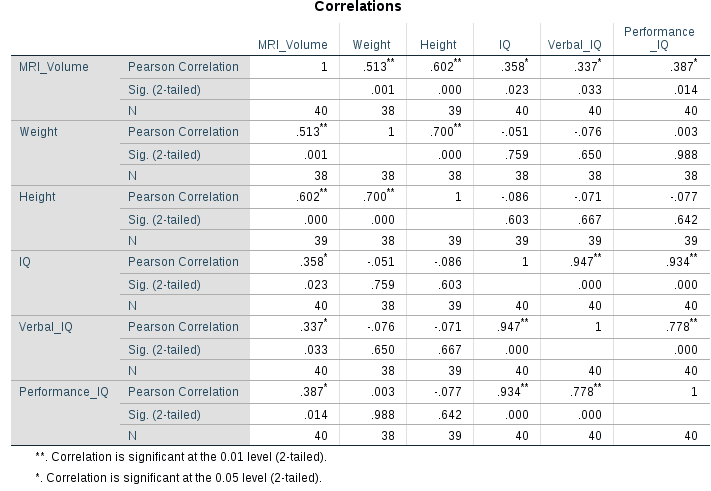
**COGS 536 Homework 4**

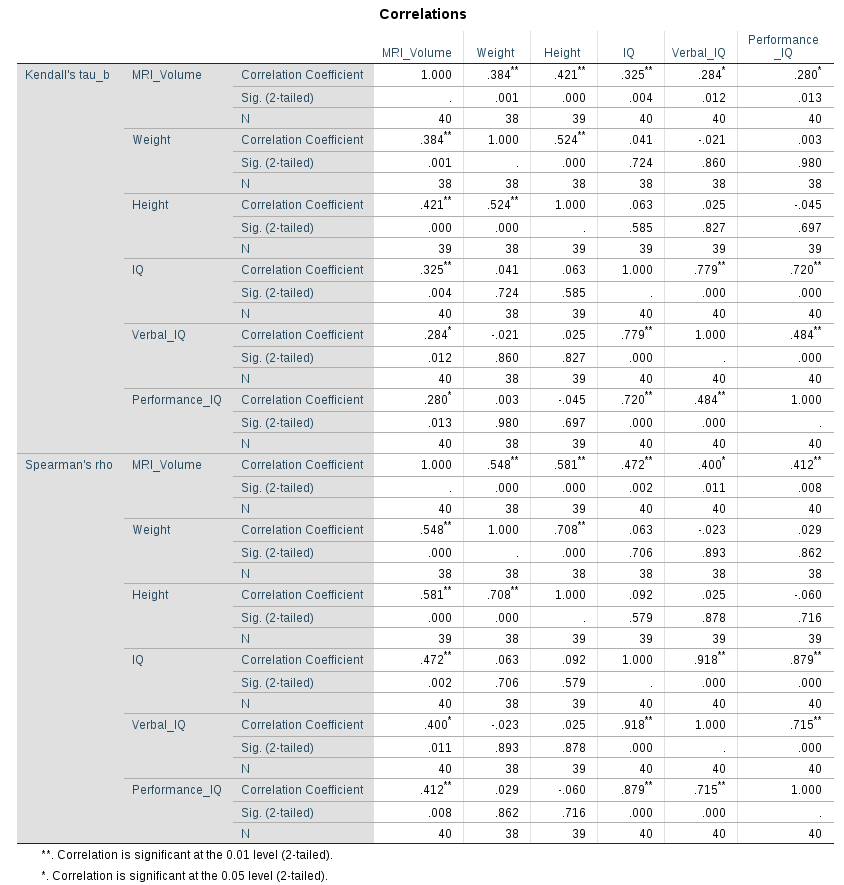
**Ayse Ozdemir – e2340008**



Parametrik correlation coefficients:

Pearson correlation coefficient values can be read from the table below:



  
  
  
  
  
  
  
  
  
  
  
  
Non-parametrik correlation coefficients:  
  
Kendall‘s tau\_b and spearman‘s rho values can be read from the table below:

* 1. **IQ:** Interval

**Verbal\_IQ:** Interval

**Performance\_IQ:** Interval

**MRI\_Volume:** Interval

**Height:** Interval

**Weight:** Interval

**Pearson‘s correlation** requires that data are interval. As all the variables are Interval, we can use pearson‘s correlation for variables.

