Statistical Inference Using SAS

SAS Code Workshop Series

Qiang Wu

Department of Biostatistics, East Carolina University

February 18, 2011

Analyzing Categorical Data

Survey Data

SAMPLE QUESTIONAIRE

For office use only

ID ☐☐☐☐☐

1. Age in years _____

(Questions 2–4: Please check the appropriate category.)

2. Gender _____ 1 = Male
_____ 2 = Female ☐

3. Race _____ 1 = White
_____ 2 = African American
_____ 3 = Hispanic
_____ 4 = Other ☐

4. Marital Status:
_____ 1 = Single
_____ 2 = Married
_____ 3 = Widowed
_____ 4 = Divorced ☐

5. Education Level:
_____ 1 = High school or less
_____ 2 = Two-year college
_____ 3 = Four-year college (BA or BS)
_____ 4 = Post-graduate degree ☐

For each of the following statements, please enter a NUMBER from the following list on the line to the LEFT of each question. Use the following codes:

1 = Strongly disagree 2 = Disagree 3 = No opinion
4 = Agree 5 = Strongly agree

_____ **6.** The president of the U.S. has been doing a good job. ☐

_____ **7.** The arms budget should be increased. ☐

_____ **8.** There should be more federal aid to big cities. ☐

Variable Description

Col.	Description	Var. Name	Type
1-3	Subject ID	ID	
4-5	Age in years	AGE	Quantitative
6	Gender	GENDER	Categorical
7	Race	RACE	Categorical
8	Marital status	MARITAL	Categorical
9	Education level	EDUCATION	Categorical
10	President doing good job	PRESIDENT	Likert Scale
11	Arms budget increased	ARMS	Likert Scale
12	Federal aid to cities	CITIES	Likert Scale

Data Format

```
PROC FORMAT;
     VALUE $SEXFMT '1' = 'Male'
                   '2' = 'Female'
                  OTHER = 'Miscoded';
     VALUE $RACE   '1' = 'White'
                   '2' = 'African Am.'
                   '3' = 'Hispanic'
                   '4' = 'Other';
     VALUE $OSCAR  '1' = 'Single'
                   '2' = 'Married'
                   '3' = 'Widowed'
                   '4' = 'Divorced';
     VALUE $EDUC   '1' = 'High Sch or Less'
                   '2' = 'Two Yr. College'
                   '3' = 'Four Yr. College'
                   '4' = 'Graduate Degree';
     VALUE LIKERT   1  = 'Str Disagree'
                    2  = 'Disagree'
                    3  = 'No Opinion'
                    4  = 'Agree'
                    5  = 'Str Agree';
RUN;
```

Data Input

```
DATA QUEST;
     INPUT ID          $ 1-3
           AGE           4-5
           GENDER      $   6
           RACE        $   7
           MARITAL     $   8
           EDUCATION   $   9
           PRESIDENT       10
           ARMS            11
           CITIES          12;
LABEL MARITAL    = "Marital Status"
      EDUCATION  = "Education Level"
      PRESIDENT  = "President Doing a Good Job"
      ARMS       = "Arms budget Increase"
      CITIES     = "Federal Aid to Cities";
FORMAT GENDER     $SEXFMT.
       RACE       $RACE.
       MARITAL    $OSCAR.
       EDUCATION  $EDUC.
       PRESIDENT ARMS CITIES LIKERT.;
DATALINES;
001091111232
002452222422
003351324442
004271111121
005682132333
006651243425
;
RUN;
```

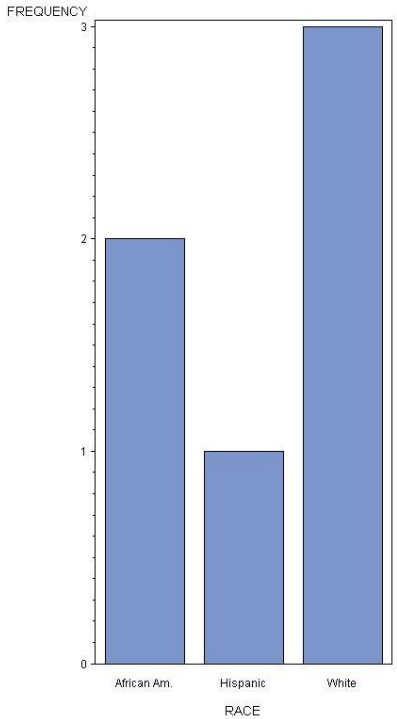
Charts and Frequency Tables

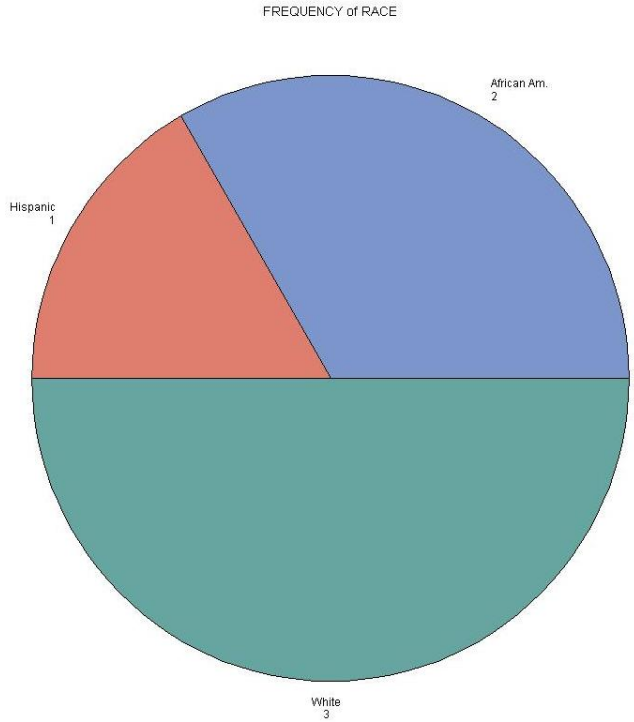
```
PROC GCHART DATA=QUEST;
     VBAR GENDER RACE EDUCATION;
     PIE GENDER RACE EDUCATION;
RUN;

PROC FREQ DATA=QUEST;
     TITLE "Frequency Counts for Categorical Variables";
     TABLES GENDER RACE MARITAL EDUCATION
            PRESIDENT ARMS CITIES;
RUN;

PROC FREQ DATA=QUEST;
     TITLE "One-sample test for binomial proportions";
     TABLES GENDER / BINOMIAL;
RUN;
```

Bar Charts and Pie Charts





Frequency Tables

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	4	66.67	4	66.67
Female	2	33.33	6	100.00

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
White	3	50.00	3	50.00
African Am.	2	33.33	5	83.33
Hispanic	1	16.67	6	100.00

Marital Status

MARITAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Single	2	33.33	2	33.33
Married	2	33.33	4	66.67
Widowed	1	16.67	5	83.33
Divorced	1	16.67	6	100.00

Education Level

EDUCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High Sch or Less	2	33.33	2	33.33
Two Yr. College	2	33.33	4	66.67
Four Yr. College	1	16.67	5	83.33
Graduate Degree	1	16.67	6	100.00

President Doing a Good Job

PRESIDENT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Str Disagree	1	16.67	1	16.67
Disagree	1	16.67	2	33.33
No Opinion	1	16.67	3	50.00
Agree	3	50.00	6	100.00

The FREQ Procedure

Arms budget Increase

ARMS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Disagree	3	50.00	3	50.00
No Opinion	2	33.33	5	83.33
Agree	1	16.67	6	100.00

Federal Aid to Cities

CITIES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Str Disagree	1	16.67	1	16.67
Disagree	3	50.00	4	66.67
No Opinion	1	16.67	5	83.33
Str Agree	1	16.67	6	100.00

Z-tests for Proportions

One-sample test for binomial proportions 3
13:59 Monday, May 10, 2010

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	4	66.67	4	66.67
Female	2	33.33	6	100.00

Binomial Proportion
for GENDER = Male

Proportion	0.6667
ASE	0.1925
95% Lower Conf Limit	0.2895
95% Upper Conf Limit	1.0000
Exact Conf Limits	
95% Lower Conf Limit	0.2228
95% Upper Conf Limit	0.9567

Test of H0: Proportion = 0.5

ASE under H0	0.2041
Z	0.8165
One-sided Pr > Z	0.2071
Two-sided Pr > |Z|	0.4142

Sample Size = 6

Two-Way Tables

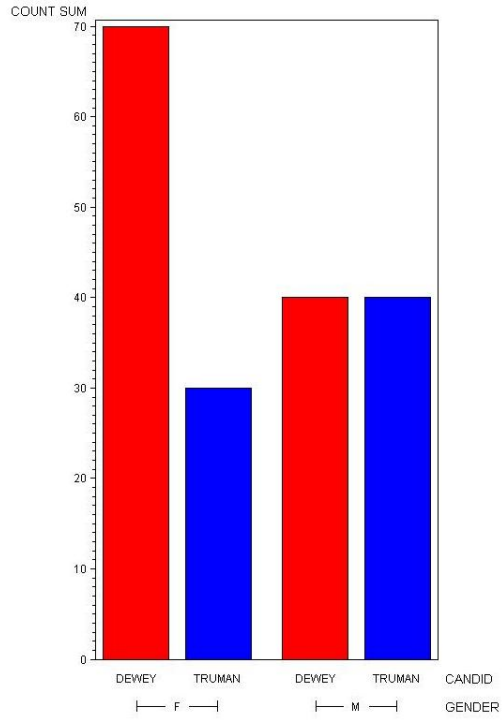
```
DATA ELECT;
     INPUT GENDER $ CANDID $ COUNT;
DATALINES;
M DEWEY 40
F DEWEY 70
M TRUMAN 40
F TRUMAN 30
;
RUN;

PROC GCHART DATA=ELECT;
     VBAR GENDER CANDID / SUMVAR=COUNT;
     PIE GENDER CANDID / SUMVAR=COUNT;
RUN;

PATTERN1 COLOR=RED;
PATTERN2 COLOR=BLUE;
PATTERN3 COLOR=GREEN;
PROC GCHART DATA=ELECT;
     VBAR CANDID / SUMVAR=COUNT GROUP=GENDER PATTERNID=MIDPOINT;
run;

PROC FREQ DATA=ELECT;
     TITLE "Two-way Tables and the Chi-square Test";
     TABLES GENDER CANDID;
     TABLES CANDID*GENDER / CHISQ;
     WEIGHT COUNT;
RUN;
```

Side-by-Side Bar Charts



Two-Way Tables and Chi-square Tests

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	100	55.56	100	55.56
M	80	44.44	180	100.00

