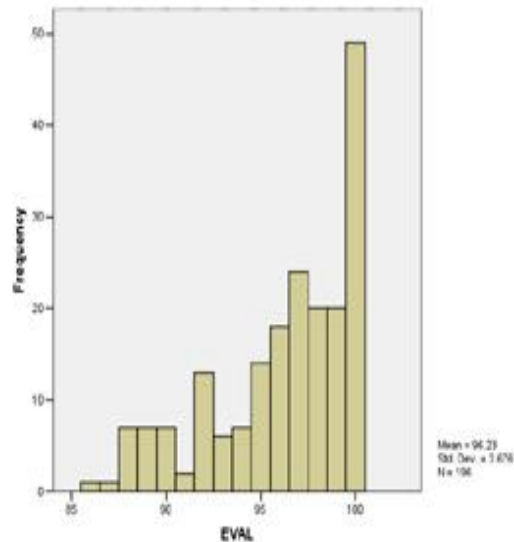**A Graduate Level Statistics Report in *G. Heart Country***

2018

ABSTRACT

This report is a post class review of the tests, exams, and statistics used in the Fundamentals of Instructing (F.O.I.) class and augments the previous report titled *Action Research: A Statistical Report on IP Upgrade* (2010). As the Action Research paper included statistics on 400 students (N = 400) an inherit limitation in this present SPSS program only allowed 200 cases as a maximum to exam. This paper will justify the use of a truncated number (N = 200) and show both samples being statistically the same samples. The smaller program will add learning knowledge about the students and answer questions not listed in the full program. The same one-sample *t* test (N = 400) was administered with the (N = 200) group using the evaluations of the course and the results were very similar except the confidence level was raised from .05 to .01 yielding significance in both as the second test was *t* (199) = 17.78, p <.01 verses the first test of *t* (399) = 21.82, p. < .01.

Motivation

There are three pedagogical motivation reasons to study/review statistics in this chapter:

1. These class work papers, surveys, tests, and evaluations can be used in graduate work – even at a PhD level – for any student interested in educational advancement…help yourself!

2. These working tools are helpful to a teacher who wants to know how well his students are doing.

3. Getting on top of some basic statistics adds to the joy of learning, the joy of teaching, and the joy of what the means mean in a measured instrument – like the one-sample *t* test on EVAL in the abstract.

Statistics (lesson)

Most people shy away from the word statistics – in Oklahoma pronounced "Stystics" – but really, the common sense issues of the discipline are very valuable in our everyday life. For instance, if the two big universities in Oklahoma were looking for a quarterback they would have a statistical list of numbers to be met by the young candidates, for instance: height, weight, GPA, and maybe a score on their attitude toward playing football. The statistics for an applicant, for example, of 5' 4'' height, 137 pounds weight, 2.01 GPA, and a low score on attitude would be easily a no choice. These numbers show present status and could predict or show the probability of success in quarterback selection or, as is discussed in this chapter, instructor pilot selection.

A sample of a course in "Systics" would include (yes, I know this looks daunting – and it is):
* Review of arithmetic concepts, algebraic concepts, coordinate geometry, notations and terminology.
* Organizing Data, Measures of Central Tendency.
* Variation, Probability, The Normal Curve, Sampling, Estimation, Hypotheses Testing.
* Inferential Statistics, Linear Regression, Nonparametic Statistics.

So, as a tool for the administration of teaching/instructing in The Fundamentals of Instruction (FOI) with my lieutenants (students) it was helpful to have a small background in statistics to monitor, review, and prognosticate on the well being of the class.

Here is a sample of the statistical results of three students (out of 200) we will be talking about:

ID	GENDER	AGE	COLL	AD	FLY	ATINST	ATFLY	PER	PRE	POST	EVAL
310	2	26[1]	2	4	8	80	57	1	60	85	96
380	2	26	1	3	8	92	66	2	55	75	100
467	1	38	1	16	41	94	82	7	55	50	99

These numbers add up to about 4,000 total numbers given the 12 variables used on each case.

