13.69 a. The model F-test has value 6.52 with p-value=0.0004. Thus, we can reject $H_o : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ and conclude that at least one of $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 \neq 0$. There is significant evidence that the set of six independent variables does provide an explaination of the variability in PULSE.

c. The coefficient of the variable PHYS1 measures the difference in average pulse between individuals who have substantial physical exercise and those who exercise physical. The estimate of this parameter is 13.43 which indicates individuals who exercise substantially have a mean increase in pulse rate of 13.43 units higher than those who exercise very little.

A 95% C.I. for the coefficient of PHYS1 is $13.43 \pm (t_{.025,23})(SE(\beta)) \Rightarrow$

$13.43 \pm (2.069)(4.2512) \Rightarrow (4.63, 22.23)$

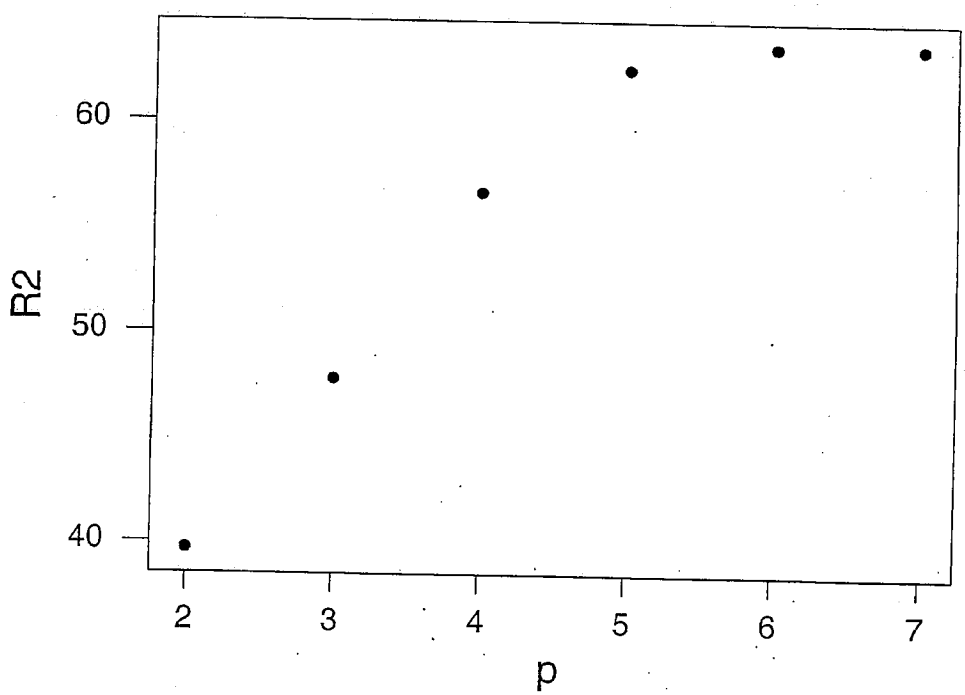
Thus, we are 95% confident that the actual average difference in the increase in pulse between the two groups is greater than zero.

13.70 a. The independent variables have values: SMOKE=0, RUN=1, PHY1=0, PHY2=0. Thus, the estimated model is

$\hat{y} = -31.688 + 11.402 + 0.1317$ HEIGHT $+ 0.02304$ WEIGHT $\Rightarrow$

$\hat{y} = -20.286 + 0.1317$ HEIGHT $+ 0.02304$ WEIGHT

3. plot of R^2 vs. p (= # of parameters in the model)



Problem #3:

Best Subsets Regression: PULSE versus RUN, SMOKE, ...

Response is PULSE

Vars	R-Sq	R-Sq(adj)	C-p	S	RUN	SMOKE	HEIGHT	WEIGHT	PHYS1	PHYS2
1	39.7	37.5	12.1	7.9557	X					
1	9.4	6.2	31.3	9.7493		X				
2	47.8	44.0	9.0	7.5343	X				X	
2	45.9	41.9	10.2	7.6760	X	X				
3	56.6	51.6	5.4	6.9999	X	X			X	
3	52.1	46.5	8.3	7.3603	X				X	X
4	62.5	56.6	3.7	6.6350	X	X			X	X
4	57.1	50.2	7.1	7.1006	X	X		X	X	
5	63.6	56.0	5.0	6.6736	X	X		X	X	X
5	62.9	55.1	5.5	6.7421	X	X	X		X	X
6	63.6	54.1	7.0	6.8169	X	X	X	X	X	X

Based on the R-Sq values (also see the plot of R-Sq vs. p, the number of parameters in the model), I would choose the model with 4 predictors (p=5, R-Sq=62.5) since the R-Sq values for models with 5 or 6 predictors are not much larger than the model with 4 predictors. Other reasons that one should choose this model include that this model has the largest R-Sq adjusted, the smallest residual standard deviation S and that weight is not a significant predictor when one fits the model with R-Sq=63.6.

Problem #4:

The model picked in (3) includes 4 predictors: run, smoke, phys1 and phys2. There are 5 possible interactions for this model:
run*smoke, run*phys1, run*phys2, smoke*phys1, smoke*phys2
(There is no interaction between phys1 and phys2 because there are 3 levels of physical activities and there can be at most 2 dummy varibles based on physical activities.)

I ran a backward model selection with run, smoke, phys1 and phys2 in every model and 9 predictors (4 main effects plus 5 interactions). At the 0.05 significance level, no interaction term was found to be significant.