

# Yes, PROC FREQ does that!

AnnMaria De Mars, Ph.D., 7 Generation Games, Santa Monica, CA

## ABSTRACT

While FREQ is one of the first procedures most people learn, the full capabilities are seldom explored. PROC FREQ is a playground for statisticians and "normal people" alike. This presentation covers a dozen uses of PROC FREQ that go beyond counts and percents. These include finding the number of unique records, McNemar test for paired data, Pearson chi-square for independent samples, Fisher's test for small samples, binomial proportions, output of freq data sets to enable "output as input", tests of binomial proportions, relative risk, confidence intervals, dot-plots and bonus features you'll have to attend (or read the paper) to learn.

## INTRODUCTION

Reaching SAS guru status takes years and knowing "all of SAS" is flat impossible (trust me, I've tried). However, anyone can jump start the transition from novice to expert status by learning a few key procedures. One of these is PROC FREQ.

Start with a basic presumption that the purpose of data analysis is to answer questions. A good presentation of that analysis will provide the answer in a manner that is easy for the end users to read and understand. PROC FREQ is a winner on both dimensions.

A word on presentation – just because SAS will print excruciating levels of details by default, it does not mean you should expect your end users to wade through all of it. Give your audience what they need and no more.

Examples used in this paper are from the Dakota Math research project funded by a U.S. Department of Agriculture Small Business Innovation Research grant on the effectiveness of computer games in raising students' math scores. Our example data set comes from a study in which players who missed a math challenge in a game are re-routed to review and reattempt to earn his or her way back into the game. Data are collected on multiple aspects of student performance, including a pre- and post-test of knowledge of Common Core math standards, outcomes of student math challenges, and results of in-game quizzes.

With no further ado, let's dive into what PROC FREQ can do.

## PROC FREQ FOR DATA CLEANING

While PROC FREQ is extremely useful for statistical analysis, the first step is always to evaluate the quality of the data. Here, PROC FREQ can also be very helpful. The first several examples below use the quizzes data set from for Fish Lake, the second game in the series. This data set has 770 observations. An individual student could have zero quiz records, having answered every math question correctly, up to, theoretically, infinity, as students can navigate the loop of problem – study – quiz as many times as necessary.

## QUESTION 1: ARE THERE MORE OCCURRENCES OF A CATEGORY THAN THERE SHOULD BE?

This is one of those questions where SAS enables a programmer to be too smart for one's own good. The answer can be arrived at through some impressive gymnastics with sorting data sets, LAG function and the like. More efficiently, the question can be answered with two simple statements using PROC FREQ.

In our sample data, it is very likely that a player might have the same quiz recorded more than once. Failing it the first time, he or she would be redirected to study and then have a chance to try again. Still, a player shouldn't have TOO many of the same quiz. This problem was supposed to have been fixed after the initial release of our gaming software, but an experienced programmer learns to check data quality both early and often, never assuming a problem has been resolved without data testing that assumption.

To check if we had the same quiz an excessive number of times, I simply did this:

```
PROC FREQ DATA = datasetname;
    TABLES variable1*variable2 / OUT=check (WHERE = (COUNT > 10)) NOPRINT ;
```

The example above uses two variables but you could use one-way tables or n-way cross-tabulations and it will work just as well.<sup>1</sup> The OUT = option will create a dataset named "check". The WHERE clause will select any records where the combination of the two variables occurs more than 10 times. COUNT is the name for the variable created by PROC FREQ that includes the frequency (count) of records. Note the two sets of parentheses, leaving out either of these will cause an error.

---

<sup>1</sup> You may be wondering what the limit is for "N" with N-way tables. There is not a single number. The maximum for N depends on your computer's memory, whether your variables are numeric, versus formatted or character or numeric formatted variables. Longer formatted lengths require more memory. It also depends on the number of levels of each variable.

## OUTPUT AS INPUT

The data set created by PROC FREQ can be used just like any other SAS data set. In this case, I used a PROC PRINT on the data set created, with a SUM statement to give the total number of records these duplicate observations represented, and a FORMAT statement to reduce the number of decimals shown for percent to two.

```
PROC PRINT DATA = check ;
SUM COUNT PERCENT;
FORMAT PERCENT 8.2 ;
```

This code produces the table below.

Obs	username	quiztype	COUNT	PERCENT
1	1429	quiz11	13	1.30
2	1429	quiz5	25	2.50
3	1619	quiz42	12	1.20
4	AZURE4	quiz4	12	1.20
5	AZURE9	quiz4	11	1.10
6	TTCARLSON34	quiz4	19	1.90
			<b>92</b>	<b>9.18</b>

**Table 1. Output from PROC PRINT of Dataset Created by OUT= Option of PROC FREQ**

The table above identifies the five individuals who accounted for 9.18% of the records. These may represent a problem with the server, i.e., a dropped connection resulting in the same record being sent repeatedly, or this may show a user who had a problem at a particular point in the game. In either case, these five individuals warrant further investigation.

## QUESTION 2: HOW MANY UNIQUE LEVELS DO I HAVE?

There are many occasions where a user may wish to know how many unique levels of a variable are included in a data set, whether that variable be classrooms, cities or users.

