REGRESSION OUTPUT EXPLAINED

Basic Information about model fit ↑ | Goodness of fit (table) ↑

DEPENDENT VARIABLE	~~Diabetes~~ Houseprice	R Squared	0.967
MODEL	OLS	Adjusted R Square	0.965
METHOD	Least Square	F statistic	415.1
DATE	Tue, 07 Dec 21	Prob (F statistic)	8.36e-12
NUMBER OF OBSERVATIONS	16	Log-Likelihood	-14.9
DF Residuals	14	AIC	33.8
DF Model	1	BIC	35.3
Covariance Type	non robust		

	Coefficient	Standard Error	t	P>|t|	[0.025	0.975]
Intercept	51.39	0.68	76.08	0.00	50.38	53.305
X1	0.75					
GNP	0.0348	0.002	20.374	0.000	0.31	0.038

→ Report of each coefficient

Omnibus	1.925	Durbin-Watson	1.619
Prob(omnibus)	0.382	Jaque-Bera (JB)	1.215
Skew	0.664	Prob(JB)	0.595
Kurtosis	2.759	Condition Number	1.66e+03

→ Table of distribution of residuals

OUTPUT EXPLAINED.

1. **Dependent Variable** → Which variable is the response in the model.
2. **Model** → what model you are using in the fit.
3. **Method** → How the parameter of the models were calculated.
4. **No of Observations** → Total rows in the ~~column~~ dataset (model).
5. **DF ~~Residuals~~ Models.** → Total columns in the model. ~~Degree~~ Degree of Freedom of ~~residuals~~. columns.
6. **DF Residuals** → Degree of freedom of residuals.
   = Number of observations (rows) − No of parameters (columns) + 1

Because formula of Degree of Freedom = N − K + 1

N → Sample size (No of observations)

K → Number of Variables

7. **R Squared** → **Coefficient of Determination** denoted by R^2 is the proportion of the variation in the dependent variable that is predictable from the independent variable.

8. **Adjusted R Square** → R square value adjusted based on the number of observations and the degree of freedom of residuals.

$$\text{Adjusted } R^2 = 1 (1 - R^2) \frac{n-1}{n-k-1}$$

9. **F Statistic** → F test of overall significance indicates whether linear regression model provides a better fit to the data than a model contains no independent variable. In other word, it compares the fit of the intercept only model.

Null hypothesis : Model with no independent variable fits the data well as your model.

Alternate hypothesis : Model fits the data better than the intercept only model.

$$F = \frac{\frac{R^2}{(k-1)}}{\frac{(1-R^2)}{(n-k)}}$$

Directly check Prob (F statistic)

10. **Prob (F-statistic)** → If P value < 0.05, accept ~~the~~ Alternate hypothesis (Most Cases) that is Model fits the data better than the intercept only model.

11. **Log likelihood** →
- Likelihood Ratio test (often termed as LR test) is a test to compare two models concentrating on the improvement with respect to likelihood values.
- If we **keep on adding predictor variables** to a linear model, **R square will improve**. This **holds true for model likelihood** as well. But the objective is to check if the **improvement in likelihood is good enough or not.**
- **Higher the value of likelihood**, better is the **fit of the model**.

Whenever we use Model = MLE (Maximum likelihood Estimation) we get log likelihood real value, in Model = OLS it will also give Log likelihood but not correct.

12. AIC → ~~Akaike~~ Akaike Information Criterion.

13. BIC → Bayesian Information Criterion.

AIC / Akaike Information Criterion is a method for scoring and selecting a model.

$$\boxed{AIC = -2/N * LL + 2 * K/N}$$

or $\boxed{AIC = -2(LL) + 2*(K)}$

N → Number of observations

K → Number of parameters

LL → Log-likelihood.

The score is minimized, the ~~log~~ model with lowest AIC is selected.

Compare to BIC method, AIC penalize complex model less, meaning it may put more emphasis on model performance on the training dataset and in turn select more complex models.

BIC / Bayesian Information Criterion is also used for scoring & selecting a model.

$$\boxed{BIC = -2 * LL + \log(N) * K}$$

Score is minimized, the model with the lowest BIC is selected.

Unlike AIC, BIC penalize more for model complexity and is less likely to be selected.

13. Coefficient → The coefficient terms tells the change in Y for a unit change in X. In the above output, if X rises by 1 unit then Y rises by 0.75. In other word, rate of change of Y with respect to X.

14. Standard Error → Standard Error of the estimate of the coefficient. Standard error shows samples variability of these parameters.

Continuation of Coefficient →

- Sign of the coefficient (positive / negative) gives direction of the effect.
- In simple linear regression (single independent variable), coefficient tells us how much dependent variable is expected to increase (if coefficient is +ve) or decrease (if coefficient is -ve) when independent variable increase by one.
- In multiple regression (more than one independent variable) coefficient tells us how much dependent variable is expected to increase when independent variable increased by one keeping all other independent variable constant.

	Coefficient	Standard Error	t stat	P value(t)
Intercept				
	0.515	0.026	19.92	0.00

Standard error of the regression coefficient how much uncertainty is associated with coefficient.
Suppose confidence Interval is at 95%.
So, lower frontier of Confidence Interval will be → 0.51 − 0.02 (Standard Error) = 0.49.

Higher frontier of CI = 0.51 + 0.02 = 0.53 So, coefficient will range from 0.49 to 0.53

15) T stat is equal to regression coefficient divided by standard error
= 0.51 / 0.02 = 19.4.

In other word, your regression coefficient stands 19 Standard Error from zero.

16) P(t) → If P value is less than 0.05, then 5% only chance that result would have come up with random distribution in other word we can say 95% probability that variable has some effect

Null hypothesis - Variable (x) don't have any effect on Y.

Alternate hypothesis - X have some effect on Y.

17) Skewness → symmetry of error around mean. Normally error should be symmetrically distributed about the mean. 0 is perfect symmetry.

18) Kurtosis → Measure shape of distribution. High kurtosis implies less outlier.

19) Omnibus → Describe normalcy of error / residual using skew & kurtosis. '1' will indicate perfectly normal distribution of error.

20) Prob (Omnibus) → Probability of Omnibus

21) Jarque Bera → Alternate method to measure normalcy of error using skewness and kurtosis.

22) Prob (JB) → Probability of JB.

23) Durbin Watson → Presence of autocorrelation (errors are not independent)
- Measurement of homoscedasticity. Heteroscedasticity would imply uneven distribution for example data point grow higher, error grow higher.
- Ideally, homoscedasticity will lie between 1 and 2.

24) Condition Number → A test for multicollinearity. Multicollinearity strongly implies high condition number.