

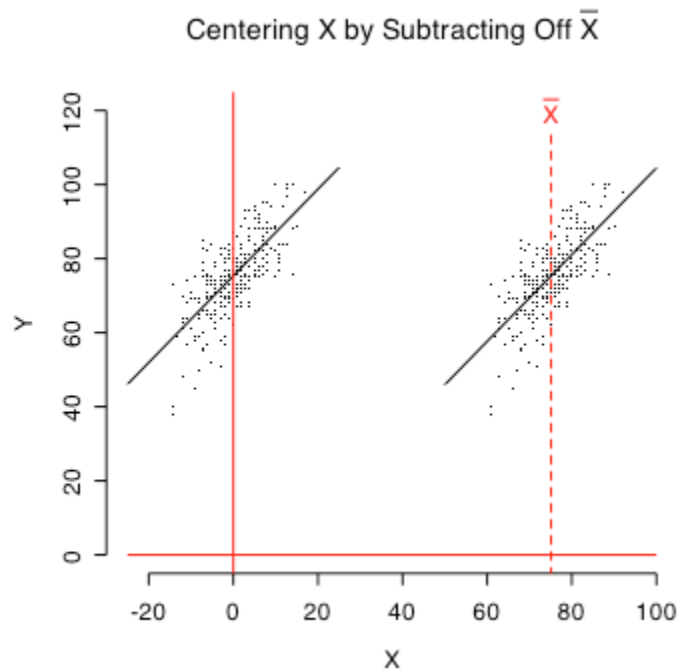
Centering and Interactions: The Training Data

A random sample of 150 technical support workers were first given a test of their technical skill and knowledge, and then randomly assigned to one of three training programmes. Six weeks later, their performance was assessed.

training.data

```
      group Skill Performance
1      1  83  60
2      1  81  57
3      1  78  51
4      1  70  53
5      1  95  72
6      1  81  58
7      1  68  39
8      1  74  53
...
147    3  71  36
148    3  75  44
149    3  77  38
150    3  73  44
```

There are 2 SAS programs. `training1.sas` illustrates some statistical ideas and SAS details. `training2.sas` is more focused on analyzing the data. Recall the idea of centering covariates at the mean.



```

/* training1.sas */
options linesize=79 pagesize=100 noovp formdlim=' ';
title 'Centering and Interactions (Customer Support Training)';

/* FIRST, DEMONSTRATE CENTERING THE SMART WAY, AND CREATION OF INTERACTION
TERMS. */

data support;
  infile 'training.data' firstobs=2;
  input id Group Skill Performance;
  cskill = skill;
  label cskill = 'Skill Minus Mean'; /* Center it below in proc standard */

proc standard out=support2 mean=0; /* Could also say std=1 to standardize */
  var cskill;

proc means;
  var skill cskill;

data support3;
  set support2;
  if group=. then g1=.;
  else if group=1 then g1=1;
  else g1=0;
  if group=. then g2=.;
  else if group=2 then g2=1;
  else g2=0;
  g1skill = g1*skill; g2skill = g2*skill;
  g1cskill = g1*cskill; g2cskill = g2*cskill;

/* NOW VERIFY THAT CENTERING AFFECTS ONLY THE INTERCEPTS, AND THAT FOR AN
EQUAL SLOPES MODEL, PROC GLM GIVES THE SAME RESULTS WHETHER THE
COVARIATE IS CENTERED OR NOT. */

proc reg;
  title2 'Equal slopes uncentered with proc reg';
  model performance = skill g1 g2;
  Group: test g1=g2=0;

proc reg;
  title2 'Equal slopes Centered with proc reg';
  model performance = cskill g1 g2;
  Group: test g1=g2=0;

proc glm;
  title2 'Equal slopes uncentered with proc glm';
  class group;
  model performance = skill group;
  lsmeans group / pdiff adjust=bon;

proc glm;
  title2 'Equal slopes Centered with proc glm';
  class group;
  model performance = cskill group;
  lsmeans group / pdiff adjust=bon;