CANDID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DEWEY	110	61.11	110	61.11
TRUMAN	70	38.89	180	100.00

Table of CANDID by GENDER

CANDID / GENDER	F	M	Total
DEWEY	70	40	110
	38.89	22.22	61.11
	63.64	36.36	
	70.00	50.00	
TRUMAN	30	40	70
	16.67	22.22	38.89
	42.86	57.14	
	30.00	50.00	
Total	100	80	180
	55.56	44.44	100.00

The FREQ Procedure

Statistics for Table of CANDID by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	7.4805	0.0062
Likelihood Ratio Chi-Square	1	7.4930	0.0062
Continuity Adj. Chi-Square	1	6.6626	0.0098
Mantel-Haenszel Chi-Square	1	7.4390	0.0064
Phi Coefficient		0.2039	
Contingency Coefficient		0.1998	
Cramer's V		0.2039	

Fisher's Exact Test

Cell (1,1) Frequency (F)	70
Left-sided Pr <= F	0.9981
Right-sided Pr >= F	0.0049
Table Probability (P)	0.0030
Two-sided Pr <= P	0.0087

Sample Size = 180

Paired Comparisons

```
PROC FORMAT;
     VALUE $OPINION 'P' = 'Positive'
                    'N' = 'Negative';
RUN;

DATA MCNEMAR;
     LENGTH BEFORE AFTER $ 1;
     INPUT BEFORE $ AFTER $ COUNT;
     FORMAT BEFORE AFTER $OPINION.;
DATALINES;
P P 23
P N 30
N P 15
N N 32
;
RUN;

PROC FREQ DATA=MCNEMAR;
     TITLE "McNemar's Test for Paired Samples";
     TABLES BEFORE*AFTER / AGREE;
     WEIGHT COUNT;
RUN;
```

McNemar's Tests

McNemar's Test for Paired Samples 6
13:59 Monday, May 10, 2010

The FREQ Procedure

Table of BEFORE by AFTER

BEFORE AFTER

Frequency Percent Row Pct Col Pct	Negative	Positive	Total
Negative	32 32.00 68.09 51.61	15 15.00 31.91 39.47	47 47.00
Positive	30 30.00 56.60 48.39	23 23.00 43.40 60.53	53 53.00
Total	62 62.00	38 38.00	100 100.00

Statistics for Table of BEFORE by AFTER

McNemar's Test

Statistic (S)	5.0000
DF	1
Pr > S	0.0253

Simple Kappa Coefficient

Kappa	0.1128
ASE	0.0946
95% Lower Conf Limit	-0.0727
95% Upper Conf Limit	0.2983

Sample Size = 100

Odds Ratio

```
DATA ODDS;
     INPUT OUTCOME $ EXPOSURE $ COUNT;
DATALINES;
CASE 1-YES 50
CASE 2-NO 100
CONTROL 1-YES 20
CONTROL 2-NO 130
;
RUN;

PROC FREQ DATA=ODDS;
     TITLE "Program to Compute an Odds Ratio";
     TABLES EXPOSURE*OUTCOME / CHISQ CMH;
     WEIGHT COUNT;
RUN;
```

Odds Ratio

Program to Compute an Odds Ratio 13:59 Monday, May 10, 2010 7

The FREQ Procedure

Table of EXPOSURE by OUTCOME

EXPOSURE OUTCOME

Frequency Percent Row Pct Col Pct	CASE	CONTROL	Total
1-YES	50 16.67 71.43 33.33	20 6.67 28.57 13.33	70 23.33
2-NO	100 33.33 43.48 66.67	130 43.33 56.52 86.67	230 76.67
Total	150 50.00	150 50.00	300 100.00

Statistics for Table of EXPOSURE by OUTCOME

Statistic	DF	Value	Prob
Chi-Square	1	16.7702	<.0001
Likelihood Ratio Chi-Square	1	17.2071	<.0001
Continuity Adj. Chi-Square	1	15.6708	<.0001
Mantel-Haenszel Chi-Square	1	16.7143	<.0001
Phi Coefficient		0.2364	
Contingency Coefficient		0.2301	
Cramer's V		0.2364	

Fisher's Exact Test

Cell (1,1) Frequency (F)	50
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	3.125E-05
Table Probability (P)	2.220E-05
Two-sided Pr <= P	6.249E-05

Sample Size = 300

Program to Compute an Odds Ratio 13:59 Monday, May 10, 2010 8

The FREQ Procedure

Summary Statistics for EXPOSURE by OUTCOME

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	16.7143	<.0001
2	Row Mean Scores Differ	1	16.7143	<.0001
3	General Association	1	16.7143	<.0001

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	3.2500	1.8189	5.8070
	Logit	3.2500	1.8189	5.8070
Cohort (Col1 Risk)	Mantel-Haenszel	1.6429	1.3331	2.0246
	Logit	1.6429	1.3331	2.0246
Cohort (Col2 Risk)	Mantel-Haenszel	0.5055	0.3432	0.7446
	Logit	0.5055	0.3432	0.7446

Total Sample Size = 300

Stratified Tables

```
DATA ABILITY;
     INPUT GENDER $ RESULTS $ SLEEP $ COUNT;
DATALINES;
BOYS FAIL 1-LOW 20
BOYS FAIL 2-HIGH 15
BOYS PASS 1-LOW 100
BOYS PASS 2-HIGH 150
GIRLS FAIL 1-LOW 30
GIRLS FAIL 2-HIGH 25
GIRLS PASS 1-LOW 100
GIRLS PASS 2-HIGH 200
;
RUN;

PROC FREQ DATA=ABILITY;
     TITLE "Mantel-Haenszel Chi-square Test";
     TABLES GENDER*SLEEP*RESULTS / ALL;
     WEIGHT COUNT;
RUN;
```

Mantel-Haenszel Chi-square Tests

Mantel-Haenszel Chi-square Test 13:59 Monday, May 10, 2010 9

The FREQ Procedure

Table 1 of SLEEP by RESULTS
Controlling for GENDER=BOYS

SLEEP RESULTS

Frequency / Percent / Row Pct / Col Pct	FAIL	PASS	Total
1-LOW	20	100	120
	7.02	35.09	42.11
	16.67	83.33	
	57.14	40.00	
2-HIGH	15	150	165
	5.26	52.63	57.89
	9.09	90.91	
	42.86	60.00	
Total	35	250	285
	12.28	87.72	100.00

Statistics for Table 1 of SLEEP by RESULTS
Controlling for GENDER=BOYS

Statistic	DF	Value	Prob
Chi-Square	1	3.7013	0.0544
Likelihood Ratio Chi-Square	1	3.6494	0.0561
Continuity Adj. Chi-Square	1	3.0315	0.0817
Mantel-Haenszel Chi-Square	1	3.6883	0.0548
Phi Coefficient		0.1140	
Contingency Coefficient		0.1132	
Cramer's V		0.1140	

Fisher's Exact Test

Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	0.9819
Right-sided Pr >= F	0.0415
Table Probability (P)	0.0235
Two-sided Pr <= P	0.0674

Mantel-Haenszel Chi-square Test 13:59 Monday, May 10, 2010 10

The FREQ Procedure

Statistics for Table 1 of SLEEP by RESULTS
Controlling for GENDER=BOYS

Statistic	Value	ASE
Gamma	0.3333	0.1623
Kendall's Tau-b	0.1140	0.0598
Stuart's Tau-c	0.0739	0.0397
Somers' D C|R	0.0758	0.0407
Somers' D R|C	0.1714	0.0892
Pearson Correlation	0.1140	0.0598
Spearman Correlation	0.1140	0.0598
Lambda Asymmetric C|R	0.0000	0.0000
Lambda Asymmetric R|C	0.0417	0.0483
Lambda Symmetric	0.0323	0.0374
Uncertainty Coefficient C|R	0.0172	0.0179
Uncertainty Coefficient R|C	0.0094	0.0099
Uncertainty Coefficient Symmetric	0.0122	0.0127

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	2.0000	0.9777	4.0911
Cohort (Col1 Risk)	1.8333	0.9795	3.4313
Cohort (Col2 Risk)	0.9167	0.8349	1.0064

Sample Size = 285

Mantel-Haenszel Chi-square Tests

Mantel-Haenszel Chi-square Test 13:59 Monday, May 10, 2010 11

The FREQ Procedure

Table 2 of SLEEP by RESULTS
Controlling for GENDER=GIRLS

SLEEP RESULTS

Frequency Percent Row Pct Col Pct	FAIL	PASS	Total
1-LOW	30 8.45 23.08 54.55	100 28.17 76.92 33.33	130 36.62
2-HIGH	25 7.04 11.11 45.45	200 56.34 88.89 66.67	225 63.38
Total	55 15.49	300 84.51	355 100.00

Statistics for Table 2 of SLEEP by RESULTS
Controlling for GENDER=GIRLS

Statistic	DF	Value	Prob
Chi-Square	1	9.0106	0.0027
Likelihood Ratio Chi-Square	1	8.7000	0.0032
Continuity Adj. Chi-Square	1	8.1199	0.0044
Mantel-Haenszel Chi-Square	1	8.9852	0.0027
Phi Coefficient		0.1593	
Contingency Coefficient		0.1573	
Cramer's V		0.1593	

Fisher's Exact Test

Cell (1,1) Frequency (F)	30
Left-sided Pr <= F	0.9990
Right-sided Pr >= F	0.0025
Table Probability (P)	0.0015
Two-sided Pr <= P	0.0037

Mantel-Haenszel Chi-square Test 13:59 Monday, May 10, 2010 12

The FREQ Procedure

Statistics for Table 2 of SLEEP by RESULTS
Controlling for GENDER=GIRLS

Statistic	Value	ASE
Gamma	0.4118	0.1234
Kendall's Tau-b	0.1593	0.0549
Stuart's Tau-c	0.1111	0.0396
Somers' D C|R	0.1197	0.0425
Somers' D R|C	0.2121	0.0724
Pearson Correlation	0.1593	0.0549
Spearman Correlation	0.1593	0.0549
Lambda Asymmetric C|R	0.0000	0.0000
Lambda Asymmetric R|C	0.0385	0.0559
Lambda Symmetric	0.0270	0.0395
Uncertainty Coefficient C|R	0.0284	0.0193
Uncertainty Coefficient R|C	0.0187	0.0128
Uncertainty Coefficient Symmetric	0.0225	0.0153

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	2.4000	1.3404	4.2973
Cohort (Col1 Risk)	2.0769	1.2789	3.3728
Cohort (Col2 Risk)	0.8654	0.7792	0.9611

