$$\boxed{Y = \boxed{\beta_0 + \beta_1 X} + \boxed{\varepsilon_i}}$$

parte ②

parte ③

parte ①

Parte ①

↳ hacemos inferencia

→ P.H

→ I.C

→ estimación puntual

→ interpretaciones

β_0

β_1

coeficientes.

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0.$

P.H → β_1

P- significancia.

Analisis varianza (ANOVA)

verificación del modelo. (R^2)

Parte ③ → $\boxed{\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X}$

inferencia.

↳ P.H.

→ I.C

↳ estimación puntual

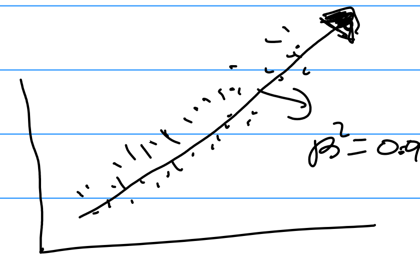
→ I. Pre

→ I. valor medio.

→ predición de dato.

$R^2 \rightarrow$ Proporción de variabilidad explicada por el modelo

$$R^2 = \frac{SSR}{\boxed{SST}}$$



Parte teorica

1 Respoder falso o verdadero y argumentar en caso de ser verdadero o dar un contra ejemplo en caso de ser falso

1.1 Un R^2 alto indica que el modelo puede hacer predicciones útiles. F

1.2 Un R^2 alto indica que la recta de regresión tiene buen ajuste. F

ojo

1.3 Un R^2 cercano a cero indica que X y Y no están relacionados. F puede presentar relaciones cuadraticas y un $R^2 \approx 0$

1.4 La formula del intervalo de predicción es:

1.4 La formula del intervalo de predicción es:

a) $\hat{y}_0 \pm t_{\alpha/2,n-2} \times \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$

b) $\hat{y}_0 \pm t_{\alpha/2,n-5} \times \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$

para la respuesta media

ojo

c) $\hat{y}_0 \pm t_{\alpha/2,n-2} \times \sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$

para un I. C de predicción.

d) $\hat{y}_0 \pm \cancel{t_{\alpha,n-3}} \times \sqrt{\sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$

$$\sqrt{\sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

1.5 Estimación puntual y por intervalo de la respuesta media $E[Y|x_0] + \varepsilon_0$ F

1.6 Predicción de valores futuros $y_0 = \beta_0 + \beta_1 x_0 = E[Y|x_0]$ F

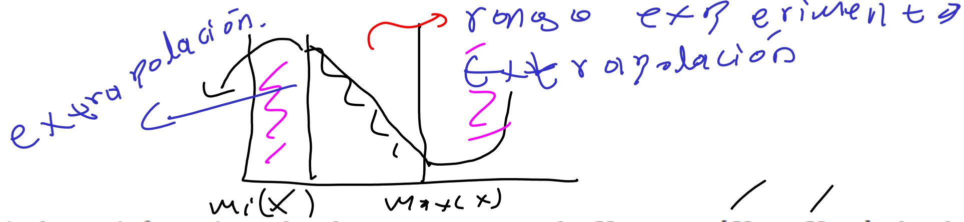
Estimación para la respuesta media:

$$E(Y|x_0) = \boxed{y_0 = \beta_0 + \beta_1 x_0}$$

$\rightarrow$ respuesta media

Estimación para $\boxed{\text{valores futuros}}$

$$E(Y|x_0) = y_0 + \beta_0 + \beta_1 x_0 + \boxed{\varepsilon_0}$$



1 ¿Sólo se podrán hacer inferencias sobre la respuesta cuando $X = x_0 \in [X_{\min}, X_{\max}]$, donde $X_{\min}$ y $X_{\max}$ son los valores mínimo y máximo de la variable predictora, que fueron fijados en la muestra.? $\checkmark$

2 Completar la siguiente tabla anova.

Table 2: Tabla ANOVA parcialmente completada

Fuente de Variación	SS (Suma de Cuadrados)	df (Grados de Libertad)	MS (Cuadrado Medio)	F
Entre Grupos	10.25	1	$MSB = \frac{SSB}{1}$	$F_0 = \frac{MSB}{MSE}$
Dentro de Grupos	20.50	$n-2 = 28$	$MSE = \frac{SSE}{n-2}$	
Total	30.75	29		