```

```
/* NOW THE MODEL INCLUDES AN INTERACTION, ALLOWING FOR UNEQUAL
SLOPES. COMPARE RESULTS WITH THE COVARIATE CENTERED AND UNCENTERED. THE
MORAL OF THE STORY IS THAT YOU CAN GET WHAT YOU WANT EITHER WAY IF YOU KNOW
WHAT YOU'RE DOING, BUT IT CAN BE EASIER WITH CENTERED IVs. */
```

```
proc reg;
  title2 'Unequal Slopes Uncentered with proc reg';
  model performance = skill g1 g2 g1skill g2skill;
  GroupAtZero: test g1=g2=0;
  Interaction: test g1skill=g2skill=0;
  GroupAtMean: test g1 + 75*g1skill = g2 + 75*g2skill = 0;
  Group1vs2AtMean: test g1 + 75*g1skill = g2 + 75*g2skill;
```

```
proc reg;
  title2 'Unequal Slopes Centered with proc reg';
  model performance = cskill g1 g2 g1cskill g2cskill;
  GroupAtMean: test g1=g2=0;
  Group1vs2AtMean: test g1=g2;
  Interaction: test g1cskill=g2cskill=0;
```

```
/* HERE, WE SEE THAT WHEN PROC GLM TESTS A CATEGORICAL INDEPENDENT VARIABLE
IN THE PRESENCE OF COVARIATES, IT IS TESTING FOR DIFFERENCES BETWEEN
INTERCEPTS. THIS IS TRUE EVEN IF THE MODEL HAS AN INTERACTION BETWEEN
THE CATEGORICAL VARIABLE AND THE COVARIATE(S). WHEN THE COVARIATES ARE
UNCENTERED, THIS IS SELDOM WHAT YOU WANT, AND YOU NEED TO WATCH OUT. */
```

```
proc glm;
  title2 'Unequal slopes uncentered with proc glm';
  class group;
  model performance = skill group skill*group;
  lsmeans group / pdiff adjust=bon;
```

```
proc glm;
  title2 'Unequal slopes Centered with proc glm';
  class group;
  model performance = cskill group cskill*group;
  lsmeans group / pdiff adjust=bon;
```

The MEANS Procedure

| Variable | Label | N | Mean | Std Dev | Minimum |
|----------|------------------|-----|------------|-----------|-------------|
| Skill | | 150 | 75.0000000 | 7.0320455 | 58.0000000 |
| cskill | Skill Minus Mean | 150 | 0 | 7.0320455 | -17.0000000 |

| Variable | Label | Maximum |
|----------|------------------|------------|
| Skill | | 95.0000000 |
| cskill | Skill Minus Mean | 20.0000000 |

Just display parameter estimates for centered and uncentered data. R^2 , overall F tests and tests for Group are identical.

Equal slopes uncentered with proc reg
Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -19.51425 | 4.70654 | -4.15 | <.0001 |
| Skill | 1 | 0.76152 | 0.06229 | 12.23 | <.0001 |
| g1 | 1 | 13.12694 | 1.06923 | 12.28 | <.0001 |
| g2 | 1 | 7.33355 | 1.06891 | 6.86 | <.0001 |

Equal slopes Centered with proc reg
Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|------------------|----|--------------------|----------------|---------|---------|
| Intercept | Intercept | 1 | 37.59984 | 0.75589 | 49.74 | <.0001 |
| cskill | Skill Minus Mean | 1 | 0.76152 | 0.06229 | 12.23 | <.0001 |
| g1 | | 1 | 13.12694 | 1.06923 | 12.28 | <.0001 |
| g2 | | 1 | 7.33355 | 1.06891 | 6.86 | <.0001 |

For this equal slopes model, output from proc glm is identical whether the covariate is centered or uncentered. JUst show least squares means.

