

Pregunta 1. Se tiene un conjunto de datos que se usa ampliamente en la investigación médica, este contiene información relacionada el avance de la diabetes en pacientes y de estos se tiene la siguiente información edad, sexo, genero, índice de masa corporal, presión arterial y seis medidas diferentes de serología sanguínea. En este ya se realizo un modelo de regresión lineal teniendo como variables independiente el índice de masa corporal y la edad, ¿Con los resultados de los modelos de regresión lineal cual diría que es factor que mejor explica el avance de la enfermedad?

OLS Regression Results							OLS Regression Results						
Dep. Variable:	avance_de_la_enfermedad		R-squared:		0.344		Dep. Variable:	avance_de_la_enfermedad		R-squared:		0.035	
Model:	OLS		Adj. R-squared:		0.342		Model:	OLS		Adj. R-squared:		0.033	
Method:	Least Squares		F-statistic:		230.7		Method:	Least Squares		F-statistic:		16.10	
Date:	Mon, 11 Sep 2023		Prob (F-statistic):		3.47e-42		Date:	Mon, 11 Sep 2023		Prob (F-statistic):		7.06e-05	
Time:	19:36:31		Log-Likelihood:		-2454.0		Time:	19:37:34		Log-Likelihood:		-2539.2	
No. Observations:	442		AIC:		4912.		No. Observations:	442		AIC:		5082.	
Df Residuals:	440		BIC:		4920.		Df Residuals:	440		BIC:		5091.	
Df Model:	1						Df Model:	1					
Covariance Type:	nonrobust						Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]		coef	std err	t	P> t	[0.025	0.975]
Intercept	152.1335	2.974	51.162	0.000	146.289	157.978	Intercept	152.1335	3.606	42.192	0.000	145.047	159.220
bmi	949.4353	62.515	15.187	0.000	826.570	1072.301	age	304.1831	75.806	4.013	0.000	155.196	453.170
Omnibus:	11.674		Durbin-Watson:		1.848		Omnibus:	52.996		Durbin-Watson:		1.921	
Prob(Omnibus):	0.003		Jarque-Bera (JB):		7.310		Prob(Omnibus):	0.000		Jarque-Bera (JB):		26.909	
Skew:	0.156		Prob(JB):		0.0259		Skew:	0.438		Prob(JB):		1.43e-06	
Kurtosis:	2.453		Cond. No.		21.0		Kurtosis:	2.167		Cond. No.		21.0	

Pregunta 2. A partir del resumen de la regresión del número de bateos de un equipo de beisbol sobre el número de carreras (runs), responda a lo siguiente:

OLS Regression Results						
Dep. Variable:	runs	R-squared:	0.373			
Model:	OLS	Adj. R-squared:	0.350			
Method:	Least Squares	F-statistic:	16.65			
Date:	Mon, 11 Sep 2023	Prob (F-statistic):	0.000339			
Time:	19:29:32	Log-Likelihood:	-167.44			
No. Observations:	30	AIC:	338.9			
Df Residuals:	28	BIC:	341.7			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-2789.2429	853.696	-3.267	0.003	-4537.959	-1040.526
bateos	0.6305	0.155	4.080	0.000	0.314	0.947
Omnibus:	2.579	Durbin-Watson:	1.524			
Prob(Omnibus):	0.275	Jarque-Bera (JB):	1.559			
Skew:	0.544	Prob(JB):	0.459			
Kurtosis:	3.252	Cond. No.	3.89e+05			

- ¿Cuál es la correlación entre los rendimientos el # de bateos y las carreras?
- ¿Cuál es el intervalo de confianza del 95% para el numero de bateos?
- ¿Qué valor se puede predecir del modelo si el número de bateos del equipo es 5550?

Pregunta 3. La siguiente tabla resumen muestra la relación que tiene el número medio de habitaciones por vivienda (RM) y el precio medio de la vivienda ocupada por sus habitantes en miles de dólares (MEDV), responda o siguiente:

OLS Regression Results						
=====						
Dep. Variable:	MDEV	R-squared:	0.484			
Model:	OLS	Adj. R-squared:	0.483			
Method:	Least Squares	F-statistic:	471.8			
Date:	Mon, 11 Sep 2023	Prob (F-statistic):	2.49e-74			
Time:	23:23:52	Log-Likelihood:	-1673.1			
No. Observations:	506	AIC:	3350.			
Df Residuals:	504	BIC:	3359.			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	-34.6706	2.650	-13.084	0.000	-39.877	-29.465
RM	9.1021	0.419	21.722	0.000	8.279	9.925
=====						
Omnibus:	102.585	Durbin-Watson:	0.684			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	612.449			
Skew:	0.726	Prob(JB):	1.02e-133			
Kurtosis:	8.190	Cond. No.	58.4			

- (a) ¿Cómo se ve afectada la medida estimada del MEDV por RM
 (b) ¿Qué nos indica el valor 'F-statistic' y el valor 'Prob (F-statistics)' frente a la significancia del modelo?

Pregunta 4. ¿En qué caso el coeficiente (R^2) sería igual a 1 en un modelo de regresión lineal?

- (c) Cuando el tamaño de muestra es muy grande
 (d) Cuando la variable dependiente Y no se explica con el modelo
 (e) Cuando el modelo de regresión pasa por todos los puntos del set de datos, sin errores de predicción
 (f) No podemos saberlo sin el tamaño de la muestra

Pregunta 5. Un gerente de una cadena de restaurantes quiere determinar si hay una diferencia significativa en la satisfacción del cliente entre dos sucursales. Para esto recopiló datos calificaciones de satisfacción de los clientes en las dos sucursales.

Con esto realizo una prueba de hipótesis para determinar si hay una diferencia significativa en la satisfacción del cliente entre las sucursales, usando un nivel de significancia del 0.05.

Considere la siguiente secuencia:

```
import numpy as np
from scipy import stats

# Datos de los dos grupos
sucursal_a = np.array([8, 9, 7, 6, 8, 9, 8, 7, 6, 7, 9, 8, 7, 6, 8, 9, 7, 6, 8, 9, 8, 7, 6, 7, 9, 8, 7, 6, 8, 9, 7, 6, 8, 9, 7, 6, 7, 9])
sucursal_b = np.array([7, 8, 6, 5, 7, 8, 7, 6, 5, 6, 8, 7, 6, 5, 7, 8, 6, 5, 7, 8, 7, 6, 5, 6, 8, 7, 6, 5, 7, 8, 6, 5, 6, 8, 7, 6, 5, 7, 8, 6, 5, 6])

# Realizar la prueba t de Student para muestras independientes
t_statistic, p_valor = stats.ttest_ind(sucursal_a, sucursal_b)

# Imprimir los resultados
print("Estadístico t:", t_statistic)
print("Valor p:", p_valor)
```

Estadístico t	Valor p
4.623	1.228

¿Se rechaza o se acepta la hipótesis nula?