When basing decisions on results from these data, it is reasonable to ask how many unique individuals are represented. The following code is what you need. The NLEVELS option on the PROC FREQ statement requests the number of unique levels. The NOPRINT option on the TABLES statement suppresses the printing of the frequency distribution, because I'm not interested in that at the moment.

```
PROC FREQ DATA = datasetname NLEVELS ;
TABLES variable / NOPRINT ;
```

This code produces Table 2.

Number of Variable Levels	
Variable	Levels
username	163

**Table 2. Output from PROC FREQ with NLEVELS Option**

From the output in Table 2, we can see very easily that there are data for 163 individual players in this data set.

## QUESTION 3: ARE OUTLIERS IN MY DATA?

One way to answer this question would be to look at a chart.

```
PROC FREQ DATA= datasetname ;
TABLES variable / PLOTS=FREQPLOT ;
```

This code produces your standard frequency distribution table, as shown in Table 3 below.

grade	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	8	1.09	8	1.09
1	1	0.14	9	1.22
3	2	0.27	11	1.49
4	280	37.99	291	39.48
5	286	38.81	577	78.29
6	160	21.71	737	100.00

Table 3: Table Output of PROC FREQ

The PLOTS = FREQPLOT option produces the chart shown in Figure 1. From this it is clear that the grades below grade 4 are outliers.

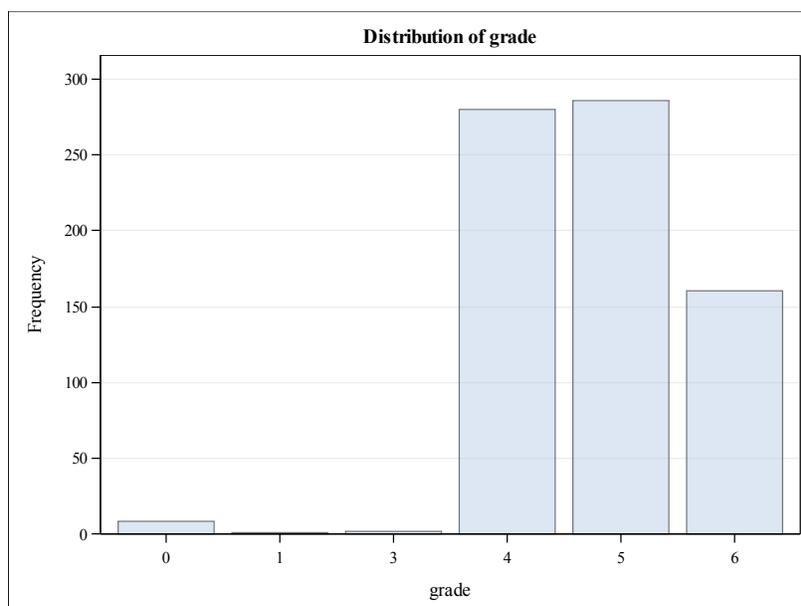


Figure 1: Plot Produced by PLOTS = FREQPLOT Option from PROC FREQ

## PROC FREQ FOR DATA ANALYSIS

Having addressed the most basic data quality issues, we are now ready to turn to common questions in data analysis.

### QUESTION 4: IS THE PROPORTION OBSERVED SIGNIFICANTLY DIFFERENT FROM SOME HYPOTHESIZED VALUE?

After we deleted a few gifted students in grades 1-3 who played Fish Lake, this left a total of 726 records, of which 39% were from fourth-graders, 39% from fifth-graders and 22% from sixth-graders. Given that 75% of the quizzes assessed mathematics standards at the third- or fourth-grade level, and 25% at fifth-grade level, what percentage of students should have passed the quizzes? Does “fourth-grade standard” mean that all fourth-grade students should pass an item? Does it mean that 50% should pass, i.e., that the median student in fourth-grade comprehends an item?

There is not a definitive answer to this question. Results from various states at various grades have shown pass rates of Common Core tests ranging from 22% to 75%. For the sake of argument (that sounds better than the fact that I don't have a good answer to that question), let's assume that in a sample of fifth-graders, 75% should pass quizzes that are at the fourth-and fifth-grade levels.

To test our null hypothesis that the passing rate is no different than a hypothesized population value of 75%, the following code was used:

```
PROC FREQ DATA = datasetname ;
    TABLES variabe / BINOMIAL (P=.25);
    WHERE grade = 5 ;
```

The option BINOMIAL with no value specified will test for the default null hypothesis that the population proportion = .50. Instead of accepting the default, any number between 0 and 1.0 can be specified. Why specify P = .25 ? Wasn't the hypothesis that 75% would pass? By default, the proportion is tested for the first variable that appears in the output, which is passed = 0. Testing the hypothesis that 25% failed is the same as testing the hypothesis that 75% passed.<sup>2</sup>

The WHERE statement will only select students in grade 5 for this analysis.

This analysis gives the standard frequency distribution, shown below.

passed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	184	64.34	184	64.34
1	102	35.66	286	100.00

**Table 4: Frequency Distribution for Pass for Quizzes for Fifth-Grade Students**

The BINOMIAL = option produces the two tables below. For the proportion of fifth-graders who failed, 64.34%, the ASE (the asymptotic standard error), lower and upper limits for the 95% confidence interval are provided. Wald and Exact confidence limits are displayed by default.

