Análisis de la asociación espacial

Regresión

Gerardo Martín

2022-06-29

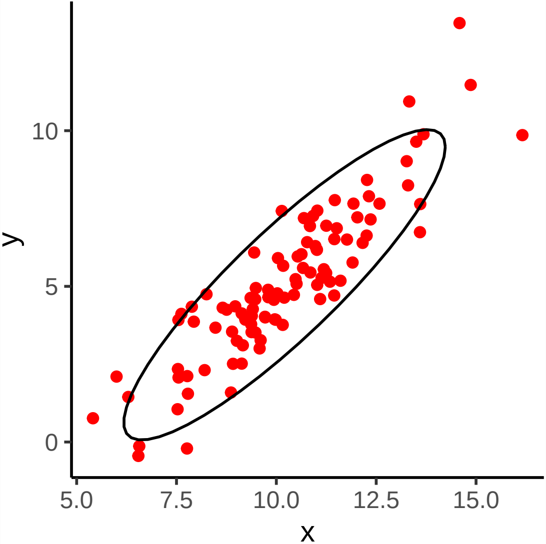
Diferencias y similitudes con correlación

- Correlación mide dependencia lineal entre dos fenómenos
- No hay causa - efecto:
  - $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
- No permite estimar cambio de una con respecto de la otra
- Regresión, sí se asume causa y efecto:
  - $y(x) = a + bx$

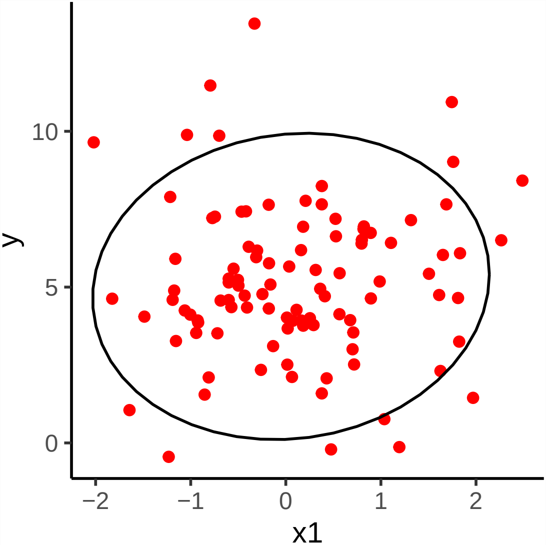
Diferencias y similitudes con correlación

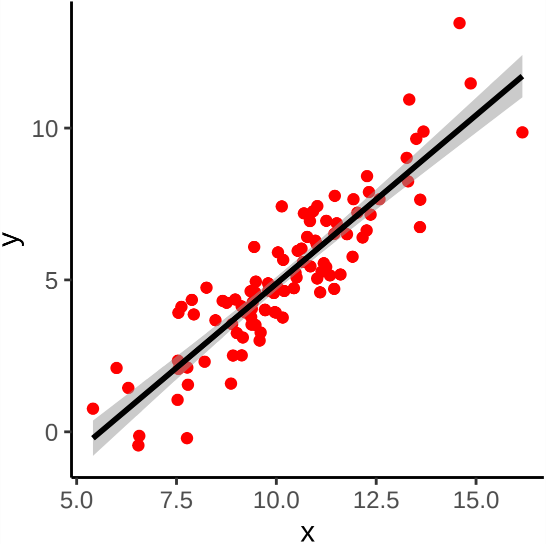
- Correlación, sólo conocemos r, ó ρ (correlación de Spearman)
  - Signo de r indica si una disminuye ó aumenta en relación a la otra
  - P indica probabilidad de que $r = 0$
- Regresión
  - Coeficientes a y b cuantifican relación entre x y y
  - a, valor de y cuando $x = 0$
  - b, aumento ó disminución de y cuando x aumenta en 1 unidad
  - R^2, cuadrado del coeficiete de correlación r