$n =$ ojo

$29 = n - 1$

$n = 30$

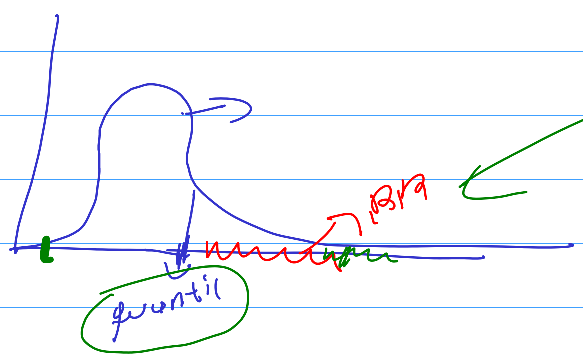
$MSB = 10.25$

$MSE = \frac{20.50}{28} = 0.375$

$H_0: B_1 = 0$

$H_1: B_1 \neq 0$

$$\boxed{F_0 = \frac{10.25}{0.375} = 27.33}$$



ubico en la RR 0.15

Nota: cuando el F_0 es grande generalmente se rechaza H_0

Parte practica

Punto 3

$$\boxed{Y_i = B_0 + B_1 X + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)}$$

E. general.

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -41627.883   4889.428  -8.514   <2e-16 ***
train_data$sqft_living  279.714      2.151 130.055   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$$\boxed{\hat{Y} = -41627.883 + 279.714X}$$

Modelo ajustado.

Punto 4

1 $H_0: B_1 = 0$

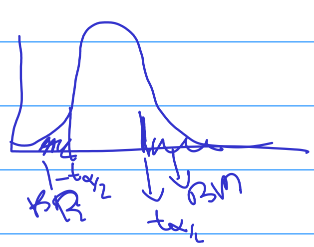
$H_a: B_1 \neq 0$

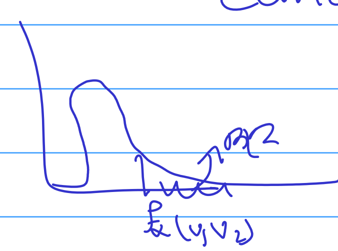
Prueba de significancia para la pendiente

Prueba de significancia para la regresión

2 $T = \frac{B_0}{Se(B_0)}$ (tabla de coeficientes)

2 $F_0 = \frac{MSR}{MSE}$ (anova)

3 

3 

$H_0: B_1 = 0.$

$H_a: B_1 \neq 0$

Prueba de la significancia para la regresión

```
Analysis of Variance Table

Response: train_data$price
                        Df     Sum Sq    Mean Sq F value    Pr(>F)
train_data$sqft_living     1 1.1427e+15 1.1427e+15   16914 < 2.2e-16 ***
Residuals              17288 1.1680e+15 6.7559e+10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

VP.

$\boxed{F_0 = 16914}$

VP < α.

↳ rechazo Ho

Prueba de la significancia para la pendiente.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -41627.883   4889.428  -8.514   <2e-16 ***
train_data$sqft_living    279.714      2.151 130.055   <2e-16 ***
---
```

VP=

T=130.055

VP < α

↳ rechazo Ho.

Se rechaza Ho por los dos caminos.

Nota: la prueba tipo t (pendiente) es una prueba individual, es decir mide la significancia de una sola variable.

* La prueba tipo F (anova) es una prueba general, evalua la significancia general del modelo.

$Y = \beta_0 + \boxed{\beta x}$

t

$Y = \beta_0 + \boxed{\beta_1 x + \beta_2 x}$ → una sola prueba F

dos pruebas tipo t

5. De una interpretación de los parámetros β_0 y β_1 del modelo, claro está, si es posible hacerlo.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -41627.883   4889.428  -8.514   <2e-16 ***
train_data$sqft_living  279.714      2.151 130.055   <2e-16 ***
---
```

Para mirar si β_0 es interpretable necesitamos evaluar si el $X=0$ está entre $(\min(x), \max(x))$

```{r}
min(train_data$sqft_living)
max(train_data$sqft_living)
```

```
[1] 380
[1] 13540
```

β_0 No tiene interpretación.

β_1

↳

Por cada unidad de aumento en el area cuadrada habitable el promedio del precio del la propiedad cambia en 279.1

$\hat{\beta}_1$

```
Call:
lm(formula = train_data$price ~ train_data$sqft_living)

Residuals:
     Min       1Q   Median       3Q      Max
-1465693  -147174   -24984   106153  4371080

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -41627.883   4889.428  -8.514   <2e-16 ***
train_data$sqft_living    279.714      2.151 130.055   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 259900 on 17288 degrees of freedom
Multiple R-squared:  0.4945,	Adjusted R-squared:  0.4945
F-statistic: 1.691e+04 on 1 and 17288 DF,  p-value: < 2.2e-16
```

R-squared = 0.4945

$$R^2 = \frac{SSR}{SST} = \frac{1.14 \times 10^{15}}{1.14 \times 10^{15} + 1.168 \times 10^{15}} \approx 0.4945$$

```
Analysis of Variance Table

Response: train_data$price
                          Df     Sum Sq    Mean Sq F value    Pr(>F)
train_data$sqft_living     1 1.1427e+15 1.1427e+15   16914 < 2.2e-16 ***
Residuals              17288 1.1680e+15 6.7559e+10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

el modelo explica el 49.45 % de la variabilidad total de la variable explica