Multiple Regression Lab: Categorical Predictors

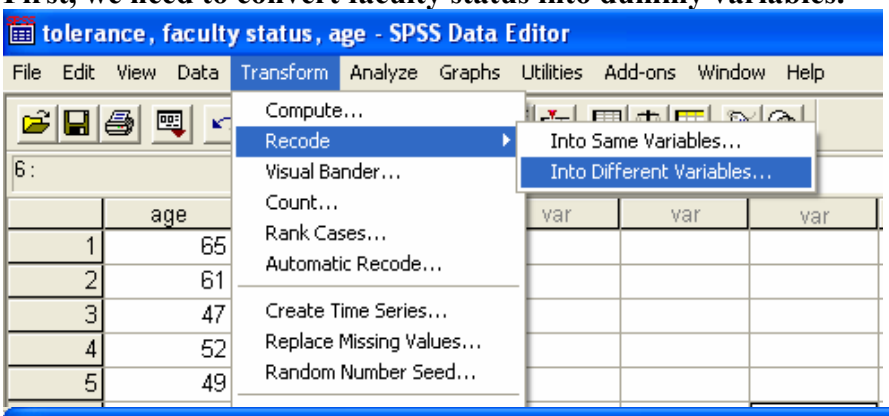
A social psychologist wanted to determine whether faculty members' age and status (1 = full professor, 2 = associate professor, 3 = assistant professor) are predictors of social tolerance. Further, the social psychologist thought the relationship between age and tolerance would differ by faculty status.

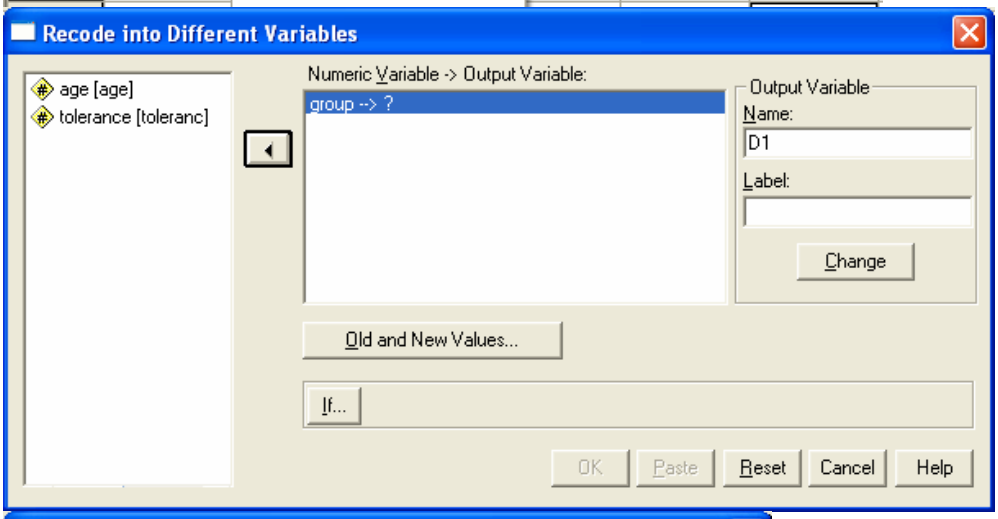
Age	Tolerance	Status
65	3	1
61	3	1
47	4	1
52	3	1
49	5	1
45	4	1
41	4	1
41	6	1
40	5	1
39	5	1
34	5	2
31	5	2
30	5	2
35	5	2
49	5	2
31	4	2
42	5	2
43	5	2
39	5	2
49	5	2
26	5	3
33	6	3
48	5	3
32	5	3
25	4	3
33	6	3
42	5	3
30	6	3
31	4	3
27	5	3

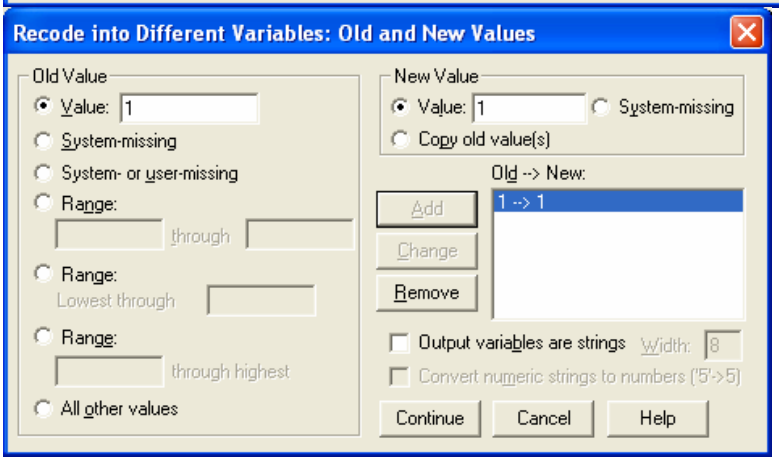
We will conduct several tests:

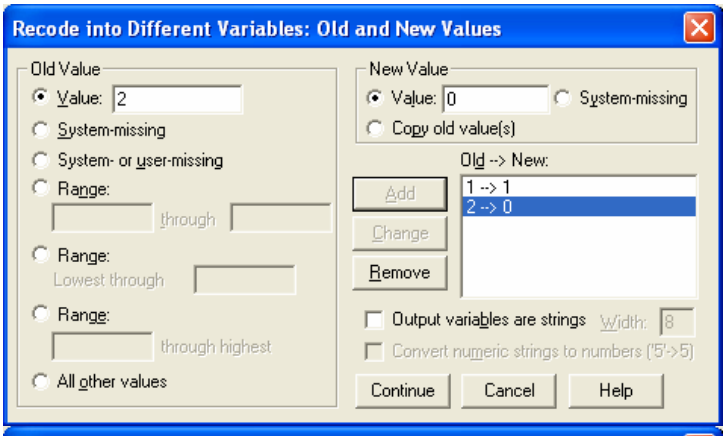
1. Is status a significant predictor of tolerance?
2. Is age a significant predictor of tolerance?
3. Is status a significant predictor of tolerance when controlling for age?
4. Is age a significant predictor of tolerance when controlling for faculty status?
5. Is the relationship between age and tolerance the same across faculty status?

