ARDC TR 11

AD732428

ARDC

AD

TECHNICAL REPORT NO. 11

THE STATISTICAL COMPARISON OF MEASURING INSTRUMENTS

(A Study of Precision and Accuracy Problems for NATO Velocity Measuring Chronographs)

by

F. E. Grubbs
J. F. O'Bryon

July 1971

Approved for public release; distribution unlimited.

U.S. ARMY MATERIEL COMMAND
ABERDEEN RESEARCH AND DEVELOPMENT CENTER
ABERDEEN PROVING GROUND, MARYLAND

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
USA Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
The Statistical Comparison of Measuring Instruments
(A Study of Precision and Accuracy Problems for NATO Velocity Measuring Chronographs)

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
Frank E. Grubbs
James F. O'Bryon

6. REPORT DATE
July 1971

7a. TOTAL NO. OF PAGES
37

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.
RDT&E Project No. 1T061102B14A

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Report No. 11

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
U. S. Army Materiel Command
Washington, D. C.

13. ABSTRACT

The problem discussed and solved here is that of determining whether the precision and accuracy of a velocity measuring chronograph is suitable for general use. The study arose in connection with the need to compare a "test" chronograph of one of the NATO countries with that of two NATO "standard" chronographs, in order to ascertain whether the "National" instrument submitted for acceptance may be used as an approved chronograph for measuring velocities generally. Since the methodology developed herein covers both the estimation problem and the various statistical tests of significance required on variances in errors of measurement (or imprecision), as well as tests on accuracy, then the results are believed to be of rather wide applicability.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Errors of Measurement						
Instrument Errors						
Statistical Comparison of Instruments						

Unclassified

Security Classificati

A B E R D E E N   R E S E A R C H   A N D   D E V E L O P M E N T   C E N T E R

TECHNICAL REPORT NO. 11

July 1971

THE STATISTICAL COMPARISON OF MEASURING INSTRUMENTS (A Study of Precision and Accuracy Problems for NATO Velocity Measuring Chronographs)

F. E. Grubbs
J. F. O'Bryon

Office of the Chief Operations Research Analyst

and

Ballistic Research Laboratories

Approved for public release; distribution unlimited.

RDT&E Project No. 1T061102B14A

A B E R D E E N   P R O V I N G   G R O U N D,   M A R Y L A N D

A B E R D E E N   R E S E A R C H   A N D   D E V E L O P M E N T   C E N T E R
TECHNICAL REPORT NO. 11

FEGrubbs/JFO'Bryon/tbs
Aberdeen Proving Ground, Md.
July 1971

THE STATISTICAL COMPARISON OF MEASURING INSTRUMENTS (A Study of Precision and Accuracy Problems for NATO Velocity Measuring Chronographs)

ABSTRACT

The problem discussed and solved here is that of determining whether the precision and accuracy of a velocity measuring chronograph is suitable for general use. The study arose in connection with the need to compare a "test" chronograph of one of the NATO countries with that of two NATO "standard" chronographs, in order to ascertain whether the "National" instrument submitted for acceptance may be used as an approved chronograph for measuring velocities generally. Since the methodology developed herein covers both the estimation problem and the various statistical tests of significance required on variances in errors of measurement (or imprecision), as well as tests on accuracy, then the results are believed to be of rather wide applicability.

Preceding page blank

TABLES OF CONTENTS

I. INTRODUCTION

The problem considered here is that of comparing a "test" chronograph of one of the NATO countries with that of two accepted NATO "standard" velocity chronographs, in order to determine whether the "test" or "National" chronograph may be used as an approved instrument for measuring velocities generally. In particular, the problem is approached from the standpoint concerning whether the precision of measurement of the "test" chronograph is about equal to that of the two NATO standards on the average, i. e. whether the variation of errors of measuring velocity is suitably small for the "test" instrument, and also whether the accuracy of measurement of velocity level for the test chronograph is proper, or needs calibration, where the accepted standard velocity level is taken to be that of the average of the two NATO standards. The theory of the suggested procedure is covered in Sections III and IV, whereas in the immediately following paragraphs we give some examples for illustrative purposes.

Our approach to the problem involves the concept of considering each measured velocity as the sum of two components, one the true velocity of the round fired and the other an error of measurement of the chronograph used. Thus, although each measured velocity consists of the sum of two inseparable components, we see that if the error of measurement is small relative to the level of velocity determined and if the variation in errors of measuring velocity is small compared to the variation in true level of velocity itself from round-to-round, then we have an acceptable chronograph for determining velocities. Now if two or more chronographs are used to measure velocities simultaneously on each round of a series of rounds fired, then it can be seen that the difference in readings of any two of the instruments consists of only the differences in (1) errors of measurement and (2) constant biases for the two instruments, and is hence free of the level of velocity measured. The model used here is based on this concept. Indeed, from the variation or variance of the differences in errors of measurement from round-to-round of pairs of the three instruments making simultaneous velocity measurements, one can strip out and hence estimate the variance in errors of measurement of each of the three chronographs. These figures therefore are basic to determining whether the precision of measurement is suitably good as compared with an estimate of the variance of true velocities among rounds in order to judge precision of measurement properly. This, along with a comparison of levels of velocity measured by the chronographs will determine accuracy or the need for calibration. Moreover, statistical tests of significance for precision and accuracy of measurement are rather easily developed and applied, as will be seen.

It happens that the suggested NATO procedure of comparing the velocities of a "National" instrument submitted for test with that of the two NATO "standard" instruments fits in well with existing theory for testing the significance of results statistically, and also the various available procedures for estimating precision of measurement, as well as the estimation of variation of the true velocities measured. Furthermore, on the basis of the velocity measurements taken we may check the agreement of the two NATO standards in any current test in order to see if they are properly calibrated before any judgement on a National chronograph tested for acceptance is actually made.