Sample Size = 355

Mantel-Haenszel Chi-square Tests

Mantel-Haenszel Chi-square Test 13:59 Monday, May 10, 2010 13

The FREQ Procedure

Summary Statistics for SLEEP by RESULTS
Controlling for GENDER

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	12.4770	0.0004
2	Row Mean Scores Differ	1	12.4770	0.0004
3	General Association	1	12.4770	0.0004

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.2289	1.4185	3.5024
	Logit	2.2318	1.4205	3.5064
Cohort (Col1 Risk)	Mantel-Haenszel	1.9775	1.3474	2.9021
	Logit	1.9822	1.3508	2.9087
Cohort (Col2 Risk)	Mantel-Haenszel	0.8891	0.8283	0.9544
	Logit	0.8936	0.8334	0.9582

Breslow-Day Test for
Homogeneity of the Odds Ratios

Chi-Square	0.1501
DF	1
Pr > ChiSq	0.6985

Total Sample Size = 640

T-tests and Nonparametric Comparisons

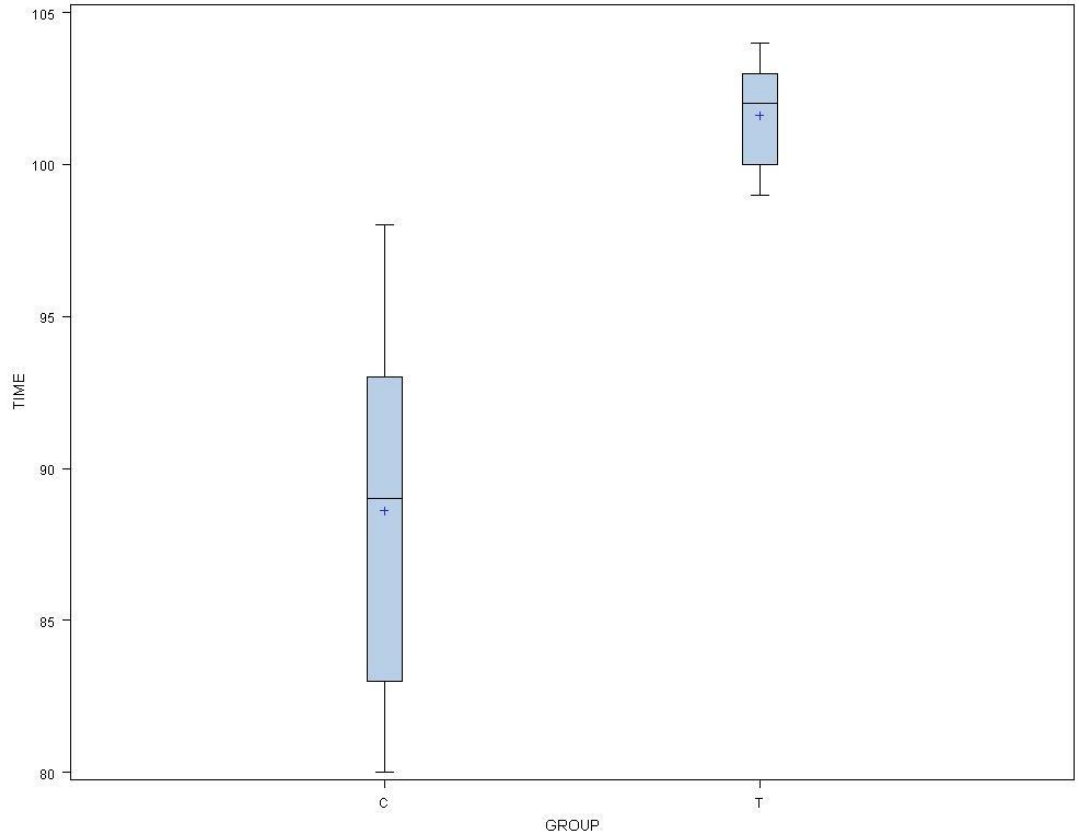
Comparing Two Groups

```
DATA RESPONSE;
     INPUT GROUP $ TIME;
DATALINES;
C 80
C 93
C 83
C 89
C 98
T 100
T 103
T 104
T 99
T 102
;
RUN;

PROC SORT DATA=RESPONSE;
     BY GROUP;
RUN;

PROC BOXPLOT DATA=RESPONSE;
     PLOT TIME*GROUP;
RUN;
```

Side-by-Side Boxplots



T-tests

```
PROC MEANS DATA=RESPONSE;
     CLASS GROUP;
     VAR TIME;
RUN;

PROC TTEST DATA=RESPONSE h0=100;
     TITLE "One-sample T-test Example";
     VAR TIME;
RUN;

PROC TTEST DATA=RESPONSE;
     TITLE "Two-sample T-test Example";
     CLASS GROUP;
     VAR TIME;
RUN;
```

T-tests

One-sample T-test Example 13:59 Monday, May 10, 2010 15

The TTEST Procedure

Variable: TIME

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	95.1000	8.5173	2.6934	80.0000	104.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
95.1000	89.0071	101.2	8.5173	5.8585	15.5493

DF	t Value	Pr > |t|
9	-1.82	0.1022

Two-sample T-test Example 13:59 Monday, May 10, 2010 16

The TTEST Procedure

Variable: TIME

GROUP	N	Mean	Std Dev	Std Err	Minimum	Maximum
C	5	88.6000	7.3007	3.2650	80.0000	98.0000
T	5	101.6	2.0736	0.9274	99.0000	104.0
Diff (1-2)		-13.0000	5.3666	3.3941		

GROUP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
C		88.6000	79.5350	97.6650	7.3007	4.3741	20.9789
T		101.6	99.0252	104.2	2.0736	1.2424	5.9587
Diff (1-2)	Pooled	-13.0000	-20.8268	-5.1732	5.3666	3.6249	10.2811
Diff (1-2)	Satterthwaite	-13.0000	-21.9317	-4.0683			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	-3.83	0.0050
Satterthwaite	Unequal	4.6412	-3.83	0.0141

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	4	12.40	0.0318

Wilcoxon Rank Sum Tests

```
DATA TUMOR;
     INPUT GROUP $ MASS @@;
DATALINES;
A 3.1 A 2.2 A 1.7 A 2.7 A 2.5
B 0.0 B 0.0 B 1.0 B 2.3
;
RUN;

PROC NPAR1WAY DATA=TUMOR WILCOXON;
     TITLE "Nonparametric Test to Compare Tumor Masses";
     CLASS GROUP;
     VAR MASS;
     EXACT WILCOXON;
RUN;
```

Wilcoxon Rank Sum Tests

Nonparametric Test to Compare Tumor Masses 17
13:59 Monday, May 10, 2010

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable MASS
Classified by Variable GROUP

GROUP	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
A	5	33.0	25.0	4.065437	6.60
B	4	12.0	20.0	4.065437	3.00

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic (S)	12.0000
Normal Approximation	
Z	-1.8448
One-Sided Pr < Z	0.0325
Two-Sided Pr > |Z|	0.0651
t Approximation	
One-Sided Pr < Z	0.0511
Two-Sided Pr > |Z|	0.1023
Exact Test	
One-Sided Pr <= S	0.0317
Two-Sided Pr >= |S - Mean|	0.0635

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test

Chi-Square	3.8723
DF	1
Pr > Chi-Square	0.0491

Paired T-tests

```
DATA PAIRED;
     INPUT CTIME TTIME;
DATALINES;
90 95
87 92
100 104
80 89
95 101
90 105
;
RUN;

PROC TTEST DATA=PAIRED;
     TITLE "Demonstrating a Paired T-test";
     PAIRED CTIME*TTIME;
RUN;
```

Paired T-tests

Demonstrating a Paired T-test 13:59 Monday, May 10, 2010 18

The TTEST Procedure

Difference: CTIME - TTIME

N	Mean	Std Dev	Std Err	Minimum	Maximum
6	-7.3333	4.1312	1.6865	-15.0000	-4.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-7.3333	-11.6687	-2.9979	4.1312	2.5787	10.1322

DF	t Value	Pr > |t|
5	-4.35	0.0074

Analysis of Variance

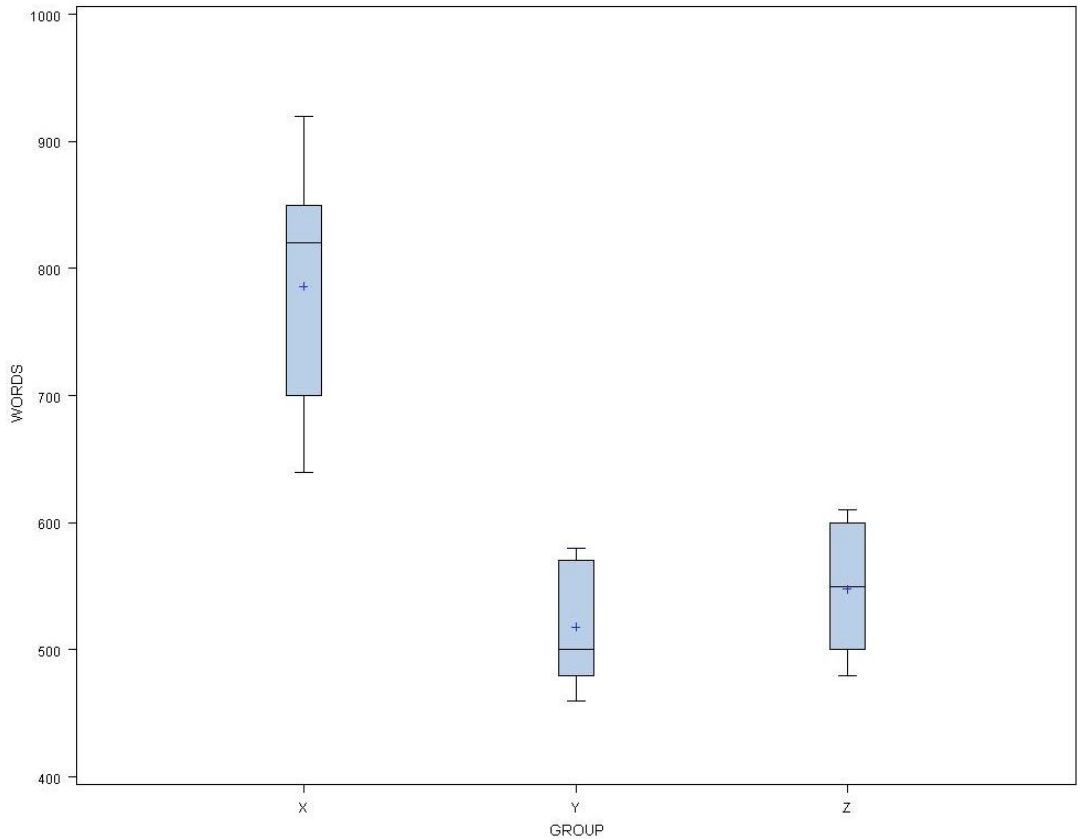
Comparing More Groups

```
DATA READING;
     INPUT GROUP $ WORDS @@;
DATALINES;
X 700 X 850 X 820 X 640 X 920
Y 480 Y 460 Y 500 Y 570 Y 580
Z 500 Z 550 Z 480 Z 600 Z 610
;
RUN;

PROC SORT DATA=READING;
     BY GROUP;
RUN;

PROC BOXPLOT DATA=READING;
     PLOT WORDS*GROUP;
RUN;
```

