

Stat 6620: Multicollinearity Exercise

	Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Model 1	Intercept	117.08469	99.78240	1.17	0.2578
	X1	4.33409	3.01551	1.44	0.1699
	X2	-2.85685	2.58202	-1.11	0.2849
	X3	-2.18606	1.59550	-1.37	0.1896
Model 2	Intercept	-1.49610	3.31923	-0.45	0.6576
	X1	0.85719	0.12878	6.66	<.0001
Model 3	Intercept	-23.63449	5.65741	-4.18	0.0006
	X2	0.85655	0.11002	7.79	<.0001
Model 4	Intercept	-19.17425	8.36064	-2.29	0.0348
	X1	0.22235	0.30344	0.73	0.4737
	X2	0.65942	0.29119	2.26	0.0369

1. In Model 1, X_1 is not significant ($p = 0.1699$). However in Model 2, X_1 is significant ($p < .0001$). Which result should I believe, is X_1 significant or not?

Ans: X_1 is a statistically significant predictor of body fat, but does not add much to what (X_2, X_3) already provide.

2. In Model 1, the p-value for X_3 ($p = 0.1896$) is smaller than the p-value for X_2 ($p = 0.2849$). Does this mean X_3 is more important than X_2 ?

Ans: No. First, what do you mean by *importance*? Statistical tests do not provide information on *clinical* importance, only information on strength of (simultaneous, in this case) linear relationship. In the presence of multicollinearity where the variables explain the same thing, there is an element of luck as to which variable grabs the large portion of the effect, so coefficients and t-ratios and p-values should be interpreted with skepticism.

3. In Model 1, the predictor X_2 has negative slope ($b_2 = -2.85685$). However in Model 3, X_2 has positive slope ($b_2 = 0.85655$). Which result should I believe, does Y increase or decrease with X_2 ?

Ans: In the presence of multicollinearity where the variables explain the same thing, there is an element of luck as to which variable grabs the large portion of the effect, so coefficients and t-ratios and p-values should be interpreted with skepticism.

4. I think X_2 should have positive slope, because thigh circumference tends to increase with body weight. So I prefer Model 4 with $R^2 = 0.7781$ even if Model 1 has higher $R^2 = 0.8014$. Do you agree?

Ans: Since you are the medical expert, yes I agree. The two models are statistically equivalent, so medical interpretation and usefulness makes Model 4 better.