Reviewing Main Effects in General Linear Models
(as estimated using restricted maximum likelihood in SAS PROC MIXED)

The models for this example come from Hoffman (2015) chapter 2. We will be examining the extent to which cognition (as measured by an information test outcome) can be predicted from age (centered at 85 years) grip strength (centered at 9 pounds), sex (with men as the reference group) and dementia status (none = 1, future = 2, and current = 3) in a sample of 550 older adults.

SAS Syntax and Output for Data Manipulation and Data Description:

```
* Defining global variable for file location to be replaced in code below;
  %LET filesave= C:\Dropbox\17_CLDP944\CLDP944_Example02;
* Location for SAS files for these models (uses macro variable filesave);
  LIBNAME filesave "&filesave.";

* Import chapter 2 example data into work library;
DATA work.Chapter2; SET filesave.SAS_Chapter2;
* Center continuous predictors;
age85 = age - 85;
grip9 = grip - 9;
* Creating manual contrasts for dementia groups (to be treated as continuous);
     IF demgroup=1 THEN DO; demNF=0; demNC=0; END; * None group is reference;
ELSE IF demgroup=2 THEN DO; demNF=1; demNC=0; END; * Future group difference;
ELSE IF demgroup=3 THEN DO; demNF=0; demNC=1; END; * Current group difference;
* Labeling new variables - note semi-colon is only at the end of all labels;
LABEL
age85=  "age85: Age in Years (0=85)"
grip9=  "grip9: Grip Strength in Pounds (0=9)"
sexMW=  "sexMW: Sex (0=Men, 1=Women)"
demNF=  "demNF: Dementia Contrast for None=0 vs Future=1"
demNC=  "demNC: Dementia Contrast for None=0 vs Current=1";
RUN;

TITLE1 "Chapter 2: Descriptive Statistics for Example Variables";
PROC MEANS NDEC=2 DATA=work.Chapter2; VAR age grip cognition; RUN;
PROC FREQ  DATA=work.Chapter2; TABLE sexMW demgroup; RUN;
PROC CORR  DATA=work.Chapter2; VAR age grip sexMW cognition; RUN;
TITLE1;
```

New-school default SAS HTML output:

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
age	age: Age in Years	550	84.93	3.43	80.02	96.97
grip	grip: Grip Strength in Pounds	550	9.11	2.98	0	19.00
cognition	cognition: Information Test Cognitive Outcome	550	24.82	10.99	0	44.00

sexMW: Sex (0=Men, 1=Women)				
sexMW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	227	41.27	227	41.27
1	323	58.73	550	100.00

demgroup: Dementia Diagnosis (1=None, 2=Future, 3=Current)				
demgroup	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	399	72.55	399	72.55
2	109	19.82	508	92.36
3	42	7.64	550	100.00

SAS MIXED Syntax and Output for Empty Model in Equation 2.3

$$Cognition_i = \beta_0 + e_i$$

```
TITLE1 'Eq 2.3: Empty Means Model';

PROC MIXED DATA=work.Chapter2 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;

     MODEL cognition =  / SOLUTION DDFM=BW;

     ODS OUTPUT CovParms=CovEmpty;

RUN; TITLE1;
```

MODEL y = fixed effects of predictors

ODS OUTPUT saves CovParms table to dataset called "CovEmpty" for use in R^2 macro.

Old-school default SAS listing output (much easier to paste into handouts):

Number of Observations

Number of Observations Read	550
Number of Observations Used	550
Number of Observations Not Used	0

This table tells you how many cases were removed due to incomplete data—make sure you pay attention to this if you are doing any model comparisons (which will need to be based on the same cases to be valid).

Covariance Parameter Estimates → CovParms

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Residual	120.76	7.2887	16.57	<.0001

This table will list all estimated parameters within the model for the variance. Right now all we have is residual variance, the variance of the e_i residuals. Because this is an empty model with no predictors, this is ALL the variance to be predicted in the cognition outcome.

