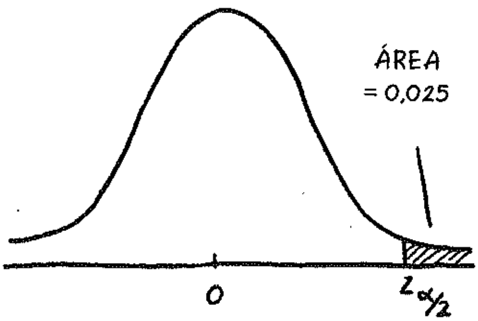
QUIZÁ HAYA LLEGADO LA HORA DE VER EXACTAMENTE CÓMO ENCONTRAR LOS EXTREMOS DE LOS INTERVALOS DE CONFIANZA...

AQUÍ, NORMALMENTE, AL NÚMERO IMPORTANTE LO LLAMAMOS α, Y MIDE LA DIFERENCIA ENTRE EL NIVEL DESEADO DE CONFIANZA Y CERTEZA. POR EJEMPLO, CUANDO EL NIVEL DE CONFIANZA ES 95%, O 0,95, α ES 0,05. ASÍ QUE HABLAMOS DEL NIVEL DE CONFIANZA $(1-\alpha)\cdot 100\%$.

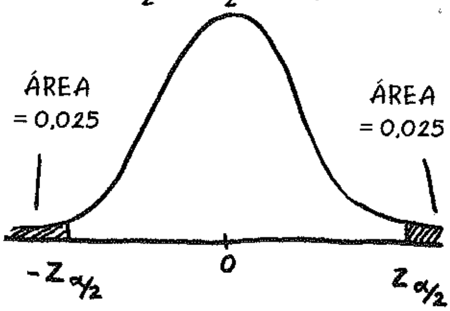
ENCONTRAR EL NIVEL DE CONFIANZA $(1-\alpha)\cdot 100\%$ IMPLICA OBSERVAR LA CURVA DE LA NORMAL TIPIFICADA Y BUSCAR LOS PUNTOS $\pm z$ ENTRE LOS QUE EL ÁREA ES $1-\alpha$.



ESTE PUNTO, LLAMADO $z_{\frac{\alpha}{2}}$, ES EL VALOR z MÁS ALLÁ DEL CUAL EL ÁREA ES $0{,}025 = \frac{\alpha}{2}$.



ESTO PASA PORQUE CORTAMOS LAS «COLAS» DE LOS DOS EXTREMOS DE LA CURVA, QUE TIENEN UN ÁREA TOTAL DE $\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$.



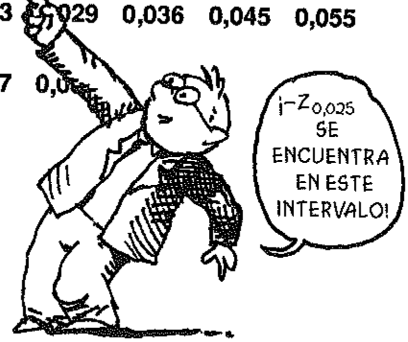
PODEMOS CALCULAR $z_{\alpha/2}$ DIRECTAMENTE A PARTIR DE LA TABLA DE LA NORMAL TIPIFICADA (PÁGINA 84). ES EL PUNTO CON LA PROPIEDAD

$$Pr(z \geq z_{\alpha/2}) = \frac{\alpha}{2}$$

EN ESTE CASO,

$$Pr(z \geq z_{0{,}025}) = 0{,}025$$

z	-2,5	-2,4	-2,3	-2,2	-2,1
F(z)	**0,006**	**0,008**	**0,011**	**0,014**	**0,018**
z	-2,0	-1,9	-1,8	-1,7	-1,6
F(z)	**0,023**	**0,029**	**0,036**	**0,045**	**0,055**
z	-1,5				
F(z)	**0,067**	**0,0**[illegible]			



AQUÍ TIENES UNA PEQUEÑA TABLA DE LOS VALORES CRÍTICOS DE VARIOS NIVELES DE CONFIANZA...

$1-\alpha$	0,80	0,90	0,95	0,99
α	0,20	0,10	0,05	0,01
$\alpha/2$	0,10	0,05	0,025	0,005
$z_{\frac{\alpha}{2}}$	1,28	1,64	1,96	2,58

PARA ESTABLECER UN INTERVALO DE CONFIANZA DEL 99%, UTILIZAMOS ESA TABLA Y ESCRIBIMOS

$$0{,}99 = Pr(\hat{p} - 2{,}58\,SE(\hat{p}) \leq p \leq \hat{p} + 2{,}58\,SE(\hat{p}))$$

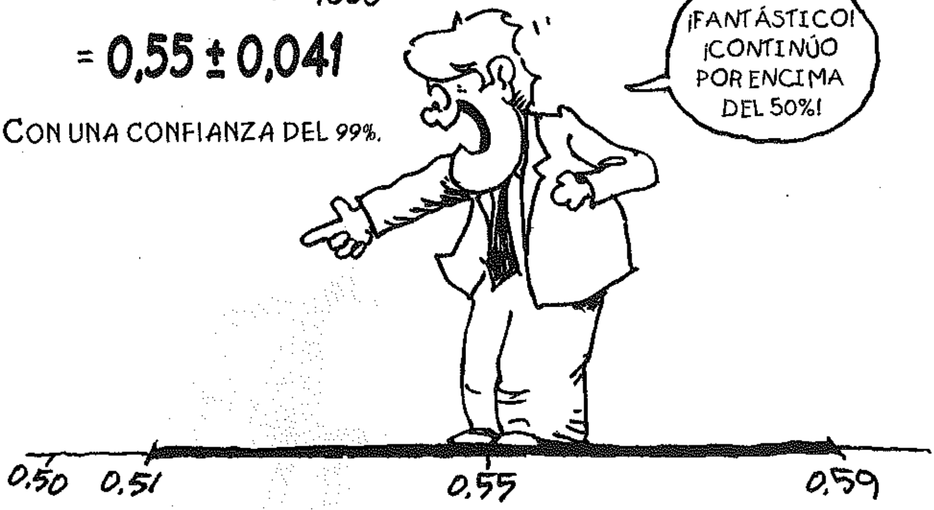
Y LO ABREVIAMOS COMO

$$p = \hat{p} \pm 2{,}58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

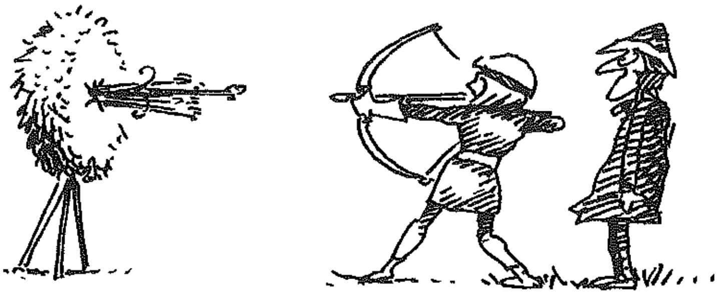
$$= 0{,}55 \pm 2{,}58\sqrt{\frac{(0{,}55)(0{,}45)}{1000}}$$

$$= \mathbf{0{,}55 \pm 0{,}041}$$

CON UNA CONFIANZA DEL 99%.

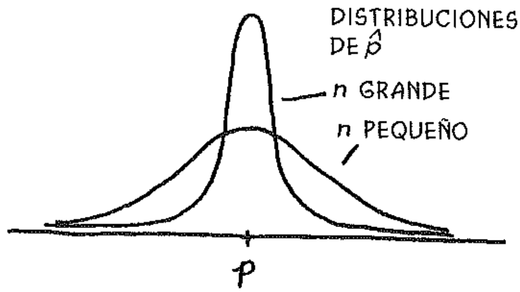


AUMENTAR EL INTERVALO ES UNA FORMA DE AUMENTAR LA CONFIANZA EN EL RESULTADO. COMO YA HEMOS DICHO ANTES, OTRA FORMA SERÍA DISPARAR LAS FLECHAS **CON MÁS PRECISIÓN**. SI SUPIÉRAMOS QUE LA ARQUERA CONSIGUE QUE UN 95% DE LAS FLECHAS DEN **A 1 CENTÍMETRO** DEL BLANCO DE LA DIANA, ¡NUESTRAS ESTIMACIONES PODRÍAN SER MUCHO MÁS PRECISAS!



¿CÓMO PODEMOS CONSEGUIRLO? ¡AUMENTANDO EL TAMAÑO DE LA MUESTRA! LA AMPLITUD DEL INTERVALO DE CONFIANZA DEPENDE DEL TAMAÑO MUESTRAL: EL INTERVALO TIENE LA FORMA $\hat{p} \pm$ E, EN LA QUE E, EL ERROR, VIENE DADO POR

$$E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



ASÍ QUE CUANTO MAYOR SEA n, EL ERROR SERÁ MENOR. (ES DECIR, SI MULTIPLICAMOS n POR CUATRO, LA AMPLITUD DEL INTERVALO SE REDUCE A LA MITAD.)

ASTUTO LE PIDE A HOLMES QUE LE DÉ UN ERROR PEQUEÑO Y MUCHA CONFIANZA. PONGAMOS UN 99% DE CONFIANZA CON E = ± 0,01. HOLMES CALCULA n.

$$n = \frac{z_{\frac{\alpha}{2}}^{2}\, p^*(1-p^*)}{E^2}$$

(p^* ES UNA APROXIMACIÓN A LA PROPORCIÓN REAL DE p; ¡RECUERDA QUE AÚN NO HEMOS REALIZADO EL MUESTREO!)

CON UNA SUPOSICIÓN CONSERVADORA DE $p^* = 0{,}5$, HOLMES DESCUBRE QUE

$$n = \frac{(2.58)^2(0{,}5)^2}{(0{,}01)^2}$$

$$= \frac{(6{,}65)(0{,}25)}{0{,}0001}$$

$$= \mathbf{16{,}641}$$

MIL VOTANTES DIERON UN 3% DE ERROR Y UNA CONFIANZA DE UN 95%. PARA CONSEGUIR UN ERROR DE UN 1% CON UNA CONFIANZA DEL 99%, ¡HOLMES TIENE QUE TOMAR UNA MUESTRA DE 16.641 VOTANTES!

ASÍ QUE LLEVAN A CABO LA ENCUESTA Y SE PRESENTAN A LAS ELECCIONES CON UN **99% DE CONFIANZA.**

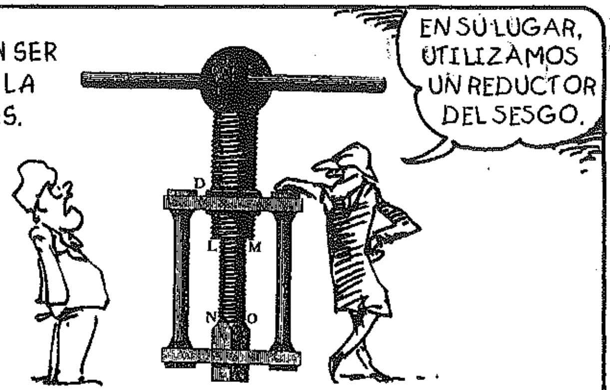
SIN EMBARGO... TODO ESTO DE LAS PROBABILIDADES SÓLO SIRVE **ANTES** DE UNAS ELECCIONES. ¡**DESPUÉS**, EL SENADOR ESTÁ **ELEGIDO** AL 100% O **NO ELEGIDO** AL 100%! Y A PESAR DE TODO, EL SENADOR ASTUTO PIERDE LAS ELECCIONES...

¡LO QUE HA PASADO ES QUE A LOS POLÍTICOS NO LOS ELIGEN LOS SONDEOS DE OPINIÓN!

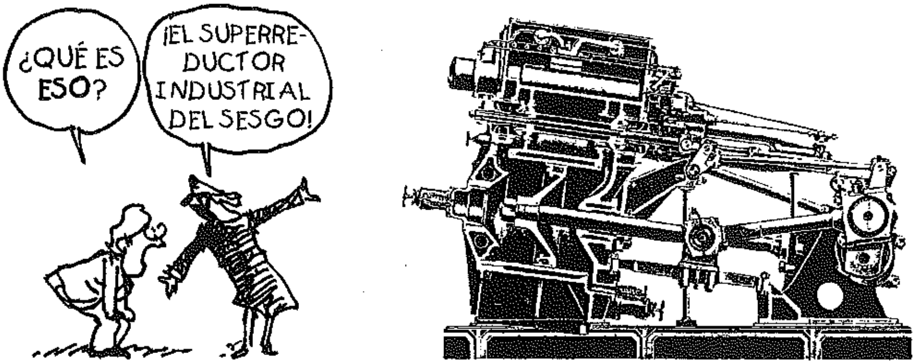
ALGUNOS DE LOS PROBLEMAS DE LOS SONDEOS FRENTE A LAS ELECCIONES SON:

NO EXISTE FORMA ALGUNA DE METERSE EN LA CABEZA DEL VOTANTE EN POTENCIA Y SABER SI VA A VOTAR, SI MIENTE O SI VA A CAMBIAR DE OPINIÓN ANTES DEL DÍA DE LAS ELECCIONES. INCLUSO CON MUESTRAS MAYORES NO SE PUEDE REDUCIR ESTE TIPO DE ERRORES.

Y A QUE ESTOS ERRORES PUEDEN SER MUY GRANDES, APENAS MERECE LA PENA PAGAR MUESTRAS MAYORES.



EN LAS ÚLTIMAS CINCO ELECCIONES PRESIDENCIALES DE LOS ESTADOS UNIDOS, EL SONDEO GALLUP HIZO UNA ENCUESTA CON MENOS DE 4.000 VOTANTES CADA VEZ. SIN EMBARGO, EN LAS CINCO ELECCIONES, LOS ERRORES DE LA ORGANIZACIÓN GALLUP EN SU PREDICCIÓN DE LOS RESULTADOS FUERON DE MENOS DE UN 2%.

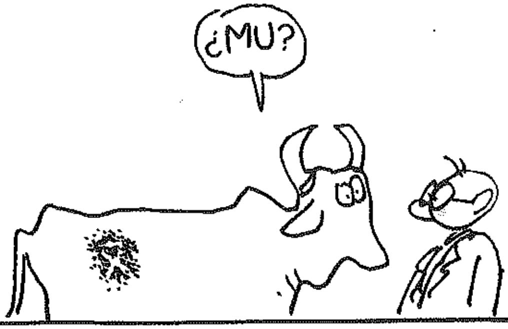


SU ÉXITO SE DEBE A LA UTILIZACIÓN DE ESTIMADORES QUE COMPENSAN LA FALTA DE RESPUESTA Y A LA DESESTIMACIÓN DE LOS VOTANTES QUE PROBABLEMENTE NO EJERCERÁN SU DERECHO EN LAS URNAS.

EN RESUMEN, PROPORCIÓN ESTIMADA = PROPORCIÓN REAL + SESGO + ERROR DE MUESTREO ALEATORIO.
INCLUSO LOS ENCUESTADORES TIENEN FONDOS LIMITADOS, ASÍ QUE DECIDEN SABIAMENTE GASTAR EL DINERO EN **REDUCIR EL SESGO** EN LUGAR DE INTENTAR AUMENTAR LA MUESTRA A MÁS DE 4.000 VOTANTES.

Intervalos de confianza de μ

HASTA AHORA, HEMOS VISTO INTERVALOS DE CONFIANZA DE UNA PROPORCIÓN p DE UNA POBLACIÓN. EXACTAMENTE EL MISMO RAZONAMIENTO FUNCIONA PARA LA MEDIA POBLACIONAL μ.



