

YOU DEN ADDRESS

BY JOHN A. CORNELL, UNIVERSITY OF FLORIDA

W.J. YOU DEN – THE MAN AND HIS METHODOLOGY



Jack Youden

"Competitiveness Through Continuous Improvement" was the theme of this year's 36th Annual Fall Technical conference and that is why I selected the topic "Jack Youden – The Man and His Methodology" for this address. Jack Youden believed in continuous improvement; he spent his life improving the ways measurements are taken.

Once I had decided to talk about Jack Youden, I began to wonder, "Who was this person for whom

- an award on the topic of Interlaboratory Testing is given each year by the American Statistical Association;
- an address is presented at the Fall Technical Conference each year (and has been since 1973);
- an award (prize) is given by the Chemical and Process Industries Division of ASQC each year for the best expository paper on statistical methods or philosophy that appeared in *Technometrics* during the previous year; and,
- who was this man for whom the January 1972 issue of the *Journal of Quality Technology* was memorialized."

In what follows, I shall try to familiarize you with a brief biographical sketch of Jack Youden's life; present some examples of Jack Youden's methodology; give some measures of the impact of Jack Youden and his methodology on the scientific community; and finally, share some statements about Jack Youden made by those who worked and/or interacted with Jack Youden. Regrettably, I never had the good fortune of meeting Jack Youden in person.

A BRIEF BIOGRAPHICAL SKETCH (1900-1971) OF JACK YOU DEN'S LIFE

As with so many others who have contributed much to our profession, Jack Youden began his career, not as a statistician, but rather in a related discipline, as a physical chemist. Born in Townsville, Australia, in 1900, Jack's family came to America in 1907 and resided in Niagara Falls, NY, where Jack attended the local public schools. During Jack's senior year of high school, the family moved to Rochester, NY, and the following year Jack enrolled at the University of Rochester. In 1921, Jack graduated with a B.S. in Chemical Engineering having been elected to Phi Beta Kappa.

Following graduation Jack served a year as an instructor in chemistry at the University of Rochester and then enrolled at Columbia University as a graduate fellow in chemistry earning an M.A. in 1923 and a Ph.D. in 1924. Upon graduation, Jack joined the staff of the Boyce Thompson Institute for Plant Research, located in Yonkers, NY, as a physical chemist. *Thus, during the first third of his life, Jack showed an interest in and was formally trained in the field of chemistry.*

The next seven years were a transition period in which Jack, a physical chemist, was slowly becoming more of a statistician. In 1931-32, Jack commuted from Yonkers to Morning Heights, NY to attend Professor H. Hotelling's lectures on *Statistical Inference* at Columbia University. And, while Jack was now assuming the role of a statistician more in his work, his laboratory experience was always to remain a treasured asset enabling him to communicate with scientists on their own grounds.



Joe Voelkel presenting the Youden Address award to John Cornell

The paper that launched Jack to new heights appeared in 1937 and was titled *Use of Incomplete Block Replications in Estimating Tobacco Mosaic Virus*, Youden (1937, 1972). In this paper Jack introduced a new class of symmetrically Balanced Incomplete Block Designs that possess the characteristic double control (rows and columns) of a Latin Square design without requiring the number of replications of each treatment equal the number of treatments. Shown in Figure 1 is one such design in which seven treatments (A, B, C, D, E, F, and G) are assigned, three to each block (plant), to the top, middle, and bottom leaves of seven plants. The incomplete block property of the design follows from each plant (block) accommodating only three of the seven treatments while all treatments are present with the top leaves, middle leaves, and bottom leaves of the plants. Such a design was later termed a "Youden Square" by Fisher and Yates (1957). It is interesting to note that a Youden Square can be formed by removing one or more rows of a Latin Square.

| | | Blocks (plants) | | | | | | |
|---------------|---|-----------------|---|---|---|---|---|---|
| leaf position | T | A | B | C | D | E | F | G |
| | M | B | D | F | E | G | A | C |
| | B | C | F | E | A | B | G | D |

FIGURE 1 – A Youden Square containing seven treatments (A, B, ..., G) set up in seven blocks of size three. The leaf position T, M, and B refers to the top, middle, and bottom leaves of each plant.

During the 1937-38 academic year, Jack took a leave of absence from the Boyce Thompson Institute to work at the Galton Laboratory, University College, London, under the guidance of R.A. Fisher. One day, Jack and Sir Ronald were walking in a garden located behind the laboratory and Jack was smoking a pipe. A piece of ember was blown from Jack's pipe and landed on Fisher's neck whereupon Fisher exclaimed some words that are

unprintable. The next day, Fisher presented Jack with a pipe with a wind cover on top. When Jack quit smoking some years later, he gave the pipe to Dr. Harry Ku, a colleague at the National Bureau of Standards, who later donated the pipe to the Youden memorabilia which are presently housed at the National Bureau of Standards.

A very interesting account of Jack's involvement during World War II as operations analyst and Head of the Bombing Accuracy Section of the U.S. Eighth Air Force in Britain is covered in Miser (1992). Jack adapted statistical tools of experimental design and graphical analysis in two Air Forces reports *How to Improve Formation Bombing* and *Bombing Charts*, Youden (1945a and 1945b). As a result of his efforts during the war, Jack was awarded the *Medal of Freedom* in 1946.

