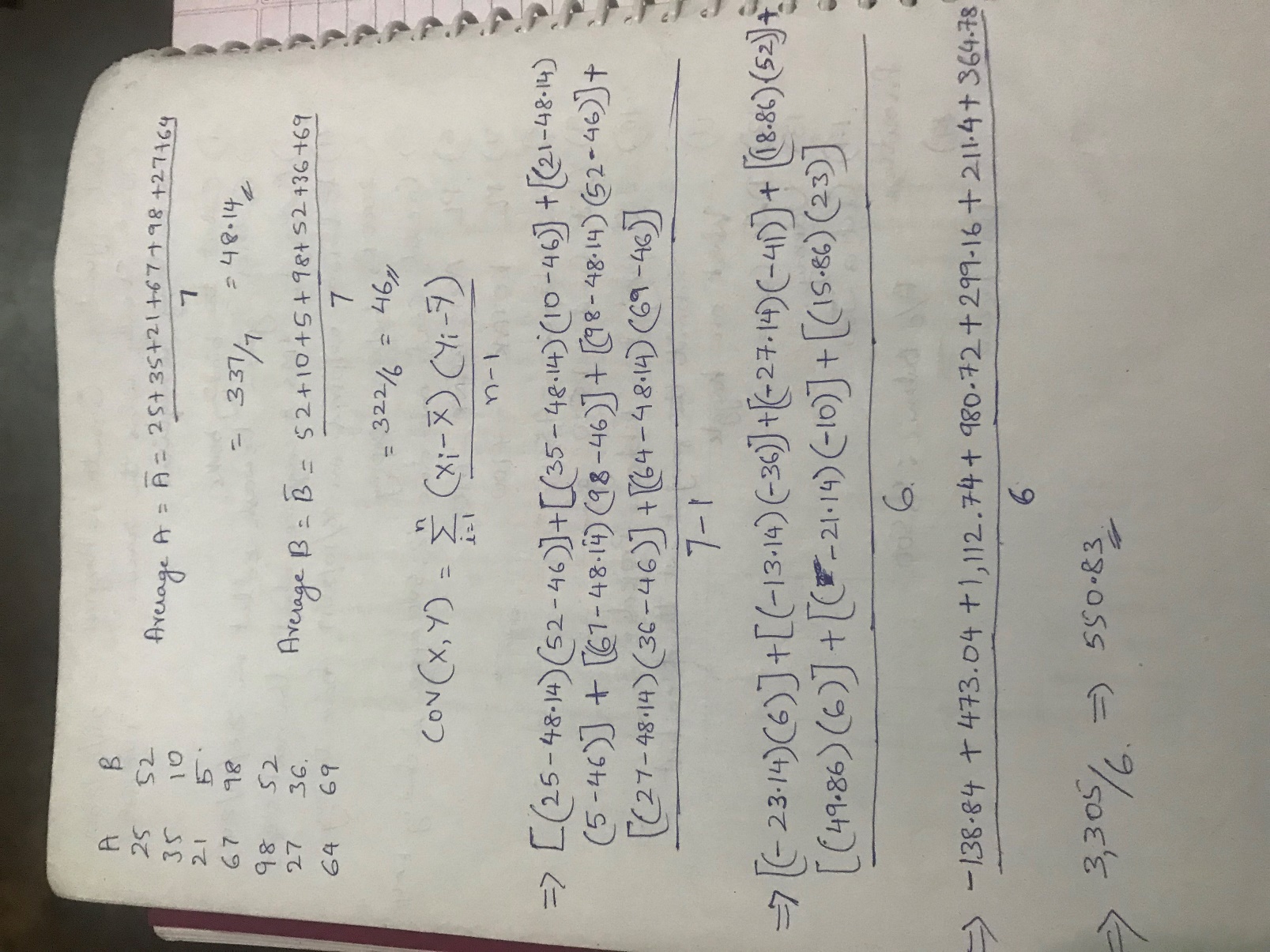
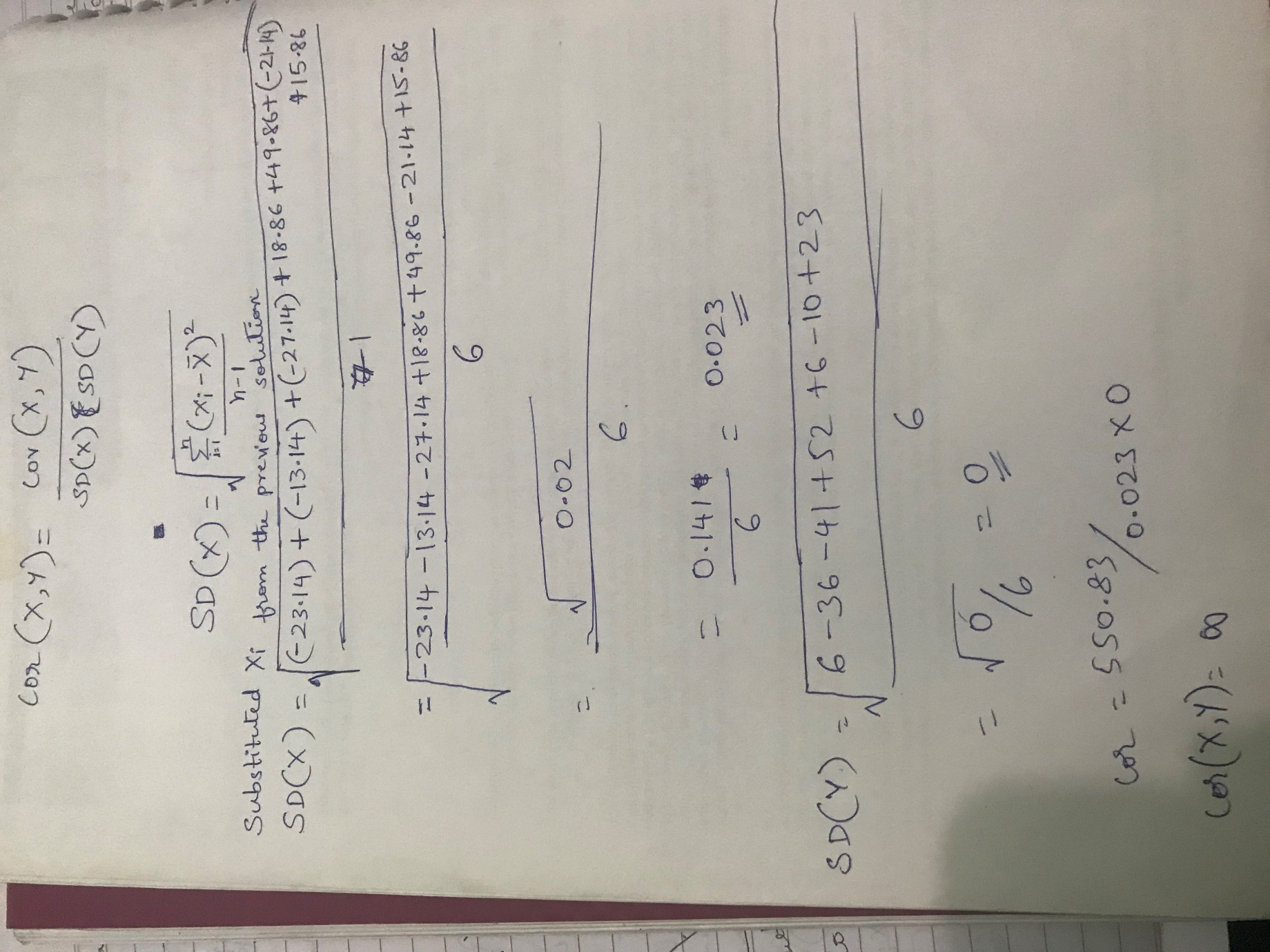
|  |  |
| --- | --- |
| A | B |
| 25 | 52 |
| 35 | 10 |
| 21 | 5 |
| 67 | 98 |
| 98 | 52 |
| 27 | 36 |
| 64 | 69 |
|  |  |