Side-by-Side Boxplots



One-way ANOVA

```
PROC MEANS DATA=READING;
     CLASS GROUP;
     VAR WORDS;
RUN;

PROC ANOVA DATA=READING;
     TITLE "Analysis of Reading Data";
     CLASS GROUP;
     MODEL WORDS = GROUP;
     MEANS GROUP / TUKEY;
RUN;

PROC NPAR1WAY DATA=READING WILCOXON;
     TITLE "The Kruskal-Wallis Test";
     CLASS GROUP;
     VAR WORDS;
     EXACT WILCOXON;
RUN;
```

One-Way ANOVA

Analysis of Reading Data 13:59 Monday, May 10, 2010 20

The ANOVA Procedure

Class Level Information

Class	Levels	Values
GROUP	3	X Y Z

Number of Observations Read	15
Number of Observations Used	15

Analysis of Reading Data 13:59 Monday, May 10, 2010 21

The ANOVA Procedure

Dependent Variable: WORDS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	215613.3333	107806.6667	16.78	0.0003
Error	12	77080.0000	6423.3333		
Corrected Total	14	292693.3333			

R-Square	Coeff Var	Root MSE	WORDS Mean
0.736653	12.98256	80.14570	617.3333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GROUP	2	215613.3333	107806.6667	16.78	0.0003

Multiple Comparisons

Analysis of Reading Data 13:59 Monday, May 10, 2010 22

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for WORDS

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	6423.333
Critical Value of Studentized Range	3.77293
Minimum Significant Difference	135.23

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	GROUP
A	786.00	5	X
B	548.00	5	Z
B			
B	518.00	5	Y

Kruskal-Wallis Tests

The Kruskal-Wallis Test 13:59 Monday, May 10, 2010 23

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable WORDS
Classified by Variable GROUP

GROUP	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
X	5	65.0	40.0	8.150372	13.00
Y	5	23.0	40.0	8.150372	4.60
Z	5	32.0	40.0	8.150372	6.40

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square		9.8151
DF		2
Asymptotic Pr >	Chi-Square	0.0074
Exact	Pr >= Chi-Square	0.0012

GLM and Contrasts

```
PROC GLM DATA=READING;
     TITLE "Analysis of Reading Data - Planned Comparisons";
     CLASS GROUP;
     MODEL WORDS = GROUP;
     LSMEANS GROUP / PDIFF ADJUST=TUKEY;
     CONTRAST 'X VS. Y AND Z' GROUP -2 1 1;
     CONTRAST 'METHOD Y VS. Z' GROUP 0 1 -1;
RUN;

PROC GLM DATA=READING;
     TITLE "Analysis of Reading Data - Comparisons Vs. a Control";
     CLASS GROUP;
     MODEL WORDS = GROUP;
     LSMEANS GROUP / PDIFF=CONTROL("X") ADJUST=DUNNETT;
RUN;
```

GLM

The GLM Procedure

Class Level Information

Class	Levels	Values
GROUP	3	X Y Z

Number of Observations Read	15
Number of Observations Used	15

The GLM Procedure

Dependent Variable: WORDS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	215613.3333	107806.6667	16.78	0.0003
Error	12	77080.0000	6423.3333		
Corrected Total	14	292693.3333			

R-Square	Coeff Var	Root MSE	WORDS Mean
0.736653	12.98256	80.14570	617.3333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
GROUP	2	215613.3333	107806.6667	16.78	0.0003

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	2	215613.3333	107806.6667	16.78	0.0003

Contrasts

Analysis of Reading Data - Planned Comparisons 26
13:59 Monday, May 10, 2010

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

GROUP	WORDS LSMEAN	LSMEAN Number
X	786.000000	1
Y	518.000000	2
Z	548.000000	3

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

Dependent Variable: WORDS

i/j	1	2	3
1		0.0005	0.0014
2	0.0005		0.8270
3	0.0014	0.8270	

Analysis of Reading Data - Planned Comparisons 27
13:59 Monday, May 10, 2010

The GLM Procedure

Dependent Variable: WORDS

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
X VS. Y AND Z	1	213363.3333	213363.3333	33.22	<.0001
METHOD Y VS. Z	1	2250.0000	2250.0000	0.35	0.5649

Contrasts vs. a Control

Analysis of Reading Data - Comparisons Vs. a Control 30
13:59 Monday, May 10, 2010

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Dunnett

GROUP	WORDS LSMEAN	HO:LSMean= Control Pr > |t|
X	786.000000	
Y	518.000000	0.0004
Z	548.000000	0.0010

Two-way ANOVA

```
DATA TWOWAY;
     INPUT GROUP $ GENDER $ WORDS @@;
DATALINES;
X M 700 X M 850 X M 820 X M 640 X M 920
Y M 480 Y M 460 Y M 500 Y M 570 Y M 580
Z M 920 Z M 550 Z M 480 Z M 600 Z M 610
X F 900 X F 880 X F 899 X F 780 X F 899
Y F 590 Y F 540 Y F 560 Y F 570 Y F 555
Z F 520 Z F 660 Z F 525 Z F 610 Z F 645
;
RUN;

PROC GLM DATA=TWOWAY;
     TITLE "Analysis of Reading Data";
     CLASS GROUP GENDER;
     MODEL WORDS = GROUP GENDER GROUP*GENDER;
     LSMEANS GROUP GENDER / PDIFF ADJUST=TUKEY;
     LSMEANS GROUP*GENDER / SLICE=GENDER;
RUN;
```

GLM

Analysis of Reading Data 13:59 Monday, May 10, 2010 31

The GLM Procedure

Class Level Information

Class	Levels	Values
GROUP	3	X Y Z
GENDER	2	F M

Number of Observations Read	30
Number of Observations Used	30

Analysis of Reading Data 13:59 Monday, May 10, 2010 32

The GLM Procedure

Dependent Variable: WORDS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	478152.1667	95630.4333	11.06	<.0001
Error	24	207459.2000	8644.1333		
Corrected Total	29	685611.3667			

R-Square	Coeff Var	Root MSE	WORDS Mean
0.697410	14.07770	92.97383	660.4333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
GROUP	2	450771.2667	225385.6333	26.07	<.0001
GENDER	1	6840.3000	6840.3000	0.79	0.3825
GROUP*GENDER	2	20540.6000	10270.3000	1.19	0.3221

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	2	450771.2667	225385.6333	26.07	<.0001
GENDER	1	6840.3000	6840.3000	0.79	0.3825
GROUP*GENDER	2	20540.6000	10270.3000	1.19	0.3221

LSmeans

Analysis of Reading Data 13:59 Monday, May 10, 2010 33

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

GROUP	WORDS LSMEAN	LSMEAN Number
X	828.800000	1
Y	540.500000	2
Z	612.000000	3

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

Dependent Variable: WORDS

i/j	1	2	3
1		<.0001	<.0001
2	<.0001		0.2186
3	<.0001	0.2186	

Analysis of Reading Data 13:59 Monday, May 10, 2010 34

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

GENDER	WORDS LSMEAN	HO:LSMean1= LSMean2 Pr > |t|
F	675.533333	0.3825
M	645.333333	

The “Slice” Option

Analysis of Reading Data 13:59 Monday, May 10, 2010 35

The GLM Procedure
Least Squares Means

GROUP	GENDER	WORDS LSMEAN
X	F	871.600000
X	M	786.000000
Y	F	563.000000
Y	M	518.000000
Z	F	592.000000
Z	M	632.000000

Analysis of Reading Data 13:59 Monday, May 10, 2010 36

The GLM Procedure
Least Squares Means

GROUP*GENDER Effect Sliced by GENDER for WORDS

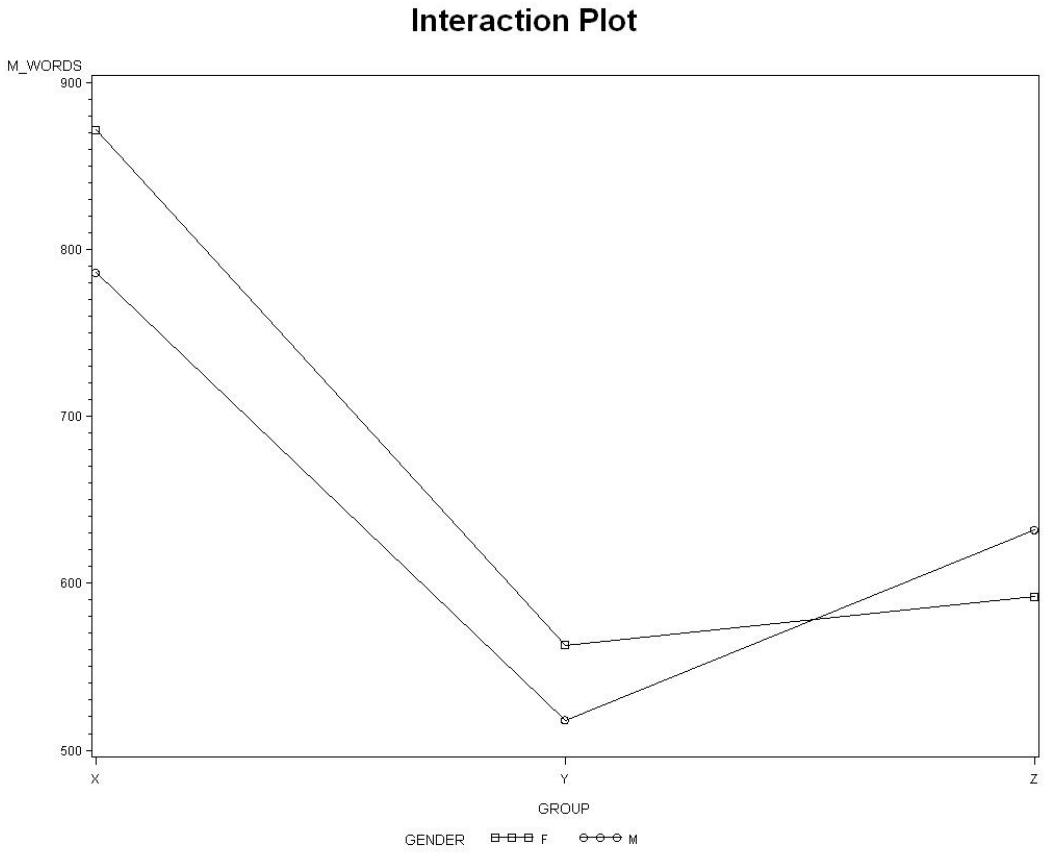
GENDER	DF	Sum of Squares	Mean Square	F Value	Pr > F
F	2	290419	145209	16.80	<.0001
M	2	180893	90447	10.46	0.0005