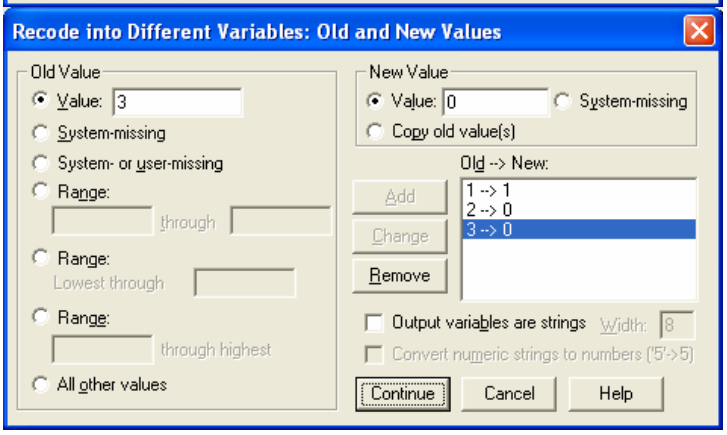
First, we need to convert faculty status into dummy variables.

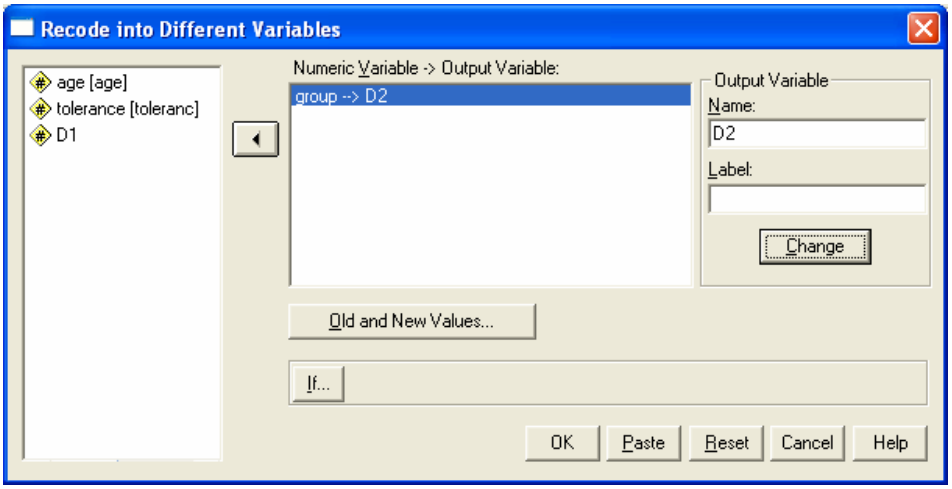


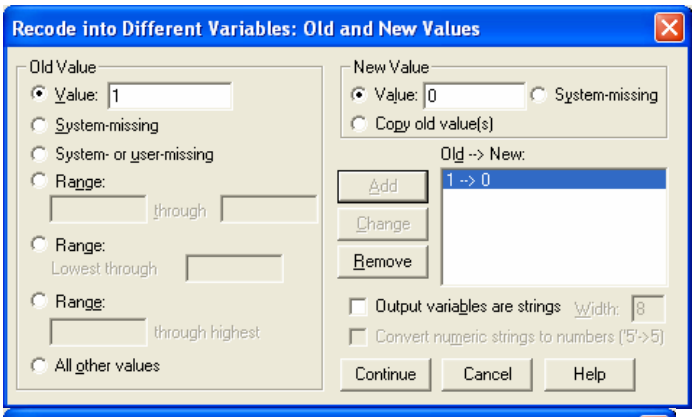


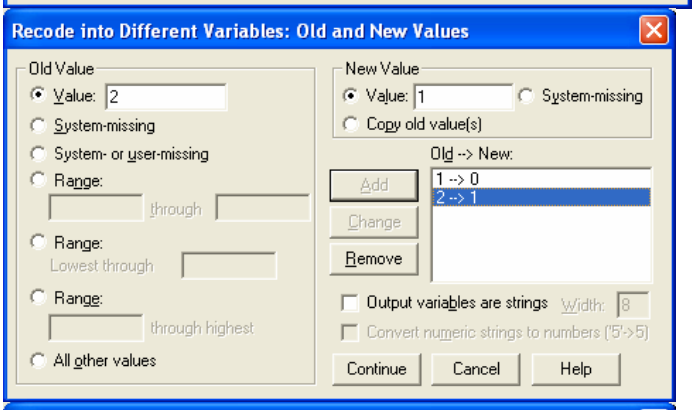


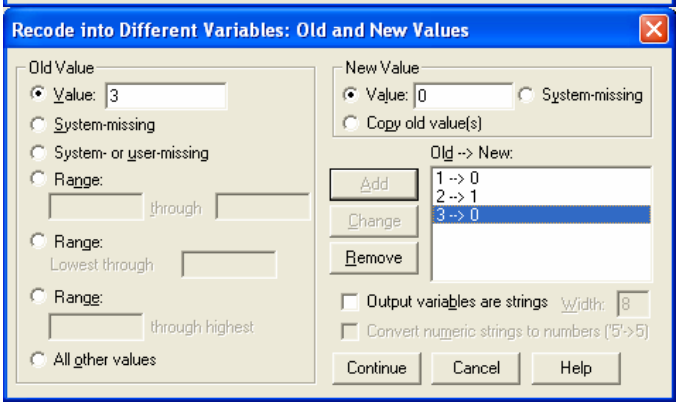

Recode into Different Variables: Old and New Values
Old Value
Value: 2
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 0
System-missing
Copy old value(s)
Old --> New:
1 --> 1
2 --> 0
Add
Change
Remove
Output variables are strings
Width: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

