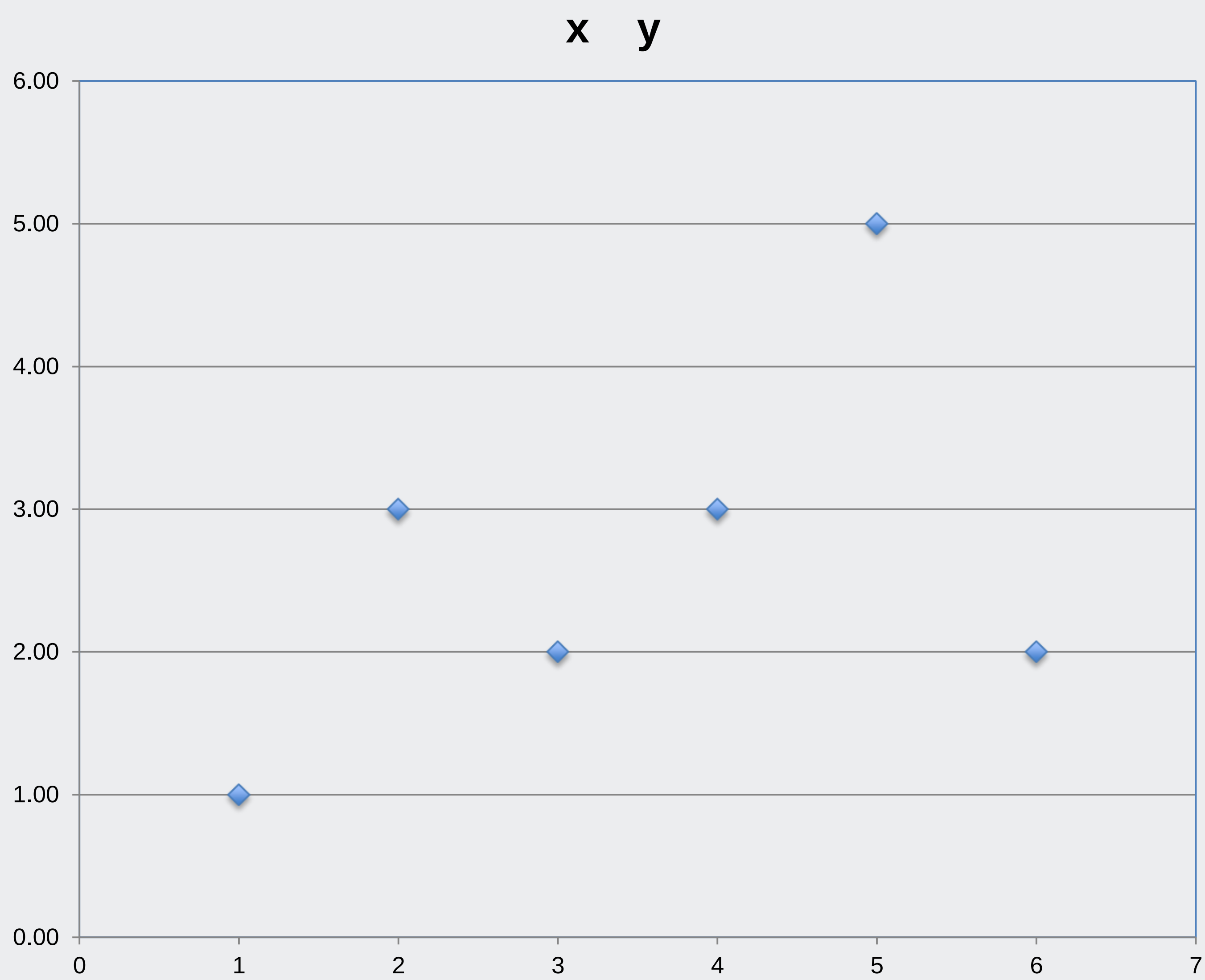
# TA 1- Ej1 Regresión lineal (1)

utilizando una planilla electrónica:

- supongamos un conjunto de datos con 1 sóla variable de entrada ***x***, y una variable de salida ***y***

  – x = {1, 3, 2, 4, 6, 5} ;
  – y = {1, 2, 3, 3, 2, 5}

- representar estos valores en 2 columnas de la planilla
- generar un gráfico de puntos para estos datos


x y
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
1
2
3
4
5
6
7

# TA 1- Ej1 (2)

Crear un modelo simple de regresión lineal
el modelo será

$$\mathbf{y} = B0 + B1 \times \mathbf{x}$$

- necesitamos estimar los coeficientes de esta ecuación (de una recta)
- el objetivo es encontrar las mejores estimaciones de los coeficientes tales que minimicen los errores al predecir ***y*** a partir de ***x***

# TA 1- Ej1 (3)

- los coeficientes se pueden calcular por las siguientes fórmulas

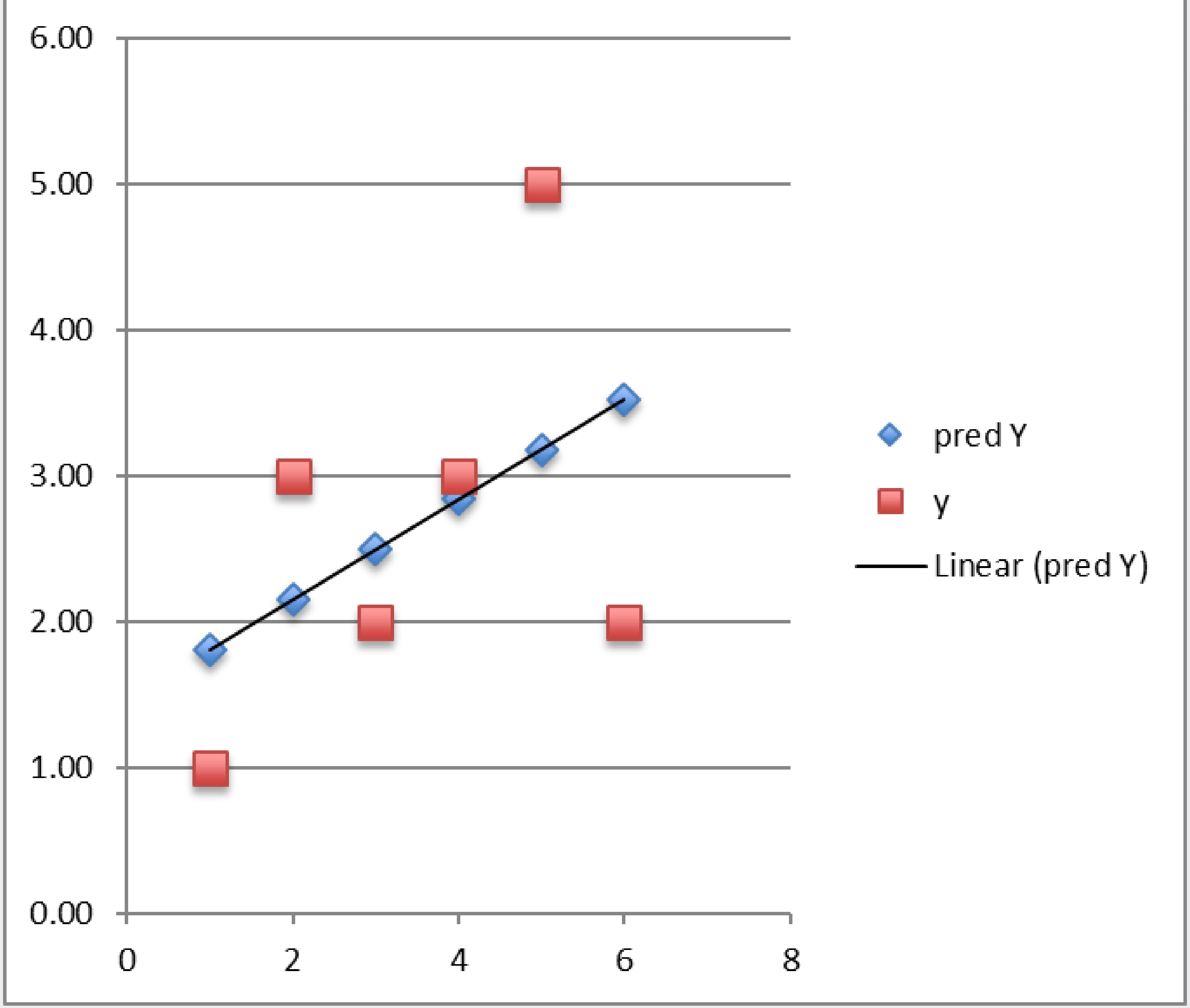
$$B1 = \frac{\sum_{i=1}^{n}(x_i - mean(x)) \times (y_i - mean(y))}{\sum_{i=1}^{n}(x_i - mean(x))^2}$$