The procedures are best described in terms of actual data and will be given in sufficient detail below as a computational and analytical procedure.

II. EXAMPLES ON PRECISION AND ACCURACY

Examples on Estimating Precision of Measurement of Three Chronographs

Tables I and III give for twelve consecutive rounds fired from each of two 105mm Guns the three velocities as measured simultaneously by the "FOTOBALK", the "COUNTER" and the "TERMA" type Chronographs. The Foto and the Counter represent the two accepted NATO "standards", whereas the Terma represents a "National" instrument submitted for test. As indicated in Section III, each velocity measured or the ith velocity may be considered as the sum of (1) the true unknown velocity, (2) the bias of the chronograph and (3) a random error of measurement. The first reference gives methods of estimating the variance (and hence standard deviation) in errors of measurement of the three chronographs (or the imprecision of measurement for each chronograph) and also the unbiased estimate of the variance (and standard deviation) of the true velocities measured. Tables I and III give for firings in each of two weapons the individual and mean velocities of the three chronographs, the individual differences in errors of measurement for the three chronographs taken two at a time and the computational procedures for estimating σ_{e1}, σ_{e2}, σ_{e3}, the standard deviations in errors of measurement for each of the three chronographs, and σ_x, the estimated standard deviation of the true velocities measured by each of the instruments. Tables II and IV give pertinent computations for conducting statistical tests of significance which will be described below.

The statistical or mathematical model used for the observed measurements of the three instruments is indicated by the quantities

$$\beta_1 + x_i + e_{i1} = r_i \qquad \beta_2 + x_i + e_{i2} = s_i \qquad \beta_3 + x_i + e_{i3} = t_i$$

where β_1, β_2 and β_3 are respectively the unknown instrumental biases, e_{i1}, e_{i2}, and e_{i3} the random errors of measurement of the three instruments, and x_i the true unknown velocity of the ith round or characteristic measured in a series of measurements on n rounds or items.

In order to estimate the standard deviation of the true velocities measured, one may use either formula (17) or formula (19) of the first reference. Formula (17), for example, gives as an unbiased estimate of the true velocity variance from the quantity

$$\text{est. } (\sigma_x^2) = (S_{x+e1,\ x+e2} + S_{x+e2,\ x+e3} + S_{x+e1,\ x+e3})/3$$

where the S's are sample covariances of the three chronographs, taken two at the time, each based on n-1 degrees of freedom.

It is noted for Table I (rounds 68-79 for the first gun) that the estimated standard deviation of the true velocities measured (2.7m/s) is some three times the greatest of the three standard deviations in errors of measurement for the three chronographs, thereby indicating sufficiently good precision of measurement. For Table III (rounds 20-31 of another weapon), the estimated standard deviation in true velocities measured (1.42 m/s) is also some three times the largest standard deviation in errors of measurement of the chronographs, thereby indicating acceptable precision. (For a standard error of measurement of velocity of .5 m/s the true value 1.42 m/s would be inflated to 1.50 m/s for the observed values of an instrument.

Sometimes the estimated variance in errors of measurement may turn out to be negative, as is the case for the Counter chronograph on rounds 68-79 (Table I). This may be due to smallness of sample size, bad observations, errors in recording, outlying observations, etc. Since it is physically impossible for a true variance in errors of measurement to be negative, then that of the Counter, i. e. - .0525, may be taken as zero, or perhaps until sufficient information is available (see below) the estimated variance in errors of measurement for the Terma chronograph, i. e. .24909 may be subtracted from the total variance in errors of measurement for the three chronographs, i. e. (.55174 + .85333 + .19659)/2 - .24909 to give .55174 which could be split for the time being between the standard Foto and Counter chronographs, in which case the Terma or test chronograph still shows up to be quite acceptable indeed on precision of measurement. For the data of Table I, the difference in errors of measurement for the two standard chronographs, the Foto and the Counter, varied from -1.1 to +1.8 m/s, or 2.9 m/s, which seems unusually large (see Table III). The lack of good agreement between the two standards, therefore, would naturally lead to poorer precision of measurement than would otherwise be expected, and might well call for an investigation of the two "standards" to determine the cause of the "outliers", -1.1 and 1.8 m/s.

In Table III for rounds 20-31 fired from another gun, we see that the standard error of measurement for the Terma instrument remains about the same, whereas the estimates for the Foto and Counter have reversed, indicating considerable variability not normally expected of standards. In any event, the data from both firings (Tables I and III) could be combined and averaged, giving

$$\bar{S}^2_{e_1-e_2} = .305379 \qquad \bar{S}^2_{e_2-e_3} = .233864 \qquad \bar{S}^2_{e_3-e_1} = .539243$$

so that

$$\text{est } (\sigma^2_{e_1}) = .30538 \quad \text{or est } \sigma_{e_1} = .553 \text{ m/s (Foto)}$$

$$\text{est } (\sigma^2_{e_2}) = .00000 \quad \text{or est } \sigma_{e_2} = .000 \text{ m/s (Counter)}$$

$$\text{est } (\sigma^2_{e_3}) = .23386 \quad \text{or est } \sigma_{e_3} = .484 \text{ m/s (Terma)}$$

Thus, the Counter chronograph, due to the averaging of two widely different variances, shows up as having excellent precision of measurement, and the Foto and Terma instruments possess about equal precision, each actually being capable of measuring a velocity standard deviation of either the estimated 1.4 m/s or 2.7 m/s. In the sequel, we recommend that it may be desirable to have two standards which have equal precision of measurement. Thus, the use of two Counter type chronographs as standards if feasible might well be considered.