In May of 1948, Jack joined the National Bureau of Standards (NBS) as assistant chief of the Statistical Engineering Laboratory, Applied Mathematics Division. Three years later he became a consultant on the statistical design and analysis of experiments to the chief of the division and remained in that capacity until his retirement from NBS in 1965. *During the middle third of his life, Jack made the transition from chemist to statistician where from this point on he would devote his energies to improving the methods of data collection and analysis.*

SOME EXAMPLES OF JACK YODEN'S METHODOLOGY

The development of the field of design of experiments was inspired largely by the needs of agriculture and biology. Jack believed that one of the reasons for the delay in the adoption of experimental designs in the physical sciences is that the classical designs (for agriculture and biol-

ogy) often do not meet the experimental situations encountered in the physical sciences. In fact, when one compares the experimental conditions which exist in the biological and agricultural sciences with those of the physical and chemical sciences, one is at once struck by certain fundamental differences. A difference which is of paramount importance, noted Jack, is the *magnitude of the errors of measurement*.

In agricultural and biological experiments, the experimental material often is land or animals and the variation over a field or between litters is likely to be large. Thus replicates are needed to effect a reduction in the error of the measurements.

In the physical and chemical sciences, measurements can often be made with high precision and the experimental material usually is relatively homogeneous. Jack did not believe it was necessary to put great reliance on replication in order to achieve good results. Excellent estimates may be obtained quite often from one measurement, or at most, two or three, Youden (1953, 1972).

After joining NBS in 1948, Jack developed some new families of experimental designs called *linked block designs* and *chain block designs*. These designs were proposed for experiments in which the number of treatments greatly exceeds the block size while the number of replications of each treatment is small. Shown in Figure 2 is a linked block design in which ten treatments (A, B, C, ..., J) are assigned to five blocks, each of size four. Note that in this design every pair of blocks have exactly one treatment in common (the blocks are linked). These designs were invented to meet the experimenter's needs; they were not defined according to combinatorial principles.

| | | | | |
|---|---|---|---|---|
| A | A | B | C | D |
| B | E | E | F | G |
| C | F | H | H | I |
| D | G | I | J | J |

FIGURE 2 – A linked block design for ten treatments in five blocks of size four each. Each pair of blocks is linked by a single treatment.

In Figure 3 are shown three chain block designs for accommodating 22 treatments in only four blocks varying in sizes from 7 to 10. Some of the treatments (in capital letters) are replicated twice while the others (in lower case letters) are replicated only once. Chain block designs are discussed in Youden and Connor (1953).

| Block | | | | Block | | | | Block | | | |
|-------|---|---|---|-------|---|---|---|-------|---|---|---|
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| A | C | E | G | A | D | G | J | A | E | I | M |
| B | D | F | H | B | E | H | K | B | F | J | N |
| C | E | G | A | C | F | I | L | C | G | K | O |
| D | F | H | B | D | G | J | A | D | H | L | P |
| i | m | q | t | E | H | K | B | E | I | M | A |
| j | n | r | u | F | I | L | C | F | J | N | B |
| k | o | s | v | m | P | s | u | G | K | O | C |
| l | p | | | n | q | t | v | H | L | P | D |
| | | | | o | r | | | q | s | u | |
| | | | | | | | | r | t | | |

FIGURE 3 – Three chain block designs for 22 treatments in four blocks of various sizes. Treatments in capital letters are replicated twice while the others are replicated only once.

"It is inevitable that the Latin Square arrangement will be tried when the rows and columns are used for factors that not only are likely to interact with treatments but also with each other," Youden and Hunter (1955). Shown in Figure 4 is a partially replicated 3 x 3 Latin Square where the treatments on the diagonal are replicated twice. By replicating the diagonal treatments one is able to obtain an estimate of the pure error variance from the replicates thus permitting a test to be made on the possible presence of interaction.

| | | cols | | |
|------|---|------|-----|-----|
| | | 1 | 2 | 3 |
| rows | 1 | AA' | B | C |
| | 2 | B | CC' | A |
| | 3 | C | A | BB' |

FIGURE 4 – A partially replicated 3 x 3 Latin Square design.

"In any kind of measurement setting, only when we realize what the contributors are to the overall error will the magnitude of the overall error have meaning." The following is a scenario that I put together that illustrates the ideas in Youden (1965, 1972).

In a laboratory setting, suppose we have two instruments I and II for measuring soil samples, and two technicians that use the instruments to run soil tests.

- Questions:
- Are the instruments similar, i.e., do they give the same test results when the same soil sample is tested on both by the same technician?
 - Do the technicians get the same test results when they read the same soil sample on the same instrument?
 - How many trials must we run in order to answer i) and ii)?

Figure 5a illustrates the simplest strategy (n=3 trials denoted by ☉) while Figure 5b displays a better strategy (n=4 trials). In 5a and 5b, I_{1A} symbolizes the measurement obtained on instrument 1 by technician A.

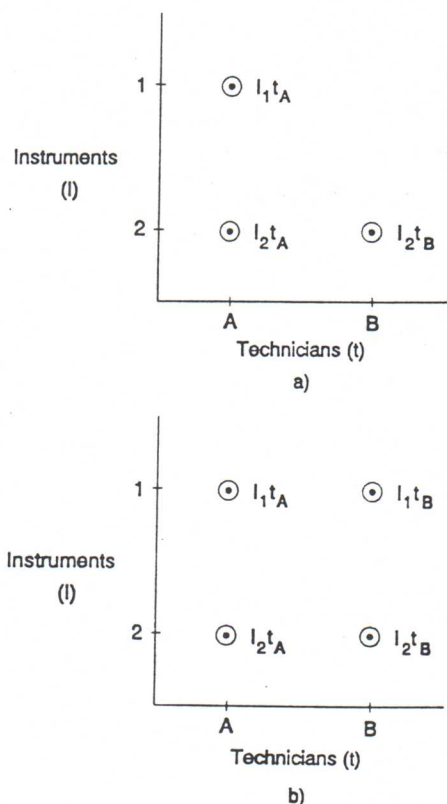


FIGURE 5 – Two instruments and two technicians. In a), technician A reads both instruments while B reads only instrument 2. In b) both technicians read both instruments.