Fit Statistics

-2 Res Log Likelihood	4196.1
AIC (Smaller is Better)	4198.1
AICC (Smaller is Better)	4198.1
BIC (Smaller is Better)	4202.4

This first "Fit Statistics" table will index relative model fit (stay tuned). The second "Information Criteria" table reports the same info plus other indices. The "parms" in REML refers to the number of estimated parameters in the model for the variance (just 1 residual variance now).

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
4196.1	1	4198.1	4198.1	4199.8	4202.4	4203.4

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	24.8218	0.4686	549	52.97	<.0001

This "Solution for Fixed Effects" table will list all estimated parameters in the model for the means. It is not printed by default in PROC MIXED.

Intercept β_0 =

SAS MIXED Syntax and Output for Age, Grip, and Sex (0=M, 1=W) Model in Equation 2.7

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) + e_i$$

```
TITLE1 'Eq 2.7: Age + Grip + Adding Sex (0=M 1=W, as continuous predictor)';
PROC MIXED DATA=work.Chapter2 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
     MODEL cognition = age85 grip9 sexMW / CHISQ SOLUTION DDFM=BW;
     CONTRAST "Model R2 F-Test with df=3" age85 1, grip9 1, sexMW 1 / CHISQ;
ODS OUTPUT CovParms=CovAgeGripSex;
RUN; TITLE1;
```

CONTRAST provides a multivariate Wald test for the significance of the model (reduction in error variance from adding three fixed effects). Here I have also requested a chi-square value for illustration purposes (F*df = chi-square).

ODS OUTPUT saves CovParms table to dataset called "CovAgeGripSex" for use in R^2 macro. Thus, you can name each saved dataset whatever you want (but max 32 characters)

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Residual	109.38	6.6200	16.52	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC
4141.1	1	4143.1	4143.1	4144.8

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	26.9594	0.7389	546	36.49	<.0001	is B0
age85	-0.4338	0.1325	546	-3.27	0.0011	is B1
grip9	0.5460	0.1663	546	3.28	0.0011	is B2
sexMW	-3.7988	0.9904	546	-3.84	0.0001	is B3

Interpret each fixed effect:

Intercept β_0 =

Main effect of Age β_1 =

Main effect of Grip Strength β_2 =

Main effect of Sex β_3 =

Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Model R2 F-Test with df=3	3	546	60.11	20.04	<.0001	<.0001

```
* Call macro to calculate R2 for overall model and change in R2 between models;
  %ModelR2(CovBase=CovEmpty, CovFewer=CovEmpty, CovMore=CovAgeGripSex);
```

This table is created by my macro program to calculate R^2 and change in R^2.

R2 (% Reduction) Overall and for CovEmpty vs. CovAgeGripSex

Name	CovParm	Estimate	StdErr	ZValue	ProbZ	R2_from_Base	R2_Increment
CovEmpty	Residual	120.76	7.2887	16.57	<.0001	0.000000	.
CovAgeGripSex	Residual	109.38	6.6200	16.52	<.0001	0.094221	0.094221

Calculate model R^2 = (empty σ^2_e – current σ^2_e) / (empty σ^2_e) = (120.76 − 109.38) / (120.76) = .09

The df=3 CONTRAST above says that this R^2 is significantly > 0, $F(3,546) = 22.04$, $p < .0001$.