MRI\_Volume and IQ has a positive weak correlation (.358)  
MRI\_Volume and Verbal\_IQ has a positive weak correlation (.337)

MRI\_Volume and Performance\_IQ has a positive weak correlation (.387)   
  
Weight and IQ has a negative weak correlation (-.051)  
Weight and Verbal\_IQ has a negative weak correlation (-.076)

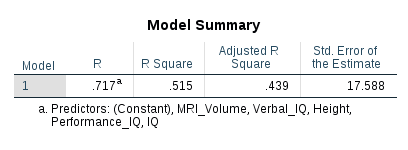
Weight and Performance\_IQ has a negative weak correlation (.003)

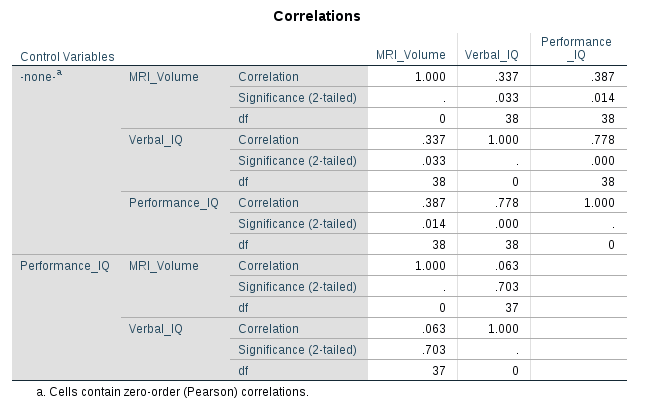
Height and IQ has a negative weak correlation (-.086)  
Height and Verbal\_IQ has a negative weak correlation (-.071)

Height and Performance\_IQ has a negative weak correlation (-.077)

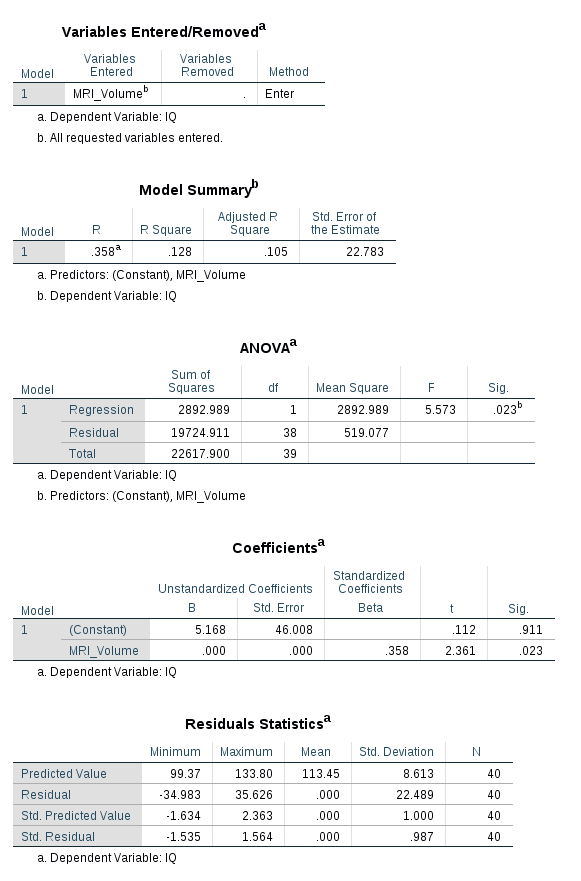
* 1. The correlation coefficient squared (known as the coefficient of determination, R2) is a measure of the amount of variability in one variable that is shared by the other.   
       
     (.358)2 of the variability in IQ is shared by MRI\_Volume.  
     (.337)2 of the variability in Verbal\_IQ is shared by MRI\_Volume.  
     (.387)2 of the variability in Performance\_IQ is shared by MRI\_Volume.

(-.051)2 of the variability in IQ is shared by Weight.  
(-.076)2 of the variability in Verbal\_IQ is shared by Weight.  
( .003)2 of the variability in Performance\_IQ is shared by Weight.  
  
(-.086)2 of the variability in IQ is shared by Height.   
(-.071)2 of the variability in Verbal\_IQ is shared by Height.  
(-.077)2 of the variability in Performance\_IQ is shared by Height.

