UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
ENES MÉRIDA
LICENCIATURA EN ECOLOGÍA

ESTADÍSTICA APLICADA
Tema III. Prueba de Hipótesis (Parte 2)

Prof. Edlin J. Guerra Castro

Método hipotético Deductivo (Falsacionismo)

Observación
Patrón en espacio y tiempo

Modelo
Explicación o teoría

Hipótesis
Predicción basada en el modelo

Hipótesis nula
lógica opuesta a la hipótesis

Experimento/muestreo
Prueba crítica a la hipótesis nula

Estadística

Interpretación

Caso de estudio: Accidente nuclear de Fukushima y salud de peces

Japón, 2011. Terremoto de 9.0

Modelo: La radioactividad modifica la estructura molecular de las células, lo que compromete su funcionamiento. El accidente de Fukushima pudo afectar la salud de las poblaciones de peces.

Hipótesis: Si el accidente de Fukushima afectó la salud de los peces de la zona, se esperaría observar una modificación de al menos ***20 %*** en los patrones de crecimiento de los peces respecto a información previa.

Hipótesis Nula: Los patrones de crecimiento entre los peces de Fukusima luego del accidente no superan ***20%*** respecto a los patrones previos al accidente.

Hipótesis estadísticas:

$$H_i\colon \mu_{antes} > \mu_{despues}$$

$$H_i\colon \mu_{antes} - \mu_{despues} > 20\%$$

$$H_0\colon \mu_{antes} - \mu_{despues} \leq 20\%$$

Modelos lineales

Modelo lineal para la hipótesis nula

$$H_0\colon\ y_{ij} = \mu + \varepsilon_{ij}$$

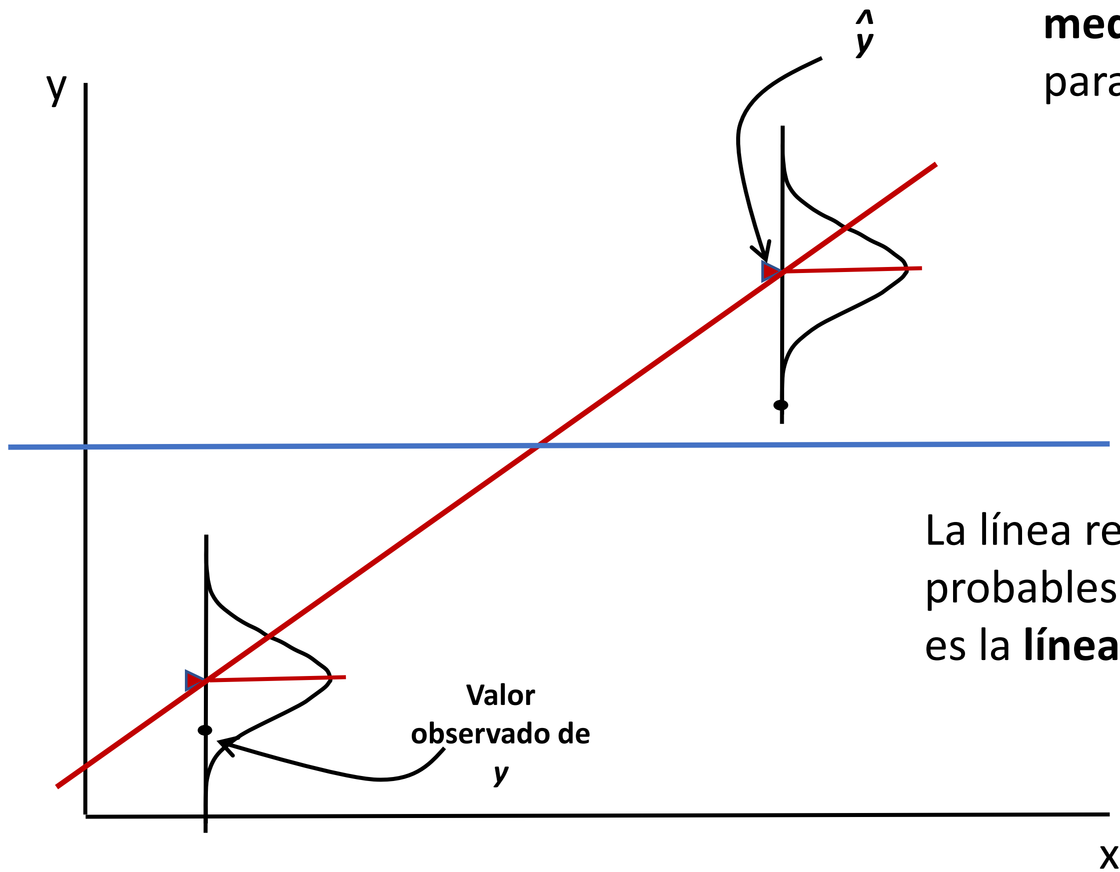
Modelo lineal para la hipótesis alternativa

$$H_A\colon y_{ij} = \mu + A_i + \varepsilon_{ij}$$

Modelos lineales

Si pensamos que los datos de *y* para cada *x* son una muestra representativa de una población de datos para ese valor de *x;*