In Figure 5a, question i) is answered by comparing the two test results made by technician A, $I_1 - I_2 = I_1 t_A - I_2 t_A$, while question ii) is answered by comparing the results of the two tests by the technicians made on instrument 2, $t_A - t_B = I_2 t_A - I_2 t_B$. In Figure 5b, question i) is answered by taking the difference between the two averages

$$I_1 - I_2 = \frac{I_1 t_A + I_1 t_B}{2} - \frac{I_2 t_A + I_2 t_B}{2}$$

while question ii) is answered by taking the difference

$$t_A - t_B = \frac{I_1 t_A + I_2 t_A}{2} - \frac{I_1 t_B + I_2 t_B}{2} \quad (2)$$

The difference in (1) gives equal weight to both technicians while the difference in (2) gives equal weight to both instruments. Further, the differences (1) and (2) with Figure 5b contain twice as much information as the differences used with Figure 5a.

Suppose now that two soil samples, S_1 and S_2 , are available to answer the question "Are the samples different?" Again, using only $n=4$ trials, we might try the strategy shown in Figure 6a where one starts with $I_1 t_B S_1$ and changes one-factor-at-a-time to produce

$$\begin{aligned} 1. & I_1 t_B S_1 \\ 2. & I_1 t_A S_1 \\ 3. & I_1 t_B S_2 \\ 4. & I_2 t_B S_1 \end{aligned} \quad (3)$$

An average that gives equal weight to both instruments, both technicians, and both soil samples is

$$\text{Average} = \frac{I_1 t_A S_1 + I_1 t_B S_2 + I_2 t_B S_1 - I_1 t_B S_1}{2}$$

Figure 6b illustrates a better strategy than Figure 6a where in Figure 6b two factors are changed at a time. Why is the strategy in Figure 6b better? Simply because both instruments are equally replicated, both soil samples are equally tested, and both technicians are equally represented. Furthermore,

$$\begin{aligned} \text{Average} &= \frac{1. + 2. + 3. + 4.}{4} \\ &= \frac{1}{2}(I_1 + I_2) + \frac{1}{2}(t_A + t_B) + \frac{1}{2}(S_1 + S_2) \end{aligned} \quad (4)$$

and

$$\begin{aligned} I_1 - I_2 &= \frac{1. + 2. - 3. - 4.}{2} = \frac{I_1(t_B S_1 + t_A S_2)}{2} - \frac{I_2(t_A S_1 + t_B S_2)}{2} \\ t_A - t_B &= \frac{2. + 3. - 1. - 4.}{2}, \quad S_1 - S_2 = \frac{1. + 3. - 2. - 4.}{2} \end{aligned} \quad (1)$$

where the numbers 1., 2., 3., and 4. signify the numbered vertices in Figure 6b. These differences are twice as precise as those in (3).

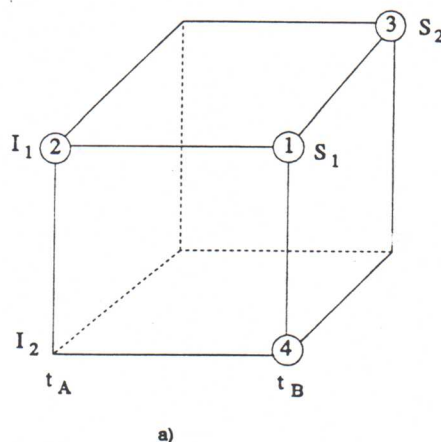


FIGURE 6 – (See caption on next page)

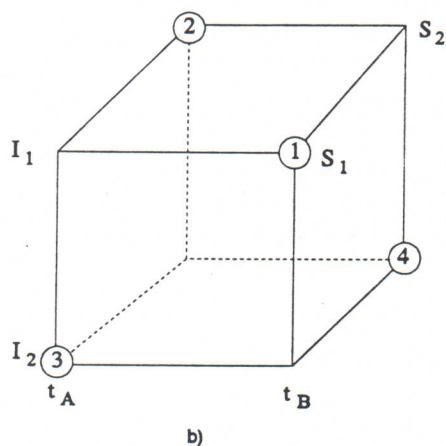


FIGURE 6 – Two instruments, two technicians, and two soil samples. Starting with $I_1 t_B S_1$, in a) one changes one factor-at-a-time; in b) one changes two factors-at-a-time.