EN EL CAPÍTULO ANTERIOR (PÁGINA 105), VIMOS QUE LA DISTRIBUCIÓN DE LA MEDIA MUESTRAL $\bar{X}$ ES APROXIMADAMENTE **NORMAL**, CON EL CENTRO EN LA **AUTÉNTICA** MEDIA POBLACIONAL μ Y DESVIACIÓN TÍPICA $\sigma/\sqrt{n}$, EN LA QUE σ ES LA DESVIACIÓN TÍPICA DE LA POBLACIÓN. DE ESTE MODO, PARA UN VALOR n MUY GRANDE,

$$0{,}95 = Pr(-1{,}96 \leq Z \leq 1{,}96)$$

$$\approx Pr\left(-1{,}96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1{,}96\right)$$

DE NUEVO, YA QUE NO CONOCEMOS EL VALOR DE σ, LA SUSTITUIMOS POR s, LA DESVIACIÓN TÍPICA MUESTRAL:

$$0{,}95 \approx Pr\left(-1{,}96 \leq \frac{\bar{X}-\mu}{s/\sqrt{n}} \leq 1{,}96\right)$$

EL TÉRMINO $s/\sqrt{n}$ RECIBE EL NOMBRE DE **ERROR TÍPICO DE MUESTREO**, Y SE ESCRIBE $SE(\bar{X})$. LLEGAMOS A LA SIGUIENTE CONCLUSIÓN:

$$0{,}95 \approx Pr(\bar{X} - 1{,}96\,SE(\bar{X}) \leq \mu \leq \bar{X} + 1{,}96\,SE(\bar{X}))$$

DONDE

$$SE(\bar{X}) = \frac{s}{\sqrt{n}}$$

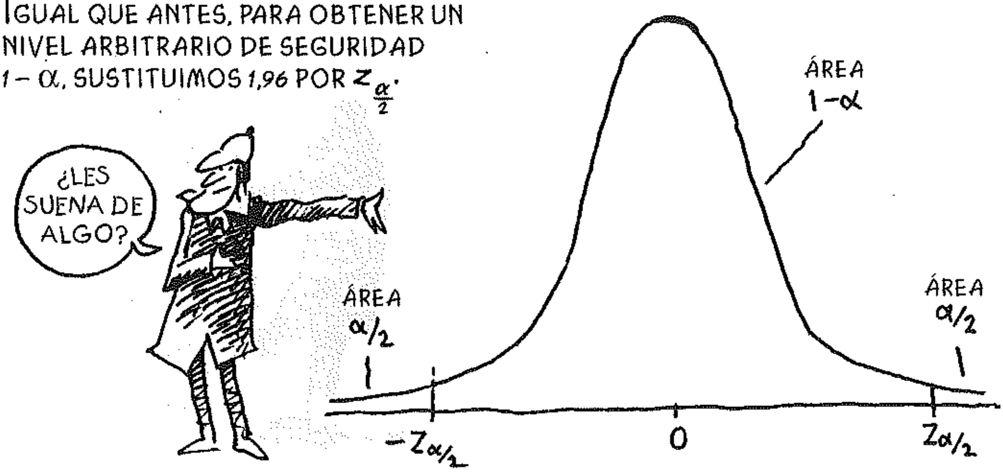
IGUAL QUE ANTES, HEMOS DESCUBIERTO QUE EL INTERVALO ALEATORIO

$$\bar{X} \pm 1.96\,SE(\bar{X})$$

INCLUYE LA MEDIA REAL, μ, CON UNA PROBABILIDAD DE 0,95... ASÍ QUE AHORA PODEMOS LLAMAR A SHERLOCK PARA QUE HAGA UNA INFERENCIA ESTADÍSTICA BASADA EN UNA SOLA MUESTRA DE TAMAÑO n Y MEDIA $\bar{x}$.

ÉL (Y NOSOTROS) ESTÁ UN 95% SEGURO DE QUE LA MEDIA μ SE ENCUENTRA EN EL INTERVALO $\bar{x} \pm 1,96SE(X)$.

IGUAL QUE ANTES, PARA OBTENER UN NIVEL ARBITRARIO DE SEGURIDAD $1-\alpha$, SUSTITUIMOS 1,96 POR $z_{\frac{\alpha}{2}}$.



VAMOS A REGRESAR A LOS DATOS DE LOS ESTUDIANTES DEL CAPÍTULO 2, SUPONIENDO QUE $n = 92$ ESTUDIANTES FUERA UNA MUESTRA ALEATORIA SIMPLE DE TODOS LOS ESTUDIANTES DEL ESTADO DE PENNSYLVANIA.

LA MEDIA MUESTRAL $\bar{x}$ ERA 145,2 LIBRAS Y LA DESVIACIÓN TÍPICA MUESTRAL s ERA 23,7. ASÍ QUE EL ERROR TÍPICO ES

$$SE(\bar{X}) = \frac{23{,}7}{\sqrt{92}} = 2{,}47$$

Y AHORA TENEMOS UNA CONFIANZA DEL 95% DE QUE EL PESO MEDIO DE TODOS LOS ESTUDIANTES DE ESE ESTADO QUEDA DENTRO DEL INTERVALO

$$\bar{x} \pm 1{,}96\,SE(\bar{X})$$
$$= 145{,}2 \pm (1{,}96)(2{,}47)$$
$$= 145{,}2 \pm 4{,}8 \text{ LIBRAS}$$

EN RESUMEN: EN UNA MUESTRA ALEATORIA SIMPLE (MAS) DE TAMAÑO GRANDE, EL INTERVALO DE CONFIANZA $(1-\alpha)\cdot 100\%$ ES:

MEDIA POBLACIONAL μ

$$\mu = \bar{x} \pm z_{\frac{\alpha}{2}} SE(\bar{x})$$

DONDE $SE(\bar{x}) = s/\sqrt{n}$

PROPORCIÓN POBLACIONAL, p

$$p = \hat{p} \pm z_{\frac{\alpha}{2}} SE(\hat{p})$$

DONDE $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

EL TAMAÑO DE LOS DOS INTERVALOS ESTÁ CONTROLADO POR EL NIVEL DE CONFIANZA $(1-\alpha)\cdot 100\%$ Y EL TAMAÑO DE LA MUESTRA, n.

La *t* de Student (¡otra vez!)

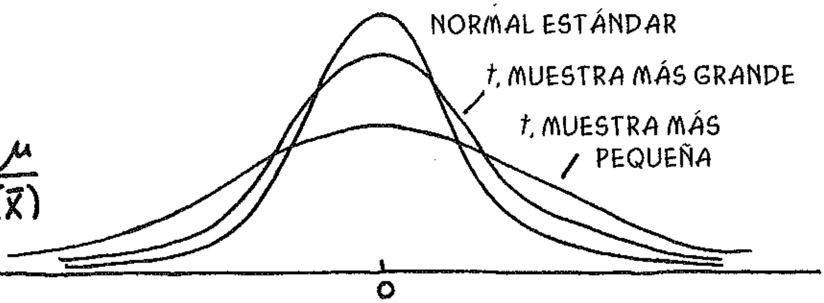
COMO YA VIMOS EN EL CAPÍTULO 6, EL ESTADÍSTICO

$$\frac{\bar{X}-\mu}{SE(\bar{X})}$$

SÓLO TIENE UNA DISTRIBUCIÓN APROXIMADAMENTE NORMAL CUANDO SE CALCULA UTILIZANDO UNA **MUESTRA MUY GRANDE**. PARA MUESTRAS MÁS PEQUEÑAS (n = 5, 10, 25...), ÉSE YA NO ES EL CASO Y TENEMOS QUE UTILIZAR LA t DE STUDENT.

VAMOS A OBSERVAR MÁS DE CERCA LA t. YA MENCIONAMOS QUE LA DISTRIBUCIÓN t ES MÁS DISPERSA QUE LA NORMAL, Y QUE LA CANTIDAD DE DISPERSIÓN DEPENDE DEL TAMAÑO DE LA MUESTRA.

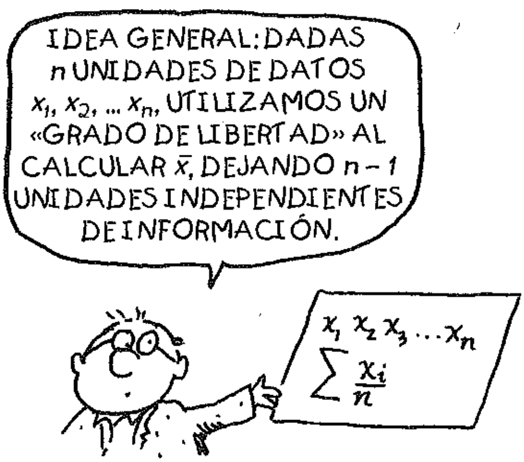
$$t = \frac{\bar{X}-\mu}{SE(\bar{X})}$$



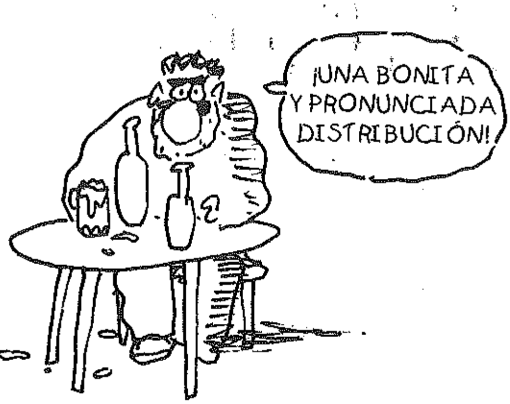
LO QUE HIZO SU DESCUBRIDOR, GOSSET, FUE **CUANTIFICAR** ESTA RELACIÓN. SI n ES EL TAMAÑO MUESTRAL, DECÍA, ENTONCES LLAMAREMOS A $n-1$ EL NÚMERO DE

grados de libertad

DE LA MUESTRA.

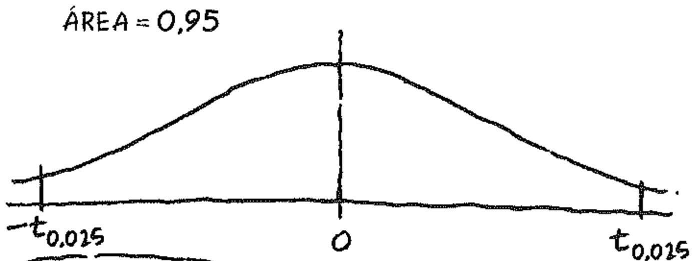


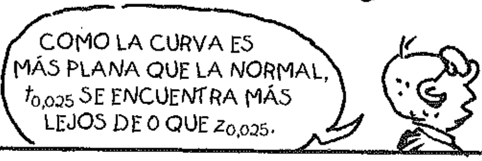
GOSSET CALCULÓ TABLAS DE LA DISTRIBUCIÓN t PARA DIFERENTES TAMAÑOS MUESTRALES; ES DECIR, GRADOS DE LIBERTAD. Y NOSOTROS REPETIMOS, A **MÁS GRADOS DE LIBERTAD**, MÁS CERCA SE ENCUENTRA t DE LA **NORMAL TIPIFICADA**.



SI CONOCEMOS EL TAMAÑO MUESTRAL n, ESCOGEMOS LA DISTRIBUCIÓN t CON $n-1$ GRADOS DE LIBERTAD.

IGUAL QUE CON LA DISTRIBUCIÓN z (LA NORMAL TIPIFICADA), OBTENEMOS UN NIVEL DE CONFIANZA DEL 95% AL ENCONTRAR EL VALOR CRÍTICO $t_{0,025}$ MÁS ALLÁ DEL CUAL EL ÁREA DE LA CURVA ES 0,025.





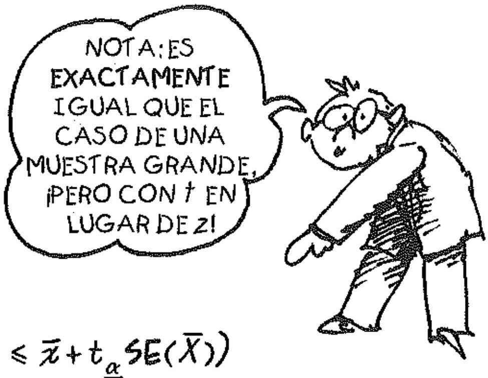
PARA OBTENER UN INTERVALO DE CONFIANZA $(1-\alpha)\cdot 100\%$, ENCONTRAMOS UN VALOR CRÍTICO $t_{\frac{\alpha}{2}}$ TAL QUE $Pr(t \geq t_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. AQUÍ TIENES UNA PEQUEÑA TABLA DE VALORES CRÍTICOS PARA LA DISTRIBUCIÓN t:

	$1-\alpha$	0,80	0,90	0,95	0,99
	α	0,20	0,10	0,05	0,01
	$\alpha/2$	0,10	0,05	0,025	0,005
GRADOS DE LIBERTAD	1	3,09	6,31	12,71	63,66
	10	1,37	1,81	2,23	4,14
	30	1,31	1,70	2,04	2,75
	100	1,29	1,66	1,98	2,63
	∞	1,28	1,65	1,96	2,58

CADA COLUMNA REPRESENTA UN NIVEL FIJO DE CONFIANZA, CON NÚMEROS CRECIENTES DE GRADOS DE LIBERTAD. CUANTOS MÁS GRADOS DE LIBERTAD, MÁS SE ACERCA EL VALOR CRÍTICO A $z_{\alpha/2}$, EL VALOR CRÍTICO DE LA DISTRIBUCIÓN NORMAL.

DERIVAMOS LA AMPLITUD DE NUESTRO INTERVALO DE CONFIANZA DIRECTAMENTE DE LA DEFINICIÓN DE t:

$$t = \frac{\bar{X}-\mu}{SE(\bar{X})}$$



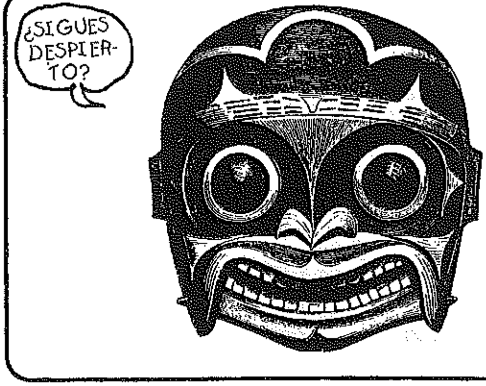
ENTONCES, PARA OBTENER UN NIVEL DE CONFIANZA $(1-\alpha)\cdot 100\%$,

$$(1-\alpha) = Pr(\bar{x} - t_{\frac{\alpha}{2}} SE(\bar{X}) \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} SE(\bar{X}))$$

DE LO QUE INFERIMOS: DADA UNA SOLA MUESTRA DE TAMAÑO n Y MEDIA $\bar{x}$, PODEMOS ESTAR $(1-\alpha)\cdot 100\%$ SEGUROS DE QUE LA MEDIA POBLACIONAL μ QUEDA EN EL INTERVALO

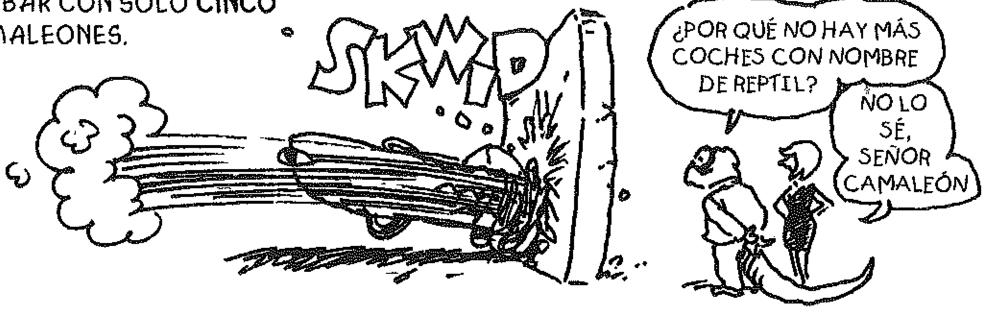
$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}} SE(\bar{x})$$

DONDE $SE(\bar{x}) = s/\sqrt{n}$ Y $t_{\frac{\alpha}{2}}$ ES EL VALOR CRÍTICO DE LA DISTRIBUCIÓN t CON $n-1$ GRADOS DE LIBERTAD.



NOTA: SI HABLAMOS CON EXACTITUD, LA DERIVACIÓN DE LA DISTRIBUCIÓN t DEPENDE DE LA PRESUNCIÓN DE QUE LA MUESTRA ERA DE UNA POBLACIÓN NORMAL. EN LA PRÁCTICA, LOS INTERVALOS DE CONFIANZA BASADOS EN t DAN RESULTADOS BASTANTE BUENOS, INCLUSO CUANDO LA DISTRIBUCIÓN POBLACIONAL TIENE UNA FORMA SÓLO APROXIMADAMENTE PARECIDA A UNA MONTAÑA.

Ejemplo: SUPONGAMOS QUE LA **CAMALEÓN MOTORS** TIENE QUE HACER PRUEBAS DE CHOQUE CON SUS COCHES PARA DETERMINAR EL COSTE MEDIO DE REPARACIÓN TRAS UNA COLISIÓN FRONTAL A UNOS 20 KILÓMETROS POR HORA. ¡RESULTA MUY CARO! ASÍ QUE DECIDEN PROBAR CON SÓLO **CINCO** CAMALEONES.



LOS DATOS DE LOS DESPERFECTOS EN DÓLARES SON 150, 400, 720, 500 Y 930.

LA MEDIA MUESTRAL:

$\bar{x}$ = **540 DÓLARES**

LA DESVIACIÓN ESTÁNDAR:

s = **299 DÓLARES**

PUEDES COMPROBAR s CON UNA CALCULADORA. ES

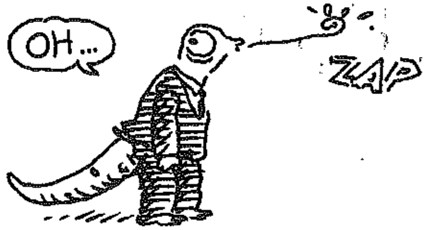
$$\sqrt{\frac{1}{4}\left((150-540)^2+(400-540)^2+(720-540)^2+(500-540)^2+(930-540)^2\right)}$$

ASÍ QUE, ¿DÓNDE PODEMOS SITUAR LA MEDIA CON UNA CONFIANZA DEL 95%? ENCONTRAMOS NUESTRO VALOR CRÍTICO $t_{0,025}$ CON 4 GRADOS DE LIBERTAD:

	$1-\alpha$	0,80	0,90	0,95	0,99
	α	0,20	0,10	0,05	0,01
	$\alpha/2$	0,10	0,05	0,025	0,005
GRADOS DE LIBERTAD	1	3,09	6,31	12,71	63,66
	2	1,89	2,92	4,30	9,92
	3	1,64	2,35	3,18	5,84
	4	1,53	2,13	2,78	4,60
	5	1,48	2,01	2,57	4,03

Y ALLÁ VAMOS:

$$\mu = \bar{x} \pm 2.78\, s/\sqrt{n}$$
$$= 540 \pm 2.78\left(299/\sqrt{5}\right)$$
$$= \mathbf{540 \pm 372}$$



ASÍ QUE LO MEJOR QUE PODEMOS DECIR CON UN 95% DE CONFIANZA ES QUE EL COSTE MEDIO POR REPARACIÓN DE DAÑOS ESTARÁ ENTRE 168 Y 912 DÓLARES.