My personal task was to find some gold in these Air Force pilots but I found out there was more than just a little gold. This was the joy of teaching these young and able men and women for the most of 15 years and I now have the statistics to suggest[2] the real golden color of the gold…

Review[3]

The working papers that the numbers come from for the statistical analysis are:

1. Questionnaire – ID number, age, gender, college, active duty time, flying time – DESCRIPTIVE STATS.

---

[1] Correlate these three highlighted numbers (26, 26, 38) with the highlighted word "age" in line "1. Questionnaire" listed the next Review section .

[2] Sometimes in statistical research a reporting author will claim his data 'proved' his work. Well, we don't prove anything in statistics. The data (plural) are going to suggest an outcome but proving it is beyond the scope. Borg and Gall (1989) say "…a researcher can never 'prove' a theory, only 'support' it" (p. 30).

[3] An 'interested student' will review *The Action Research* paper to get a feel, perhaps in more detail, as to the nitty-gritty of the material. I will plan on using some of the documents in this paper so both papers are worthy of your time.

a. includes <u>Attitude Toward Instructing</u> (ATTINST) – using 10 bi-modal adjectives (a descriptive scale) to determine their feelings/attitude toward ground school and classroom instruction.

b. includes <u>Attitude Toward Flying</u> (ATTFLY)– 11 statements on flying given and the student makes a judgment on how he feels about each statement.

2. <u>Personality test</u> (PER) – One of nine personality traits is determined for each student.

3<u>. PRE TEST</u> (PRE) – series of C-21A equipment questions to set a base line for knowledge of the airplane.

4. <u>POST TEST</u> (POST) – matched series (either could be pre/post) airplane questions at end of the course.

5. <u>EVALUATION</u> (EVAL) - Student evaluation and reaction – There are 20 questions covering four areas.

FIRST THREE PAGES OF QUESTIONNAIRE

FOI QUESTIONNAIRE – Page 1 NOTE: This is the form for first data line in the SPSS program:

ID, GENDER, AGE, COLLEGE, ACTIVE DUTY, and FLYING TIME.

(cut off is intentional)

The first part of the Questionnaire is for the DESCRIPTIVE statistics in the program. The computer will get the TOTALS of each variable like GENDER = MALE – 177, FEMAIL – 23; the codes for COLLEGE are AFA - 1, ROTC – 2, OTHER – 3. Note FLY is flying time and truncated two places for example 8 = 800 hours, and 41 = 4100 hours. FREQUENCY of these variables is given from the computer program[4] in the following table format:

---

[4] The program is *SPSS* vs 13.0 and was quite popular in the last 25 or so years. This is a student version and is truncated in capacity and production.

This box shows GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALE	177	88.5	88.5	88.5
	FEMALE	23	11.5	11.5	100.0
	Total	200	100.0	100.0	

This box shows COLLEGE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	USAFA	90	45.0	45.0	45.0
	CIVIL ROTC	107	53.5	53.5	98.5
	OTHER	3	1.5	1.5	100.0
	Total	200	100.0	100.0	

Upping the statistics SPSS shows descriptive mean/average of the MALES and FEMALES in COLLEGE

In the GENDER case we could have added up the total 1s for 177 or the total 2s for 23 to get the total of 200. A little more math with the 1s and 2s codes to determine the AF Academy graduates in both MALE and FEMALE is shown below. Figuring the percentages is placing the little number on top in each category and dividing by the total or 82/177 = 46.3%

GENDER = MALE

COLLEGE	Mean	N	Std. Deviation	Sum	Range	Skewness	% of Total N
USAFA	1.00	82	.000	82	0	.	46.3%
CIVIL ROTC	1.00	92	.000	92	0	.	52.0%
OTHER	1.00	3	.000	3	0	.	1.7%
Total	1.00	177	.000	177	0	.	100.0%

GENDER = FEMALE

COLLEGE	Mean	N	Std. Deviation	Sum	Range	Skewness	% of Total N
USAFA	2.00	8	.000	16	0	.	34.8%
CIVIL ROTC	2.00	15	.000	30	0	.	65.2%
Total	2.00	23	.000	46	0	.	100.0%

A good instructor is making notes who went to one school over the other if for no other reason than to know and be able to keep in mind one group were in Colorado and are known as "Zoomies" (politely).

