

Models for Group Comparisons – Summary

Since we are estimating and comparing several models, it will be helpful to list several of them all in one place. This handout summarizes how to do group comparisons both by running separate models for each group and by using interaction terms. We won't go through this handout separately, but it may help you to keep everything straight.

I. **Pooled (Constrained) Model.** Blacks and whites are combined into a single analysis, hence the coefficient estimates are constrained to be the same for both racial groups, i.e. the intercepts and the effect of education and job experience are the same for both groups.

```
. use "http://www.nd.edu/~rwilliam/stats2/statafiles/blwh.dta"
. reg income educ jobexp
```

Source	SS	df	MS			
Model	32798.4018	2	16399.2009	Number of obs =	500	
Residual	7382.84742	497	14.8548238	F(2, 497) =	1103.96	
				Prob > F =	0.0000	
				R-squared =	0.8163	
				Adj R-squared =	0.8155	
				Root MSE =	3.8542	
Total	40181.2493	499	80.5235456			

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.94512	.0436998	44.51	0.000	1.859261	2.03098
jobexp	.7082212	.0343672	20.61	0.000	.6406983	.775744
_cons	-7.382935	.8027781	-9.20	0.000	-8.960192	-5.805678

Hence, $SSE_c = 7382.84680$, $R^2_c = .81626$, $N = 500$, $K = 2$, $DFE_c = 497$.

II. **Unconstrained Models: Separate Models for each group.** Here, we estimate separate models, first for whites, then blacks. This makes it possible for the intercepts and slope coefficients to freely differ across populations. This is equivalent to Model IIIC.

Whites:

```
. reg income educ jobexp if !black
```

Source	SS	df	MS			
Model	18361.9894	2	9180.99472	Number of obs =	400	
Residual	5878.16991	397	14.8064733	F(2, 397) =	620.07	
				Prob > F =	0.0000	
				R-squared =	0.7575	
				Adj R-squared =	0.7563	
				Root MSE =	3.8479	
Total	24240.1594	399	60.7522791			

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.893338	.0562591	33.65	0.000	1.782735	2.003941
jobexp	.722255	.0412236	17.52	0.000	.6412111	.8032988
_cons	-6.461189	1.089219	-5.93	0.000	-8.602547	-4.31983

Hence, $N_w = 400$, $SSE_w = 5878.16929$, $DFE_w = 397$.

Blacks:

```
. reg income educ jobexp if black
```

Source	SS	df	MS			
Model	4924.27286	2	2462.13643	Number of obs =	100	
Residual	891.81705	97	9.19399021	F(2, 97) =	267.80	
Total	5816.08991	99	58.748383	Prob > F =	0.0000	
				R-squared =	0.8467	
				Adj R-squared =	0.8435	
				Root MSE =	3.0322	

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.677949	.0725479	23.13	0.000	1.533962	1.821936
jobexp	.421975	.0581021	7.26	0.000	.3066585	.5372915
_cons	-3.0512	1.154604	-2.64	0.010	-5.342771	-.7596302

Hence, $N_b = 100$, $SSE_b = 891.81711$, $DFE_b = 97$. Combining the black and white numbers, for the unconstrained model we get $N_u = 500$, $SSE_u = 6770$, $DFE_u = 494$. We now compute the incremental F:

$$F_{K+1, N_1+N_2-2K-2} = \frac{(SSE_c - SSE_u) * (N_1 + N_2 - 2K - 2)}{SSE_u * (K + 1)} = \frac{(7383 - 6770) * 494}{6770 * 3} = 14.91$$

III. **Unconstrained and Partially Unconstrained Models: Dummy Variables and Interaction Effects.** In this approach, interaction effects and dummy variables are used to allow for racial differences in parameter effects. The dummy variable BLACK is coded 1 if black, 0 otherwise. BLACKED = BLACK * EDUC. BLACKJOB = BLACK * JOBEXP. Interaction effects allow more flexibility in model specification. Rather than letting all parameters freely vary across groups versus constraining all effects to be the same, interaction effects allow for intermediate approaches.