Interaction Plot

```
PROC MEANS DATA=TWOWAY NWAY NOPRINT;
     CLASS GROUP GENDER;
     VAR WORDS;
     OUTPUT OUT=MEANS MEAN=M_WORDS;
RUN;

SYMBOL1 V=SQUARE COLOR=BLACK I=JOIN;
SYMBOL2 V=CIRCLE COLOR=BLACK I=JOIN;
PROC GPLOT DATA=MEANS;
     TITLE "Interaction Plot";
     PLOT M_WORDS*GROUP=GENDER;
RUN;
```

Interaction Plot



Contrasts

```
PROC GLM DATA=TWOWAY;
     TITLE "Demonstrating the CONSTRAST Statement of GLM";
     CLASS GROUP GENDER;
     MODEL WORDS = GROUP GENDER GROUP*GENDER / SS3;
     CONTRAST 'X VS. Y FEMALE ONLY' GROUP 1 -1 0 GROUP*GENDER 1 0 -1 0 0 0;
     CONTRAST 'X VS. Y MALE ONLY' GROUP 1 -1 0 GROUP*GENDER 0 1 0 -1 0 0;
RUN;
```

Contrasts

Demonstrating the CONSTRAST Statement of GLM 37
13:59 Monday, May 10, 2010

The GLM Procedure

Class Level Information

Class	Levels	Values
GROUP	3	X Y Z
GENDER	2	F M

Number of Observations Read	30
Number of Observations Used	30

Demonstrating the CONSTRAST Statement of GLM 38
13:59 Monday, May 10, 2010

The GLM Procedure

Dependent Variable: WORDS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	478152.1667	95630.4333	11.06	<.0001
Error	24	207459.2000	8644.1333		
Corrected Total	29	685611.3667			

R-Square	Coeff Var	Root MSE	WORDS Mean
0.697410	14.07770	92.97383	660.4333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	2	450771.2667	225385.6333	26.07	<.0001
GENDER	1	6840.3000	6840.3000	0.79	0.3825
GROUP*GENDER	2	20540.6000	10270.3000	1.19	0.3221

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
X VS. Y FEMALE ONLY	1	238084.9000	238084.9000	27.54	<.0001
X VS. Y MALE ONLY	1	179560.0000	179560.0000	20.77	0.0001

ANCOVA Data

```
DATA COVAR;
     LENGTH GROUP $ 1;
     INPUT GROUP MATH IQ @@;
DATALINES;
A 260 105 A 325 115 A 300 122 A 400 125 A 390 138
B 325 126 B 440 135 B 425 142 B 500 140 B 600 160
;
RUN;

PROC CORR DATA=COVAR NOSIMPLE;
     TITLE "Covariate Example";
     VAR MATH IQ;
RUN;
PROC TTEST DATA=COVAR;
     CLASS GROUP;
     VAR IQ MATH;
RUN;
```

ANCOVA Data

Covariate Example 13:59 Monday, May 10, 2010 39

The CORR Procedure

2 Variables: MATH IQ

Pearson Correlation Coefficients, N = 10
Prob > |r| under H0: Rho=0

	MATH	IQ
MATH	1.00000	0.92456 0.0001
IQ	0.92456 0.0001	1.00000

Covariate Example 13:59 Monday, May 10, 2010 40

The TTEST Procedure

Variable: IQ

GROUP	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	5	121.0	12.2270	5.4681	105.0	138.0
B	5	140.6	12.4820	5.5821	126.0	160.0
Diff (1-2)		-19.6000	12.3552	7.8141		

GROUP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		121.0	105.8	136.2	12.2270	7.3256	35.1350
B		140.6	125.1	156.1	12.4820	7.4784	35.8677
Diff (1-2)	Pooled	-19.6000	-37.6193	-1.5807	12.3552	8.3454	23.6697
Diff (1-2)	Satterthwaite	-19.6000	-37.6207	-1.5793			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	-2.51	0.0365
Satterthwaite	Unequal	7.9966	-2.51	0.0365

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	4	1.04	0.9691

Variable: MATH

GROUP	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	5	335.0	59.5819	26.6458	260.0	400.0
B	5	458.0	101.3	45.2935	325.0	600.0
Diff (1-2)		-123.0	83.0888	52.5500		

GROUP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		335.0	261.0	409.0	59.5819	35.6975	171.2
B		458.0	332.2	583.8	101.3	60.6798	291.0
Diff (1-2)	Pooled	-123.0	-244.2	-1.8195	83.0888	56.1229	159.2
Diff (1-2)	Satterthwaite	-123.0	-249.3	3.3435			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	-2.34	0.0474
Satterthwaite	Unequal	6.4726	-2.34	0.0547

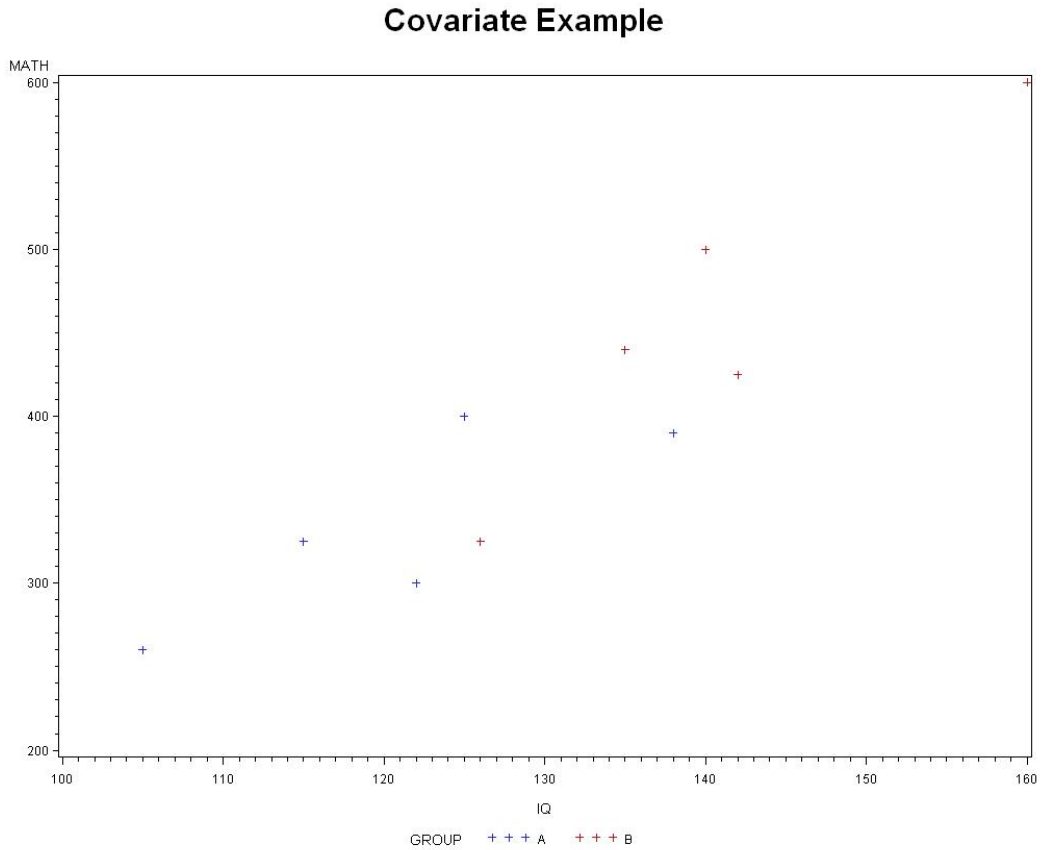
ANCOVA

```
PROC GPLOT DATA=COVAR;
     PLOT MATH*IQ=GROUP;
RUN;

PROC GLM DATA=COVAR;
     CLASS GROUP;
     MODEL MATH = IQ GROUP IQ*GROUP / SS3;
RUN;

PROC GLM DATA=COVAR;
     CLASS GROUP;
     MODEL MATH = IQ GROUP / SS3;
     LSMEANS GROUP / PDIFF;
RUN;
```

Scatter Plots



ANCOVA

Covariate Example 13:59 Monday, May 10, 2010 42

The GLM Procedure

Class Level Information

Class	Levels	Values
GROUP	2	A B

Number of Observations Read	10
Number of Observations Used	10

Covariate Example 13:59 Monday, May 10, 2010 43

The GLM Procedure

Dependent Variable: MATH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	83455.14993	27818.38331	17.39	0.0023
Error	6	9597.35007	1599.55834		
Corrected Total	9	93052.50000			

R-Square	Coeff Var	Root MSE	MATH Mean
0.896861	10.08688	39.99448	396.5000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
IQ	1	41278.21495	41278.21495	25.81	0.0023
GROUP	1	3634.41141	3634.41141	2.27	0.1824
IQ*GROUP	1	3816.96372	3816.96372	2.39	0.1734

ANCOVA

Covariate Example 13:59 Monday, May 10, 2010 44

The GLM Procedure

Class Level Information

Class	Levels	Values
GROUP	2	A B

Number of Observations Read	10
Number of Observations Used	10

Covariate Example 13:59 Monday, May 10, 2010 45

The GLM Procedure

Dependent Variable: MATH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	79638.18621	39819.09311	20.78	0.0011
Error	7	13414.31379	1916.33054		
Corrected Total	9	93052.50000			

R-Square	Coeff Var	Root MSE	MATH Mean
0.855841	11.04058	43.77591	396.5000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
IQ	1	41815.68621	41815.68621	21.82	0.0023
GROUP	1	96.59793	96.59793	0.05	0.8288

