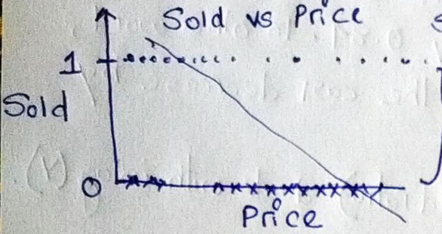
LOGIT MODELS

What if Y (Dependent Variable) is categorical?

→ Example of a car, $Sold = \beta_0 + \beta_1 (Price) + \varepsilon_i$



Suppose $Sold = 0.7989 - 0.0285(Price)$

So in this case, as the price increases, selling of car decrease.

Probability (Car being sold), suppose 1 - Yes, 0 - No.

But the problem is if line (value) > 1 and line (value) < 0, only two answer it can take, either yes or NO.

Midpoint - 0.5

So we cannot use OLS (Ordinary least Square) and limit the probability to [0,1].

Therefore we will transform Probability P to $\boxed{\frac{P_i}{1-P_i}}$ (First Modification) this will change the range from

This is also known as the ODDS ← New Range - $[0,\infty]$

→ So, if we have probability, $p = 50\% = 0.5$ odd $= \frac{0.5}{1-0.5} = 1$. so ODDS = 1.

$p = 80\% = 0.8$ odd $= \frac{0.8}{1-0.8} = \frac{0.8}{0.2} = 4$ so ODDS = 4.

But ODD can get to very large number.

If we log the ODD $\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 (Price) + e$. Range $= [-\infty, \infty]$ Midpoint $= 0$.

This is known as Log ODD.

This is also known as binomial logistic regression. Binomial because we will have only two option. In our case either car is sold or not sold. and logistic because we have log odd situation. So whenever we use log odd then it will be logistic or logit.

BINOMIAL LOGISTIC REGRESSION $= \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 (Price) + \varepsilon_i$

Suppose we add another variable pink slip.

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 (Price) + \beta_2 (Pink\ Slip) + \varepsilon_i$$

Model output →

Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	−0.173	0.057	−3.04	0.0023	0.84	0.75	0.94
Pink Slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

By seeing p value < 0.05, we can say Price and pink slip are very important variable.

$$\log\left(\frac{\hat{P}}{1-P_i}\right) = 0.396 - 0.173 (Price) + 1.555 (Pink\ Slip)$$

Inference, for a $1000 increase in price, the log-odds of selling the car decreased by 0.173 on average holding all else constant.

So what is log odd of selling the car?

OR in the table, that is odd ratio.

OR is the multiplicative effect of one extra unit of our X variable on the odds of selling a car so if price increases by one unit or one thousand dollars the odds of selling the car will be multiplied by 0.84. In other word, for a \$1000 increase in price, the odds of selling the car decrease by 16%, on average holding all else constant.

Coeff in the table deal with the log (Y), OR (odd ratio) deals with unlog (Y).

For pink slip, odd ratio (OR)

Cars with a pink slip have 4.73 times the odds of being sold compared to cars without pink slip on average holding all else constant.

Remember → "Chance" = "Probability" ≠ "Odds"

If the probability of rain tomorrow is 20%, what are the odds of rain tomorrow?

probability = $1/5$ odds $= \frac{p_i}{1-p_i} = \frac{0.2}{0.8} = 0.25$ or $1/4$.

Q) Find the probability that a \$4,500 car with a pink slip will sell?

$$\ln\left(\frac{p}{1-p}\right) = 0.396 - 0.173(4.5) + 1.555(1) = 1.1725.$$

$$\ln\left(\frac{p}{1-p}\right) = 1.1725$$

$$\frac{p}{1-p} = e^{1.1725}$$

$$\frac{p}{1-p} = e^{1.1725}, \quad p = (e^{1.1725})(1-p), \quad p = 3.23(1-p)$$

$$p = 3.23 - 3.23p$$

$$p + 3.23p = 3.23 \Rightarrow 4.23p = 3.23$$

$$\Rightarrow p = \frac{3.23}{4.23} = 0.764.$$

So, a car with a sale price of \$4,500 and a pink slip has a probability of sale 76.4%.

~~Q) Find the probability that a \$4,000 car~~