Binomial Proportion	
passed = 0	
Proportion	0.6434
ASE	0.0283
95% Lower Conf Limit	0.5878
95% Upper Conf Limit	0.6989
Exact Conf Limits	
95% Lower Conf Limit	0.5848
95% Upper Conf Limit	0.6989

**Table 5: Confidence Limits and ASE for Test of Binomial Proportion for Quizzes for Fifth-Graders**

The next table gives the test of the null hypothesis that the proportion = .25 – a Z-value, ASE and probabilities for one-sided and two-sided tests. As can be seen, the probability is less than .0001.

Test of H0: Proportion = 0.25	
ASE under H0	0.0256
Z	15.3627
One-sided Pr > Z	<.0001
Two-sided Pr >  Z	<.0001

**Table 6: Test of H0: Proportion = 0.25 for Quizzes for Fifth-Graders**

Thus, can we conclude that between 59% and 70% of fifth-grade students on American Indian reservations in the Great Plains would probably fail quizzes in mathematics at their purported grade level? NO!

There are several options for tests of binomial proportions with PROC FREQ but all assume randomly sampled set of independent observations. These data violate both assumptions. Students were not randomly sampled – only players who failed a mathematics challenge in the game would be routed to take a quiz. Observations were not independent. A student could take multiple quizzes or take the same quiz multiple times.

---

<sup>2</sup> If that really bothers you, you can use the LEVEL= option to specify a different level for the proportion.

Yes, PROC FREQ Can Do That

While PROC FREQ can do many things, one thing that it cannot do is to verify that assumptions of statistical tests have been met.

Yes, that was a dirty trick to pull as an example, but it is important to keep in mind the limits of statistical software as a tool that does require some knowledge of basic statistics.

**QUESTION 5: IS THERE A SIGNIFICANT RELATIONSHIP BETWEEN GRADE AND PASSING?**

It would seem that a valid test would be related to grade level, with students in higher grades more likely to pass. To test for a relationship between two categorical variables with a random sample of independent observations, chi-square is the obvious choice.

```
PROC FREQ DATA = datasetname ;
    TABLES variable1*variable2 / CHISQ;
    WHERE grade in (5,6) ;
```

Table of grade by pass60			
grade	pass60		
Frequency Percent Row Pct Col Pct	0	1	Total
5	78 58.21 97.50 59.54	2 1.49 2.50 66.67	80 59.70
6	53 39.55 98.15 40.46	1 0.75 1.85 33.33	54 40.30
Total	131 97.76	3 2.24	134 100.00

**Table 7: Cross-tabulation of grade and passing**

Even with a passing score set at 60% and excluding fourth-grade students, less than 3% of those tested passed the pretest. It actually appears to be the opposite as might be expected, with students in sixth grade less likely to pass. Is this relationship beyond what would occur by chance? The chi-square option produces several tests of association of categorical variables, including a Pearson chi-square, continuity adjusted chi-square and phi-coefficient. With regard to statistical decisions of significance versus non-significance or effect size, the differences among these are usually trivial. You can find discussions of each in Severino (2000) or De Mars (2013).

Statistic	DF	Value	Prob
Chi-Square	1	0.0619	0.8036
Likelihood Ratio Chi-Square	1	0.0633	0.8014
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0614	0.8043
Phi Coefficient		-0.0215	
Contingency Coefficient		0.0215	
Cramer's V		-0.0215	
<b>WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.</b>			

**Table 8: Chi-Square Statistics Produced by Chi-square Option**

It appears from the results that there is a non-significant relationship between grade and passing the test, but there is a warning that with 50% of cells having a count less than 5, chi-square may not be a valid test. In a situation such as this, three options exist: collect more data, modify the categories, or use Fisher's Exact Test. In the present example, more data collection is not possible, since the project has ended. Fortunately, the exact test is produced by default for 2 x 2 tables when the CHISQ option is used.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	78
Left-sided Pr <= F	0.6448
Right-sided Pr >= F	0.7905
Table Probability (P)	0.4352
Two-sided Pr <= P	1.0000

**Table 9: Fisher's Exact Test Produced by Chi-square Option**

Fisher's exact test is helpful in two instances. One is when you have a small sample size and chi-square tests are of questionable validity. The second is when, regardless of sample size, you want an exact probability computed by SAS. The probability of getting a table showing this degree of relationship or less is 43.5%, not really necessary in this case but it is quite helpful when giving expert witness to be able to say that not only is the probability greater than 5 in 100 but over 40% of the time a relationship this large would occur by chance. This point serves as a reminder that PROC FREQ applies across an enormous range of disciplines. The examples used in this paper are from research on the effectiveness of video games in teaching mathematics, a combination of disciplines in itself. These identical analyses are also applicable to law, health care, marketing and just about any other field imaginable.

