See R code attached for calculations

Question 1

a)

Null Hypothesis: $B_1 = 0$
Alternative Hypothesis: $B_1 =/= 0$

Partial F-Stat = 47.02891, the p-value is 0.00000. Therefore, at this p-value, x_1 contributes significantly to the prediction of Y.

Null Hypothesis: $B_2 = 0$
Alternative Hypothesis: $B_2 =/= 0$

Partial F - Stat = 32.79028
P-Value : 0.00000
At a p-value of 0, X_2 also contributes significantly to the prediction of Y.

b)

F- Stat = 126.4159
P-value = 0.0000
At a p-value of zero, the independent variables also significantly contribute to the prediction of Y.

c)

The two models being compared are

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + E$$

and

$$Y = B_0 + E$$

This is because of the variables used in the formula. x1,x2,x3 are used in the numerator.

Question 2:

Perform the overall F test for the regression of Y on both independent variables. Interpret your results:

H_0: $B_1 = B_2 = 0$

H_a: $B_1 =/= B_2 =/= 0$

F-Statistic: 20.03

P-value: 0.008

Since the p-value is less than 0.05, we fail to reject the null hypothesis.

Preform Variables-added-in-order-tests for both independent variables, with X1 first

For X_1:

H_o: B1 = 0

H_a: B1 =/= 0

F-Statistic: 40.06

P-Value: 0.0352

Since the p-value is less than 0.05, we uphold the null hypothesis. So, X_1 is a significant variable.

For X_2:

H_o: B2 = 0

H_a: B2 =/= 0

F-Statistic: 0.01

P-Value 0.9344

Since the p-value is more than 0.05, then we reject the null hypothesis.

So X_2 is not a significant variable.

Preform Variables-added-in-order-tests for both independent variables, with X1 first

Looking at the associated computer output, provide a table of variables-added-last- tests:

Variation	df	SS	MS	F	R^2
$X_1\|X_2$	1	1402.32	1402.32	9.80	0.9092
$X_1\|X_2$	1	1.098	1.098	0.01	
Residual	4	572.40	143.10		

Which predictors appear to be necessary? Why?

Question 3

a)

Since AGE has the highest R^2 value at 0.7752 in the correlation matrix, it has the strongest linear relationship with the independent variable. In other words, it explains the largest variation in SBP.

b)

$r^2_{SBP, SMK \mid AGE} = 0.33229$

Taking the square root yields the partial correlation coefficient as

$r_{SBP, SMK \mid AGE} = 0.5682429.$

$r^2_{SBP, QUET \mid AGE} = 0.10099$

Again, taking the square root yields:

$r_{SBP, QUET \mid AGE} = 0.3177892$

c)

Null Hypothesis: $P_{SBP, SMK} = 0$

Alternative Hypothesis: $P_{SBP, SMK \mid Age} =/= 0$

SS = 4689.684229 - 3861.6304

The residual MS = 59.871880

F-val = SS/ ResMS = 13.83043

p-value of 0.0008

F-Critical value = 4.18

Since the f-value is greater than the f-critical value, the null hypothesis is rejected.

d)

Null: $P_{SBP, QUET \mid AGE,SMK} = 0$

Alternative: $P_{SBP, QUET \mid AGE, SMK} =/= 0$

T = 1.91. In referencing a t-table, $t_{28, 0.975} = 2.048$. Since 1.91 < 2.048, we do not reject the null hypothesis. Therefore, the difference between the partial correlations is not significant.

e)

1. AGE
2. SMK
3. QUET

Based on the tests done throughout the problem, QUET would become the least important.

f)

Null Hypothesis: B11 = B 12 = B22 = 0
Alternative Hypothesis: B11 =/= B12 =/= B22 =/= 0

$R^2_{SBP \mid AGE} = 0.601$
$R^2_{SBP \mid AGE,SMK, QUET} = 0.761$

$R^2_{SBP(QUET,SMK)|AGE} = 0.4010025$

The F-value is 9.370691; the p-value is 0.0008. However, in calculating the F-Critical value in order to further check our results, we find that it is equal to 3.340386. Since the F-value is more than this critical value, we would reject the null hypothesis.