1.Calculate covariance and correlation between below two columns A and B Mention all step by step formula calculations in the answer sheet.

Solution:





2. What are the different ways to deal with multi collinearity? Solution:

* Remove highly correlated predictors from the model. If you have two or more factors with a high VIF, remove one from the model. ...
* Use Partial Least Squares Regression (PLS) or Principal Components Analysis, regression methods that cut the number of predictors to a smaller set of uncorrelated components.

3. What should be the correlation threshold value based on which we determine the highly

collinear variables?

Solution:

If two predictor variables X1 and X2 have a correlation of above about .70, this suggests that one of two things may be occurring.

• X1 and X2 may be two different ways of measuring the same thing.

• X1 and X2 are so strongly confounded that their predictive contributions will be very difficult to separate.

If X1 and X2 are just two measures of the same ‘construct’, the decision what to do is relatively easy. You may be able to combine or average these variables into one measure, or select just one of them to represent the construct.

If X1 and X2 clearly represent different constructs, but they are strongly confounded (highly correlated), their effects are very difficult to separate no matter what analysis you use. The decision what to do in that case is more difficult.

If X1 and X2 are highly correlated with each other, they provide redundant information; they will compete to explain a lot of the same variance in Y.

4. What are the two different types of variable we used in ANOVA?

Solution:

* VarianceWithinSamples(SSE)
* VarianceBetweenSamples(SSC)

5. What are the null and alternate hypothesis in chi-square test?

Solution: Null hypothesis: Assumes that there is no association between the two variables. Alternative hypothesis: Assumes that there is an association between the two variables.