This problem has elements of a two-way analysis of variance, or 2 x 2 factorial arrangements, of fruit fly sex and continent of residence. In addition to the factorial, there is a quantitative variable (latitude) whose slope is to be tested as the primary research hypothesis. A number of ways of fitting this model are examined below.

```
1          ************************************************************;
2          *** Wing size of the fruit fly Drosophila subovscura.     ***;
3          *** In europe this fruit fly has long been known to       ***;
4          *** display a "cline", that is a trend of increasing      ***;
5          *** wing size with latitude.  The fly was accidentally    ***;
6          *** introduced to North America around 1980.  After       ***;
7          *** about 10 latitudes the flies were examined to see if  ***;
8          *** a cline had developed, but not trend was found.       ***;
9          *** The data is from 20 latitudes after introduction.  It ***;
10         *** is for 11 locations in western NA and 10 sites in     ***;
11         *** Europe.  Is there now a cline for NS, and is it       ***;
12         *** the same as the cline for Europe?                     ***;
13         ************************************************************;
14
15         dm'log;clear;output;clear';
16         options nodate nocenter nonumber ps=512 ls=99 nolabel;
17         ODS HTML style=minimal rs=none
17       ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\WingSize01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\WingSize01.html
18
19         Title1 'Chapter 9 : Relationship of WingSize size to selected variables';
20         filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex0918.csv';
21
```

The original dataset had both male and female on the same line. To get male and females on different lines, I created two datasets, one called “original” that maintained the original dataset and a second modified dataset called “WingSize”. Note the “keep=” statements in the data step. These retain different variables for each of the two data sets. Also note that there are two “output” statements for each observation in the original data set, one for the males and one for the females. For each output we need an identifier for sex and we need to create single variables for wing size and standard error.

```
22         data WingSize (keep=continent latitude sex wingsize se)
23           Original (keep=continent latitude females seF males seM ratio seR);
24              infile input1 missover DSD dlm="," firstobs=2;
25            input continent $ latitude females seF males seM ratio seR;
26              label continent = 'North America (na) or Europe (eu)'
27                    latitude = 'North latitude (35 to 56 degrees)'
28                    females = 'Wing size for females (1000 log mm)'
29                    males = 'Wing size for males (1000 log mm)'
30                    ratio = 'Ratio of female basal length to Wing size';
31            *** se is standard error for females, males and the ratio **;
32            output original;
33            sex = 'Female'; Wingsize = females; se = sef; output wingsize;
34            sex = 'Male'; Wingsize = males; se = sem; output wingsize;
35         datalines;
NOTE: The infile INPUT1 is:
      File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex0918.csv, RECFM=V,LRECL=256
NOTE: 21 records were read from the infile INPUT1.
      The minimum record length was 31.
      The maximum record length was 35.
NOTE: The data set WORK.WINGSIZE has 42 observations and 5 variables.
NOTE: The data set WORK.ORIGINAL has 21 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.09 seconds
      cpu time            0.09 seconds
36         run;
37
```

Once the modified data set was created, with separate observations for male and female, additional modifications created a single letter variable to identify each group (m and f for males and females, lower case for North America and upper case for Europe). Also, indicator, or dummy, variables (0, 1) are created for each sex and continent combination and an interaction with the quantitative variable.

```
38         data wingsize; set wingsize;
39            if continent eq 'na' and sex eq 'Female' then ID = 'f';
40            if continent eq 'na' and sex eq 'Male'   then ID = 'm';
41            if continent eq 'eu' and sex eq 'Female' then ID = 'F';
42            if continent eq 'eu' and sex eq 'Male'   then ID = 'M';
43            if ID eq 'F' then group1 = 1; else group1 = 0;
44            if ID eq 'M' then group2 = 1; else group2 = 0;
45            if ID eq 'f' then group3 = 1; else group3 = 0;
46            if ID eq 'm' then group4 = 1; else group4 = 0;
47            X1 = group1*latitude;
48            X2 = group2*latitude;
49            X3 = group3*latitude;
50            X4 = group4*latitude;
51         run;
NOTE: There were 42 observations read from the data set WORK.WINGSIZE.
NOTE: The data set WORK.WINGSIZE has 42 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
52
53         PROC sort DATA=Wingsize; BY ID wingsize; RUN;
NOTE: There were 42 observations read from the data set WORK.WINGSIZE.
NOTE: The data set WORK.WINGSIZE has 42 observations and 14 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.05 seconds
      cpu time            0.05 seconds
54         PROC PRINT DATA=Original; TITLE2 'Raw data Listing'; RUN;
NOTE: There were 21 observations read from the data set WORK.ORIGINAL.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.18 seconds
      cpu time            0.06 seconds

Chapter 9 : Relationship of WingSize size to selected variables
Raw data Listing

Obs    continent    latitude    females    seF    males    seM    ratio     seR
  1       na          35.5        901      2.5     797     3.8    0.831    0.010
  2       na          37.0        896      3.5     806     3.0    0.834    0.014
  3       na          38.6        906      3.0     812     3.2    0.836    0.012
  4       na          40.7        907      3.5     807     3.2    0.833    0.013
  5       na          40.9        898      3.6     818     2.7    0.830    0.012
  6       na          42.4        893      3.4     809     3.3    0.828    0.015
  7       na          45.0        913      4.3     810     4.3    0.834    0.024
  8       na          46.8        915      3.8     819     3.3    0.825    0.014
  9       na          48.8        927      2.0     800     4.9    0.832    0.009
 10       na          49.8        924      4.5     823     2.2    0.824    0.011
 11       na          50.8        930      3.4     814     4.1    0.826    0.013
 12       eu          36.4        905      6.1     789     5.9    0.822    0.016
 13       eu          39.3        889      4.7     803     3.5    0.809    0.021
 14       eu          41.3        915      2.1     812     3.7    0.821    0.007
 15       eu          43.4        930      2.6     820     3.1    0.831    0.016
 16       eu          45.5        895      4.4     808     2.9    0.821    0.017
 17       eu          47.3        926      3.2     815     6.2    0.826    0.010
 18       eu          48.5        944      4.8     855     3.1    0.825    0.032
 19       eu          50.4        925      3.2     842     4.3    0.820    0.013
 20       eu          52.1        920      3.9     819     3.1    0.834    0.019
 21       eu          56.1        934      4.2     839     4.4    0.825    0.012
```

