Gobierno del Estado de Sonora | Secretaría de Salud Pública

La prueba de hipótesis

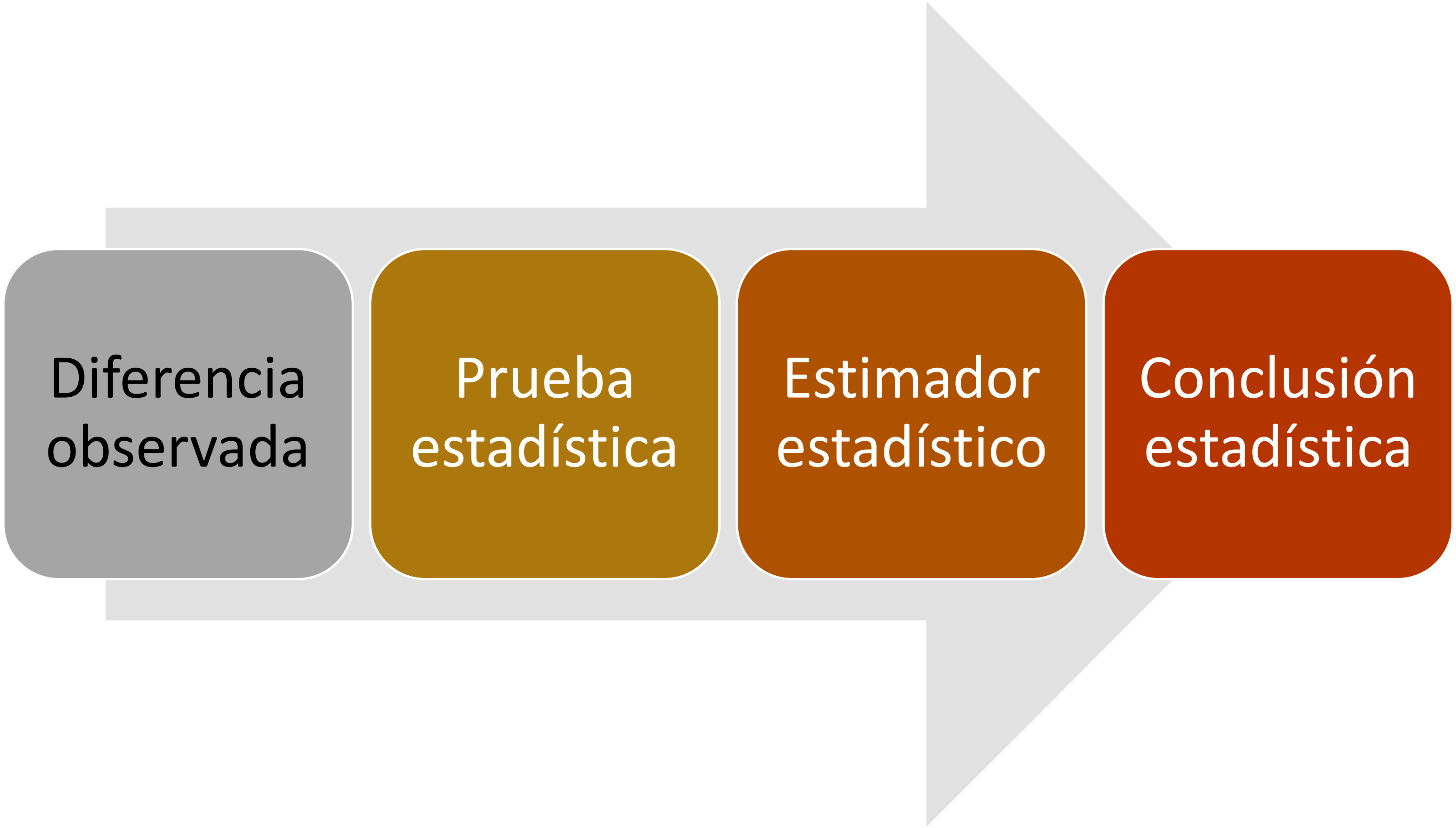
Por: Gerardo Álvarez Hernández PhD

SONORA
UNIDOS LOGRAMOS MÁS

Probando causalidad

- Si tenemos dos grupos de estudio, los dos grupos de resultados que se obtengan nunca serán idénticos
- Habrá diferencias aun cuando sean muestras representativas extraídas del mismo universo, debido a errores de muestreo (rol del azar)
- Los resultados deben entonces ser sometidos a procedimientos estadísticos, para probar su ***significancia***
- Significancia implica que la diferencia es real y no es debida al azar

El principio básico



- ¿La diferencia observada es real y no es debida al azar?

