report

Student ID: 11484265

2024-09-23

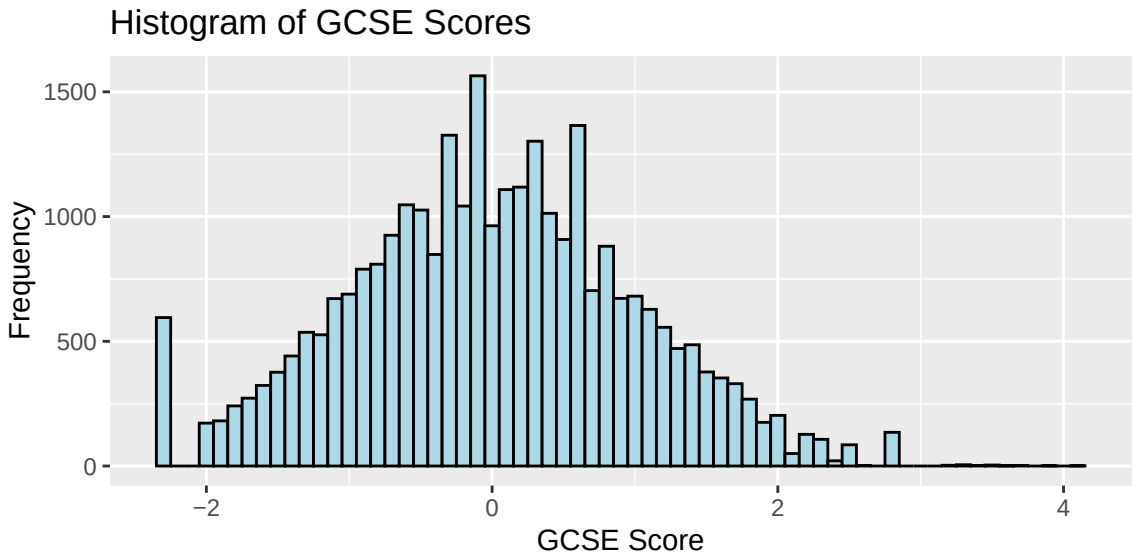
Question 1: Descriptive Analysis (10%)

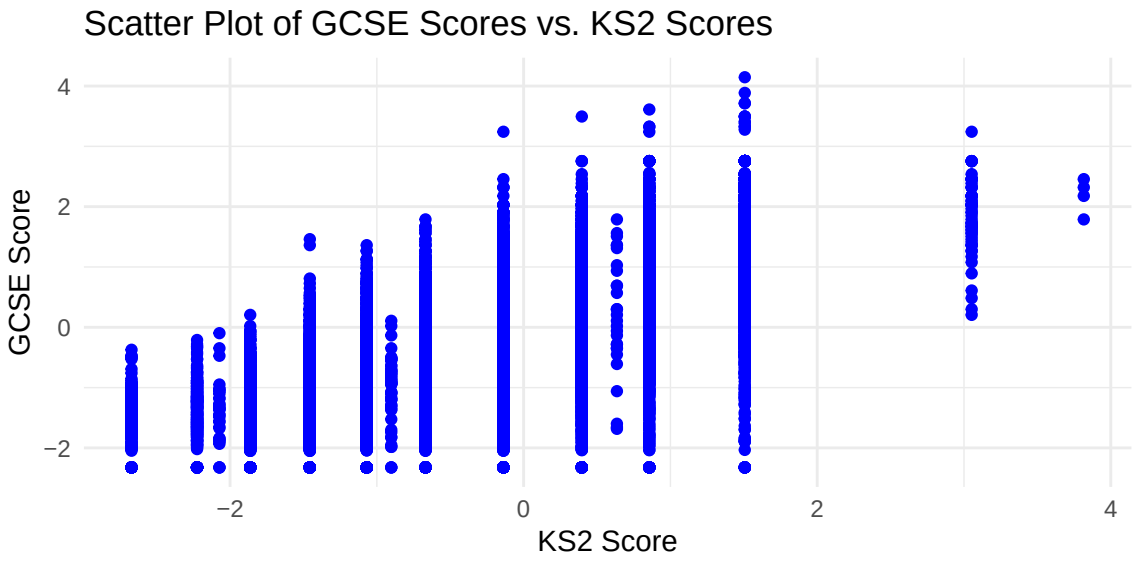
Describe the dataset: -How many pupils and schools are there in the dataset?