#### QUESTION 4 REVISITED: IS THE PROPORTION OBSERVED SIGNIFICANTLY DIFFERENT FROM SOME HYPOTHESIZED VALUE?

What proportion of students in fifth- and sixth-grade should pass a test of common core mathematics standards related to fractions at the fourth- and fifth-grade level? Pre-tests were given in the second semester of the school year, so it seemed reasonable that the median student of the fifth- and sixth-graders should pass, that is 50% of the students.

At this point in the research, our teacher-consultants from the beta test schools pointed out that a) over 97% of students failed with a passing score set at 60%, b) with a mean score on the pretest of less than 30%, a student could double his or her score and still show "no improvement", i.e., still failing. While not wholly convinced by this argument to 'grade on the curve', for demonstration purposes, the research staff agreed to set "passing" at 30% correct. (And, yes, Repeated Measures Analyses of Variance were performed with test score as a dependent variable, but these results are not salient to PROC FREQ).

All students in the participating schools were tested, and each student took the pretest only once, thus the test of binomial proportions is valid. The following code was used, as the default is to test the population proportion = .50, no test proportion was specified.

```
PROC FREQ DATA = datasetname ;
    TABLES variablename / BINOMIAL ;
    WHERE grade in (5,6) ;
```

Results from this procedure are shown in Tables 10, 11 and 12 below. It can be seen from Table 10 that even with a passing rate of 30%, less than half of students earned a passing score. In Table 11, it can be seen that the 95% confidence interval ranged from 53.7% to 70.2% failing the pretest. Again, remember that the default is to test the proportion of the first (lowest) category, which in this case is pass30 = 0, or failing.

The null hypothesis that 50% of students would pass the test was rejected, even for this very low bar set for passing.

pass30	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	83	61.94	83	61.94
1	51	38.06	134	100.00

**Table 10: Frequency Distribution for Pass for Pretest for Fifth- and Sixth-Grade Students**

Binomial Proportion	
pass30 = 0	
Proportion	0.6194
ASE	0.0419
95% Lower Conf Limit	0.5372
95% Upper Conf Limit	0.7016
Exact Conf Limits	
95% Lower Conf Limit	0.5316
95% Upper Conf Limit	0.7018

Table 11: Confidence Limits and Standard Error for Binomial Test

Test of H0: Proportion = 0.5	
ASE under H0	0.0432
Z	2.7644
One-sided Pr > Z	0.0029
Two-sided Pr >  Z	0.0057

Table 12: Test of H0: Proportion = 0.50 for Fifth- and Sixth-Grade Pretest Passing Scores

At this point, you may be wondering about the extreme difficulty of this test and whether it does in fact test items at the fourth- and fifth-grade level. A sample item is shown below.

It takes  $\frac{2}{3}$  of a deer skin to make a warm robe for one small child. Grandmother tells the hunters she needs to make three robes for the winter. How many deer skins must they bring back?

ANSWER

In fact, the test is precisely aligned with Common Core standards for fourth- and fifth-grade mathematics and the finding that students even at the sixth-grade level could not pass it was somewhat shocking even for staff working in low-performing schools. Thus, repeated analyses were conducted to validate the results.

**QUESTION 6: WHAT IS THE RELATIVE RISK?**

**QUESTION 7: WHAT IS THE ODDS RATIO?**

Using the 30% cut-off for “passing”, perhaps there is a difference between grades. One way to answer this question is by comparing the relative risk of failure for fifth-graders versus sixth-graders. Although relative risk is predominantly used in medical applications, dividing disease prevalence of the exposed group by prevalence of the unexposed, it can be applied to other types of risks as well. What is the relative risk of failure by grade? What is the odds ratio? Both of these questions can be answered with three simple lines of code.

```
PROC FREQ DATA=datasetname ;
    TABLES variable1*variable2 / NOPERCENT NOCOL CMH;
```

First, note the NOPERCENT NOCOL options. Remember all the way back up in the introduction the suggestion to give your audience what they need to know and no more? Relative risk is computed by dividing the risk of one group by the risk of a second group. The relative risk of failure is the percentage failing in fifth-grade divided by the percentage failing in sixth-grade. These are the two row percentages, so why not just show the row percentages and leave out the values not used in the calculation, as shown in Table 13 below?

Table of grade by pass30			
grade	pass30		
Frequency Row Pct	0	1	Total
5	45 56.25	35 43.75	80
6	38 70.37	16 29.63	54
<b>Total</b>	83	51	134

**Table 13: Cross-tabulation of Pass on Pretest by Grade**

The next option on the TABLES statement, CMH, produces a table of relative risks and odds ratio, shown in Table 14. It can be easily seen that the relative risk of  $.799 = .56/.70$ , the risks of failure for grades five and six, respectively. Similarly, the relative risk for passing is 1.47, or  $.43/.29$ . With a relative risk of 1.0 falling within the 95% confidence limits, the risks are not significantly different.

An odds ratio of .54 indicates lower odds of failure for students in fifth-grade compared to sixth-grade, and again, with 1.0 falling within the 95% confidence intervals, the null hypothesis of a difference in pass rates is rejected.