Significancia clínica vs. Significancia estadística

- Significancia estadística no implica que la diferencia observada es desde el punto de vista clínico, “relevante”
- Con muestras grandes, diferencias muy pequeñas con poca o ninguna importancia clínica, pueden ser estadísticamente significativas
- Con muestras de tamaño pequeño, una diferencia clínicamente significativa puede no serlo estadísticamente
- Las implicaciones prácticas de cualquier hallazgo deben ser juzgadas en otros terrenos más allá del estadístico

Regla de Oro

- Una diferencia estadísticamente significativa podría no ser de importancia clínica
- Una diferencia clínicamente significativa (relevante) podría no ser estadísticamente significativa
- Siempre comenzar por un juicio clínico/epidemiológico y apoyarlo estadísticamente
- La determinación del tamaño de la muestra es un paso inicial crucial

Pruebas de hipótesis

- Una hipótesis estadística es una proposición acerca de los parámetros de una o más poblaciones
- Un parámetro tiene el objetivo de crear un modelo de la realidad
- Las hipótesis nulas plantean una igualdad
    - La prevalencia de paludismo en ambos grupos es de 12%
    - La tasa de recurrencia es igual entre ambos tratamientos
    - La desnutrición es independiente del nivel de pobreza

Gutiérrez-Villafuerte, 2008

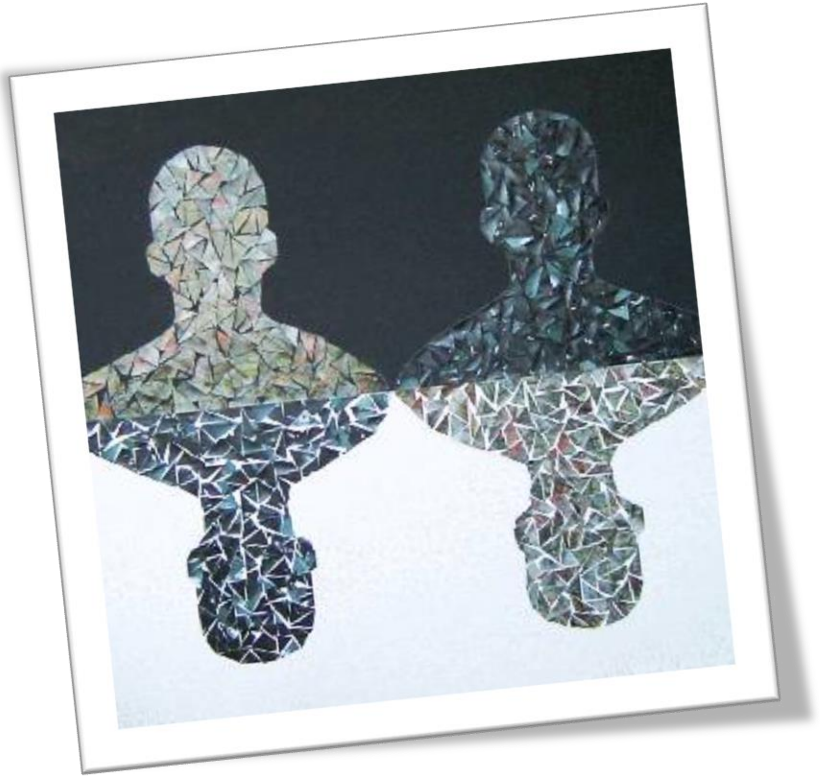
Prueba de hipótesis

- Meta: hacer un pronunciamiento (s) acerca de parámetros poblacionales desconocidos ***basados en datos muestrales***
- Elementos de una prueba de hipótesis:

1. **Hipótesis nula (Ho) –** Enunciado respecto a los valores de parámetros desconocidos
    - Típicamente implica que **no hay asociación entre las variables o no existen diferencias** (contienen las mismas cantidades)

Prueba de hipótesis

2. **Hipótesis alterna (Ha)** - Enunciado contrario a la Ho (siempre contendrá una diferencia)
3. **Prueba estadística-** cantidad basada en una muestra de datos, usada para probar la hipótesis nula
4. **Región de rechazo –** Valores de la prueba estadística por la que rechazaremos la Ho en favor de Ha



¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

- **La prueba de hipótesis es una parte integral de la bioestadística y comprende los siguientes pasos:**

1 Especificar la hipótesis nula (Ho) y la alterna (Ha)

- ***Ho: el valor hipotetizado de μ que pretendemos rechazar***

- Ho: la hipótesis de no asociación (p.e. la enfermedad y la exposición no están asociadas)

- Ho: $\beta_1 = 0$; Ho: $\mu_1 = \mu_2$

- Ha: Otros valores que pretendemos aceptar

- Ha: la hipótesis de asociación (p.e. la enfermedad y la exposición están asociadas)

- Ha: $\beta_1 \neq 0$; Ha: $\mu_1 \neq \mu_2$

¿Cómo podemos determinar qué tan probable es que la hipótesis causal es correcta?

- El colesterol sérico tiene una distribución aproximadamente normal en hombres sanos de 20-74 años, con una media poblacional de 211 mg/dl
- Suponga que tenemos una muestra de 25 sujetos hipertensos con un nivel medio de colesterol sérico de 220 mg/dl y una desviación estándar de 38.6 mg/dl

- Ho: $\mu = 211$ mg/dl
- Ha: $\mu \neq 211$ mg/dl

¿Cómo podemos determinar qué tan probable es que la hipótesis causal es correcta?