There are 54 schools and 29506 pupils.

-What is the average number of pupils per school? 546

-Create histograms for “gcsescore” and “ks2score”.





While there are pupils who score low on the GCSE on a wide range of KS2 scores, the general trend is that higher GCSE scorers generally have higher KS2. Those who scored extremely high on the KS2 score may not always have the highest GCSE, but are all above average.

Question 2: Model Estimation (30%)

Estimate the following multilevel models and summarise the results in a table:

-Model 0 (m0): An empty model with "gcsescore" as the dependent variable and a random intercept at the school level.

-Model 1 (m1): m0 + a fixed effect for "ks2score".

-Model 2 (m2): m1 + "ks2score" squared to capture non-linear effects.

-Model 3 (m3): m2 + a random slope for "ks2score" to allow the relationship between "ks2score" and "gcsescore" to vary by school.

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: gcsescore ~ ks2score + (1 | school)
##    Data: dataw
##
##      AIC      BIC   logLik deviance df.resid
##  61681.1  61714.3 -30836.6  61673.1    29502
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2744 -0.5746  0.0347  0.6353  4.8956
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept) 0.02755  0.1660
##  Residual             0.47047  0.6859
## Number of obs: 29506, groups:  school, 54
##
```

```
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) -0.0005187  0.0229688  -0.023
## ks2score     0.7175608  0.0042911 167.222
##
## Correlation of Fixed Effects:
##          (Intr)
## ks2score 0.006

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: gcsescore ~ ks2score + ks2_sq + (1 | school)
##    Data: dataw
##
##      AIC      BIC   logLik deviance df.resid
##  61612.1  61653.5 -30801.0  61602.1    29501
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.3463 -0.5726  0.0334  0.6370  4.9378
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept) 0.02768  0.1664
##  Residual             0.46934  0.6851
## Number of obs: 29506, groups:  school, 54
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.026398   0.023223  -1.137
## ks2score     0.721979   0.004318 167.208
## ks2_sq       0.028154   0.003337   8.436
##
## Correlation of Fixed Effects:
##          (Intr) ks2scr
## ks2score -0.010
## ks2_sq   -0.132  0.121

## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: gcsescore ~ ks2score + ks2_sq + (ks2score | school)
##    Data: dataw
##
##      AIC      BIC   logLik deviance df.resid
##  61548.4  61606.4 -30767.2  61534.4    29499
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.4575 -0.5753  0.0289  0.6355  4.9397
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  school   (Intercept) 0.028139 0.16775
##           ks2score    0.002482 0.04982  0.34
##  Residual             0.467260 0.68356
## Number of obs: 29506, groups:  school, 54
```

```
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.026964   0.023406  -1.152
## ks2score     0.718022   0.008103  88.614
## ks2_sq       0.023710   0.003434   6.903
##
## Correlation of Fixed Effects:
##          (Intr) ks2scr
## ks2score  0.277
## ks2_sq   -0.126  0.075
```

Compare how the inclusion of "ks2score" and its squared term "ks2score2" affects the model fit and the interpretation of the relationship between "ks2score" and "gcsescore". Specifically, analyse how the introduction of non-linearity in Model 2 improves the model fit over Model 1.

including ks2: variance in schools went down by 0.05; residual went down by half. including ks2 sq: var and res did not change. relationship between "ks2score" and "gcsescore": positive... ~.72 m2 over m1: ???

Additionally, evaluate the significance and impact of allowing the slope of "ks2score" to vary by school in Model 3. Discuss how these modifications influence the explanation of variance between schools and within schools and how they reflect the underlying educational processes.

not that impactful... ????? wtf