Estimates of the Common Relative Risk (Row1/Row2)				
Type of Study	Method	Value	95% Confidence Limits	
<b>Case-Control</b> (Odds Ratio)	Mantel-Haenszel	0.5414	0.2603	1.1260
	Logit	0.5414	0.2603	1.1260
<b>Cohort</b> (Col1 Risk)	Mantel-Haenszel	0.7993	0.6167	1.0361
	Logit	0.7993	0.6167	1.0361
<b>Cohort</b> (Col2 Risk)	Mantel-Haenszel	1.4766	0.9134	2.3870
	Logit	1.4766	0.9134	2.3870

**Table 14: Relative Risks and Odds Ratio Statistics Produced by CMH Option**

The failure to find a significant difference in failure rates between grades might seem to indicate a problem with test validity as one might expect that students in sixth-grade would be more likely to pass a test of fourth- and fifth-grade mathematics. In fact, these results are consistent with our previous work showing students in fourth-grade outperformed fifth-graders on a test of third- and fourth-grade mathematics. There is a “recency effect” at work with students who have just learned a concept or skill showing better performance than older students who had the same lessons over a year ago. Yet again, there is an obvious need for knowledge of SAS and knowledge of the field of research to which the statistics are applied.

### QUESTION 8: IS THERE A SIMPLE WAY TO VISUALIZE THE RELATIONSHIP BETWEEN TWO CATEGORICAL VARIABLES?

Having found that fifth-graders did not perform significantly worse than sixth-graders, and in fact, performed non-significantly better, the next question we ask is whether there is a difference in pass rates of fourth-graders from fifth- and sixth-graders. This question could be answered with a dot plot.

```
ODS GRAPHICS ON ;
PROC FREQ DATA=datasetname ;
    TABLES variable1*variable2 / PLOTS=FREQPLOT(TYPE=DOTPLOT SCALE=PERCENT) ;
```

Note that ODS GRAPHICS ON is required. The option PLOTS= FREQPLOT (TYPE= DOTPLOT) produces the dot plot. The SCALE=PERCENT is not required but when comparing across categories with unequal distribution, you may find it easier to compare percentages than frequencies. The option produces the code in Figure 2. It is important to remember that the percentages shown are not within category, but rather of the total sample.

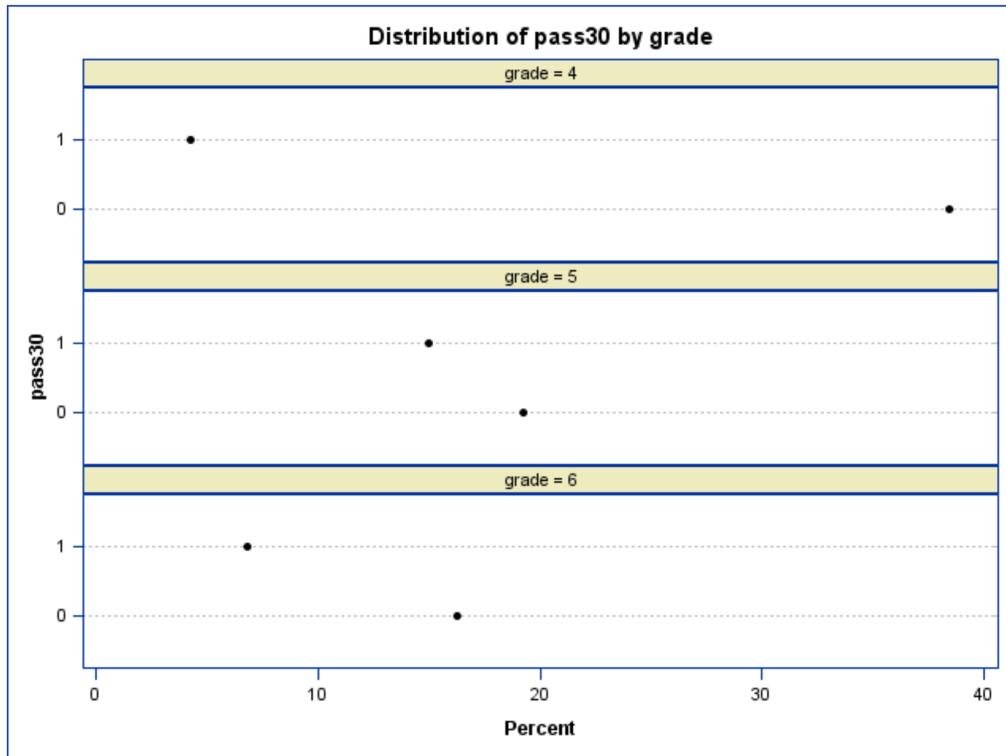


Figure 2: Plot Produced by PLOTS = FREQPLOT Option from PROC FREQ

This plot provides several pieces of information at a glance. For example, there are far more fourth-grade students who failed the test than students in any other combination. Also, fourth-graders were far more likely to fail than pass the test. The differences between pass and failure percentages were not as large for fifth and sixth-graders.

Slightly closer inspection shows that fourth-graders make up over 40% of the sample, with sixth-graders comprising only around 20%.

