Special Silver Anniversary Address
A Quarter Century of Army Design of Experiments Conferences

Frank E. Grubbs
Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

Welcome to the 25th Conference on the Design of Experiments in Army Research, Development and Testing. A whole quarter of a century has passed since Sam Wilks recommended that the Army start this series of conferences, and what an excellent idea he had, with all the vision for the future of Army statistics. As we all know, Wilks was a very remarkable man: a gentleman, a good leader, an outstanding scholar and research statistician, a man who also had very vital interests in applications, and he liked to see people work together. Sam traveled much for the Department of Defense and he consulted widely on all probable areas of statistical application for the Government. He missed none of these conferences, and we remember him so well in his role of selecting many of the key statisticians in the universities to participate in these conferences, as we met at the Cosmos Club in Washington, and drank and dined with Sam. I think we have a better pay-off from these statistical conferences than the other DOD conferences, because of the close interface with university statisticians, including, of course, the eminent statistician who gives the keynote address next. To the memory of Sam Wilks we owe so much, and therefore in 1964 we devoted the Army Design of Experiments Conference to Sam's memory.

This particular conference, the landmark 25th, is also dedicated to our good friend, fellow mathematician and statistician, teacher, and excellent administrator, Dr. Francis G. Dressel. How would the design of experiments conferences ever have survived if it had not been for Francis?, who carried always the big part of the load. I am asking him to make a few remarks next. I am glad Francis has trained Bob Launer so well too!

Now how in the world did I ever get saddled with a "Special Silver Anniversary Address," including the fancy title that came from, I might say, a former friend? It is probably because some of my colleagues saw me enjoying being too much a "free lancer" at these conferences, so they thought! I must stretch the exact title of my talk a bit to cover more statistics. Back in the mid-thirties, I was a timid, very illiterate Southerner, trying to eke out a living by teaching engineering math at (now) Auburn University, and it became starkly clear that we lowly instructors would hardly ever be promoted unless we got a Ph.D.! But it was also made quite clear to us that getting a Ph.D. would not make us a better teacher! A quick OR study (not so-called then) convinced us that we should seek something in applied math, as an outlet, and statistics was the subject to study, for it was needed and spreading fast too, for example to our agricultural experiment station problems. I had a friend, who got a Ph.D. at Cornell University, and had a good job at our experiment station, so that he introduced me to the analysis of variance, which seemed to be a misnomer, and he even alarmed me with the idea of the analysis of covariance! I later heard that Karl Pearson was the greatest statistician of all, but that one R.A. Fisher was not sprinkling holy water on all the things Karl Pearson had done. My friend wanted to know if I had read any of the reputable journals on the subject of statistics, and I hadn't, of course. In fact, in about 1934 a paper by a young genius on the distribution of quadratic forms in a normal system, with applications to the analysis of covariance had appeared in a (strange) journal called *The Proceedings of the Cambridge Philosophical Society*, written by

our keynote speaker of today. Later, in the early 1940s when I had been in uniform at the BRL during World War II, we had trouble justifying a journal called "*Biometrika*" at first, and there was no way the Army would ever approve *The Annals of Eugenics* in our Library even if it contained statistical papers!

In any event, back in the mid- to late-thirties it did seem that statistics would be a good choice of graduate study with an outlet. But where in the United States could one study statistics? He certainly could not do so down South, and in fact there were only two, maybe three, places to go for statistics courses – Iowa State University, the University of Iowa, and perhaps the University of Michigan. The latter was mostly an actuarial school with well-known voices such as Menge and Glover, with Cecil C. Craig and Paul Dwyer coming along, and perhaps most interesting of all an athletic statistician and actuary, Harry Clyde Carver, who would challenge his graduate students to beat him at any sport of their own choice. If Carver won, there would be a stiff final exam and no A's! One had to beat Carver at his own game!