| Group | Performance LSMEAN | LSMEAN Number |
|-------|--------------------|---------------|
| 1 | 50.7267741 | 1 |
| 2 | 44.9333870 | 2 |
| 3 | 37.5998389 | 3 |

| Group | x_2 | x_3 | $E(Y \mathbf{x})$ |
|-------|-------|-------|--|
| 1 | 1 | 0 | $(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1$ |
| 2 | 0 | 1 | $(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1$ |
| 3 | 0 | 0 | $\beta_0 + \beta_1 x_1$ |

```
proc reg;
  title2 'Unequal Slopes Uncentered with proc reg';
  model performance = skill g1 g2 g1skill g2skill;
  GroupAtZero: test g1=g2=0;
  Interaction: test g1skill=g2skill=0;
  GroupAtMean: test g1 + 75*g1skill = g2 + 75*g2skill = 0;
  Group1vs2AtMean: test g1 + 75*g1skill = g2 + 75*g2skill;
```

Centering and Interactions (Customer Support Training)
 Unequal Slopes Uncentered with proc reg

12

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 150
 Number of Observations Used 150

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-----|----------------|-------------|---------|--------|
| Model | 5 | 9194.71754 | 1838.94351 | 67.38 | <.0001 |
| Error | 144 | 3929.82246 | 27.29043 | | |
| Corrected Total | 149 | 13125 | | | |

Root MSE 5.22402 R-Square 0.7006
 Dependent Mean 44.42000 Adj R-Sq 0.6902
 Coeff Var 11.76052

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -1.79639 | 8.41130 | -0.21 | 0.8312 |
| Skill | 1 | 0.52395 | 0.11235 | 4.66 | <.0001 |
| g1 | 1 | -21.30069 | 11.75037 | -1.81 | 0.0720 |
| g2 | 1 | -10.37307 | 11.17456 | -0.93 | 0.3548 |
| g1skill | 1 | 0.45954 | 0.15622 | 2.94 | 0.0038 |
| g2skill | 1 | 0.23742 | 0.14870 | 1.60 | 0.1125 |

The REG Procedure

Test GroupAtZero Results for Dependent
Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 44.88509 | 1.64 | 0.1967 |
| Denominator | 144 | 27.29043 | | |

Test Interaction Results for Dependent
Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 118.07192 | 4.33 | 0.0150 |
| Denominator | 144 | 27.29043 | | |

Test GroupAtMean Results for Dependent
Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 2171.92438 | 79.59 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Test Groupplvs2AtMean Results for
Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 1 | 820.33666 | 30.06 | <.0001 |
| Denominator | 144 | 27.29043 | | |

| Group | x_2 | x_3 | $E(Y \mathbf{x})$ |
|-------|-------|-------|--|
| 1 | 1 | 0 | $(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1$ |
| 2 | 0 | 1 | $(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1$ |
| 3 | 0 | 0 | $\beta_0 + \beta_1 x_1$ |

```
proc reg;
title2 'Unequal Slopes Centered with proc reg';
model performance = cskill g1 g2 glcskill g2cskill;
GroupAtMean: test g1=g2=0;
Group1vs2AtMean: test g1=g2;
Interaction: test glcskill=g2cskill=0;
```

Centering and Interactions (Customer Support Training)
 Unequal Slopes **Centered** with proc reg

17

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 150
 Number of Observations Used 150

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-----|----------------|-------------|---------|--------|
| Model | 5 | 9194.71754 | 1838.94351 | 67.38 | <.0001 |
| Error | 144 | 3929.82246 | 27.29043 | | |
| Corrected Total | 149 | 13125 | | | |

Root MSE 5.22402 R-Square 0.7006
 Dependent Mean 44.42000 Adj R-Sq 0.6902
 Coeff Var 11.76052

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|------------------|----|--------------------|----------------|---------|---------|
| Intercept | Intercept | 1 | 37.50006 | 0.74029 | 50.66 | <.0001 |
| cskill | Skill Minus Mean | 1 | 0.52395 | 0.11235 | 4.66 | <.0001 |
| g1 | | 1 | 13.16456 | 1.04631 | 12.58 | <.0001 |
| g2 | | 1 | 7.43335 | 1.04596 | 7.11 | <.0001 |
| glcskill | | 1 | 0.45954 | 0.15622 | 2.94 | 0.0038 |
| g2cskill | | 1 | 0.23742 | 0.14870 | 1.60 | 0.1125 |

