

1) a) for linear regression, the dependent variable is continuous while for logistic regression, the dependent variable is binary.

Linear regression is used to establish a linear relationship between dependent and independent variables while logistic regression is used to find the odds ratio or used to find the probability of an event.

$$(b) i) \text{logit}(p) = -3.0231 + 0.0065X_1 + 0.0795X_2 - 0.0702X_3 + 0.0489X_4$$

where, X_1 - sbp

X_2 - tobacco

X_3 - obesity

X_4 - age

$$\therefore \text{logit}(p) = -3.0231$$

$$\log\left(\frac{p}{1-p}\right) = -3.0231$$

$$\begin{aligned} \text{odds} &= \exp(-3.0231) \\ &= 4.87\% \end{aligned}$$

(b) ii) Intercept is the value when sbp=0, tobacco=0, obesity=0, age=0.

$$\begin{aligned} (b) iii) \text{logit}(p) &= -3.0231 + 0.0065(200) + 0.0795(30.2) - 0.0702(40.2) + 0.0489(60) \\ &= 0.78976 \end{aligned}$$

$$\log\left(\frac{p}{1-p}\right) = 0.78976$$

$$\frac{p}{1-p} = \exp(0.78976) = 2.2029$$

$$p = 0.6878$$

(b) iv) Model 2 is better: ① Model 2 has smaller AIC

② Model 2 has residuals much closer to the d.f. indicating that is a better model.

2)a) Measurements of the same individuals are taken repeatedly through time.

2)b) Fixed effects – the average relationships between predictor variables and the dependent variable across the entire pop.

Random effects – capture individual variability and differences between subjects that cannot be explained by the fixed effects.

2)c)i) Time 6 months: the coefficient (-7.998) suggests that on average the cholesterol levels decreased by approximately 7.998 units at the 6-month time point compared to the baseline.

Time 12-months: the coefficient (-11.887) suggests that on average the cholesterol levels decreased by approximately 11.887 units at the 12-month time point compared to the baseline.

3)a)	FA	PCA
① produces factors		produces components
② factors cause variables		components are aggregates of variable
③ analyses only the variance shared among the variables		analyses all the variance

3)b) It is a measure of how suitable the available data is to be used for factor analysis. Bartlett's test should be significant ($p < 0.05$) for the FA to be considered appropriate.

3)c)i) FA can reduce these numerous attributes into fewer factors or underlying dimensions that captures the essence of the original data. This not only simplifies the dataset, but also makes it more interpretable.

3)c)ii) ① Bartlett's test of sphericity

② Kaiser-Meyer-Olkin

3)c)iii) 6 factors due to the biggest drop is from 6 to 7.

3)c)iv) MR1: Economical choice

MR2: Family's choice

MR3: Texture

MR4: Experience

MR5: Taste

MR6: Healthiness