PARA CALCULAR ESTE INTERVALO DE CONFIANZA UTILIZANDO LA t DE STUDENT, HEMOS REALIZADO UNA **PRESUNCIÓN NO COMPROBADA**: HEMOS DADO POR HECHO QUE LOS COSTES DE REPARACIÓN DE LOS COCHES TIENEN UNA **DISTRIBUCIÓN APROXIMADAMENTE NORMAL**, ES DECIR, QUE SI HICIÉSEMOS CHOCAR 1.000 CAMALEONES, EL HISTOGRAMA DE LOS COSTES DE REPARACIÓN SERÍA SIMÉTRICO Y CON FORMA DE MONTAÑA. **NO PODEMOS SABERLO** CON TAN SÓLO CINCO DATOS... PERO QUIZÁ AÑOS DE EXPERIENCIA CON ANTIGUOS MODELOS HAN PRODUCIDO HISTOGRAMAS DE LOS COSTES DE REPARACIÓN DE LA PARTE FRONTAL DEL COCHE CON DISTRIBUCIÓN NORMAL: UNA INFORMACIÓN QUE APOYARÍA NUESTRA DECISIÓN DE UTILIZAR LA t DE STUDENT.

PARA RESUMIR (!), YA TENEMOS TRES SIMPLES RECETAS PARA ENCONTRAR LOS INTERVALOS DE CONFIANZA. PARA LA PROPORCIÓN, O LA MEDIA DE MUESTRAS DE GRAN TAMAÑO, BUSCAMOS $z_{\frac{\alpha}{2}}$ EN UNA TABLA NORMAL. PARA LA MEDIA DE UNA MUESTRA PEQUEÑA (DIGAMOS $n \leq 30$), BUSCAMOS $t_{\frac{\alpha}{2}}$ EN LA TABLA t.

EN TODOS LOS CASOS, LA AMPLITUD DEL INTERVALO ES EL VALOR CRÍTICO POR EL ERROR ESTÁNDAR:

$$z_{\frac{\alpha}{2}} SE(\hat{p}) \qquad z_{\frac{\alpha}{2}} SE(\overline{X}) \qquad t_{\frac{\alpha}{2}} SE(\overline{X})$$

Y TODOS Y CADA UNO DE ESOS ERRORES ESTÁNDAR SON PROPORCIONALES A AQUEL NÚMERO MÁGICO:

$$\frac{1}{\sqrt{n}}.$$

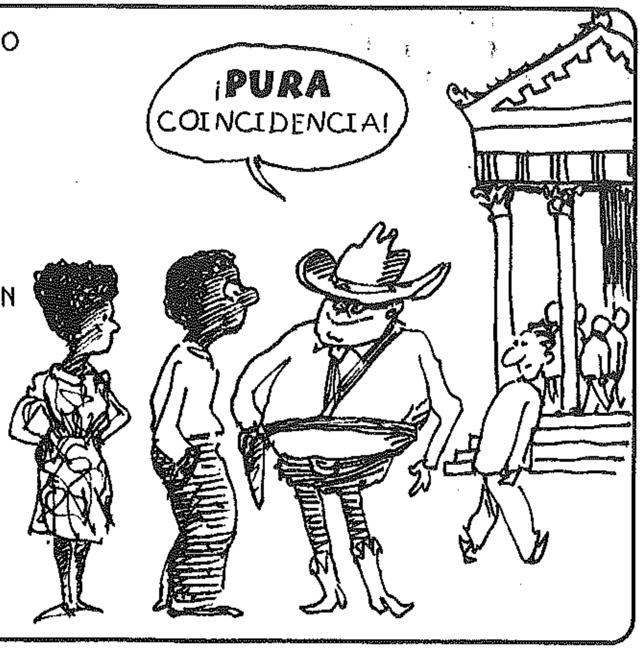
♦ Capítulo 8 ♦

CONTRASTE DE HIPÓTESIS

AHORA NOS ADENTRAREMOS EN NUEVOS TERRENOS... LA POLÍTICA, LA ECONOMÍA, Y LAS CIENCIAS EXACTAS Y LAS NO TAN EXACTAS ABUSAN MUY A MENUDO DE ESTAS PRUEBAS DE SIGNIFICACIÓN. Y PARA DESCUBRIR EL POR QUÉ, DEBEMOS PREGUNTARNOS:

«¿ESTAS OBSERVACIONES PUEDEN HABERSE DADO POR CASUALIDAD?»

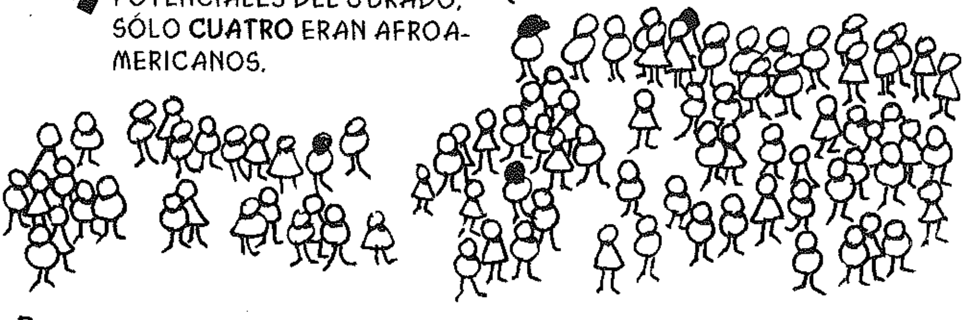
EMPEZAMOS CON UN EJEMPLO DE LA LEY: UN CONJUNTO DE DIFERENTES CASOS, QUE SE DIERON EN EL SUR DE LOS ESTADOS UNIDOS ENTRE 1960 Y 1980, EN LOS QUE TESTIGOS EXPERTOS DENUNCIABAN LA EXISTENCIA DE **DISCRIMINACIÓN RACIAL** EN LA SELECCIÓN DEL JURADO.



EN TEORÍA, LAS LISTAS DE JURADOS SE ELABORAN DE FORMA ALEATORIA A PARTIR DE UNA LISTA DE CIUDADANOS SUSCEPTIBLES DE SER ELEGIDOS. SIN EMBARGO, DURANTE LAS DÉCADAS DE LOS 50 Y 60, EN LOS ESTADOS DEL SUR, HABÍA MUY POCOS AFROAMERICANOS EN LAS LISTAS DE JURADOS. ASÍ QUE ALGUNOS ABOGADOS DE LA DEFENSA PUSIERON EN TELA DE JUICIO LOS VEREDICTOS. EN LA APELACIÓN, UN TESTIGO EXPERTO EN ESTADÍSTICA PRESENTÓ ESTA PRUEBA:

1) EL **50%** DE LOS CIUDADANOS SUSCEPTIBLES DE SER ELEGIDOS ERAN AFROAMERICANOS.

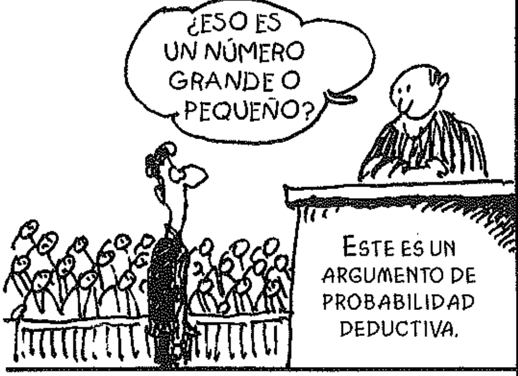
2) EN UNA LISTA DE 80 MIEMBROS POTENCIALES DEL JURADO, SÓLO **CUATRO** ERAN AFROAMERICANOS.



¿PUEDE SER ESTO FRUTO DE LA **PURA CASUALIDAD**?

PARA COMPLETAR EL ARGUMENTO, SUPONGAMOS QUE LA ELECCIÓN DEL JURADO POTENCIAL FUE **ALEATORIA**, ENTONCES, EL NÚMERO DE AFROAMERICANOS DE LA LISTA DE 80 PERSONAS SERÍA LA VARIABLE ALEATORIA **BINOMIAL X** CON $n = 80$ PRUEBAS Y $p = 0,5$.

ENTONCES, LA POSIBILIDAD DE FORMAR UN JURADO CON HASTA 4 MIEMBROS AFROAMERICANOS ES $Pr(x \leq 4)$ O SEA, ALREDEDOR DE 0,000000000000000014 (!).



COMO LA PROBABILIDAD ES TAN PEQUEÑA, LA LISTA EN CUESTIÓN, CON SÓLO CUATRO AFROAMERICANOS, RESULTA UNA **PRUEBA DE PESO** CONTRA LA **HIPÓTESIS** DE LA SELECCIÓN ALEATORIA.

PARA HABLAR EN TÉRMINOS MÁS FAMILIARES, EL ESTADÍSTICO SEÑALA QUE ESTA PROBABILIDAD ES MENOR QUE LA DE CONSEGUIR TRES **ESCALERAS REALES** SEGUIDAS EN EL POKER.

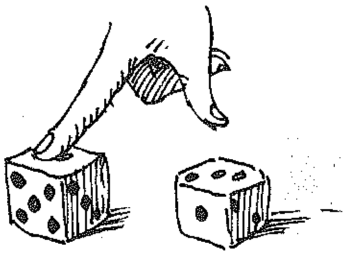
POR ESO EL JUEZ **RECHAZA LA HIPÓTESIS** DE LA SELECCIÓN ALEATORIA.

REPASEMOS, DE NUEVO, TODO EL PROCESO, PARA ENTENDER **LOS CUATRO PASOS** DEL CONTRASTE DE HIPÓTESIS ESTADÍSTICO.

Paso n.° 1. FORMULAR TODAS LAS HIPÓTESIS

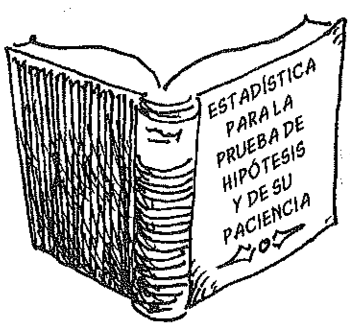
Ho, LA **HIPÓTESIS NULA**, LAS OBSERVACIONES SON EL RESULTADO DE LA PURA **CASUALIDAD**.

Ha, LA **HIPÓTESIS ALTERNATIVA**, HAY UN EFECTO REAL, LAS OBSERVACIONES SON EL RESULTADO DE ESTE EFECTO REAL, ADEMÁS DE LA VARIACIÓN CASUAL.



Paso n.° 2. ANÁLISIS ESTADÍSTICO

ENCONTRAR UN ESTADÍSTICO QUE CONFIRME LA EVIDENCIA CONTRA LA HIPÓTESIS NULA.



EN EL CASO DEL JURADO, H_0 INDICA QUE EL JURADO FUE **ELEGIDO ALEATORIAMENTE** ENTRE LA TOTALIDAD DE LA POBLACIÓN. LOS AFROAMERICANOS TIENEN UNA PROBABILIDAD $p = 0{,}5$ DE SER ELEGIDOS.

H_a INDICA QUE LOS AFROAMERICANOS TIENEN MENOS POSIBILIDADES DE SER ELEGIDOS PARA LA LISTA DE JURADOS QUE LA PROPORCIÓN DE LOS MISMOS EN LA POBLACIÓN: $p < 0{,}50$.

EN EL CASO DEL JURADO LA PRUEBA ESTADÍSTICA ES LA VARIABLE ALEATORIA BINOMIAL X CON $p = 0{,}5$ Y $n = 80$.

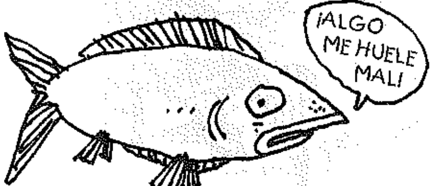
Paso n.°3. VALOR p:
UNA AFIRMACIÓN DE PROBABILIDAD QUE RESPONDE A LA PREGUNTA: SI LA HIPÓTESIS NULA FUERA VERDADERA, ENTONCES ¿CUÁL SERÍA LA PROBABILIDAD DE OBSERVAR UN VALOR DE LA PRUEBA ESTADÍSTICA TAN EXAGERADO COMO EL QUE HEMOS VISTO?

Paso n.°4. COMPARAR EL VALOR p CON UN **RIESGO α FIJO.**

α ACTÚA COMO UN PUNTO DE CORTE POR DEBAJO DEL CUAL ACEPTAMOS QUE UN EFECTO ES ESTADÍSTICAMENTE SIGNIFICATIVO. O SEA, SI

VALOR P $\leq \alpha$

ENTONCES, DESCARTAMOS LA HIPÓTESIS NULA H_0 Y DECIDIMOS QUE PASA ALGO RARO.



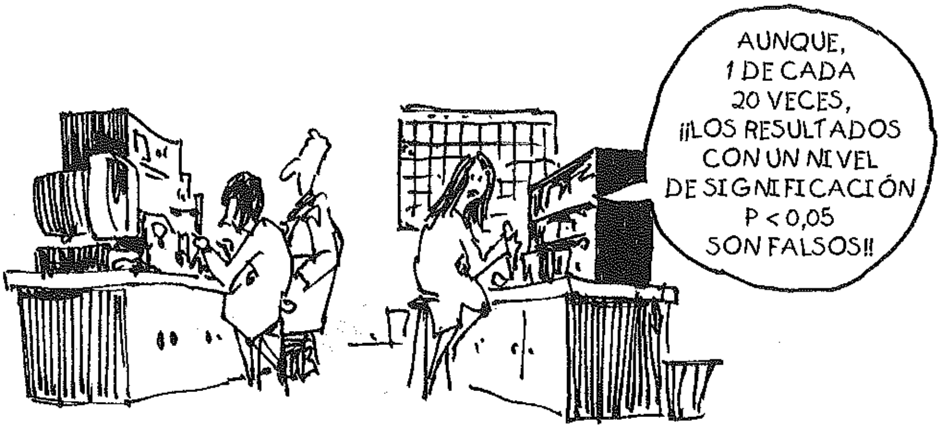
EN EL EJEMPLO, EL VALOR p ERA

$$\Pr(x \leq 4 \mid p = 0{,}50 \text{ Y } n = 80) = 1{,}4 \times 10^{-18}$$

CALCULAMOS ESTE VALOR p DE FORMA MODERNA, USANDO UN PAQUETE DE SOFTWARE ESTADÍSTICO.

EN EL CASO DEL JURADO, EL ESTADÍSTICO ENCONTRÓ $p = 3{,}6 \times 10^{-18}$, EL MISMO NÚMERO DE OPORTUNIDADES DE QUE TE SALGAN TRES ESCALERAS REALES SEGUIDAS.

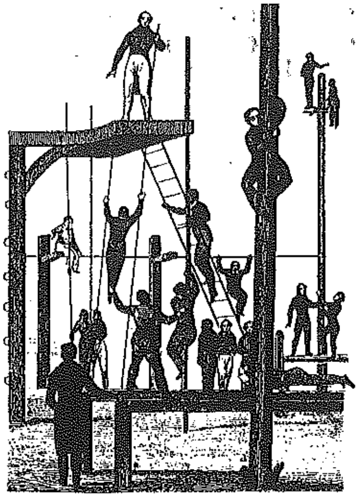
En los estudios científicos, se usa con frecuencia un valor α fijo de 0,05 o 0,01. Podemos decir que estos valores fijos son reliquias de la era preinformática, cuando nos referíamos a tablas que sólo se publicaban para determinados valores críticos. Aún hoy, en algunas publicaciones científicas sólo aparecen los resultados si el valor $p \leq 0{,}05$.



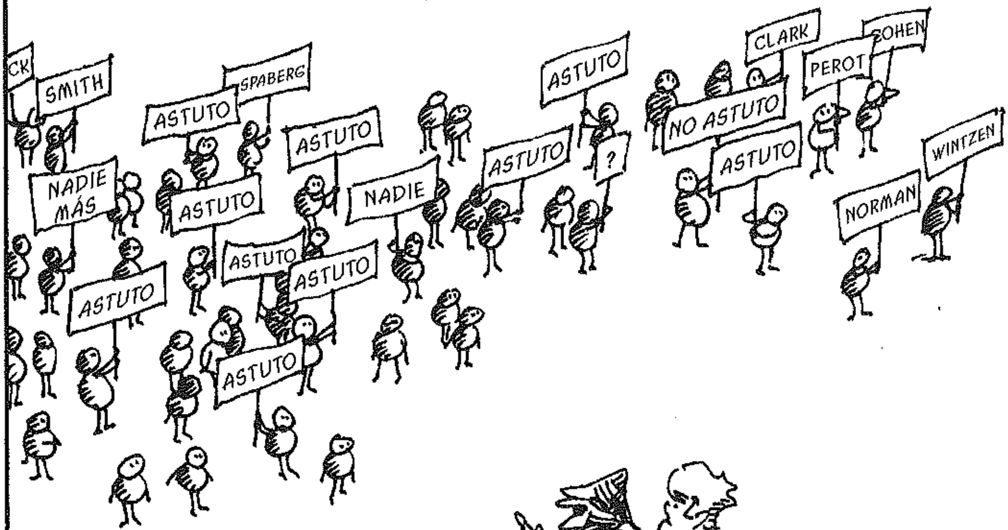
En los procedimientos legales, el estándar es más flexible

MUESTRA GRANDE
PRUEBA DE SIGNIFICACIÓN PARA PROPORCIONES

EL EJEMPLO DEL JURADO ERA UN CASO ESPECÍFICO DE UN PROBLEMA GENERAL. LA HIPÓTESIS NULA ERA $p = p_0$, DONDE p_0 ERA UNA PROBABILIDAD (EN ESTE CASO, 0,5). VEAMOS AHORA PROBLEMAS COMO ÉSTE DESDE UN PUNTO DE VISTA GENERAL. **CONTRASTEMOS LA HIPÓTESIS** $p = p_0$.