Coming up next are the first two instruments on attitude in instructing and flying. We can determine the mean or average score in each event by adding all the scores up and dividing by 200, or by 177, or by 23. After that we will be comparing these means to see if there is any **correlation** – another big stat term!

The Second page of the QUESTIONNAIRE – Attitude Toward Instructing (The score under ATTINST)

This instrument is (has been) one of the most important tool used in the class. We will see later on when doing correlation statistics just how much significance this finding does to the overall reports or numbers taken in the class. An instructor pilot candidate who has a high attitude toward instruction will certainly be more successful than one who does not. This simple picking of a bi-model adjective, critical as it is, will yield a very important part of the student's whole picture.

Note: an applicant must take the time to compare the direction of the words as the strong side may be on the left - or the right - as this was designed to keep the student from going down one side of the page in a hurry.

At the graduate level this simple instrument will/could be the most important document a student can design in his statistical research work as a subject's attitude can be compared or correlated to many of the other variables to 'support' the researcher's research question(s) or maybe his hypothesis.

Third page of the QUESTIONNAIRE – Attitude Toward Flying – (the score under ATTFLY)

This instrument is the second most important document in the course as it too will correlate well with the outcomes, goals, and tasks of the course. This will tell at a glance just how much the student feels about flying and taking this course on becoming an instructor pilot.

My first review glance is at number 39 or fourth from the bottom *"I am potentially the best officer, pilot, and instructor candidate in this class."* I have stood up on my desk top and loudly challenged the lieutenants (in my tertiary language) on anything less than a score of 9…actually, this maybe the only real question that needs to be responded to.

Course, this goes into some cognate, pertinent, and collateral input. An instructor should know the personality of the student he is going to stand up and holler at – it helps…just like the coach getting into the face of his new quarterback.

Before I jump down from my podium pulpit I ask "What number do you think I would put there?" Then we look at number 34 *"Really, I'd rather be back home than sitting through this particular class."* I really do miss this class and these lieutenants. Looking back to review and play again in my ole head is a joy.

USAF C-21 FOI QUESTIONNAIRE

Attitude Toward Flying

Below is a list of statements about your flying and career in the Air Force (you may have heard some of them before). Please answer each one of the comments by placing the number that best reflects your personal attitude or "feeling" toward each statement. Use this scale below:

9	8	7	6	5	4	3	2	1
Strongly Agree		Agree		Neutral		Disagree		Strongly Disagree

(32) _____ CRM is a very fascinating subject.

(33) _____ Being able to brief systems is the most important part of being an instructor.

(34) _____ Really, I'd rather be back home than sitting through this particular class.

(35) _____ Becoming an instructor is more for my career than my personal desire.

(36) _____ As an IP I can be a positive influence on other pilots.

(37) _____ Having to sit through most CRM training takes the fun out of flying the jet.

(38) _____ We spend way too much time on minutia while briefing systems.

(39) _____ I am potentially the best officer, pilot, and instructor candidate in this class.

(40) _____ A good instructor should know as much about the jet as the mechanics do.

(41) _____ For systems briefings, if it is not in the dash-one, we should not brief it.

(42) _____ Studying human factor issues may keep me alive longer than studying aircraft systems.

- Truncated by design

Now, we are still in the morning of the first day (somewhere) after about 26 minutes on statistics from the Questionnaire…

A Note on STATS

So far the data variables AGE, ACTIVE DUTY, and FLYING TIME for each case (subject, person, pilot) has been data called QUALITATIVE as they are measures or quantities that can be put into an order or ranked in some way. However, GENDER, and COLLEGE are QUALITIVE data and cannot be comparable by arithmetic relations. This is more explanatory than anything else but we need to know; however, the numbers or scores the MALES and FEMALES respond or report in the tests, exams, or survey can be used in a QUANTITIVE manner.