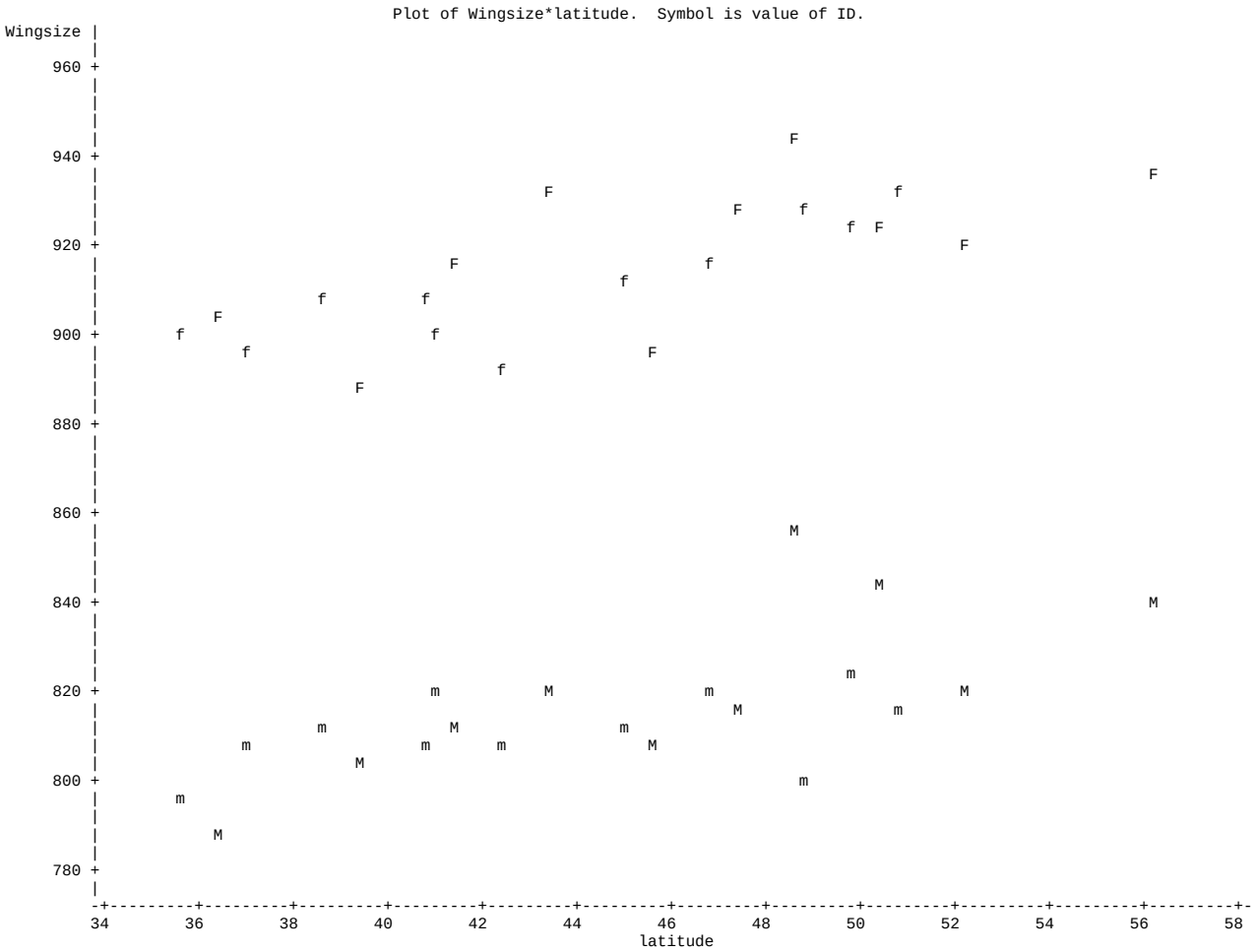
```
55         PROC PRINT DATA=WingSize; TITLE2 'Processed data Listing'; RUN;
NOTE: There were 42 observations read from the data set WORK.WINGSIZE.
NOTE: The PROCEDURE PRINT printed page 2.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.17 seconds
      cpu time            0.10 seconds
56

Chapter 9 : Relationship of WingSize size to selected variables
Processed data Listing

Obs continent latitude  sex   Wingsize  se ID group1 group2 group3 group4  X1   X2   X3   X4
  1    eu       39.3   Female    889   4.7 F     1      0      0      0   39.3  0.0  0.0  0.0
  2    eu       45.5   Female    895   4.4 F     1      0      0      0   45.5  0.0  0.0  0.0
  3    eu       36.4   Female    905   6.1 F     1      0      0      0   36.4  0.0  0.0  0.0
  4    eu       41.3   Female    915   2.1 F     1      0      0      0   41.3  0.0  0.0  0.0
  5    eu       52.1   Female    920   3.9 F     1      0      0      0   52.1  0.0  0.0  0.0
  6    eu       50.4   Female    925   3.2 F     1      0      0      0   50.4  0.0  0.0  0.0
  7    eu       47.3   Female    926   3.2 F     1      0      0      0   47.3  0.0  0.0  0.0
  8    eu       43.4   Female    930   2.6 F     1      0      0      0   43.4  0.0  0.0  0.0
  9    eu       56.1   Female    934   4.2 F     1      0      0      0   56.1  0.0  0.0  0.0
 10    eu       48.5   Female    944   4.8 F     1      0      0      0   48.5  0.0  0.0  0.0
 11    eu       36.4   Male      789   5.9 M     0      1      0      0    0.0 36.4  0.0  0.0
 12    eu       39.3   Male      803   3.5 M     0      1      0      0    0.0 39.3  0.0  0.0
 13    eu       45.5   Male      808   2.9 M     0      1      0      0    0.0 45.5  0.0  0.0
 14    eu       41.3   Male      812   3.7 M     0      1      0      0    0.0 41.3  0.0  0.0
 15    eu       47.3   Male      815   6.2 M     0      1      0      0    0.0 47.3  0.0  0.0
 16    eu       52.1   Male      819   3.1 M     0      1      0      0    0.0 52.1  0.0  0.0
 17    eu       43.4   Male      820   3.1 M     0      1      0      0    0.0 43.4  0.0  0.0
 18    eu       56.1   Male      839   4.4 M     0      1      0      0    0.0 56.1  0.0  0.0
 19    eu       50.4   Male      842   4.3 M     0      1      0      0    0.0 50.4  0.0  0.0
 20    eu       48.5   Male      855   3.1 M     0      1      0      0    0.0 48.5  0.0  0.0
 21    na       42.4   Female    893   3.4 f     0      0      1      0    0.0  0.0 42.4  0.0
 22    na       37.0   Female    896   3.5 f     0      0      1      0    0.0  0.0 37.0  0.0
 23    na       40.9   Female    898   3.6 f     0      0      1      0    0.0  0.0 40.9  0.0
 24    na       35.5   Female    901   2.5 f     0      0      1      0    0.0  0.0 35.5  0.0
 25    na       38.6   Female    906   3.0 f     0      0      1      0    0.0  0.0 38.6  0.0
 26    na       40.7   Female    907   3.5 f     0      0      1      0    0.0  0.0 40.7  0.0
 27    na       45.0   Female    913   4.3 f     0      0      1      0    0.0  0.0 45.0  0.0
 28    na       46.8   Female    915   3.8 f     0      0      1      0    0.0  0.0 46.8  0.0
 29    na       49.8   Female    924   4.5 f     0      0      1      0    0.0  0.0 49.8  0.0
 30    na       48.8   Female    927   2.0 f     0      0      1      0    0.0  0.0 48.8  0.0
 31    na       50.8   Female    930   3.4 f     0      0      1      0    0.0  0.0 50.8  0.0
 32    na       35.5   Male      797   3.8 m     0      0      0      1    0.0  0.0  0.0 35.5
 33    na       48.8   Male      800   4.9 m     0      0      0      1    0.0  0.0  0.0 48.8
 34    na       37.0   Male      806   3.0 m     0      0      0      1    0.0  0.0  0.0 37.0
 35    na       40.7   Male      807   3.2 m     0      0      0      1    0.0  0.0  0.0 40.7
 36    na       42.4   Male      809   3.3 m     0      0      0      1    0.0  0.0  0.0 42.4
 37    na       45.0   Male      810   4.3 m     0      0      0      1    0.0  0.0  0.0 45.0
 38    na       38.6   Male      812   3.2 m     0      0      0      1    0.0  0.0  0.0 38.6
 39    na       50.8   Male      814   4.1 m     0      0      0      1    0.0  0.0  0.0 50.8
 40    na       40.9   Male      818   2.7 m     0      0      0      1    0.0  0.0  0.0 40.9
 41    na       46.8   Male      819   3.3 m     0      0      0      1    0.0  0.0  0.0 46.8
 42    na       49.8   Male      823   2.2 m     0      0      0      1    0.0  0.0  0.0 49.8

57         options ps=65 ls=132;
58         proc plot data=WingSize;  TITLE2 'Plot of the raw data with ID variable';
59            plot WingSize * latitude = ID;
60         RUN;
60       !      OPTIONS PS=256;
61
62         Title2 'Fit of WingSize on Latitude one line';
NOTE: There were 42 observations read from the data set WORK.WINGSIZE.
NOTE: The PROCEDURE PLOT printed page 3.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.12 seconds
      cpu time            0.02 seconds
```