COMO SIEMPRE, IMAGINAMOS TENER UNA POBLACIÓN NUMEROSA... OBSERVAMOS UNA MUESTRA GRANDE... Y VEMOS QUE ALGUNAS CARACTERÍSTICAS SE PRESENTAN CON PROBABILIDAD $\hat{p}$.



BASÁNDONOS EN ESTA OBSERVACIÓN, QUEREMOS SABER SI LA PROBABILIDAD REAL POBLACIONAL ES (POR EJEMPLO) MAYOR QUE LA DE OTRO VALOR p_0.
POR EJEMPLO, AL SENADOR ASTUTO, QUE TIENE UNA $\hat{p}$ DE 0,55 LE GUSTARÍA SABER QUE $p > 0,5$, O SEA, MAYORÍA ABSOLUTA.

Paso n.°1.

La hipótesis **nula** es

$$H_0 : p = p_0$$

La hipótesis **alternativa** depende de la dirección tomada por el efecto que estamos buscando. En el caso del senador Astuto,

$$H_a : p > p_0$$

Pero, en otros casos la hipótesis alternativa podría ser:

$$H_a : p < p_0$$

o

$$H_a : p \neq p_0$$

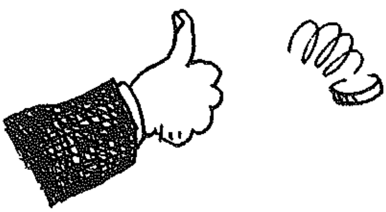
En el ejemplo de la selección del jurado, la hipótesis alternativa era

$$H_a : p < 0.5$$

Y otras veces, nos interesa saber que p es diferente a algún valor p_0. Por ejemplo, en el contraste del lanzamiento de una moneda, la hipótesis alternativa es

$$H_a : p \neq 0.5$$

Pero no tenemos una opinión a priori de si saldrán más caras o cruces.



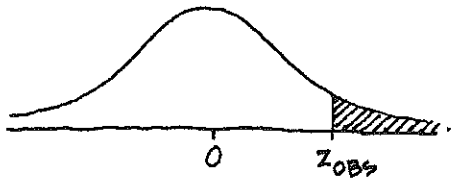
Paso n.°2. La prueba estadística es

$$z_{OBS} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$$

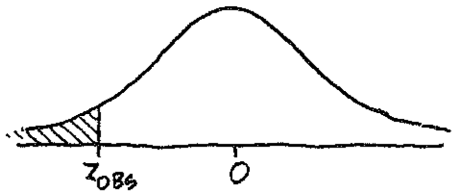
Lo que mide la gran desviación de p con respecto a p_0. Si H_0 es cierta, z_{OBS} tiene una distribución normal tipificada.

Paso n.°3. El valor p depende de cuál sea la hipótesis alternativa relevante:

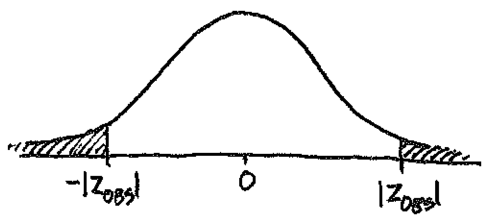
a) H_a «a la derecha»: $p > p_0$ utiliza un valor p $Pr(z > z_{OBS})$



b) H_a «a la izquierda»: $p < p_0$ utiliza un valor p $Pr(z < z_{OBS})$



c) H_a «bilateral»: p p_0 utiliza un valor p $Pr(|z| > |z_{OBS}|)$



EN EL EJEMPLO DEL SENADOR ASTUTO:

1) LAS HIPÓTESIS SON:

$$H_0 : p = 0{,}5$$

$$H_a : p > 0{,}5$$

2) LA PRUEBA ESTADÍSTICA ES:

$$z_{OBS} = \frac{0{,}55 - 0{,}50}{\sqrt{(0{,}5)(0{,}5)}/\sqrt{1.000}} = 3{,}16$$

3) EL VALOR p ES:

$$Pr(z > z_{OBS}) = Pr(z \geq 3{,}16) = 0{,}0008$$

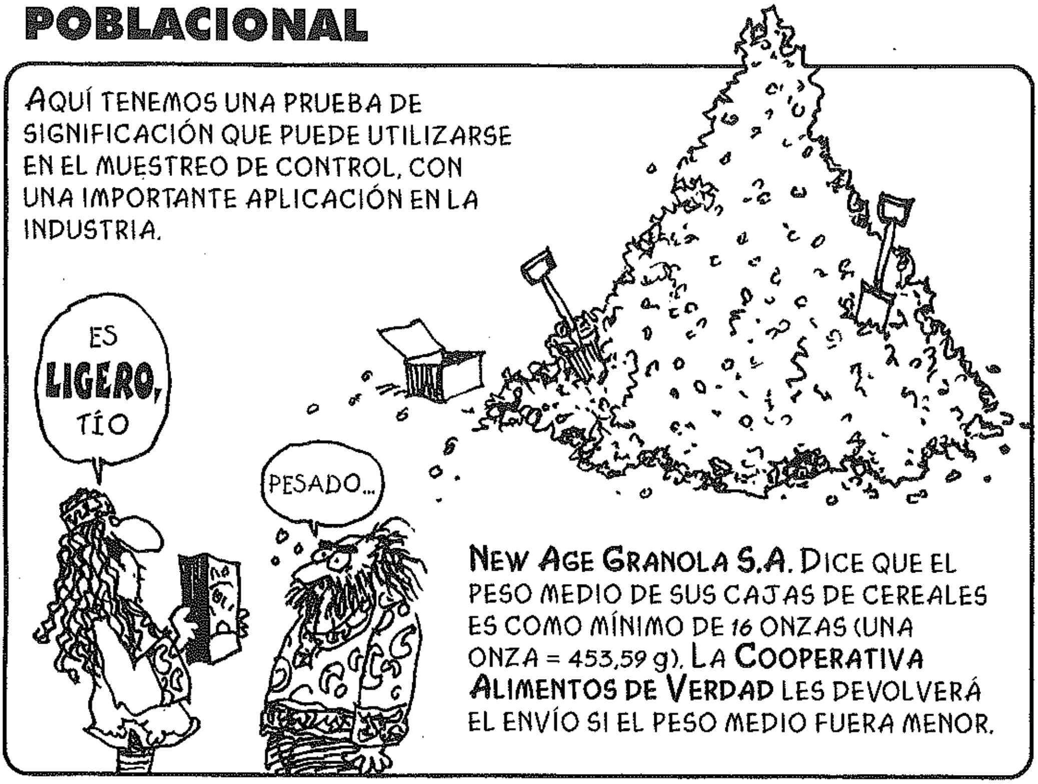
(DE LA TABLA NORMAL)

4) ASTUTO, QUE ES BASTANTE CONSERVADOR, TOMA UN NIVEL DE SIGNIFICACIÓN α DE 0,01 Y OBSERVA QUE

$$Pr(z > z_{OBS}) = 0.0008 < \alpha$$

ASÍ, EL SENADOR RECHAZA LA HIPÓTESIS NULA. ÉL (Y SUS PARTIDARIOS) PUEDEN ESTAR SEGUROS DE SU VENTAJA ELECTORAL.

MUESTRA GRANDE
PRUEBA PARA LA MEDIA POBLACIONAL



OBVIAMENTE, ALIMENTOS DE VERDAD NO PIENSA PESAR CADA CAJA DEL ENVÍO. ¡VAN A UTILIZAR LA ESTADÍSTICA!

PRIMERO, ESCOGEN SUS HIPÓTESIS.

$$H_0: \mu = 16 \text{ ONZAS}$$

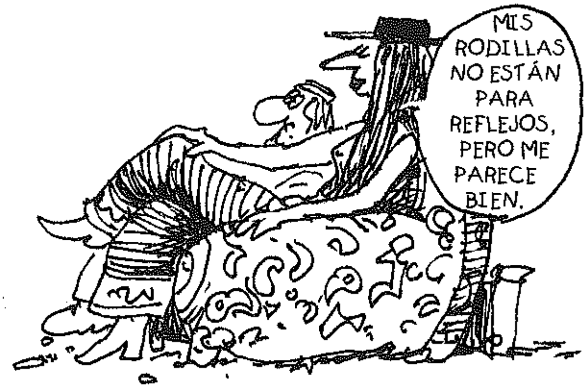
$$H_a: \mu < 16 \text{ ONZAS}$$

RECHAZAR LA HIPÓTESIS NULA SUPONE RECHAZAR A GRANOLA

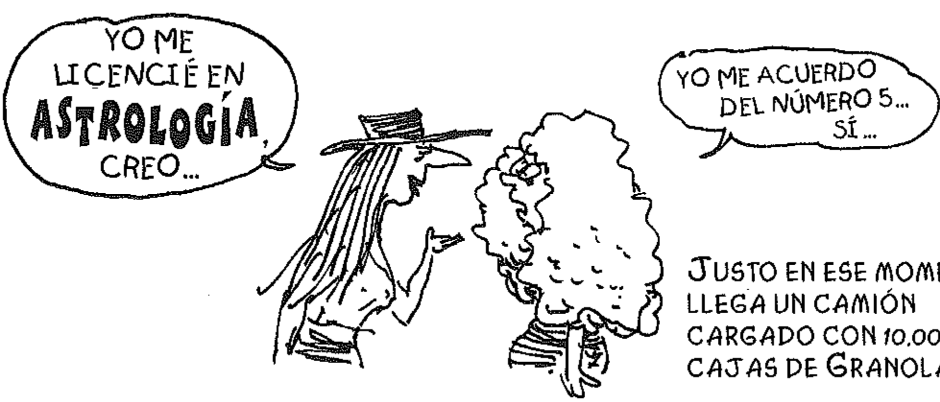
DESPUÉS, ELIGEN UN ANÁLISIS ESTADÍSTICO. AHORA, YA DEBERÍA SER UN ACTO REFLEJO SABER QUE LA DISPERSIÓN MUESTRAL DESDE LA MEDIA ES

$$\frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$$

DÓNDE s ES LA DESVIACIÓN TÍPICA MUESTRAL. EN LA HIPÓTESIS NULA, ESTA SE APROXIMA A LA NORMAL ESTÁNDAR CUANDO LA MUESTRA ES GRANDE, POR EL TEOREMA CENTRAL DEL LÍMITE.



VOLVIENDO UN MOMENTO AL PASO N.º 3, ESTABLECEN UN LÍMITE PARA EL RIESGO α. COMO NO ACABARON LA CARRERA DE CIENCIAS, LOS DE ALIMENTOS DE VERDAD CREEN QUE $\alpha = 0{,}05$ SUENA BIEN.



JUSTO EN ESE MOMENTO, LLEGA UN CAMIÓN CARGADO CON 10.000 CAJAS DE GRANOLA.

COGEN UNA PEQUEÑA MUESTRA ALEATORIA SIMPLE DE 49 CAJAS, LAS PESAN POR SEPARADO Y DETERMINAN EL RESUMEN ESTADÍSTICO:

$\bar{x} = 15{,}90$ ONZAS

$s = 0{,}35$ ONZAS

UN POCO LIGERO PERO, ¿SIGNIFICATIVO?

INSERTAN LOS VALORES EN LA PRUEBA ESTADÍSTICA PARA DESCUBRIR QUE:

$$z_{OBS} = \frac{15{,}9 - 16}{0{,}35/\sqrt{49}} = -2$$

AHORA CALCULAN EL VALOR p:

$$Pr(z < -2 \mid H_0) = 0{,}0227$$

AL SER ÉSTE MENOR QUE EL RIESGO $\alpha = 0{,}05$, ALIMENTOS DE VERDAD **RECHAZA** LA HIPÓTESIS NULA Y EL ENVÍO.

MUESTRA PEQUEÑA

PRUEBA PARA LA MEDIA POBLACIONAL

VOLVEMOS A CAMALEÓN MOTORS, Y A SU PRUEBA DE UN ACCIDENTE A UNOS 20 KILÓMETROS POR HORA. LA **COMPAÑÍA DE SEGUROS LA HONRADA** CUBRIRÁ AL ASEGURADO SÓLO SI EL COSTE MEDIO DE LA REPARACIÓN DE SU COCHE TRAS UN ACCIDENTE A 20 KILÓMETROS POR HORA ES INFERIOR A 1.000 DÓLARES. LA COMPAÑÍA UTILIZA EL ESTÁNDAR $\alpha = 0{,}05$. ASÍ QUE...

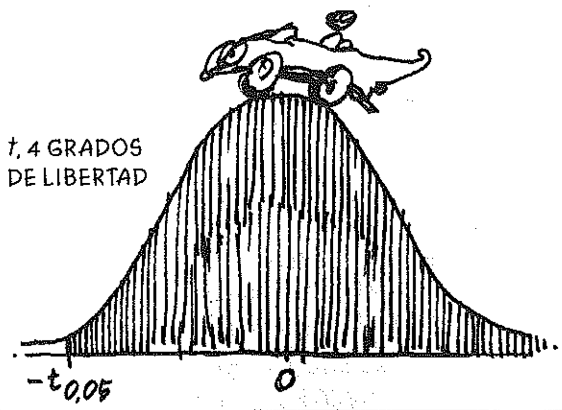
H_0: $\mu \geq 1.000$ DÓLARES. EL COSTE MEDIO ES DEMASIADO ALTO.

H_a: $\mu < 1.000$ DÓLARES. EL COSTE MEDIO ESTÁ BIEN.

EL ANÁLISIS ESTADÍSTICO SE BASA EN LA DISTRIBUCIÓN t

$$t = \frac{\bar{X} - \mu_0}{SE(\bar{X})}$$

EN LA QUE μ_0 ES LA MEDIA HIPOTÉTICA DE 1.000 DÓLARES.



Y QUEREMOS QUE NUESTRO VALOR t OBSERVADO PERMANEZCA A LA IZQUIERDA DE $-t_{0,05}$ (YA QUE LA $\bar{x}$ DE VALOR REDUCIDO ES PREFERIBLE, $\bar{x} - \mu_0$ DEBERÍA SER NEGATIVA PARA APOYAR A H_a).

		α		
		0,05	0,025	0,005
GRADOS DE LIBERTAD	1	6,31	12,71	63,66
	2	2,92	4,30	9,92
	3	2,35	3,18	5,84
	4	2,13	2,78	4,60
	5	2,01	2,57	4,03

POR LA TABLA DE VALORES CRÍTICOS t, VEMOS QUE $t_{0,05} = 2,13$, ASÍ QUE DECIDIMOS RECHAZAR LA H_0 SI

$$t_{OBS} \leq -t_{0,05} = -2,13$$

DEL CAPÍTULO 7 TENEMOS $\bar{x} = 540$ DÓLARES Y $s = 299$ DÓLARES PARA UNA MUESTRA PEQUEÑA DE CINCO COCHES, ASÍ QUE OBTENEMOS

$$t_{OBS} = \frac{540-1.000}{299/\sqrt{5}}$$

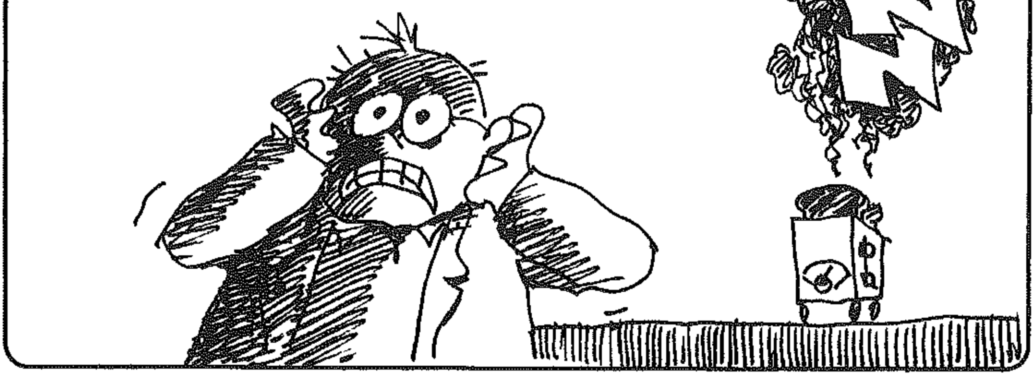
$$= -3,44 < -t_{0,05}$$

EL COCHE PASA LA PRUEBA... H_0 ES RECHAZADA... Y LA PÓLIZA DE SEGUROS ES TRAMITADA.

ESTE ES UN EJEMPLO DE **ACEPTACIÓN DE MUESTREO**. LA HIPÓTESIS NULA CONSISTE EN QUE LOS COSTES DE REPARACIÓN SON INADMISIBLES, Y LA CASA AUTOMOVILÍSTICA ASUME LA RESPONSABILIDAD HASTA QUE SE PRESENTEN PRUEBAS DE INOCENCIA SUFICIENTES, ES DECIR, QUE EL ACCIDENTE ENTRE DENTRO DEL PRESUPUESTO.

