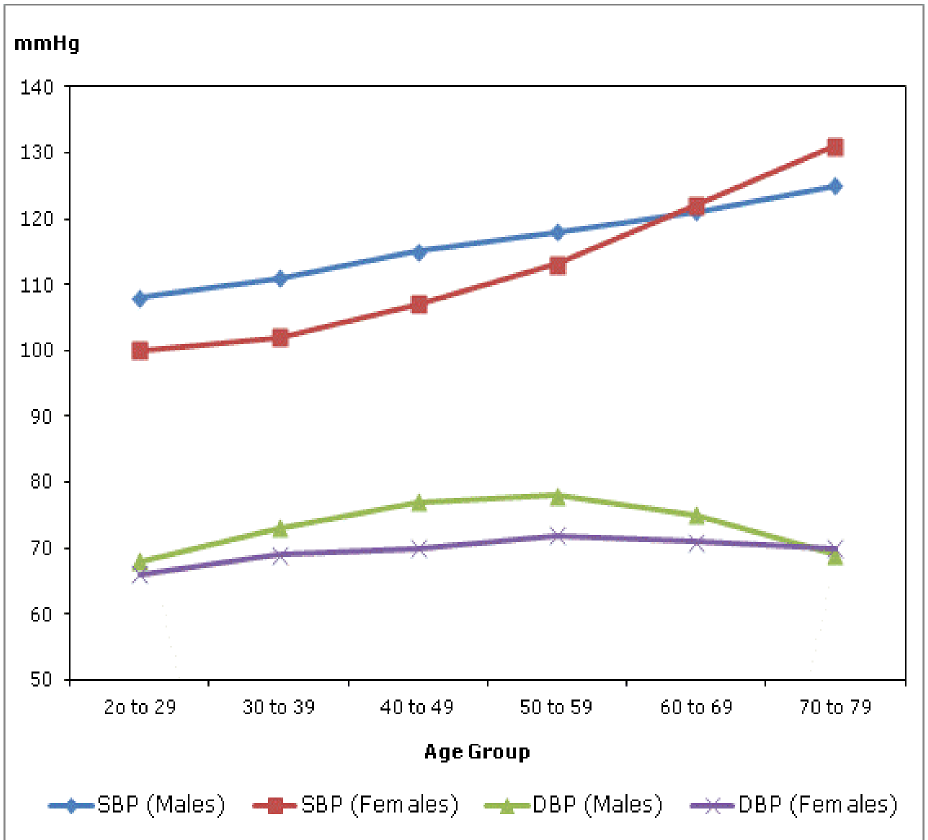
Practice Problem Topic 5: Multiple Linear Regression

Continuing on the Theme of Different Factors/Variables Affecting Blood Pressure

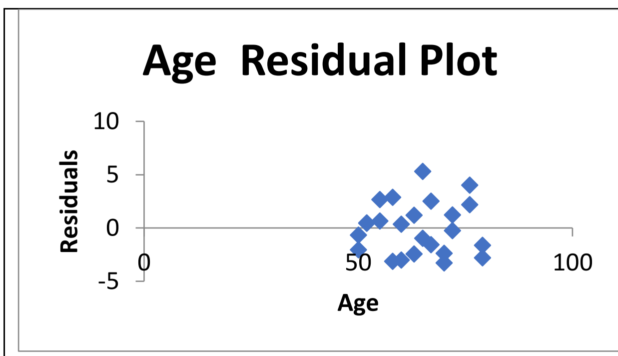
Do age, gender, and interaction (explanatory variables) have an effect on blood pressure (response variable)?

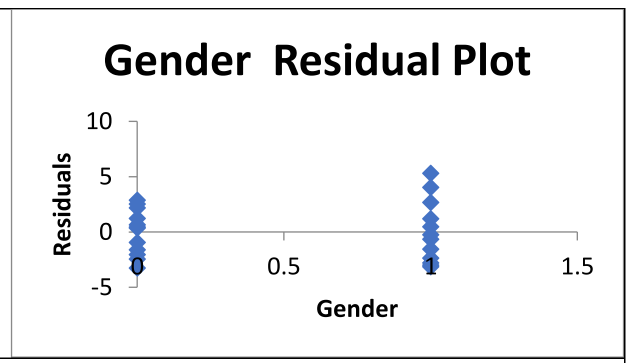
Do age (in years), gender (coded as: gender = 1 for females, and gender = 0 for males), and the interaction between age and gender have an effect on systolic blood pressure (SBP) (in mmHg)? Measurements are based on a random sample of 24 people (12 females and 12 males) between the ages of 50 and 79. Use the analysis shown in the SPSS output below (with some values missing), to answer parts **(a) – (d)**. All assumptions are met for the required analysis. [This is a hypothetical data set, but it is patterned after real data collected by Canadian Health Measures Survey, 2009 to 2011.]