Chapter 9 : Relationship of WingSize size to selected variables
Plot of the raw data with ID variable




```
63         PROC REG DATA=WingSize lineprinter;
64            MODEL WingSize = latitude;
65         RUN;
66         Title2 'Fit of WingSize on Latitude separate intercepts';
NOTE: The PROCEDURE REG printed page 4.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.12 seconds
      cpu time            0.07 seconds
```

Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on Latitude one line

The REG Procedure
Model: MODEL1
Dependent Variable: Wingsize

Number of Observations Read 42
Number of Observations Used 42

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4621.44765	4621.44765	1.72	0.1971
Error	40	107425	2685.62571		
Corrected Total	41	112046			

Root MSE	51.82302	R-Square	0.0412
Dependent Mean	864.52381	Adj R-Sq	0.0173
Coeff Var	5.99440		

```
Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     1      780.53173       64.52572      12.10      <.0001
latitude      1        1.88323        1.43561       1.31      0.1971
```

Note the lack of a significant linear trend above, even though there is a discernable slope in the scatter plot. This is probably due to the relatively large error (MSE) 2685.63, making for a relatively large standard error 1.883. Separate intercepts (level adjustments for the 4 groups) have been added to the model below in the form of the 4 indicator variables. The adjustment of the levels, or intercepts, accounts for considerable variation, reducing the MSE to 125.35. This model has a significant slope, likely due to a much reduced standard error (0.320).