What about testing the four factors instruments, technicians, soil samples, and days 1 and 2 with only $n=5$ trials? Figure 7a illustrates the strategy where trials 1 and 2 are performed on day 1 and trials 3, 4, and 5 are performed on day 2. An average that gives equal weight to both instruments, technicians, soil samples, and days, is

$$\text{Average} = \frac{2(1.) + 2. + 3. + 4. + 5.}{6}$$

and

$$I_1 - I_2 = \frac{1. + 2(4.) - 2. - 3. - 5.}{3}, \quad t_A - t_B = \frac{2. + 3. + 4. - 1. - 2(5.)}{3}$$

$$S_1 - S_2 = \frac{1. + 2(3.) - 2. - 4. - 5.}{3}, \quad d_1 - d_2 = \frac{1. + 2(2.) - 3. - 4. - 5.}{3}$$

Personally I prefer the strategy in Figure 7b which is a 2^{4+1} fractional replicate where each instrument is used by both technicians the same number of times, each soil is tested by each technician on each instrument the same number of times, and on each day each instrument is used by each technician, each soil is tested on each instrument, and each soil is tested by each technician. And, with three additional trials ($n=8$) in Figure 7b, each difference $I_1 - I_2, \dots, d_1 - d_2$, is 1.78 times as precise as those in (5). Further, there exists sufficient information for measuring three additional factors such as soil types, time of day, and laboratory 1 versus laboratory 2, etcetera.

"Probably the oldest way to ascertain the error in a measurement is to repeat it," Youden (1962, 1972). Suppose we have a known standard weight S and we wish to measure the weights of two other materials A and B using the standard. Two ways to calibrate or compare A and B against the standard S are,

1. Weigh S and A and then repeat the test, by reversing their positions on the scale. Compute the difference between the two test results for an estimate of the error. Now, weigh S and B and repeat the test. Compute the difference between the two test results.

"Generally it is much better to devise some indirect way of measuring the error of a comparison. Preferably the indirect way should make it impossible or quite diffi-

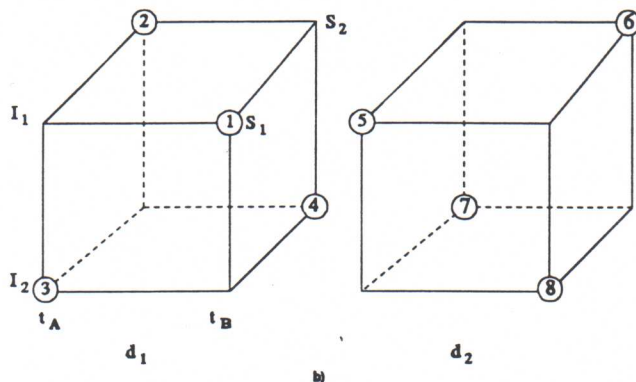
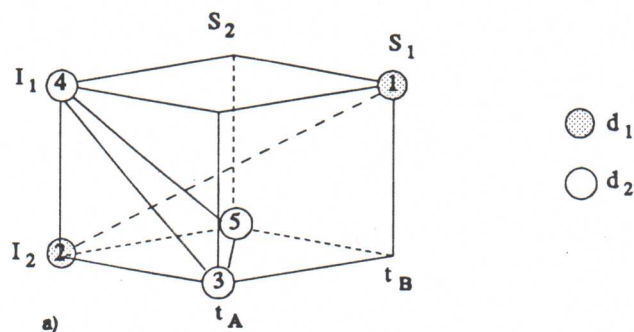


FIGURE 7 – Strategies for studying four factors each at two levels. In a) $n = 5$, while in b) $n = 2^3 = 8$.

cult for the operator (tester) to have any idea what his/her error is as he/she makes their reading."

2. Weigh S and A , S and B , and A and B . Then an estimate of the differences $S - A$ and $S - B$ are

$$\text{Estimate}(S - A) = \frac{2(S - A) + [(S - B) + (B - A)]}{3}$$

$$\text{Estimate}(S - B) = \frac{2(S - B) + [(S - A) - (B - A)]}{3}$$

The errors of measure are the differences $(S - A) - \{(S - B) + (B - A)\}$, or, $(S - B) - \{(S - A) - (B - A)\}$

Test procedures are used to ascertain whether a product meets the specification set down for the product. When performing test procedures, a double problem confronts the producer as well as the tester: (i) there is bound to be a certain amount of variation in the product, and (ii) there is bound to be a certain amount of variation in the test results, stemming from the procedure itself, made on a given sample of the product. What we would like to know is, "Is there a simple method of determining whether the test procedure as set forth is capable of yielding acceptable agreement among results from different parties (such as different technicians, different laboratories, etcetera)? Furthermore, if the results are not acceptable, we would like some specific indication of what is

wrong with the procedure. On the other hand, if the procedure appears to be reasonably good but there are some disturbing discrepancies, we would like to know which parties are having trouble and if possible, why they are having trouble. And most importantly, we want to be able to get this information back to the parties concerned in such a form that the diagnosis is believed. FOR ONLY THEN WILL THE PARTIES TAKE ANY ACTION TO CORRECT THE DIFFICULTIES.

The following is a synopsis of a typical day at the office that took place during the spring of 1992

My initial reaction upon looking at my schedule of this Monday's activities was, "Oh my, a typical Monday. I guess I might as well get started." The appointment book read as listed in Figure 8.

| 9 | | Monday March 1992 |
|-------|--|-------------------------|
| 7:00 | | |
| 7:30 | | |
| 8:00 | | |
| 8:30 | Ed Hanlon - Soil Science | |
| 9:00 | | |
| 9:30 | | |
| 10:00 | Joe Jones - Food Science & HN | |
| 10:30 | | |
| 11:00 | Harry Pierce - Entomology | |
| 11:30 | | |
| 12:00 | | |
| 1:00 | C. Courtney - Vet. Medicine | |
| 1:30 | | |
| 2:00 | | |
| 2:30 | S. Rezaian - Agronomy (RSM) | |
| 3:00 | | |
| 3:30 | J. Williams - Root dist. of Ponds Trees | |
| 4:00 | | |
| 4:30 | Mixtures paper - hopefully | |
| 5:00 | | |

FIGURE 8 - Appointment schedule of the day's activities.