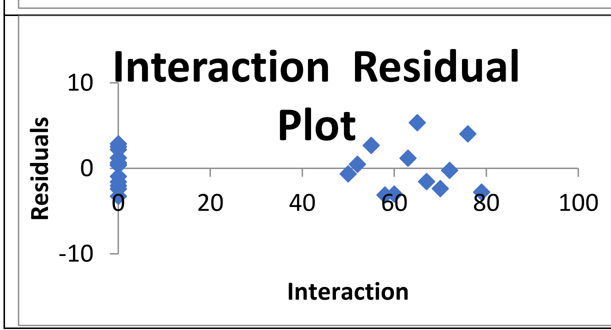
Age (years)	Gender (F=1,M=0)	Interaction	SBP (mm Hg)
50	1	50	107
52	1	52	110
55	1	55	115
58	1	58	112
60	1	60	114
63	1	63	121
65	1	65	127
67	1	67	122
70	1	70	124
72	1	72	128
76	1	76	136
79	1	79	132
50	0	0	114
52	0	0	117
55	0	0	118
58	0	0	121
60	0	0	119
63	0	0	117
65	0	0	119
67	0	0	123
70	0	0	118
72	0	0	123
76	0	0	125
79	0	0	122



SBP = systolic blood pressure; DBP = diastolic blood pressure
Source: Canadian Health Measures Survey, 2009 to 2011. The CHMS collects health information on the Canadian household population aged 3 to 79.







Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.933[a]	.871	.852	2.589
a. Predictors: (Constant), Interaction, Age, Gender				

Model 1:

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	905.244	3	301.748	45.007	<.001[b]
	Residual	134.089	20	6.704		
	Total	1039.333	23			
a. Dependent Variable: SBP						
b. Predictors: (Constant), Interaction, Age, Gender						

Coefficients[a]								
Model		Unstandardized Coefficients		Stand. Coeff.	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	102.987	5.413		19.025	<.001	91.696	114.279
	Age	.261	.084		3.111	.006	.086	.436
	Gender	-42.129	7.655		-5.503	<.001	-58.098	-26.160
	Interaction (Age*gender)	.675	.119		5.688	<.001	.427	.922
a. Dependent Variable: SBP								

(a) Find the standard error of the model.

Standard error of the model is: $\hat{\sigma} = \sqrt{MSE} = \sqrt{\dfrac{SSE}{n-(k+1)}} = \sqrt{\dfrac{134.089}{24-(3+1)}} = \sqrt{6.70445} = 2.58930$

(b) What percentage of variation in blood systolic blood pressure (SBP) is explained by this regression model? (Determine the **adjusted** percentage.)

$$SST = SSR + SSE = 905.244 + 134.089 = 1039.333$$

$$MST = \frac{SST}{n-1} = \frac{1039.333}{24-1} = 45.188391 \qquad MSE = \frac{SSE}{n-(k+1)} = \frac{134.089}{24-(3+1)} = 6.70445$$

$$R^2_{adj} = 1 - \frac{MSE}{MST} = 1 - \frac{6.70445}{45.188391} = 1 - 0.148367 = 0.85163$$

Therefore, 85.2% of the variation in SBP is explained by the regression model.

(c) Find the predicted blood systolic blood pressure (SBP) for an individual who is 60 years old and female. Then, determine a 95% prediction interval for all single observation responses of SBP at the given values of the predictor variables. **Note: For your standard error, you can use an appropriate approximation or calculate it with SE(Fit) = 0.816. Show ALL steps.**

The point estimate for the SBP an individual who is 60 years old and female is:

“Fit” = $\hat{y}_p$ = 102.987 + 0.261(60) – 42.129 (1) + 0.675(60)(1) = 117.018 mmHg

At $df = n-(k+1) = 24-(3+1) = 20$, $t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.086$

The endpoints of the prediction interval are:

$$\hat{y}_p \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}^2 + [SE(Fit)]^2}$$
$$117.018 \pm 2.086 \times \sqrt{(2.58930)^2 + (0.816)^2}$$
$$117.018 \pm 2.086 \times \sqrt{6.70445 + 0.665856}$$
$$117.018 \pm 2.086 \times 2.71483$$
$$117.018 \pm 5.6631$$

(111.355, 122.681) mmHg

We can be 95% confident that the prediction interval for all single observation responses of SBP at the given values of the predictor variables is between 111.36 and 122.68 mmHg.

(d) Suppose the systolic blood pressure (SBP) for an individual who is 60 years old and female (as given in part (c)) was actually 110 mmHg. What is the residual or error of this observation? Is this observed (actual) blood hemoglobin of this woman within the prediction interval you calculated in part (c)? Does it have to be within the prediction interval? Explain.

Residual = actual – predicted = $y_i = \hat{y}_p$ (determined in part (c))

= 110 – 117.018 = -7.018 mmHg

No, the actual value of blood hemoglobin level was NOT within the prediction interval calculated in part (c). However, it does not have to be within the prediction interval because we are only 95% confident that it will be within that interval.