Correlation and Simple Regression

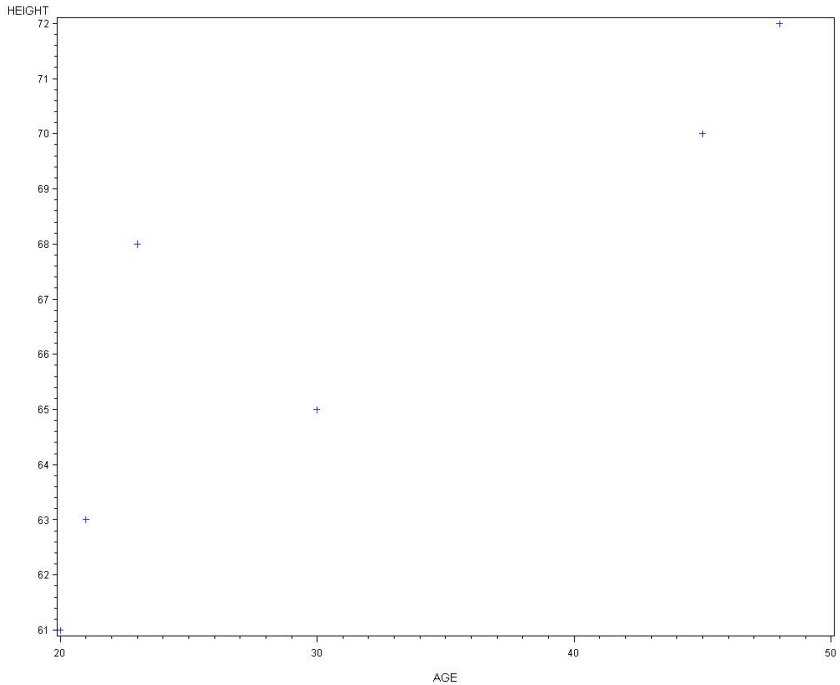
Correlations

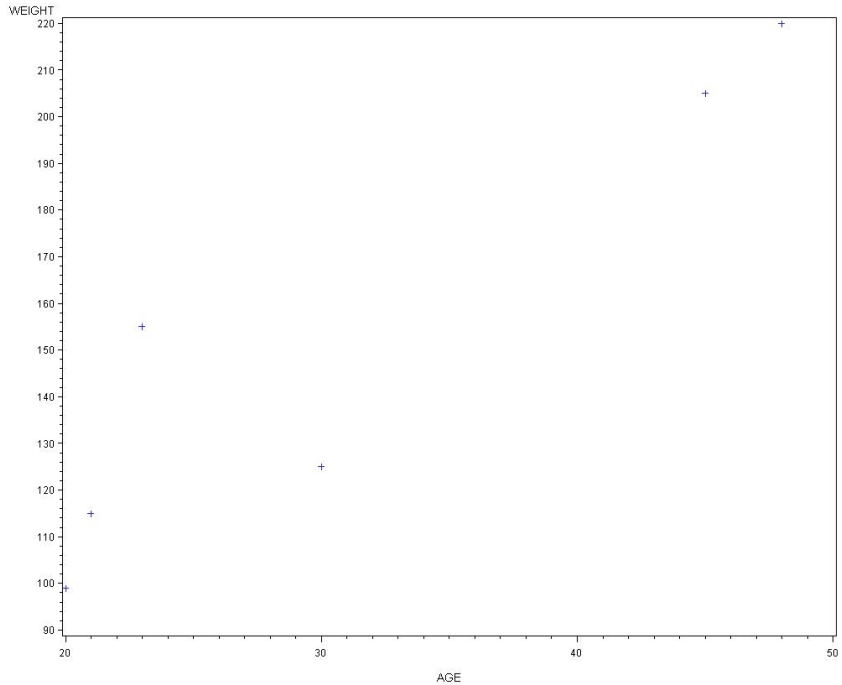
```
DATA CORR_EG;
     INPUT GENDER $ HEIGHT WEIGHT AGE;
DATALINES;
M 68 155 23
F 61 99 20
F 63 115 21
M 70 205 45
M 69 170 .
F 65 125 30
M 72 220 48
;
RUN;

GOPTIONS RESET=ALL;
PROC GPLOT DATA=CORR_EG;
     PLOT (HEIGHT WEIGHT)*AGE;
RUN;

PROC CORR DATA=CORR_EG PEARSON SPEARMAN;
     TITLE "Example of a Correlation Matrix";
     VAR HEIGHT WEIGHT AGE;
RUN;
```

Scatter Plots





Correlation Matrices

Example of a Correlation Matrix 13:59 Monday, May 10, 2010 47

The CORR Procedure

3 Variables: HEIGHT WEIGHT AGE

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
HEIGHT	7	66.85714	3.97612	68.00000	61.00000	72.00000
WEIGHT	7	155.57143	45.79613	155.00000	99.00000	220.00000
AGE	6	31.16667	12.41639	26.50000	20.00000	48.00000

Pearson Correlation Coefficients
Prob > |r| under HO: Rho=0
Number of Observations

	HEIGHT	WEIGHT	AGE
HEIGHT	1.00000	0.97165	0.86614
		0.0003	0.0257
	7	7	6
WEIGHT	0.97165	1.00000	0.92496
	0.0003		0.0082
	7	7	6
AGE	0.86614	0.92496	1.00000
	0.0257	0.0082	
	6	6	6

Spearman Correlation Coefficients
Prob > |r| under HO: Rho=0
Number of Observations

	HEIGHT	WEIGHT	AGE
HEIGHT	1.00000	1.00000	0.94286
		<.0001	0.0048
	7	7	6
WEIGHT	1.00000	1.00000	0.94286
	<.0001		0.0048
	7	7	6
AGE	0.94286	0.94286	1.00000
	0.0048	0.0048	
	6	6	6

Partial Correlations

```
PROC CORR DATA=CORR_EG PEARSON SPEARMAN;
     VAR AGE;
     WITH HEIGHT WEIGHT;
RUN;

PROC CORR DATA=CORR_EG NOSIMPLE;
     TITLE "Example of Partial Correlation";
     VAR HEIGHT WEIGHT;
     PARTIAL AGE;
RUN;
```

Partial Correlations

Example of a Correlation Matrix 13:59 Monday, May 10, 2010 48

The CORR Procedure

2 With Variables: HEIGHT WEIGHT
1 Variables: AGE

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
HEIGHT	7	66.85714	3.97612	68.00000	61.00000	72.00000
WEIGHT	7	155.57143	45.79613	155.00000	99.00000	220.00000
AGE	6	31.16667	12.41639	26.50000	20.00000	48.00000

Pearson Correlation Coefficients
Prob > |r| under HO: Rho=O
Number of Observations

	AGE
HEIGHT	0.86614 0.0257 6
WEIGHT	0.92496 0.0082 6

Spearman Correlation Coefficients
Prob > |r| under HO: Rho=O
Number of Observations

	AGE
HEIGHT	0.94286 0.0048 6
WEIGHT	0.94286 0.0048 6

Example of Partial Correlation 13:59 Monday, May 10, 2010 49

The CORR Procedure

1 Partial Variables: AGE
2 Variables: HEIGHT WEIGHT

Pearson Partial Correlation Coefficients, N = 6
Prob > |r| under HO: Partial Rho=0

	HEIGHT	WEIGHT
HEIGHT	1.00000	0.91934 0.0272
WEIGHT	0.91934 0.0272	1.00000

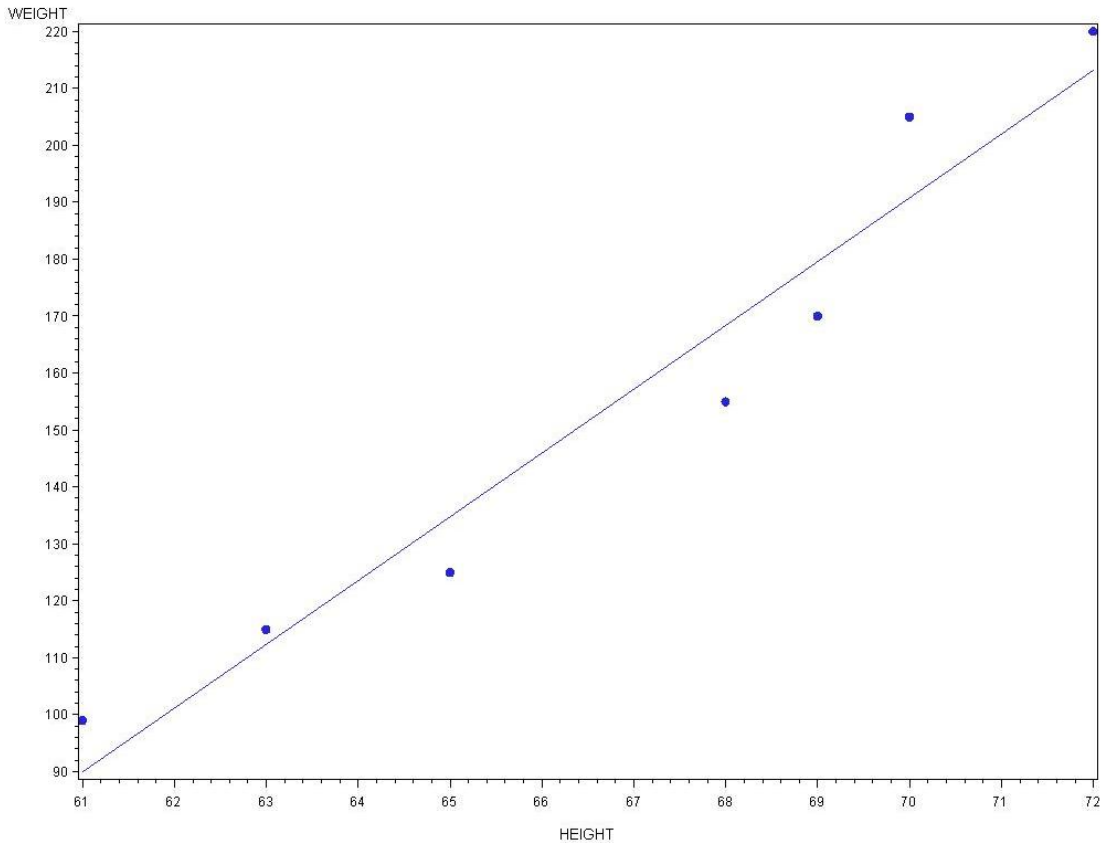
Regressions

```
SYMBOL1 V=DOT I=R;
PROC GPLOT DATA=CORR_EG;
     PLOT WEIGHT*HEIGHT;
RUN;

PROC REG DATA=CORR_EG;
     TITLE "Regression Line for Height-Weight Data";
     MODEL WEIGHT = HEIGHT;
     PLOT RESIDUAL.*HEIGHT;
RUN;

GOPTIONS CSYMBOL=BLUE;
SYMBOL1 VALUE=DOT;
SYMBOL2 VALUE=NONE I=RLCLM95;
SYMBOL3 VALUE=NONE I=RLCLI95 LINE=3;
PROC GPLOT DATA=CORR_EG;
     TITLE "Regression Lines and 95% CI's";
     PLOT WEIGHT*HEIGHT=1
          WEIGHT*HEIGHT=2
          WEIGHT*HEIGHT=3 / OVERLAY;
RUN;
```