Such is the schedule of an experiment station statistician working in a large research-oriented university setting. A typical day involves crossing many different disciplines where each client has the expectation that you will be able to help him or her solve their particular problem. Today's activities promised to be no different from the norm.

Owing to the diverse nature of the clientele, a successful statistician's toolbox must contain a plethora of techniques. On this particular day, I was especially grateful to Jack Youden. I envisioned having to apply some of the methods and procedures he developed; specifically, to evaluate the performances of several laboratories, and, in using his index for rating and comparing two diagnostic tests. The morning's activities began with Dr. Hanlon, a Professor of Soil Science, showing me a list of soil test results performed by the students in his soils class. Each student had performed the Mehlich III test procedure on two Florida soils (Okeelanta and Torry) to measure the amount of Phosphorous (ppm P converted to Kg P ha^{-1}) in a sample of each soil. The data are listed in Table 1. "The purpose behind having the students perform the actual soil tests themselves," Dr. Hanlon said, "was to see if they clearly understood the Mehlich III soil test procedure." I tended to think more along the lines stated by Jack in his booklet entitled *EXPERIMENTATION AND MEASUREMENT*, "The best way to find out about some of the difficulties in making measurements is to make measurements" Youden (1962), page 12. (See Table 1. BELOW)

Dr. Hanlon expressed an interest in finding out if the variability in the students' soil test results was typical. By typical is meant, "Can we determine that the Mehlich III procedure is capable of yielding acceptable agreement among the test results obtained by different people?" I recalled having seen a similar question involving the agreement among test results from different laboratories in a paper by Youden (1959, 1972). Youden had suggested a very simple graphical procedure for plotting the results from different laboratories and I wondered if perhaps the graphical procedure might help to answer Dr. Hanlon's question about whether or not the students understood the procedure.

Table 1. Phosphorous (kg P ha^{-1}) Test Results on Okeelanta and Torry Soils.

| Student | Okeelanta | Torry | Student | Okeelanta | Torry | Student | Okeelanta | Torry |
|---------|-----------|-------|---------|-----------|-------|---------|-----------|-------|
| 1 | 22 | 30 | 9 | 65 | 79 | 17 | 29 | 35 |
| 2 | 31 | 34 | 10 | 39 | 49 | 18 | 40 | 44 |
| 3 | 28 | 35 | 11 | 57 | 61 | 19 | 47 | 39 |
| 4 | 41 | 50 | 12 | 28 | 42 | 20 | 60 | 52 |
| 5 | 62 | 64 | 13 | 30 | 40 | 21 | 53 | 65 |
| 6 | 39 | 51 | 14 | 61 | 69 | 22 | 38 | 46 |
| 7 | 52 | 49 | 15 | 47 | 53 | | | |
| 8 | 44 | 62 | 16 | 65 | 66 | | | |

The graph or plot is constructed by drawing the customary x-axis horizontally at the bottom of the page and the y-axis vertically along the lefthand side of the page. Along the y-axis are recorded the Phosphorous values for the Okeelanta soil and along the x-axis are recorded the Phosphorous values for the Torry soil. Along each axis, the scale units are the same and cover the range of values obtained for each of the soils. The pair of results reported by a student are then used to plot a point. There will be as many points (22 in our case) as there are students. The soil test results listed in Table 1 are plotted in Figure 9.

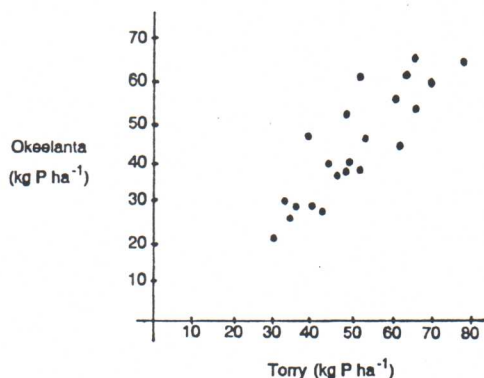


FIGURE 9 – Plot of Okeelanta and Torry soil test results by 22 students.

After the points are plotted, a horizontal median line is drawn parallel to the x-axis so that there are as many points above the line as there are below it. A second median line is then drawn parallel to the y-axis so that there are as many points lying to the left of this line as there are lying to the right of this line. The two median lines divide the plot into four quadrants, I, II, III, and IV, as shown in Figure 10.

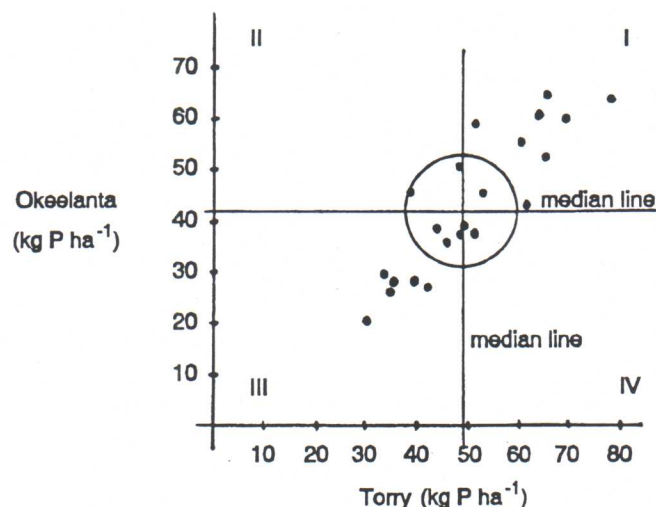


FIGURE 10 – Median lines dividing the plot of soil test results into four quadrants.