Procedure for Determining the Acceptability of a National Chronograph Submitted for Test

Here, we turn to the main problem of determining the acceptability for use of a "National" chronograph submitted for test, in this case the Terma instrument. Before actually proceeding to tests of significance of the precision and accuracy of the Terma chronograph, we may as well use the present data available on the Fotobalk and Counter chronographs to check on their precision of measurement and agreement in level of measuring velocities. The needed data are given on Table II for rounds 68-79.

In order to compare the imprecision of measurement, or variances in errors of measurement $\sigma^2_{e_1}$ vs. $\sigma^2_{e_2}$, of the Foto and Counter chronographs for rounds 68-79, we use formula (29) of Section IV:

$$t = r(yz)\sqrt{n-2}/[1 - r^2(yz)]^{1/2} = -.197\sqrt{10}/[1 - (-.197)^2]^{1/2} = -.635$$

Since the 95% significance level of Student's t for 10 degrees of freedom (d.f.) is $t_{.95}$ (10) = 1.812, we conclude that the Fotobalk and Counter have equal precision (or imprecision) of measurement from this significance test. As indicated above, the Counter instrument may be more precise, but there are not enough rounds and hence degrees of freedom in this particular test to establish differences in precision of measurement for rounds 68-79, and the variance $S^2(y)$ of Table II is large.

We next check on the agreement of the Fotobalk and Counter in average level. Here, we use Student's t-test or formula (31). We obtain

$$t_0 = \bar{z}\sqrt{n}/S(z) = \bar{v}_1\sqrt{n}/S(v_1) = (-.19167)\sqrt{12}\ /\sqrt{.55174} = -.894$$

which for 11 degrees of freedom is not significant, since $t_{.95}$ (11) = 1.80. We conclude from this significance test that the Fotobalk and Counter agree also in level of measuring velocity.

Actually, in the significance tests which follow for the Terma or "test" chronograph, the two "standard" chronographs could have unequal precision and also not agree in level, since the formulas used make a comparison of the precision and accuracy of the test or Terma instrument with the <u>average</u> precision and accuracy of the two standards.

We now proceed with the testing of the Terma instrument for acceptability, having run a preliminary test on the two standards. First, using the data of Table II and the computations thereon, we construct a test of significance for judging whether the imprecision of measurement of the Terma instrument is in sufficient agreement with the average of the two standard chronographs. In order to do this, we use formula (36), i. e.

$$t_0(n-2) = \frac{[S^2(u_1)/S^2(v_1) - .75]\sqrt{n-2}}{[3(1-r^2\{u_1v_1\})(S^2\{u_1\}/S^2\{v_1\})]^{1/2}}$$

$$= \frac{[(.387/.552) - .75]\sqrt{10}}{[3(1 - \{-.710\}^2)(.387/.552)]^{1/2}} = -.152$$

which is not significant, since $-t_{.95}(10) = -1.81$. We conclude therefore on the basis of rounds 68-79 that the Terma chronograph has variance in errors of measurement about equal to the average of the Foto and Counter, i. e. .249 vs (.604 - .053)/2 = .276.

Finally, for rounds 68-79, we check on the level of measurement of velocity of the Terma instrument as compared to the average of the Foto and Counter. We now use formula (37), i. e.

$$t_0(n-1) = \bar{u}_1\sqrt{n}/S(u_1) = -1.53\sqrt{12}/.622 = -8.51$$

which for 11 d. f. is very highly significant, indicating the Terma chronograph reads low by 1.53 m/s. In summary, for rounds 68-79, we conclude the Terma may be precise enough, but it reads low by 1.53 m/s, and may require calibration. Est $\sigma_{e_3} = .499$ m/s $= \sqrt{S^2(u_1) - S^2(v_1)/4}$

We now proceed to study rounds 20-31 fired from the other weapon and conduct similar comparisons. First, for the two standards, the Foto and Counter on rounds 20-31, we find from formula (29)

$$t = \frac{r(yz)\sqrt{n-2}}{\sqrt{1-r^2(yz)}} = \frac{.2626\sqrt{10}}{\sqrt{1-(.2626)^2}} = .861$$

so that the Foto and Counter are equally precise. Then, we check levels from formula (31)

$$t_0 = \bar{z}\sqrt{n}/S(z) = -.608\sqrt{12}/(.2429) = -8.67$$

which indicates great significance, so that in this case the Foto reads low by .608 m/s as compared to the other standard, the Counter. Ordinarily, we would look for the cause of this disagreement, run a retest if desirable, or calibrate the two standards. In this particular case, however, we note that the sample variance of the difference in measured velocities (or errors of measurement) is very small, $S^2_{e1-e2} = .059$ and therefore our t-test is sensitive enough to pick up the difference of .608 m/s. For illustrative purposes, we will proceed nevertheless to check out the Terma chronograph on rounds 20-31, since its average will be compared with the average velocity of the Foto and Counter anyway, thereby making no difference since the Foto and Counter instruments are the accepted "standards".

To ascertain for rounds 20-31 whether the variance in errors of measurement of the Terma chronograph (.219) is equal to that of the average of the Fotobalk and Counter instruments (.030), we use formula (36), i. e.

$$t_0\ (n-2) = \frac{[(.2334)/(.0590) - 0.75]\sqrt{10}}{[3(1-(.1959)^2)(.2334)/.0590)]^{1/2}} = 3.00$$

We therefore conclude on the basis of rounds 20-31 that there is considerable evidence the Terma chronograph is not as precise as the (average of the) Fotobalk and Counter instruments, since $t_{.95}\ (10)$ = 1.812.