2

Especificar un nivel de significancia estadística

- α = nivel de significancia
- Es la probabilidad de que rechazaremos Ho cuando en realidad es verdadera
- Esto se conoce como *error tipo I*
- ***¿Cuán probable es que cometamos un error y concluyamos que Ho es falsa cuando en realidad es correcta?***

- Por supuesto, queremos que esta probabilidad sea lo mas "pequeña" posible
- Usualmente es = 0.05, pero puede ser cualquier valor que el investigador desee

Suponga que estamos en un juicio

Sentencia del Juez	Condición real del acusado: Inocente	Condición real del acusado: Culpable
Inocente	Correcto	Error Tipo II β
Culpable	Error Tipo I α	Correcto

Gutiérrez-Villafuerte, 2008

El error tipo II (beta) significa que *estamos negando un efecto que en realidad existe*

		Población (Hipótesis Nula)	
		Verdadera	Falsa
Resultado de la Prueba de Ho	No rechazar Ho	Correcto	Error tipo II β
	Rechazar Ho	Error tipo I α	Correcto

Es la probabilidad de rechazar Ho cuando en realidad es verdadera

Gutiérrez-Villafuerte, 2008

Tipos de errores

- Recuerda: nunca podemos estar seguros de que estamos correctos aceptando o rechazando una hipótesis… estamos % confiados
- Ejemplo
  - Ho: El medicamento ‘X’ no tiene ningún efecto.
    - Es decir, no hay diferencia en la tasa de mortalidad entre pacientes usando el medicamento ‘X’ y los que no lo usan
  - Ha: El medicamento ‘X’ tiene un efecto
    - El medicamento reduce la tasa de mortalidad en quienes lo usan

Tipos de errores

	El medicamento no tiene efecto **Ho es verdadera**	El medicamento tiene efecto **Ho es falsa** Ha es verdadera
No rechazamos Ho No efecto del medicamento	**NO ERROR**	**ERROR TIPO II (Beta)**
Rechazamos Ho Hay un efecto del medicamento (aceptamos Ha)	**ERROR TIPO I (Alfa)**	**NO ERROR**

- Si no rechazamos Ho, concluimos que no hay una asociación entre el medicamento y la mortalidad
- Si rechazamos Ho y aceptamos Ha, concluimos que hay una relación entre el medicamento y la mortalidad

Consecuencias de los errores tipo I y II

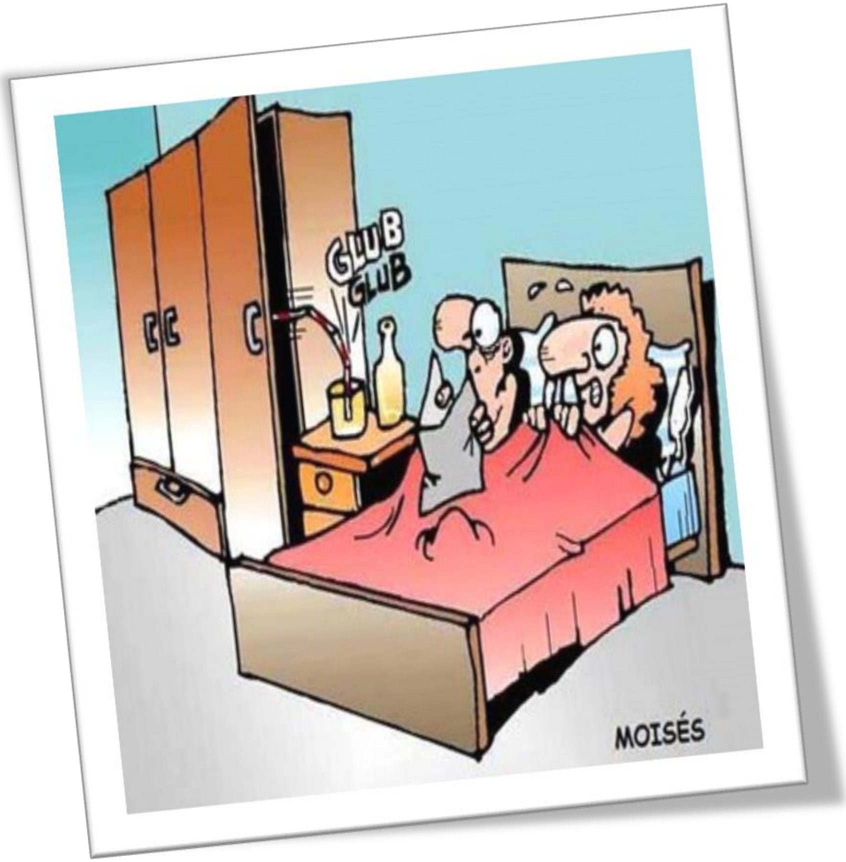
- La seriedad del error depende del problema investigado
- El error tipo I (**α**) significa que **estamos aceptando la presencia de un efecto cuando en realidad no existe**
- El error tipo II (**β**) significa que **negamos un efecto que en realidad existe**