SAS MIXED Syntax and Output for Dementia Group Model in Equation 2.8

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + e_i$$

We can use the model equation to calculate the **dementia group means** for predicted cognition:

Cognition for None = β_0

Cognition for Future = $\beta_0 + \beta_4$

Cognition for Current = $\beta_0 + \beta_5$

We can determine the **differences between the dementia group means** as follows:

None vs. Future = Future – None = $(\beta_0 + \beta_4) - (\beta_0) = \beta_4$

None vs. Current = Current – None = $(\beta_0 + \beta_5) - (\beta_0) = \beta_5$

Future vs. Current = Current – Future = $(\beta_0 + \beta_5) - (\beta_0 + \beta_4) = \beta_5 - \beta_4 = -\beta_4 + \beta_5$

These values are then requested via the SAS ESTIMATE statements below…

```
TITLE1 'Eq 2.8: Adding Dementia Group';
TITLE2 'Using Manual Group Contrasts so Reference=None';
TITLE3 'sexMW, demNF, and demNC are all treated as continuous predictors';
PROC MIXED DATA=work.Chapter2 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
     MODEL cognition = age85 grip9 sexMW demNF demNC / CHISQ SOLUTION DDFM=BW;
     ODS OUTPUT CovParms=CovDem;
```

The first CONTRAST below includes all fixed effects, and thus tests the model R^2.
The second CONTRAST below includes only the new fixed effects, and thus tests the increment to the model R2 from adding dementia group. In this case this is also an "omnibus" ANOVA test for group.

```
     CONTRAST "Model R2 F-Test df=3" age85 1, grip9 1, sexmw 1, demNF 1, demNC 1;
     CONTRAST "Omnibus F-Test for Dementia Group df=2"          demNF 1, demNC 1;
```

The first 3 ESTIMATEs request predicted outcomes, so they include the intercept.
The last 3 ESTIMATES request slopes for group differences, so they do NOT include the intercept.

```
* Request conditional (adjusted) group means (hold age=85, grip=9, men);
  ESTIMATE "Intercept for None Group"    intercept 1 demNF 0 demNC 0; * Given as B0;
  ESTIMATE "Intercept for Future Group"  intercept 1 demNF 1 demNC 0; * Not given (B0+B4);
  ESTIMATE "Intercept for Current Group" intercept 1 demNF 0 demNC 1; * Not given (B0+B5);
* Request group differences (unconditional because there are no interactions);
  ESTIMATE "None   vs. Future Group"   demNF  1 demNC 0; * Given as B4;
  ESTIMATE "None   vs. Current Group"  demNF  0 demNC 1; * Given as B5;
  ESTIMATE "Future vs. Current Group"  demNF -1 demNC 1; * Not given (B5-B4);
RUN; TITLE1; TITLE2; TITLE3;
```

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Residual	88.0709	5.3401	16.49	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
4016.3	1	4018.3	4018.3	4019.9	4022.6	4023.6

Interpret each fixed effect below:
Intercept β_0 =

Main effect of Age β_1 =

Main effect of Grip Strength β_2 =

Main effect of Sex β_3 =

Main effect of DemNF β_4 =

Main effect of DemNC β_5 =

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	29.2643	0.6985	544	41.90	<.0001	is B0
age85	-0.4057	0.1189	544	-3.41	0.0007	is B1
grip9	0.6042	0.1498	544	4.03	<.0001	is B2
sexMW	-3.6574	0.8914	544	-4.10	<.0001	is B3
demNF	-5.7220	1.0191	544	-5.61	<.0001	is B4
demNC	-16.4798	1.5228	544	-10.82	<.0001	is B5

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept for None Group	29.2643	0.6985	544	41.90	<.0001	B0
Intercept for Future Group	23.5424	1.0785	544	21.83	<.0001	B0+B4
Intercept for Current Group	12.7845	1.5302	544	8.35	<.0001	B0+B5
None vs. Future Group	-5.7220	1.0191	544	-5.61	<.0001	B4
None vs. Current Group	-16.4798	1.5228	544	-10.82	<.0001	B5
Future vs. Current Group	-10.7578	1.7080	544	-6.30	<.0001	B5-B4

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Model R2 F-Test with df=5	5	544	41.75	<.0001
Omnibus F-Test for Dementia Group with df=2	2	544	67.06	<.0001

```
* Call macro to calculate R2 for overall model and change in R2;
  %ModelR2(CovBase=CovEmpty, CovFewer=CovAgeGripSex, CovMore=CovDem);
```