Small point worth making as for example the MALES were recorded with a mean of 95.96 and the FEMALES recorded a mean of 96.35 – actual means in the EVAL or the FINAL EVALUATION for this course. If you look back to the first page score box the MEAN for both (total subjects/students) is 96.01.

ATTINST vs. ATTFLY

The scores for ATTINST and ATTFLY are the next two to work a quick example of expanding our 'Systics.' We have already said we could ourselves add up the score total for the ATTINST column and divide by 200 and come up with a MEAN. That number by the way is 87.34. The MEAN for ATTFLY is 75.01.   The MEAN of either GROUP could be compared to some sort of average, a cut-off point, or another MEAN like the MEAN in ATTFLY. Again statistically comparing the MEANS…

Later on we will let the computer determine what the relationship is between the MEANS of both of these variables. We will be looking to see if the scores/numbers are close or far apart and let the computer tell us if there is any significance – meaning is there a true relationship between the two.

Course, statistically the computer can determine the difference between the groups like the MALES/FEMALES or the COLLEGES in the variables like ATTINST, ATTFLY, or the last three (two tests, and one evaluation).

Below is bar graph of the MALE – FEMALE (left-right). Left bar each is ATTINST (lower bars are ATTFLY)
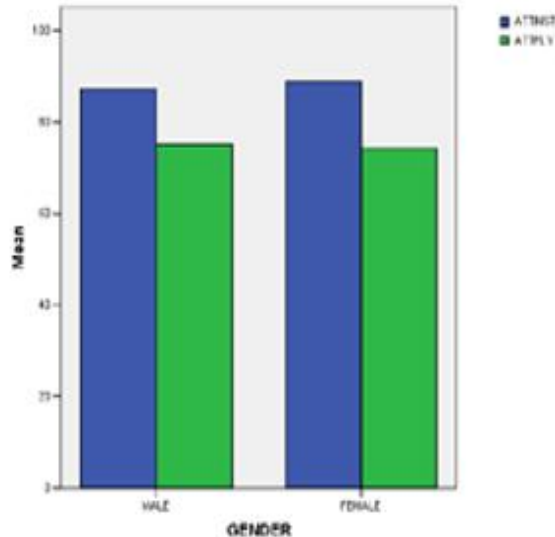
The Training Variable Called PERSONALITY

Another little test given in class on the first day is from a small program I picked up at the University of North Texas when I went back for some post graduate work – in Psychology. It is helpful from a teaching point to have an idea/feeling or knowing about what kind of personality a student has, and of course, a student will realize gain from knowing which personality.

It should not surprise very many of us to see just how many in the class are perfectionists, helper/caretakers, and powerful individualists as these three constitute about 60% of the total. It pays from a teaching view or sitting next to him in the jet to know where he is coming from. Yes, it is a toss up using the 29 powerful Individualists over the 26 Observer/Philosophers. In statistics a spade is a spade and the numbers reflect this order.

However, I want to compare this subset (N = 200) to the original set (N = 400) and see just how close it is.

Personality (N = 200)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Perfectionist	55	27.5	27.5	27.5
	Helper/Caretaker	38	19.0	19.0	46.5
	Achiever/Performer	8	4.0	4.0	50.5
	Authentic Individualist	9	4.5	4.5	55.0
	Observer/Philosopher	26	13.0	13.0	68.0
	Loyal Team Member	4	2.0	2.0	70.0
	Optimistic Generalist	20	10.0	10.0	80.0
	Powerful Individualist	29	14.5	14.5	94.5
	Team Player	11	5.5	5.5	100.0
	Total	200	100.0	100.0	

Let US compare the four with frequency and percent from the smaller sample to the large:

	N = 200		N = 400	
Perfectionist	55	27.5%	105	26.3%
Helper/Caretaker	38	19.0%	68	17.0%
Powerful Individualist	29	14.5%	73	18.3%*
Observer/Philosopher	26	13.0%	46	11.5%

* Notice in the larger sample the Powerful Individualist I was 18.3% and in second place – by five numbers. Well the question in the last part will be to compare and see if there is <u>any significance</u> in one mean over the other while we look to see how close the means are.