Consecuencias de los errores tipo I y II

- Si buscamos una cura para el cáncer, el error tipo II sería bastante serio (estaríamos negando algo que es realmente útil)
- Por otro lado, si investigamos un medicamento costoso para tratar un resfriado, querríamos evitar el error tipo I, es decir no quisiéramos aceptar un efecto falso de un remedio caro para el resfriado

Consecuencias de los errores tipo I y II

- Un cuento para distinguir ambos errores:
- "Érase una vez un rey bastante celoso de su reina. Este rey tenía dos caballeros. Alfa, que era muy guapo y Beta que era bastante feo. Sucedió sin embargo, que la reina se enamoró de Beta. El rey sospechaba que la reina sostenía un romance con Alfa y que este tenía que ser decapitado"
- El rey cometió ambos tipos de errores: sospechaba una relación (con Alfa) que en realidad no existía, y falló en detectar la relación (con Beta) que en verdad existía



Wasserthail-Smoller, 2003

Consecuencias de los errores tipo I y II

- "La reina se escapó del reino con Beta y vivieron por siempre felices, mientras el rey sufría sentimientos de culpa acerca de su error y fatal rechazo de Alfa"
- Si se pretende disminuir la probabilidad del error tipo I, automáticamente se incrementa la probabilidad de ocurrencia del error tipo II
- Para bajar las probabilidades de ambos errores en un estudio, es necesario incrementar el número de observaciones (poder estadístico)

Wasserthail-Smoller, 2003

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

3 Especificar la prueba estadística

- Puede usar cualquier prueba, sin embargo, con frecuencia dicha prueba tomará la forma de un valor de ‘t’:
- $$T = \frac{\text{Estimador muestral – valor de la hipótesis nula}}{\text{Desviación estándar del estimador}}$$
- En nuestro ejemplo estamos probando la media poblacional (μ), así que la prueba estadística será:

$$T = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}$$

Media poblacional

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

4 **Especificar la distribución nula de la prueba estadística**

- Cuando Ho es verdadera, T tiene una distribución *t* con (n-1) grados de libertad
- Para nuestro ejemplo, la prueba estadística tendrá una distribución *t* con 24 grados de libertad

5 **Calcular la prueba estadística**

$$T = \frac{220 - 211}{38.6 / \sqrt{25}}$$

$$T = \frac{220 - 211}{38.6 / \sqrt{25}} = 1.17$$

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

6 Estime el valor de p

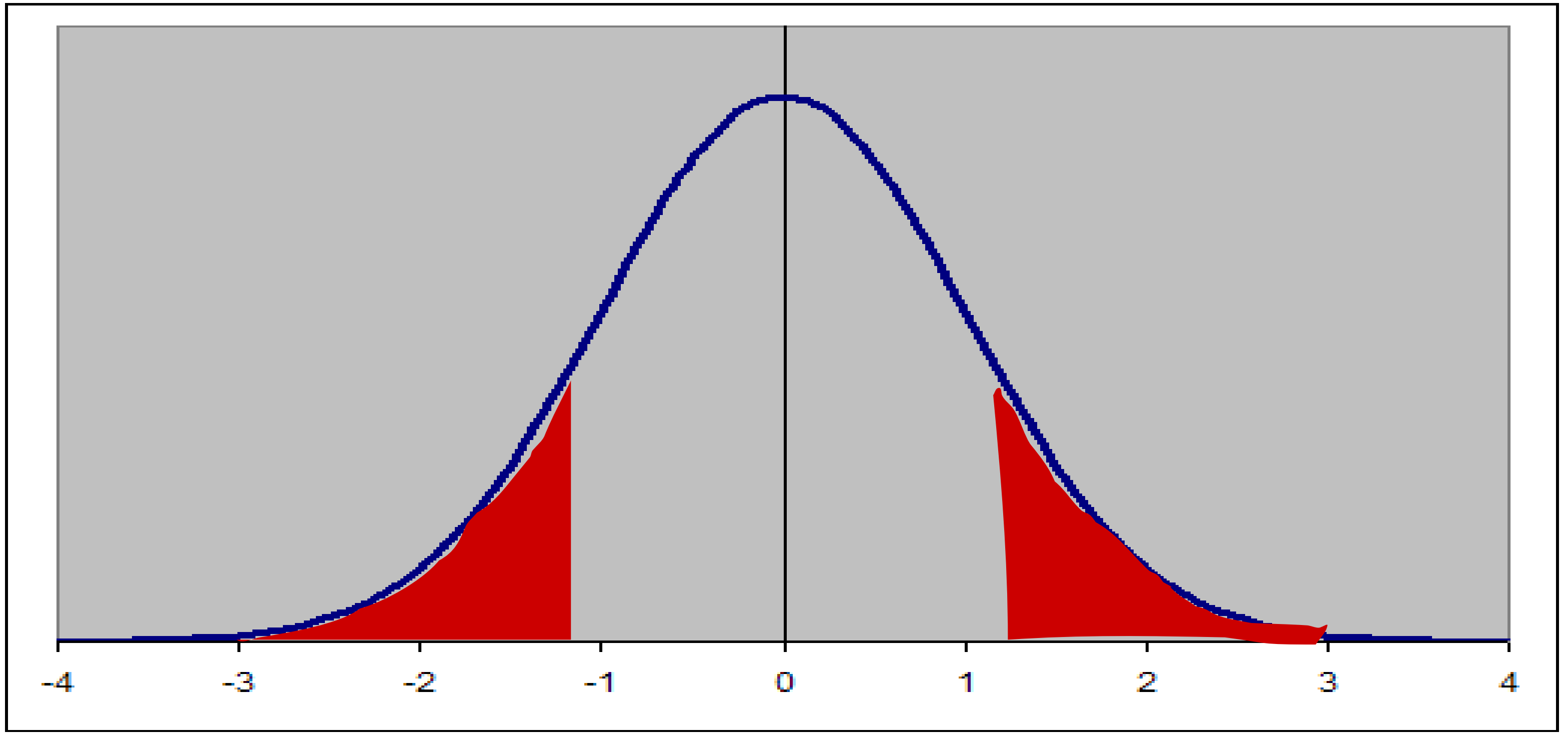
- Necesitamos decidir si nuestra prueba estadística proporciona suficiente evidencia para elegir *Ha* en lugar de *Ho*
- En nuestro ejemplo:

  - ¿Cuál es la probabilidad de que si reclutamos 25 sujetos conseguiríamos un estadístico *'t'* tan grande <u>en magnitud</u> como el que observamos en la población?

  - Usamos nuestra distribución nula y determinamos qué tanta área está a la derecha de 1.17 y a la izquierda de (-1.17)

  Valor de p = Prob (/T/ $\geq$ 1.17)

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?



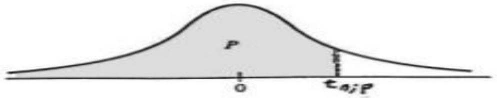
El área sombreada es = 0.24 (valor de p del estadístico T = 1.17)

Esta es la probabilidad de observar nuestro resultado, o incluso cualquier otro valor,

que no apoye la hipótesis nula (si Ho es verdadera)

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

Distribución t de Student



La tabla A.4 da distintos valores de la función de distribución en relación con el número de grados de libertad; concretamente, relaciona los valores p y $t_{n;p}$ que satisfacen

$$P(t_n \leq t_{n;p}) = p.$$

n	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	0, 1584	0, 3249	0, 7265	1, 3764	3, 0777	6, 3138	12, 7062	31, 8205	63, 6567
2	0, 1421	0, 2887	0, 6172	1, 0607	1, 8856	2, 9200	4, 3027	6, 9646	9, 9248
3	0, 1366	0, 2767	0, 5844	0, 9785	1, 6377	2, 3534	3, 1824	4, 5407	5, 8409
4	0, 1338	0, 2707	0, 5686	0, 9410	1, 5332	2, 1318	2, 7764	3, 7469	4, 6041
5	0, 1322	0, 2672	0, 5594	0, 9195	1, 4759	2, 0150	2, 5706	3, 3649	4, 0321
6	0, 1311	0, 2648	0, 5534	0, 9057	1, 4398	1, 9432	2, 4469	3, 1427	3, 7074
7	0, 1303	0, 2632	0, 5491	0, 8960	1, 4149	1, 8946	2, 3646	2, 9980	3, 4995
8	0, 1297	0, 2619	0, 5459	0, 8889	1, 3968	1, 8595	2, 3060	2, 8965	3, 3554
9	0, 1293	0, 2610	0, 5435	0, 8834	1, 3830	1, 8331	2, 2622	2, 8214	3, 2498
10	0, 1289	0, 2602	0, 5415	0, 8791	1, 3722	1, 8125	2, 2281	2, 7638	3, 1693
11	0, 1286	0, 2596	0, 5399	0, 8755	1, 3634	1, 7959	2, 2010	2, 7181	3, 1058
12	0, 1283	0, 2590	0, 5386	0, 8726	1, 3562	1, 7823	2, 1788	2, 6810	3, 0545
13	0, 1281	0, 2586	0, 5375	0, 8702	1, 3502	1, 7709	2, 1604	2, 6503	3, 0123
14	0, 1280	0, 2582	0, 5366	0, 8681	1, 3450	1, 7613	2, 1448	2, 6245	2, 9768
15	0, 1278	0, 2579	0, 5357	0, 8662	1, 3406	1, 7531	2, 1314	2, 6025	2, 9467
16	0, 1277	0, 2576	0, 5350	0, 8647	1, 3368	1, 7459	2, 1199	2, 5835	2, 9208
17	0, 1276	0, 2573	0, 5344	0, 8633	1, 3334	1, 7396	2, 1098	2, 5669	2, 8982
18	0, 1274	0, 2571	0, 5338	0, 8620	1, 3304	1, 7341	2, 1009	2, 5524	2, 8784
19	0, 1274	0, 2569	0, 5333	0, 8610	1, 3277	1, 7291	2, 0930	2, 5395	2, 8609
20	0, 1273	0, 2567	0, 5329	0, 8600	1, 3253	1, 7247	2, 0860	2, 5280	2, 8453
21	0, 1272	0, 2566	0, 5325	0, 8591	1, 3232	1, 7207	2, 0796	2, 5176	2, 8314
22	0, 1271	0, 2564	0, 5321	0, 8583	1, 3212	1, 7171	2, 0739	2, 5083	2, 8188
23	0, 1271	0, 2563	0, 5317	0, 8575	1, 3195	1, 7139	[illegible]	2, 4999	2, 8073
24	0, 1270	0, 2562	0, 5314	0, 8569	1, 3178	1, 7109	2, 0639	2, 4922	2, 7969
25	0, 1269	0, 2561	0, 5312	0, 8562	1, 3163	1, 7081	[illegible]	2, 4851	2, 7874
26	0, 1269	0, 2560	0, 5309	0, 8557	1, 3150	1, 7056	2, 0555	2, 4786	2, 7787
27	0, 1268	0, 2559	0, 5306	0, 8551	1, 3137	1, 7033	2, 0518	2, 4727	2, 7707
28	0, 1268	0, 2558	0, 5304	0, 8546	1, 3125	1, 7011	2, 0484	2, 4671	2, 7633
29	0, 1268	0, 2557	0, 5302	0, 8542	1, 3114	1, 6991	2, 0452	2, 4620	2, 7564
30	0, 1267	0, 2556	0, 5300	0, 8538	1, 3104	1, 6973	2, 0423	2, 4573	2, 7500
40	0, 1265	0, 2550	0, 5286	0, 8507	1, 3031	1, 6839	2, 0211	2, 4233	2, 7045
50	0, 1263	0, 2547	0, 5278	0, 8489	1, 2987	1, 6759	2, 0086	2, 4033	2, 6778
60	0, 1262	0, 2545	0, 5272	0, 8477	1, 2958	1, 6706	2, 0003	2, 3901	2, 6603
80	0, 1261	0, 2542	0, 5265	0, 8461	1, 2922	1, 6641	1, 9901	2, 3739	2, 6387
100	0, 1260	0, 2540	0, 5261	0, 8452	1, 2901	1, 6602	1, 9840	2, 3642	2, 6259
120	0, 1259	0, 2539	0, 5258	0, 8446	1, 2886	1, 6577	1, 9799	2, 3578	2, 6174
∞	0, 126	0, 253	0, 524	0, 842	1, 282	1, 645	1, 960	2, 327	2, 576