We note for rounds 20-31 that the standard deviation in errors of measurement for the Terma chronograph is estimated as .468 m/s and this instrument is measuring an estimated standard deviation in true velocity of 1.42 m/s, so that it may be sufficiently precise nevertheless since $\sqrt{(1.42)^2 + (.47)^2} \approx 1.50$ m/s. Hence, we may want to check on level of velocity of the Terma chronograph anyway by using formula (37). Here, we find

$$t_0(n-1) = \bar{u}_1\sqrt{n}/S(u_1) = -\ .421\sqrt{12}/(.483) = -3.02$$

and since $t_{.95}\ (11) = -1.796$, we conclude that the Terma chronograph reads low by .421 m/s.

The variance in errors of measurement of the Terma or third chronograph for rounds 20-31 may be estimated from

Est $\sigma^2_{e_3} = S^2(u_1) - S^2(v_1)/4 = .2334 - .0590/4 = .2187$ m/s which agrees with the value of .2186 computed by a somewhat different formula on Table III.

The standard deviation of the mean velocities listed in the fifth column of Table III is found to be 1.43 m/s as compared to the estimated true value of 1.42 m/s. In summary, therefore, we conclude that the overall measuring process on rounds 20-31 is efficient as far as precision of measurement is concerned, although some calibration of instruments may be desirable.

Finally, we turn to an analysis of results based on combining the data from firings in both weapons.

Analysis of Combined Results

We now apply signfiicance tests to the combined data for the firings in both guns. We noted previously that if the data on variances in errors of measurement from Tables I and III are combined, we obtain estimated standard errors of measurement of est σ_{e1} = .553 m/s (Foto), est σ_{e2} = .000 m/s (Counter), and est σ_{e3} = .484 m/s (Terma). These estimates are obtained from only 24 rounds, and we would continue to revise them as more data are accummulated in order to have accurate information on the performance and reliability of the measuring instruments. Nevertheless, it would appear from these available data that two Counter chronographs could well be used as standards, and it is known from other experience that the Counter is a precise, highly reliable chronograph.

Looking at Tables II and IV, we combine the variances and covariances appropriately, finding

$$\bar{S}^2(y) = (29.053 + 7.508)/2 = 18.281$$

$$\bar{S}^2(z) = (.5517 + .0590)/2 = .3054 = \bar{S}^2(v_1)$$

$$\bar{S}(yz) = (-.7884 + .1748)/2 = -.3068$$

$$\bar{S}^2(u_1) = (.3870 + .2334)/2 = .3102$$

$$\bar{S}(u_1 v_1) = (-.3284 + .0230)/2 = -.1527$$

$\bar{r}\,(yz) = -\ .3068/\sqrt{(18.281)(.3054)} = -\ .1298$

$\bar{r}\,(u_1v_1) = -\ .1527/\sqrt{(.3102)(.3054)} = -\ .4961$

Testing the significance of the relative imprecisions of measurement of the Foto and Counter for the 24 rounds, we compute from formula (29)

$$t = \frac{\bar{r}(yz)\sqrt{2(n-2)}}{\sqrt{1-\bar{r}^2(yz)}} = \frac{-.1298\sqrt{20}}{\sqrt{1-(-.1298)^2}} = -.585$$

so that we still don't pick up any difference in imprecision of measurement for the Foto and Counter instruments. This is rather surprising indeed, but the data of rounds 68-79 show considerable heterogeneity as compared to rounds 20-31. Had all rounds been fired in the latter gun, we probably could have established that the Counter instrument is more precise than the Fotobalk. [Formula (30) does accept the hypothesis that $\sigma_{e2} = 0$ on 1 d.f.]

Next, we check on accuracy or agreement in level of the Foto and Counter for all rounds. We have that the average difference in readings is

$$(-.1917\ -\ .6083)/2 = -\ .400$$

and computing from (31)

$$t_0 = \bar{v}_1\sqrt{24}/\bar{S}(v_1) = -\ .400(4.899)/(.5526) = -\ 3.55$$

which is highly significant, since $-\ t_{.95}\ (22) = -\ 1.72$ for 22 degrees of freedom. We are now able from all 24 rounds to conclude therefore that the Fotobalk reads on the average .40 m/s lower than the Counter.

Now we proceed to significance tests on the Terma for both series of rounds, 68-79 and 20-31. First, we test the imprecision of measurement of the Terma versus the average of the Foto and Counter. Here, we find

$$t_0(2n-4) = \frac{[\bar{S}^2(u_1)/\bar{S}^2(v_1)\ -\ .75]\sqrt{2n-4}}{\sqrt{3[1-\bar{r}^2(u_1v_1)]\bar{S}^2(u_1)/\bar{S}^2(\bar{v}_1)}}$$

or

$$t_0(20) = \frac{[(.3102)/(.3054)\ -\ .75]\sqrt{20}}{\sqrt{3[1-(-.496)^2](.3102)/(.3054)}} = .784$$

so that the Terma's imprecision (.484 m/s) is not significantly different from the average imprecision of the Foto and Counter ($\sqrt{.3054/2}$ = .391 m/s).

Finally, we check on overall calibration or accuracy of the Terma. Here, the overall difference in average velocity of the Terma and the average of the Foto and Counter is

$$(-1.53 - .421)/2 = - .976 \text{ m/s}$$

and using the t-test of (37), we find

$$t_0(22) = \bar{u}_1 \sqrt{2n}/\bar{S}/(u_1) = - .976 \sqrt{24}/.557 = - 8.58$$

so that we conclude the Terma reads 0.976 m/s lower than the average of the Foto and Counter.