There

are three

items to discuss

at this point to help

us with the catch up: 1. The

statistics will give us the mean

of each one of the variables like age,

as the average AGE of the lieutenants is

28.6 years, and the number of females is 23

which is 11.5%. These are ==DESCRIPITIVE== STATS.

2. Involves with the scores of a couple of instruments

Like the ATTINST which gives us the mean score of 87.3

and the ATTFLY score which was 75.0. We have looked at these

two and compared the means using a Paired Samples Test to show the

correlation of .243** which means it is significant. These tests are into what

is called ==INFERENTIAL== statistics (to make a guess from samples as to what the

population may be). 3. The third part has to do with the way statistics are

explained or shown in text form and written up in journal articles or like

this chapter is being reported. We are using tables and charts to

visually show the means, standard deviations, reliability,

and coefficients of the surveys, tests, exams, and

evaluations. This section, demonstrated in

the FOI Questionnaire is called No. 3

--- ==EDUCATIONAL RESEARCH== ---

Next, for the last part of our

paper, it has to do, with

yes, all with the one

(no help on this…)

word, ready?

"STYSTICS"!

Correlation

(And Computing Correlation Coefficients)

This was the beginning in statistics where the hard work began to pay off as I could now see how to do the math and correlation was the means to determine not only the relationship between variables but more precise "how the value of one variable changes when the value of another variable changes" (Salkind, p. 77). The numerical index that reflects the liner relationship between variables is called **correlation coefficient.** It is represented by the small letter *r*. The value ranges between a value of -1 and +1 or in indirect relation or a positive relation.

Another way of saying this is the correlation coefficient "reflects the amount of variability that is shared between two variables and what they have in common" (p. 80). The text book example of this is height and weight would normally to be correlated as they have a lot of other shared characteristics. Also IQ and GPA have relationship that will reflect high, medium, or no relationship meaning for example the more IQ you have the more you would expect to see a higher GPA.

We have been looking at ATTINST and ATTFLY statistically and have the numbers so what does the software, SPSS, computer, or straight math tell us about the how much they have in common? The table below, in simple terms, says the correlation is .272 or about a third is shared. The table also reflects the correlation as being significant (**).

Correlations		ATTINST	ATTFLY		
ATTINST	Pearson Correlation	1	.272(**)		
	Sig. (2-tailed)		.000		
	N	200	200		
ATTFLY	Pearson Correlation	.272(**)	1		
	Sig. (2-tailed)	.000			
	N	200	200		

** Correlation is significant at the 0.01 level (2-tailed).

Let me explain a little more about this procedure and I want to use the data and results from the **PRE** and **POST** tests. What happed was these 200 students took these two tests and SPSS reported the two scores by each student were statistically compared.

From my teaching perspective I asked the questions "What was the difference in the scores between the two tests? Did my treatment of teaching the class cause the scores on the POST test to be higher? And if it was higher, was the difference statistically significant?

I remember the professor in my first 'stystics' class had us do BY HAND more than one of "The Karl Pearson Product-Moment Correlation Coefficient" problems. I loved it. He called the procedure "The big old hairy one." Course, we didn't have 200 subjects in our problems but having a large number is what computers are for.

Surprising (or not?) the two ATTINST-ATTFLY and the two PRE-POST correlated well and both were significant showing in the latter that the treatment was **significant**.

Correlation Matrix

As you can see below, there are seven variables illustrated in this correlation matrix. The pairs each have a correlation coefficient. For example, there is the PRE-POST coefficient of .243(**) seen across from PRE and down from POST. You could do the same thing going from PRE at the top and going down to POST for the same coefficient. The variables match perfectly with themselves, noting the 1.00s down the diagonal, and you can see the matrix creates a mirror image of itself.

