

Solución de la Práctica dirigida 1 de Análisis de Regresión

Estrella Guerra, Danilo David

2025-01-30

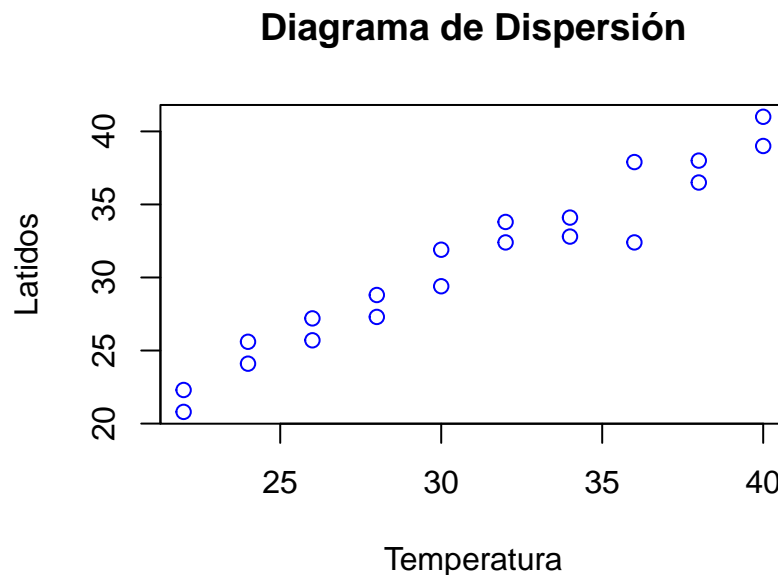
Caso: Zoológico de Australia Se hizo un experimento en un zoológico de Australia para analizar distintas temperaturas en °C (X) de un recinto cerrado y los correspondientes ritmos cardiacos en latidos/minuto (Y) de una especie de lagarto:

Análisis de Correlación

a) Elabore el gráfico de dispersión

Para elaborar un grafico de dispersión vamos a usar la libreria plot (la función plot es una función genérica para la representación gráfica de objetos en R)

```
plot(PD1$X,PD1$Y,main="Diagrama de Dispersión",  
     col="blue",xlab="Temperatura",ylab="Latidos")
```



Se puede apreciar una relación directa entre la temperatura °C del recinto y la cantidad de latidos/minuto de los lagartos.

b) Estime el coeficiente de correlación

La función `cor()` en R se utiliza para calcular el coeficiente de correlación entre dos variables. Este coeficiente mide la fuerza y la dirección de la relación lineal entre las dos variables.

```
cor(PD1$X,PD1$Y,method="p")
```

```
## [1] 0.9721215
```

Hay una fuerte asociación lineal directa entre la temperatura del recinto y la cantidad de latidos/minuto del lagarto.

c) Pruebe la significancia del coeficiente de correlación

1)

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2)