In the ideal situation where only random errors of precision are present, the points should be equally numerous in all four quadrants. In the case of the 22 students, nine points are positioned in quadrants I and III, while only 2 points are positioned in each of the quadrants II and IV. This suggests that some of the students (those represented by the rightmost points in I and by the leftmost points in III) tended to record high P values on both soils or low P values on both soils, which is evidence of individual student biases; not an unexpected result. [Youden (1959, 1972) reported similar findings of individual laboratory biases in an interlaboratory study.] Dr. Hanlon commented that the students were asked to perform the Mehlich III test on both soils at a single sitting and that they were not informed that the soils were different. "It is possible that some of the students probably thought the two samples were from the same soil," he said after being shown the plot.

Of the two Florida soils, Torry is thought to possess approximately 20% more P than Okeelanta. Thus the vertical median line is slightly farther to the right on the Torry-axis than the horizontal median line is high on the Okeelanta-axis. On the other hand, assuming the two soils to be similar and nearly equal in magnitude of P, the scatter of points, or variation in the data, within each soil should be approximately the same for the two soils. If indeed this is the case, then drawing a line at 45 degrees through the intersection of the median lines and measuring the perpendicular distance from each point to the 45 degree line makes it possible to obtain an estimate of the precision of the procedure from the data. The precision (or standard deviation of a single result) is obtained by multiplying the average length of the perpendiculars by $\sqrt{\pi}/2$ or 1.2533. Fortunately, a simpler and quicker method for estimating the standard deviation involves taking the difference $(T_i - O_i) = d_i$ for each student and calculating the average \bar{d} . The average \bar{d} is then subtracted from each d_i to obtain a set of corrected differences $d'_i = d_i - \bar{d}$. The average of the absolute values $|d'|$ when multiplied by: $\sqrt{\pi}/2$ or 0.886, gives an estimate of the standard deviation. For the data listed in Table 1, the estimate is: $\hat{\sigma} = 4.41$.

A test on the randomness of the students' test results can be made by constructing a circle centered at the intersection of the median lines and counting the percentage of points falling inside the circle. In other words, if the variation in the students' results is strictly random error (without bias or some other type of systematic error) with the magnitude of error being the same for both soils, the plot should resemble a random sample taken from a Circular Normal distribution (symmetrical bivariate Normal). The multiples of the standard deviation that include various percentages of the points, determined by the formula

$$\text{Percent} = 100[1 - \exp(-b^2/2)]$$

are given in Table 2.

Table 2. Probability Table for Circular Normal Distribution.

| Percent of the Points Within Circle | Multiple b of the Standard Deviation |
|--|---|
| 50 | 1.177 |
| 60 | 1.350 |
| 70 | 1.552 |
| 75 | 1.665 |
| 80 | 1.794 |
| 90 | 2.146 |
| 95 | 2.448 |
| 99 | 3.035 |

A circle whose radius is about 2.5 to 3.0 times the standard deviation gives a fair idea of the smallest circle that could be expected to contain nearly all of the points in the absence of bias. With the soil test data from the 22 students, a circle of radius $r = 2.448(4.41) = 10.8$ is drawn in Figure 10. If random error only is present, the circle should contain approximately 95 percent of the points. Over half (14 out of 22) of the points are outside the circle which clearly suggests those students whose points lie outside the circle have bias incorporated in their performance of the testing procedure. Although Youden professed that generally a fair number of points will lie outside such a circle, when shown the results of Figure 10, Dr. Hanlon expressed concern over whether or not he had described the Mehlich III test procedure clearly enough. He later repeated the lecture on conducting the Mehlich III test procedure and a follow-up exercise by the same students using the same two soils produced an estimate of the standard deviation of $\hat{\sigma} = 3.38$ with only five of the 22 students falling outside the 95 percent circle on the second attempt.

In a later paper, Mandel and Lashof (1974) examine the assumptions underlying the "Youden Diagram" and present a geometrical argument for interpreting and separating the variation among laboratories (students in our case) and within laboratories. They generalize the plot to cover situations where the two samples (soils O and T) are not equal in magnitude of the property measured and/or where the major axis of the elongated ellipse of plotted points does not bisect median lines. When their method was applied to the students' test results listed in Table 1, the 95 percent confidence ellipsoid assumed the form shown in Figure 11. While the scatter in the data values when taken along the minor axis of the ellipse represents a measure of random error (σ_{ϵ}^2), the scatter in the data values when taken along the major axis of the ellipse is a measure of the systematic student differences plus random error ($2\sigma_{\text{students}}^2 + \sigma_{\epsilon}^2$). Estimates of the two variance components for random error and among the students are $\hat{\sigma}_{\epsilon}^2 = 22.12$ and $\hat{\sigma}_{\text{students}}^2 = 153.34$, respectively. A test for "sphericity" using the ratio of the lengths of the major to minor axes clearly rejected the hypothesis that $\sigma_{\text{students}}^2 = 0$.

This supported our earlier findings in Figure 10 that suggested the students' test results were not simply random error.

