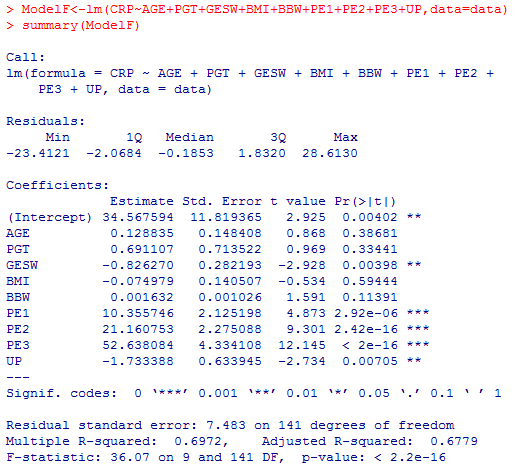
**Regression Model**

1. **Full Model – with normal people (PE=0) as the base level**
2. **Check Criteria for Selecting Model:**

(a) No big difference for R2 between full and reduced models. (69.72% vs. 68.73%)

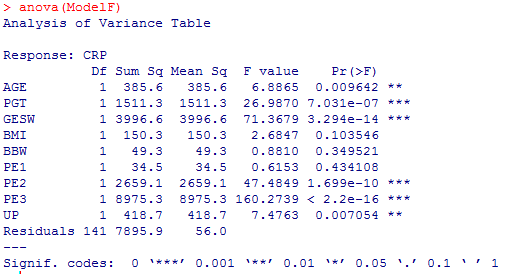
(b)

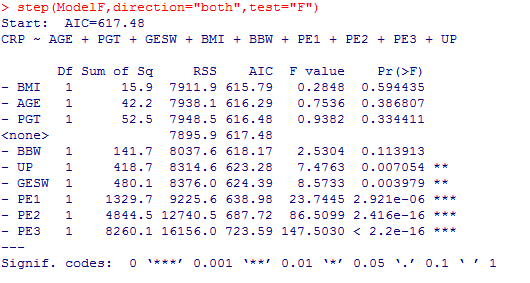
Not significant, retain p variables

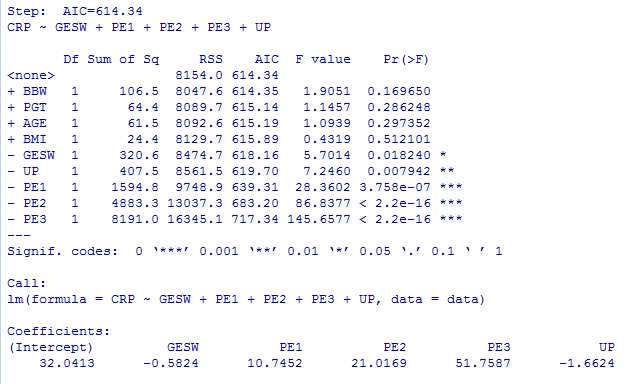
(c) MSE is 56.2344.

(d) AIC: For full model, the AIC is 617.48. The best (smallest) AIC we could get is 614.34, for model = β0 + β1 \* GESW + β2 \* PE1 + β3 \* PE2 + β4 \* PE3 + β5\* UP + ε

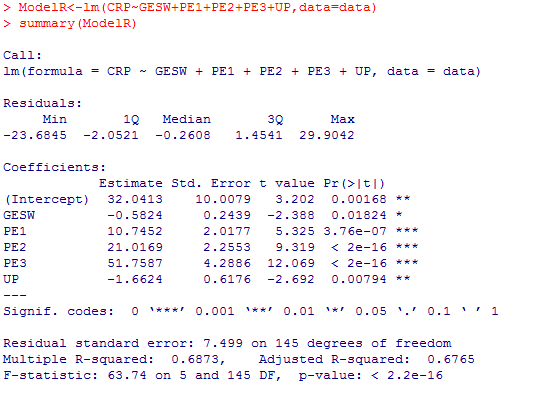
(With normal people (PE = 0) as the base level).



1.  **Stepwise Strategy**

…stepwise steps…

1. **Conducting the Analysis**

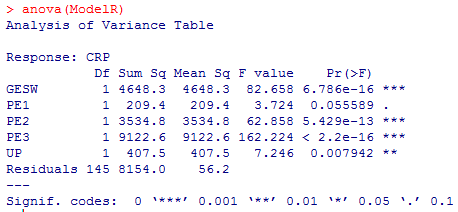


In general, the explanatory variables GESW, PE and UP are more related to the response variable CRP. The correlation is higher and the model is much simpler.

The best reduced multiple linear regression model is:

= 32.0413 + (-0.5824) \* GESW + 10.7452 \* PE1 + 21.0169 \* PE2 + 51.7587 \* PE3 + (-1.6624) \* UP

(With normal people (PE = 0) as the base level).



The coefficient of Determination R2 is 68.73%. According to the reduced linear model, 68.73% of the variability in CRP is accounted for by variation in GESW, PE and RP.

