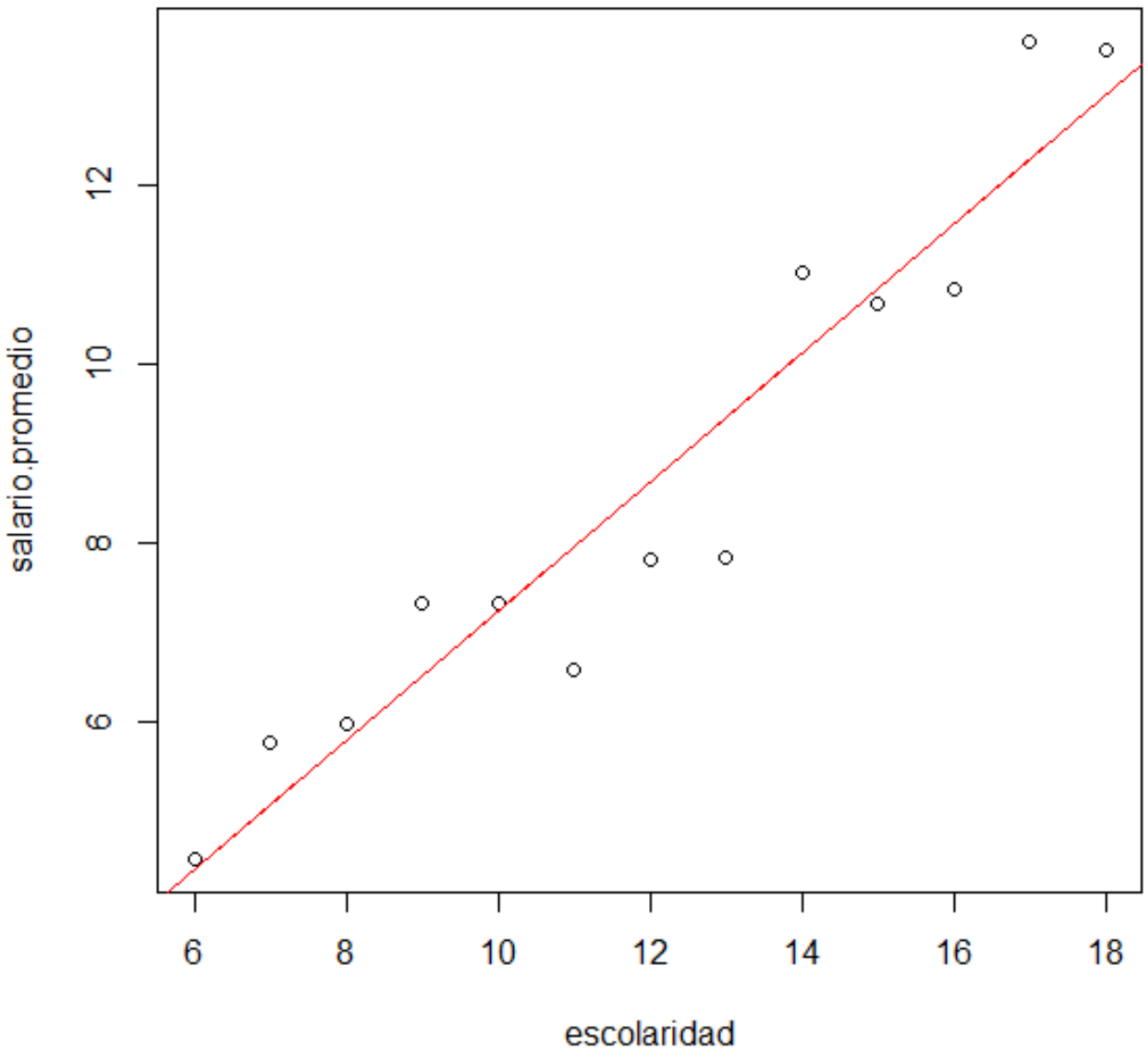
Estimación por intervalos y pruebas de hipótesis

Regresión con dos variables

Estimación por intervalos

Para entrar en contexto, tomemos el ejemplo de la tabla 2.6.