Centering and Interactions (Customer Support Training)
 Unequal Slopes Centered with proc reg

18

The REG Procedure
 Model: MODEL1

Test GroupAtMean Results for Dependent
 Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 2171.92438 | 79.59 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Test Group1vs2AtMean Results for
 Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 1 | 820.33666 | 30.06 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Test Interaction Results for Dependent
 Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 118.07192 | 4.33 | 0.0150 |
| Denominator | 144 | 27.29043 | | |

| | Uncentered with proc reg | Centered with proc reg |
|-----------------|--------------------------|------------------------|
| GroupAtZero | F = 1.64 | |
| Interaction | F = 4.33 | F = 4.33 |
| GroupAtMean | F = 79.59 | F = 79.59 |
| Group1vs2AtMean | F = 30.06 | F = 30.06 |

Compare tests based on Type III Sums of Squares from proc glm

Unequal slopes uncentered with proc glm

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|-------------|----|-------------|-------------|---------|--------|
| Skill | 1 | 4144.661016 | 4144.661016 | 151.87 | <.0001 |
| Group | 2 | 89.770177 | 44.885088 | 1.64 | 0.1967 |
| Skill*Group | 2 | 236.143846 | 118.071923 | 4.33 | 0.0150 |

Unequal slopes Centered with proc glm

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------------|----|-------------|-------------|---------|--------|
| cskill | 1 | 4144.661016 | 4144.661016 | 151.87 | <.0001 |
| Group | 2 | 4343.848752 | 2171.924376 | 79.59 | <.0001 |
| cskill*Group | 2 | 236.143846 | 118.071923 | 4.33 | 0.0150 |

| | Uncentered with proc reg | Centered with proc reg |
|-----------------|--------------------------|------------------------|
| GroupAtZero | F = 1.64 | |
| Interaction | F = 4.33 | F = 4.33 |
| GroupAtMean | F = 79.59 | F = 79.59 |
| Group1vs2AtMean | F = 30.06 | F = 30.06 |

So proc glm is testing differences between intercepts, even when there is an interaction.

Centering and Interactions (Customer Support Training)

26

Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

| Group | Performance LSMEAN | LSMEAN Number |
|-------|-----------------------|------------------|
| 1 | 50.6646230 | 1 |
| 2 | 44.9334080 | 2 |
| 3 | 37.5000601 | 3 |

Least Squares Means for effect Group
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Performance

| i/j | 1 | 2 | 3 |
|-----|--------|--------|--------|
| 1 | | <.0001 | <.0001 |
| 2 | <.0001 | | <.0001 |
| 3 | <.0001 | <.0001 | |

And the multiple comparison tests of differences between least squares means are not follow-ups to the test for Group.

