INFERENCIA REGRESIÓN

ESTIMADORES DE LOS PARÁMETROS

- La base para inferencia referida a los parámetros de la regresión la proporcionan las propiedades de la distribución de muestreo de *a* y *b*, obtenidos mediante el método de mínimos cuadrados.

- Considerar qué ocurriría si para el mismo estudio de regresión se usaran muestras aleatorias diferentes. No puede esperarse que se obtenga exactamente la misma ecuación. Los estimadores *a* y *b*, obtenidos por el método de mínimos cuadrados, son estadísticos muestrales que tienen su propia distribución de muestreo.
- Se presentan las propiedades de la distribución de muestreo de cada uno:

• Distribución muestral de *b*:

Valor esperado: se obtiene de aplicar el operador esperanza a la expresión de b y trabajarlo matemáticamente:

$$E(b) = \beta$$

Desvío estándar: se obtiene de aplicar el operador varianza y trabajar matemáticamente

$$\sigma(b) = \frac{\sigma}{\sqrt{(x_i - \bar{x})^2}}$$

y con el desvío estimado para σ :

$$S(b) = \frac{S}{\sqrt{(x_i - \bar{x})^2}}$$

Distribución muestral Normal

- Así es posible encontrar intervalos en particular para la pendiente de la regresión:

$$\left(b \pm z_{1-\alpha} S(b)\right) \quad ó$$
$$\left(b \pm t_{1-\alpha} S(b)\right)$$

- # Distribución muestral de a:

Esperanza:

$$E(a) = E(\overline{Y} - b\overline{X}) = E(\overline{Y}) - E(b\overline{X}) = \alpha + \beta\overline{X} - \beta\overline{X} = \alpha$$

$$E(a) = \alpha$$

Desvío estándar: se obtiene de aplicar el operador varianza y trabajar matemáticamente:

$$\sigma^2(a) = \sigma^2 \left[\frac{1}{n} + \frac{X^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right]$$

Distribución muestral Normal

- Esto permite encontrar el intervalo de confianza para la ordenada del modelo de regresión lineal:

$$\left(a \pm z_{1-\alpha}\, S(a)\right) \quad \text{ó}$$

$$\left(a \pm t_{1-\alpha}\, S(a)\right)$$

- **Estos intervalos se obtienen en la salida de regresión en los softwares:**

- Por ejemplo:

	Coeficientes	Error típico	Inferior 95%	Superior 95%
Intercepción	-1032,201975	432,099881	**-1930,80287**	**-133,6010803**
Pendiente	437,7332939	89,50327239	**251,6010495**	**623,8655383**

- Las estimaciones puntuales no proporcionan información alguna acerca de la precisión de una estimación. Para eso es necesario obtener estimaciones por intervalo que son muy parecidas a las ya estudiadas.

- El primer tipo de estimación por intervalo, es el intervalo de confianza de una estimación del *valor medio de las* ***y*** que corresponden a un valor dado de x.

- El segundo tipo, el intervalo de predicción, se usa cuando se necesita una estimación por intervalo de un *solo valor de* ***y*** para un valor dado de x.

- La estimación puntual del valor medio de *y* es igual a la estimación puntual de un solo valor de *y*. Pero las estimaciones por intervalo que se obtienen para estos dos casos son diferentes. En un intervalo de predicción el margen de error es mayor ya que los errores para caso son diferentes.

PREDICCIÓN

- ***Predicción***: es la estimación del valor medio de Y dado un valor particular de X:

$$\hat{Y}_h = a + bX_h$$

- Se considera la recta de regresión y su precisión.

  a + bX es el estimador insesgado de **a + βX**, su distribución es normal, ya que es una combinación lineal de variables aleatorias normales.

  Entonces la variación depende de la variación o error en ambos estimadores, o sea en **a y b**:

- En general, no se puede esperar que $\hat{y}_p$ sea exactamente igual a $E(\hat{y}_p)$