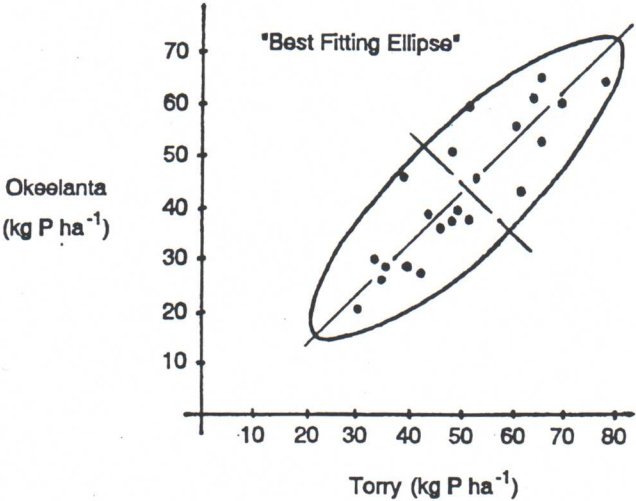


FIGURE 11 – The 95 percent confidence ellipsoid of the soil test results.

"The ideal diagnostic test should discriminate unerringly between diseased and healthy individuals," Youden (1950). Suppose we have a grouping of individuals some of which are known to have a disease while the others are known to be healthy (not have the disease). Further, suppose all of the individuals are classified by a test as either testing positive (having the disease) or testing negative (not having the disease). Such results are tabulated as follows:

| Individuals | Classified by Test | | Total |
|----------------|--------------------|----------|-------|
| | Positive | Negative | |
| Known Diseased | a | b | a+b |
| Healthy | c | d | c+d |

From the table above, the proportion of diseased individuals correctly classified is $a/(a+b)$, while the proportion of healthy individuals correctly classified is $d/(c+d)$. An intuitive measure of the success or goodness of a diagnostic test would seem to be the difference between the proportion of individuals correctly classified minus the proportion of individuals incorrectly classified averaged over the known diseased and healthy groups." Assuming false positives (c of the healthy individuals) to be as undesirable as false negatives (b of the diseased individuals), Jack's index for rating a diagnostic test is

$$J = \frac{1}{2} \left[\frac{a-b}{a+b} + \frac{d-c}{c+d} \right] \quad (6)$$
$$= \frac{ad-bc}{(a+b)(c+d)}$$

One might also interpret the index, by looking at the table entries, as the product of the numbers of correctly classified individuals minus the product of the numbers of incorrectly classified individuals divided by the product of the total numbers of diseased and healthy individuals. The index in (6) has certain desirable features which are described in Youden (1950) and reiterated later by Nelson (1974):

In my duties as a statistical consultant to the faculty in the College of Veterinary Medicine, I have used Jack's index on numerous occasions to evaluate diagnostic testing procedures. But just as important to the veterinarian is the determination of what constitutes a truly infected animal and a truly healthy or uninfected animal. Measures such as Predictive value of a positive test = Prob (infected animal given a positive test result)

$$\text{or} \quad = \frac{a}{a+c}$$

Predictive value of a negative test = Prob (healthy animal given a negative test result)

$$\text{or} \quad = \frac{d}{b+d}$$

$$\text{Accuracy of a test} = \frac{a+d}{(a+b) + (c+d)} = \text{Probability of correctly identifying the infection status of an animal}$$

are all related to the classification scheme leading to Jack's index. These measures are discussed in Courtney and Cornell (1990) as is the use of McNemars' paired χ^2 test for comparing two test procedures when the infection status of an animal is unknown.

THE IMPACT OF JACK'S WORK ON THE SCIENTIFIC COMMUNITY

The impact of Jack's work on the scientific community is most clearly defined in two main areas: experimental design and statistical analysis. The experimental designs that are most highly acclaimed are the "Youden Squares (BIBD's)", the linked block and chain block designs, the partially replicated Latin Squares, calibration designs, and constrained randomization sequences. The methods of analyses most frequently cited are the index (J) for rating diagnostic tests, the two-sample chart or "Youden Diagram", and the extreme rank sum test for outliers. Over a forty-one year period (1924-1964) this work as well as other works appeared in over forty different journals, five small books, and in more than ten conference proceedings and transactions. In addition, during a six-year period (1954-1959), Jack wrote a bimonthly column (36 articles) for *Industrial and Engineering Chemistry*. During the period from 1948 to 1965, Jack delivered 211 talks

under 125 different titles around the country on topics in statistical methodology and experimental design.

So, how does one measure the impact on others left by such an individual? I searched the Citation Index to find that during the six-year period from 1986 to 1991, citations to Jack's work exceeded 64 per year on the average which is rather remarkable when one realizes that many of the citations are to papers published over forty years ago. Many of Jack's methods are as relevant today as they were when they were developed.

REMEMBERING JACK YODEN

Since I had never met Jack Youden in person, I felt I needed to solicit the help of others who had in saying a few words to remember Jack by. The following are excerpts of statements about Jack from others.

"I met and talked with Jack Youden at several Gordon Research Conferences, and was always impressed with his ingenuity and appeal to practitioners. It was down-to-earth people like Jack Youden that got me interested in statistics rather late in life."

John W. Gorman

"When I worked with Jack Youden on partially replicated Latin Squares, he didn't care to get involved with the mathematics of the problem. He just looked at the block-treatment pattern and would say "it's right or it's not right." "He had an intuition about balance and symmetry of designs; he was amazing." J. Stu Hunter

"When compiling a subject index for the NBS Special Publication 300, Vol. 1, a volume that contains 15 of Youden's publications, I had great difficulty in finding terms in Youden's writing to include in the index, but no trouble at all for the other authors. Youden took great pains to avoid words that needed to be technically defined."

"Youden's handwriting was extremely neat and it seems that his first draft was his final draft (after thinking about what to write for a long period of time). Enclosed is an example, ..."

