1. What does ϵ account for? Give example.

> ϵ accounts for non-linearity, measurement errors and other latent variables which are not observed in our data. Consider a linear process given as $y = 5x + 2$, if due to inaccuracies in instruments if we have random noise added while taking readings we can denote this noise as epsilon with the process thus, indicated as $y = 5x + 2 + \epsilon$, where ϵ is the unobserved error added by the instrument into our data.

2. Why is the case study in this notes is a regression problem?

> The problem stated above is a regression problem because the labels associated with the data are numeric / real values.

3. Is the case study a supervised or unsupervised problem? Why?

> This is a supervised learning problem, since associated with each data point (here budget invested in TV, radio and newspaper ads) there is a label / ground truth (here the number of units sold).

4. When is scatter plot useful?

> Scatter plots are used for estimating relationships between predictor and response variable and also to find the influence of a predictor on the response. They are also used to find trends in the residuals plots, that is find trends in error (where there should be none).

5. What is the meaning of $\hat{a}$ and $\hat{b}$ being unbiased estimators of a and b?

> $\hat{a}$ and $\hat{b}$ are said to unbiased estimator of a and b if the mean of their sampling distribution is same as a and b. Formally, $\hat{a}$ and $\hat{b}$ are statistic of the sample and can be considered as random variables we say that $\hat{a}$ and $\hat{b}$ are unbiased estimators if $E[\hat{a}] = a$ and $E[\hat{b}] = b$

6. What is the z-value so that $P(-z \leq \frac{x-\mu}{\sigma} \leq z) = 0.99$?

> z=2.576

7. What is the z-value so that $P(-z \leq \frac{x-\mu}{\sigma} \leq z) = 0.90$?

> z=1.645

8. when is z-statistic used and when is t-statistic used when testing a pair of hypotheses for mean value?

> We use z-statistic when we have the standard deviation (σ) of the population for whose mean we are performing a hypothesis test, while we use t-statistic when we have only the estimate of standard deviation $(\hat{\sigma})$ obtained from the sample. This is due to the fact that the sampling distribution of the random varible $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ which we use to perform the test of hypothesis comparing means follows normal distribution if standard deviation of the population is known else follows t-distribution when we only have a sample based estimate for standard deviation.

9. What does $n-1$ degrees of freedom for t-statistic mean?

> Degrees of freedom is the number of values in the final calculation of a statistic that are free to vary. When we calculate mean with n samples, if the mean is known we can at the most vary n-1 samples and not more, since the nth sample is determined with the help of mean and n-1 samples, thus, t-statistic which represents the density of the sampling distribution of mean has n-1 degrees of freedom.

10. Why y has same variance as ϵ?

> We have the noise ϵ associated with our predictions which gives us the formulation for the target variable as $y = a + \sum_{i=1}^{n} b_i x_i + \epsilon$. We have,
>
> $$Var(y) = Var(a + \sum_{i=1}^{n} b_i x_i + \epsilon) = Var(c + \epsilon) = Var(\epsilon)$$

11. Why $E[y|x] = f(x)$?

> We have,
>
> $$\begin{aligned} E[y|x] &= E[f(x) + \epsilon|x] \\ &= E[f(x) + \epsilon] \\ &= E[f(x)] + E[\epsilon] \\ &= E[f(x)] + 0 \\ &= f(x) \end{aligned}$$

12. What happens when $n \leq p+1$ in multivariate regression where p is the number of predictors?

> When $n \leq p+1$ there is no unique solution to the Normal Equation.

13. What happens if the data matrix $X_{n\times(p+1)}$ is rank deficient?

> When the matrix $X_{n\times(p+1)}$ is rank deficient then some of the predictors are not linearly independent thus, the response will depend not only on the predictors but also, on the interactions between the dependent ones.

14. Why multivariate regression is preferred over single variable regression when multiple predictors are involved?

> Multiple linear model have the drawback that if there is an interaction between two predictors V1 and V2 then in the simple linear model constructed using the V1 predictor the interaction term also plays an influence, as against in the Multivariate model where this interaction can be taken into account.

15. Define linearity in the linear regression model.

> The linearity of the model is considered with respect to the coefficients of the predictor variables, that is the model is linear with respect to coefficients not the data.