Note the "not full rank" message because we have only 3 degrees of freedom for level differences, but we added 4 variables (one too many). Also, recall that the "intercept" is a true intercept and that the other dummy variable values are **adjustments** to that intercept value (not additional intercept estimates).

```
67         PROC REG DATA=WingSize lineprinter;
68            MODEL WingSize = latitude group1 group2 group3 group4;
69         RUN;
70         Title2 'Fit of WingSize on Latitude separate slopes and intercepts';
NOTE: The PROCEDURE REG printed page 5.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.10 seconds
      cpu time            0.05 seconds

Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on Latitude separate intercepts

The REG Procedure
Model: MODEL1
Dependent Variable: Wingsize

Number of Observations Read          42
Number of Observations Used          42

Analysis of Variance
                                  Sum of           Mean
Source                     DF        Squares         Square    F Value    Pr > F
Model                       4         107409          26852     214.22    <.0001
Error                      37     4637.94684      125.34991
Corrected Total            41         112046

Root MSE             11.19598    R-Square     0.9586
Dependent Mean      864.52381    Adj R-Sq     0.9541
Coeff Var             1.29505

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some
statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.
NOTE: The following parameters have been set to 0, since the variables are a linear
combination of other variables as shown.

group4 =  Intercept - group1 - group2 - group3

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     B      732.83456       14.25644      51.40      <.0001
latitude      1        1.79261        0.31988       5.60      <.0001
group1        B      102.95163        4.96921      20.72      <.0001
group2        B        4.85163        4.96921       0.98      0.3352
group3        B       99.54545        4.77398      20.85      <.0001
group4        0              0              .          .          .
```

```
71         PROC REG DATA=WingSize lineprinter;
72            MODEL WingSize = latitude group1 group2 group3 group4 x1 x2 x3 x4;
73         RUN;
NOTE: The PROCEDURE REG printed page 6.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.12 seconds
      cpu time            0.06 seconds
```

This model includes the slope interactions. The "not full rank" message now has two lines, one for the intercepts and one for the slopes. As with the intercepts, there is only one true slope. The other dummy variable interaction values are **adjustments** to that slope value.

```
Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on Latitude separate slopes and intercepts

The REG Procedure
Model: MODEL1
Dependent Variable: Wingsize

Number of Observations Read          42
Number of Observations Used          42

Analysis of Variance
                                  Sum of           Mean
Source                    DF     Squares         Square    F Value    Pr > F
Model                      7      107913          15416     126.80    <.0001
Error                     34  4133.68477      121.57896
Corrected Total           41      112046

Root MSE             11.02629    R-Square     0.9631
Dependent Mean      864.52381    Adj R-Sq     0.9555
Coeff Var             1.27542

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some
statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.
NOTE: The following parameters have been set to 0, since the variables are a linear
combination of other variables as shown.

group4 =  Intercept - group1 - group2 - group3
    X4 =  latitude - X1 - X2 - X3

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     B      779.86417       28.73352      27.14      <.0001
latitude      B        0.70648        0.65913       1.07      0.2913
group1        B       56.32623       40.15128       1.40      0.1697
group2        B      -72.93864       40.15128      -1.82      0.0781
group3        B       39.43235       40.63533       0.97      0.3387
group4        0              0              .          .          .
X1            B        1.07735        0.89439       1.20      0.2367
X2            B        1.75441        0.89439       1.96      0.0580
X3            B        1.38829        0.93216       1.49      0.1456
X4            0              0              .          .          .


75         Title2 'Fit of WingSize on indicator variables with GLM';
76         PROC GLM DATA=WingSize; class id;
77            MODEL WingSize = latitude id id*latitude / solution;
78         RUN;
79
80         Title2 'Fit of WingSize on original variables with GLM';
NOTE: The PROCEDURE GLM printed pages 7-8.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.18 seconds
      cpu time            0.10 seconds
```