Scatter Plot and Regression Line



Linear Regression

Regression Line for Height-Weight Data 50
13:59 Monday, May 10, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: WEIGHT

Number of Observations Read 7
Number of Observations Used 7

Analysis of Variance

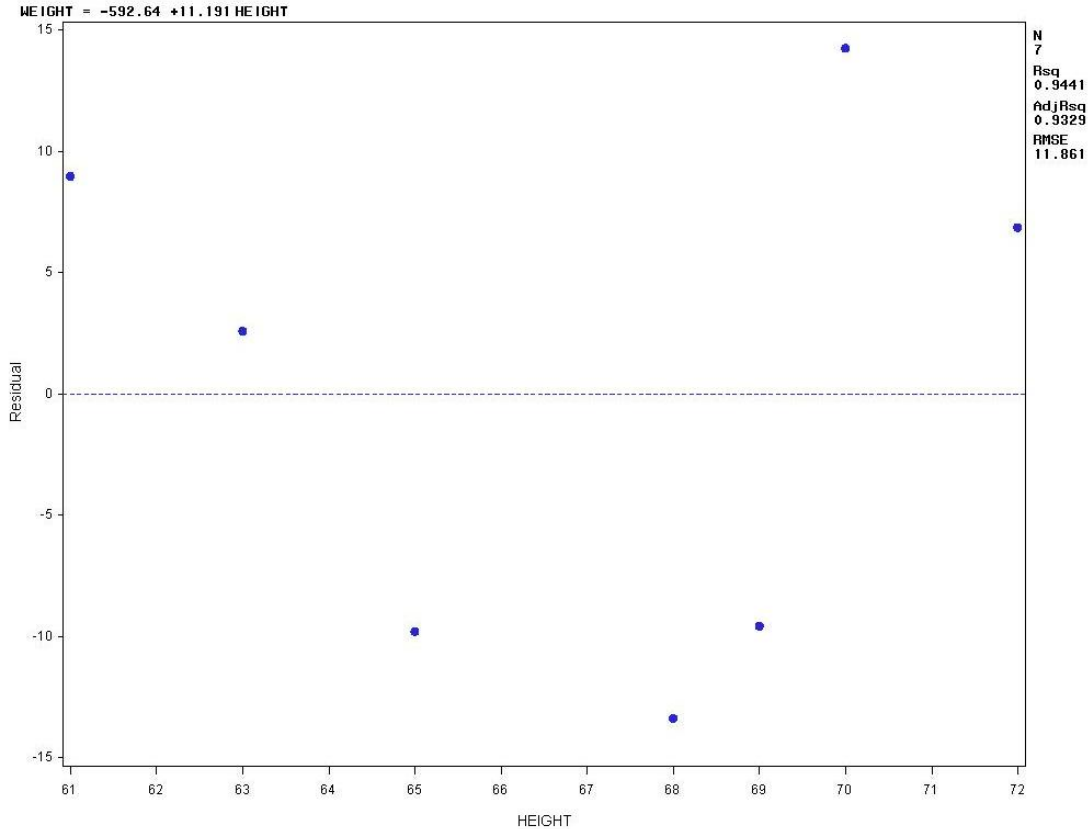
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	11880	11880	84.45	0.0003
Error	5	703.38705	140.67741		
Corrected Total	6	12584			

Root MSE	11.86075	R-Square	0.9441
Dependent Mean	155.57143	Adj R-Sq	0.9329
Coeff Var	7.62399		

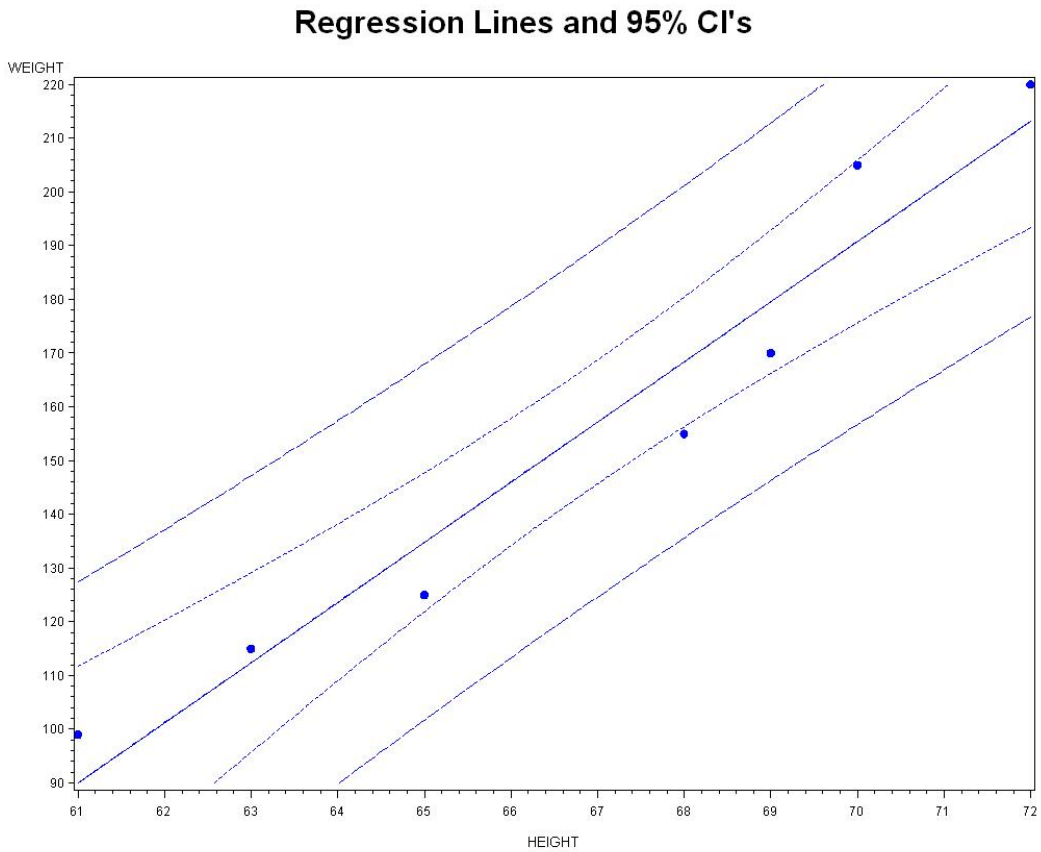
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-592.64458	81.54217	-7.27	0.0008
HEIGHT	1	11.19127	1.21780	9.19	0.0003

Residual Plot



Confidence Region



Adding a Quadratic Term

```
DATA CORR_EG;
     SET CORR_EG;
     HEIGHT2 = HEIGHT**2;
RUN;

PROC REG DATA=CORR_EG;
     TITLE "Regression Line for Height-Weight Data";
     MODEL WEIGHT = HEIGHT HEIGHT2;
     PLOT RESIDUAL.*HEIGHT;
RUN;
```

Quadratic Regression

Regression Line for Height-Weight Data 51
13:59 Monday, May 10, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: WEIGHT

Number of Observations Read 7
Number of Observations Used 7

Analysis of Variance

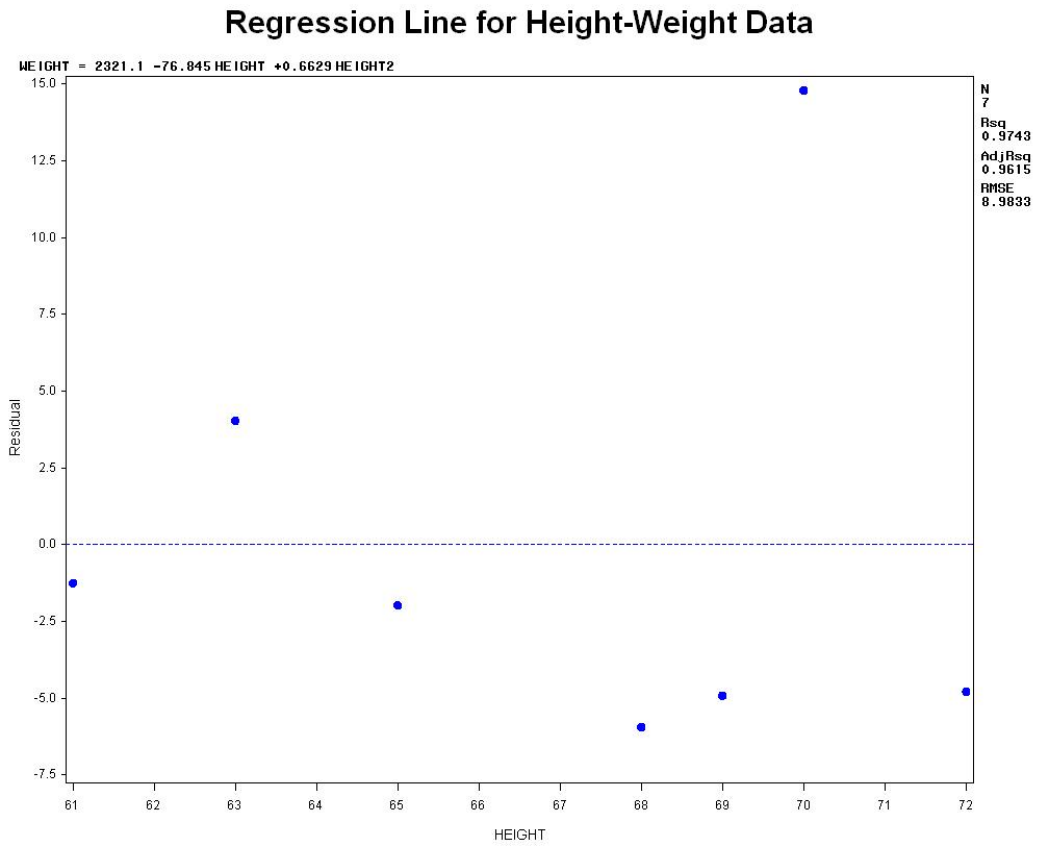
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	12261	6130.45691	75.97	0.0007
Error	4	322.80046	80.70012		
Corrected Total	6	12584			