$$\alpha = 0.02$$

3)

```
cor.test(PD1$X,PD1$Y,method="p",alternative = "t")
```

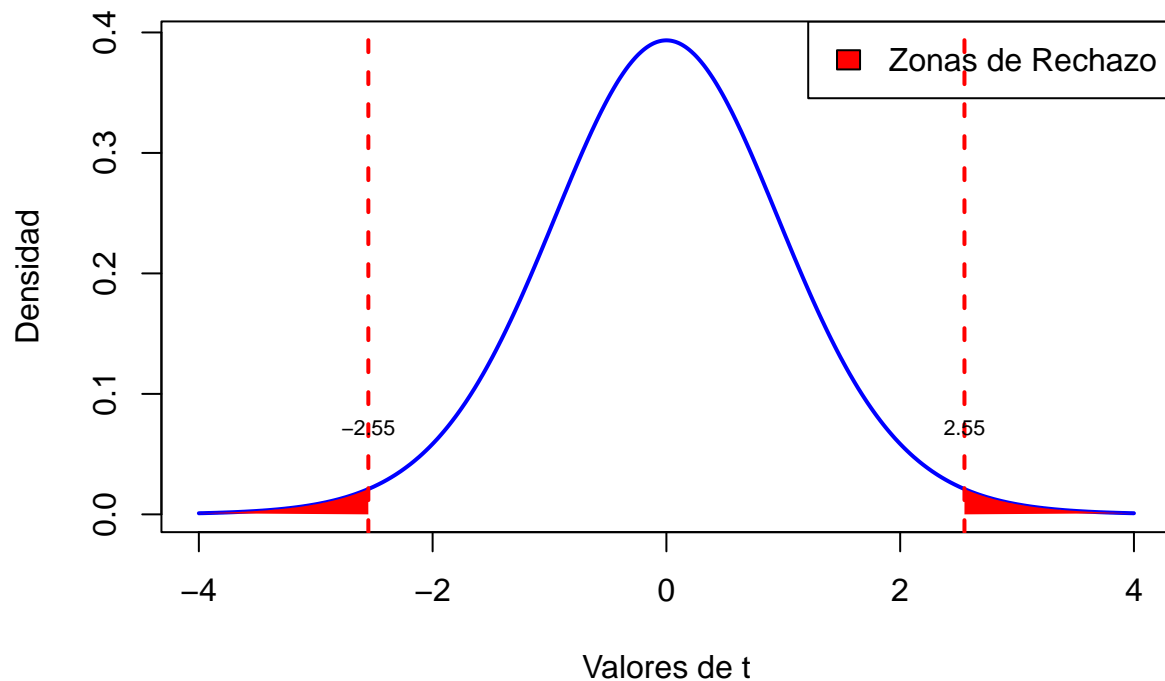
```
##
## Pearson's product-moment correlation
##
## data: PD1$X and PD1$Y
## t = 17.59, df = 18, p-value = 8.732e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9294241 0.9891331
## sample estimates:
## cor
## 0.9721215
```

del anterior código hemos obtenido el valor de:

$t = 17.59$, $p - value = 8.732e - 13 \approx 0.000 < \alpha$ se rechaza H_0

4)

Distribución t con Zonas de Rechazo



```
qt(0.02/2,18)
```

```
## [1] -2.55238
```

```
qt(c(0.02/2,1-0.02/2),18)
```

```
## [1] -2.55238 2.55238
```

5) Conclusión

A un $\alpha = 0.02$, se puede afirmar que existe correlación significativa entre la temperatura del recinto y el número de latidos/minuto de los lagartos.

d) Obtenga un intervalo del 98% de confianza para el coeficiente de correlación

```
cor.test(PD1$X,PD1$Y,method="p",alternative = "t",
         conf.level = 0.98)$conf
```

```
## [1] 0.9162722 0.9908944
```

```
## attr(,"conf.level")
```

```
## [1] 0.98
```

$IC(\rho) = [0.9162722, 0.9908944]$

El intervalo que va de 0.916 a 0.991 brinda un 98% de confianza de contener a la correlación entre la temperatura del recinto y el número de latidos/minuto de los lagartos.

Análisis de Regresión Lineal Simple

e) Obtenga la ecuación de regresión estimada. Interprete los coeficientes estimados.

```
modelo<-lm(Y~X,data=PD1)
print(modelo)

##
## Call:
## lm(formula = Y ~ X, data = PD1)
##
## Coefficients:
## (Intercept)          X
##      1.4591      0.9545
```

$$\hat{Y} = 1.4591 + 0.9545x$$

El código $lm(Y \sim X, data = PD1)$ nos permite conocer los coeficientes de nuestro modelo.