In summary, using data for both groups of 12 rounds, 68-79 and 20-31, we have not been able to pick up any great differences in precision of measurement of one instrument over the other, although the Counter appears the more precise instrument from estimation procedures. As far as level of measuring velocity is concerned, it is rather clear that calibration is required for all three chronographs. The three instruments together, or individually, perform a suitably good job of measuring velocity variation as they seem to be sufficiently precise for this purpose.

If any test chronographs fail to record velocity data or lose rounds, then they may not be acceptable on such grounds alone, of course, irrespective of precision and accuracy otherwise.

Finally, the above analyses based on just 24 rounds represent only a beginning, especially for the chronographs considered as standards. As more firings are carried out, and the data analyzed, then the information accumulated should be combined so that the precise characteristics of the standard instruments are known in great detail in so far as precision, accuracy and reliability are concerned.

TABLE I

ESTIMATES OF PRECISION OF MEASUREMENT BASED ON THREE SIMULTANEOUS VELOCITY MEASUREMENTS OF THE FOTOBALK, COUNTER AND TERMA CHRONOGRAPHS

Round No.	Foto r	Counter s	Terma t	Mean Velocity m/s	r-s	s-t	t-r
68	813.3	812.8	811.2	812.43	+0.5	+1.6	-2.1
69	820.4	821.2	819.4	820.33	-0.8	+1.8	-1.0
70	816.4	816.7	815.0	816.03	-0.3	+1.7	-1.4
71	819.6	820.0	818.0	819.20	-0.4	+2.0	-1.6
72	817.8	817.7	815.4	816.97	+0.1	+2.3	-2.4
73	819.5	819.8	817.7	819.00	-0.3	+2.1	-1.8
74	814.3	814.8	813.2	814.00	-0.5	+1.6	-1.1
75	821.3	819.5	817.8	819.53	+1.8	+1.7	-3.5
76	819.1	819.5	818.1	818.90	-0.4	+1.4	-1.0
77	819.1	819.6	818.9	819.20	-0.5	+0.7	-0.2
78	817.1	817.5	816.5	817.03	-0.4	+1.0	-0.6
79	821.9	823.0	821.4	822.10	-1.1	+1.6	-0.5

$S^2_{e_1-e_2} = .551742$* $\qquad \bar{e}_1 - \bar{e}_2 = -.191667 = \bar{z} = \bar{v}_1$

$S^2_{e_2-e_3} = .196591 \qquad \bar{e}_2 - \bar{e}_3 = 1.62500$

$S^2_{e_3-e_1} = .853333 \qquad \bar{e}_3 - \bar{e}_1 = -1.43333$

Using Equation (9) of the first Reference:

$\text{est}(\sigma^2_{e_1})$ $1/2(.551742 + .853333 - .196591) = .60424$ or est $\sigma_{e_1} = .777$ m/s(Foto)

$\text{est}(\sigma^2_{e_2}) = 1/2(.551742 + .196591 - .853333) = -.0525$ or est $\sigma_{e_2} = 0$ m/s(Counter)

$\text{est}(\sigma^2_{e_3}) = 1/2(-.551742 + .196591 + .853333) = .24909$ or est $\sigma_{e_3} = .499$ m/s(Terma)

est $\sigma_x = 2.70$ m/s = Estimated standard deviation of true velocity

*$S^2_{e_1-e_2} = S^2_r + S^2_s - 2S_{rs}$, etc, also.

TABLE II

SIMULTANEOUS VELOCITIES OF THE FOTOBALK, COUNTER AND TERMA CHRONOGRAPHS ON EACH OF TWELVE ROUNDS IN METERS PER SECOND (m/s)

Round No.	Foto r	Counter s	Terma t	(r+s) - 1620=y	(r-s) = z or v_1	t - (r+s)/2 = u_1
68	813.3	812.8	811.2	6.1	0.5	-1.85
69	820.4	821.2	819.4	21.6	-0.8	-1.40
70	816.4	816.7	815.0	13.1	-0.3	-1.55
71	819.6	820.0	818.0	19.6	-0.4	-1.80
72	817.8	817.7	815.4	15.5	0.1	-2.35
73	819.5	819.8	817.7	19.3	-0.3	-1.95
74	814.3	814.8	813.2	9.1	-0.5	-1.35
75	821.3	819.5	817.8	20.8	1.8	-2.60
76	819.1	819.5	818.1	18.6	-0.4	-1.20
77	819.1	819.6	818.9	18.7	-0.5	-0.45
78	817.1	817.5	816.5	14.6	-0.4	-0.80
79	821.9	823.0	821.4	24.9	-1.1	-1.05

$S^2(y) = [12(3716.55) - (201.9)^2]/132 = 29.053$*

$S^2(z) = S^2(v_1) = [12(6.51) - (-2.3)^2]/132 = .5517$

$S^2(u_1) = [12(32.3175) - (18.35)^2]/132 = .3870$

$S(yz) = [12(-47.37) - (-2.3)(201.9)]/132 = -.7884$**

$S(u_1 v_1) = [12(-.095) - (-2.3)(-18.35)]/132 = -\ .3284$

$r(yz) = -.788/\sqrt{(29.053)(.5517)} = -\ .197$

$r(u_1 v_1) = -\ .3284/\sqrt{(.3870)(.5517)} = -.710$

Mean (r-s) = $\bar{z} = \bar{v} = -.1917$ $\bar{u}_1 = -1.529$

* $S^2(y) = [n\ \Sigma y_i^2 - (\Sigma y_i)^2]/n(n-1)$

** $S(yz) = [n \Sigma y_i z_i - (\Sigma y_i)(\Sigma z_i)]/n(n-1)$

TABLE III

ESTIMATES OF PRECISION OF MEASUREMENT BASED ON THREE SIMULTANEOUS VELOCITY MEASUREMENTS OF THE FOTOBALK, COUNTER AND TERMA CHRONOGRAPHS