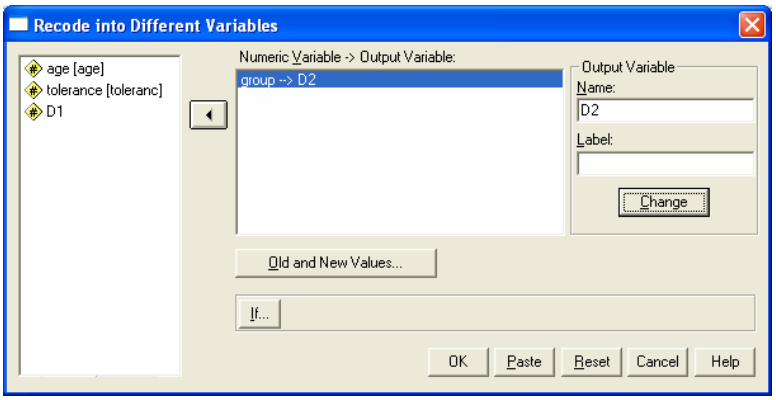

Recode into Different Variables: Old and New Values
Old Value
Value: 3
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 0
System-missing
Copy old value(s)
Old --> New:
1 --> 1
2 --> 0
3 --> 0
Add
Change
Remove
Output variables are strings
Width: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help


Recode into Different Variables
age [age]
tolerance [toleranc]
D1
Numeric Variable -> Output Variable:
group --> D2
Output Variable
Name:
D2
Label:
Change
Old and New Values...
If...
OK
Paste
Reset
Cancel
Help

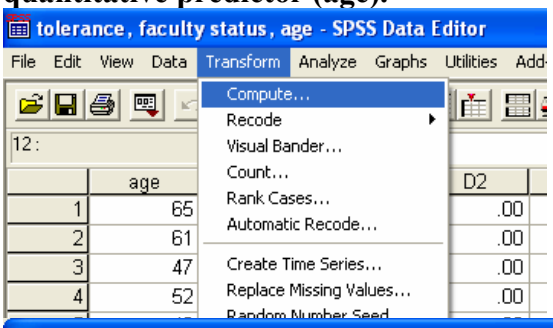

Recode into Different Variables: Old and New Values
Old Value
Value: 1
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 0
System-missing
Copy old value(s)
Old --> New:
1 --> 0
Add
Change
Remove
Output variables are strings
Width: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

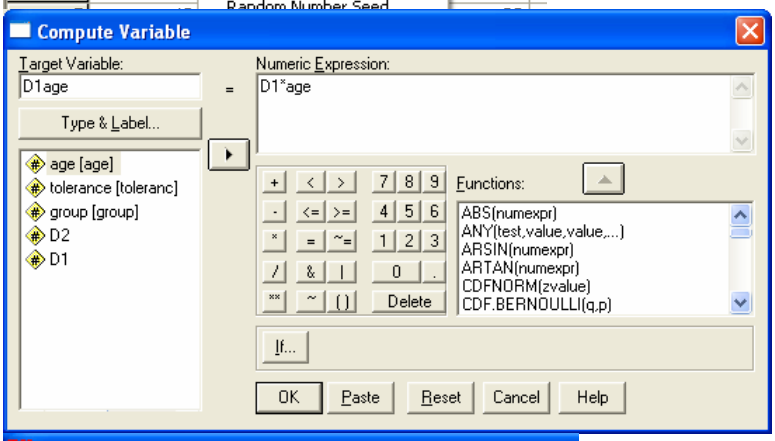

Recode into Different Variables: Old and New Values
Old Value
Value: 2
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 1
System-missing
Copy old value(s)
Old --> New:
1 --> 0
2 --> 1
Add
Change
Remove
Output variables are strings
Width: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

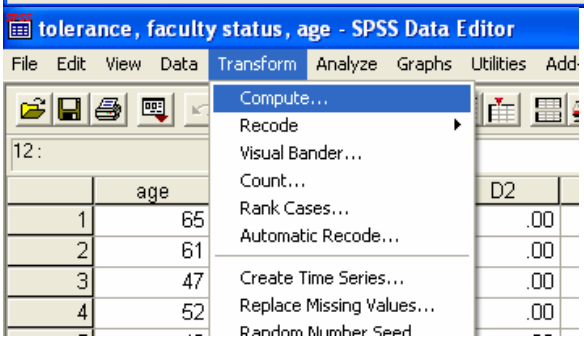

Recode into Different Variables: Old and New Values
Old Value
Value: 3
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 0
System-missing
Copy old value(s)
Old --> New:
1 --> 0
2 --> 1
3 --> 0
Add
Change
Remove
Output variables are strings
Width: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

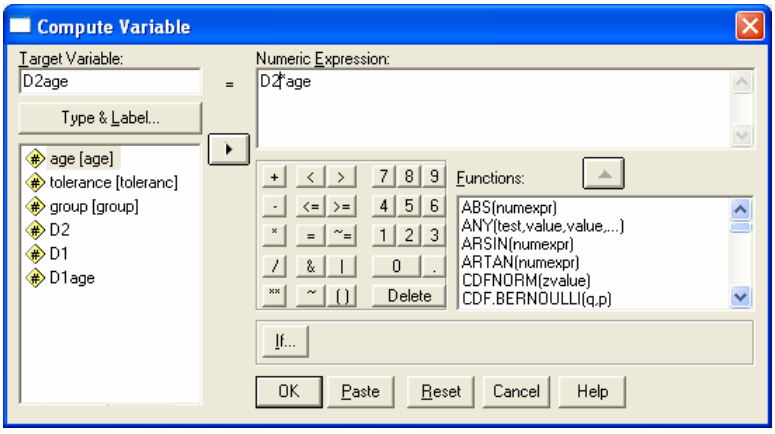

Recode into Different Variables
age [age]
tolerance [toleranc]
D1
Numeric Variable -> Output Variable:
group --> D2
Output Variable
Name:
D2
Label:
Change
Old and New Values...
If...
OK
Paste
Reset
Cancel
Help

Then, we need to create interaction terms for the categorical predictor (D1, D2) with the quantitative predictor (age).