With proc glm, you should almost always center the covariates if the model includes interactions between factors and covariates.

```

/* training2.sas */
options linesize=79 noovp formdlim=' ';
title 'Customer Support Training';

/* THIS PROGRAM IS FOCUSED ON UNDERSTANDING THE DATA. training1.sas WAS
MORE ABOUT STATISTICAL IDEAS AND SOME SAS DETAILS */

data support;
  infile 'training.data' firstobs=2;
  input id Group Skill Performance;
  cskill = skill;
  label cskill = 'Skill Minus Mean'; /* Center it below in proc standard */

proc standard out=support2 mean=0; /* Could also say std=1 to standardize */
  var cskill;

data support3;
  set support2;
  if group=. then g1=.;
  else if group=1 then g1=1;
  else g1=0;
  if group=. then g2=.;
  else if group=2 then g2=1;
  else g2=0;
  g1skill = g1*skill; g2skill = g2*skill;
  g1cskill = g1*cskill; g2cskill = g2*cskill;

/* EQUAL REGRESSIONS SAYS THAT THE THREE REGRESSION LINES ARE RIGHT ON TOP
OF EACH OTHER. THAT IS, THERE ARE NO DIFFERENCES AMONG TRAINING
PROGRAMMES FOR ANY SKILL LEVEL. THIS IS THE NULL HYPOTHESIS FOR THE
EQUAL SLOPES MODEL, TOO. */

proc reg data = support3;
  title2 'Test Equal Regressions (and equal slopes)';
  model performance = skill g1 g2 g1skill g2skill;
  EqualRegressions: test g1=g2=g1skill=g2skill = 0;
  Interaction: test g1skill=g2skill = 0;

proc iml;
  title2 'Proportion of remaining variation explained by';
  title3 'Unequal regressions and Unequal slopes';
  F = 41.74; s = 4; NminusP = 144; a1 = s*F/(NminusP + s*F);
  F = 4.33; s = 2; NminusP = 144; a2 = s*F/(NminusP + s*F);
  print "Unequal Regressions: " a1 ", Unequal Slopes: " a2;

/* LOOK TO SEE WHAT'S GOING ON. A CONVENIENT WAY TO GET THE THREE
REGRESSIONS IS TO JUST FIT THEM DIRECTLY. ONLY THE ESTIMATED REGRESSION
COEFFICIENTS ARE WHAT WE WOULD GET FROM A SINGLE REGRESSION MODEL WITH
PRODUCT TERMS. EVERYTHING ELSE IS "WRONG." */

proc sort;
  by group;
proc reg;
  title2 'Separate regressions';
  model performance = skill;
  by group; /* Data must be sorted by this variable. */

/* LOOK AT A ROUGH PLOT OF THE REGRESSION LINES OVER THE RANGE OF THE
DATA. OUTPUT FROM THE SEPARATE REGRESSIONS GIVES

Group 1: Yhat = -23.09708 + 0.98349*X
Group 2: Yhat = -12.16946 + 0.76137*X
Group 3: Yhat = -1.79639 + 0.52395*X

```

WHAT IS THE RANGE OF THE DATA? A REAL DATA ANALYSIS JOB
WOULD START WITH DESCRIPTIVE STATISTICS AND YOU'D KNOW THIS ALREADY. */

```
proc univariate;  
  var skill;
```

```
/* CREATE A SAS DATA SET OF POINTS TO PLOT. THIS WOULD BE BETTER IN R. */
```

```
data pts;  
  do skill = 60 to 95;  
    group = 1; Yhat = -23.09708 + 0.98349*skill; output;  
    group = 2; Yhat = -12.16946 + 0.76137*skill; output;  
    group = 3; Yhat = -1.79639 + 0.52395*skill; output;  
  end;  
  label Yhat = 'Predicted Performance';
```

```
options pagesize=500;  
proc print;  
  title2 'Look at the data set pts';
```

```
options pagesize=35;  
proc plot;  
  title2 'Rough Plot of the Three Regression Lines';  
  plot Yhat * skill = group; /* Plotting symbol is the value of group. */
```

```
/* ARE THE DIFFERENCES BETWEEN TRAINING PROGRAMMES SIGNIFICANT EVEN AT LOW  
SKILL LEVELS? FROM PROC UNIVARIATE, MINIMUM IS 58 AND 25TH PERCENTILE IS 71. */
```

```
proc reg data = support3;  
  title2 'Test group differences at lower skill levels';  
  model performance = skill g1 g2 g1skill g2skill;  
  DiffAt58: test g1 + 58*g1skill = g2 + 58*g2skill = 0;  
  DiffAt71: test g1 + 71*g1skill = g2 + 71*g2skill = 0;  
  /* 67 is the 10th percentile. */  
  DiffAt67: test g1 + 67*g1skill = g2 + 67*g2skill = 0;  
  /* 64 is the 5th percentile. */  
  DiffAt64: test g1 + 64*g1skill = g2 + 64*g2skill = 0;  
  Group1vs2At64: test g1 + 64*g1skill = g2 + 64*g2skill;  
  Group1vs3At64: test g1 + 64*g1skill = 0;  
  Group2vs3At64: test g2 + 64*g2skill = 0;
```

```
/* HOW MIGHT THESE RESULTS BE DESCRIBED IN PLAIN LANGUAGE? YOU DON'T HAVE  
TO (AND SHOULD NOT) SAY EVERYTHING. HERE'S A POSSIBILITY.
```