We selected Michigan, for Iowa State University seemed too far away, and Alan T. Craig of Iowa was scheduled to give the basic graduate statistics course at Michigan in the summer of 1937, and what a good start to learn to throw dice, et al! That summer, I tried to learn what a random variable was. I had known Clifford Cohen for years back at Auburn, and at Michigan, Clifford of all things, had elected to write his dissertation on the very obscure subject of truncated sample theory! But how in the world could there ever be much interest, let alone wide applications, of such an odd topic? A colleague, who had treaded this mill before, assured me that "Clifford Cohen was a very smart man – he chose a topic no one else would ever work on!" The advice continued, "Don't ever pick a 'hot' topic, for a genius will beat you to it, and you'll never get your degree!" The non-statistical graduate students in mathematics frowned at anyone studying statistics, as it was not as important as topology or even differential geometry either. One of them was somewhat friendly though, and on occasion would drop by our "flat" with his wife. He would pick up and continue my wife's needlepoint with much enthusiasm, but didn't care for statistics then, or any part of it! His name – Jimmy Savage. And it goes to show you what can happen to a pure, rigorous mathematician, once he is "bitten by the bug" or otherwise the clever ideas of the Reverend Thomas Bayes!

It was not easy to find and settle upon a dissertation topic without some guidance, but all the professors already had too many graduate students, and they had passed along topics to some they never heard from, so that I had "better look around in the library." No one then told me, for example, that concerning truncated sample theory this would develop into the field of order statistics, and moreover blossom into reliability, life-testing, et al, and it was in fact many, many years before that did occur. You see, no advice I had been given really sunk in, for I decided to work on outliers, and the international situation had gotten so gloomy that writing a dissertation would not be done very quickly anyway. So being a reserve officer, I was introduced to the Army.

The physical and engineering sciences were just beginning to make some uses of statistics, although Walter Shewhart had made applications of statistical quality control. In 1941 at Michigan I got a hold of a new book, <u>An Engineer's Manual of Statistical Methods</u> by one Major Leslie E. Simon, and as I read it and was enlightened by the book, some correspondence

developed, for here was an authority who had the vision, the wisdom and the courage, of all things, at the time of a very low cycle on the Bayesian revolution (so that we later kidded Les Simon that he had the unmitigated gall) to publish in the back of his book some I_Q Charts to estimate the fraction of defectives in a lot by using Bayes' equally likely hypothesis! Today, there seem to be no 100% classical statisticians, so that we can chalk up another win for Les.* And this introduced me to the Army's Ballistic Research Laboratories, on active duty in uniform as a Lieutenant, where there was never to be an end to all kinds of knotty statistical problems.

As pointed out in Les Simon's book, An Engineer's Manual of Statistical Methods, Dr. L.S. Dederick had worked out the probability of the sample range (largest minus smallest observation) back in 1926, and had partially tabulated its distribution, but wouldn't submit it for publication! Sam Wilks had on occasion consulted with the personnel of the Ballistic Section of BRL at Aberdeen Proving Ground on various statistical problems. Also, since the dispersion of shots on a target, as from rifle firings, was often measured by the "extreme spread," or bivariate range, this little nasty statistical distribution had eluded statisticians, and Mr. Philip G. Rust, an industrialist and "rifle accuracy bug," established by sampling shot patterns the distribution of the extreme spread for small sample sizes. Also, on the train from Washington to Wilmington, Phil Rust had told Sam Wilks about it and had suggested that he look into the theory of the probability distribution of the extreme spread in order to study its properties, as it was widely used in ballistics and rifle accuracy competitions also.

When I arrived at the BRL at Aberdeen Proving Ground in 1941, John von Neumann, Robert H. Kent, H.R. Bellinson and B.I. Hart had just worked out and published in the *Annals of Mathematical Statistics* the distribution of the mean square successive difference, and the mean square successive difference to the variance, and B.I. Hart had calculated percentage points of both.

The "real word data," coming out of ballistic testing of all kinds, often defied any good or "normal" analysis, and were loaded with outliers! There was thus an applicable dissertation topic! In the mid-forties, there existed a critical need to do something about speeding up the production of firing tables, as about 100 female "computers" were always busy running computations on those big, heavy desk electric Friden or Monroe calculators.