Tabla A.4: Tabla de la distribución t de Student.

- De manera simple:
- Para rechazar la Ho, la ‘T’ calculada debe ser mayor a la ‘T’ crítica
- T calculada = 1.17
- T crítica = 2.0639
- ¿Podemos rechazar la Ho?

¿Cómo podemos determinar qué tan probable es que la hipótesis causal sea correcta?

7 **Establezca su conclusión**

- Dado que T calculada es menor a T crítica fallamos en rechazar Ho
- Toda vez que nuestro valor de ‘p’ (0.24) es mucho mayor que el nivel de significancia (0.05), *no podemos rechazar la hipótesis nula*
- Por lo tanto, concluimos que:

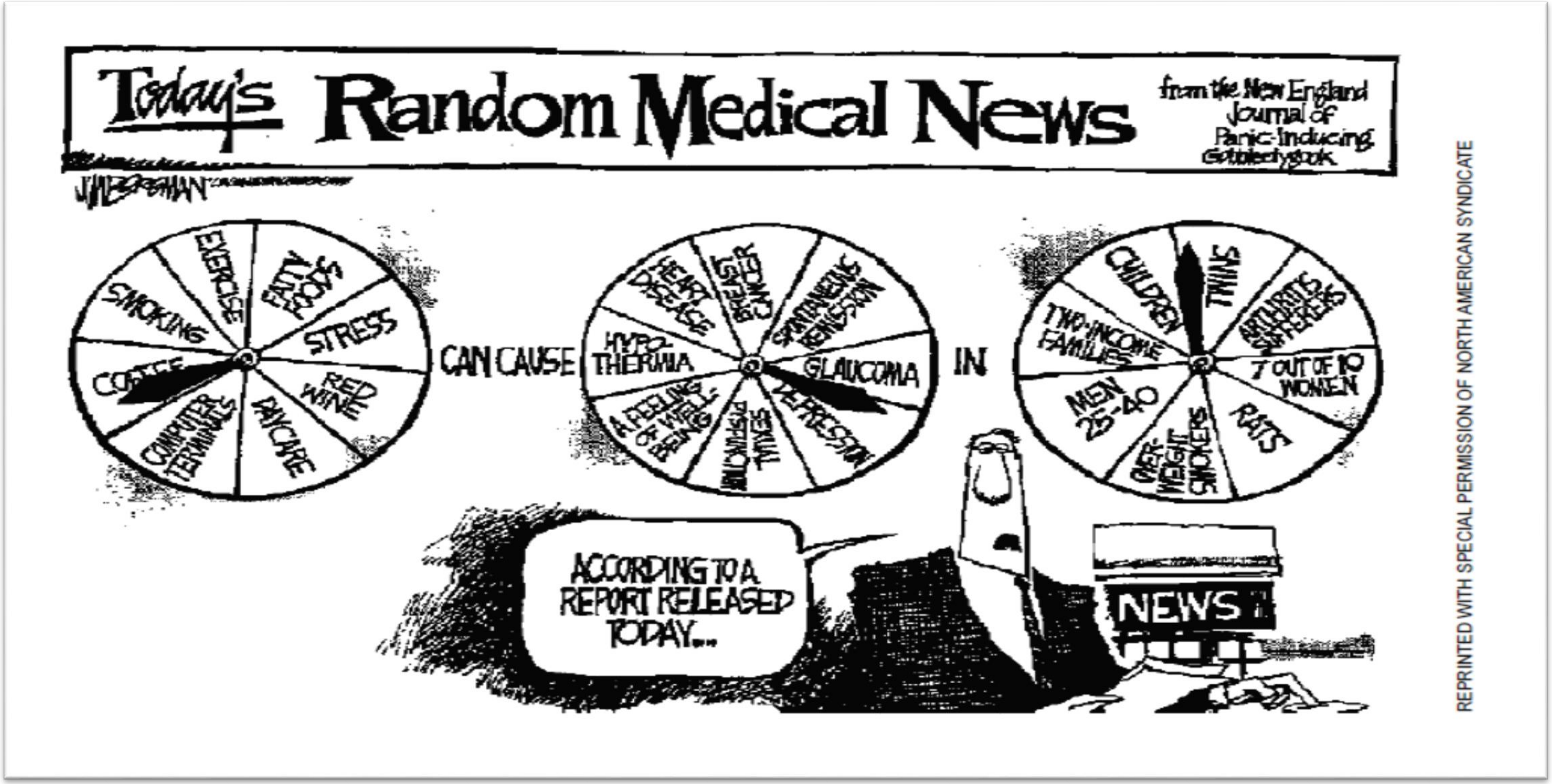
  - No tenemos suficiente evidencia de que el nivel medio de colesterol sérico de los hombres hipertensos de 20-74 años sea diferente al de un hombre sano.