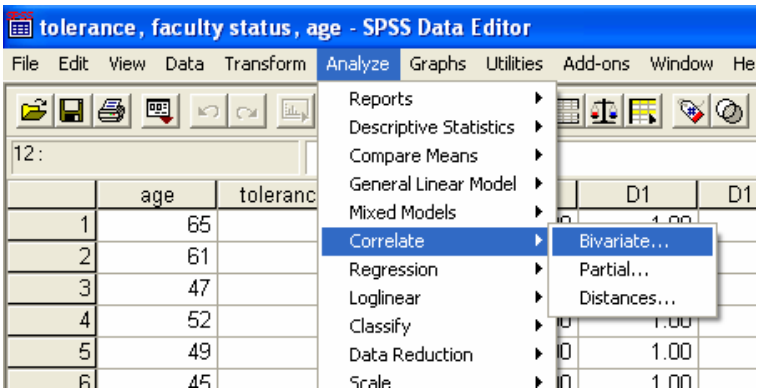


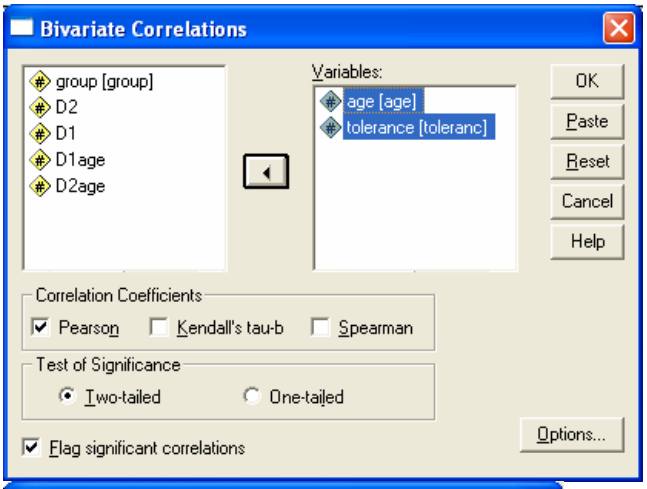
If you are lucky, your data file now resembles the data shown below.

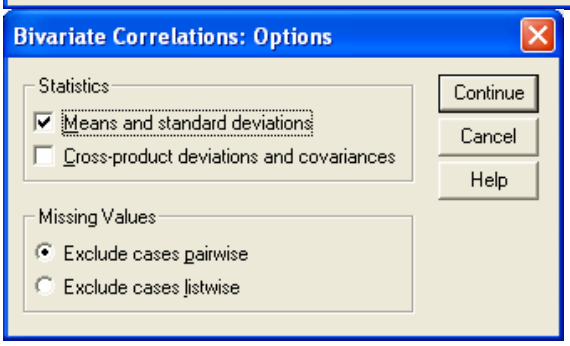
	age	toleranc	group	D2	D1	D1age	D2age
1	65	3	1	.00	1.00	65.00	.00
2	61	3	1	.00	1.00	61.00	.00
3	47	4	1	.00	1.00	47.00	.00
4	52	3	1	.00	1.00	52.00	.00
5	49	5	1	.00	1.00	49.00	.00
6	45	4	1	.00	1.00	45.00	.00
7	41	4	1	.00	1.00	41.00	.00
8	41	6	1	.00	1.00	41.00	.00
9	40	5	1	.00	1.00	40.00	.00
10	39	5	1	.00	1.00	39.00	.00
11	34	5	2	1.00	.00	.00	34.00
12	31	5	2	1.00	.00	.00	31.00
13	30	5	2	1.00	.00	.00	30.00
14	35	5	2	1.00	.00	.00	35.00
15	49	5	2	1.00	.00	.00	49.00
16	31	4	2	1.00	.00	.00	31.00
17	42	5	2	1.00	.00	.00	42.00
18	43	5	2	1.00	.00	.00	43.00
19	39	5	2	1.00	.00	.00	39.00
20	49	5	2	1.00	.00	.00	49.00
21	26	5	3	.00	.00	.00	.00
22	33	6	3	.00	.00	.00	.00
23	48	5	3	.00	.00	.00	.00
24	32	5	3	.00	.00	.00	.00
25	25	4	3	.00	.00	.00	.00
26	33	6	3	.00	.00	.00	.00
27	42	5	3	.00	.00	.00	.00
28	30	6	3	.00	.00	.00	.00
29	31	4	3	.00	.00	.00	.00
30	27	5	3	.00	.00	.00	.00

Data View | Variable View

Now that the variables are created, we need to request descriptive statistics and correlations.







Correlations

Descriptive Statistics

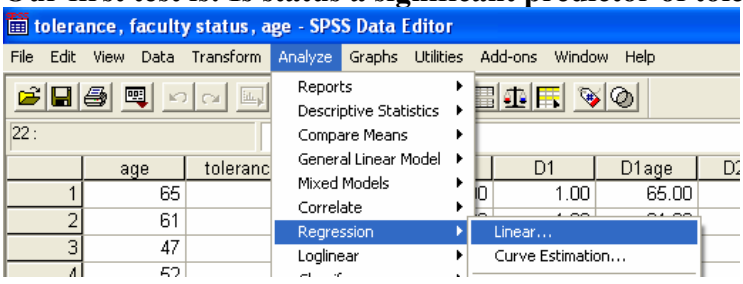
	Mean	Std. Deviation	N
age	39.6667	9.93889	30
tolerance	4.7333	.82768	30