Correlations		*AGE*	***FLY***	*ATTINST*	*ATTFLY*	*PRE*	*POST*	*EVAL*		
AGE	*Pearson Correlation*	*1*	***.808(**)***	*.139(*)*	*.100*	*-.136*	***-.157(*)***	*.046*		
	Sig. (2-tailed)		*.000*	*.049*	*.161*	*.055*	*.026*	*.517*		
FLY	*Pearson Correlation*	*.808(**)*	*1*	*.101*	*.145(*)*	*-.190(**)*	*-.074*	*.060*		
	Sig. (2-tailed)	*.000*		*.153*	*.041*	*.007*	*.298*	*.401*		
ATTINST	*Pearson Correlation*	*.139(*)*	*.101*	*1*	*.272(**)*	*.019*	*.070*	*.115*		
	Sig. (2-tailed)	*.049*	*.153*		*.000*	*.785*	*.323*	*.106*		
ATTFLY	*Pearson Correlation*	*.100*	*.145(*)*	*.272(**)*	*1*	*-.079*	*-.032*	*.104*		
	Sig. (2-tailed)	*.161*	*.041*	*.000*		*.264*	*.649*	*.143*		
PRE	*Pearson Correlation*	*-.136*	*-.190(**)*	*.019*	*-.079*	*1*	*.243(**)*	*-.068*		
	Sig. (2-tailed)	*.055*	*.007*	*.785*	*.264*		*.001*	*.338*		
POST	*Pearson Correlation*	*-.157(*)*	*-.074*	*.070*	*-.032*	*.243(**)*	*1*	*-.054*		
	Sig. (2-tailed)	*.026*	*.298*	*.323*	*.649*	*.001*		*.450*		
EVAL	*Pearson Correlation*	*.046*	*.060*	*.115*	*.104*	*-.068*	*-.054*	*1*		
	Sig. (2-tailed)	*.517*	*.401*	*.106*	*.143*	*.338*	*.450*			

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

At a glance the matrix shows, for instance, a high direct relationship with **FLY** and **AGE** (the strong relationship with a positive **.808(**).** This makes sense as more flying time is associated with age or years flying airplanes. Note the indirect relationship with **AGE** and the POST exam. Says the older pilots do not have a strong relation with taking tests on aircraft equipment toward the end of class (the negative, or directional minus number **-.157(*).** This 'elegant solution' also shows how much or how strong the relationship is between the two equipment exams in the class or the PRE and POST exams with the correlation of .243(**). The statistics shows the relationship between the student's scores in both tests meaning the tests have a lot of variability and a lot in common.

More later on the (.xx* and .yy**) as to significance - has to do with probability in sample to population.

To show another and more specific way to illustrate correlation coefficient for just the two variables PRE and POST seen the matrix below. This says the same thing as the matrix above with all of the seven variables listed.

Correlations		PRE	POST		
PRE	Pearson Correlation	1	.243(**)		
	Sig. (2-tailed)		.001		
	N	200	200		
POST	Pearson Correlation	.243(**)	1		
	Sig. (2-tailed)	.001			
	N	200	200		

** Correlation is significant at the 0.01 level (2-tailed).

Below is a table for the means and standard deviation of the descriptive and quantifiable variables. Standard deviation (SD) is another one of my favorite ("do it by hand!") statistical procedures as it adds to what MODE, MEAN, and MEDIAN (measures of central tendency) of a set of scores will tell you by adding in measures of variation. These are called range, quartile deviation, mean deviation, variance, and standard deviation.

"**Standard deviation** (abbreviated as *s* or SD) represents the average amount of variability in a set of scores…it's the average distance from the mean. The larger the standard deviation, the larger the average distance each data point is from the mean of the distribution" (Salkind, p. 37). Gay says "…standard deviation indicates how spread out a set of scores is, that is, whether the scores are relatively close together and clustered around the mean or spread out covering a wide range of scores" (p. 328).

To explain SD look at the means below and see the PRE mean score of 51.93. The SD is 10.78 which shows a pretty good spread. The grades are spread out and not clustering around the mean. Same thing with POST scores: they are also spread out but the mean is higher. To compare another variable we can see the EVAL standard deviation – as well as the AGE standard deviation – both are closer to being clustered around the mean and is not spread out.