$$\hat{Y}_i = -0.0144 + 0.7240\, X_i$$

- *“Cuando la escolaridad (X_i) aumenta un año, el incremento <u>promedio</u> del salario es de 0.7240 ($\hat{\beta}_2$).*
  - $\hat{\beta}_2$ es una **estimación puntual** del valor verdadero, β_2.
- ¿Qué tan confiable es un estimador?
  - Lo medimos a través de su **error estándar**.
  - $\Pr\left(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta\right) = 1 - \alpha$
    - $\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta$ es el intervalo de confianza
    - $1 - \alpha$ es el nivel de confianza *(la probabilidad de que el valor verdadero de β_2 se encuentre dentro de ese intervalo)*
    - α es el nivel de significancia

- Si conocemos las distribuciones de probabilidad de los estimadores, podemos encontrar sus intervalos de confianza.
  - Gracias al supuesto de normalidad de las perturbaciones, los estimadores de MCO-OLS, $\hat{\beta}_1, \hat{\beta}_2$ están normalmente distribuidos.

- Puesto que no se conoce σ y solo tenemos $\hat{\sigma}$, podemos expresarlo como

$$t = \frac{\hat{\beta}_2 - \beta_2}{ee(\hat{\beta}_2)} = \frac{Estimador - parámetro}{Error\ estándar\ estimado\ del\ estimador}$$

  - La variable t sigue una distribución t con $n - k\ g.l.$ (grados de libertad), donde
    - n es el número de observaciones de la muestra y
    - k la cantidad de parámetros a estimar.

- Entonces, construyendo el intervalo de confianza para β_2 mediante la distribución t tenemos:

$$\Pr\left(-t_{\alpha/2} \leq t \leq t_{\alpha/2}\right) = 1 - \alpha$$

  - Donde:
  - $t_{\alpha/2}$ es el valor de la variable t obtenida de la distribución para un nivel de significancia de $\alpha/2$ y $n - k\ g.l.$ (se le conoce como **nivel crítico**).

- Sustituyendo el valor de t, tenemos que:

$$\Pr\left(-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2}{ee(\hat{\beta}_2)} \leq t_{\alpha/2}\right) = 1 - \alpha$$

- Despejando β_2, obtenemos el **intervalo de confianza** para β_2

$$\Pr\left(\hat{\beta}_2 - t_{\alpha/2}\, ee(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\, ee(\hat{\beta}_2)\right) = 1 - \alpha$$

- Escrito de forma compacta:

$$\hat{\beta}_2 \pm t_{\alpha/2}\, ee(\hat{\beta}_2)$$

- El intervalo de confianza para β_1:

$$\hat{\beta}_1 \pm t_{\alpha/2}\, ee(\hat{\beta}_1)$$

Pruebas de hipótesis

- ¿Es compatible o no una observación, de acuerdo a las hipótesis que planteamos?

- La hipótesis planteada:
  - Hipótesis nula, H_0
- Se suele contrastar vs. una
  - Hipótesis alternativa, H_1

Teoría de las pruebas de hipótesis

Diseño de reglas o procedimientos que permitan decidir si se rechaza o no la hipótesis nula.

Para diseñar las reglas, tenemos los métodos *mutuamente complementarios*:

