

Lab06 - PRONÓSTICO PARA VALOR MEDIO E INDIVIDUAL

Escriba Flores, Daniel Agustin

2025-04-24

Contents

Contexto:	1
Pregunta A: Analisis de regresion y diagrama de dispersion	1
Regresion Lineal	2
Grafico de dispersion	3
Pregunta B: Intervalo de confianza de la media, intervalo de prediccion.	3
Intervalo de confianza para la media.	3
Intervalo de prediccion	4
Interpretacion	4
Pregunta C: Intercarlo para valores de “y” y grafico	4
Intervalos para todos los y	4
Cuadro final de los datos originales con sus respectivos intervalos de confianza	5
Grafico de intervalos	5

Contexto:

La gerencia de una empresa embotelladora de bebidas gaseosas desea desarrollar un método para distribuir los costos de entrega a los clientes. Aunque un costo se relaciona claramente con el tiempo de viaje en una ruta específica, otra variable de costo refleja el tiempo requerido para descargar las cajas de bebida refrescante en el punto de entrega. El tiempo de entrega y el número de cajas entregadas están registradas en la base de datos ‘delivery.xlsx’.

Deseando predecir el tiempo de entrega (en min) con base al número de cajas entregadas, haga lo siguiente:

Pregunta A: Analisis de regresion y diagrama de dispersion

- Lea la base de datos y haga un análisis de regresión que muestre el diagrama de dispersión junto con su recta de regresión.

```
library(readxl)
library(pander)

data = read_excel("delivery.xlsx", sheet="Hoja1")
data = data[-1]
data = as.data.frame(data)

# Usar pander para formatear la tabla
pander(data, caption = "Tabla de datos de delivery")
```

Table 1: Tabla de datos de delivery

Cajas	Tiempo
52	32.1
64	34.8
73	36.2
85	37.8
95	37.8
103	39.7
116	38.5
121	41.9
143	44.2
157	47.1
161	43
184	49.4
202	57.2
218	56.8
243	60.6
254	61.2
267	58.2
275	63.1
287	65.6
298	67.3

Regresion Lineal

```
x = data$Cajas
y = data$Tiempo

modelo = lm(y ~ x)
summary(modelo)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3788 -0.4173  0.3620  1.0186  4.0802
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.834531   1.054219   23.56 5.61e-15 ***
## x           0.140026   0.005627    24.88 2.15e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.987 on 18 degrees of freedom
## Multiple R-squared:  0.9718, Adjusted R-squared:  0.9702
## F-statistic: 619.2 on 1 and 18 DF,  p-value: 2.152e-15

modelo$coefficients

## (Intercept)          x
```

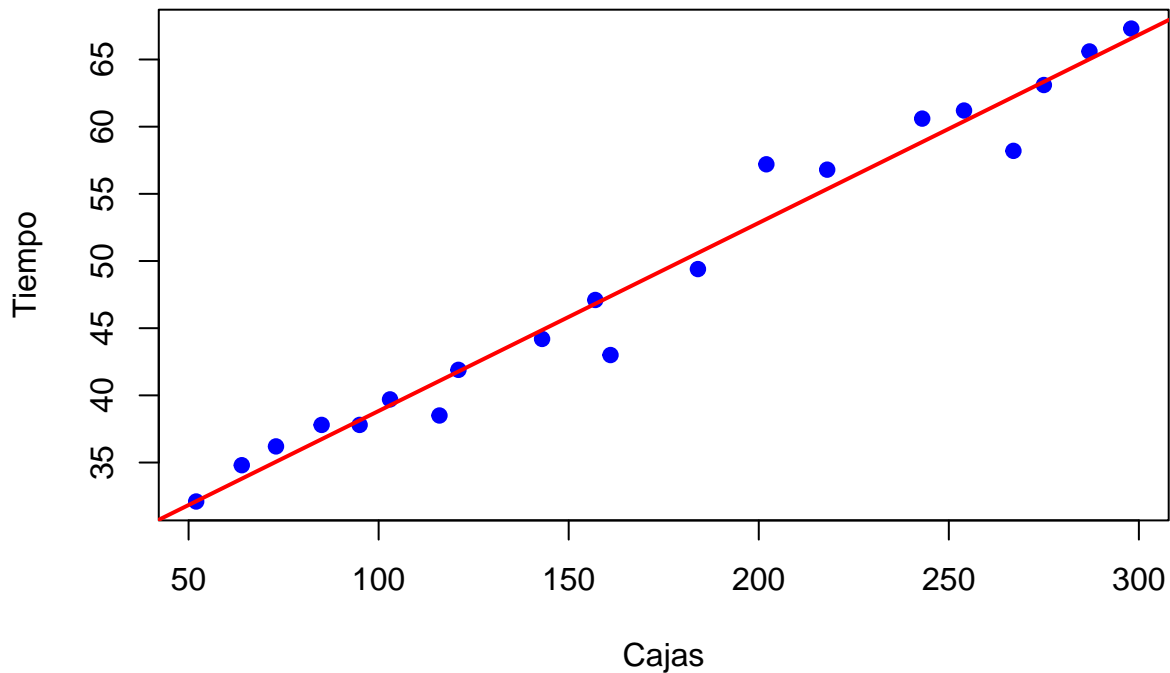
```
## 24.8345309 0.1400263
```

Grafico de dispersion

```
# Gráfico de dispersión
plot(x, y, main = "Diagrama de Dispersion",
     xlab = "Cajas", ylab = "Tiempo", pch = 19, col = "blue")

# Agregar línea de regresión
abline(modelo, col = "red", lwd = 2)
```

Diagrama de Dispersion



Pregunta B: Intervalo de confianza de la media, intervalo de prediccion.