Leslie E. Simon valued brains to solve the Army's problems in ballistics, and he had established a scientific advisory committee with some of the best brains in the physical sciences in the country. What a wonderful and stimulating place to work, less much time to be in uniform and fight the "battle of Aberdeen!"

Back in the early thirties, Simon was Chief of Manufacture at Picatinny Arsenal, and had cultivated the interest and expertise of Walter Shewhart to apply the principles of statistical quality control to the manufacture of ammunition at Picatinny Arsenal. At the BRL in World War II, Simon saw the pressing need to pass on the principles of statistical quality control to industry in the production of ammunition and weapons for the U.S. Army, and he also worked

* In a panel discussion on Bayesian methods on reliability one time, I stated that statistically I was 50% classical, 25% fiducial and 25% Bayesian, but Frank Proschan promptly branded me as a hermaphrodite.

with Harold Dodge of the Bell Telephone Laboratories to start computations of the standard sampling inspecting tables for the Army Ordnance Corps, later put into Military Standard 105A.

Sam Wilks had long been aware of the need for well-designed experiments and hence suggested that the Army start a series of annual conferences to promote statistical methods. Sam suggested that the Design of Experiments Conferences should have three types of sessions: first, there would be some special invited papers by well-known authorities on the philosophy and general principles of statistical design of experiments, then there would be some technical papers presented by Army statisticians, and finally there would be clinical sessions with suggestions from the experts – and we still stick to this format today. These conferences had their beginning 19-21 October 1955 at the Diamond Ordnance Fuze Laboratory and National Bureau of Standards in Washington, D.C. We note that Sam's conferences were Army wide, and attracted DOD interest, while a conference the Ballistic Research Laboratories put on a year earlier (1954) on the use of statistical methods was primarily for Army Ordnance personnel.

Within Ordnance and the Army, Leslie E. Simon certainly was the great stimulus to the advancement of statistical methods, for at the BRL Les was not only its Director, but he also prepared a large number of papers on engineering statistics or statistical engineering – whatever you want to call it. Moreover, there was a pressing need for these very papers to acquaint industry to the methods of quality control and statistics in connection with the World War II effort. And Les helped promote the short courses on statistical methods in industry. There was a great deal of interest during this period concerning the concept of "economical lot sizes," and also the concept of producing very large "homogeneous lots" so that for ammunition at least we could get rid of the situation where at a field artillery battery site there existed a mixture of rounds from several or many lots with different levels of muzzle velocity and degrees of surface-finish roughness.

In the mid-nineteen forties, a very significant and worldwide development occurred due to an idea of our imminent and esteemed Scientific Advisory Committee member, John von Neumann. He had suggested the construction of the ENIAC or Electronic Numerical Integrator and Calculator – a digital computer – at the BRL. We saw the handwriting on the wall: the ENIAC could be used to Monte Carlo anything to death, in addition to the more straight-forward computations of mathematical and statistically tractable functions. And so many statistical problems were planned! First, however, there had to be some calculations on the distributions of outliers, and the ENIAC staff was looking for work! It was then that I learned about priorities and the real importance of any statistical problem to the country!

Once they got the ENIAC wired for my outlier problem, the Atomic Energy Commission called on Gen Simon and Johnny von Neumann to use the ENIAC to obtain an optimum solution to the problem of imploding the core of a nuclear warhead, with the result, of course, that even though our Computing Laboratory had begged for work, my suggestion and the start of some extensive calculations to keep them busy immediately got a vanishing priority!

In those days of a great scientific effort at the BRL and many of the country's great physicists, chemical physicists, mathematicians, engineers, et al, I felt just like Les Simon had always said, "The engineers would call him a statistician, and the statisticians would call him an

engineer!" Indeed, we were trying to apply statistical techniques to many knotty physical problems for which there was a physical model that applied well. Perhaps I would have been much better off in agriculture! We survived some way or the other and hence got away from the use of primarily the probable error which was never to be deleted from firing tables though!

