Dhanvanth Voona

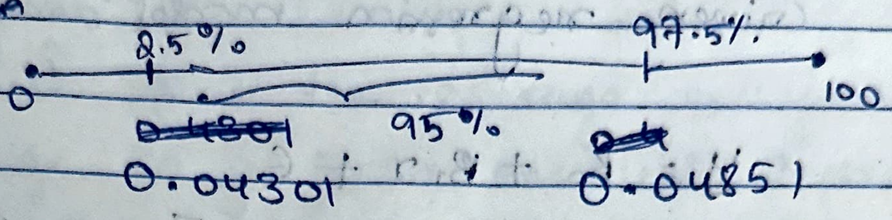
A20543395

Homework 5

1A (a) β_1 is the coefficient of TV feature.

We have for β_1

~~Given~~



So 2.5%, 97.5% are the lower and upper bounds of 95% Confidence Interval of β_1. That is model is 95% confident that true effect of TV advertising lies between 0.04301 & 0.0485) assuming other features remain constant.

(b) If we look at the lower and upper bound β_3 which is $[-0.012, 0.0105]$. Here the Confidence Interval 0 (zero) in it. ~~Which means~~ If $\beta_3 = 0$, it shows there is no effect of amount spent on newspapers on sales.

Because the confidence Interval includes 0, we do not have enough evidence to reject null hypothesis. Hence we fail to reject null hypothesis (H_0).

2 (a) Given regression model:

$$y = \beta_0 + \beta_1 x + \epsilon$$

We know $E(\epsilon) = 0$ [for unbiased regression]

Then:

$$E[Y/x] = E[\beta_0 + \beta_1 x + \epsilon/x]$$

$$= \underbrace{E[\beta_0]}_{\text{constant}} + E[\beta_1 x] + \underbrace{E(\epsilon)}_{0}$$

$$E[Y/x] = \beta_0 + \beta_1 x \quad \text{--- eq (1)}$$

lets suppose x is increased by unit (1)

then $E[Y/x+1] = \beta_0 + \beta_1 (x+1)$

$$= \beta_0 + \beta_1 x + \beta_1 \quad \text{--- eq (2)}$$

So, lets calculate expected change in y for one unit increase in x

i.e.

$$\Delta E[Y/x] = E[Y/x+1] - E[Y/x]$$

from eq (1) & eq (2)

$$\Delta E[Y/x] = \cancel{\beta_0} + \cancel{\beta_1 x} + \beta_1 - \cancel{\beta_0} - \cancel{\beta_1 x}$$

$$= \beta_1$$

So $\hat{\beta}_1$ gives the average or expected change in y for a one unit increase in x, rather than simply change in y for one unit increase in x.

2(b) In multiple linear regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

the value of $\hat{\beta}_1$ (estimate coefficient of x_1) is interpreted as expected change in target variable y for a one unit increase in x_1, while keeping x_2 as constant.

for example Take house pricing (y) & house size (x_1) & no. of bedrooms (x_2): Here $\hat{\beta}_1$ is expected change in house price (y) for one-unit increase in house size (ex: for an extra square foot) but only if no. of bedrooms (x_2) remains same.

3A

(a) from previous homework

$R^2 = 0.7501$ for regression model:

muscle mass $= \beta_0 + \beta_1 Age + \epsilon$

We know Adjusted $R^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1}\right)$

→ $n = 59$ [found it using nrow function in R]

∴ Adjusted $R^2 = 1 - \left(\frac{0.2499 \times 58}{57}\right)$

$= 1 - \frac{14.49}{57} = 1 - 0.254$

Adjusted $R^2 = 0.745$ — eq ①

Adjusted R^2 slightly lesser than R^2 due to considering only one predictor.

In case ② given adjusted $R^2 = 0.8223$

→ Combination of age & estrogen explains 82.23% of variability in muscle mass.

The increase in adjusted R^2 from 0.7453 to 0.8223 indicates estrogen is significant predictor of muscle mass.

3(b) 3rd model [Adjusted R^2 = 0.851] ⇒ greater than 2nd mode [Adj R^2 = 0.822]

This suggests predictors like sleep, protein intake & exercise combined together able to explain variability of muscle mass better than age & estrogen level together. This means these 3 features are significant & needed to be included in feature selection for muscle mass prediction

15A In case of (i) confidence Interval :-
It estimates the mean response ($\hat{y}$) for a given value of x. i.e In finding expectation of y, the error will be diminished and reduced to zero.
ii. The prediction interval predicts the individual new response y_{new} for a given value of x.

So PI accounts for both uncertainity in estimating population mean & addition variability that comes with predicting new ~~va~~ individual response. To account for this addition variability (ϵ) P.I needs to be wider.

So basically when you are talking about C.I, we are giving range of average of that ~~t~~ unit [for ex: temperature]. C.I give 95% ~~ra~~ range of average temperature for tomorrow, here it does not have to account for individual uncertainity of temperatures since they are vanished while finding expectancy, but in case of P.I

It has consider individual uncertainity too. So it requires more length in the interval to be confident.