This model is the same as the PROC REG above. It is an AnCova with 4 separate intercepts (one intercept and 3 adjustments) and 4 separate slopes (one slope and 3 adjustments). However, this application used the "CLASS" statement from PROC GLM instead of the dummy variables that were created in the data step. The results are the same.

```
Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on indicator variables with GLM

The GLM Procedure

Class Level Information
Class         Levels    Values
ID                 4    F M f m

Number of Observations Read          42
Number of Observations Used          42


Dependent Variable: Wingsize
                                     Sum of
Source                       DF         Squares     Mean Square    F Value    Pr > F
Model                         7     107912.7914      15416.1131     126.80    <.0001
Error                        34       4133.6848        121.5790
Corrected Total              41     112046.4762

R-Square     Coeff Var      Root MSE    Wingsize Mean
0.963107      1.275417      11.02629         864.5238

Source                       DF       Type I SS     Mean Square    F Value    Pr > F
latitude                      1       4621.4476       4621.4476      38.01    <.0001
ID                            3     102787.0817      34262.3606     281.81    <.0001
latitude*ID                   3        504.2621        168.0874       1.38    0.2647

Source                       DF     Type III SS     Mean Square    F Value    Pr > F
latitude                      1     3772.720501     3772.720501      31.03    <.0001
ID                            3     1524.400042      508.133347       4.18    0.0127
latitude*ID                   3      504.262074      168.087358       1.38    0.2647

                                                Standard
Parameter              Estimate                    Error    t Value    Pr > |t|
Intercept          779.8641724 B     28.73352016      27.14      <.0001
latitude             0.7064751 B      0.65913492       1.07      0.2913
ID          F       56.3262256 B     40.15127904       1.40      0.1697
ID          M      -72.9386446 B     40.15127904      -1.82      0.0781
ID          f       39.4323542 B     40.63533390       0.97      0.3387
ID          m        0.0000000 B       .                .          .
latitude*ID F        1.0773529 B      0.89438968       1.20      0.2367
latitude*ID M        1.7544085 B      0.89438968       1.96      0.0580
latitude*ID f        1.3882933 B      0.93215755       1.49      0.1456
latitude*ID m        0.0000000 B       .                .          .

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to
solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.


81         PROC GLM DATA=WingSize; class sex continent;
82            MODEL WingSize = latitude sex continent sex*continent sex*latitude
83                  continent*latitude sex*continent*latitude / solution;
84         RUN;
85
86         Title2 'Fit of WingSize on original variables with GLM model reordered';
NOTE: The PROCEDURE GLM printed pages 9-10.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.18 seconds
      cpu time            0.12 seconds
```

Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on original variables with GLM

The GLM Procedure

Class Level Information

Class	Levels	Values
sex	2	Female Male
continent	2	eu na

Number of Observations Read	42
Number of Observations Used	42

Dependent Variable: Wingsize

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	107912.7914	15416.1131	126.80	<.0001
Error	34	4133.6848	121.5790		
Corrected Total	41	112046.4762			

R-Square	Coeff Var	Root MSE	Wingsize Mean
0.963107	1.275417	11.02629	864.5238

Source	DF	Type I SS	Mean Square	F Value	Pr > F
latitude	1	4621.4476	4621.4476	38.01	<.0001
sex	1	102613.7143	102613.7143	844.01	<.0001
continent	1	167.8953	167.8953	1.38	0.2481
sex*continent	1	5.4721	5.4721	0.05	0.8333
latitude*sex	1	21.7609	21.7609	0.18	0.6749
latitude*continent	1	158.3395	158.3395	1.30	0.2618
latitud*sex*continen	1	324.1617	324.1617	2.67	0.1117

Source	DF	Type III SS	Mean Square	F Value	Pr > F
latitude	1	3772.720501	3772.720501	31.03	<.0001
sex	1	1073.113225	1073.113225	8.83	0.0054
continent	1	118.440410	118.440410	0.97	0.3306
sex*continent	1	304.296390	304.296390	2.50	0.1229
latitude*sex	1	38.441861	38.441861	0.32	0.5776
latitude*continent	1	158.339480	158.339480	1.30	0.2618
latitud*sex*continen	1	324.161738	324.161738	2.67	0.1117

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		779.8641724	B	28.73352016	27.14	<.0001
latitude		0.7064751	B	0.65913492	1.07	0.2913
sex	Female	39.4323542	B	40.63533390	0.97	0.3387
sex	Male	0.0000000	B	.	.	.
continent	eu	-72.9386446	B	40.15127904	-1.82	0.0781
continent	na	0.0000000	B	.	.	.
sex*continent	Female eu	89.8325160	B	56.78248337	1.58	0.1229
sex*continent	Female na	0.0000000	B	.	.	.
sex*continent	Male eu	0.0000000	B	.	.	.
sex*continent	Male na	0.0000000	B	.	.	.
latitude*sex	Female	1.3882933	B	0.93215755	1.49	0.1456
latitude*sex	Male	0.0000000	B	.	.	.
latitude*continent	eu	1.7544085	B	0.89438968	1.96	0.0580
latitude*continent	na	0.0000000	B	.	.	.
latitud*sex*continen	Female eu	-2.0653489	B	1.26485802	-1.63	0.1117
latitud*sex*continen	Female na	0.0000000	B	.	.	.
latitud*sex*continen	Male eu	0.0000000	B	.	.	.
latitud*sex*continen	Male na	0.0000000	B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The models above and below are similar to the analyses above with 4 separate intercepts and 4 separate slopes. This application also uses the "CLASS" statement from PROC GLM, but here we retain the identity of the sexes and continents instead of using the generic "ID" separation. The overall results are still the same in that we account for exactly the same amount of variation.