TEORÍA DE LA DECISIÓN

PODEMOS PENSAR EN EL CONTRASTE DE HIPÓTESIS Y EN LA PRUEBA DE SIGNIFICACIÓN EN TÉRMINOS DE **DETECTORES DE HUMO DOMÉSTICOS**. SI TIENES UNO DE ESTOS APARATOS EN CASA, TE HABRÁS DADO CUENTA DE QUE SUELEN DISPARARSE CADA VEZ QUE SE CHAMUSCAN LAS TOSTADAS.



ESTO ES LO QUE SE LLAMA UN **ERROR DE TIPO I**: UNA ALARMA SIN FUEGO. POR EL CONTRARIO, UN **ERROR DE TIPO II** ES UN FUEGO SIN ALARMA. TODOS LOS COCINEROS SABEN CÓMO EVITAR UN ERROR DE TIPO I: **QUITANDO LAS PILAS**. POR DESGRACIA ESTO AUMENTA LA INCIDENCIA DE ERRORES DE TIPO II.

DE IGUAL MODO, LA DISMINUCIÓN DE POSIBILIDADES DE ERRORES DE TIPO II, POR EJEMPLO, HACIENDO QUE LA ALARMA SEA HIPERSENSIBLE, PUEDE AUMENTAR EL NÚMERO DE FALSAS ALARMAS.

PODEMOS RESUMIR TODO ESTO EN UNA **TABLA DE DECISIONES** DE 2×2:

	SIN FUEGO	CON FUEGO
SIN ALARMA	NO HAY ERROR	TIPO II
CON ALARMA	TIPO I	NO HAY ERROR

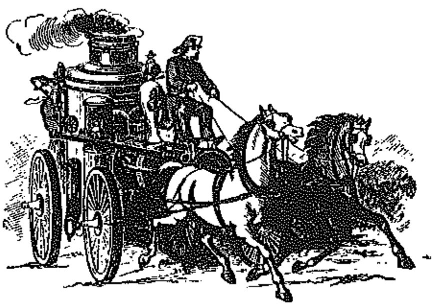
AHORA PENSEMOS EN LA HIPÓTESIS NULA COMO CONDICIÓN DE «SIN FUEGO», MIENTRAS QUE LA HIPÓTESIS ALTERNATIVA ES QUE HAY UN INCENDIO. LA ALARMA CORRESPONDE AL RECHAZO DE LA HIPÓTESIS NULA.

ESTADO REAL

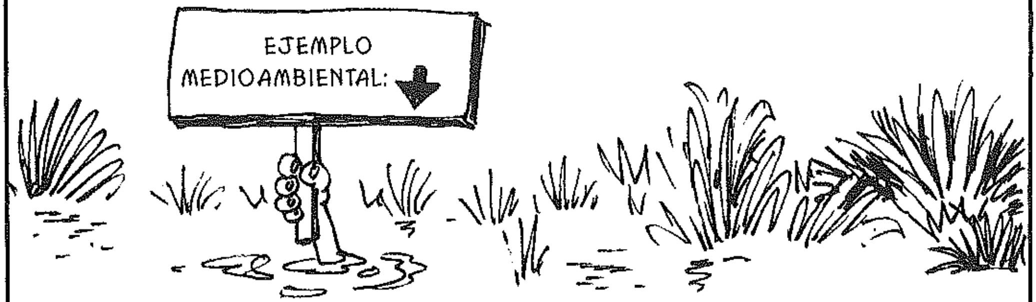
	H_0	H_a
ACEPTAR H_0	NO HAY ERROR	TIPO II
RECHAZAR H_a	TIPO I	NO HAY ERROR

TODAS LAS PRUEBAS DE SIGNIFICACIÓN QUE HICIMOS ANTERIORMENTE EN ESTE CAPÍTULO SUBRAYABAN LA PROBABILIDAD DE ENFRENTARSE A UN ERROR DE TIPO I, ES DECIR, LA PROBABILIDAD DE QUE SE DIERAN NUESTRAS OBSERVACIONES SI H_0 FUERA CIERTA. PEDIMOS QUE:

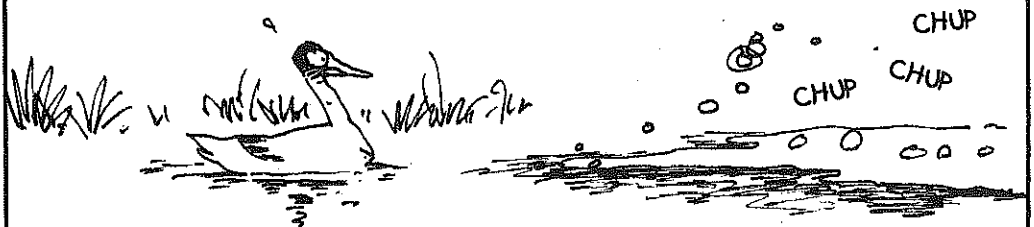
$$\text{Pr(DE RECHAZO DE } H_0 \mid H_0) = \text{Pr(ERROR TIPO I} \mid H_0) = \alpha$$



PERO, OTRAS VECES, LO QUE EN REALIDAD QUEREMOS CONOCER ES LA POSIBILIDAD DE QUE SE PRODUZCA UN **ERROR DE TIPO II**. ES DECIR, ¿QUÉ NIVEL DE SENSIBILIDAD TIENE NUESTRO «SISTEMA DE ALARMA» CUANDO LA HIPÓTESIS ALTERNATIVA ES **VERDADERA**?



ANTES, LAS FÁBRICAS QUE VERTÍAN SUS DESECHOS EN LAS VÍAS FLUVIALES DEBÍAN DEMOSTRAR QUE EL VERTIDO NO TENÍA EFECTOS NOCIVOS PARA LA VIDA ANIMAL. ESTO ES LA H_0. EL QUE CONTAMINABA PODÍA SEGUIR HACIÉNDOLO HASTA QUE LA HIPÓTESIS NULA FUERA RECHAZADA POR ALCANZAR EL NIVEL DE SIGNIFICACIÓN 0,05.



ASÍ QUE, CUANDO EL RESPONSABLE DE LA CONTAMINACIÓN CREÍA ESTAR VIOLANDO LOS LÍMITES ESTABLECIDOS POR LA AGENCIA PARA LA PROTECCIÓN DEL MEDIO AMBIENTE, LLEVABA A CABO UN PLAN DE SEGUIMIENTO DE CONTAMINACIÓN NADA EFECTIVO.

EL CULPABLE DE LA CONTAMINACIÓN SE SIENTE MUY SATISFECHO, PORQUE, COMO EN NUESTRA ALARMA DE HUMO SIN PILAS, SU ANÁLISIS TIENE POCAS POSIBILIDADES O NINGUNA DE HACER SALTAR LA ALARMA.

FORMALICEMOS ESTA IDEA. PARA DESCRIBIR LA PROBABILIDAD DE ERROR DE TIPO II, AÑADIMOS UNA NUEVA LETRA GRIEGA: BETA O β.

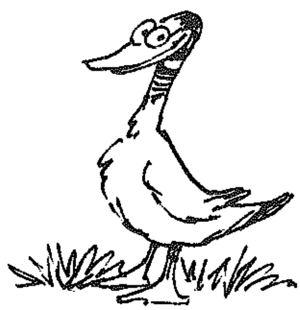
$$\beta = Pr(\text{DE ACEPTACIÓN } H_0 | H_a)$$

$$= Pr(\text{ERROR TIPO II} | H_a)$$

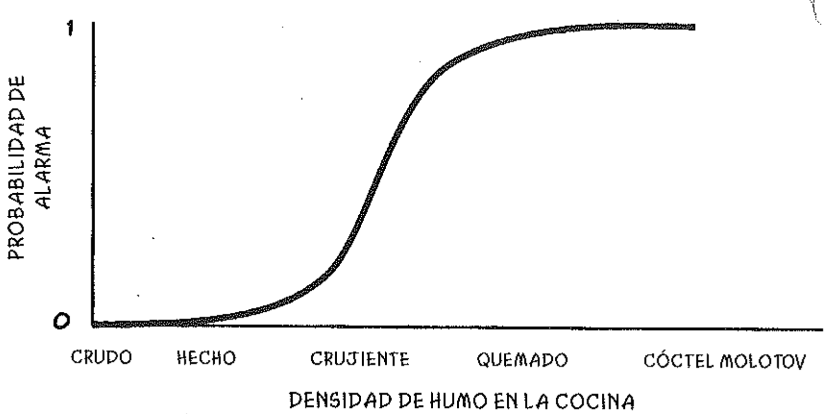
LA **POTENCIA** DE UNA PRUEBA SE DEFINE COMO $1 - \beta$.

$$Pr(\text{DE RECHAZO DE } H_0 | H_a).$$

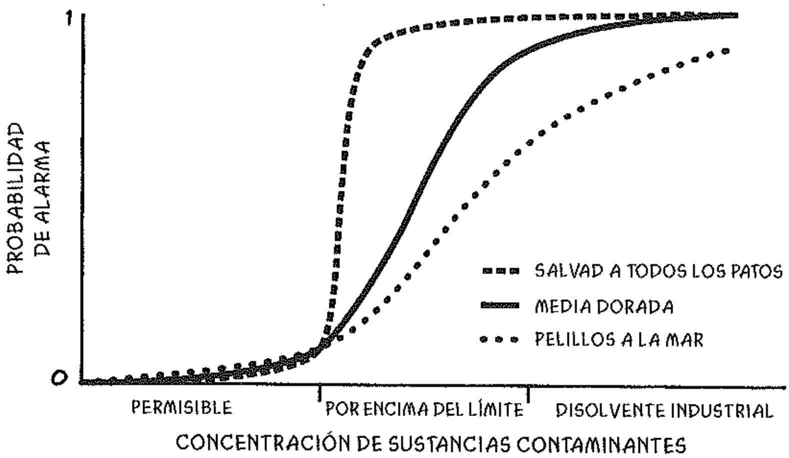
TE GUSTARÁ SABER QUE LOS LEGISLADORES MEDIOAMBIENTALES CADA VEZ EXIGEN MÁS PROGRAMAS DE SEGUIMIENTO PARA DEMOSTRAR QUE TIENEN UNA PROBABILIDAD MUY ALTA DE DETECTAR GRAVES CASOS DE CONTAMINACIÓN. EL **ANÁLISIS DE POTENCIA** REVELA A MENUDO DEFECTOS OCULTOS EN LOS PROGRAMAS DE SEGUIMIENTO.



UNA FORMA DE VISUALIZAR EL EFECTO DE LOS ANÁLISIS DE POTENCIA ES DIBUJAR LA GRÁFICA DE LA PROBABILIDAD DEL RECHAZO DE H_0 Y EL ESTADO REAL DEL SISTEMA DE ALARMA. EN EL CASO DE LA ALARMA PARA HUMOS, LA PROBABILIDAD ASCIENDE HASTA *1* A MEDIDA QUE EL HUMO SE HACE MÁS DENSO.



PARA EL EJEMPLO DE LA CALIDAD DEL AGUA DE LA AGENCIA PARA LA PROTECCIÓN DEL MEDIO AMBIENTE, EL EJE HORIZONTAL REPRESENTA LA CONCENTRACIÓN REAL DE CONTAMINANTE EN EL AGUA.



AQUÍ, ESTÁN REPRESENTADAS LAS CURVAS DE EFECTIVIDAD DE LOS TRES PROGRAMAS DE SEGUIMIENTO. LA DE **SALVAD A TODOS LOS PATOS** (CON UN COSTE DE 5 MILLONES DE DÓLARES), LA **MEDIA DORADA** (CON UN COSTE DE 500.000 DÓLARES) Y **PELILLOS A LA MAR** (TAMBIÉN, CON UN COSTE DE 500.000 DÓLARES). CUANTO MAYOR SEA LA POTENCIA DE LA PRUEBA MAYOR SERÁ LA PRONUNCIACIÓN DE LA CURVA.

ENTONCES ¿POR QUÉ TIENES ESA **SENSACIÓN DE VACÍO** EN EL ESTÓMAGO? PORQUE, PARA PONER EN PRÁCTICA ESTAS IDEAS, TENEMOS QUE APRENDER A APLICARLAS EN DIFERENTES SITUACIONES DE LAS QUE NI SIQUIERA HEMOS HABLADO HASTA AHORA. ES LO QUE VAMOS A HACER A CONTINUACIÓN, MEDIANTE LA **COMPARACIÓN** DE **DOS POBLACIONES**.

◆ Capítulo 9 ◆

COMPARACIÓN DE DOS POBLACIONES

DONDE APRENDEREMOS NUEVAS RECETAS
USANDO VIEJOS INGREDIENTES...

PERO, LO QUE CONVIERTE A LA ESTADÍSTICA EN ALGO CASI TAN DESAFIANTE COMO LA COCINA, ES LA VARIEDAD. AL IGUAL QUE UN COCINERO EXPERTO, EL ESTADÍSTICO PUEDE DEGUSTAR O «PROBAR» LOS INGREDIENTES EN UN PROBLEMA, PARA DESCUBRIR CUÁL ES LA FORMA MÁS EFECTIVA DE COMBINARLOS EN UNA RECETA ESTADÍSTICA.

(LA RAZÓN POR LA QUE TANTO LOS LIBROS DE COCINA COMO LOS DE ESTADÍSTICA SON TAN VOLUMINOSOS ES PORQUE AMBOS APORTAN SOLUCIONES EN UNA GRAN VARIEDAD DE SITUACIONES.)

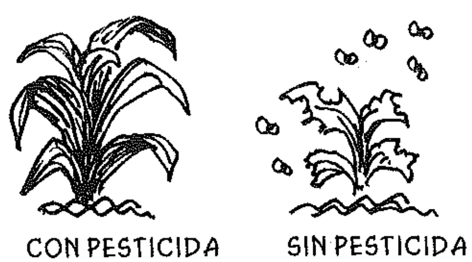
EN ESTE CAPÍTULO UTILIZAREMOS NUESTROS MÉTODOS DEL PLATO COMBINADO CON ALGUNAS RECETAS NUEVAS QUE NOS AYUDARÁN A CONTESTAR LAS SIGUIENTES PREGUNTAS:

¿TOMAR ASPIRINAS CON REGULARIDAD PUEDE REDUCIR EL RIESGO DE INCIDENCIA DE INFARTO?

¿PUEDE UN PESTICIDA DETERMINADO AUMENTAR EL CRECIMIENTO DE MAÍZ POR HECTÁREA?

¿SON DIFERENTES LOS SALARIOS DE HOMBRES Y MUJERES QUE DESEMPEÑAN UN MISMO TRABAJO?

EL INGREDIENTE EN COMÚN DE TODAS ESTAS PREGUNTAS ES ESTE: PUEDEN SER CONTESTADAS MEDIANTE **LA COMPARACIÓN DE DOS MUESTRAS ALEATORIAS INDEPENDIENTES**, UNA DE CADA POBLACIÓN.



Y AL FINAL DEL CAPÍTULO, CONSIDERAREMOS LAS DIFERENTES FORMAS DE COMPARAR DOS MEDIAS, LO CUAL NO IMPLICA ÚNICAMENTE TOMAR DOS MUESTRAS ALEATORIAS...

Comparando TASAS DE ÉXITO (o de fracaso) en dos poblaciones.

EMPECEMOS CON UN EXPERIMENTO, QUE FORMÓ PARTE DE UN ESTUDIO DE LA UNIVERSIDAD DE HARVARD, CON EL QUE SE PRETENDÍA DECIDIR QUÉ GRADO DE EFECTIVIDAD TENÍA LA ASPIRINA EN LA REDUCCIÓN DE RIESGO DE INFARTO. Y COMO OCURRE EN MUCHAS PRUEBAS MÉDICAS, LAS PROBABILIDADES DE QUE ALGÚN INDIVIDUO SUFRA LA ENFERMEDAD, EN ESTE CASO UN INFARTO, EN EL TRANSCURSO DE UN AÑO, SON MUY PEQUEÑAS. PERO QUEREMOS RESPUESTAS RÁPIDAS. ¿QUÉ PODEMOS HACER?

LA SIMPLE, AUNQUE CARA, SOLUCIÓN ES EXAMINAR A UN GRAN NÚMERO DE INDIVIDUOS DURANTE UN PERÍODO REDUCIDO DE TIEMPO. EN ESTE ESTUDIO, SE FORMARON DOS GRUPOS A PARTIR DE 22.071 INDIVIDUOS (TODOS ELLOS MÉDICOS VOLUNTARIOS).

AL GRUPO 1 SE LE ADMINISTRA UN PLACEBO, UNA PASTILLA IDÉNTICA A LA ASPIRINA PERO QUE NO CONTIENE ASPIRINA.

AL GRUPO 2 SE LE ADMINISTRA UNA ASPIRINA DIARIA.