  - ¿Puede explicar por qué?

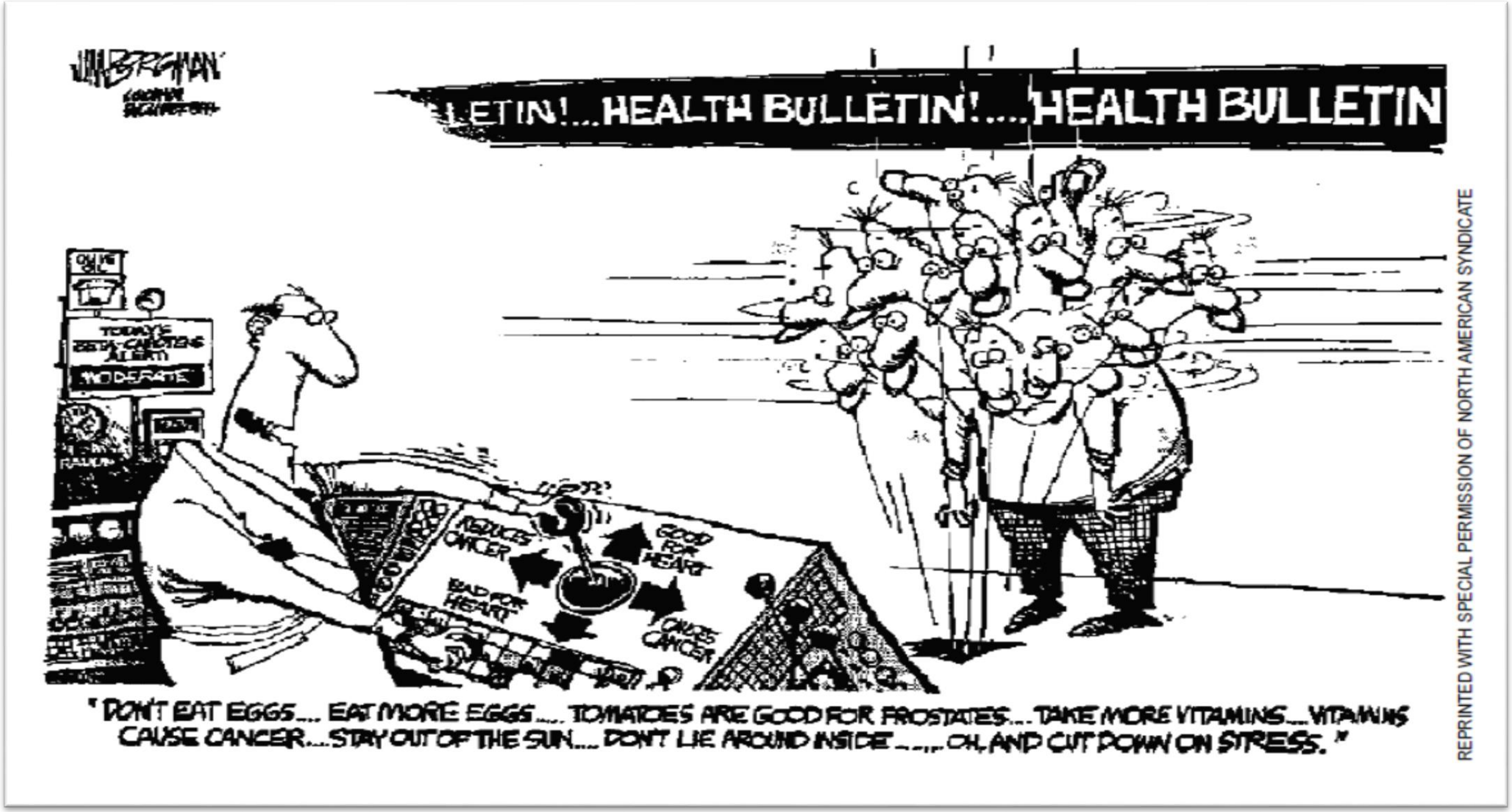
Artículo de revisión

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests?
BMJ 2001; 322: 226-31

The epidemiological findings are deceptive



…What medical research, particularly epidemiologists, produce is gobbledegook



Summary points

P values, or significance levels, measure the strength of the evidence against the null hypothesis; the smaller the P value, the stronger the evidence against the null hypothesis

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests? BMJ 2001; 322: 226-31

An arbitrary division of results, into “significant” or “non-significant” according to the P value, was not the intention of the founders of statistical inference

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests? BMJ 2001; 322: 226-31