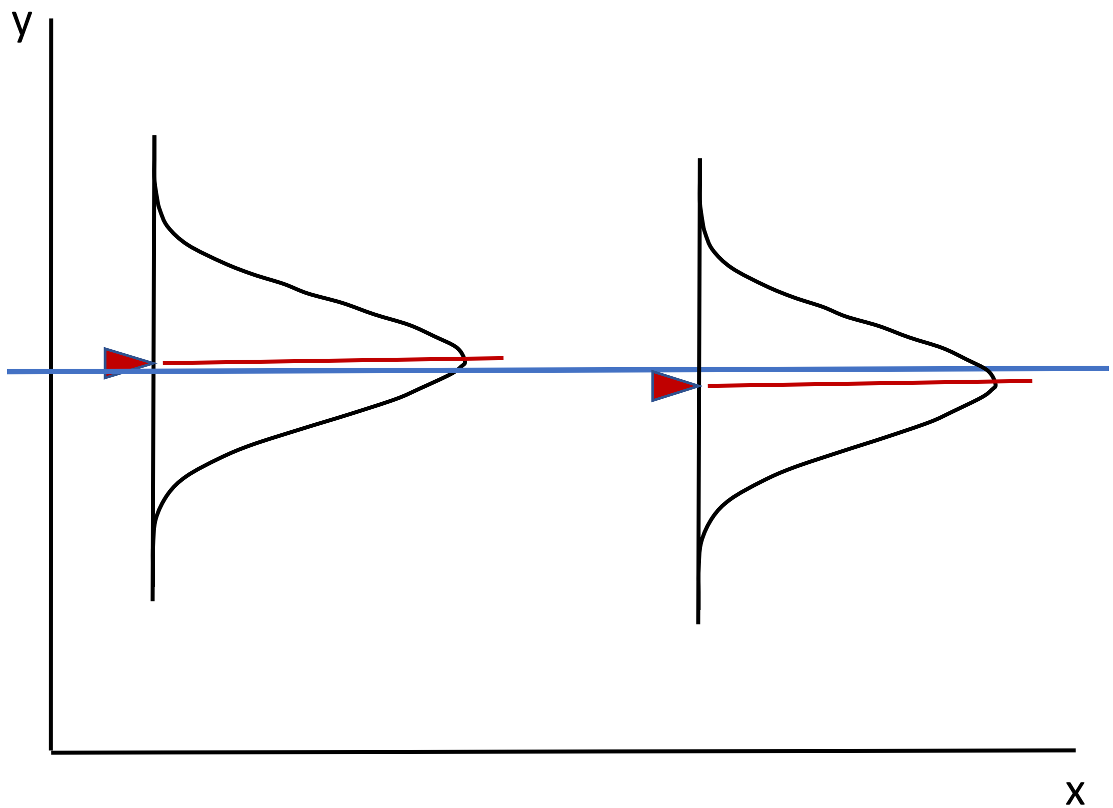
y que dicha población tiene una distribución normal cuya moda (y **media**) es el valor más probable de *y* para esa *x*; entonces



La línea recta que une los valores más probables de *y* para todos los valor de *x* es la **línea más probable**.

Modelos lineales

Si la hipótesis nula es cierta:



Modelos lineales

Casos simples: modelo de la media: una cola, dos colas, tipo de varianzas, tipo de distribución.

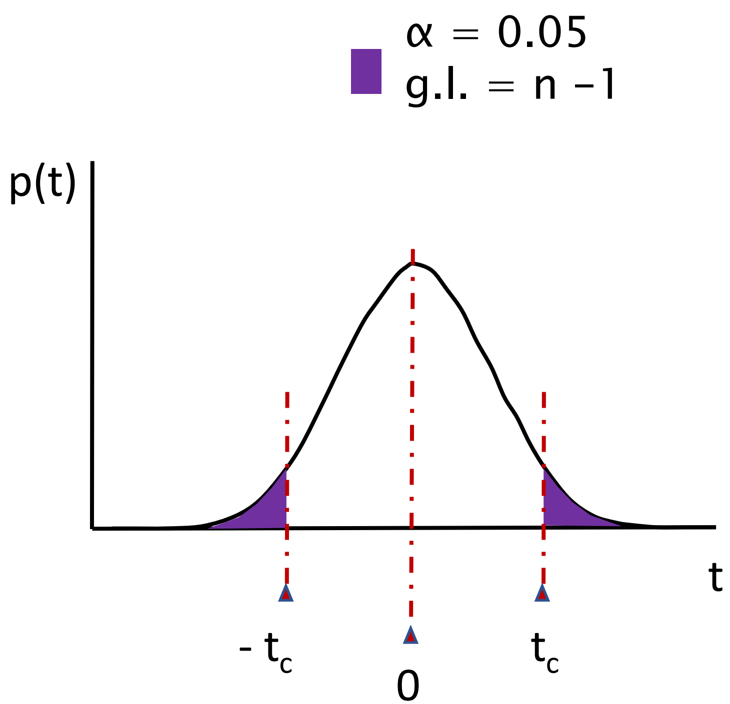
Casos más complejos:
1. Análisis de varianza: un factor
2. Análisis de varianza: dos o más factores
   a) Diseños ortogonales
   b) Diseños anidados
   c) Diseños mixtos
3. Regresión lineal simple
4. Regresión lineal múltiple
5. Análisis de covarianza