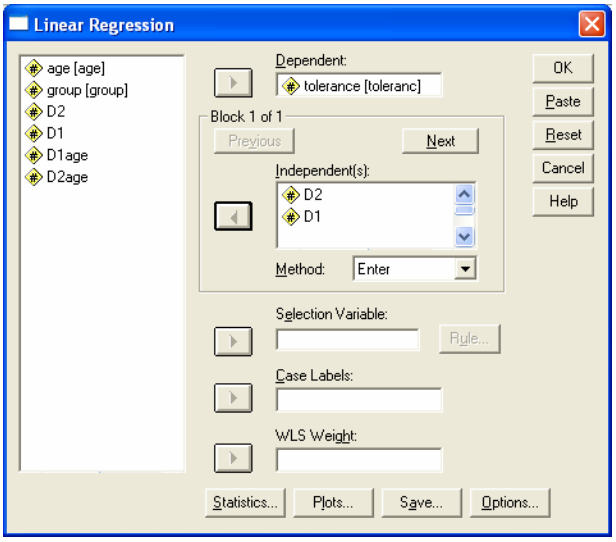
Correlations

		age	tolerance
age	Pearson Correlation	1	-.510**
	Sig. (2-tailed)	.	.004
	N	30	30
tolerance	Pearson Correlation	-.510**	1
	Sig. (2-tailed)	.004	.
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Our first test is: Is status a significant predictor of tolerance?





Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	D2, D1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: tolerance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.474[a]	.225	.167	.75523

a. Predictors: (Constant), D2, D1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.467	2	2.233	3.916	.032[a]
	Residual	15.400	27	.570		
	Total	19.867	29			

a. Predictors: (Constant), D2, D1

b. Dependent Variable: tolerance

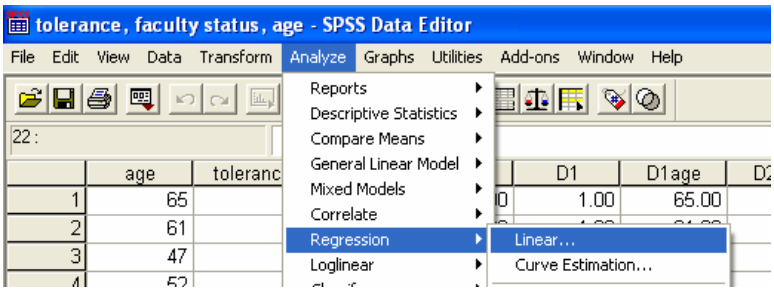
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.100	.239		21.355	.000
	D1	-.900	.338	-.521	-2.665	.013
	D2	-.200	.338	-.116	-.592	.559

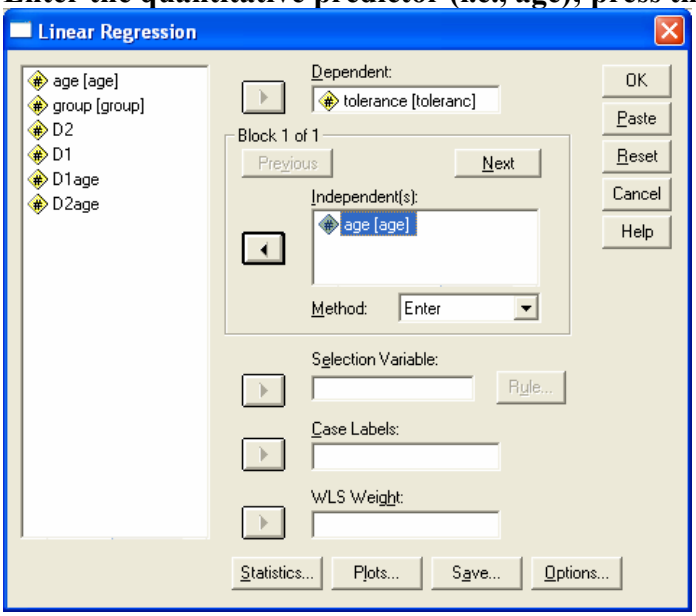
a. Dependent Variable: tolerance

The ANOVA F would tell us whether faculty status is a significant predictor of tolerance.

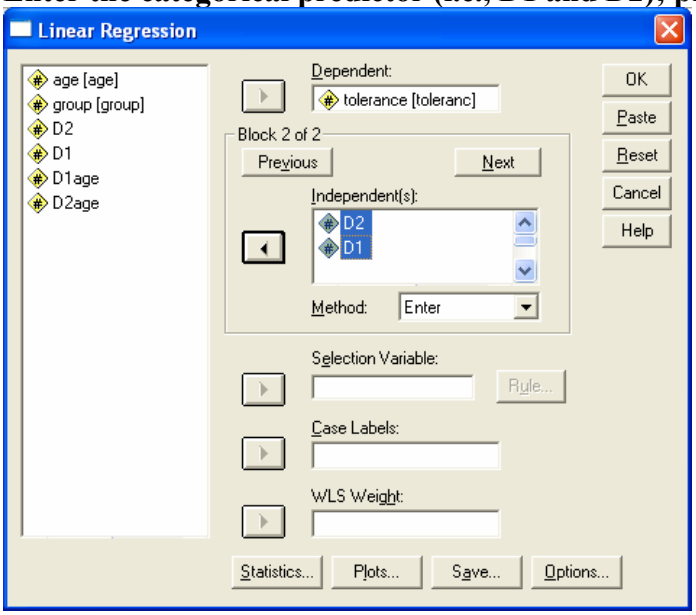
The remaining questions can be obtained through one set of analyses if you do it right! The order of the variables entered as predictors is VERY important.



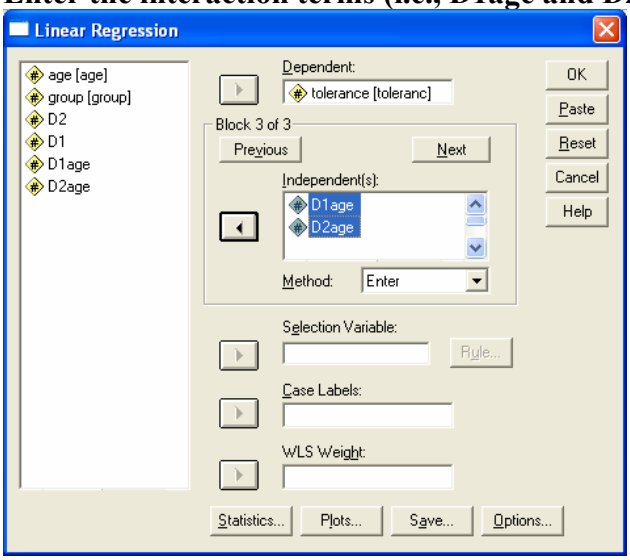
Enter the quantitative predictor (i.e., age); press the NEXT button.