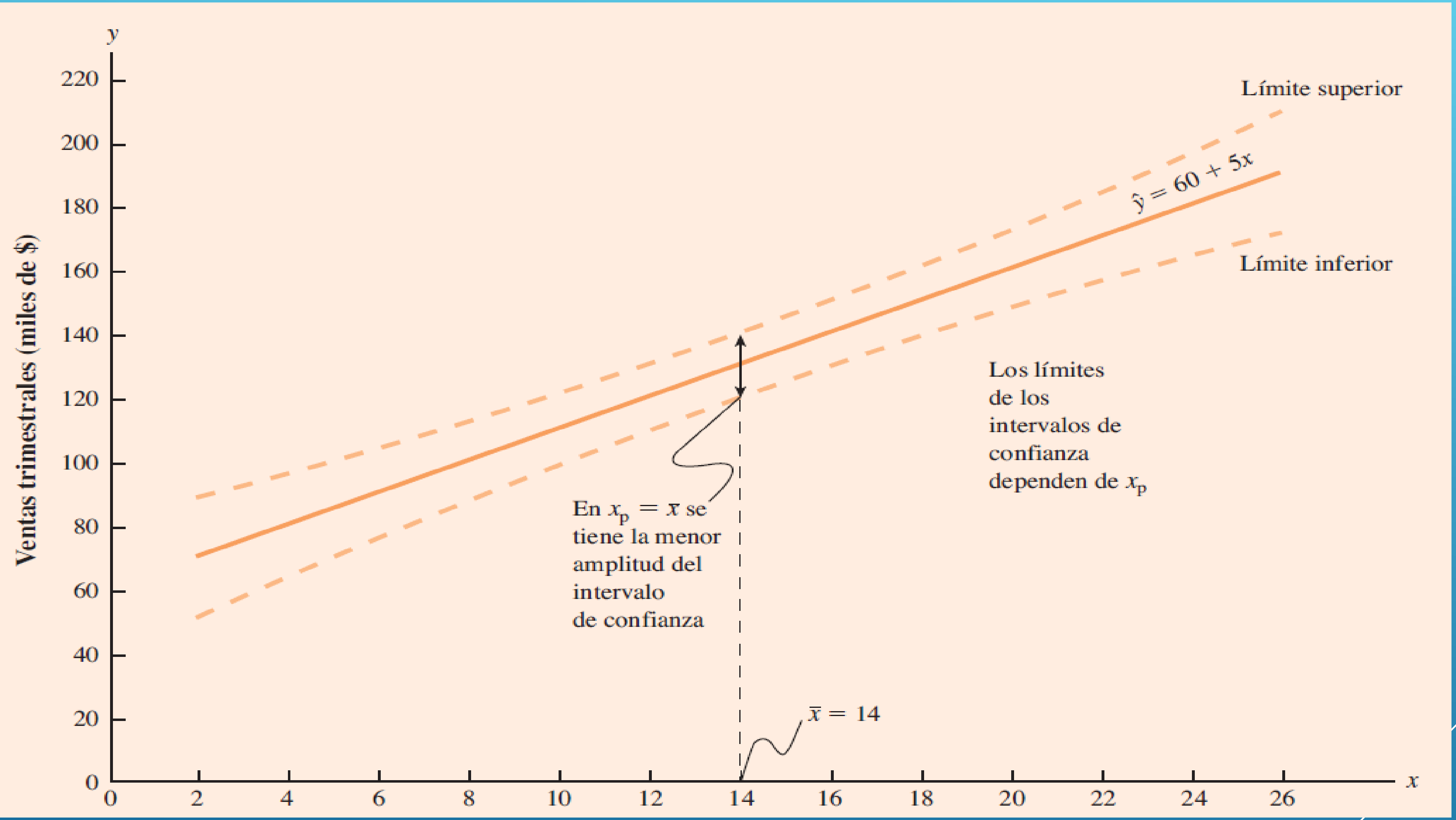
- Para hacer una inferencia acerca de qué tan cerca está $\hat{y}_p$ de la media verdadera $E(\hat{y}_p)$ es necesario estimar la varianza de $\hat{y}_p$.
- La fórmula para estimar la varianza de $\hat{y}_p$ para un xp dado, es $s^2(\hat{y}_p)$

$$s_{\hat{y}_p}^2 = s^2\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right]$$

INTERVALO DE CONFIANZA PARA EL VALOR MEDIO

$$\hat{y}_p \pm z_{1-\alpha} S_{\hat{y}_p} \qquad \text{ó}$$

$$\hat{y}_p \pm t_{1-\alpha;n-2} S_{\hat{y}_p}$$



y
Ventas trimestrales (miles de $)
Límite superior
$\hat{y} = 60 + 5x$
Límite inferior
Los límites de los intervalos de confianza dependen de x_p
En $x_p = \bar{x}$ se tiene la menor amplitud del intervalo de confianza
$\bar{x} = 14$
x

INTERVALO DE PREDICCIÓN PARA UN SOLO VALOR DE *Y* *(PRONÓSTICO)*

- Para obtener un intervalo de predicción, es necesario determinar primero la varianza correspondiente al uso de $\hat{y}_p$ como estimación de un valor individual de y cuando a x= xp. Esta varianza está formada por la suma de los dos componentes siguientes:
- **1.** La varianza de los valores individuales de y respecto de la media $E(\hat{y}_p)$ para la cual una estimación está dada por s^2
- **2.** La varianza correspondiente al uso de $\hat{y}_p$ para estimar $E(\hat{y}_p)$ para la cual una estimación está dada por $s^2(\hat{y}_p)$