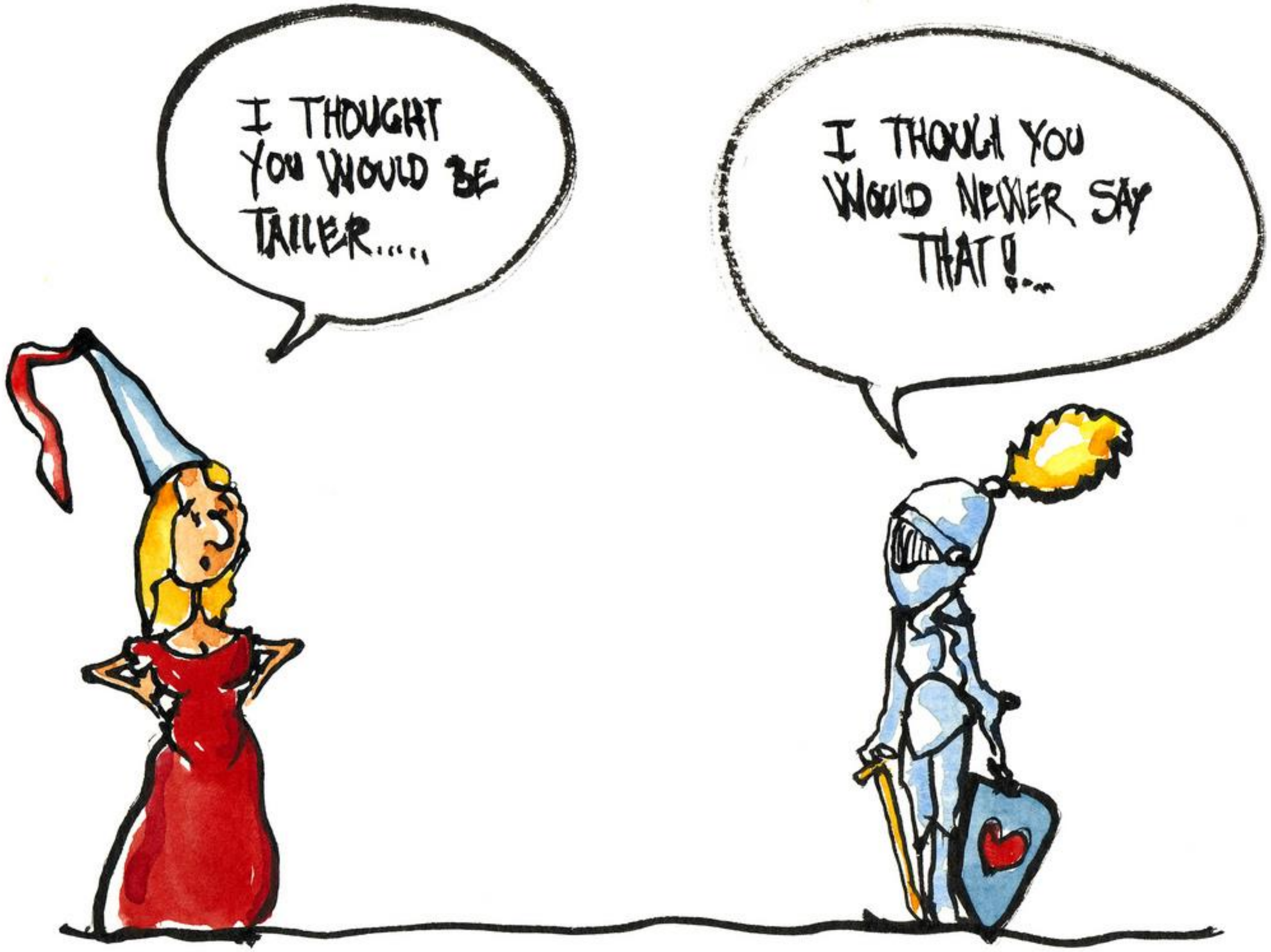
- Intervalos de confianza
- Pruebas de significancia

Método del intervalo de confianza

- **Prueba de dos colas (bilateral)**
  - Si, para el ejemplo de salarios, donde $\hat{\beta}_2 = 0.724$, suponemos que:
    - $H_0: \beta_2 = 0.5$
    - $H_1: \beta_2 \neq 0.5$
  - Si β_2 se encuentra dentro del intervalo de confianza, **no rechazamos la $\boldsymbol{H_0}$**.
  - *Aceptar ≠ no rechazar.*
  - Cuando se rechaza la H_0, decimos que el resultado es **estadísticamente significativo.**

Método del intervalo de confianza

- Prueba de una cola (unilateral):
  - $\exists n$ expectativas *a priori* de que la H_1 sea unilateral:
    - $H_0 : \beta_2 \leq 0.5$
    - $H_1 : \beta_2 > 0.5$

