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

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?Chart, scatter chart

Description automatically generated

Q3. What is the difference between *R2* and adjusted *R2*?

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

* Yes
* No

Q5. Justify the above answer.

Q6. Do residual data show homoscedasticity?

* Yes
* No

Q7. Justify the above answer.

Q8. What is the value of *R2* and what does this tell you?

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

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

Q11. In your own words, describe what the b1 is.

Q12. What does the value of b1 tell you in practical terms?

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

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

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

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

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

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

Q19. Justify your choice.

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

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

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

Q23. For the above case, what is the null and alternative hypothesis?

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

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.