Round No.	Foto r	Counter s	Terma t	Mean Velocity m/s	r-s	s-t	t-r
20	793.8	794.6	793.2	793.87	-0.8	+1.4	-0.6
21	793.1	793.9	793.3	793.43	-0.8	+0.6	+0.2
22	792.4	793.2	792.6	792.73	-0.8	+0.6	+0.2
23	794.0	794.0	793.8	793.93	0.0	+0.2	-0.2
24	791.4	792.2	791.6	791.73	-0.8	+0.6	+0.2
25	792.4	793.1	791.6	792.37	-0.7	+1.5	-0.8
26	791.7	792.4	791.6	791.90	-0.7	+0.8	-0.1
27	792.3	792.8	792.4	792.50	-0.5	+0.4	+0.1
28	789.6	790.2	788.5	789.43	-0.6	+1.7	-1.1
29	794.4	795.0	794.7	794.70	-0.6	+0.3	+0.3
30	790.9	791.6	791.3	791.27	-0.7	+0.3	+0.4
31	793.5	793.8	793.5	793.60	-0.3	+0.3	0.0

$S^2_{e_1 - e_2} = .059015$ $\quad\quad \bar{e}_1 - \bar{e}_2 = -.608333$

$S^2_{e_2 - e_3} = .271136$ $\quad\quad \bar{e}_2 - \bar{e}_3 = +.725000$

$S^2_{e_3 - e_1} = .225152$ $\quad\quad \bar{e}_3 - \bar{e}_1 = -.116667$

Using Equation (9) of the first Reference:

est($\sigma^2_{e_1}$) = 1/2(.059015+.225152-.271136)=.00652 or est σ_{e_1} = .0255 m/s(Foto)

est($\sigma^2_{e_2}$) = 1/2(.059015-.225152+.271136)= .05250 or est σ_{e_2}= .229 m/s(Counter)

est($\sigma^2_{e_3}$) = 1/2(-.059015+.225152+.271136)= .21863 or est σ_{e_3}= .468 m/s(Terma)

est $\sigma_{\wedge}$ = 1.42 m/s = Estimated standard deviation of true velocity

TABLE IV

SIMULTANEOUS VELOCITIES OF THE FOTOBALK, COUNTER AND TERMA CHRONOGRAPHS ON EACH OF TWELVE ROUNDS IN METERS PER SECOND (m/s)

Round No.	Foto r	Counter s	Terma t	(r+s)-1580=y	r-s =z or v_1	t-(r+s)/2 = u_1
20	793.8	794.6	793.2	8.4	-.8	-1.00
21	793.1	793.9	793.3	7.0	-.8	- .20
22	792.4	793.2	792.6	5.6	-.8	- .20
23	794.0	794.0	793.8	8.0	.0	- .20
24	791.4	792.2	791.6	3.6	-.8	- .20
25	792.4	793.1	791.6	5.5	-.7	-1.15
26	791.7	792.4	791.6	4.1	-.7	- .45
27	792.3	792.8	792.4	5.1	-.5	- .15
28	789.6	790.2	788.5	-0.2	-.6	-1.40
29	794.4	795.0	794.7	9.4	-.6	.00
30	790.9	791.6	791.3	2.5	-.7	+ .05
31	793.5	793.8	793.5	7.3	-.3	- .15

$S^2(y) = [n\Sigma y_i^2 - (\Sigma y_i)^2]/n(n-1) = [12(448.89) - (66.3)^2]/132 = 7.508$

$S^2(z) = S^2(v_1) = [12(5.09) - (-7.3)^2]/132 = .0590$

$S^2(u_1) = [12(4.6925) - (-5.05)^2]/132 = .2334$

$S(yz) = [n\Sigma y_i z_i - (\Sigma y_i)(\Sigma z_i)]/n(n-1) = [12(-38.41)-(66.3)(-7.3)]/132 = .1748$

$S(u_1 v_1) = [12(3.325 - (-5.05)(-7.3)]/132 = .0230$

$r(yz) = (.1748)/\sqrt{(7.508)(.0590)} = .2626$

$r(u_1 v_1) = (.0230)/\sqrt{(.2334)(.0590)} = .1959$

Mean(r-s) $= \bar{z} = \bar{v}_1 = -.608$ $\qquad \bar{u}_1 = -.421$

III. THEORY*

We consider three instruments which are used to make simultaneous or the same measurements on each of a series of n items. Their measurements are indicated by the quantities

$$(1) \quad \beta_1 + x_i + e_{i1} = r_i \qquad \beta_2 + x_i + e_{i2} = s_i \qquad \beta_3 + x_i + e_{i3} = t_i$$

where β_1, β_2 and β_3 are respectively the (unknown) instrumental biases, e_{i1}, e_{i2} and e_{i3} the random errors of measurement of the three instruments, and x_i the true, unknown value of the ith item or characteristic measured in a series of measurements on n items. For the significance tests indicated below, the x_i and the e_{ij} are assumed to be normally and independently distributed.