"Average job performance depends on level of skill and technical knowledge prior to training. Naturally, those with higher prior levels of skill tend to perform better. Overall, average job performance was best for employees receiving Training Programme 1, followed by 2 and 3 in that order. The advantage of Programme 1 was greatest for those with higher levels of prior skill, but was still apparent for those with relatively low skill levels."

A HIGH-RESOLUTION PLOT OF THE THREE REGRESSION LINES WOULD BE GOOD, EVN FOR A NON-TECHNICAL AUDIENCE.

THERE ARE SOME MORE INTERESTNG ISSUES THAT COULD BE EXPLORED WITH THESE DATA. ONE EXAMPLE IS TESTS OF PAIRWISE DIFFERENCES BETWEEN SLOPES. ANOTHER ONE IS TO LOCATE THE EXACT SKILL LEVEL AT WHICH GROUP DIFFERNCES BECOME CLEAR, TREATING ALL TESTS AS SCHEFFE FOLLOW-UPS TO THE INITIAL TEST OF EQUAL REGRESSIONS. IT'S LIKELY NOT GOING TO BE AT THE 5TH PERCENTILE, BECAUSE OF THE PENALTY FOR PROTECTING INFINITELY MANY TESTS AT A SINGLE JOINT SIGNIFICANCE LEVEL. */

Customer Support Training
 Test Equal Regressions (and equal slopes)

1

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 150
 Number of Observations Used 150

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-----|----------------|-------------|---------|--------|
| Model | 5 | 9194.71754 | 1838.94351 | 67.38 | <.0001 |
| Error | 144 | 3929.82246 | 27.29043 | | |
| Corrected Total | 149 | 13125 | | | |

Root MSE 5.22402 R-Square 0.7006
 Dependent Mean 44.42000 Adj R-Sq 0.6902
 Coeff Var 11.76052

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -1.79639 | 8.41130 | -0.21 | 0.8312 |
| Skill | 1 | 0.52395 | 0.11235 | 4.66 | <.0001 |
| g1 | 1 | -21.30069 | 11.75037 | -1.81 | 0.0720 |
| g2 | 1 | -10.37307 | 11.17456 | -0.93 | 0.3548 |
| g1skill | 1 | 0.45954 | 0.15622 | 2.94 | 0.0038 |
| g2skill | 1 | 0.23742 | 0.14870 | 1.60 | 0.1125 |

Customer Support Training
 Test Equal Regressions (and equal slopes)

2

The REG Procedure
 Model: MODEL1

Test Equal Regressions Results for
 Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|-------------|---------|--------|
| Numerator | 4 | 1139.07990 | 41.74 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Test Interaction Results for Dependent
 Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|-------------|---------|--------|
| Numerator | 2 | 118.07192 | 4.33 | 0.0150 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training 4
 Proportion of remaining variation explained by
 Unequal regressions and Unequal slopes

a1 a2

Unequal Regressions: 0.5369179 , Unequal Slopes: 0.0567274

Customer Support Training 5
 Separate regressions

----- Group=1 -----

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 50
 Number of Observations Used 50

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 1 | 2240.23137 | 2240.23137 | 102.94 | <.0001 |
| Error | 48 | 1044.58863 | 21.76226 | | |
| Corrected Total | 49 | 3284.82000 | | | |

Root MSE 4.66500 R-Square 0.6820
 Dependent Mean 50.94000 Adj R-Sq 0.6754
 Coeff Var 9.15784

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -23.09708 | 7.32694 | -3.15 | 0.0028 |
| Skill | 1 | 0.98349 | 0.09693 | 10.15 | <.0001 |