(e) For this part of the question, refer to the SPSS output at the start of this question as well as any of the output below. Note: You will NOT need to use all of the tables. At the 5% significance level, perform the most appropriate test, to determine whether gender and/or the interaction between gender and age have an effect on systolic blood pressure (SBP), after accounting for age.
Show ALL steps.

Model 1: $\mu(SBP\,|\,age, gender, \text{interaction}) = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3(age \times gender)$

Model 2: $\mu(SBP\,|\,age, gender) = \beta_0 + \beta_1 age + \beta_2 gender$

Model 3: $\mu(SBP\,|\,age) = \beta_0 + \beta_1 age$

Model 4: $\mu(SBP\,|\,gender) = \beta_0 + \beta_1 gender$

Model 1: [Note: The SPSS output for Model 1 is given at the start of this question]

Model 2:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	688.305	2	344.153	20.589	<.001
	Residual	351.028	21	16.7156		
	Total	1039.333	23			

a. Dependent Variable: SBP

b. Predictors: (Constant), Age, Gender

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	81.423	6.101		13.345	<.001
	Age	.598	.094	.810	6.389	<.001
	Gender	1.000	1.669	.076	.599	.556

a. Dependent Variable: SBP

Model 3:

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	682.305	1	682.305	42.044	<.001
	Residual	357.028	22	16.229		
	Total	1039.333	23			

a. Dependent Variable: SBP

b. Predictors: (Constant), Age

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	81.923	5.955		13.757	<.001
	Age	.598	.092	.810	6.484	<.001

a. Dependent Variable: SBP

Model 4:

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.000	1	6.000	0.128	.724
	Residual	1033.333	22	46.970		
	Total	1039.333	23			

a. Dependent Variable: SBP

b. Predictors: (Constant), Gender

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	119.667	1.978		60.486	<.001
	Gender	1.000	2.798	.076	.357	.724

a. Dependent Variable: SBP

>>>>>>>>>>

Solution for (e):

H0: $\beta_2 = \beta_3 = 0$ (Reduced model) Gender and/or the interaction have no effect on SBP.

OR: $\mu(SBP \mid age) = \beta_0 + \beta_1 age$ (Model 3)

Ha: At least one of the $\beta_i \neq 0, i = 2,3$ (Full model model)

OR: $\mu(SBP \mid age, gender, \text{interaction}) = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3 (age \times gender)$ (Model 1)

$$df_E(reduced) = n-(k+1) = 24-(1+1) = 22$$

$$df_E(full) = n-(k+1) = 24-(3+1) = 20$$

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

$$= \frac{[357.028 - 134.089]/[22-20]}{134.089/20} = \frac{222.939/2}{6.70445} = \frac{111.4695}{6.70445} = 16.626$$

F-distribution, with $df = [Extra\ df, df_{ERROR}(Full)] = [\text{Number of selected } \beta_i\text{'s}, n-(k+1)] = (2,20)$

P-value: P < 0.001, which provides extremely strong evidence against the null hypothesis.
Since P < α (0.05), reject H0

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that gender and/or the interaction between gender and age have an effect on systolic blood pressure (SBP), after accounting for age.

(f) For this part of the question, refer to the four models defined in part (e) and any of the SPSS output shown above. At the 5% significance level, <u>perform the most appropriate test</u>, to determine whether the interaction between gender and age has an effect on systolic blood pressure (SBP), <u>after accounting for</u> age and gender. <u>**Show ALL steps**</u>.

H0: $\beta_3 = 0$ (Reduced model) Interaction has no effect on SBP.

OR: $\mu(SBP \mid age, gender) = \beta_0 + \beta_1 age + \beta_2 gender$ (Model 2)

Ha: $\beta_3 \neq 0$ (Full model model)

OR: $\mu(SBP \mid age, gender, \text{interaction}) = \beta_0 + \beta_1 age + \beta_2 gender + \beta_3 (age \times gender)$ (Model 1)

$$df_E(reduced) = n-(k+1) = 24-(2+1) = 21$$

$$df_E(full) = n-(k+1) = 24-(3+1) = 20$$

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

$$= \frac{[351.028 - 134.089]/[21-20]}{134.089/20} = \frac{216.939/1}{6.70445} = 32.357464 \approx 32.357$$

F-distribution, with $df = [Extra\ df, df_{ERROR}(Full)] = [\text{Number of selected } \beta_i\text{'s}, n-(k+1)] = (1,20)$

P-value: P < 0.001, which provides extremely strong evidence against the null hypothesis.
Since P < α (0.05), reject H0

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that the interaction between gender and age has an effect on systolic blood pressure (SBP), after accounting for age and gender. [**Note:** Take $\sqrt{F}$ and compare it with the Full Model Table of Coefficients.]

>>>>>>>>