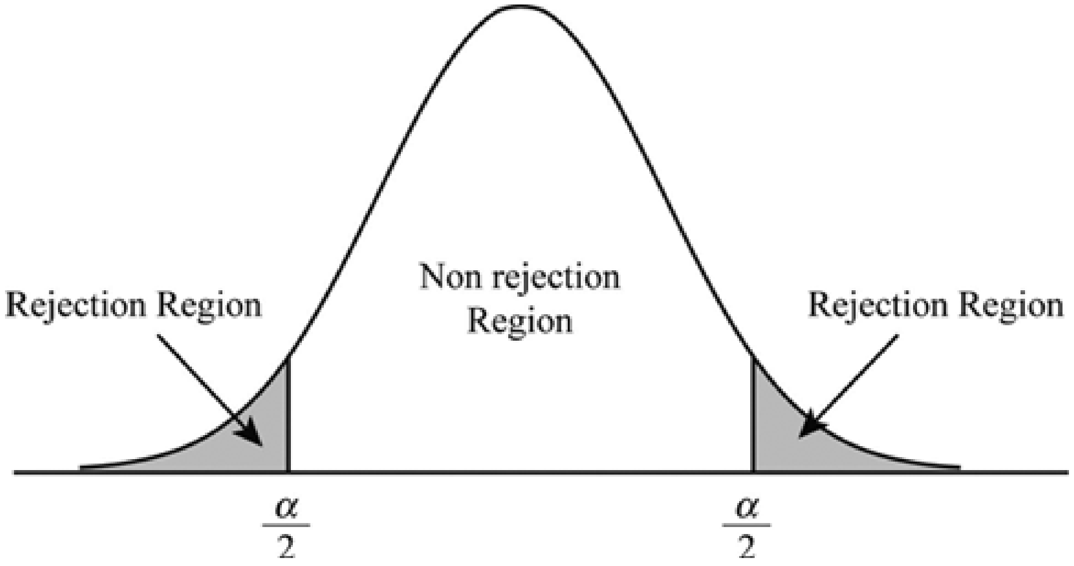


Pruebas de significancia

Pruebas de significancia de los coeficientes de regresión: la prueba t

- Prueba de significancia: Procedimiento que utiliza resultados muestrales para verificar una H_0.
- Intervalo de confianza = **región de aceptación.**
- Área(s) fuera del int. De confianza = **región(es) de rechazo/ críticas.**

La prueba t



- Para nuestro ejemplo, asumiendo:
  - $H_0 \colon \beta_2 = \beta_2^* = 0.5$
  - $H_1 \colon \beta_2 \neq 0.5$

- Calculamos el valor de t

$$t = \frac{0.724 - 0.5}{0.07} = 3.2$$

- Un estadístico es estadísticamente significativo si cae en la región crítica.

Prueba *t* de significancia

Tipo de hipótesis	H_0: Hipótesis nula	H_1: Hipótesis alternativa	Regla de decisión
Dos colas	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t\rvert > t_{\alpha/2}, gl$
Cola derecha	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_\alpha, gl$
Cola izquierda	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_\alpha, gl$

Hipótesis nula "cero"