However, with this separate the fit into these separate components (sex and continent) it is possible to determine if both contribute significantly to the model. The results below appear to indicate that the adjustment of sex accounts for nearly all of the variation in the adjustment for intercepts. The continents appear to account for very little of the variation. Since the sex effect appears to be significant and continents do not the model below was ordered with sex and its interaction first, followed by continent and then the sex by continent interaction terms. From this analysis is appears that all that is needed is a sex level adjustment, nothing more. This interpretation is based on the TYPE I SS.

```
87         PROC GLM DATA=WingSize; class sex continent;
88            MODEL WingSize = latitude sex sex*latitude continent continent*latitude
89                  sex*continent sex*continent*latitude / solution;
90         RUN;
91
NOTE: The PROCEDURE GLM printed pages 11-12.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.19 seconds
      cpu time            0.12 seconds


Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on original variables with GLM model reordered

The GLM Procedure

Class Level Information
Class          Levels    Values
sex                 2    Female Male
continent           2    eu na

Number of Observations Read          42
Number of Observations Used          42


Dependent Variable: Wingsize
                                    Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        7     107912.7914      15416.1131     126.80    <.0001
Error                       34       4133.6848        121.5790
Corrected Total             41     112046.4762

R-Square     Coeff Var      Root MSE    Wingsize Mean
0.963107      1.275417      11.02629         864.5238

Source                      DF       Type I SS     Mean Square    F Value    Pr > F
latitude                     1       4621.4476       4621.4476      38.01    <.0001
sex                          1     102613.7143     102613.7143     844.01    <.0001
latitude*sex                 1         15.6052         15.6052       0.13    0.7224
continent                    1        167.8953        167.8953       1.38    0.2481
latitude*continent           1        158.3395        158.3395       1.30    0.2618
sex*continent                1         11.6278         11.6278       0.10    0.7590
latitud*sex*continen         1        324.1617        324.1617       2.67    0.1117

Source                      DF     Type III SS     Mean Square    F Value    Pr > F
latitude                     1     3772.720501     3772.720501      31.03    <.0001
sex                          1     1073.113225     1073.113225       8.83    0.0054
latitude*sex                 1       38.441861       38.441861       0.32    0.5776
continent                    1      118.440410      118.440410       0.97    0.3306
latitude*continent           1      158.339480      158.339480       1.30    0.2618
sex*continent                1      304.296390      304.296390       2.50    0.1229
latitud*sex*continen         1      324.161738      324.161738       2.67    0.1117
```

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		779.8641724	B	28.73352016	27.14	<.0001
latitude		0.7064751	B	0.65913492	1.07	0.2913
sex	Female	39.4323542	B	40.63533390	0.97	0.3387
sex	Male	0.0000000	B	.	.	.
latitude*sex	Female	1.3882933	B	0.93215755	1.49	0.1456
latitude*sex	Male	0.0000000	B	.	.	.
continent	eu	-72.9386446	B	40.15127904	-1.82	0.0781
continent	na	0.0000000	B	.	.	.
latitude*continent	eu	1.7544085	B	0.89438968	1.96	0.0580
latitude*continent	na	0.0000000	B	.	.	.
sex*continent	Female eu	89.8325160	B	56.78248337	1.58	0.1229
sex*continent	Female na	0.0000000	B	.	.	.
sex*continent	Male eu	0.0000000	B	.	.	.
sex*continent	Male na	0.0000000	B	.	.	.
latitud*sex*continen	Female eu	-2.0653489	B	1.26485802	-1.63	0.1117
latitud*sex*continen	Female na	0.0000000	B	.	.	.
latitud*sex*continen	Male eu	0.0000000	B	.	.	.
latitud*sex*continen	Male na	0.0000000	B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

```
92         Title2 'Fit of WingSize - best model?';
93         PROC GLM DATA=WingSize; class sex continent;
94            MODEL WingSize = latitude sex / solution;
95            output out=next2 r=resid p=YHat;
96         RUN;
NOTE: The data set WORK.NEXT2 has 42 observations and 16 variables.
NOTE: The PROCEDURE GLM printed pages 13-14.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.20 seconds
      cpu time            0.12 seconds
```

Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize - best model?

The GLM Procedure

Class Level Information

Class	Levels	Values
sex	2	Female Male
continent	2	eu na

Number of Observations Read	42
Number of Observations Used	42

Dependent Variable: Wingsize

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	107235.1619	53617.5810	434.62	<.0001
Error	39	4811.3143	123.3670		
Corrected Total	41	112046.4762			

R-Square	Coeff Var	Root MSE	Wingsize Mean
0.957060	1.284762	11.10707	864.5238

Source	DF	Type I SS	Mean Square	F Value	Pr > F
latitude	1	4621.4476	4621.4476	37.46	<.0001
sex	1	102613.7143	102613.7143	831.78	<.0001