Root MSE	8.98332	R-Square	0.9743
Dependent Mean	155.57143	Adj R-Sq	0.9615
Coeff Var	5.77440		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2321.12131	1343.15025	1.73	0.1590
HEIGHT	1	-76.84468	40.54924	-1.90	0.1310
HEIGHT2	1	0.66290	0.30525	2.17	0.0956

Residual Plot

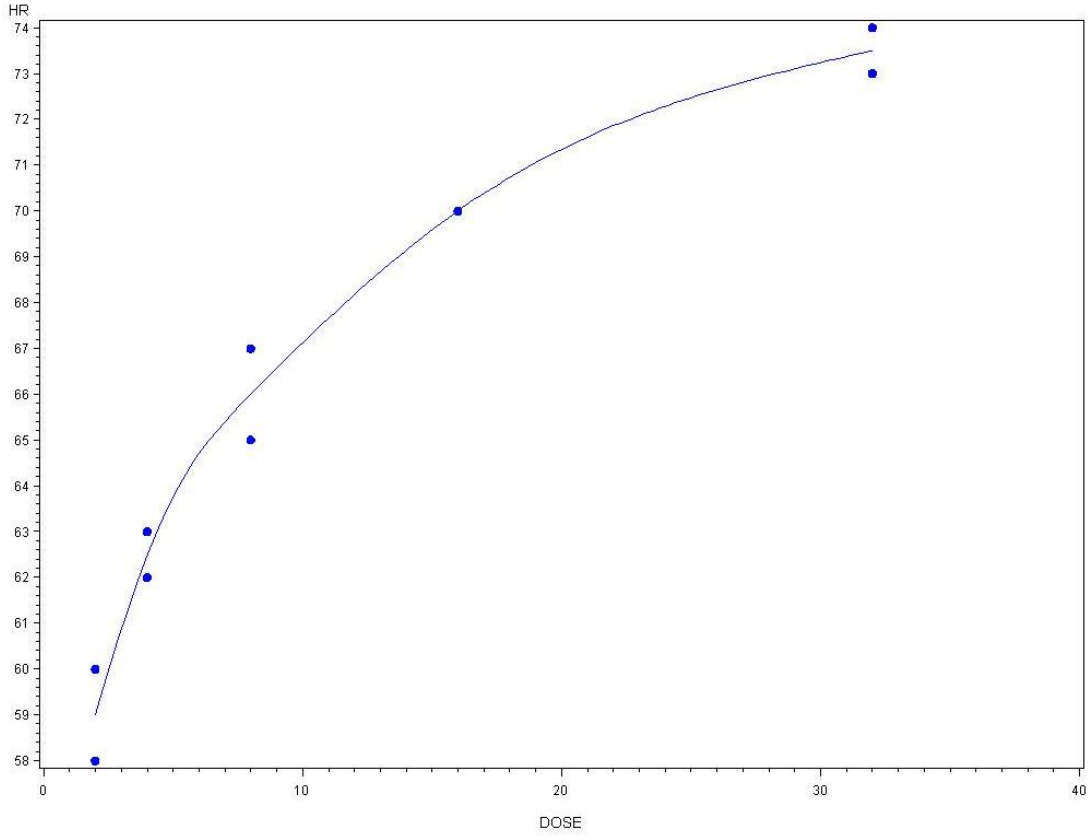


Data Transformation

```
DATA HEART;
     INPUT DOSE HR @@;
DATALINES;
2 60 2 58 4 63 4 62 8 67 8 65 16 70 16 70 32 74 32 73
;
RUN;

SYMBOL VALUE=DOT I=SM;
PROC GPLOT DATA=HEART;
     PLOT HR*DOSE;
RUN;
```

Scatter Plot



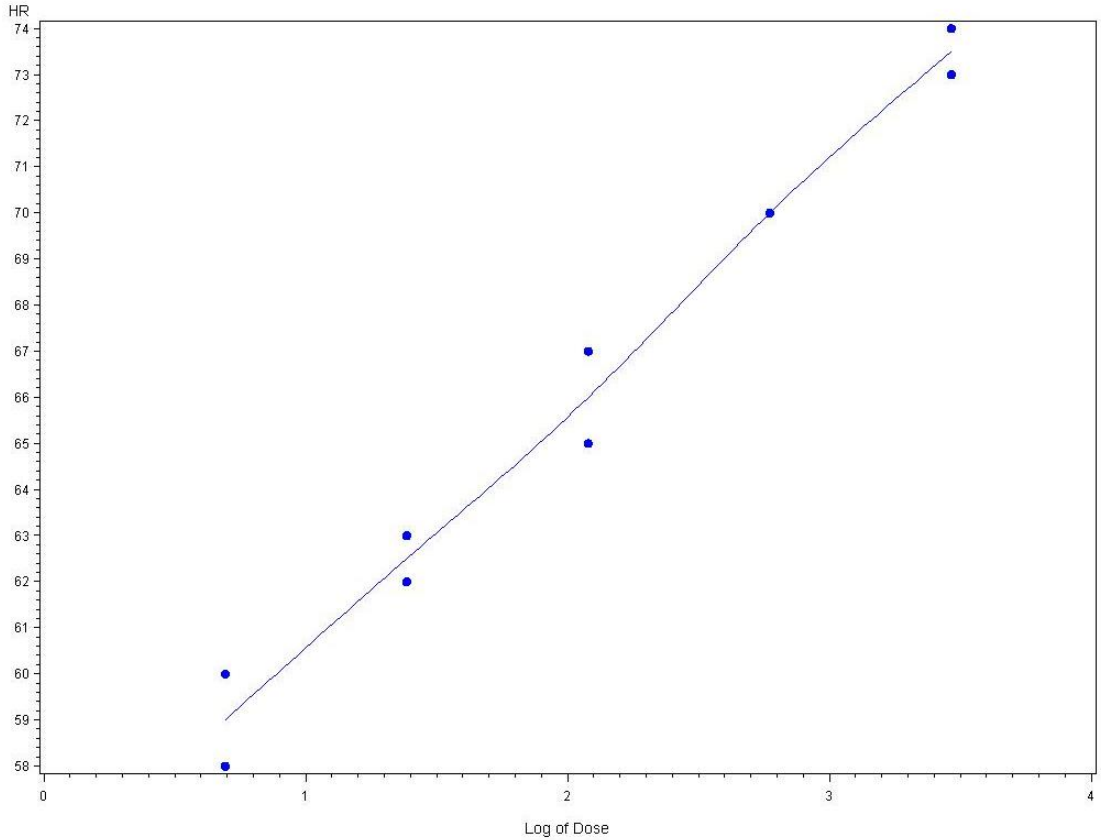
Regression on Transformed Data

```
DATA HEART;
     SET HEART;
     LDOSE = LOG(DOSE);
     LABEL LDOSE = "Log of Dose";
RUN;

PROC GPLOT DATA=HEART;
     PLOT HR*LDOSE;
RUN;

PROC REG DATA=HEART;
     TITLE "Investigating the Dose/HR Relationship";
     MODEL HR = LDOSE;
     PLOT R.*LDOSE;
RUN;
```

Plot of Transformed Data



Regression on Transformed Data

Investigating the Dose/HR Relationship 52
13:59 Monday, May 10, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: HR

Number of Observations Read 10
Number of Observations Used 10

Analysis of Variance

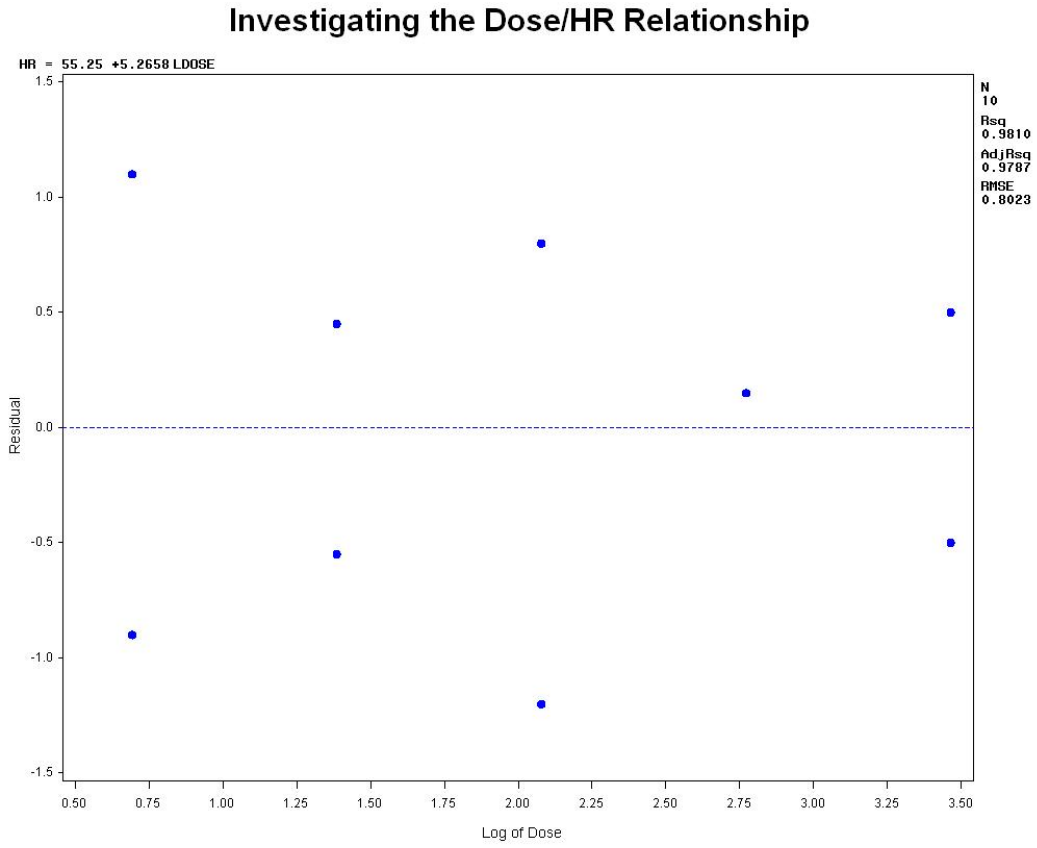
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	266.45000	266.45000	413.90	<.0001
Error	8	5.15000	0.64375		
Corrected Total	9	271.60000			

Root MSE	0.80234	R-Square	0.9810
Dependent Mean	66.20000	Adj R-Sq	0.9787
Coeff Var	1.21199		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	55.25000	0.59503	92.85	<.0001
LDOSE	Log of Dose	1	5.26584	0.25883	20.34	<.0001

Residual Plot



Reference Text