Question 4:

a) The calculated F-Statistic is 5.327738. The p-value for the F-statistic is 0.03180264. However, the F-Critical value is 4.35. Even though the p-value is less than a significance level of 0.05, the F-statistic is greater than the F-critical value. Therefore, we reject the null hypothesis.

b) The calculated F-Statistic is 4.495102. The p-value for this statistic is 0.04738016. However, since the F-critical value is 4.38, and the F-Value is 4.49, we would still reject the null hypothesis.

c) The F-value for this test is 5.3769, its p-value is 0.1353005, and the F-critical value at the 0.05 significance level is 3.492828. Since the f-value is more than this critical value, we would reject this null hypothesis.

d) Based on the tests done above, X1, X2, X3 would be the most important predictors.

Homework 4

Jacob Plaza

3/23/2022

''' # #QUESTION 1 #

#A #Null Hypothesis: B1 = 0 –> #Alternative Hypothesis: B1 =/= 0
''' MS.X1 = (7010.03 + 10.93 + 2248.23)/(1 + 1 + 21) MS.X1

#calculating the residual ms

partial_f_1 = 18953.04/MS.X1 partial_f_1 #dividing the X1 SS by the residual MS to get the partial F - stat

pval1 = 1-pf (47.02891, 1, 25) pval1 #finding the p-value, which rounds to 0. #At a p-value of zero, X1 contributes significantly to the prediction of Y.

#Null Hypothesis for Second test: B2 = 0 #Alternative Hypothesis for Second Test = B2 =/= 0

numerator = (7010.03 + 10.93)/2 denominator = (2248.23/21) partial_f_2 = numerator/denominator partial_f_2

pval2 = 1- pf(32.79028, 2, 21) pval2 #At a p-value of 0, X2 also contributes significantly to the prediction of Y.

#B

num = (18953.04 + 7010.03)/2 denom = (2248.23 + 10.93)/22 F_stat = num/denom F_stat

p_value = 1 - pf(126.4169, 2, 22) p_value

#Y = B0 + B1X1 + B2X2 + B3X3 + E #and #Y = B0 + E #We know this because x1,x2,x3 are used in the numerator, and the denominator is the #MSE

#QUESTION 2 # r_sq = (6305.7142 - 572.3929)/6305.7142 r_sq

#QUESTION 3 #

#b)

sqrt(0.3229) #partial correlation coefficient r_SBP,SMK|AGE = 0.5682429

#r^2_sbp,quet|AGE = 0.10099 sqrt(0.10099) #partial correlation coefficient r^2sbp,quet|AGE = 0.3177892

#c)

SS = 4689.684229 - 3861.6304 Res_MS = 59.871880 SS/Res_MS

p_val_c = 1 - pf(13.83043, 1, 29) p_val_c

f_crit_3c = qf (0.95,1, 29) f_crit_3c #f) r2_sbp_on_age = (6425.9687 - 2564.3383)/6425.9786 r2_sbp_on_age

r2sasq = (6425.9687 - 1536.1430)/6435.9687 r2sasq

partial_correlation = (0.761 - 0.601)/(1- 0.601) partial_correlation

FVAL = ((4889.8257 - 3861.6304)/2)/(54.8623) FVAL p_val_f = 1 - pf(9.370691, 2, 28) p_val_f F_CRIT_F = qf(0.95, 2, 28) F_CRIT_F

#QUESTION 4 #

#a) Res.MS = (130.529+551.723)/20 Res.MS FStat = 181.743/Res.MS FStat 1 - pf(FStat, 1, 20) F_crit_a = qf(0.95, 1, 20) F_crit_a

#b)

ResMS.B = 551.723/19 ResMS.B

F.Val.B = 130.529 / ResMS.B F.Val.B p.val.b = 1 - pf(F.Val.B, 1, 19) p.val.b F_crit_b = qf(0.95,1, 19) F_crit_b

#c) numerator_c = ((1523.658 + 181.743 + 130.529) - 1523.658)/2 denominator_c = (551.723/19)

F_val_4c = numerator_c / denominator_c F_val_4c p_val_4c = 1 - pf(F_val_4c, 2, 20) p_val_4c F_crit_4c = qf(0.95, 2, 20) F_crit_4c