**Exploratory Data Analysis – Selby and Travis**

**Normality checks (Shapiro Wilk’s test):**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Statistic** | **P-value** |
| MilkAUC | 0.991244 | 0.6610 |
| precalveBW | 0.992768 | 0.8004 |
| precalveBCS | 0.888353 | <0.0001 |
| prebreedBW | 0.990196 | 0.5644 |
| prebreedBCS | 0.944617 | 0.0001 |
| breedBW | 0.984093 | 0.1785 |
| breedBCS | 0.954406 | 0.0005 |
| weanBW | 0.988029 | 9.3879 |
| weanBCS | 0.892917 | <0.0001 |
| prebreedBWchange | 0.953913 | 0.0005 |
| breedBWchange | 0.984855 | 0.2086 |
| weanBWchange | 0.982181 | 0.1200 |
| Calf30 | 0.956514 | 0.0008 |
| Calf60 | 0.967794 | 0.0062 |
| Calf90 | 0.988089 | 0.3922 |
| Calf120 | 0.983767 | 0.1669 |
| Calfwean | 0.979862 | 0.0737 |

What I highlighted above indicates that these variables violate normality assumption (p-value<0.05), we may need to use a different distribution for these variables.

**Scatter matrix for body weight measures (BW):**

**Diagram, engineering drawing

Description automatically generated**

Body weight at different stages of cow life are positively correlated. Appears to be no association between the milkAUC and the body weights, same conclusion as in Table 2 in the grant.

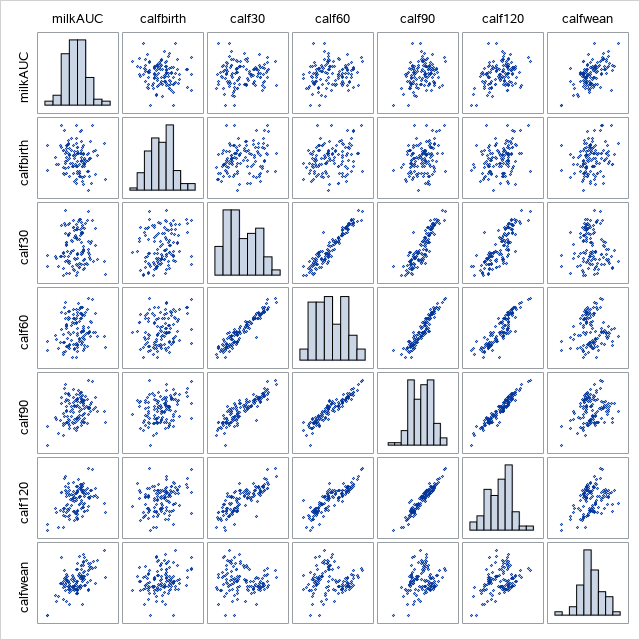
**Scatter matrix for body condition scores (BCS)**

**A picture containing text, crossword puzzle

Description automatically generated**

Quite interesting patterns in these scatter plots. There appears to be no association between the milkAUC and the body condition scores.

**Scatter matrix for calf weights at all ages:**

****

Some correlation between the calf weight measures days, will need to be careful about this when fitting appropriate model. There appears to be a positive association between the milkAUC and the calf weight, bot this positive association doesn’t show until about 90 days after birth.

**Something to note:**

| **preg** | | | | |
| --- | --- | --- | --- | --- |
| **preg** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **0** | 10 | 8.47 | 10 | 8.47 |
| **1** | 108 | 91.53 | 118 | 100.00 |

91.53% of the cows are pregnant. Will we be using this preg variable in the model?

**Things done so far**

* This exploratory data analysis, other things to explore?
* Computing wilk’s lambda?
* Date from birth column created
* Tried some models, but running into singularity issues

**Questions**

* What’s the AUC calculation?
* Using the “change” variables or not? Other variables we aren’t using?
* Types of models? Multivariate multiple regressions? Organizing these….