**Question 8.11**

In order to determine which variables most closely predict the height of the oldest child, we run several versions of the stepwise regression. To gain an overview of the different combinatorics, we run a best subset regression and view all the possible permutations of variables that may determine the height of the oldest child. Although the result is that it reaches the highest adjusted R2 when all of them are included, the option of using them is tempered with the decision to avoid using a large amount of variables, minimizing the CP, and AIC, both which are minimized when the fifth variable is added.

Running a forward stepwise regression we find that the order of highest correlation to the height of the oldest child is their age, then their weight, height of the father, then height of the mother (which slightly supersedes the height of the father at that point), and then weight of the mother. The weight of the father does not reach the minimum F – to - Enter, and so it is not added.

Running a backward stepwise results in removing only the weight of the father before the F value reaches the maximum F – to - remove.

Lastly, we run a general stepwise analysis, which mimics the forward selection procedure and does not remove any variables during the process. The results of the process are shown below:

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | Age\_oldest\_child |  | 1 | 0.8304 | 0.8304 | 159.951 | 724.61 | <.0001 |
| **2** | Weight\_oldest\_child |  | 2 | 0.0614 | 0.8918 | 51.1454 | 83.47 | <.0001 |
| **3** | Height\_father |  | 3 | 0.0170 | 0.9088 | 22.5495 | 27.15 | <.0001 |
| **4** | Height\_mother |  | 4 | 0.0084 | 0.9172 | 9.3150 | 14.79 | 0.0002 |
| **5** | Weight\_mother |  | 5 | 0.0028 | 0.9201 | 6.1778 | 5.13 | 0.0250 |

Our final regression line (eliminating the weight of the father) is:

Height\_oldest\_child = 0.52776 + (1.21129\* Age\_oldest\_child) + (0.07769\* Weight\_oldest\_child) + (0.32174\* Height\_mother) + (-0.01282\* Weight\_mother) + (0.25345\* Height\_father)

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 0.52776 | 5.36450 | 0.03841 | 0.01 | 0.9218 |
| **Age\_oldest\_child** | 1.21129 | 0.09720 | 616.35434 | 155.30 | <.0001 |
| **Weight\_oldest\_child** | 0.07769 | 0.00825 | 351.65930 | 88.60 | <.0001 |
| **Height\_mother** | 0.32174 | 0.07276 | 77.59926 | 19.55 | <.0001 |
| **Weight\_mother** | -0.01282 | 0.00566 | 20.36365 | 5.13 | 0.0250 |
| **Height\_father** | 0.25345 | 0.06237 | 65.53135 | 16.51 | <.0001 |

**Question 8.12**

**We find that although the predictor variables remain the same in all cases, they are picked at different stages and hold different weights.**

OCSEX = ‘Male’

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | Age\_oldest\_child |  | 1 | 0.8661 | 0.8661 | 71.6892 | 472.11 | <.0001 |
| **2** | Weight\_oldest\_child |  | 2 | 0.0554 | 0.9215 | 14.6344 | 50.84 | <.0001 |
| **3** | Height\_father |  | 3 | 0.0101 | 0.9316 | 5.8851 | 10.47 | 0.0018 |

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 17.16893 | 6.01483 | 30.91167 | 8.15 | 0.0056 |
| **Age\_oldest\_child** | 1.35946 | 0.13828 | 366.70633 | 96.66 | <.0001 |
| **Weight\_oldest\_child** | 0.07105 | 0.01047 | 174.66775 | 46.04 | <.0001 |
| **Height\_father** | 0.27090 | 0.08372 | 39.72657 | 10.47 | 0.0018 |

Regression Line:

Height\_oldest\_child (1) = 17.16893 + (Age\_oldest\_child\*1.35946) + (0.07105\*Weight\_oldest\_child) + (0.27090\*Height\_father)

OCSEX = ‘Female’

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | Age\_oldest\_child |  | 1 | 0.8129 | 0.8129 | 70.2380 | 317.08 | <.0001 |
| **2** | Height\_father |  | 2 | 0.0388 | 0.8517 | 42.9286 | 18.85 | <.0001 |
| **3** | Weight\_oldest\_child |  | 3 | 0.0274 | 0.8791 | 24.2478 | 16.09 | 0.0001 |

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 11.23349 | 6.55727 | 14.32502 | 2.93 | 0.0911 |
| **Age\_oldest\_child** | 1.21067 | 0.15771 | 287.62998 | 58.93 | <.0001 |
| **Weight\_oldest\_child** | 0.06347 | 0.01582 | 78.54440 | 16.09 | 0.0001 |
| **Height\_father** | 0.38494 | 0.09345 | 82.81388 | 16.97 | 0.0001 |

Regression Line:

Height\_oldest\_child (2) = 11.23349 + (1.21067\* Age\_oldest\_child) + (0.06347\* Weight\_oldest\_child) + (0.38494\* Height\_father)

Both sexes

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | Age\_oldest\_child |  | 1 | 0.8304 | 0.8304 | 159.951 | 724.61 | <.0001 |
| **2** | Weight\_oldest\_child |  | 2 | 0.0614 | 0.8918 | 51.1454 | 83.47 | <.0001 |
| **3** | Height\_father |  | 3 | 0.0170 | 0.9088 | 22.5495 | 27.15 | <.0001 |

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 14.03591 | 4.50096 | 43.44601 | 9.72 | 0.0022 |
| **Age\_oldest\_child** | 1.21966 | 0.10259 | 631.52269 | 141.35 | <.0001 |
| **Weight\_oldest\_child** | 0.07545 | 0.00864 | 340.98721 | 76.32 | <.0001 |
| **Height\_father** | 0.33073 | 0.06348 | 121.28277 | 27.15 | <.0001 |

Height\_oldest\_child = 14.03591+ (1.21966\* Age\_oldest\_child) + (0.07545\* Weight\_oldest\_child) + (0.33073\* Height\_father)

**However, adding in the sex of the oldest child does not affect the regression line.**

With Adding In OCSEX

| **Summary of Stepwise Selection** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Variable Removed** | **Number Vars In** | **Partial R-Square** | **Model R-Square** | **C(p)** | **F Value** | **Pr > F** |
| **1** | OCAGE |  | 1 | 0.8304 | 0.8304 | 169.023 | 724.61 | <.0001 |
| **2** | OCWEIGHT |  | 2 | 0.0614 | 0.8918 | 56.9316 | 83.47 | <.0001 |
| **3** | FHEIGHT |  | 3 | 0.0170 | 0.9088 | 27.4285 | 27.15 | <.0001 |

| **Variable** | **Parameter Estimate** | **Standard Error** | **Type II SS** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 14.03591 | 4.50096 | 43.44601 | 9.72 | 0.0022 |
| **OCAGE** | 1.21966 | 0.10259 | 631.52269 | 141.35 | <.0001 |
| **OCWEIGHT** | 0.07545 | 0.00864 | 340.98721 | 76.32 | <.0001 |
| **FHEIGHT** | 0.33073 | 0.06348 | 121.28277 | 27.15 | <.0001 |

Height\_oldest\_child = 14.03591+ (1.21966\* Age\_oldest\_child) + (0.07545\* Weight\_oldest\_child) + (0.33073\* Height\_father)