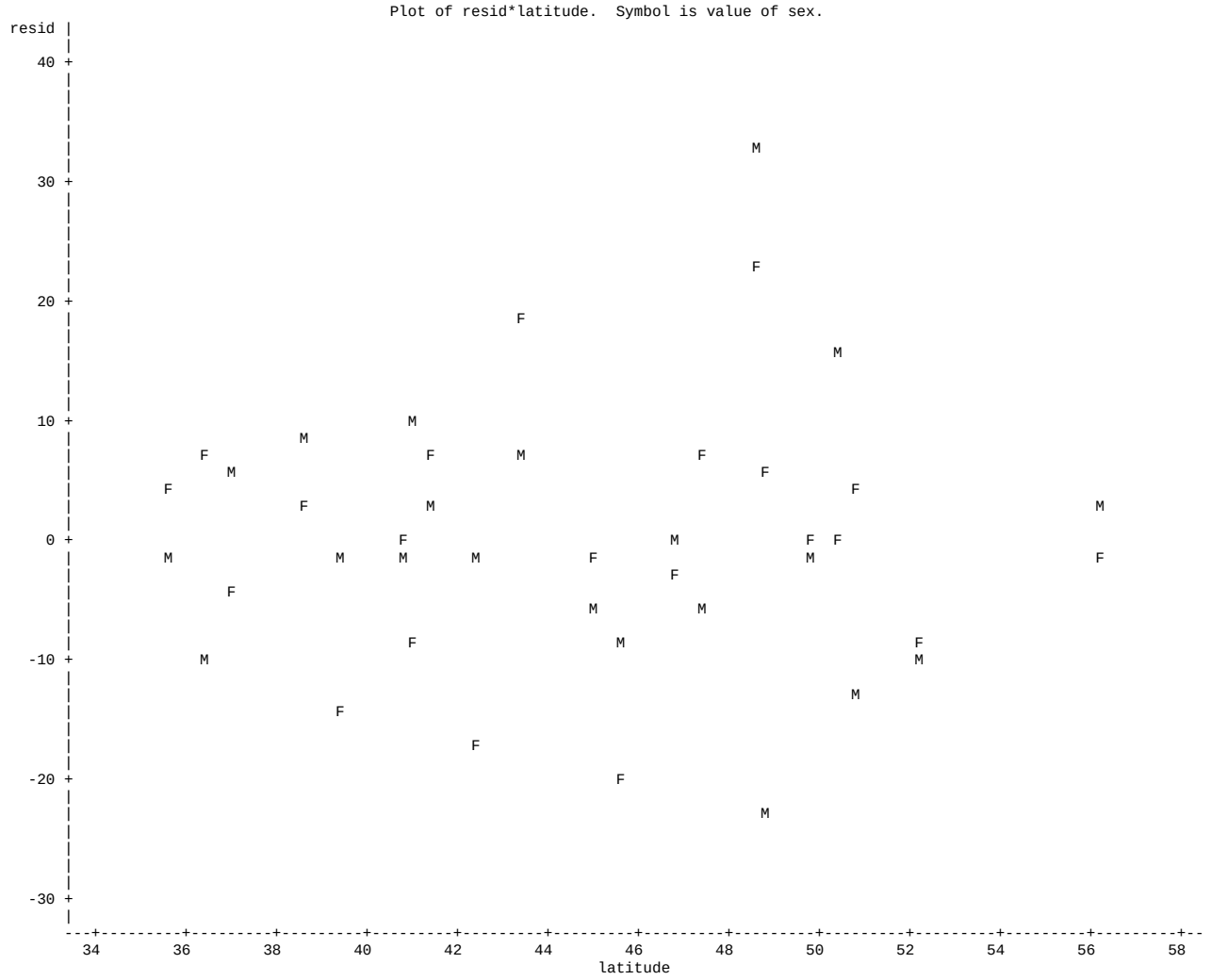
Source	DF	Type III SS	Mean Square	F Value	Pr > F
latitude	1	4621.4476	4621.4476	37.46	<.0001
sex	1	102613.7143	102613.7143	831.78	<.0001

```
                                                 Standard
Parameter                 Estimate                  Error    t Value    Pr > |t|
Intercept              731.1031578 B     13.93539338      52.46      <.000latitude
1.8832305        0.30769045       6.12       <.0001
sex       Female       98.8571429 B      3.42771661      28.84      <.0001
sex       Male          0.0000000 B       .                .          .

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to
solve the normal equations.  Terms whose estimates are followed by the letter 'B' are not
uniquely estimable.


98         options ps=65 ls=132;
99         proc plot data=next2;  TITLE2 'Residual plot with sex ID variable';
100           plot resid * latitude = sex;
101        RUN;
101      !      OPTIONS PS=256;
102
NOTE: There were 42 observations read from the data set WORK.NEXT2.
NOTE: The PROCEDURE PLOT printed page 15.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.10 seconds
      cpu time            0.02 seconds


Chapter 9 : Relationship of WingSize size to selected variables
Residual plot with sex ID variable
```



Examination of the assumptions of homogeneity (from the plot above) and normality (below) indicate possible problems with the first assumption but no issues with the second.

```
103        PROC UNIVARIATE DATA=NEXT2 NORMAL PLOT; VAR resid;RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 16.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.10 seconds
      cpu time            0.04 seconds


Chapter 9 : Relationship of WingSize size to selected variables
Residual plot with sex ID variable

The UNIVARIATE Procedure
Variable:  resid

Moments
N                          42    Sum Weights                42
Mean                        0    Sum Observations            0
Std Deviation      10.8327803    Variance           117.349128
Skewness           0.48361594    Kurtosis            1.3897594
Uncorrected SS     4811.31426    Corrected SS       4811.31426
Coeff Variation             .    Std Error Mean     1.67153428

Tests for Normality
Test                   --Statistic---     -----p Value------
Shapiro-Wilk           W     0.966654     Pr < W      0.2539
Kolmogorov-Smirnov     D     0.108445     Pr > D     >0.1500
Cramer-von Mises       W-Sq  0.084476     Pr > W-Sq   0.1825
Anderson-Darling       A-Sq  0.505494     Pr > A-Sq   0.2004

Extreme Observations
------Lowest-----        ------Highest-----
   Value      Obs             Value      Obs
-23.0048       33           9.87271       40
-20.6473        2          15.98203       19
-16.8093       21          18.30750        8
-14.9713        1          22.70302       10
-12.7713       39          32.56016       20

   Stem Leaf  Boxplot                      Normal Probability Plot
      3 3                      1     0        32.5+                                                *
      2                                          |                                              ++
      2 3                      1     0           |                                          *+++++
      1 68                     2     |           |                                      **+++
      1 0                      1     |           |                                  ++++
      0 5567778                7  +-----+        |                            +*******
      0 00023344               8  |  +  |        |                       +*****
     -0 4322222110            10  *-----*        |                  ******
     -0 99865                  5  +-----+        |              ****+
     -1 310                    3     |           |          *+***
     -1 75                     2     |           |       +++*+
     -2 31                     2     0       -22.5+  +*+++*
        ----+----+----+----+                        +----+----+----+----+----+----+----+----+----+----+----+
    Multiply Stem.Leaf by 10**+1                        -2        -1         0        +1        +2
```