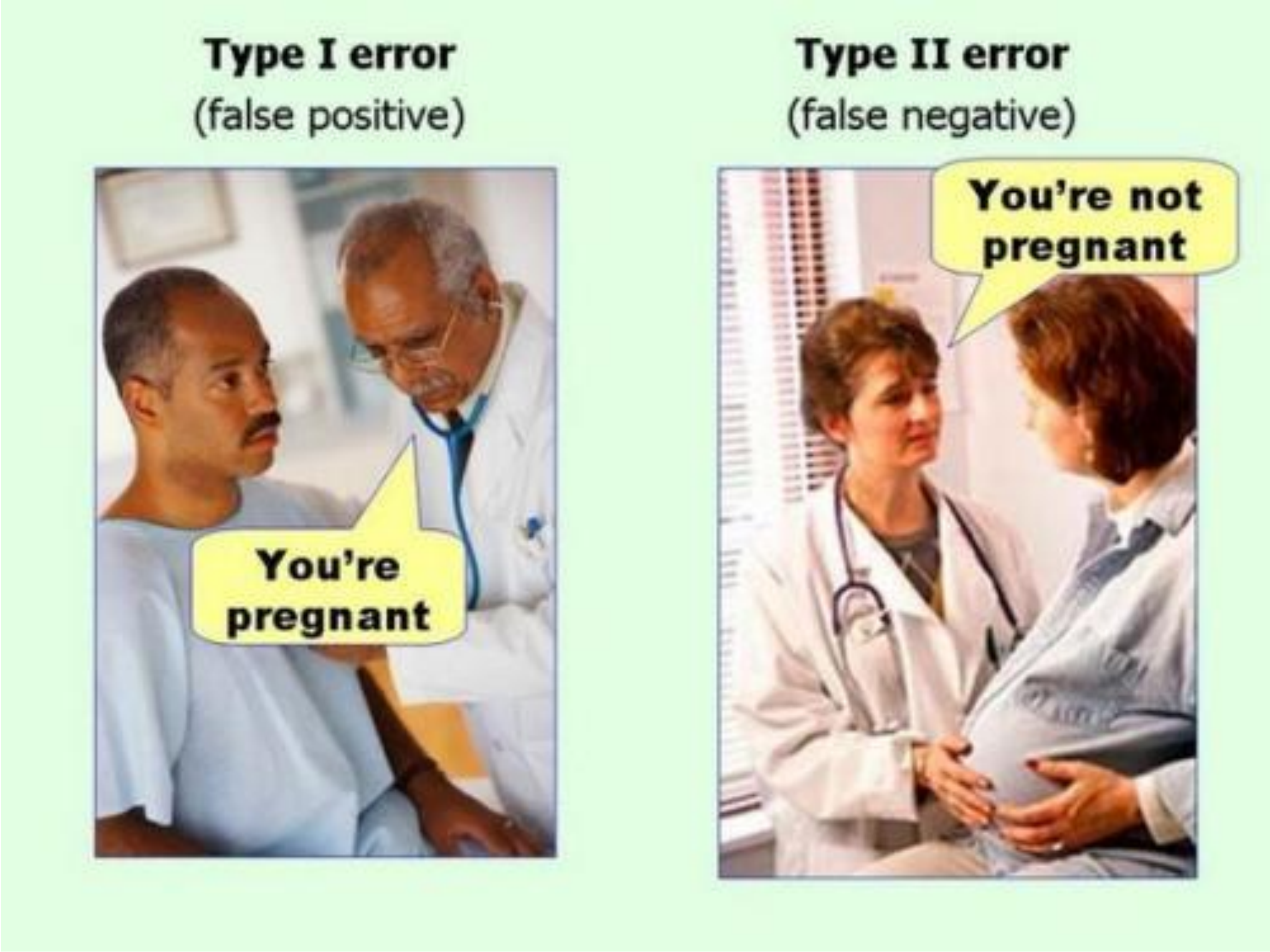
- Es muy común que comencemos con una hipótesis nula del estilo: $H_0: \beta_2 = 0$
- Es un mecanismo para determinar si Y tiene relación con X.
- Si no se rechaza esta H_0, entonces no tiene sentido plantearse ninguna otra hipótesis.

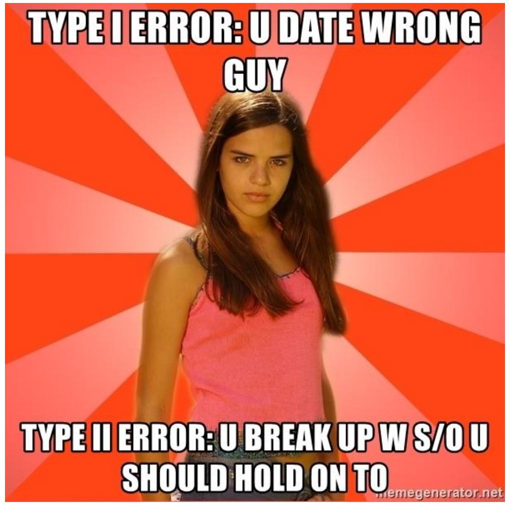
Selección del nivel de significancia, α

Rechazar o no una H_0 depende de α.

α= Probabilidad de cometer un error de **tipo I (P de rechazar H_0** cuando es verdadera).

Error **tipo II**: No rechazar H_0 cuando es falsa.





Error tipo I y tipo II

- Si se disminuye la probabilidad de cometer un error tipo I, se aumenta el tipo II y viceversa.

P-value

- El p-value expresa la probabilidad **exacta** de cometer un error de tipo I.

- Se define al p-value como *“el nivel de significancia más bajo al cual puede rechazarse una hipótesis nula”*.