Diferencias gráficas



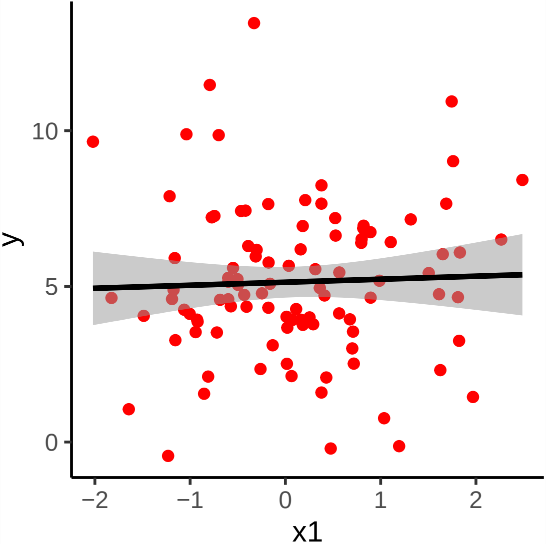
Diferencias gráficas



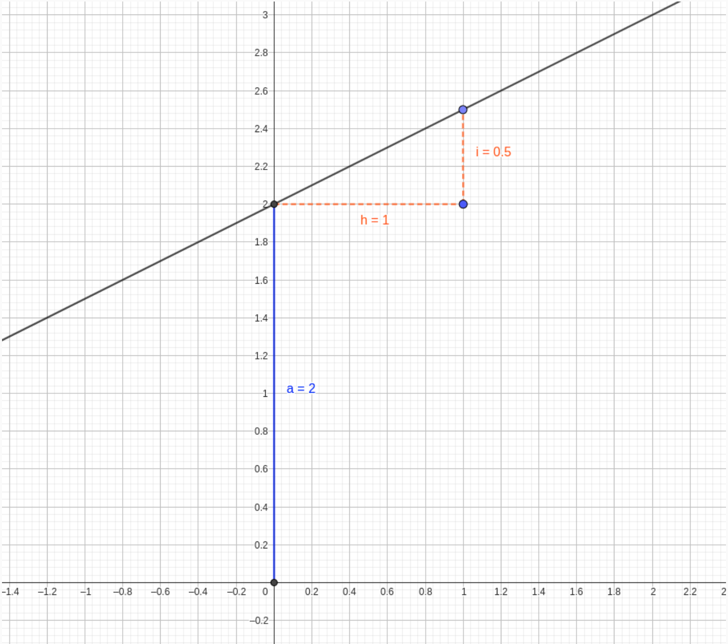


y
10
5
0
5.0
7.5
10.0
12.5
15.0
x

Diferencias gráficas



Significato de parámetros



Significado de parámetros

$$y(x) = a + bx$$

- $a = 2$
- $b = i/h = 0.5$

$$y(x) = 2 + 0.5x$$

Implementando en R

- Función nativa para estimar a, b y R^2, `lm`
- Hipótesis que queremos rechazar, x no afecta a y
  - $a \neq 0, b \neq 0$
- Uso:

```
lm(x ~ y, data = df.1) #df.1 es tabla que contiene datos de x
```

```r
reg.lin <- lm(y ~ x, data = df.1)
summary(reg.lin)
```

```
##
## Call:
## lm(formula = y ~ x, data = df.1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6090 -0.8366 -0.0884  0.6772  3.4914
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.20626    0.58606  -10.59   <2e-16 ***
## x            1.10859    0.05621   19.72   <2e-16 ***
## ---
```

Columnas

- `Estimate` contiene el valor estimado
- `Pr(>|t|)` es la probabilidad de que $a = 0$ ó $b = 0$

Filas

- $a = (\texttt{Intercept})$
- $b = \texttt{x}$
- $R^2 = \texttt{Adjusted R} - \texttt{squared}$

Implementando en R

```r
reg.lin1 <- lm(y ~ x1, data = df.1)
summary(reg.lin1)
```

```
##
## Call:
## lm(formula = y ~ x1, data = df.1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4589 -1.2876 -0.2732  1.3428  8.3563
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.13185    0.24895  20.614   <2e-16 ***
## x1           0.09689    0.25581   0.379    0.706
## ---
```

Al igual que en correlación, debemos:

1. Extraer
2. Añadir a base de coordenadas
3. Ajustar
4. Interpretar
5. Predecir

Retomaremos el ejemplo de la correlación espacial:

Table 1: Primeras seis filas de una base de datos de mediciones colectadas en campo.

x	y	mediciones
-102.7928	29.57881	49.62024
-103.6011	27.38053	41.14992
-104.5670	25.53772	137.04156
-101.8276	29.87109	51.70786
-100.5730	27.40428	103.70981
-101.0474	29.20245	75.43274

