ANOVA/Regression I
Solutions to Session 2 : Multiple Regression

1.

Model	Variables Used	SSR	d.f.	SSE	d.f.	Overall F	p	R^2
1	HGT	588.9225	1	299.3275	10	19.67	0.0013	.6630
2	AGE	526.3929	1	361.8571	10	14.55	0.0034	.5926
3	$(AGE)^2$	521.9320	1	366.3180	10	14.25	0.0036	.5876
4	HGT, AGE	692.8226	2	195.4274	9	15.95	0.0011	.7800
5	HGT, $(AGE)^2$	689.6500	2	198.6000	9	15.63	0.0012	.7764
6	HGT, AGE, $(AGE)^2$	693.0605	3	195.1895	8	9.47	0.0052	.7803

2.
F(AGE | HGT)= (692.8226-588.9225)/(195.4274/9)
= (299.3275-195.4274)/(195.4274/9)=4.78

If you look at the **`anova, sequential`** output of **model 4** and look at the *age* row you will see the F-test =4.78 and a corresponding p-value = 0.0565. Notice that the t-test p-value for β_2 (the regression coefficient associated with age, is 0.056, and $T^2=(2.187)^2=4.78=F$. Since the p-value of F(AGE|HGT) is 0.056 which is greater than a=0.05, the addition of AGE doesn't significantly contribute to the model.

3. To answer the same question about $(AGE)^2$ after controlling both height and age, the *partial Type I F-test* $F((AGE)^2|HGT,AGE)=0.01$ with corresponding p-value=0.924 which is not significant. Thus, even though $(AGE)^2$ was significant as a single predictor of weight, it is not significant after controlling for height and age. So, a quadratic relationship between weight and age is probably not born out by the data.

4.
$SS((AGE)^2 \mid HGT, AGE) = 0.24$ from model 6
$SS(AGE \mid HGT,\{AGE)^2) = 3.41$ from model 7
$SS(HGT \mid AGE,(AGE)^2) = 166.58$ from model 8

5.
$F((AGE)^2 \mid HGT,AGE) = 0.24/(195.19/8) = 0.01$ p-value = 0.924
$F(AGE \mid HGT,(AGE)^2) = 3.41/(195.19/8) = 0.14$ p-value = 0.718
$F(HGT \mid AGE,(AGE)^2) = 166.58/(195.19/8) = 6.83$ p-value = 0.031

6. According the above results, *HGT* seems to be the most important predictor of weight since even after controlling for both *AGE* and *$(AGE)^2$* the *partial Type III F-test* is significant (0.031 < a=0.05).