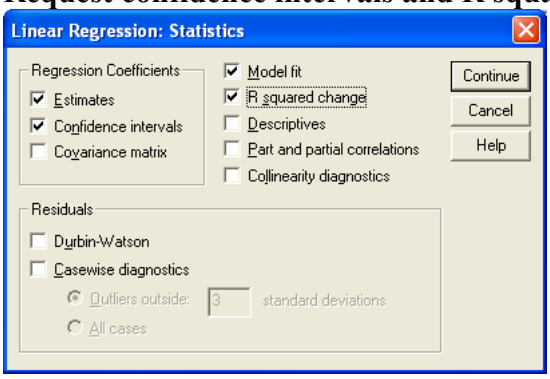
Enter the categorical predictor (i.e., D1 and D2); press the NEXT button.



Enter the interaction terms (i.e., D1age and D2age).



Request confidence intervals and R squared Change.



Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	age[a]	.	Enter
2	D2, D1[a]	.	Enter
3	d2age, d1age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: tolerance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.510[a]	.260	.234	.72455	.260	9.843	1	28	.004
2	.546[b]	.299	.218	.73210	.038	.713	2	26	.500
3	.721[c]	.519	.419	.63069	.221	5.517	2	24	.011

a. Predictors: (Constant), age

b. Predictors: (Constant), age, D2, D1

c. Predictors: (Constant), age, D2, D1, d2age, d1age

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.167	1	5.167	9.843	.004[a]
	Residual	14.699	28	.525		
	Total	19.867	29			
2	Regression	5.931	3	1.977	3.689	.025[b]
	Residual	13.935	26	.536		
	Total	19.867	29			
3	Regression	10.320	5	2.064	5.189	.002[c]
	Residual	9.546	24	.398		
	Total	19.867	29			

a. Predictors: (Constant), age

b. Predictors: (Constant), age, D2, D1

c. Predictors: (Constant), age, D2, D1, d2age, d1age

d. Dependent Variable: tolerance

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.418	.553		11.605	.000	5.285	7.551
	age	-.042	.014	-.510	-3.137	.004	-.070	-.015
2	(Constant)	6.069	.631		9.626	.000	4.773	7.366
	age	-.030	.018	-.356	-1.653	.110	-.067	.007
	D1	-.446	.427	-.259	-1.045	.306	-1.324	.432
	D2	-.034	.342	-.020	-.099	.922	-.738	.670
3	(Constant)	4.590	.974		4.713	.000	2.580	6.600
	age	.016	.029	.187	.535	.598	-.045	.076
	D1	3.830	1.499	2.219	2.555	.017	.736	6.924
	D2	-.285	1.493	-.165	-.191	.850	-3.366	2.797
	d1age	-.104	.037	-2.945	-2.770	.011	-.181	-.026
	d2age	.000	.041	-.002	-.002	.999	-.085	.085

a. Dependent Variable: tolerance

Excluded Variables[c]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	D1	-.245[a]	-1.212	.236	-.227	.636
	D2	.095[a]	.576	.569	.110	.990
	d1age	-.409[a]	-1.888	.070	-.342	.515
	d2age	.153[a]	.937	.357	.177	1.000
2	d1age	-2.945[b]	-3.390	.002	-.561	.025
	d2age	1.429[b]	1.629	.116	.310	.033

a. Predictors in the Model: (Constant), age

b. Predictors in the Model: (Constant), age, D2, D1

c. Dependent Variable: tolerance

Is age a significant predictor of tolerance?
Model 1 contains only age. Thus, the *t* test in the coefficients section **OR** the *F* in the ANOVA summary section for **MODEL 1** would be used to answer this question. [Since there is only one predictor in Model 1, the *t* and the *F* are equivalent.]

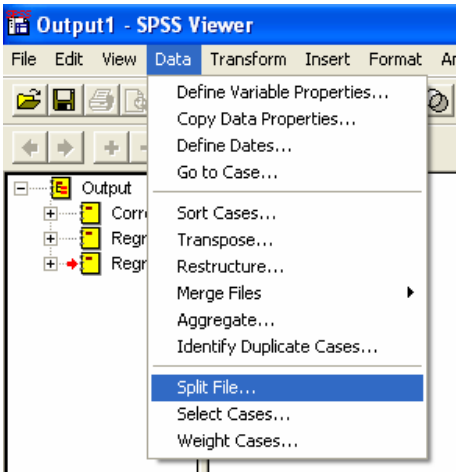
Is status a significant predictor of tolerance when controlling for age?
Model 2 contains status (D1, D2) and age. The *F CHANGE and Sig. F Change* in the MODEL SUMMARY for **Model 2** would be used to answer this question.

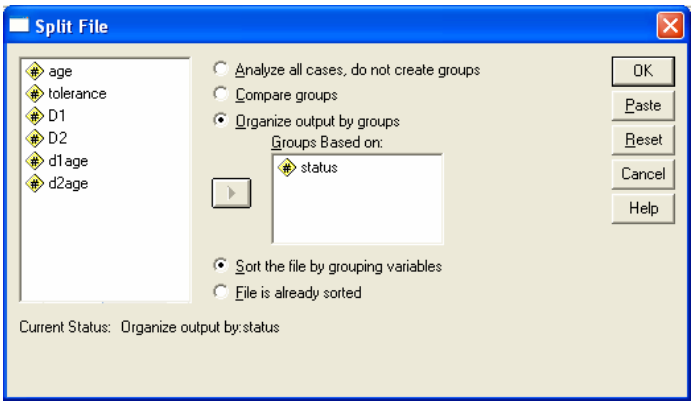
Is age a significant predictor of tolerance when controlling for faculty status?
Model 2 contains status (D1, D2) and age. The *t* test for age in the Coefficients section for **Model 2** would be used to answer this question.