Evaluación de supuestos

Normalidad
Homocedasticidad
Independencia
Aditividad

Casos simples: modelo de la media con la prueba *t*

Distribución de *t* de Student



Distribución continua que surge cuando se estima la media de una población (con distribución aproximadamente normal) en situaciones en las que no se conoce la desviación estándar de la población.

Es una distribución simétrica alrededor del 0 y con forma de campana, y lleva siempre asociados grados de libertad gl=n-1 (por estimar el valor de **σ** a partir de **s**).

$$\overline{X} - (t_{\alpha,gl} \bullet \sqrt{\frac{S^2}{n}}) \leq \mu \leq \overline{X} + (t_{\alpha,gl} \bullet \sqrt{\frac{S^2}{n}})$$

El 95% de los valores de **μ** se encuentran entre un valor inferior y uno superior de $\overline{\mathbf{X}}$

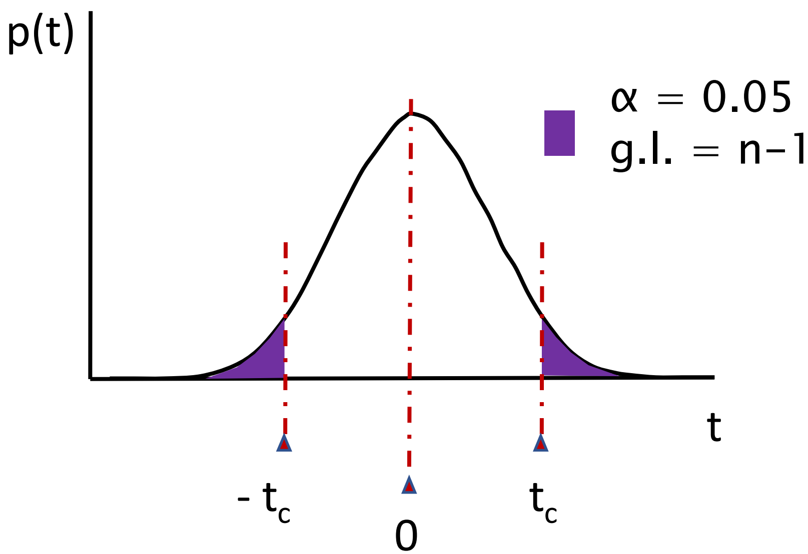
Distribución de t de Student

(una sola muestra)

$$t = \frac{\overline{X} - \mu_0}{S_{\overline{X}}}$$

Diferencia entre media de la muestra y un valor de referencia (fijo)

Error estándar: medida de la precisión para estimar la media a partir de la muestra



La distribución de probabilidades de valores de t Student bajo la hipótesis nula de que $t = \mu_0$

Bajo la H_0

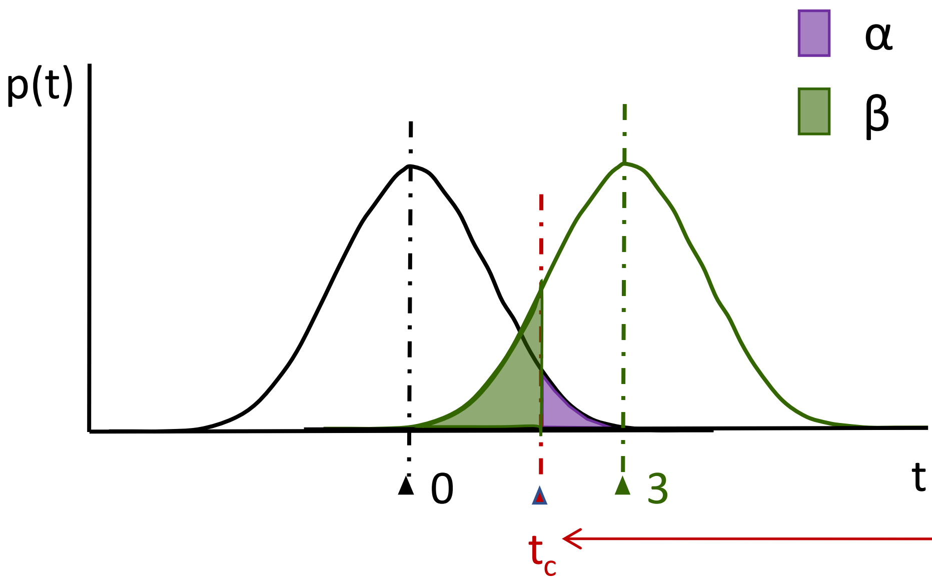
t ≈ 0 Tienen probabilidades altas de ocurrir