From the common or simultaneous measurements made by the three instruments, we form for items i = 1, 2, ..., n the following three columns of differences:

$$(2) \quad \beta_1 - \beta_2 + e_{i1} - e_{i2} \qquad \beta_2 - \beta_3 + e_{i2} - e_{i3} \qquad \beta_3 - \beta_1 + e_{i3} - e_{i1}$$

Instruments 1 and 2 may be considered as standards, or instruments of known precision, and our interest centers around whether the "test" instrument 3 may be precise enough (i. e. possesses sufficiently small variance in errors of measurements), and whether instrument 3 reads at the correct level, as compared to the two standards, or needs calibration. Methods are given by Grubbs (1948) for estimating the true product variance , and the true variances in errors of measurement for the three instruments, σ^2_{e1}, σ^2_{e2} and σ^2_{e3}. These variance estimates, which

*Sections III and IV were prepared by the first-named author.

unfortunately are sometimes negative, nevertheless give an initial idea of the sizes of σ^2_{e1}, σ^2_{e2} and σ^2_{e3} preliminary to a test of significance.

It is easy to see that if the covariances of normally distributed quantities are zero then such quantities are independently distributed. Hence, ignoring for the moment the constant differences in biases, we see that the covariance of $e_{i1} - e_{i2}$ and $l_1 (e_{i2} - e_{i3}) + l_2(e_{i3} - e_{i1})$, where l_1 and l_2 are constants, is given by

(3) $$-l_1 \sigma^2_{e2} - l_2 \sigma^2_{e1}$$

and this is zero if and only if

(4) $$-l_1/l_2 = \sigma^2_{e1}/\sigma^2_{e2}$$

Thus, if $\sigma^2_{e1} = \sigma^2_{e2}$, then we may use weights of -1/2 and + 1/2 (adding to one) as follows for the second and third pairs of differences.

(5) $$u_{1i} = - 1/2(e_{i2} - e_{i3}) + 1/2(e_{i3} - e_{i1}) = e_{i3} - (e_{i1} + e_{i2})/2$$

and this quantity is therefore, for normally distributed variates, independent of

(6) $$v_{1i} = e_{i1} - e_{i2}$$

which establishes the validity of an F-test on σ^2_{e3} and a t-test on level of measurement of the third instrument as found by Hahn and Nelson (1970).

On the other hand, even though σ_{e1} and σ_{e2} are not equal, then the u_{1i} and v_{1i} of (5) and (6) may still be used to advantage. In this case, the population covariance of (5) and (6) is clearly

(7) $$\sigma(u_1 v_1) = (\sigma^2_{e2} - \sigma^2_{e1})/2$$

or the true correlation coefficient is

(8) $\rho(u_1 v_1) = (\sigma^2_{e_2} - \sigma^2_{e_1})/[(\sigma^2_{e_1} + \sigma^2_{e_2})(4\sigma^2_{e_3} + \sigma^2_{e_1} + \sigma^2_{e_2})]^{1/2}$

which in many practical instances may be somewhat near zero. In any event, we may compute the variance and covariances

(9) $S^2(u_1) = \sum_{i=1}^{n}(u_{1i} - \bar{u}_1)^2/(n - 1) = [n \Sigma u^2_{1i} - (\Sigma u_{1i})^2]/n(n - 1)$

(10) $S^2(v_1) = \sum_{i=1}^{n}(v_{1i} - \bar{v}_1)^2/(n - 1) = [n \Sigma v^2_{1i} - (\Sigma v_{1i})^2]/n(n - 1)$

(11) $S(u_1 v_1) = \sum_{i=1}^{n} (u_{1i} - \bar{u}_1)(v_{1i} - \bar{v}_1)/(n - 1)$

$= [n \Sigma u_{1i}v_{1i} - (\Sigma u_{1i})(\Sigma v_{1i})]/n(n - 1)$

and use the Pitman-Morgan (1939) test for correlated variances:

(12) $$t(n - 2) = \frac{[S^2(u_1)/S^2(v_1) - \lambda]\sqrt{n-2}}{[4(1 - r^2)\,\lambda\,\{S^2(u_1)/S^2(v_1)\}]^{1/2}}$$

where t(n-2) is Student's t with n-2 degrees of freedom, the observed or sample correlation coefficient r is given by

(13) $r = S(u_1 v_1)/[S(u_1)\ S(v_1)]$

and the ratio of population variances of u_1 and v_1 is

(14) $\lambda = [\sigma^2_{e_3} + (\sigma^2_{e_1} + \sigma^2_{e_2})/4](\sigma^2_{e_1} + \sigma^2_{e_2}) = \sigma^2(u_1)/\sigma^2(u_2)$

Thus, to test the null-hypothesis that $\sigma^2_{e_3} = (\sigma^2_{e_1} + \sigma^2_{e_2})/2$, we use the value $\lambda = 3/4$ in formula (12).

In case $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ are rather accurately known, perhaps from past data, it is easy to see that the quantity

(15) $$w_i = [-\sigma_{e1}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)](e_{i2} - e_{i3}) + [\sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)](e_{i3} - e_{i1})$$

$$= e_{i3} - [\sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)]e_{i1} - [\sigma_{e1}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)]e_{i2}$$

is also independent of $e_{i1} - e_{i2}$, so that a valid F test can therefore still be developed for three different instruments. In this case of known σ_{e1} and σ_{e2}, we define and compute the following:

(16) $$u_{2i} = [-\sigma_{e1}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)][\beta_2 - \beta_3 + e_{i2} - e_{i3}] + [\sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)][\beta_3 - \beta_1 + e_{i3} - e_{i1}]$$