Is the relationship between age and tolerance the same across faculty status?
Model 3 contains status (D1, D2), age, and the interaction terms (D1age, D2age). The *F CHANGE and Sig. F Change* in the MODEL SUMMARY for **MODEL 3** would be used to answer this question.

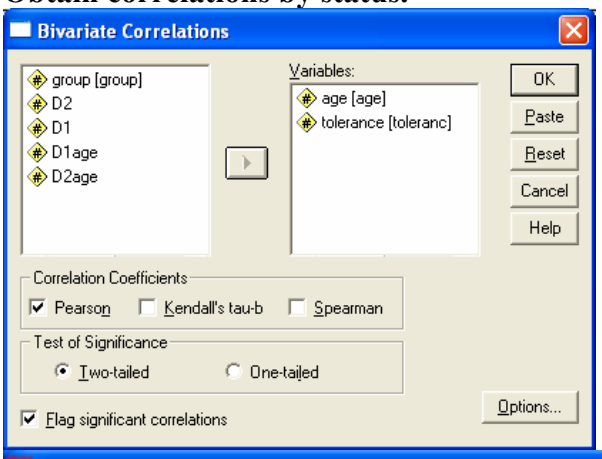
If the relationship between age and tolerance was similar across faculty status, we would have all the output that we would need for the results section.

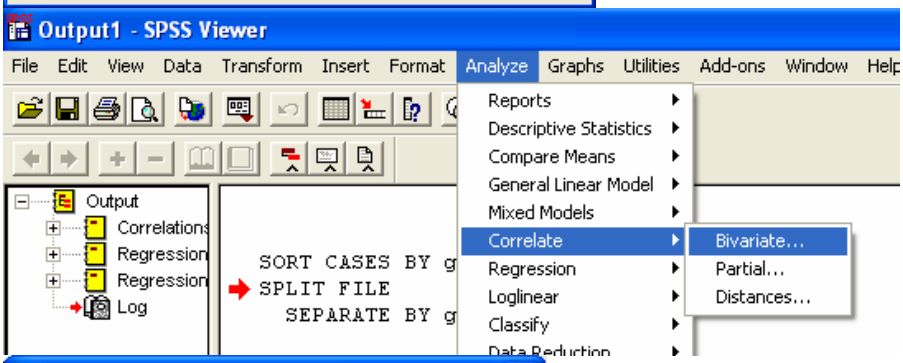
Since the relationship between age and tolerance was NOT the same across faculty status, we need to obtain descriptive statistics, correlations, and regression analyses predicting tolerance using age for each faculty status group.

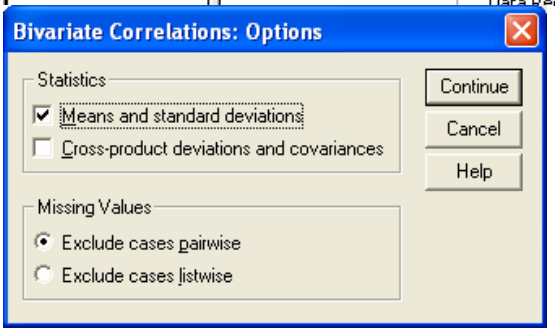




Obtain correlations by status.







Correlations
status = Full Professor

Descriptive Statistics[a]

	Mean	Std. Deviation	N
age	48.0000	8.99383	10
tolerance	4.2000	1.03280	10

a. status = Full Professor

Correlations[a]

		age	tolerance
age	Pearson Correlation	1	-.766**
	Sig. (2-tailed)	.	.010
	N	10	10
tolerance	Pearson Correlation	-.766**	1
	Sig. (2-tailed)	.010	.
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

a. status = Full Professor

status = Associate Professor

Descriptive Statistics[a]

	Mean	Std. Deviation	N
age	38.3000	7.22726	10
tolerance	4.9000	.31623	10

a. status = Associate Professor

Correlations[a]

		age	tolerance
age	Pearson Correlation	1	.355
	Sig. (2-tailed)	.	.314
	N	10	10
tolerance	Pearson Correlation	.355	1
	Sig. (2-tailed)	.314	.
	N	10	10

a. status = Associate Professor

status = Assistant professor

Descriptive Statistics[a]

	Mean	Std. Deviation	N
age	32.7000	7.21187	10
tolerance	5.1000	.73786	10

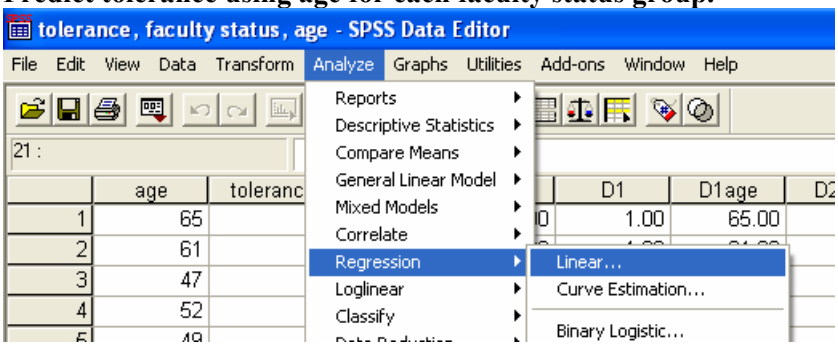
a. status = Assistant professor

Correlations[a]