IIIa. **Intercepts vary across groups, all other parameters the same.** To do this, you regress Y on the IVs and include a dummy variable for race.

```
. reg income educ jobexp black
```

Source	SS	df	MS			
Model	33206.4588	3	11068.8196	Number of obs =	500	
Residual	6974.79047	496	14.0620776	F(3, 496) =	787.14	
Total	40181.2493	499	80.5235456	Prob > F =	0.0000	
				R-squared =	0.8264	
				Adj R-squared =	0.8254	
				Root MSE =	3.7499	

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.840407	.0467507	39.37	0.000	1.748553	1.932261
jobexp	.6514259	.0350604	18.58	0.000	.5825406	.7203111
black	-2.55136	.4736266	-5.39	0.000	-3.481921	-1.620798
_cons	-4.72676	.9236842	-5.12	0.000	-6.541576	-2.911943

```
. * Wald test of differences in intercepts
. test black
```

```
( 1) black = 0
```

```
F( 1, 496) = 29.02
Prob > F = 0.0000
```

IIIb. Intercepts and some slopes vary across groups, other slopes are the same. Regress Y on the IVs, a dummy variable for race, and (in this example) one interaction term:

```
. gen blacked = black*educ
. gen blackjob = black*jobexp

. reg income educ jobexp black blackjob
```

Source	SS	df	MS	Number of obs =	500
Model	33352.2559	4	8338.06397	F(4, 495) =	604.39
Residual	6828.99339	495	13.7959462	Prob > F =	0.0000
Total	40181.2493	499	80.5235456	R-squared =	0.8300
				Adj R-squared =	0.8287
				Root MSE =	3.7143

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.834776	.0463385	39.60	0.000	1.743732	1.925821
jobexp	.7128145	.0395293	18.03	0.000	.6351486	.7904805
black	.4686862	1.040728	0.45	0.653	-1.576103	2.513475
blackjob	-.2556117	.0786289	-3.25	0.001	-.4100993	-.1011242
_cons	-5.514076	.9464143	-5.83	0.000	-7.373561	-3.654592

```
. * Wald test of racial difference in effect of jobexp
. test blackjob
```

```
( 1) blackjob = 0
```

```
F( 1, 495) = 10.57
Prob > F = 0.0012
```

```
. * Wald test of racial differences in jobexp effect and/or race intercept
. test black blackjob
```

```
( 1) black = 0
```

```
( 2) blackjob = 0
```

```
F( 2, 495) = 20.07
Prob > F = 0.0000
```

IIIc. Intercepts and all slopes free to vary across groups (totally unconstrained). Equivalent to Model II, where we estimated separate models for each group.

```
. reg income educ jobexp black blacked blackjob
```

Source	SS	df	MS	Number of obs =	500
Model	33411.2623	5	6682.25246	F(5, 494) =	487.60
Residual	6769.98696	494	13.7044271	Prob > F =	0.0000
Total	40181.2493	499	80.5235456	R-squared =	0.8315
				Adj R-squared =	0.8298
				Root MSE =	3.7019

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	1.893338	.054125	34.98	0.000	1.786994 1.999681
jobexp	.722255	.0396598	18.21	0.000	.6443322 .8001777
black	3.409988	1.756477	1.94	0.053	-.0410984 6.861075
blacked	-.2153886	.1038015	-2.08	0.039	-.4193354 -.0114418
blackjob	-.3002799	.0812705	-3.69	0.000	-.4599584 -.1406015
_cons	-6.461189	1.0479	-6.17	0.000	-8.520079 -4.402298

```
. * Wald test of differences in slopes
. test blacked blackjob
```

```
( 1) blacked = 0
( 2) blackjob = 0
```

```
F( 2, 494) = 7.47
Prob > F = 0.0006
```

```
. * Wald test of any differences across groups, including intercepts
. test black blacked blackjob
```

```
( 1) black = 0
( 2) blacked = 0
( 3) blackjob = 0
```

```
F( 3, 494) = 14.91
Prob > F = 0.0000
```

Note that $N_u = 500$, $SSE_u = 6770$, $DFE_u = 494$. These are the exact same numbers we got using the Model II procedure of estimating separate models for each racial group. Further, the regression coefficients estimated under this approach can easily be converted to the betas estimated under the previous approach, and vice versa.