Descriptive Statistics

	Mean	Std. Deviation	N
AGE	28.60	3.901	200
FLY	13.85	8.551	200
ATTINST	87.34	7.967	200
ATTFLY	75.01	7.102	200
PRE	51.93	10.782	200
POST	70.63	11.174	200
EVAL	96.01	3.982	200

Knowing about SD, for some of us 'statisticians,' is the precursor to success in the understanding basic statistics. One just needs to know and it is not that hard - beyond the scope of this paper - but I did teach it to my oldest granddaughter when she was in the ninth grade. It took just about 13 minutes – using the numbers 7, 8, 9, 10, 11, 12, 13.[5]

---

[5] The methodology and answer are on my Web site: www.gheart.net/reminiscences then click part IV, page 307.

Final Statistics: The *t* test

This paper has been discussing post class statistics using a sample of 200 subjects (pilots) from a sample of 400 subjects (pilots) from a paper on FOI statistics in 2010. I want to clarify the numbers, samples, and how important the *t* test was to the seemingly success of the class noting the final evaluation scores.

First of all the number of students from 1998 to 2014 was 869 students. After a few years the tests, exams, and procedures developed to the point where I could start with a standardized group meaning all had the questionnaire, pre test, attinst, attfly, personality, post, and evaluation data. When about 400 total were complete I administered the statistics to see what the sample (N = 400) would yield. So, for the first few years and the last few years not everyone was given the 'treatment' noting (1) is was no longer needed to support my research questions, and (2) I no longer had some of the testing material available.

The results, again, for the N = 400 is in *Action Research: A Statistical Report on IP Upgrade* (2010) and is available on request but will be in the upcoming book on this 16 years teaching FOI. The N= 200 is a sample of the N = 400. My SPSS would not let me work with more than 200 subjects so I took a sample of the large group and ran the same *t* test to see if the results were the same, close, or something else. I was not surprised to see how close the numbers were. We discussed this at length with comparing the personality scores of both groups and in the *t* tests noted in the abstract. This means the first sample findings (400) matched the second (subset) findings (200).

Statistically the minimum sample required (S) for a given population size (N):

N	S
100	89
200	132
400	196
869	267
1500	306
3000	341
15000	375
100000	384

A couple of observations about these numbers: (1) each subject from the sample must have the same equal choice of being accepted – everyone has the same chance to be randomly selected and (2) the N number is almost irrelevant as long as everyone 'could be picked.'

Going back to my sample of 400 for 869, or 200 for 400, my numbers were almost overkill and that is just fine as most books say the larger the sample size the better the data will be for running the stats. I could almost be accused of the 400 actually sampling the population and not being the sample. If this is true then some of the tests do not or should not count. For instance, if one is using the population doing a *t* test to determine if the sample results that could be representative of the population the *t* test is not applicable because you have already determined the results for the whole population. The significance of the test would not be meaningful as the prediction is not applicable. Gay says this best: "Thus, a test of significance is made when we wish to determine how probable it is that the differences we have found between our samples will also be found in the populations from which they are draw" (p. 353.

Both the 400 and 200 test results show, or better, suggest, they are the same samples.

<center>Evaluation</center>

One of the variables to interest a teacher is the final evaluation (EVAL). For the N = 400 the scores were quite high reflecting a mean of 95.6 with a standard deviation of 3.46 and, of course, no surprise with the N = 200 reflecting a very close mean of 96.01 and a standard deviation of 3.98. Those are very high grades for the class and I always wanted to be able to take the credit and show statistically how significant these data were; however, referring to the correlation coefficient on the page 11 matrix there are no or darn little correlation with any of the other variables – and of course no significance.

I was able to show in the N = 400 paper that the mean increase from the first set of 200 students to the final total of 400 there was significance and this would reflect the population would also have the higher score.