* 1.   
     We conduct a partial correlation analysis on SPSS.

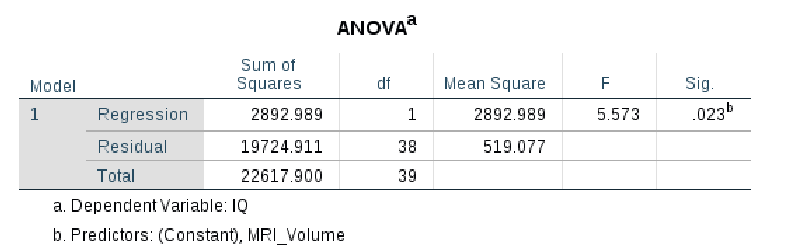


The results of the partial correlation is shown in the table above. In the second part of  
 the table which is labeled with Performance\_IQ, shows that there is a **statistically not significant, weak, positive partial correlation between the MRI\_Volume and Verbal\_IQ**, whilst controlling for Performance\_IQ.  **(r(37) = .063, n=40, p=.703)**.   
  
 However, when we refer to the Pearson's product-moment correlation – also known as the zero-order correlation – between "MRI\_Volume" and "Verbal\_IQ", *without controlling for "Performance\_IQ*", we can see that there is a statistically significant, moderate, positive correlation between "MRI\_Volume" and "Verbal\_IQ" (r(38) = -.337, n = 40, p = . 033).   
  
 This suggests that "Performance\_IQ" **has an influence** in controlling for the relationship between "MRI\_Volume" and "Verbal\_IQ".

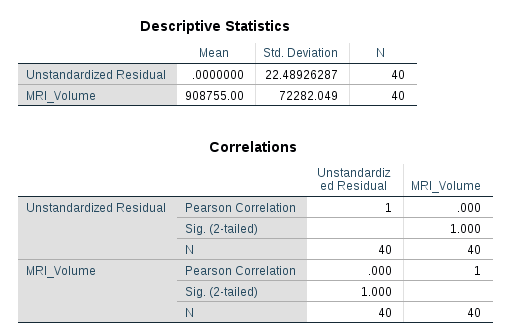
1. 
2. When we look at the Model Summary table:

**R Field:** Correlation between MRI-Volume and IQ

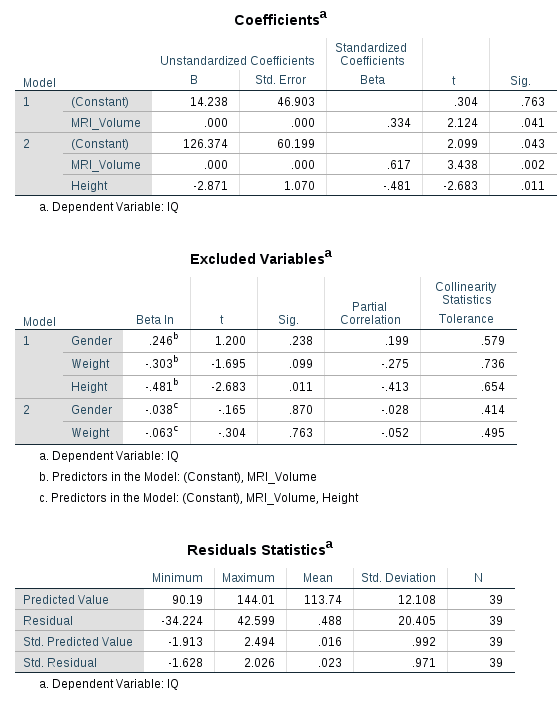
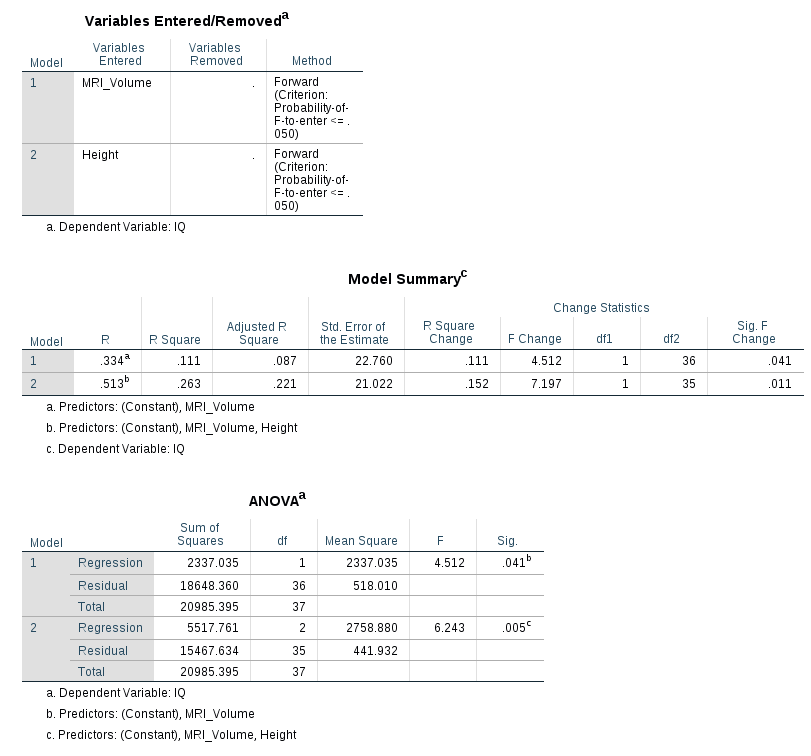
**R Square Field:** The amount of variance in the IQ (criterion-DV) that is explained by the MRI\_Volume(predictor-IV)   
  