- $b_o = 1.4591$ No tiene interpretación
- $b_1 = 0.9545$ Al aumentarse la temperatura en un grado °C, se espera que el número medio de latidos se incremente en 0.9545 latidos/minuto.

f y h

- f) Realice el Análisis de Varianza. Use $\alpha = 0.02$
g) Pruebe la significancia individual de β_1 . Use $\alpha = 0.02$

1a)

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2a)

$$\alpha = 0.02$$

3a) Prueba estadística

$$F = \frac{CMReg}{CME} \sim F_{(1,n-2)}$$

4a) Desarrollo de la prueba

```
anova(modelo)

## Analysis of Variance Table
##
## Response: Y
##      Df Sum Sq Mean Sq F value    Pr(>F)
## X      1  601.36   601.36   309.39 8.732e-13 ***
## Residuals 18   34.99     1.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F= 309.39, p-valor = 0.0000000000008732

5a) Región Crítica 1b)

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2b)

$$\alpha = 0.02$$

3b) Prueba estadística

$$t = \frac{\hat{\beta}_1}{EE(\hat{\beta}_1)} \sim t_{(n-2)}$$

4b) Desarrollo de la prueba

```
summary(modelo)
```

```
##
## Call:
## lm(formula = Y ~ X, data = PD1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4227 -0.7432  0.0136  1.0000  2.0773
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.45909    1.71094   0.853   0.405
## X            0.95455    0.05427  17.590 8.73e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.394 on 18 degrees of freedom
## Multiple R-squared:  0.945, Adjusted R-squared:  0.942
## F-statistic: 309.4 on 1 and 18 DF, p-value: 8.732e-13
```

$t_{cal} = 17.590$, p-valor = $8.73e-13$

5b) Región Crítica

6ab)

A un $\alpha = 0.02$, se puede afirmar que el número de latidos/minuto es explicado por la temperatura en el recinto.

g) Halle intervalos del 98% de confianza para β_0 y β_1

```
confint(modelo, level = 0.98)
```

```
##              1 %      99 %  
## (Intercept) -2.9078734 5.826055  
## X           0.8160338 1.093057
```

$$IC(\beta_1) = [0.8160338, 1.093057]$$

El intervalo va de 0.816 a 1.093 brinda un 98% de confianza de contener a β_1

i) Halle un intervalo del 98% de confianza para σ_ε^2

$$IC(\sigma_\varepsilon^2) = \left[\frac{(n-2)CMRes}{x_{(1-\alpha/2, n-2)}^2}, \frac{(n-2)CMRes}{x_{(\alpha/2, n-2)}^2} \right]$$

```
sig<-summary(modelo)$sigma  
n<-nrow(PD1)  
LI<-(n-2)*sig^2/qchisq(1-0.02/2, n-2)  
LS<-(n-2)*sig^2/qchisq(0.02/2, n-2)  
c(LI, LS)
```

```
## [1] 1.005202 4.987428
```

El intervalo que va de 1.005 a 4.987 brinda un 98% de confianza de contener a σ_ε^2

j) Halle e interprete el coeficiente de determinación

```
summary(modelo)$r.sq*100
```

```
## [1] 94.50203
```

$$r^2 = 94.5$$

El 94.5% de la variabilidad del número de latidos/minuto es explicado por la temperatura en el recinto.

k) Halle un intervalo del 98% de confianza para el número medio de latidos por minuto a una temperatura de 33 °C

```
predict(modelo, data.frame(X = 33), level = 0.98, interval = "confidence")
```

```
##      fit      lwr      upr  
## 1 32.95909 32.11656 33.80162
```

$$IC(U_{y/x}) = [32.116, 33.801]$$

El intervalo que va de 32.116 a 33.801 brinda un 98% de confianza de contener al número medio de latidos/minuto cuando el recinto tiene una temperatura de 33°C

l) Halle un intervalo del 98% de confianza para el número de latidos por minuto, de un recinto que tiene 39 °C

```
predict(modelo, data.frame(X = 33), level = 0.98, interval = "prediction")
```

```
##          fit      lwr      upr  
## 1 32.95909 29.30228 36.6159
```

$IC(Y) = [34.874, 42.497]$

```
lista = [1, 2, 3, 4, 5, 6]  
print(lista[2])
```

```
## 3
```