Now, getting back to the uses of Army Statistics, which led up the the Design of Experiments Conferences, we record that a very good account of the statistics in the Army from the very beginning has been prepared by our good friend Clifford Maloney (*The American Statistician*, June 1962), who traced various statistical interests in one way or the other from very early times – he started out with Daniel Bernoulli in 1777. As pointed out by Maloney, there certainly was much vital interest in medical statistics of the Army from the beginning, and at West Point the graduates needed to know something about the dispersion of shots on a target, and least squares and the adjustment of data. In fact, in early times, the best engineers in the USA were really coming from West Point. Then again, there was always considerable interest in the sensitivity of explosives to impact or shock, the sensitivity of primers and other items of ammunition, which no doubt brought about the so-called "Bruceton Method" of sensitivity analysis, and later developed into the Dixon-Mood "Up and Down" technique at Princeton, and since has been widely used. Naturally, Dixon and Mood were students of Sam Wilks, who again enters the general picture! So you see, the Army did indeed have the most natural needs and demands for the application of statistical methods, and Sam Wilks was the first university professor to recognize this vital development for the good of all concerned, as he was always in touch with so many important applications.

I think that the Army Design of Experiments (DOE) Conferences Sam Wilks started have performed the vital task of fulfilling the need for cross-fertilization of statistical theory and practice, even though these conferences occurred only once a year. It is through the Army DOE conferences that we have become acquainted with each other, discussed common statistical problems, presented solutions to others, learned a lot from the eminent university statisticians and gotten their best suggestions during the clinical sessions. Moreover, this has all stimulated Army statisticians to perform very good work and publish a number of useful results for others to apply. If it had not been for Wilks' vision and the DOE conferences, we would have been off to ourselves, no doubt, working very much more inefficiently.

Now all of you have the nice little pamphlet prepared by Bob Launer for this, the twenty-fifth anniversary of the Army DOE conferences, and we note that the series got off with a bang in 1955 as Bill Cochran led with the philosophy underlying the design of experiments. Churchhill Eisenhart spoke on principles of randomization (isn't that still an unsettled topic?) and John Tukey headed a panel on "Where do statisticians fit in?". Jack Youden aided in his most interesting way of talking about the design of experiments in industrial research and development. The application of order statistics and problems in subjective testing came into the Second DOE conference, and we were fortunate to have R.A. Fisher at the Third conference! Also at the Third conference HO Hartley spoke on changes in the outlook of statistics brought about by modern computers, and Ben Epstein, who at one time even worked at Frankford Arsenal, covered what was to become a very important Army field "life-testing" – and later reliability and reliability growth. Here at Natick for the Fourth conference, it was appropriate to have L.H.C. Tippett discuss statistical methods in the textile industry, and the Fifth conference

taught me a lot about smoking and lung cancer (now forgotten) because of the lively debates between Joe Berkson of the Mayo Clinic and Jerzy Neyman (both smoking, I believe!).

The theme or title of these conferences was stretched many times, I am reminded, to include many important topics of the day or time, and this was necessary and good too! In fact, I note that Egon Pearson gave the keynote address of the Eighth conference on a statistician's place in assessing the likely operational performance of Army weapons and equipment, or the need for statistics in military operations research and weapon systems analysis. In fact, the Army has a parallel series of conferences, started in 1961, known as the Army Operations Research Symposia. I found that the field of operations research was being staffed primarily by mathematicians, physicists, engineers and others, but not enough statisticians, who could aid in their modeling problems of stochastic processes. For example, for probability of hitting problems there was often the need to have simple approximations to the distribution of quadratic forms in normal variables, and techniques like the Wilson-Hilferty transformation of Chi-square to approximate normality and the Polya-Wilson approximation to those darn cut-off normal integrals that were found to be very useful. Moreover, we also saw that the theories of life-testing would apply to Lanchester type combat theory. Because of the critical need for the evaluation of weapon systems, and later many other military operations research topics, the Army OR symposia have attracted a large number of "high brass" type visitors. Statistical topics have been often discussed at the Army OR conferences and OR topics at the Army DOE conferences. Forget titles!