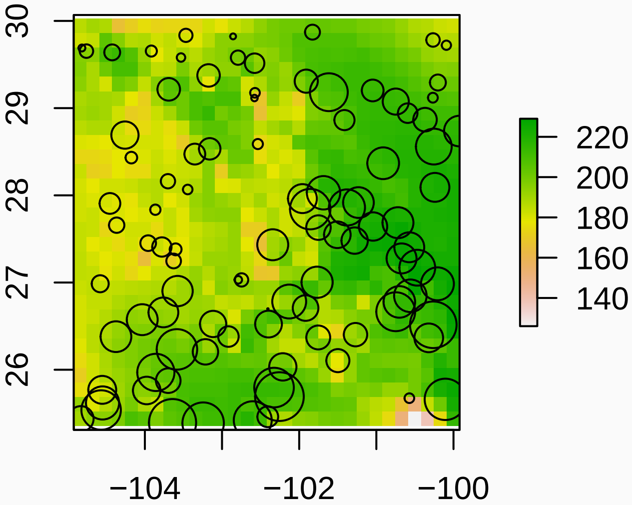


Figure 1: Diámetro de burbujas indica magnitud

Extrayendo valores de capas raster

```r
valores.capas <- extract(r, puntos[, c("x", "y")])
puntos <- data.frame(puntos, valores.capas)
```

Valores extraídos de capa raster

x	y	mediciones	Var.1	Var.2	Var.3
-102.7928	29.57881	49.62024	199	106	161
-103.6011	27.38053	41.14992	179	105	143
-104.5670	25.53772	137.04156	187	126	193
-101.8276	29.87109	51.70786	203	105	149
-100.5730	27.40428	103.70981	224	135	143
-101.0474	29.20245	75.43274	213	121	131
-102.5358	28.58792	35.05952	174	101	132
-101.4132	28.86275	69.94669	209	125	134
-102.3974	26.52351	92.91473	211	134	153
-102.6020	25.42059	131.78474	218	147	148

Ajustando el modelo lineal

Con base en análisis de correlación sabemos que es más probable que `Var.2` explique las mediciones:

```
m1 <- lm(mediciones ~ Var.2, puntos)
```

Ajustando el modelo lineal

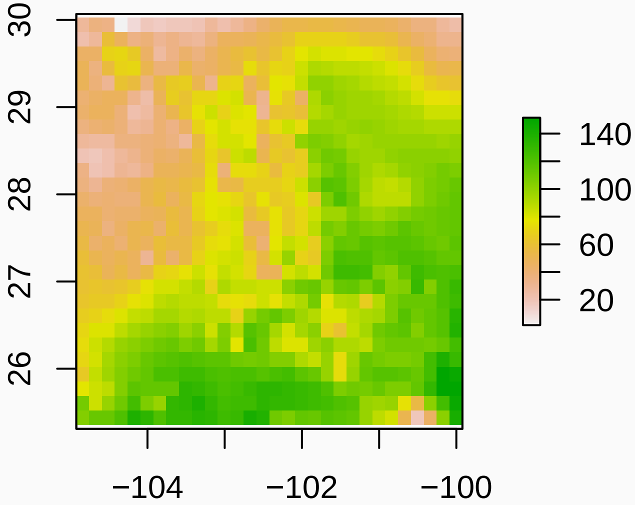
```r
summary(m1)
```

```
##
## Call:
## lm(formula = mediciones ~ Var.2, data = puntos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -70.751 -13.179  -0.278  13.347  56.566
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -160.5896    18.0366  -8.904 3.12e-14 ***
## Var.2          2.0531     0.1505  13.641  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Necesitamos crear base con valores para predecir:

```
val.nuevos <- data.frame(rasterToPoints(r))
preds <- predict(m1, val.nuevos)
preds.r <- rasterFromXYZ(data.frame(val.nuevos[, c("x", "y")]
```

Gráfico de la predicción



Comparación con otros modelos

```r
m2 <- lm(mediciones ~ Var.1, puntos)
m3 <- lm(mediciones ~ Var.3, puntos)
```

Comparación con otros modelos

```
##
## Call:
## lm(formula = mediciones ~ Var.1, data = puntos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.153 -17.571  -3.082  18.632  70.786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -193.0459    39.1002  -4.937 3.30e-06 ***
## Var.1          1.3950     0.1966   7.096 2.12e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##
## Call:
## lm(formula = mediciones ~ Var.3, data = puntos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -108.501  -31.980    2.111   24.895   85.851
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.1195    35.5229   1.664   0.0993 .
## Var.3         0.1614     0.2335   0.691   0.4913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Comparación con criterio de Akaike

```r
AIC(m1)
```

```
## [1] 900.5037
```

```r
AIC(m2)
```

```
## [1] 965.1384
```

```r
AIC(m3)
```

```
## [1] 1006.045
```

Modelo con AIC menor, `m1`, por lo tanto, parece más adecuado que `m2` y `m3`

Conclusión

- Regresión lineal sirve para predecir
- Correlación sirve para medir asociación