Name	CovParm	Estimate	StdErr	ZValue	ProbZ	R2_from_Base	R2_Increment
CovEmpty	Residual	120.76	7.2887	16.57	<.0001	-0.00000	.
CovAgeGripSex	Residual	109.38	6.6200	16.52	<.0001	0.09422	.
CovDem	Residual	88.0709	5.3401	16.49	<.0001	0.27069	0.17647

Previous model R^2 = (empty σ^2_e – previous σ^2_e) / (empty σ^2_e) = (120.76 – 109.38) / (120.76) = .09
Current model R^2 = (empty σ^2_e – current σ^2_e) / (empty σ^2_e) = (120.76 – 88.07) / (120.76) = .27
The df=5 CONTRAST above says that current R^2 is significantly > 0, $F(5,544) = 41.75$, $p < .0001$.

Change in model R^2 = (current R^2) − (previous R^2) = .27 − .09 = .18
The df=2 CONTRAST above says that change in R^2 is significantly > 0, $F(2,544) = 67.06$, $p < .0001$.

SAS MIXED Syntax and Output for Dementia Group Model in Equation 2.8

Using CLASS statement (SAS-coded contrasts instead of manually created contrasts)

Because the default reference group is the HIGHEST group numerically or last alphabetically, I have changed the model to reflect "Current" (group=3) as the reference:

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) + \beta_4(\text{DemCN}_i) + \beta_5(\text{DemCF}_i) + e_i$$

```
TITLE1 'Eq 2.8: Adding Dementia Group';
TITLE2 'Categorical Predictor for Dementia Group on CLASS statement';
PROC MIXED DATA=work.Chapter2 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=REML;
     CLASS demgroup; * CLASS statement demgroup replaces previous dem contrasts;
     MODEL cognition = age85 grip9 sexMW demgroup / SOLUTION CHISQ DDFM=BW;
     CONTRAST "Model R2 F-Test with df=5" age85 1, grip9 1, sexmw 1,
                                          demgroup -1 1 0, demgroup -1 0 1;
* Request conditional (adjusted) group means (hold age=85, grip=9, men) and all diffs;
     LSMEANS demgroup / DIFF=ALL AT(age85 grip9 sexMW) = (0 0 0);
* Request conditional (adjusted) group means and all differences;
     LSMEANS demgroup / DIFF=ALL;

*** All of the code below is redundant with LSMEANS, but here is how you get all the info;

  CONTRAST "Omnibus F-Test for Dementia Group with df=2"  demgroup -1 1 0, demgroup -1 0 1;
* Request conditional (adjusted) group means (hold age=85, grip=9, men);
  ESTIMATE "Intercept for None Group"    intercept 1 demgroup 1 0 0; * Not given (B0+B4);
  ESTIMATE "Intercept for Future Group"  intercept 1 demgroup 0 1 0; * Not given (B0+B5);
  ESTIMATE "Intercept for Current Group" intercept 1 demgroup 0 0 1; * Given as B0;
* Request group differences (unconditional because there are no interactions);
  ESTIMATE "None   vs. Future Group"   demgroup -1 1 0; * Not given (B5-B4);
  ESTIMATE "None   vs. Current Group"  demgroup -1 0 1; * Given as B5;
  ESTIMATE "Future vs. Current Group"  demgroup 0 -1 1; * Given as B4;
RUN; TITLE1; TITLE2;
```

New output after using the CLASS statement for demgroup:

The row with the dot indicates which group is the reference. The other rows then indicate group mean differences relative to the reference group.