In the spring of 1964, some six months before the Tenth conference, we got the shock of our lives with the untimely passing of Sam Wilks. For the Tenth conference, Les Simon came forth with an excellent and informative paper on the stimulus of S.S. Wilks to Army Statistics, and the high importance of the DOE conferences to Army statistical endeavors.

Fortunately, the DOE conferences have proceeded to cover the water front, and stimulate and train more statisticians.

Now although I have mentioned many of the key benefits and much in the way of significant progress that has resulted from the DOE conferences, let me now jump to a look at the whole series, or the view from an operations research eye. To begin with, it becomes quite clear that we have learned a lot about modeling processes (stochastic) or fitting models to data in order to make more general predictions, or to summarize. "Models?" Yes! And this reminds me of what George Box is quoted as saying, "All models are wrong, (but) some (even) work!" How true this is! Aren't models competitive, and haven't we found that the situation doesn't exist for which only one model is right and all the others wrong? In fact, we are often lucky that any of several competitive models may serve the purpose at hand very well. Yes, I think we have learned how to model many important Army areas of application, and this has also brought about model development or better theories.

Obviously, the great benefit to the Army from the DOE conferences has been the expert counseling of in-house statisticians by the eminent university statisticians who have so kindly given of their time and experience. One has only to look at the little booklet of featured speakers to be very highly impressed with the caliber of the talent. We greatly appreciate this, as their

help, including the clinical session suggestions, has been outstanding, and for very difficult areas of application.

This brings to mind another point. The U.S. Army is a very large and diversified organization. In case you need some converting on this point, just attend one of the Army Science Conferences held biennially at West Point. In addition to our little corners of application we have discussed over the years, at the Army Science conferences, they have presented papers on, for example, sampling the polar ice caps – which brought up many statistical problems of note – or even the extraction and analysis of snake venom! What are the main controversies about? You guessed it: the instrumentation, the measurements and their interpretation. Army investigators have grown increasingly aware of errors of measurement, precision and accuracy, and even just how to define these illusive concepts. And so have others. For example, Committee E-11 of the American Society for Testing and Materials has for some 20-25 years been working on the problem of standardizing the views of engineers, chemists, etc., on the subjects of precision and accuracy, and come forth with a recommended practice. I still don't see an end to this effort, for there are enough "divinely endowed," stubbornly statistical minds to bring about nothing but impasses. (Incidentally, I know that I alone am right though, and they needn't think they can sway me to a compromise!) As a passing remark, there's a full-time job for a young, competent statistician for NASA, the FAA, and such agencies, in connection with sampling the atmosphere in order to establish temperature profiles, ozone content profiles, etc., by knowing the capability of their instrumentation for the first time.

We have learned much about the statistical design and analysis of scientific type experiments, and the construction of designs – the latter, I think! Furthermore, I see evidence very frequently of some "fancy" experimental designs that Army investigators are using, with very sophisticated analyses, too. On the front cover of the program of the Tenth Conference, there is a 10 x 10 Graeco-Latin Square, and no one yet has pointed out an error in it! When the nice, balanced experiments have been violated in one way or the other, speakers like HO Hartley (who regrets that his duties as President of ASA keep him away today) have come along to help or straighten us out. We have used linear models mostly, but have been hit by nonlinear models at times, and George Box has on several occasions given us his unique approach to time series analysis.

There have been many advances over the years in the analysis of contingency tables, and count data generally. We have had many contributors on this subject speak to us, and the several approaches presented to us, including Kullback's information theory approach, and some of the recent work of Feinberg in our preceding two-day tutorial course. Perhaps the U.S. Army Operational Test and Evaluation Agency has made much use of contingency table analyses, and have benefited from them. I will continue to try and sort out that problem, and I note that my experience has been primarily in connection with the comparison of two or more binomial type proportions, and irrespective of Fisher's fixed marginals, and stuff like that, I still don't want to confuse the issue by imbedding the comparison of binomial p's in a contingency table analysis. Maybe the real experts have other views.

Hasn't the field of reliability and related applications hit us with a big bang, to say the least? And the high-level "brass" or managers have shown the greatest of interest in it too.

