



Virtually Professional

Volume 1, Number 1 — July 2000

Note to TAIR Members: This is the first in what TAIR hopes will be an on-going series of electronically distributed professional development articles about the practice of institutional research entitled “TAIR Virtually Professional.” Initially, the series will be published semiannually, but the frequency may increase if TAIR members express interest in reading and writing for the series. If you would like to submit an article for consideration, if you have suggestions for future topics that you feel would be of interest to TAIR’s members, or if you have some constructive feedback, please contact a member of TAIR’s Professional Development Committee. Members are identified on the TAIR Web site at <http://www.texas-air.org>. The Professional Development Committee extends its appreciation to Dr. Stan Adelman for his contribution in initiating this series.

Put Away Your Inferential Stat Book and Re-read Campbell and Stanley

by Stan Adelman, Ph. D.

Amarillo College

Inferential statistics are used to estimate whether population differences exist, based on the differences between random samples drawn from the population. Sound familiar? That’s what I learned when I was last in a statistics class (back in 1972!). That’s *not* what I’ve encountered in 20 years of institutional research.

For any given one of those 20 years, the women at Amarillo College have a higher percentage of courses passed with “C”s or better than do the men. I suspect the differences, based on a Chi square, would even be “statistically significant.” So what? In each of those 20 instances, the difference was between two different *populations*, not between two different *samples*. The performance difference was based on *all* students enrolled that year. Hence no sampling was involved.

Remember, samples *must* be drawn randomly from the population they represent. Consequently, inferential statistics cannot be used to predict the performance of future populations from the performance of previous ones. While the students could be considered samples of the population students, or students ever to attend Amarillo College, absolutely no effort was made by the admissions office to ensure that any year’s students were randomly drawn from all students who would ever attend Amarillo College. Consequently each

year's students must be considered a separate population. The differences measured each year are absolutely real, and, to the extent that my computer program is correct, an absolutely valid description of the difference in academic performance between men and women for the population, students attending Amarillo College during the year X.

Using inferential statistics in a case like that described above can lead to two problems. First, the use of inferential statistics can lead to a false sense of confidence in the results. If the differences are statistically significant, then they must be, to a non-technical audience, real. Well, statistically significant or not, the differences measured actually occurred. And the fact that the differences occurred in 1981, or 1985, or 1995 does not mean they will occur again in 2000. At least you cannot legitimately use any technique learned in inferential statistics to show that such a result would be highly likely. Each year's class of students is not randomly drawn from any common population, and that is, at least as I learned more than 25 years ago, the fundamental basis for the validity of using inferential statistics.

On the other hand, the fact that differences occurred not only in 1981, 1985 and 1995, but in every year from 1980 through 1998, and that in all but one instance women outperformed men, might—statistics aside—lead us to expect that the women of the year 2000 class would again outperform their male counterparts. Similarly, we come to expect that if we put water in a pot on a stove and turn the burner on high, sooner or later the water will boil. We really don't need inferential statistics to come to that expectation. I mean, after a certain age, a child, when told by mom or dad that something is hot, eventually learns to stop testing the hypothesis.

Using inferential statistics can lead to a second problem, as well. That problem centers on one of at least two meanings of the word significant. The word itself, used in the context of "statistically significant," is almost a fallacy of four terms (yes, now you've got to go dig out your logic text). Take the statement, "We have found significant differences between the performance of men and of women." The statistician means by such a statement that, were one to draw two more randomly selected samples of men and women, it is quite likely that the women would again outperform the men.

What does the lay person mean by the word significant? Typically, a lay person means, "Boy, the differences in the performance of men and women are really important." After all, a significant event in history isn't (necessarily) one which will again occur—it's an important event. The discovery of penicillin was a significant (important) event, but I doubt that discovery will occur again any time soon. (Being a devoted science fiction reader, I wish to leave open the possibility that the human race might at some time in the future forget how to

produce penicillin, thus making possible once again the discovery of penicillin by some future generation.)

By co-opting a very nice word, which has a fairly common English usage, researchers may very well say, once they've discussed statistical significance, that they've finished discussing the significance of the findings. Have they? Just how important are the differences found? That's not a statistical question at all.

Now it's table time. Here's a table of the differences in performance between all men and women enrolled each fall at Amarillo College, fall 1985 through fall 1998.

Course Pass Rates with Grades of A to C

Fall	Women	Men
1985	71%	68%
1986	73%	67%
1987	72%	65%
1988	71%	65%
1989	70%	63%
1990	74%	67%
1991	74%	67%
1992	74%	66%
1993	73%	66%
1994	72%	67%
1995	70%	66%
1996	66%	67%
1997	72%	64%
1998	75%	69%

How significant (important) are these differences? I guess my first question is what happened in 1996? That aside, what, if anything, should Amarillo College do about this? Most years, the women outperform the men by around six to seven percent.

- Should advisors call male students during the term?
- Should a mailing be sent to male students reminding them that free peer tutoring is available to all students?
- Should instructors schedule conferences with male students who perform poorly in class?

The equity of such suggestions leaves me, at least, feeling a bit queasy. Shouldn't we be doing these things for all students? What about the 25% of the hours women take that result in D's, F's, and W's? The fact that the percentage is smaller than that of men doesn't by definition make it good, does it? Does "better" always achieve "good?"

I hope I've given you pause. By and large, at our institutions, we're the keepers of the numbers, and we're the first ones to interpret them. In a very narrow meaning, we have specific expertise in knowing whether differences are significant (likely to recur). We also have the experience to spot differences. But we do *not* have any special training in knowing whether differences we spot are significant (important). We best serve our institutions when we point out differences we think likely to recur, and then join in discussions of how important those differences are, and what, if anything, should be done about them.