<center>PRE – POST *t* test results</center>

The bar chart below shows/illustrates the difference between the MALES and FEMALES and their scores on the PRE and POST tests. In both tests the scores were very close showing an increase of about the same amount of increase in the POST test.
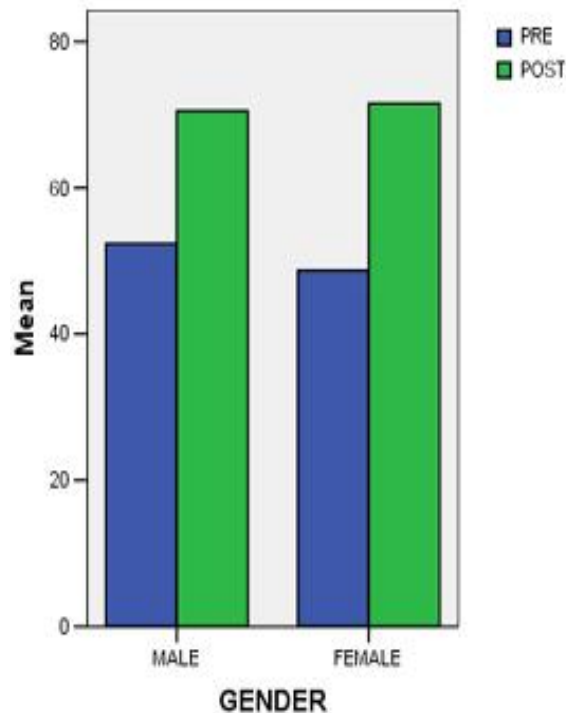
Referring to the same correlation coefficient matrix, and as we have been talking about, there is a difference in the total scores from the PRE to the POST and it shows significance. I decided with the help of Salkind again on the appropriate test to run to see how much and how significant the scores and correlation were and if it could be determined where the significance came from.

The Correct Test Statistic for examining relationship between two variables is a *t* test for the significance of the correlation coefficient and SPSS put the data through a Paired Samples Test. Below show the means and standard deviation with a correlation of .243 significant at the .001 level.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRE	51.93	200	10.782	.762
	POST	70.63	200	11.174	.790

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PRE & POST	200	.243	.001

Paired Samples Test

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 PRE-POST	-18.700	13.508	.955	-20.584	-16.816	-19.578	199	.000

      The difference in the means is 18.7 [SPSS always used the first number so the table shows a negative number (-18.70) and yes, I know you all probably don't care about that little detail]. This is the difference between the PRE and POST means.

      Salkind's guidance suggests it is unlikely the *t* score = -19.58 was obtained by chance meaning some treatment caused it.

      df stands for degrees of freedom which is "a value for different statistical tests and approximates the sample size of number of individual cells in experimental design" (Salkind, p. 383).

      The stat book has a Table to get the *t* values needed (for rejection of the null hypothesis – which says there is no difference in the means). The value (df vs. 0.05 significance) I had to look up is 1.65.

      So, the *t* score of -19.58 is greater than 1.65 which is the critical value for rejection of the null hypothesis meaning the means are (again) not the same and the change (*t* score) suggests in this case the treatment or training was the cause for the rejection.

      The last box Sig. (2-tailed) .000 is certainly beyond the scope of this paper but it indicates the probability is greater than 5% that any one test of the null hypothesis, the two groups differ because of the way they were taught (p. 167).

      What this means is the data suggests and the test statistic for the data suggests it was not by chance the POST mean was higher but it was the treatment the students received in class caused the change and it was significant. This last part is always good in academics. You should show significance!

A Statistical Finish - with Reward

Want to show you a picture of my oldest grandkid, Claire. She has the moniker "NOG" for Number One Granddaughter. She is the one I taught how to do standard deviation but it was a few years after this picture was taken in the simulator with two distinguised lieutnenants - getting some G.Heart Country training. Claire has just asked me "How come the motion is not on?"

Speaking of standard deviation I want to finish this by saying thanks for reading this paper and I hope some of it will pay off in the future; for sure, the reward for the reading is to give you the formula for doing SD. I cannot tell you how important it is with statistics that you know and understand what SD does and how often it is used in solving problems (and doing the work of the Lord).

STANDARD DEVIATION is the goo of statistics...you can't fly without it!

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

g♥