### QUESTION 9: IS THERE A SIGNIFICANT DIFFERENCE IN THE DISTRIBUTION OF MATCHED PAIRS?

Having learned our lesson that when observations are not independent we need to consider that fact, let's take a look at analyzing matched pairs. We have pre-test and post-test data on 68 students. The test consisted of 23 quantitative or multiple choice items, e.g., " Which number sentence describes the figures above?" and two, two-point short-answer questions. All students who played the game for eight weeks took the post-test. Our null hypothesis is that there is no difference in the probability of passing or failing the test from pre-test to post-test.

The mean percentage correct on the pre-test for these matched pairs was 28.7%. Although we acknowledge it is a low percentage, for demonstration purposes, we again set the percentage passing at 30% correct.

These three simple statements produce two types of tests for correlated data.

```
PROC FREQ DATA =datasetname ;
    TABLES variable1*variable2 ;
    TEST AGREE ;
```

The TABLES statement provides a simple cross-tabulation. We can see that there was an improvement from a 25% pass rate on the pretest to nearly 40% on the post-test. Was this difference significant? Of the students who passed the pretest, 82% passed the post-test, and of those who failed the pre-test, 25% passed the post-test, so it *appears* there was an improvement. What is needed is an appropriate statistical test.

Table of pre_pass by post_pass			
pre_pass	post_pass		
Frequency Percent Row Pct Col Pct	0	1	Total
0	38 55.88 74.51 92.68	13 19.12 25.49 48.15	51 75.00
1	3 4.41 17.65 7.32	14 20.59 82.35 51.85	17 25.00
Total	41 60.29	27 39.71	68 100.00

**Table 15: Cross-tabulation of Pre\_Pass and Post\_Pass Produced by TABLES Statement**

The TEST statement has an even dozen options that produce statistics to test for relationships between categorical variables, using AGREE will provide McNemar's test. What exactly does McNemar's test test? It is a test for correlated proportions and the null hypothesis being tested is that there is no difference in marginal proportions. In this case, that is a test of the null hypothesis that the distribution is the same across pre-test and post-test.

McNemar's Test	
Statistic (S)	6.2500
DF	1
Pr > S	0.0124

**Table 16: McNemar's Test Results from TEST AGREE Statement**

The null hypothesis is rejected with  $p < .05$ . A significantly higher proportion of students passed the post-test.

It's worth noting that this exact same statistic could be used for biostatistics, assessing whether patients show a difference in symptoms, fewer drug side effects, etc. following a treatment.

Another statistic produced by the same statements is the Kappa coefficient, which is irrelevant to this question, so let's turn to a question where it is relevant.

### QUESTION 9: WHAT IS THE LEVEL OF AGREEMENT BETWEEN RATERS?

There are two essay questions on the test used for pre- and post-test. Each essay question was scored twice, by two different raters. (Only the analysis with the first question is shown here.) In this case, the exact same statements are used as in the previous analysis, with only a PLOTS option added,:

```
ODS GRAPHICS ON;
PROC FREQ DATA =datasetname ;
    TABLES variable1*variable2 / PLOTS = KAPPAPLOT;
TEST AGREE ;
```

These statements produced the results below. First, Table 17, a cross-tabulation of the two ratings.

Table of ESSAY1_Score2 by ESSAY1_Score1				
ESSAY1_Score2	ESSAY1_Score1			
Frequency Col Pct	0	1	2	Total
0	176 88.00	11 35.48	2 9.09	189
1	18 9.00	16 51.61	6 27.27	40
2	6 3.00	4 12.90	14 63.64	24
Total	200	31	22	253

**Table 17: Cross-tabulation of Ratings from Two Raters**

There seems to be a relationship between the raters' scores, with 88% of players scored "0" by the first rater also scored zero by the second. While the distribution across ratings – incorrect, partially correct, correct – is approximately the same 3% of those scored incorrect by Rater 1 were scored correct by Rater 2. A quantitative level of agreement would be helpful, but we know (hopefully) that neither a phi coefficient nor chi-square would be appropriate given that these are not independent observations but rather the same 253 records rated twice. (There are more records than the quizzes because this includes both students who answered every question correctly, and thus never encountered a quiz, and students who quit before completing the first quiz.)

We can look at the McNemar test again. Recall, this is a test of whether the marginal proportions differ, and it is, as we hoped, non-significant, with  $S = 4.09$  ( $p > .25$ ).

Test of Symmetry	
Statistic (S)	4.0897
DF	3
Pr > S	0.2519

**Table 18: Cross-tabulation of Ratings from Two Raters**

To measure agreement between two raters, we use Kappa's coefficient. PROC FREQ produces two types of Kappa coefficients. The Kappa coefficient ranges from -1 to 1, with 1 indicating perfect agreement, 0 indicating exactly the agreement that would be expected by chance and negative numbers indicating less agreement than would be expected by chance (Vierra & Garrett, 2005).

When there are only two categories, PROC FREQ produces only the Kappa coefficient. When more than two categories are rated, a weighted Kappa is also produced which credits categories closer together as partial agreement and categories at the extreme ends as no agreement.