- b. Encuentre un intervalo de confianza para la media y un intervalo de predicción para el tiempo de entrega cuando el número de cajas entregadas es 150 asumiendo un nivel de confianza de 98%. Interprete ambos resultados.

Intervalo de confianza para la media.

```
# Nivel de confianza
nc = 0.98

nuevo = data.frame(x=150) # cajas entregadas = 150

predict(object=modelo, newdata=nuevo, int="confidence",
        level=nc)
```

```
##          fit          lwr          upr
## 1 45.83848 44.66925 47.00771
```

Intervalo de prediccion

```
nuevo = data.frame(x=150) # cajas entregadas= 150

predict(object=modelo, newdata=nuevo, interval="prediction",
        level=nc)
```

```
##          fit          lwr          upr
## 1 45.83848 40.6351 51.04185
```

Interpretacion

- Con un nivel de confianza del 98%, el intervalo de confianza para la media del tiempo de entrega cuando el número de cajas entregadas es 150 se estima entre 44,67 y 47,01 minutos, lo que indica que estamos seguros de que la media poblacional del tiempo de entrega para pedidos de esta cantidad de cajas se encuentra dentro de este rango.
- Asimismo, el intervalo de predicción para un nuevo pedido individual con 150 cajas se proyecta entre 40,64 y 51,04 minutos, reflejando que, con el mismo nivel de confianza, esperamos que el tiempo de entrega de un único pedido future con 150 cajas se encuentre en este intervalo.
- El intervalo de predicción es más amplio debido a la mayor incertidumbre inherente a la predicción de valores individuales en contraposición a la estimación de la media poblacional.

Pregunta C: Intercarlo para valores de “y” y grafico

- c. Halle los intervalos de predicción para todos los valores de y asumiendo un nivel de confianza de 98% y grafique los intervalos para la media y los valores de y , que contenga además el diagrama de dispersión y la recta de regresión.

Intervalos para todos los y

```
y_futuro = predict(object=modelo, interval="prediction",
                    level=0.98)

dim(y_futuro)
```

```
## [1] 20  3
```

```
y_futuro
```

```
##          fit          lwr          upr
## 1 32.11590 26.65138 37.58042
## 2 33.79621 28.38262 39.20981
## 3 35.05645 29.67775 40.43515
## 4 36.73677 31.40006 42.07347
## 5 38.13703 32.83130 43.44276
## 6 39.25724 33.97361 44.54087
## 7 41.07758 35.82470 46.33046
## 8 41.77771 36.53493 47.02049
## 9 44.85829 39.64843 50.06816
## 10 46.81866 41.61984 52.01748
```

```
## 11 47.37877 42.18167 52.57586
## 12 50.59937 45.39990 55.79884
## 13 53.11984 47.90391 58.33578
## 14 55.36027 50.11901 60.60152
## 15 58.86092 53.56038 64.16147
## 16 60.40121 55.06712 65.73530
## 17 62.22155 56.84211 67.60100
## 18 63.34176 57.93139 68.75214
## 19 65.02208 59.56111 70.48305
## 20 66.56237 61.05069 72.07405
```

Cuadro final de los datos originales con sus respectivos intervalos de confianza

```
data_nueva = cbind(data, y_futuro) # cbind
pander(data_nueva, caption = "Tabla de datos de delivery con intervalos de confianza")
```

Table 2: Tabla de datos de delivery con intervalos de confianza

Cajas	Tiempo	fit	lwr	upr
52	32.1	32.12	26.65	37.58
64	34.8	33.8	28.38	39.21
73	36.2	35.06	29.68	40.44
85	37.8	36.74	31.4	42.07
95	37.8	38.14	32.83	43.44
103	39.7	39.26	33.97	44.54
116	38.5	41.08	35.82	46.33
121	41.9	41.78	36.53	47.02
143	44.2	44.86	39.65	50.07
157	47.1	46.82	41.62	52.02
161	43	47.38	42.18	52.58
184	49.4	50.6	45.4	55.8
202	57.2	53.12	47.9	58.34
218	56.8	55.36	50.12	60.6
243	60.6	58.86	53.56	64.16
254	61.2	60.4	55.07	65.74
267	58.2	62.22	56.84	67.6
275	63.1	63.34	57.93	68.75
287	65.6	65.02	59.56	70.48
298	67.3	66.56	61.05	72.07

Grafico de intervalos

```
library(ggplot2)
ggplot(data = data_nueva, aes (x=Cajas, y=Tiempo)) +
  geom_point (col="red") +
  geom_smooth(method=lm, formula= y~x,
              se=T, level=nc,
              col="blue", fill="orange") +
  geom_line(aes(y=lwr), color="purple", linetype="dashed") +
  geom_line(aes(y=upr), color="purple", linetype="dashed") +
  theme_light()
```

