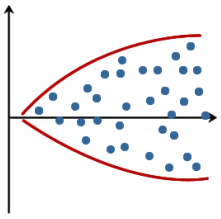
Q1. In your own words, describe what a residual is in linear regression.

= The errors between actual value and predicted value in linear regression is called residual.

Q2. If you know that your residual data follow the below pattern, are your data better approximated with a linear model for the lower values of independent variable or higher values of independent variable and why?

= lower data better approximated because this cone shape tells us that it's good to approximate lower value but for higher value it's inconsistent.



Q3. What is the difference between R^2 and adjusted R^2?

= R squared value assumes that all the independent variables considered affect the result of the model, whereas the adjusted R squared value considers only those independent variables which actually have an effect on the performance of the model.

Q4. Is there independence of observations if you are trying to predict baby length with mother's height?

- Yes
- No

= Yes

Q5. Justify the above answer.

= I think they are independent since there is no relation between them.

Q6. Do residual data show homoscedasticity?

- Yes
- No

= Yes

Q7. Justify the above answer.

= Homoscedasticity means that residual are constant or equally distributed so they show homoscedasticity. If they do not then its called heteroskedasticity.

Q8. What is the value of R^2 and what does this tell you?

= R^2 is regression coefficient in the sample. It says the proportion of variance in the dependent variable can be explained by independent variable.

Q9. Can you consider the relationship between mother's height and baby length a statistically significant linear relationship and why?

= No, it's not significant since there are some outliers in case of length of baby.

Q10. Having the ANOVA table for the linear regression in mind, what is the null and alternative hypothesis in this case?

= Null hypothesis should be zero and alternative hypothesis is not zero in that case

Q11. In your own words, describe what the b_1 is.

= b_1 is slope coefficient. The slope of a regression line (b1) represents the rate of change in y as x changes.

Q12. What does the value of b_1 tell you in practical terms?

=y is dependent on x, the slope describes the predicted values of y given x.

Q13. Could you claim the same for the mother's height in the range between 140cm and 145cm and why?

=(the question is not clear) My thought is answer should be no .

Q14. According to this model, what is the prediction of baby length for mother's height of 170cm?

=52.564

Q15. Report on your findings for predicting baby length with mother's height.

= A linear regression established that mother's height could statistically significantly predict baby length. The regression equation: predicted baby length = 15.334 + (0.219 x mother's height).

Q16. Can you predict baby length with father's age? Why?

=Yes, it's possible to predict baby length regarding father's age since baby length is dependent but father's age is not. So, there is relation between them.

Q17. What does homogeneity of variance mean and why is it important assumption of an independent t-test?

= Homogeneity of variance means the distributions of the outcomes in each group are comparable and similar. It is important assumption of an independent t-test since it makes sure whether the groups are similar or not.

Q18. Is there homogeneity of variance between head circumference for babies of smoking mothers and head circumference for babies of non-smoking mothers?

- Yes
- No

=Yes

Q19. Justify your choice.

= there is homogeneity of variance between head circumference for babies of smoking mothers and head circumference for babies of non-smoking mothers since Levene's test for equality of variances value is more than 0.05 more precisely it is 0.368.

Q20. Do smokers have lighter babies? Justify your answer.

= Yes, The smokers group have lighter babies.

Q21. Do women over 35 have lighter babies? Justify your answer.

= Yes.

Q22. Using the cholesterol dataset, was the diet effective in lowering cholesterol concentration after 8 weeks of use? Justify your answer.

= yes

Q24. Was the margarine diet more effective after 4 weeks of use or after 8 weeks of use? Justify your answer.

- 8 Weeks of use

Q25. If you know that the average cholesterol concentration in healthy adults is 3 mmol/L, would you consider your sample (N=18) significantly better or worse than average adult population? Justify your answer.