Remember, I remarked that Clifford Cohen in the late 1930s wrote his dissertation on the obscure subject of truncated sample theory? Well, finally the area came to life and how! Although the normal distribution was the "universal" one in the past, it didn't "take" with the reliability analysts at all, and they aimed for the exponential distribution. At the 1977 Monterey Conference (23rd), a paper was given by Herback, Green and Blumenthal on the "curse" of the exponential model, and they quote:

"The exponential is wrong,
But works like a song.
Beware the Weibull:
It's incorrigible."
-Anonymous

Remember George Box – All models are wrong!

There is some heavy interest in reliability growth, and the whole field of reliability will continue and will continue to expand. There are now so many methods or recommendations for obtaining confidence bounds on system reliability than an appointed committee has not been able to standardize on a technique for DOD. It might be said that sample order statistics are of much importance to the Army nowadays, and often even help to take care of the outlier problem. Finally, reliability analysts have worked on estimation and other properties of the two- and three-parameter Weibull models so much that this has actually aided in the spread of the Weibull distribution to many other areas of application than reliability. Perhaps this is because of the robustness of the Weibull model in representing a variety of shapes.

A very old statistical problem is that of bio-assay types of analysis, and it borders on the estimation of risks and safety levels in any number of other fields. There are many papers on the subject of quantal response, "sensitivity analyses," explosive sensitivity (Bruceton), ballistic limit, Up and Down method, etc. which have been aired in these conferences. Quantal response investigations, and especially the estimation of both high and low percentage points (of unknown distributions), does indeed cover a very important statistical effort for the Army, and it will continue to expand also. Maybe this is an area for which the use of physical models is needed in addition to statistical analyses, or at least a combination of both. We will face more and more safety type problems, for which there will be heavy demands for statistical treatment, while we used to avoid them altogether. Let me mention one nasty little problem having to do with armor protection in tanks, or penetration mechanics, and hence for safety of tank crews as a result of armor thickness determination and design. It is also a statistical problem the Army continues to need help on to estimate the parameters for zero chance of penetration. Do you like continuous distributions that slowly change to a series of binomial and continuous models of some kind? And even approach a binomial distribution with parameter zero? In this case, we start firing at a piece of armor plate of a certain thickness, and for the high-striking velocities we will (usually) get 100% penetrations of the projectiles through the plate, and there will be a "residual" velocity distribution for the projectiles or pieces of projectiles which have penetrated the plate and come off the back. But as the striking velocity is decreased, then the proportion of projectiles penetrating the plate will decrease, ultimately to zero for low-striking velocities, and thus we say that a safety level exists somewhere, or at least we would like to know just where, for example,

only 1 in a 1000 of the projectiles would penetrate. The curve or residual velocity versus striking velocity gets very steep near the bottom, obviously, and it's a challenge to ballisticians and statisticians to deal with the precise and highest striking velocity for which zero penetrations occur. What I am also indicating is that there are many problems of interest for which statisticians and physical scientists must work as team members, and the DOE conferences guarantee just that.

Let us not forget the field of sampling inspection or acceptance sampling inspection, and the DOD's use of standard sampling inspection tables and practices. These are important activities that the Army initiated with the original help of Harold Dodge, and our DOE Proceedings include a number of papers on the subject. This is really the area of statistics that taught us much about operating characteristic curves, or power curves, and the determination of sample sizes, and the like. Thus, many statistical areas of interest spill over into other topics, and so the process continues.

Now I have made my little choices of some of the benefits and topics of value that we have been privileged to be part of in our twenty-five years of Army Design of Experiments Conferences. Perhaps you can expand or improve on what I have covered and hence make more sense out of things. I invite you to do so. In any event, it certainly seems quite clear that these conferences have been very "cost-effective" to the Army.

I think it was Cliff Maloney who once suggested that there should be published a volume of the best papers of the proceedings of these conferences. This assignment I was given made me look through the whole shelf-wide Proceedings, and I agree that there are certain of the papers which indeed should be brought together in some kind of memoirs.

Maybe we can now get Francis Dressell to make a remark or two, as we have dedicated this landmark 25th Conference to him. Francis!