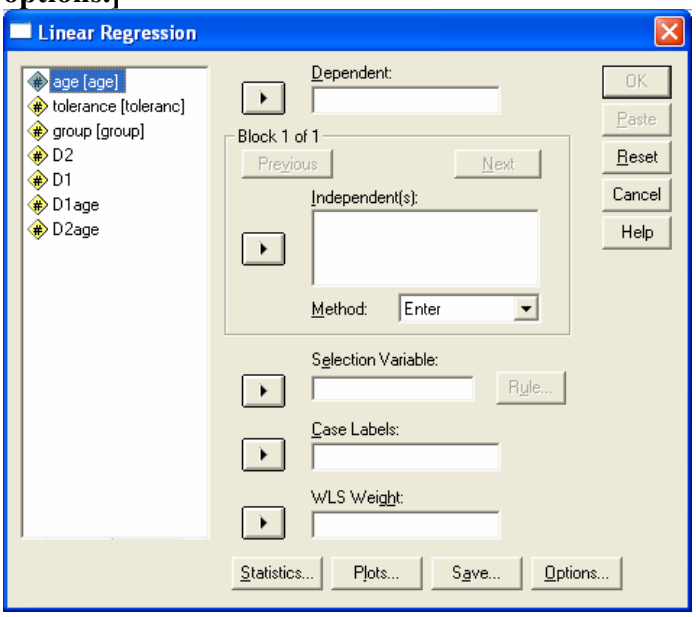
		age	tolerance
age	Pearson Correlation	1	.152
	Sig. (2-tailed)	.	.674
	N	10	10
tolerance	Pearson Correlation	.152	1
	Sig. (2-tailed)	.674	.
	N	10	10

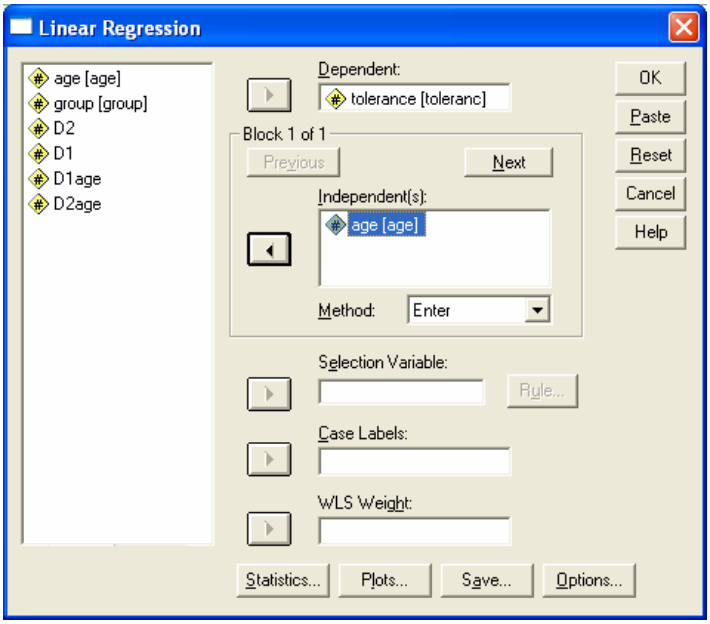
a. status = Assistant professor

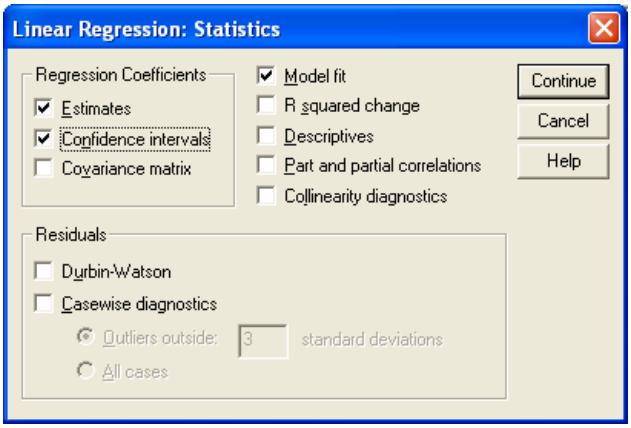
Predict tolerance using age for each faculty status group.



[It is best to press the RESET button at this point to remove all the previously specified options.]







status = Full Professor

Variables Entered/Removed[b,c]

Model	Variables Entered	Variables Removed	Method
1	age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: tolerance

c. status = Full Professor

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.766[a]	.586	.534	.70477

a. Predictors: (Constant), age

b. status = Full Professor

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.626	1	5.626	11.327	.010[a]
	Residual	3.974	8	.497		
	Total	9.600	9			

a. Predictors: (Constant), age

b. Dependent Variable: tolerance

c. status = Full Professor

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	8.420	1.273		6.612	.000	5.483	11.356
	age	-.088	.026	-.766	-3.366	.010	-.148	-.028

a. Dependent Variable: tolerance

b. status = Full Professor

Is age a significant predictor of tolerance for full professors?

status = Associate Professor

Variables Entered/Removed[b,c]

Model	Variables Entered	Variables Removed	Method
1	age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: tolerance

c. status = Associate Professor

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.355[a]	.126	.017	.31358

a. Predictors: (Constant), age

b. status = Associate Professor

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.113	1	.113	1.153	.314[a]
	Residual	.787	8	.098		
	Total	.900	9			

a. Predictors: (Constant), age

b. Dependent Variable: tolerance

c. status = Associate Professor

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.305	.563		7.651	.000	3.008	5.603
	age	.016	.014	.355	1.074	.314	-.018	.049

a. Dependent Variable: tolerance

b. status = Associate Professor

Is age a significant predictor of tolerance for associate professors?

status = Assistant professor

Variables Entered/Removed[b,c]

Model	Variables Entered	Variables Removed	Method
1	age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: tolerance

c. status = Assistant professor

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.152[a]	.023	-.099	.77348

a. Predictors: (Constant), age

b. status = Assistant professor

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.114	1	.114	.190	.674[a]
	Residual	4.786	8	.598		
	Total	4.900	9			

a. Predictors: (Constant), age

b. Dependent Variable: tolerance

c. status = Assistant professor

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.590	1.194		3.843	.005	1.836	7.344
	age	.016	.036	.152	.436	.674	-.067	.098

a. Dependent Variable: tolerance

b. status = Assistant professor

Is age a significant predictor of tolerance for assistant professors?