- La expresión para estimar la varianza de un valor individual de yp, S_{ind} es:

$$S_{ind}^2 = S^2 + S_{\hat{y}_p}^2 =$$

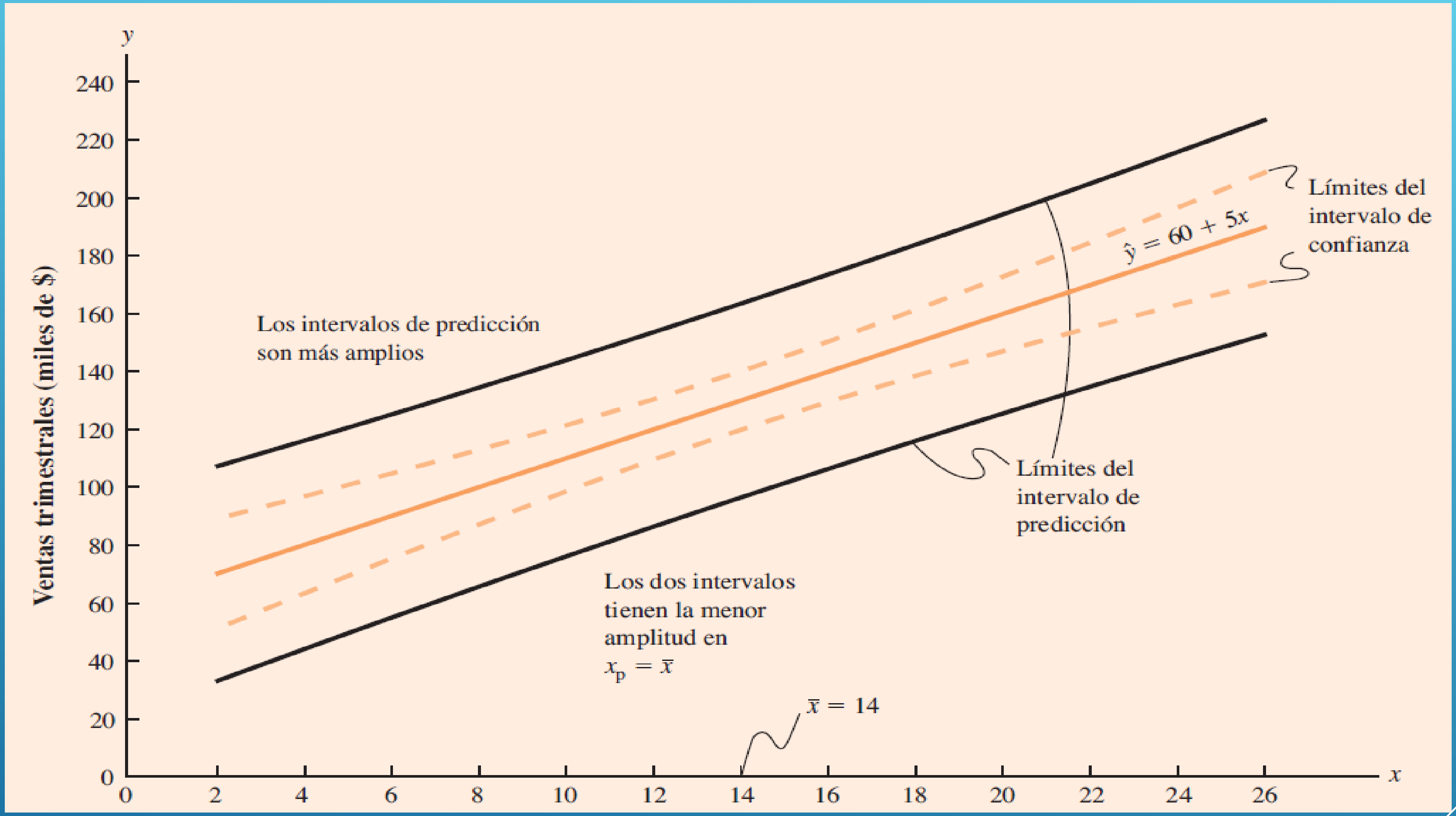
$$= S^2 + S^2\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right] =$$

$$= S^2\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]$$

- **Luego el intervalo es:**

$$\left(\hat{y}_p \pm z_{1-\alpha} S_{ind}\right) \quad ó$$

$$\left(\hat{y}_p \pm t_{1-\alpha;n-2} S_{ind}\right)$$



y
240
220
200
180
160
140
120
100
80
60
40
20
0
Ventas trimestrales (miles de $)
Los intervalos de predicción son más amplios
$\hat{y} = 60 + 5x$
Límites del intervalo de confianza
Límites del intervalo de predicción
Los dos intervalos tienen la menor amplitud en $x_p = \bar{x}$
$\bar{x} = 14$
0
2
4
6
8
10
12
14
16
18
20
22
24
26
x