t >> 0
t << 0 Tienen probabilidades bajas de ocurrir

El error Tipo II (re-visitado)

En una prueba de t para una muestra, la H_0 es que no existen diferencias entre la media de la muestra y un valor μ de referencia.

La H_A de dicha prueba es que la diferencia entre la media de la muestra y μ es igual a 3; H_A: $\mu_{EST} - \mu_{REF} = 3$



— Distribución de probabilidades de los valores de *t* cuando se cumple la H_0

— Distribución de probabilidades de los valores de *t* cuando se cumple la H_A (especificada cuantitativamente).

α

Tamaño del efecto

$$t = \frac{\overline{X} - \mu_0}{S_{\overline{X}}}$$

n

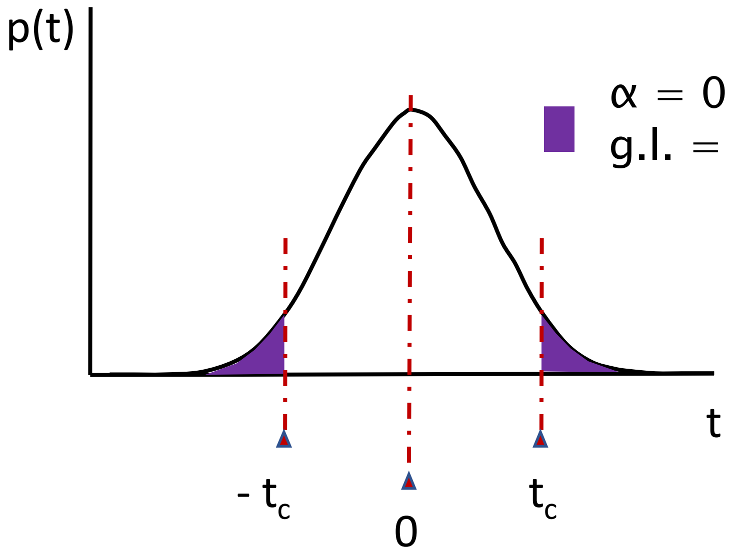
σ

Distribución de t de Student
(dos muestras independientes)

$$t = \frac{\overline{X}_A - \overline{X}_B}{S_p}$$

Diferencia entre las dos medias

Error estándar "pooled": medida de la precisión para estimar la diferencia de las medias.



$\alpha = 0.05$

$g.l. = n_A + n_B - 2$

La distribución de probabilidades de valores de t Student bajo la hipótesis nula de que $t = \mu_0$

Bajo la H_0

- $t \approx 0$ Tienen probabilidades altas de ocurrir
- $t >> 0$, $t << 0$ Tienen probabilidades bajas de ocurrir

Caso de estudio: Accidente nuclear de Fukushima y salud de peces

1. Definir la hipótesis nula (H_0)

$$H_0\colon \mu_{antes} - \mu_{despues} \leq 20\%$$

2. Elegir una prueba que mida la desviación del H_0 y que tenga un estadístico con distribución conocida.

Dos grupos, variable continua. Interesan diferencias en promedios. La prueba t-student es buena alternativa.

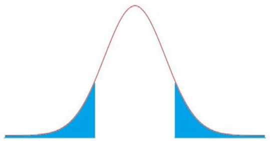
3. Definir el criterio de rechazo de la H_0

$$\alpha < 0.05 \mid t_{obs} > t_{c(gl,\,\alpha)}$$

4. Con los datos del estudio, calcular el estadístico elegido

$$t = \frac{\overline{X}_A - \overline{X}_B}{S_p}$$

5. Determinar la probabilidad asociada de obtener nuestro valor muestral del estadístico si la H_0 es cierta



6. Rechazar H_0 si se cumple el criterio de rechazo; retener de lo contrario.

La necesidad de la Prueba estadística

Los muestreos están sujetos a error muestral y a la variabilidad natural de las cosas naturales

Estamos lidiando con respuestas probabilísticas, no con conclusiones absolutas