Customer Support Training 6
 Separate regressions

----- Group=2 -----

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 50
 Number of Observations Used 50

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 1 | 1667.19066 | 1667.19066 | 49.87 | <.0001 |
| Error | 48 | 1604.72934 | 33.43186 | | |
| Corrected Total | 49 | 3271.92000 | | | |

| | | | |
|----------------|----------|----------|--------|
| Root MSE | 5.78203 | R-Square | 0.5095 |
| Dependent Mean | 45.04000 | Adj R-Sq | 0.4993 |
| Coeff Var | 12.83754 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -12.16946 | 8.14248 | -1.49 | 0.1416 |
| Skill | 1 | 0.76137 | 0.10782 | 7.06 | <.0001 |

Customer Support Training 7
 Separate regressions

----- Group=3 -----

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 50
 Number of Observations Used 50

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 1 | 593.57551 | 593.57551 | 22.25 | <.0001 |
| Error | 48 | 1280.50449 | 26.67718 | | |
| Corrected Total | 49 | 1874.08000 | | | |

| | | | |
|----------------|----------|----------|--------|
| Root MSE | 5.16500 | R-Square | 0.3167 |
| Dependent Mean | 37.28000 | Adj R-Sq | 0.3025 |
| Coeff Var | 13.85460 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -1.79639 | 8.31626 | -0.22 | 0.8299 |
| Skill | 1 | 0.52395 | 0.11108 | 4.72 | <.0001 |

Skipping proc univariate output except for the quantiles ...

| Quantile | Estimate |
|------------|----------|
| 100% Max | 95 |
| 99% | 91 |
| 95% | 87 |
| 90% | 84 |
| 75% Q3 | 80 |
| 50% Median | 74 |
| 25% Q1 | 71 |
| 10% | 67 |
| 5% | 64 |
| 1% | 59 |
| 0% Min | 58 |

Customer Support Training
Look at the data set pts

10

| Obs | skill | group | Yhat |
|-----|-------|-------|---------|
| 1 | 60 | 1 | 35.9123 |
| 2 | 60 | 2 | 33.5127 |
| 3 | 60 | 3 | 29.6406 |
| 4 | 61 | 1 | 36.8958 |
| 5 | 61 | 2 | 34.2741 |
| 6 | 61 | 3 | 30.1646 |
| 7 | 62 | 1 | 37.8793 |
| 8 | 62 | 2 | 35.0355 |
| 9 | 62 | 3 | 30.6885 |
| 10 | 63 | 1 | 38.8628 |
| 11 | 63 | 2 | 35.7969 |
| 12 | 63 | 3 | 31.2125 |
| 13 | 64 | 1 | 39.8463 |
| 14 | 64 | 2 | 36.5582 |
| 15 | 64 | 3 | 31.7364 |
| 16 | 65 | 1 | 40.8298 |
| 17 | 65 | 2 | 37.3196 |
| 18 | 65 | 3 | 32.2604 |
| 19 | 66 | 1 | 41.8133 |