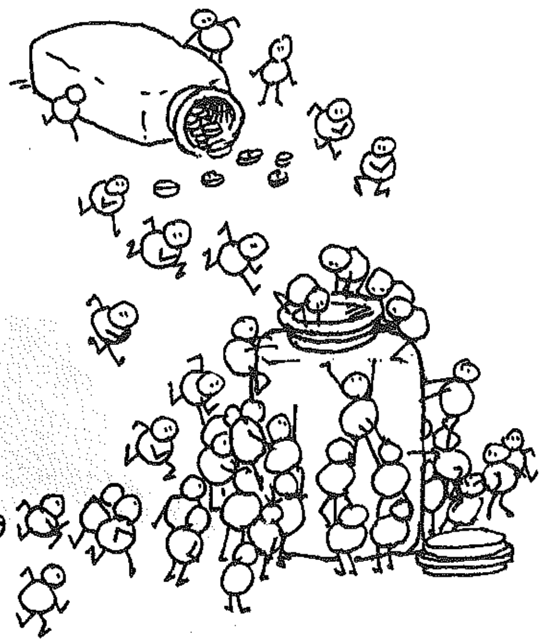
DURANTE UN PERÍODO DE APROXIMADAMENTE CINCO AÑOS*, LOS RESULTADOS REFLEJARON LAS SIGUIENTES RESPUESTAS: INFARTO O AUSENCIA DE INFARTO. EL RESULTADO: (EN LA TABLA QUE PRESENTAMOS A CONTINUACIÓN HEMOS COMBINADO INFARTOS MORTALES Y NO MORTALES.)

	INFARTO	NO INFARTO	n	INCIDENCIA DE INFARTO
PLACEBO	239	10.795	11.034	$\hat{p}_1 = \frac{239}{11.034} = 0,0217$
ASPIRINA	139	10.898	11.037	$\hat{p}_2 = \frac{139}{11.037} = 0,0126$

LA DIFERENCIA QUE SE OBSERVA EN EL NIVEL DE ÉXITO ES $\hat{p}_1 - \hat{p}_2 = 0,0091$. PARECE UNA CANTIDAD PEQUEÑA HASTA QUE NOS FIJAMOS EN EL RIESGO RELATIVO.

$$\frac{\hat{p}_1}{\hat{p}_2} = \frac{0,0217}{0,0126} = 1,72.$$

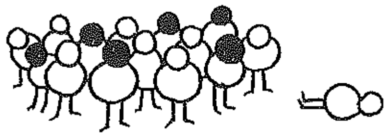
LOS INDIVIDUOS DEL GRUPO PLACEBO ERAN 1,72 VECES MÁS SUSCEPTIBLES DE SUFRIR UN INFARTO QUE LOS INDIVIDUOS DEL GRUPO DE LA ASPIRINA.



* EL EXPERIMENTO SE DETUVO PRONTO, POR SU RESULTADO POSITIVO. NO HUBIERA SIDO PRÁCTICO NI INTELIGENTE OCULTAR LOS RESULTADOS AL GRUPO DEL PLACEBO.

El modelo: LAS OBSERVACIONES HECHAS A PARTIR DE LOS GRUPOS DEL PLACEBO Y DE LA ASPIRINA SON MUESTRAS INDEPENDIENTES EXTRAÍDAS DE DOS POBLACIONES BINOMIALES. PARA LA CONSISTENCIA NOS REFERIMOS AL INFARTO COMO UN **ÉXITO** (!)

POBLACIÓN 1
DEL PLACEBO,
POSIBILIDAD DE ÉXITO = p_1



POBLACIÓN 2
DE LA ASPIRINA,
POSIBILIDAD DE ÉXITO = p_2

EL OBJETIVO ES LA ESTIMACIÓN DE LA DIFERENCIA REAL: $p_1 - p_2$.

PARA CADA POBLACIÓN (EN REALIDAD GRANDES MUESTRAS TOMADAS DE UNA POBLACIÓN EN GENERAL) TENEMOS LAS YA CONOCIDAS VARIABLES ALEATORIAS:

X_1 NÚMERO DE ÉXITOS DE LA POBLACIÓN 1

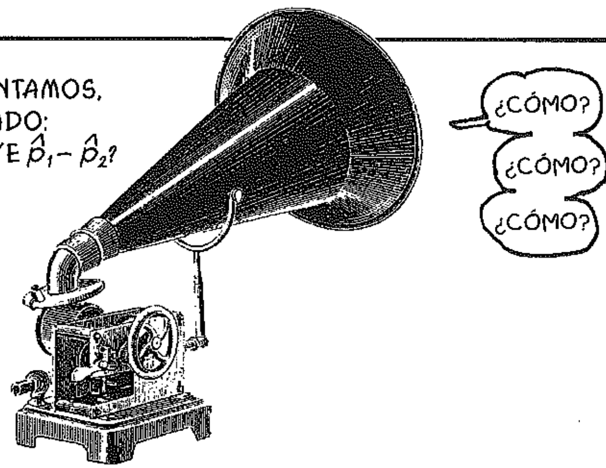
X_2 NÚMERO DE ÉXITOS DE LA POBLACIÓN 2

$\hat{P}_1 = \dfrac{X_1}{n_1}$ PROPORCIÓN DE ÉXITOS DE LA POBLACIÓN 1

$\hat{P}_2 = \dfrac{X_2}{n_2}$ PROPORCIÓN DE ÉXITOS DE LA POBLACIÓN 2

Y UN ESTIMADOR DE LA DIFERENCIA: $\hat{P}_1 - \hat{P}_2$

Y AHORA NOS PREGUNTAMOS, COMO UN DISCO RAYADO: ¿CÓMO SE DISTRIBUYE $\hat{p}_1 - \hat{p}_2$?



Distribución muestral de $\hat{P}_1 - \hat{P}_2$

PARA MUESTRAS GRANDES, $\hat{p}_1 - \hat{p}_2$ SE DISTRIBUYE CASI CON NORMALIDAD, MUCHO MÁS QUE EN EL CASO DE UNA SOLA MUESTRA. PODEMOS REALIZAR LA TÍPICA TRANSFORMACIÓN z PARA OBTENER LA VARIABLE ALEATORIA NORMAL ESTÁNDAR (APROXIMADAMENTE).

$$z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sigma(\hat{P}_1 - \hat{P}_2)}$$

PERO, ¿CÓMO CALCULAMOS LA DESVIACIÓN ESTÁNDAR EN EL DENOMINADOR?

COMO LAS DOS MUESTRAS SON INDEPENDIENTES, TAMBIÉN LO SON LAS VARIABLES ALEATORIAS $\hat{p}_1$ Y $\hat{p}_2$, Y LAS DOS VARIANZAS SE SUMAN.

$$\sigma^2(\hat{P}_1 - \hat{P}_2) = \sigma^2(\hat{P}_1) + \sigma^2(\hat{P}_2)$$

ENTONCES

$$\sigma(\hat{P}_1 - \hat{P}_2) = \sqrt{\sigma^2(\hat{P}_1) + \sigma^2(\hat{P}_2)}$$

Y AHORA, UNA VEZ QUE CONOCEMOS LA DISTRIBUCIÓN DE LA PRUEBA ESTADÍSTICA, PODEMOS PASAR A REALIZAR EL CÁLCULO DE LOS **INTERVALOS DE CONFIANZA** Y EL **CONTRASTE DE LA HIPÓTESIS**, QUE AFIRMA QUE LA INGESTIÓN DE ASPIRINA REDUCE EL RIESGO DE INFARTO.

Intervalos de confianza para $p_1 - p_2$

COMO SIEMPRE, LOS INTERVALOS DE CONFIANZA PARA NUESTRA ESTIMACIÓN SON ASÍ:

$$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}} SE(\hat{p}_1 - \hat{p}_2)$$

DIFERENCIA REAL ENTRE LAS PROPORCIONES POBLACIONALES

DIFERENCIA OBSERVADA

VALOR CRÍTICO

ERROR TÍPICO

SE SUMAN LAS VARIANZAS $\hat{P}_1$ Y $\hat{P}_2$. AHORA EL ERROR TÍPICO ES:

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

EN EL ESTUDIO SOBRE LA ASPIRINA, EL ERROR TÍPICO ES:

$$\sqrt{\frac{(0{,}0217)(0{,}9783)}{11.034} + \frac{(0{,}0126)(0{,}9874)}{11.037}}$$

$$= \mathbf{0{,}00175}$$

PARA CONSEGUIR UN INTERVALO DE CONFIANZA DEL 95% EN EL ESTUDIO DE LA ASPIRINA, SÓLO TENEMOS QUE AÑADIR LOS VALORES OBSERVADOS:

$$p_1 - p_2 = 0{,}0091 \pm (1{,}96)(0{,}00175)$$
$$= \mathbf{0{,}0091 \pm 0{,}0034}$$
$$= \mathbf{0{,}0057 ; 0{,}0125}$$

ESTAMOS SEGUROS, COMO MÍNIMO EN UN 95%, DE QUE LA DIFERENCIA EN LA INCIDENCIA DE INFARTO SE ENCUENTRA ENTRE 0,0057 Y 0,0125. SE TRATA, DEFINITIVAMENTE, DE UNA CIFRA POSITIVA. AHORA ESTAMOS SEGUROS, COMO MÍNIMO EN UN 95%, DE QUE LA ASPIRINA DISMINUYE LA INCIDENCIA DE INFARTO.

Contraste de hipótesis

LA PREGUNTA RELATIVA AL CONTRASTE DE HIPÓTESIS ES:

H_0. LA HIPÓTESIS NULA, ES QUE LA ASPIRINA NO TIENE EFECTO: $p_1 = p_2$.

H_a. LA HIPÓTESIS ALTERNATIVA, ES QUE LA ASPIRINA REDUCE LA INCIDENCIA DE INFARTO: $p_1 > p_2$.

AHORA NECESITAMOS UN ESTADÍSTICO CON UNA DISTRIBUCIÓN NORMAL CUANDO H_0 ES CIERTA...

OBSERVA QUE BAJO H_0 LAS DOS PROPORCIONES SON **IGUALES**, $p_1 = p_2 = p$... AHORA PODEMOS JUNTAR TODOS LOS DATOS PARA CONSEGUIR LA PROPORCIÓN DE INFARTO **EN AMBAS MUESTRAS A LA VEZ**:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

CUANDO LA HIPÓTESIS NULA ES VERDADERA, EL ERROR ESTÁNDAR DEPENDE ÚNICAMENTE DE ESTA ESTIMACIÓN CONJUNTA:

$$SE_0(\hat{P}_1 - \hat{P}_2) = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Y PODEMOS FORMULAR UNA PRUEBA ESTADÍSTICA

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{SE_0(\hat{P}_1 - \hat{P}_2)}$$

(NORMALMENTE, EL NUMERADOR SERÍA $\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)$, PERO H_0 ASUME QUE $p_1 - p_2 = 0$.)

EN EL ESTUDIO SOBRE LA ASPIRINA NOS ENCONTRAMOS CON:

$$\hat{p} = \frac{378}{22.071}$$

$$SE_0(\hat{P}_1 - \hat{P}_2) = 0{,}00175$$

ENTONCES

$$Z_{OBS} = \frac{0{,}0091}{0{,}00175} = 5{,}20$$

z_{OBS} ESTÁ A MÁS **DE CINCO DESVIACIONES TÍPICAS** DE CERO, UN RESULTADO ALTAMENTE SIGNIFICATIVO. HALLAREMOS EL VALOR p CON AYUDA DE UNA TABLA O UN ORDENADOR PERSONAL.

$$\text{VALOR P} = PR(Z \geq Z_{OBS}) = PR(Z \geq 5{,}2) = \mathbf{0{,}0000001}$$

SI LA HIPÓTESIS NULA FUERA VERDADERA, LA PROBABILIDAD DE OBSERVAR UN EFECTO ASÍ DE GRANDE SERÍA DE **UNA ENTRE DIEZ MILLONES**, ¡ES UNA PRUEBA DE MUCHO PESO CONTRA H_0!

La receta básica:

PARA PROBAR LA HIPÓTESIS NULA

$$H_0: p_1 = p_2$$

CALCULAMOS EL ESTADÍSTICO

$$Z_{OBS} = \frac{\hat{p}_1 - \hat{p}_2}{SE_0(\hat{P})}$$

(EN LA QUE SE_0 SE CALCULA USANDO LA PROBABILIDAD CONJUNTA EN LA COMBINACIÓN DE LOS DOS GRUPOS).

EL VALOR p RELEVANTE DEPENDE DE LA HIPÓTESIS ALTERNATIVA

A) H_a BILATERAL: $p_1 = p_2$

VALOR P= Pr (|Z|Z$_{OBS}$|)

B) H_a A LA DERECHA: $p_1 > p_2$

VALOR P= Pr (Z > Z$_{OBS}$)

C) H_a A LA IZQUIERDA: $p_1 < p_2$

VALOR P= Pr (Z< Z$_{OBS}$)

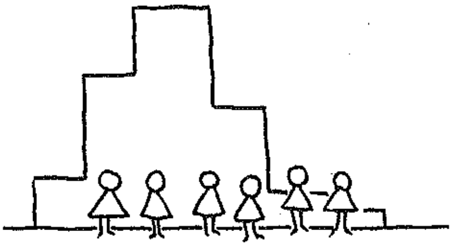
EL ANÁLISIS DEL ESTUDIO SOBRE LA ASPIRINA DEPENDE DE CIERTAS CARACTERÍSTICAS DEL EXPERIMENTO, DISEÑADAS PARA ASEGURAR LA ALEATORIEDAD Y ELIMINAR LA IMPARCIALIDAD:

LOS PUNTOS 1 Y 2 CONSTITUYEN PARTES ESENCIALES DE LA MAYORÍA DE LOS DISEÑOS DE PRUEBAS MÉDICAS CON SERES HUMANOS, PERO EL PUNTO 3 NO ES SIEMPRE NECESARIO. SE PUEDEN ENCONTRAR BUENAS PRUEBAS ESTADÍSTICAS CON MUESTRAS PEQUEÑAS EN PAQUETES DE SOFTWARE. ESTOS PROCEDIMIENTOS NO **PARAMÉTRICOS** DEPENDEN DE UNOS CÁLCULOS DE PROBABILIDAD SIMPLES PERO LARGOS, PARECIDOS A LOS CÁLCULOS DEL JUEGO QUE YA VIMOS EN EL CAPÍTULO 4...

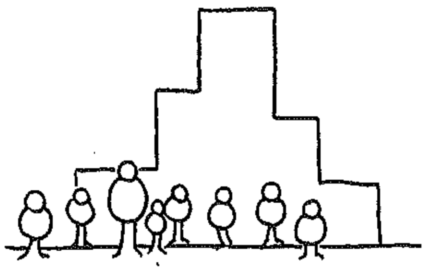
Comparación de las MEDIAS de dos poblaciones

SUPONGAMOS QUE QUEREMOS COMPARAR EL SALARIO MEDIO DE LOS TRABAJADORES Y LAS TRABAJADORAS QUE DESEMPEÑAN EL MISMO TRABAJO EN UNA EMPRESA.

LA POBLACIÓN UNO ESTÁ FORMADA POR MUJERES, LA POBLACIÓN DOS, POR HOMBRES.



LA POBLACIÓN UNO TIENE UN SALARIO MEDIO μ_1 Y UNA DESVIACIÓN TÍPICA σ_1.



LA POBLACIÓN DOS TIENE UN SALARIO MEDIO μ_2 Y UNA DESVIACIÓN TÍPICA σ_2.

DOS MUESTRAS ALEATORIAS DE TAMAÑO n_1 DEL GRUPO 1 Y UNA n_2 DEL GRUPO 2 NOS DA UNAS MEDIAS MUESTRALES $\bar{x}_1$ Y $\bar{x}_2$ Y UNAS DESVIACIONES TÍPICAS σ_1 Y σ_2 RESPECTIVAMENTE. EL ESTIMADOR DE μ_1 Y μ_2 ES

$$\bar{X}_1 - \bar{X}_2$$

¿UN ESTIMADOR $\bar{X}_1 - \bar{X}_2$ ES BUENO O NO? PARA LAS MUESTRAS GRANDES ES APROXIMADAMENTE NORMAL (POR EL TEOREMA CENTRAL DEL LÍMITE) Y EL ERROR ESTÁNDAR ES

$$SE(\bar{X}_1-\bar{X}_2) = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

(LAS VARIANZAS SE SUMAN, PORQUE LAS MUESTRAS SON INDEPENDIENTES.) AHORA PODEMOS PASAR DIRECTAMENTE A LOS

intervalos de confianza: PARA

MUESTRAS GRANDES, EL INTERVALO DE CONFIANZA $(1-\alpha)\,100\%$ PARA LA DIFERENCIA ENTRE MEDIAS ES

$$\mu_1-\mu_2 = \bar{x}_1-\bar{x}_2 \pm z_{\frac{\alpha}{2}}\, SE(\bar{X}_1-\bar{X}_2)$$

* «RIGHT IN THE FORMULA» POR «CORRECTO» Y «EN LA DERECHA». [N.T.]