Harry Ku

The example to which Dr. Ku refers is a handwritten manuscript by Jack two years after retiring from NBS entitled "The Role of Statistics in Regulatory Work". At the end of the introductory paragraph where Jack sets the stage for discussing the role of statistics in regulatory work is the following excerpt:

"There is one particular role I am determined that statistics should not have. AOAC must not serve as a playground for statisticians to exhibit their special skills at the price of bewildering the chemist. There is an important reason for insisting on simple and intuitively acceptable statistical techniques. Presentation of evidence before a court, or to a producer whose product is rejected, will be more convincing if it is understandable."

J. Youden (1967)

I believe this not only typifies the way Jack felt about many of the contributions he made throughout his illustrious career but more importantly, this excerpt exemplifies the type of person Jack Youden was.

In closing, I wish to express my sincere gratitude to the Statistics Division for giving me the opportunity to share with you today, some of the career highlights of William John (Jack) Youden; chemist, statistician, researcher, and teacher, in whose name this address is given each year. I wish also to publicly thank Dr. Harry Ku, a colleague of Jack Youden, who graciously provided me with several interesting stories of events of Jack's life and much of the material about Jack Youden that helped me in putting this talk together.

REFERENCES

- Courtney, C. H. and J. A. Cornell (1990). Evaluation of Heartworm Immunodiagnostic Tests. *Journal of American Veterinary Medical Association*, Vol. 197, No. 6, pp. 724-729.
- Fisher, R. A. and F. Yates (1957). *STATISTICAL TABLES FOR BIOLOGICAL AGRICULTURAL AND MEDICAL RESEARCH*. 5th Ed., Oliver and Boyd Ltd., London England.
- Mandel, J. and T. W. Lashof (1974). Interpretation and Generalization of Youden's Two-Sample Diagram. *Journal of Quality Technology*, Vol. 6, No. 1, pp. 22-36.
- Miser, H. J. (1992). Craft in Operations Research. *Operations Research*, Vol. 40, No. 4, pp. 633-639.
- Nelson, L. S. (1974). Rating and Discrimination of Pass-Fail Tests. *Journal of Quality Technology*, Vol. 6, No. 1, pp. 58-59.
- Youden, W. J. (1937). Use of Incomplete Block Replications in Estimating Tobacco-Mosaic Virus. *Contributions from Boyce Thompson Institute*, Vol. 9, No. 1, pp. 41-48. (1972). *Journal of Quality Technology*, Vol. 4, No. 1, pp. 50-57.
- Youden, W. J. (1950). Index for Rating Diagnostic Tests. *Cancer*, Vol. 3, No. 1, pp. 32-35.
- Youden, W. J. (1953). Sets of Three Measurements. *The Scientific Monthly*, Vol. LXXVII, No. 3, pp. 143-147. (1972). *Journal of Quality Technology*, Vol. 4, No. 1, pp. 40-44.
- Youden, W. J. and W. S. Connor (1953). The Chain Block Design. *Biometrics*, Vol. 9, No. 2, pp. 127-140.
- Youden, W. J. and J. S. Hunter (1955). Partially Replicated Latin Squares. *Biometrics*, Vol. 11, No. 4, pp. 399-405.
- Youden, W. J. (1959). Graphical Diagnosis of Interlaboratory Test Results. *Industrial Quality Control*, Vol. XV, No. 11, pp. 24-28. (1972). *Journal of Quality Technology*, Vol. 4, No. 1, pp. 29-33.
- Youden, W. J. (1962). Uncertainties in Calibration. *IRE Transactions on Instrumentation*, Vol. I-II, Nos. 3 and 4, pp. 133-138. (1972). *Journal of Quality Technology*, Vol. 4, No. 1, pp. 34-39.
- Youden, W. J. (1962). *EXPERIMENTATION AND MEASUREMENT*. Vistas of Science Series, National Science Teachers Association, Washington, DC, 128 pages.
- Youden, W. J. (1965). The Evolution of Designed Experiments. *Proceedings of the IBM Scientific Computing Symposium on Statistics*, IBM Data Processing Division, White Plains, NY, pp. 59-67. (1972). *Journal of Quality Technology*, Vol. 4, No. 1, pp. 11-19.
- Youden, W. J. (1967). The Role of Statistics in Regulatory Work. *Journal of the Association of Official Analytical Chemists*, Vol. 50, No. 5, pp. 1007-1013.

Highlights of J. Youden's career are summarized in

- Eisenhart, Churchill 1976. William John Youden. Vol. 14, pp. 552-557. *Dictionary of Scientific Biography*, Ed. by Charles C. Gillespie, New York: Schribner's.
- Eisenhart, Churchill and Rosenblatt, Joan R. 1972. W. J. Youden, 1900-1971. *Annals of Mathematical Statistics*, Vol. 43, No. 4, pp. 1035-1040.
- Hamburg, Morris (editor) 1971. W. J. Youden, Renowned Statistician, 1907-1971. *The American Statistician*, Vol. 25, No. 3, p. 51.
- Ku, Harry H. 1978. W. J. Youden. *International Encyclopedia of Statistics*, Edited by William H. Kuskal and Judith M. Taneer, The Free Press, New York, pp. 1257-1261.
- Lasater, H.A. (editor) 1972. William John Youden, 1900-1971. *Journal of Quality Technology*, Vol. 4, No. 1, pp. 1-67. This memorial issue includes reprints of nine of Youden's papers and one of his bimonthly columns from *Industrial and Engineering Chemistry* as well as an extensive bibliography. Δ