A P value of 0.05 need not provide strong evidence against the null hypothesis, but it is reasonable to say that $P < 0.001$ does. In the results sections of papers the precise P value should be presented, without reference to arbitrary thresholds

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests? BMJ 2001; 322: 226-31

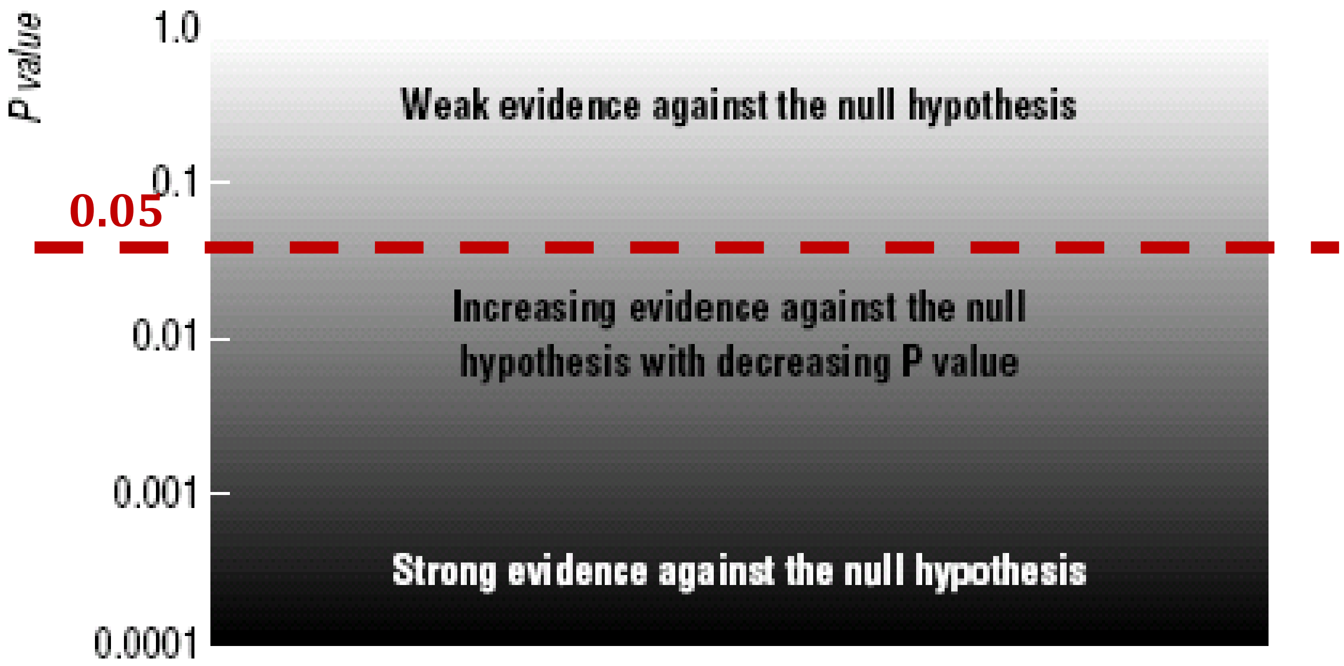
Results of medical research should not be reported as "significant" or "non-significant" but should be interpreted in the context of the type of study and other available evidence. Bias or confounding should always be considered for findings with low P values

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests? BMJ 2001; 322: 226-31

And epidemiological as well!!!!

To stop the discrediting of medical research by chance findings we need more powerful studies

Sterne JAC, Smith GD. Sifting the evidence—what´s wrong with significance tests? BMJ 2001; 322: 226-31



Suggested interpretation of P values from published medical research