Contraste de hipótesis: ESTABLECEMOS

LA HIPÓTESIS NULA DE QUE LAS MEDIAS DE LAS DOS POBLACIONES SON IGUALES:

$$H_0: \ \mu_1 = \mu_2$$

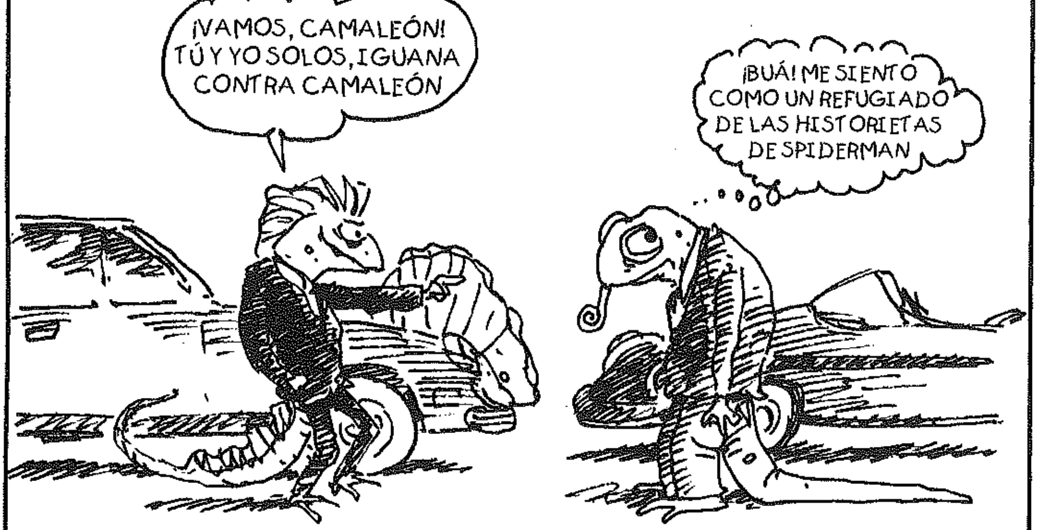
LA PRUEBA ESTADÍSTICA ES:

$$z_{OBS} = \frac{\bar{X}_1-\bar{X}_2}{SE(\bar{X}_1-\bar{X}_2)}$$

Y LOS VALORES P SON COMO SIEMPRE.

¿Y cómo se comparan las medias de MUESTRAS PEQUEÑAS?

¿TE ACUERDAS DE LA **CAMALEÓN MOTORS**? LA COMPETENCIA, AUTO IGUANA, AFIRMA QUE SU ACCESORIO DE POLIESTIRENO COLOCADO EN LA PARTE DELANTERA DE LA CARROCERÍA, PROPORCIONA UNA MAYOR PROTECCIÓN EN CASO DE CHOQUE FRONTAL. PARA DEMOSTRARLO, HAN ESTRELLADO **SIETE IGUANAS.**



ESTOS SON SUS RESULTADOS COMPARADOS CON LOS DE CAMALEÓN:

CAMALEÓN

1	*$150
2	$400
3	$720
4	$500
5	$930
n_1	5
$\bar{x}_1$	$540
s_1	$299

*(COSTE EN DÓLARES)

IGUANA

1	$50
2	$200
3	$150
4	$400
5	$750
6	$400
7	$150
n_2	7
$\bar{x}_2$	$300
s_2	$238

LA DISTRIBUCIÓN t PUEDE APLICARSE SI DOS POBLACIONES TIENEN FORMA DE MONTAÑA Y TIENEN LA MISMA DESVIACIÓN TÍPICA $\sigma = \sigma_1 = \sigma_2$. EL ÚNICO PROBLEMILLA ES QUE TENEMOS QUE JUNTAR LAS VARIANZAS MUESTRALES s_1^2 Y s_2^2 PARA FORMAR UNA ESTIMACIÓN ÚNICA DE σ:

$$s^2_{\text{CONJUNTA}} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

EL ERROR ESTÁNDAR ES EL MISMO QUE EN LAS MUESTRAS GRANDES, SUSTITUYENDO S_{CONJUNTA} A S_1 Y S_2.

$$SE(\bar{X}_1-\bar{X}_2) = \sqrt{\frac{s^2_{\text{CONJUNTA}}}{n_1} + \frac{s^2_{\text{CONJUNTA}}}{n_2^2}}$$

$$= s_{\text{CONJUNTA}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

EL INTERVALO DE CONFIANZA $(1-\alpha)\cdot 100\%$ ES:

$$\mu_1-\mu_2 = (\bar{x}_1-\bar{x}_2) \pm t_{\frac{\alpha}{2}}\, SE(\bar{X}_1-\bar{X}_2)$$

DONDE $t_{\frac{\alpha}{2}}$ ES UN VALOR CRÍTICO DE t CON n_1+n_2-2 GRADOS DE LIBERTAD.

LOS REPTILES FABRICANTES DE COCHES CONVIENEN EN QUE SUS RESPECTIVAS DESVIACIONES TÍPICAS ESTÁN MUY PRÓXIMAS Y REPARAN EN QUE LOS HISTOGRAMAS TIENEN FORMA DE MONTAÑA. Y CALCULAN:

$$s_{\text{CONJUNTA}} = \sqrt{\frac{4\cdot 299^2 + 6\cdot 328^2}{10}} = 264$$

$$SE(\bar{X}_1-\bar{X}_2) = 264\sqrt{\frac{1}{5}+\frac{1}{7}} = 154$$

EL INTERVALO DE CONFIANZA DEL 95% ES:

$$\mu_1-\mu_2 = 540-300 \pm t_{0{,}025}(154)$$

$$= 240 \pm (2{,}23)(154)$$

$$= 240 \pm 340$$

PUESTO QUE ÉSTE INCLUYE EL VALOR CERO AUTO IGUANA **NO** HA EXPERIMENTADO UNA MEJORA SIGNIFICATIVA EN LOS GASTOS DE REPARACIÓN.

A CONTINUACIÓN, UN EJEMPLO ILUSTRATIVO DE LAS PIFIAS QUE PUEDEN COMETERSE POR LEER EL LIBRO DE RECETAS CON LOS PIES: EL PROPIETARIO DE UNA GRAN FLOTA DE TAXIS QUIERE COMPARAR LA CANTIDAD DE GASOLINA CONSUMIDA CON **GASOLINA A** Y **GASOLINA B.**

EMPIEZA CON 100 TAXIS, Y ASIGNA ALEATORIAMENTE 50 A CADA TIPO DE GASOLINA, Y, TRAS UNOS DÍAS DE CONDUCCIÓN, AFIRMA

	TAMAÑO MUESTRAL	MEDIA DE MILLAS RECORRIDAS	DESVIACIÓN TÍPICA
A	50	25	5,00
B	50	26	4,00

LA DIFERENCIA ENTRE MUESTRAS ES

$$\bar{x}_1 - \bar{x}_2 = 25 - 26 = -1$$

¿DE VERDAD LA GASOLINA B ES MEJOR QUE LA A?

DEBIDO AL ELEVADO VALOR DE LAS DESVIACIONES TÍPICAS, EL ERROR ESTÁNDAR ES BASTANTE SUSTANCIAL:

$$SE(\bar{X}_1-\bar{X}_2) = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{25}{50}+\frac{16}{50}}$$

$$= 0{,}905$$

EN EL NIVEL DE CONFIANZA DEL 95%, TENEMOS

$$\mu_1-\mu_2 = \bar{x}_1-\bar{x}_2 \pm z_{0,025}(0{,}905)$$

$$= -1 \pm (1{,}96)(0{,}905)$$

$$= -1 \pm 1{,}774$$

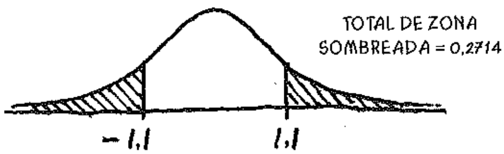
ESTO INCLUYE EL VALOR CERO, QUE CORRESPONDE A $\mu_1=\mu_2$

EL VALOR P PARA LA HIPÓTESIS ALTERNATIVA, H_a, ES $\mu_1 \neq \mu_2$

$$Pr(|z| \geq |z_{OBS}|) = Pr(|z| \geq \frac{1}{0{,}905})$$

$$= Pr(|z| \geq 1{,}1) = 2\,(0{,}1357)$$

$$= 0{,}2714$$



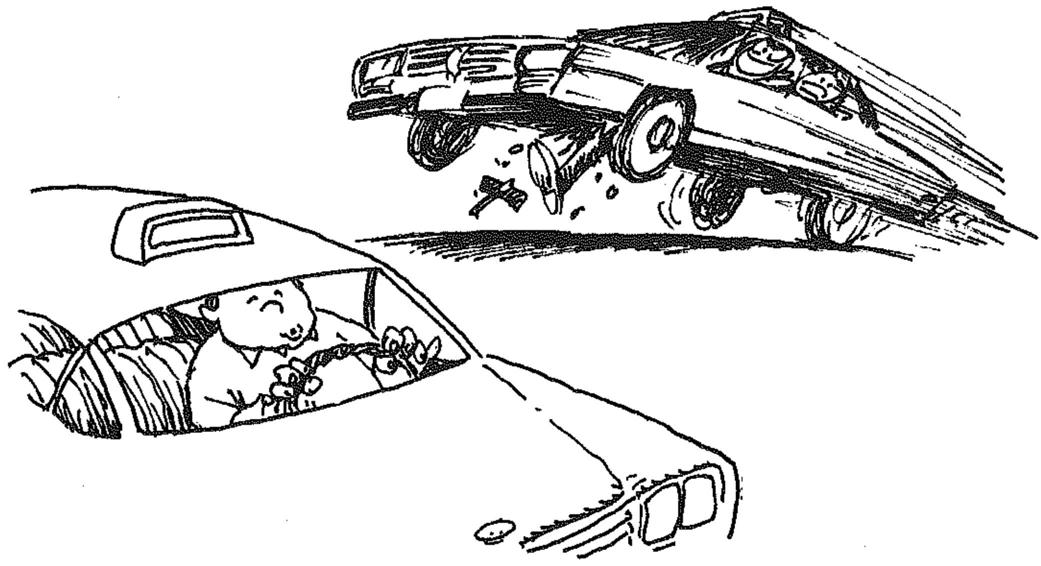
ESTA CIFRA **EXCEDE** EL VALOR DE SIGNIFICACIÓN $\alpha = 0{,}05$, ASÍ QUE LLEGAMOS A LA CONCLUSIÓN DE QUE LAS PRUEBAS A FAVOR DE UNA U OTRA GASOLINA SON POCO FIABLES.

COMPARACIONES APAREADAS

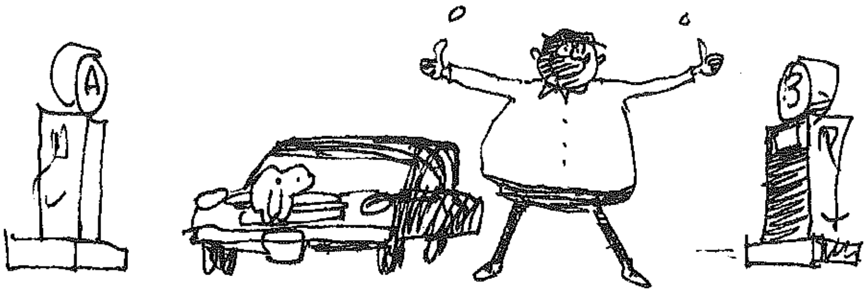
Una forma mejor de comparar tipos de gasolina

EL PROPIETARIO DE LOS TAXIS SIGUIÓ PASO A PASO EL LIBRO DE RECETAS. SUS MUESTRAS ERAN ALEATORIAS, Y LOS TAMAÑOS MUESTRALES ERAN LO SUFICIENTEMENTE GRANDES. SIMPLEMENTE, SE EQUIVOCÓ AL NO **PENSAR** CUANDO ERA NECESARIO.

AUNQUE LA GASOLINA B PARECE UN POCO MEJOR QUE LA GASOLINA A, EL INTERVALO DE CONFIANZA ES AMPLIO POR LAS GRANDES DESVIACIONES TÍPICAS. **EL NÚMERO DE MILLAS RECORRIDAS VARÍA AMPLIAMENTE DE UN TAXI A OTRO.** ¿POR QUÉ ESTA GRAN VARIABILIDAD? ¡PORQUE LOS TAXIS, Y LOS TAXISTAS, TIENEN **PERSONALIDADES DIFERENTES!**



UNA FORMA MUCHO MEJOR DE REALIZAR ESTE ESTUDIO ES PONER GASOLINA A Y GASOLINA B EN EL **MISMO TAXI** EN DÍAS DIFERENTES.



TODAVÍA TENEMOS QUE APLICAR LA ALEATORIEDAD PARA ELEGIR, LANZANDO UNA MONEDA, CUÁNDO PONEMOS GASOLINA A, EL MARTES O EL MIÉRCOLES. TAMBIÉN SE PUEDE REDUCIR LA APLICACIÓN DEL EXPERIMENTO A 10 TAXIS, Y ASÍ LE AHORRAMOS AL PROPIETARIO MUCHO TIEMPO Y DINERO.

TAXI	GASOLINA A	GASOLINA B	DIFERENCIA
1	27,01	26,95	0,06
2	20,00	20,44	−0,44
3	23,41	25,05	− 1,64
4	25,22	26,32	− 1,10
5	30,11	29,56	0,55
6	25,55	26,60	− 1,05
7	22,23	22,93	−0,70
8	19,78	20,23	−0,45
9	33,45	33,95	−0,50
10	25,22	26,01	−0,79
MEDIA	25,20	25,80	− 0,60
DESVIACIÓN TÍPICA	4,27	4,10	0,61

OBSERVA QUE LAS MEDIAS Y LAS DESVIACIONES TÍPICAS DE LA GASOLINA A Y LA GASOLINA B SON MÁS O MENOS LAS MISMAS. ESTO ERA DE ESPERAR, YA QUE POSEEN LA MISMA FUENTE DE VARIABILIDAD, COMO OCURRÍA EN EL EXPERIMENTO NO APAREADO. PERO EN ESTA OCASIÓN, LA **COLUMNA DE LA DIFERENCIA** TIENE UNA DESVIACIÓN TÍPICA **MUY PEQUEÑA**. LA COLUMNA DE LA DIFERENCIA, AL COMPARAR LA RESPUESTA DE UN SOLO COCHE CON AMBOS TIPOS DE GASOLINA, ELIMINA LA VARIABILIDAD ENTRE LOS TAXIS.

LAS DIFERENCIAS d_i PROPORCIONAN UNA MEDIDA **ÚNICA DE LA DIFERENCIA PARA CADA TAXI**, Y ASÍ, PODEMOS UTILIZARLAS PARA REALIZAR UNA PRUEBA t (DE MUESTRA PEQUEÑA)

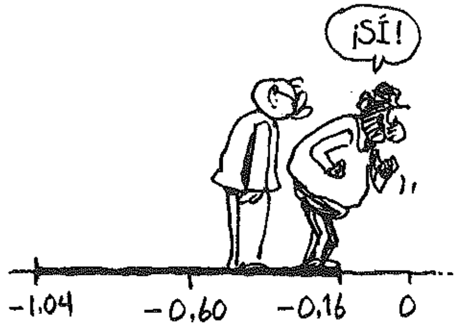
$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

EL INTERVALO DE CONFIANZA DEL 95% ALREDEDOR DE $\bar{d}$ ES

$$\mu_d = \underbrace{\bar{d}}_{\text{MEDIA MUESTRAL}} \pm \underbrace{t_{0,025}}_{\text{VALOR CRÍTICO}} \underbrace{(s_d/\sqrt{n})}_{\text{ERROR TÍPICO}}$$

$$= -0,6 \pm (2,26)\left(\frac{0,61}{\sqrt{10}}\right)$$

$$= \mathbf{-0,60 \pm 0,44}$$



AHORA TENEMOS $-1,04 \le \mu_d \le -0,16$ CON UNA SEGURIDAD DEL 95%, ES UNA BUENA PRUEBA DE QUE LA GASOLINA B ES REALMENTE MEJOR.

EL VALOR P DEL CONTRASTE DE HIPÓTESIS PUEDE SER CALCULADO CON LA AYUDA DE UN PAQUETE DE SOFTWARE:

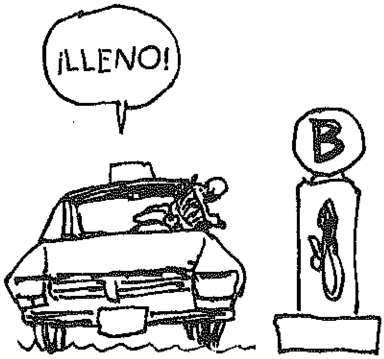
$$H_a: \mu_d \neq 0$$

$$\text{VALOR P} = Pr(|t| \geq |t_{OBS}|)$$

$$= Pr\left(|t| \geq \frac{0,6}{0,19}\right)$$

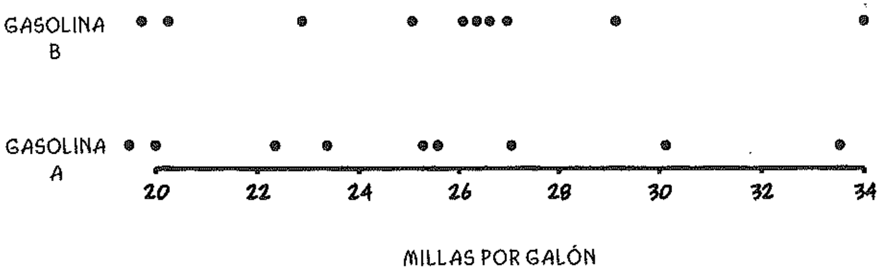
$$= Pr(|t| \geq 3,15)$$

$$= \mathbf{0,012 < 0,05}$$

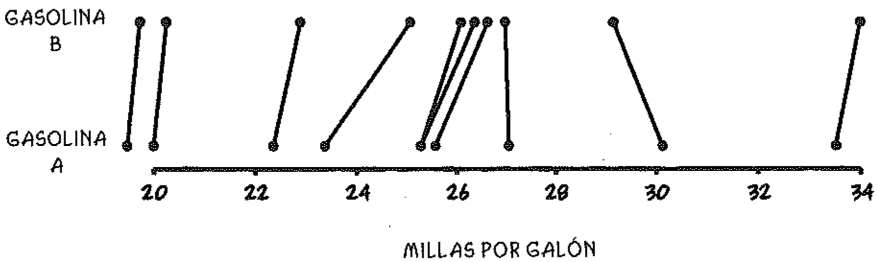


DE NUEVO, LA GASOLINA B SUPERA LA PRUEBA.