$$B0 = mean(y) - B1 \times mean(x)$$

- calcular primero las medias de **x** e **y**
- agregar columnas para
  - **($x_i$ - media (x))** y **($y_i$-media(y))**
- B1? B0?

# TA 1- Ej1 (4)

- Predicciones:
  - la ecuación del modelo es
    - $\mathbf{y} = B0 + B1 \times \mathbf{x}$
- agregar una columna “predicción” y aplicar el modelo a los valores de entrenamiento
- generar dos columnas adicionales
  - **x** entre 0 y 8, con paso de 0.1
  - **y** resultante de aplicar la ecuación (con los coeficientes calculados) a los valores de la columna con las x generadas
  - agregar una serie en el gráfico para mostrar esta recta de ajuste




Universidad Católica del Uruguay

# TA 1- Ej1 (5)

- estimación del error de predicción

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - y_i)^2}{n}}$$

- agregar las columnas necesarias en la planilla y calcular el error medio cuadrático
- ¿cuánto da?

# TA 1- Ej1 (6)

- método rápido alternativo para calcular B1 (y por tanto B0)

$$B1 = corr(x,y) \times \frac{stdev(y)}{stdev(x)}$$

- en la planilla, la correlación se puede calcular utilizando la función “PEARSON”,
  - ¿cuánto es esta correlación?

- ¿B1? ¿confirma?

# TA 1- Ej2 Regresión lineal con descenso de gradiente (1)

- minimizar una función siguiendo los gradientes de la función de costo
- necesitamos conocer la función de costo y sus derivadas
- moverse “hacia abajo” según el gradiente
- las instancias se presentan al modelo una a una
- el modelo hace una predicción, se calcula el error y se actualiza el modelo para reducir el error de la próxima predicción
- en cada iteración se actualizan los “pesos” (coeficientes)

$$w = w - alpha \times delta$$

- alpha es la “tasa de aprendizaje” y delta el error

# TA 1- Ej2 (2)

- el modelo simple de regresión lineal es

$$\mathbf{y} = B0 + B1 \times \mathbf{x}$$

- DG, primera iteración
  - inicializamos B0 y B1 con 0.0
  - $\mathbf{y} = 0.0 + 0.0 \times \mathbf{x}$
  - error = p(i) - y(i)
  - p(i) = 0.0 + 0.0 × 1 = 0.0 , error = 0.0 - 1 = -1
  - B0 (t+1) = B0(t) - alpha × error **= 0.0 - 0.01** × **-1 = 0.01**
  - B1 (t+1) = B1(t) - alpha × error × **x =**

B1 (t+1) **=** **0.0 - 0.01** × **-1** × 1 **= 0.01**

- **repetir 18 - 24 iteraciones**
  - **¿cuántas serían apropiadas?**

# TA 1- Ej2 (3)

- listar los valores de B0 y B1 de todas las iteraciones
- graficar el error de predicción vs. iteraciones
- calcular error medio cuadrático