**MRI\_Volume explains 13% of the variance in IQ.**

1.   
   SSM = total variability around the mean  
   SSResidual = The sum of squared errors in prediction  
   SSRegression = The improvement in prediction by using the predicted value of Y over just using the mean of Y. (SSTotal – SSResidual)  
     
     
   SSM = 2892.989  
   SSR = 19724.911  
   SST = 22617.900  
     
     
     
     
     
   As SSM value is comparatively small, using the regression model is better than using the mean. In short this model can be counted as a good fit.  
   We can also check F-ratio as it tells us how much variability the model can explain relative to how much it can’t explain. Here F-ratio is 5.573, so the model is a good fit.
2. Y (IQ) = ( 5.17 + 0.000119 \* (MRI\_Volume) ) + εi
3. Y(IQ) = (5.17 + 0.000119 \* (924059)) + εi

Y(IQ) = 115,133021 + εi  
  
Residual term, εi, which represents the difference between the score predicted by the line and the score that is actually obtained.  
  
  
What is the residual (i.e. error)?

1. 

Correlation coefficient = -4.958E-17  
  
Unstandardized Residual and MRI\_Volume has a negative weak correlation(-4.958E-17)  
  
MRI\_Volume does not have a strong affect on residuals.

1.   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   * 1. At the beginning, model‘s start is empty. No variables loaded into the model.

The predictor which has the lowest p-value entered into the model first. At the next steps, the predictor with the next lowest p-value is loaded in, and this process continues one variable at a time being added in as long as the p-value continues to meet a specified criterion.  
  
 It took 2 steps to stop. In the first step MRI\_Volume added into the model, in the second step Height also added into the model.

In the first step MRI\_Volume added into the model, as it has the lowest p-value.  
 In the second step Height added into the model, as it has the second lowes p-value.

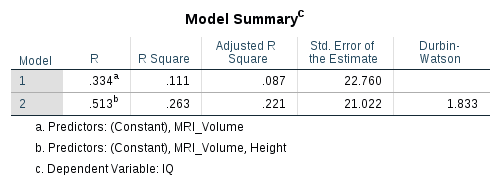
Gender and weight excluded from the model as they aren‘t statistically significant(p>0.05).  
  
 From the model summary table(column R square):  
 Both models have statistical significance as p < 0.05  
 Adjusted R Square for model 1 is 0.87 so **87%** of the variance in the dependent variable and the outcome variable is explained just **by MRI\_Volume**.

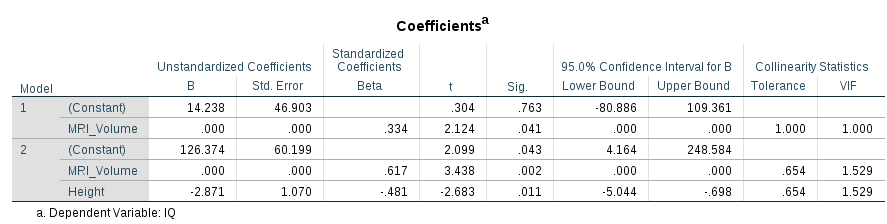
Adjusted R Square for model 2 is 0.263 so **26%** of the variance in the dependent variable and the outcome variable is explained **by MRI\_Volume and Height together**.