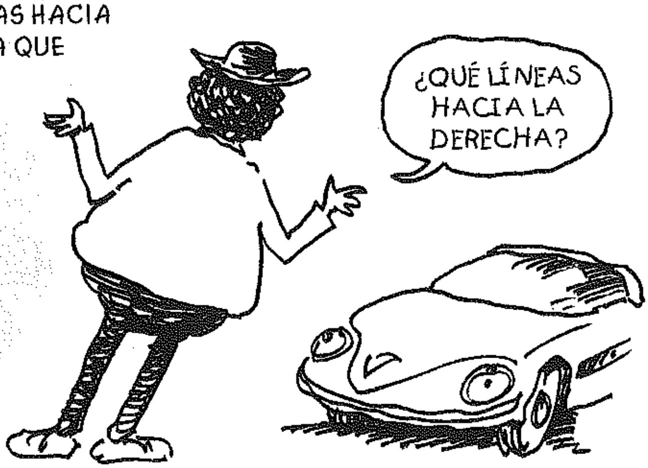
AQUÍ TENEMOS UNOS DIAGRAMAS DE PUNTOS SOBRE LOS DATOS DEL CONSUMO POR MILLAS RECORRIDAS: EL PRIMERO REPRESENTA LAS MILLAS NO APAREADAS:



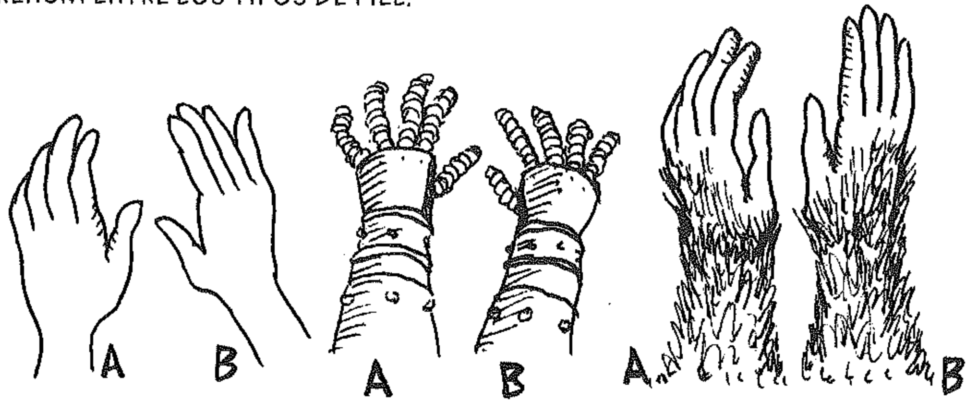
Y AHORA, LOS MISMOS DATOS APAREADOS, SEGÚN TAXIS:



EL PREDOMINIO DE LÍNEAS HACIA LA DERECHA DEMUESTRA QUE LA GASOLINA B APORTA MEJORES RESULTADOS.



UN EXPERIMENTO DE COMPARACIÓN APAREADA ES UNA DE LA FORMAS MÁS EFICACES DE REDUCIR LA VARIABILIDAD NATURAL AL COMPARAR FORMAS DE TRATAMIENTO. POR EJEMPLO, SI COMPARAMOS CREMAS PARA MANOS, LAS DOS MARCAS SE ASIGNAN DE FORMA ALEATORIA A LA MANO DERECHA O IZQUIERDA DE CADA INDIVIDUO. ESTO ELIMINA LA VARIABILIDAD DEBIDA A LA DIFERENCIA ENTRE LOS TIPOS DE PIEL.



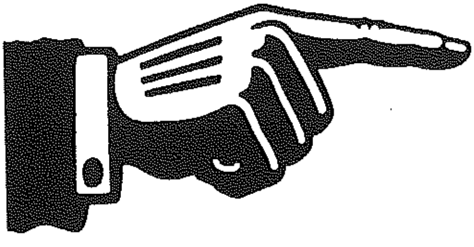
O, SI COMPARAMOS DOS MARCAS DE CEREALES, CADA CATADOR PUNTÚA AMBAS MARCAS (POR ORDEN ALEATORIO). LA COMPARACIÓN APAREADA ELIMINA EL SESGO NATURAL DE QUE AL CATADOR LE GUSTEN O NO LOS CEREALES EN GENERAL.

En este capítulo, hemos aplicado las nociones básicas sobre los intervalos de confianza y el contraste de hipótesis para la comparación de dos poblaciones. Existen innumerables posibilidades. Podríamos haber continuado describiendo comparaciones de:

- Desviaciones típicas de dos poblaciones cuando el tamaño muestral es pequeño.
- Las medias de más de dos poblaciones cuando el tamaño muestral es grande.
- Las medias de más de dos poblaciones cuando el tamaño muestral es pequeño.

¡ETC.!

En la práctica, los estadísticos profesionales determinan la naturaleza general del problema y, después, consultan el libro adecuado.



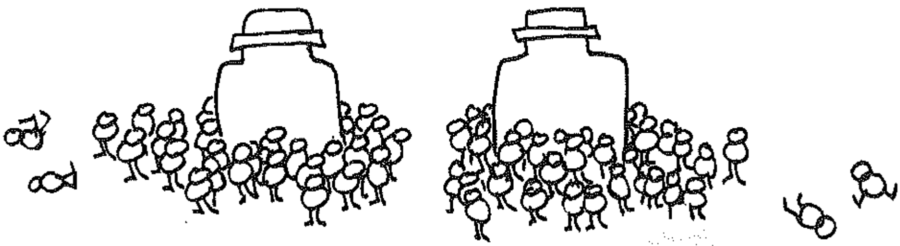
La única novedad de este capítulo ha sido la idea de **experimento de comparación por parejas**. En el capítulo siguiente, veremos otros tipos de diseños experimentales.

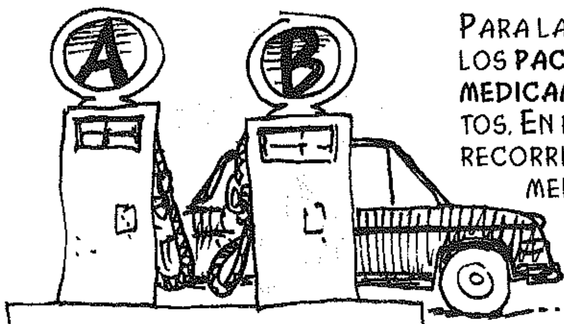
♦ Capítulo 10 ♦
DISEÑO EXPERIMENTAL

A MENUDO, EL DISEÑO DE UN EXPERIMENTO ORIGINA SU ÉXITO O SU FRACASO. EN EL EJEMPLO DE COMPARACIÓN APAREADA, NUESTRO ESTADÍSTICO PASÓ DE ACUMULAR Y ANALIZAR DATOS DE FORMA PASIVA A PARTICIPAR ACTIVAMENTE EN EL DISEÑO EXPERIMENTAL.

EN ESTE CAPÍTULO, PRESENTAMOS LAS IDEAS FUNDAMENTALES DEL DISEÑO EXPERIMENTAL, Y DEJAREMOS LOS DETALLADOS ANÁLISIS NUMÉRICOS PARA EL ÚTIL PAQUETE DE SOFTWARE DE TU ORDENADOR.

LOS ELEMENTOS DE UN DISEÑO SON LAS **UNIDADES EXPERIMENTALES** Y LOS **TRATAMIENTOS** ASIGNADOS A LAS UNIDADES. EL OBJETIVO DE TODO DISEÑO ES LA COMPARACIÓN DE LOS TRATAMIENTOS.



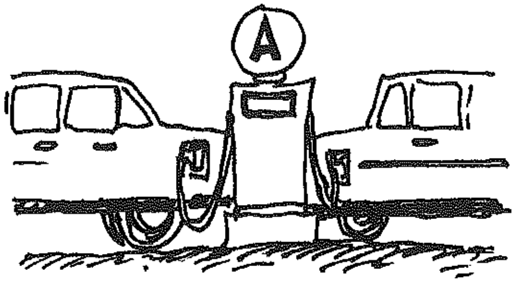


PARA LAS PRUEBAS MÉDICAS, LOS **PACIENTES** SON UNIDADES Y LOS **MEDICAMENTOS** SON LOS TRATAMIENTOS. EN EL EJEMPLO DE LAS MILLAS RECORRIDAS, LAS UNIDADES EXPERIMENTALES SON LOS TAXIS Y LOS TRATAMIENTOS QUE SE COMPARAN SON LOS TIPOS DE GASOLINA A Y B.

EN LOS EXPERIMENTOS AGRÍCOLAS, LAS UNIDADES EXPERIMENTALES SON, A MENUDO, LAS PARCELAS DE UN TERRENO, Y LOS TRATAMIENTOS PUEDEN SER LAS DIFERENTES APLICACIONES DE TIPOS DE GRANO, PESTICIDAS, FERTILIZANTES, ETC.

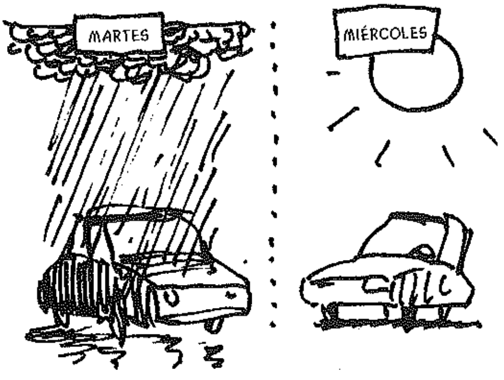
EN LA ACTUALIDAD, LAS IDEAS PARA EL DISEÑO EXPERIMENTAL SE APLICAN GENERALMENTE EN EL PROCESO DE **OPTIMIZACIÓN INDUSTRIAL, EN LA MEDICINA Y EN LAS CIENCIAS SOCIALES**. EL DISEÑO EXPERIMENTAL UTILIZA **TRES PRINCIPIOS FUNDAMENTALES**, QUE ESTÁN CLARAMENTE ILUSTRADOS EN NUESTRO EJEMPLO DE LOS TAXIS.

Repetición: SE ASIGNAN LOS MISMOS TRATAMIENTOS A LAS DIFERENTES UNIDADES EXPERIMENTALES. SIN LA REPETICIÓN, RESULTA IMPOSIBLE ESTABLECER LA VARIABILIDAD NATURAL Y EL ERROR DE LA MEDIDA.



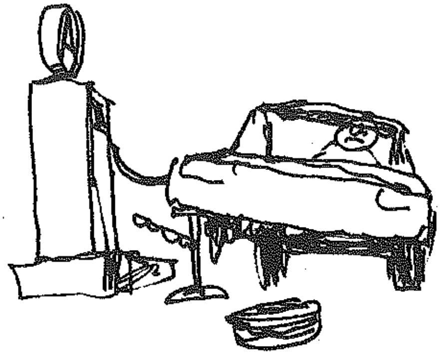
Control local: HACE REFERENCIA A CUALQUIER MÉTODO QUE REPRESENTE Y REDUZCA LA VARIABILIDAD NATURAL. UNA DE SUS FORMAS ES LA AGRUPACIÓN DE LAS UNIDADES EXPERIMENTALES EN **BLOQUES**. EN EL EJEMPLO DE LOS TAXIS, SE USARON LOS DOS TIPOS DE GASOLINA EN TODOS LOS TAXIS, Y DECIMOS QUE CADA TAXI ES UN BLOQUE.

Aleatorización: ¡ES EL PASO PRIMORDIAL DE TODAS LAS ESTADÍSTICAS! LOS TRATAMIENTOS DEBEN SER ASIGNADOS DE FORMA ALEATORIA A LAS UNIDADES EXPERIMENTALES. A CADA TAXI LE ASIGNAMOS GASOLINA A EL MARTES O EL MIÉRCOLES LANZANDO UNA MONEDA. DE NO HABERLO HECHO ASÍ, LOS RESULTADOS PODRÍAN HABERSE VISTO ARRUINADOS POR LAS DIFERENCIAS ENTRE MARTES Y MIÉRCOLES.



AHORA SUPONGAMOS QUE QUEREMOS INVESTIGAR EL EFECTO DE DOS MARCAS DE NEUMÁTICOS Y TAMBIÉN DE DOS TIPOS DE GASOLINA. TENEMOS CUATRO TRATAMIENTOS POSIBLES, QUE PODEMOS APLICAR EN UN DISEÑO FACTORIAL DOS POR DOS. LOS DOS FACTORES SON LAS MARCAS DE LOS TIPOS DE GASOLINA Y LOS NEUMÁTICOS.

	GASOLINA A	GASOLINA B
NEUMÁTICO A	*a*	*b*
NEUMÁTICO B	*c*	*d*

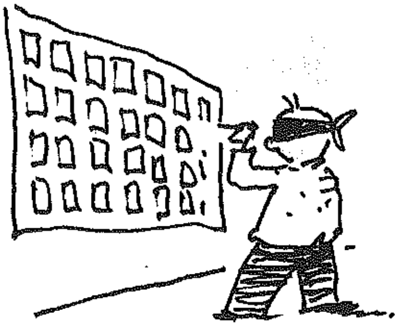


PODEMOS ASIGNAR LOS CUATRO TRATAMIENTOS A CUATRO DÍAS DIFERENTES EN CADA TAXI. LOS CUATRO TRATAMIENTOS (*A, B, C* Y *D*) SE REPITEN EN CADA BLOQUE (TAXI). ESTO SE DENOMINA DISEÑO **COMPLETO DE BLOQUES ALEATORIZADOS**.

HASTA AHORA, HEMOS ASUMIDO QUE TODOS LOS DÍAS DE LA SEMANA SON IGUALES, PERO TAMBIÉN PODEMOS CONTROLAR ESTE ASPECTO DE LA SIGUIENTE FORMA: USANDO SÓLO **CUATRO TAXIS** Y ASIGNÁNDOLES EL TRATAMIENTO COMO EN LA TABLA DE LA DERECHA:

		DÍA 1	2	3	4
TAXI	1	*a*	*b*	*c*	*d*
	2	*b*	*c*	*d*	*a*
	3	*c*	*d*	*a*	*b*
	4	*d*	*a*	*b*	*c*

UNA TABLA CUATRO POR CUATRO, CON CUATRO ELEMENTOS DIFERENTES, EN LA QUE APARECEN TODOS LOS ELEMENTOS EN LAS CUATRO COLUMNAS Y EN LAS CUATRO FILAS, RECIBE EL NOMBRE DE **cuadrado latino.** EN ESTE EXPERIMENTO, LOS CUATRO DÍAS Y LOS CUATRO TAXIS RECIBEN LOS CUATRO TRATAMIENTOS UNA VEZ.



EL PASO DE LA ALEATORIZACIÓN ELIGE AL AZAR UN ÚNICO CUADRADO LATINO ENTRE TODOS LOS POSIBLES.

SI LAS CUATRO UNIDADES NO SON SUFICIENTES, PODEMOS AUMENTAR EL NÚMERO DE UNIDADES EXPERIMENTALES REPITIENDO EL DISEÑO EXPERIMENTAL. SI EMPEZAMOS CON OCHO TAXIS, PODEMOS DIVIDIRLOS EN DOS GRUPOS DE CUATRO Y REPETIR EL DISEÑO EN CADA GRUPO.

HEMOS PROMETIDO NO ENTRAR EN DETALLES DEL ANÁLISIS DE DATOS, PERO ÉSTA ES, *GROSSO MODO*, LA FORMA DE TRATAR UN TIPO COMPLEJO DE DISEÑO COMO ESTE.

LOS DISEÑOS EXPERIMENTALES SE ANALIZAN DISTRIBUYENDO LA VARIABILIDAD TOTAL ENTRE LAS DIFERENTES FUENTES. EN EL EJEMPLO DE LOS TAXIS, LAS FUENTES DE VARIABILIDAD SON: EL TAXI, LA MARCA DEL NEUMÁTICO, EL TIPO DE GASOLINA, EL DÍA Y EL ERROR ALEATORIO. EL ANÁLISIS DE LA VARIANZA, **ANOVA*** PARA ABREVIAR, DIVIDE LA VARIACIÓN TOTAL EN PARTES Y LOCALIZA LAS PORCIONES DE CADA FUENTE.

EN EL SIGUIENTE CAPÍTULO, EXPLICAREMOS DETALLADAMENTE UN TIPO DE MODELO PARA ANALIZAR DISEÑOS COMPLEJOS: **EL MODELO DE REGRESIÓN LINEAL.** EN LA REGRESIÓN LINEAL PODRÁS VER EL **ANOVA** DE CERCA Y EXPRESADO EN FORMA NUMÉRICA...

* AUNQUE EN CASTELLANO SERÍA MÁS CORRECTO ANVA O ANDEVA, SEGUIREMOS LAS SIGLAS INGLESAS ANOVA. [N.T.]