Now consider one last issue. All of the models above fit a single intercept and slope and then estimate "differences" between those intercepts and slopes and other possible slopes and intercepts. Estimating slopes and intercept differences is the approach needed when we test for differences. However, once differences have been tested and the best model determined it is useful to obtain estimates of the actual slopes and intercepts. This could be done by fitting the models separately (e.g. BY SEX CONTINENT), but if the parameters are estimated this way we loose the pooled variance estimate and we loose power. We therefore would like to estimate the individual slopes and intercepts with standard errors in a single model. This is demonstrated below with the full model.

The program below uses PROC REG and fits 4 slopes and 4 intercepts. Since the model does not fit an intercept alone (see NOINT option) the 4 level adjustments fit 4 separate intercepts, not intercept adjustments. Also, the independent variable "latitude" is not fitted alone so the 4 interaction terms do not fit slope adjustments. Each interaction term fits its own slope. The standard errors are correct.

```
105        Title2 'Fit of WingSize on Latitude separate slopes and intercepts';
106        PROC REG DATA=WingSize lineprinter;
107           MODEL WingSize = group1 group2 group3 group4 x1 x2 x3 x4 / noint;
108        RUN;
109
NOTE: The PROCEDURE REG printed page 17.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.17 seconds
      cpu time            0.04 seconds


Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on Latitude separate slopes and intercepts

The REG Procedure
Model: MODEL1
Dependent Variable: Wingsize

Number of Observations Read          42
Number of Observations Used          42

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance
                                  Sum of           Mean
Source                     DF        Squares         Square    F Value    Pr > F
Model                       8       31498772        3937347    32385.1    <.0001
Error                      34     4133.68477      121.57896
Uncorrected Total          42       31502906

Root MSE             11.02629    R-Square     0.9999
Dependent Mean      864.52381    Adj R-Sq     0.9998
Coeff Var             1.27542

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
group1        1      836.19040       28.04479      29.82      <.0001
group2        1      706.92553       28.04479      25.21      <.0001
group3        1      819.29653       28.73352      28.51      <.0001
group4        1      779.86417       28.73352      27.14      <.0001
X1            1        1.78383        0.60454       2.95      0.0057
X2            1        2.46088        0.60454       4.07      0.0003
X3            1        2.09477        0.65913       3.18      0.0032
X4            1        0.70648        0.65913       1.07      0.2913


110        Title2 'Fit of WingSize on original variables with GLM';
111        PROC GLM DATA=WingSize; class sex continent;
112           MODEL WingSize = sex*continent sex*continent*latitude / noint solution;
113        RUN;
NOTE: Due to the presence of CLASS variables, an intercept is implicitly fitted.  R-Square
has been corrected for the mean.
116        ODS HTML close;
117        quit;
NOTE: The PROCEDURE GLM printed pages 18-19.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.21 seconds
      cpu time            0.10 seconds
```

The program below uses PROC GLM to fit 4 slopes and 4 intercepts. Again, the model does not fit an intercept alone (see NOINT option), so the 4 level adjustments fit 4 separate intercepts, instead of intercept adjustments. Also, the independent variable "latitude" is not fitted alone so the 4 interaction terms do not fit slope adjustments. Each interaction term fits its own slope. The standard errors are correct. Previously the standard errors on the adjustments have been "standard errors of differences" and not standard errors for actual slopes and intercepts.

```
Chapter 9 : Relationship of WingSize size to selected variables
Fit of WingSize on original variables with GLM

The GLM Procedure

Class Level Information
Class          Levels    Values
sex                 2    Female Male
continent           2    eu na

Number of Observations Read          42
Number of Observations Used          42


Dependent Variable: Wingsize

                                    Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        8     31498772.32      3937346.54    32385.1    <.0001
Error                       34         4133.68          121.58
Uncorrected Total           42     31502906.00

R-Square     Coeff Var      Root MSE    Wingsize Mean
0.963107      1.275417      11.02629         864.5238

Source                      DF       Type I SS     Mean Square    F Value    Pr > F
sex*continent                4     31494331.57      7873582.89    64761.1    <.0001
latitud*sex*continen         4         4440.74         1110.19       9.13    <.0001

Source                      DF     Type III SS     Mean Square    F Value    Pr > F
sex*continent                4     373743.1513      93435.7878     768.52    <.0001
latitud*sex*continen         4       4440.7425       1110.1856       9.13    <.0001

                                                        Standard
Parameter                                Estimate           Error    t Value    Pr > |t|
sex*continent        Female eu      836.1903980     28.04478612      29.82      <.0001
sex*continent        Female na      819.2965266     28.73352016      28.51      <.0001
sex*continent        Male eu        706.9255278     28.04478612      25.21      <.0001
sex*continent        Male na        779.8641724     28.73352016      27.14      <.0001
latitud*sex*continen Female eu        1.7838280      0.60454450       2.95      0.0057
latitud*sex*continen Female na        2.0947684      0.65913492       3.18      0.0032
latitud*sex*continen Male eu          2.4608836      0.60454450       4.07      0.0003
latitud*sex*continen Male na          0.7064751      0.65913492       1.07      0.2913
```