Solution for Fixed Effects

Effect	demgroup: (1=None, 2=Future 3=Current)	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept		12.7845	1.5302	544	8.35	<.0001	new B0
age85		-0.4057	0.1189	544	-3.41	0.0007	same B1
grip9		0.6042	0.1498	544	4.03	<.0001	same B2
sexMW		-3.6574	0.8914	544	-4.10	<.0001	same B3
demgroup	1	16.4798	1.5228	544	10.82	<.0001	new B4
demgroup	2	10.7578	1.7080	544	6.30	<.0001	new B5
demgroup	3	0	.	.	.	.	new ref group

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
age85	1	544	11.65	11.65	0.0006	0.0007
grip9	1	544	16.27	16.27	<.0001	<.0001
sexMW	1	544	16.83	16.83	<.0001	<.0001
demgroup	**2**	**544**	**134.11**	**67.06**	**<.0001**	**<.0001**

Contrasts

Label	Num DF	Den DF	F Value	Pr > F	
Model R2 F-Test with df=5	5	544	41.75	<.0001	
Omnibus F-Test for Dementia Group with df=2	**2**	**544**	**67.06**	**<.0001**	**now given by default**

These group means and mean differences, which we requested, are given through LSMEANS below (with less typing).

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept for None Group	29.2643	0.6985	544	41.90	<.0001
Intercept for Future Group	23.5424	1.0785	544	21.83	<.0001
Intercept for Current Group	12.7845	1.5302	544	8.35	<.0001
None vs. Future Group	-5.7220	1.0191	544	-5.61	<.0001
None vs. Current Group	-16.4798	1.5228	544	-10.82	<.0001
Future vs. Current Group	-10.7578	1.7080	544	-6.30	<.0001

Least Squares Means

Effect	demgroup: (1=None, 2=Future, 3=Current)	age85	grip9	sexMW	Estimate	Standard Error	DF	t Value	Pr > |t|
demgroup	1	0.00	0.00	0.00	29.2643	0.6985	544	41.90	<.0001
demgroup	2	0.00	0.00	0.00	23.5424	1.0785	544	21.83	<.0001
demgroup	3	0.00	0.00	0.00	12.7845	1.5302	544	8.35	<.0001

The first three rows are adjusted group means at the specified levels of the other predictors.
The second three rows are group means at the sample mean values of the other predictors instead.

demgroup	1	-0.07	0.11	0.59	27.2143	0.4702	544	57.88	<.0001
demgroup	2	-0.07	0.11	0.59	21.4923	0.9024	544	23.82	<.0001
demgroup	3	-0.07	0.11	0.59	10.7345	1.4486	544	7.41	<.0001

Differences of Least Squares Means

Effect	demgroup: (1=None, 2=Future, 3=Current)	demgroup: (1=None, 2=Future, 3=Current)	age85	grip9	sexMW	Estimate	Standard Error	DF	t Value	Pr > |t|
demgroup	1	2	0.00	0.00	0.00	5.7220	1.0191	544	5.61	<.0001
demgroup	1	3	0.00	0.00	0.00	16.4798	1.5228	544	10.82	<.0001
demgroup	2	3	0.00	0.00	0.00	10.7578	1.7080	544	6.30	<.0001

The first three rows are adjusted group mean differences at the specified levels of the other predictors.
The second three rows are group mean differences at the sample mean values of the other predictors instead.
Because demgroup does not interact with these other predictors, the group mean differences are the same.

demgroup	1	2	-0.07	0.11	0.59	5.7220	1.0191	544	5.61	<.0001
demgroup	1	3	-0.07	0.11	0.59	16.4798	1.5228	544	10.82	<.0001
demgroup	2	3	-0.07	0.11	0.59	10.7578	1.7080	544	6.30	<.0001

So to CLASS or not to CLASS? Either can work in every circumstance. The use of CLASS for categorical predictors can be more convenient in models with more than one categorical predictor (e.g., to get marginal and cell means for factorial designs), whereas manual group contrasts can be more convenient when most other predictors are continuous, or when some of your effects pertain to only some levels of the grouping variable (i.e., nested effects; stay tuned).