Kappa Statistics				
Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.5135	0.0570	0.4017	0.6252
Weighted Kappa	0.5834	0.0559	0.4739	0.6929

**Table 19: Cross-tabulation of Ratings from Two Raters**

The simple Kappa and weighted Kappa are both shown in Table 11 above, along with their associated standard errors and confidence limits. While the weighted Kappa is larger than simple Kappa, and significantly different from zero, it is still far from perfect. Better training of raters or a more detailed rubric may be recommended.

The plot produced with the PLOTS = KAPPAPLOT option is shown in Figure 2 below.

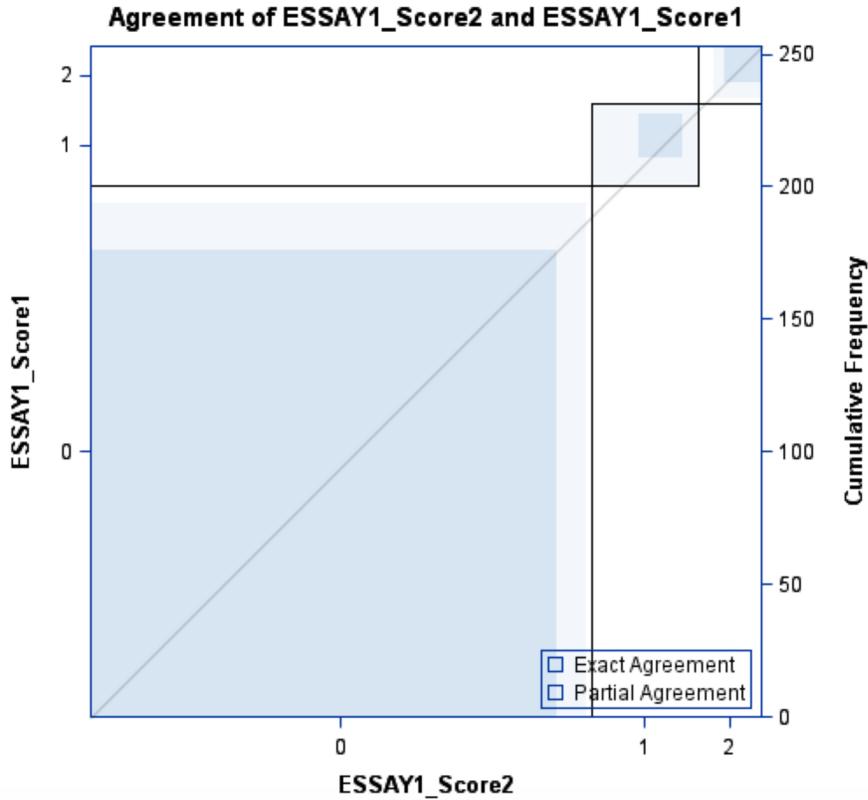


Figure 2: Plot Produced by PLOTS = KAPPAPLOT Option from PROC FREQ

This visual representation of the agreement shows that there was a large amount of exact agreement (dark blue shading) for incorrect answers, scored 0, with a small percentage partial agreement and very few with no agreement. With 3 categories, only exact agreement or partial agreement is possible for the middle category. Two other take-away points from this plot are that agreement is lower for correct and partially correct answers than incorrect ones and that the distribution is skewed, with a large proportion of answers scored incorrect. Because it is adjusted for chance agreement, Kappa is affected by the distribution among categories (Vierra & Garrett, 2005). If each rater scores 90% of the answers correct, there should be 81% agreement by chance, thus requiring an extremely high level of agreement to be significantly different from chance. The Kappa plot shows agreement and distribution simultaneously.

#### QUESTION 10: HOW CAN I MAKE A FORMATTED TABLE?

While all of the previous examples included some element of data analysis, there are situations where the greatest need is simply for presentation. In addition to a pretest and quizzes, 7 Generation Games include math problems that appear as the player navigates the virtual world, asking, for example, which direction has the highest proportion of poisonous snakes. Answers are numbered 1 through 14, but there is a complication. After the beta test, teachers requested more math problems be added to the game. Since re-numbering the problems would require changing a lot of code, and the problem numbers had to be integers, the new problems between 1 and 2 were numbered 22, between 4 and 5, 42, etc.

It would be helpful to have categories ordered 1, 2, 2a, 3 ... to 14. The solution first requires creating a user-defined format as shown below:

```
PROC FORMAT ;
VALUE quiz
1 = "01"
2 = "02"
3 = "03"
4 = "04"
5 = "05"
6 = "06"
```

## Yes, PROC FREQ Can Do That

```
7 = "07"  
8 = "08"  
9 = "09"  
10 = "10"  
11 = "11"  
12 = "12"  
13 = "13"  
14 = "14"  
22 = '02a'  
42 = '04a'  
43 = '04b'  
51 = '05b'  
61 = '06b'  
;  
VALUE correct  
0 = "Wrong"  
1 = "Right"  
;
```

In brief, PROC FORMAT invokes the format procedure. A VALUE statement is required. It names the format and then gives a formatted value for values of the variable. As shown in the example above, multiple formats can be created in a single PROC FORMAT step.