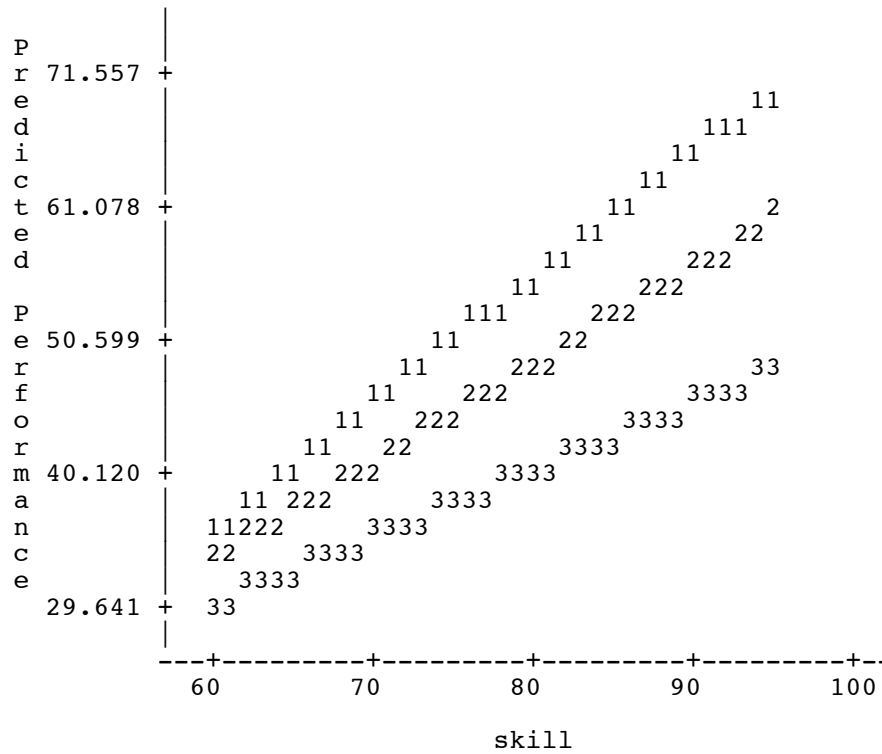
Skipping ...

| | | | |
|-----|----|---|---------|
| 100 | 93 | 1 | 68.3675 |
| 101 | 93 | 2 | 58.6380 |
| 102 | 93 | 3 | 46.9310 |
| 103 | 94 | 1 | 69.3510 |
| 104 | 94 | 2 | 59.3993 |
| 105 | 94 | 3 | 47.4549 |
| 106 | 95 | 1 | 70.3345 |
| 107 | 95 | 2 | 60.1607 |
| 108 | 95 | 3 | 47.9789 |

Customer Support Training
 Rough Plot of the Three Regression Lines

14

Plot of Yhat*skill. Symbol is value of group.



Customer Support Training
 Test group differences at lower skill levels

15

The REG Procedure
 Model: MODEL1
 Dependent Variable: Performance

Number of Observations Read 150
 Number of Observations Used 150

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-----|----------------|-------------|---------|--------|
| Model | 5 | 9194.71754 | 1838.94351 | 67.38 | <.0001 |
| Error | 144 | 3929.82246 | 27.29043 | | |
| Corrected Total | 149 | 13125 | | | |

Root MSE 5.22402 R-Square 0.7006
 Dependent Mean 44.42000 Adj R-Sq 0.6902
 Coeff Var 11.76052

Customer Support Training
Test group differences at lower skill levels

17

The REG Procedure
Model: MODEL1

Test DiffAt58 Results for Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 49.76774 | 1.82 | 0.1651 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training
Test group differences at lower skill levels

18

The REG Procedure
Model: MODEL1

Test DiffAt71 Results for Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 1204.88567 | 44.15 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training
Test group differences at lower skill levels

19

The REG Procedure
Model: MODEL1

Test DiffAt67 Results for Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 474.61887 | 17.39 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training
Test group differences at lower skill levels

20

The REG Procedure
Model: MODEL1

Test DiffAt64 Results for Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 2 | 227.79970 | 8.35 | 0.0004 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training 21
 Test group differences at lower skill levels

The REG Procedure
 Model: MODEL1

Test Group1vs2At64 Results for
 Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 1 | 78.28716 | 2.87 | 0.0925 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training 22
 Test group differences at lower skill levels

The REG Procedure
 Model: MODEL1

Test Group1vs3At64 Results for
 Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 1 | 448.28484 | 16.43 | <.0001 |
| Denominator | 144 | 27.29043 | | |

Customer Support Training 23
 Test group differences at lower skill levels

The REG Procedure
 Model: MODEL1

Test Group2vs3At64 Results for
 Dependent Variable Performance

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|-----|----------------|---------|--------|
| Numerator | 1 | 172.31830 | 6.31 | 0.0131 |
| Denominator | 144 | 27.29043 | | |