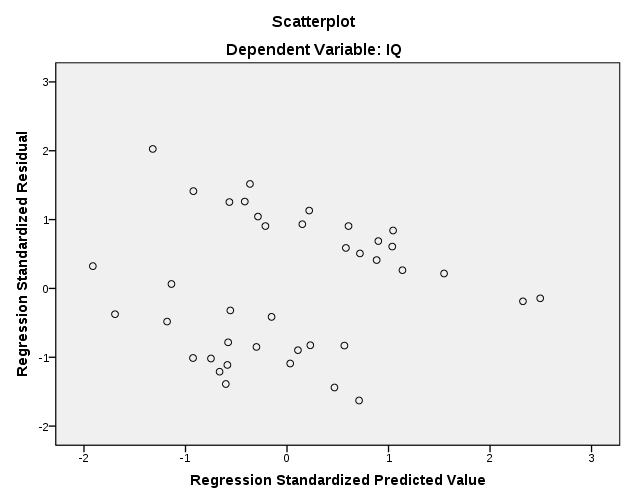
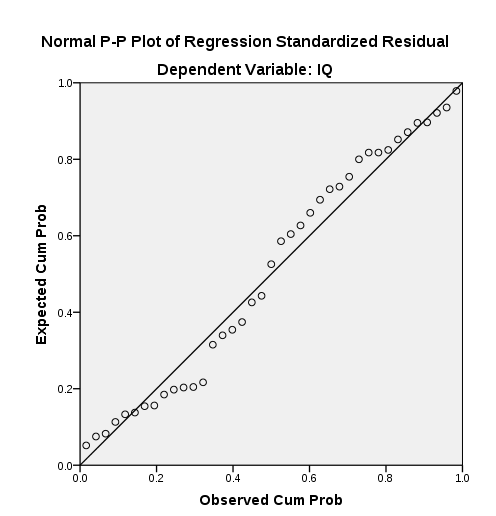
* + 1. MRI\_Volume and Height preserved in the final model.   
       Gender and weight excluded from the model as they aren‘t statistically significant(p>0.05)  
         
       MRI\_Volume (p = .002) and Height (p = .011) are significant predictors.  
         
       Height it the stronger predictor as; height increases by 1, IQ increases by 2,871 units.  
         
       Relation to the outcome variable:  
       Height increases by 1, IQ increases by 2,871 units  
       MRI\_Volume increases by 1, IQ increases by 0.000203 units.
    2. **Y (IQ) = ( 126,374 + 0.000203 \* (MRI\_Volume) – 2,871 \* (Height) ) + εi**
    3. Y (IQ) = ( 126,374 + 0,000203 \* (854258) – 2,871 \* (175,66) ) + εi   
       Y (IQ) = 126,374 + 173,414374 – 504,31986 + εi

Y (IQ) = -204.531486 + εi

What is the residual for this prediction?

* + 1.   
       **Independence of Errors:**  
         
       Normality holds if the errors are independent(uncorrelated). To check if errors are independent, we can conduct Durbin-Watson test on SPSS.  
         
         
         
         
         
         
         
         
       The Durbin-Watson d = 1.833, which is between the two critical values of 1.5 < d < 2.5. Therefore, we can assume that there is no first order linear auto-correlation in our multiple linear regression data.  
         
         
         
       **Multicollinearity:**The information in the table below also allows us to check for multicollinearity in our multiple linear regression model. Tolerance should be > 0.1 (or VIF < 10) for all variables, which they are. Multicollinearity assumption is met.



  
  
  
  
  
**Normality of Residuals:**  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
We can check for normality of residuals with a normal P-P plot. The plot shows that the points generally follow the normal (diagonal) line with no strong deviations. This indicates that the residuals are normally distributed.  
  
**Homoscedasticity:**  
  
  
As the data does not have an obvious pattern, there are points equally distributed above and below zero on the X axis, and to the left and right of zero on the Y axis. As a result; assumption of homoscedasticity is met.  
  
  
  
  
Finally, assumptions of multiple regression is met as Independence of Errors, Multicollinearity, Normality of Residuals, and Homoscedasticity is met.