The remainder of the code first creates and opens an RTF file and sets a style for the table. (You can find examples of 52 pre-defined table styles here [http://www.sascommunity.org/mwiki/images/6/66/SAS\\_Styles\\_-\\_updated\\_for\\_SAS\\_9dot2.pdf](http://www.sascommunity.org/mwiki/images/6/66/SAS_Styles_-_updated_for_SAS_9dot2.pdf).) Now, for the problems to be printed in order of the format, not the problem number, all that is needed is to add the ORDER = FORMATTED option to the PROC FREQ statement, and a FORMAT statement specifying the format.

The main point of interest in this table for the reader is the percentage of right answers for each problem, that is, the percentages of each row. The NOCUM, NOCOL and NOPERCENT options on the TABLES statement omit the unnecessary information of cumulative percents, column percentages and percent of total respectively.

```
ODS RTF FILE = "path\nice_table.rtf " STYLE=STYLES.FANCYPRINTER ;  
PROC FREQ DATA= datasetname ORDER=FORMATTED ;  
    TABLES problem*correct / NOCUM NOCOL NOPERCENT ;  
    FORMAT problem quiz. correct correct. ;  
run;  
ODS RTF CLOSE ;
```

At a glance at Table 20, the reader can see that there is not a simple, linear trend in proportion correct, with the problems getting more difficult as the game progresses. It's possible that higher-than-anticipated item difficulty levels may be due to issues with items' wording or physical requirements, such as dragging and dropping an object to the correct spot on a number line. It's also readily apparent that some of the items added later in game development, such as 5b, and the problems that occur in the last levels of the game, have small sample sizes and the reader should exercise caution in making generalizations regarding these items. Limiting the table output to only the relevant information makes the task of interpreting data much easier for the average reader.

## CONCLUSION

PROC FREQ is a procedure of many uses. Although it is considered a 'basic' statistical technique, as the examples above have shown, it has numerous practical applications. From identifying outliers to assessing inter-rate agreement, PROC FREQ can be used to evaluate data quality. The procedure offers a large number of options for analyzing diverse data types, only a fraction of which have been touched on in this paper. These include analysis of both independently sampled and correlated data, tests of association, agreement, relative risk and odds ratios, to name a few. Finally, PROC FREQ, when used with the Output Delivery System (ODS) can produce both graphics and presentation quality tables.

What PROC FREQ cannot do is determine whether the sample selection method met assumptions of a given test, identify results that indicate problems with data quality, or decide criteria requiring subject matter knowledge, such as a reasonable cut-off for passing a test or the best hypothesized value for the population.

<i>Table of problem by correct</i>			
<i>problem</i>	<i>correct</i>		
<i>Frequency Row Pct</i>	<i>Right</i>	<i>Wrong</i>	<i>Total</i>
<i>01</i>	495 88.71	63 11.29	558
<i>02</i>	338 95.48	16 4.52	354
<i>02a</i>	19 70.37	8 29.63	27
<i>03</i>	250 89.29	30 10.71	280
<i>04</i>	149 86.63	23 13.37	172
<i>04a</i>	81 39.51	124 60.49	205
<i>04b</i>	81 91.01	8 8.99	89
<i>05</i>	122 76.25	38 23.75	160
<i>05b</i>	2 100.00	0 0.00	2
<i>06</i>	134 97.81	3 2.19	137
<i>06b</i>	28 62.22	17 37.78	45
<i>07</i>	98 91.59	9 8.41	107
<i>08</i>	66 51.16	63 48.84	129
<i>09</i>	58 59.79	39 40.21	97
<i>10</i>	41 87.23	6 12.77	47
<i>11</i>	6 66.67	3 33.33	9
<i>12</i>	3 100.00	0 0.00	3
<i>13</i>	12 100.00	0 0.00	12
<i>14</i>	12 54.55	10 45.45	22
<i>Total</i>	1995	460	2455

**Table 20: Table Produced by ORDER = FORMATTED Option**

## REFERENCES

- De Mars, A. (2013). Categorical Data Analysis: Because There Really Is No Fate Worse Than Death. Proceedings of the SAS Users of Oregon conference. Salem, OR.  
[http://www.sascommunity.org/wiki/Categorical\\_Data\\_Analysis:\\_Because\\_There\\_Really\\_Is\\_No\\_Fate\\_Worse\\_Than\\_Death](http://www.sascommunity.org/wiki/Categorical_Data_Analysis:_Because_There_Really_Is_No_Fate_Worse_Than_Death)
- Severino, R. ( 2000 ) PROC FREQ: It's more than counts. Proceedings of the SAS Users Group International meeting. Indianapolis, IN. <http://www2.sas.com/proceedings/sugi25/25/btu/25p069.pdf>
- Vierra, A. J. & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine, 360-363.

## ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of the administration and teachers at Tate Topa Tribal School, Ojibwa Indian School and Turtle Mountain Elementary School in data collection and game design

This research was supported by funding from USDA SBIR grant #: 2013-02666.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

AnnMaria De Mars  
7 Generation Games  
2425 Olympic Blvd, Ste. 4000-W  
Santa Monica, CA 90404  
(310) 804-9553  
[annmaria@7generationgames.com](mailto:annmaria@7generationgames.com)  
[www.7generationgames.com](http://www.7generationgames.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.