(17) $$v_{2i} = \beta_1 - \beta_2 + e_{i1} - e_{i2}$$

(18) $$S^2(u_2) = \sum_{i=1}^{n} (u_{2i} - \bar{u}_2)^2/(n - 1)$$

(19) $$S^2(v_2) = \sum_{i=1}^{n} (v_{2i} - \bar{v}_2)^2/(n - 1)$$

(20) $$S(u_2 v_2) = \sum_{i=1}^{n} (u_{2i} - u_2)(v_{2i} - \bar{v}_2)/(n - 1)$$

The population variance of u_{2i} is given by

(21) $$\sigma_{u2}^2 = \sigma_{e3}^2 + \sigma_{e1}^2\, \sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)$$

and the unbiased estimate of σ_{u2}^2 is determined from

(22) $$\hat{\sigma}_{u2}^2 = [\sigma_{e1}^4/(\sigma_{e1}^2 + \sigma_{e2}^2)^2]\, S^2(u_2) - 2[\sigma_{e1}^2\, \sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)^2] S(u_2\, v_2) + [\sigma_{e2}^4/(\sigma_{e1}^2 + \sigma_{e2}^2)^2]\, S^2(v_2)$$

With the above definitions, therefore, it is clear that

(23) $(n-1)\, S^2(u_2)/[\sigma_{e3}^2 + \sigma_{e1}^2\, \sigma_{e2}^2/(\sigma_{e1}^2 + \sigma_{e2}^2)] = \chi^2\,(n-1)$:Chi-square w/n-1 d. f.

(24) $(n-1)\, S^2(v_2)/(\sigma_{e_1}^2 + \sigma_{e_2}^2) = \chi^2\,(n-1)$:Chi square w/n-1 d. f.

and

$$\frac{S^2(u_2) \cdot [\sigma^2_{e_1} + \sigma^2_{e_2}]}{S^2(v_2) \cdot [\sigma^2_{e_3} + \sigma^2_{e_1} \sigma^2_{e_2} / (\sigma^2_{e_1} + \sigma^2_{e_2})]} = F(n-1, n-1) \qquad (25)$$

or Snedecor's F with (n-1) and (n-1) degrees of freedom.

Thus, to test the hypothesis $\sigma^2_{e_3} = (\sigma^2_{e_1} + \sigma^2_{e_2})/2$, we compute

$$F(n-1, n-1) = S^2(u_2)/S^2(v_2)[1/2 + \sigma^2_{e_1} \sigma^2_{e_2} / (\sigma^2_{e_1} + \sigma^2_{e_2})^2] \qquad (26)$$

and compare this observed value of F with a desired probability level or percentage point of the F-distribution.

When $\sigma_{e_1} = \sigma_{e_2}$, then it may be noted that (26) reduces to special case

$$F(n-1, n-1) = 4S^2(u)/3S^2(v) \qquad (27)$$

which is the known result of Hahn and Nelson (1970). Of course, it is a rather desirable condition in applications to have $\sigma_{e_1} = \sigma_{e_2}$, for independence is assured and the computations are somewhat easier.

Finally, it is clear that there are no problems of dependence involved in using t tests to judge differences in levels of measurement, for then sample means and sample variances are independently distributed for normal variates, and the ordinary theory applies therefore.

IV. SUMMARY OF USEFUL FORMULAS AND RESULTS

A. Two-Instrument Case

For the two-instrument case, we record here for reference the results of Maloney and Rastogi (1970). Here, we take the readings from the two instruments, i. e.

$$\beta_1 + x_i + e_{i1} = r_i \qquad\qquad \beta_2 + x_i + e_{i2} = s_i$$

and form their sums and differences

$$y_i = r_i + s_i = \beta_1 + \beta_2 + 2x_i + e_{i_1} + e_{i_2}$$

$$z_i = r_i - s_i = \beta_1 - \beta_2 + e_{i_1} - e_{i_2}$$

We then compute the sample correlation coefficient of y and z:

(28) $$r(yz) = S(yz)/S(y)S(z)*$$

Finally, we compute the quantity

(29) $$t = r(yz)\ \sqrt{n-2}/[1 - r^2(yz)]^{1/2}$$

which is distributed as Student's t with n-2 degrees of freedom.

We note that the population correlation of y and z is

$$\rho(yz) = (\sigma^2_{e1} - \sigma^2_{e2})/[(4\sigma^2_X + \sigma^2_{e1} + \sigma^2_{e2})(\sigma^2_{e1} + \sigma^2_{e2})]^{1/2}$$

which is zero if and only if $\sigma_{e1} = \sigma_{e2}$. Thus, the t-statistic of (29) is precisely a test of whether $\sigma_{e1} = \sigma_{e2}$. In consequence of this, we accept the hypothesis $\sigma_{e1} = \sigma_{e2}$ at the α probability level if the observed $t = t_0$ is such that

$$-t_{1-\alpha/2} \leq t_0 \leq t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the upper $\alpha/2$ probability level of t with n-2 d. f.

We accept $\sigma_{e_1} < \sigma_{e_2}$ in a one-sided test at the α probability level if the observed

$$t_0 < - t_{1-\alpha}$$

and we accept $\sigma_{e1} > \sigma_{e2}$ at the α probability level if the observed

$$t_0 > t_{1-\alpha}$$

$*r(yz) = (S^2_r - S^2_s)/[(S^2_r + S^2_s + 2S_{rs})(S^2_r + S^2_s - 2S_{rs})]^{1/2}$ also.

Maloney and Rastogi (1970) also develop a test of whether $\sigma_{e1} = 0$, or whether $\sigma_{e2} = 0$. To test whether $\sigma_{e1} = 0$, they use

(30) $$-2 \ln \lambda = -2 \ln[(S_r^2 S_s^2 - S_{rs})/(S_r^2\{S_r^2 + S_s^2 - 2S_{rs}\})]^{n/2}$$

where S_r, S_s are the ordinary sample variances and S_{rs} the sample covariance of r_i and s_i based on n-1 d. f. The quantity $-2\ln \lambda$ for sufficiently large n is distributed as Chi-square with 1 d. f. as shown by S. S. Wilks. We therefore reject the hypothesis $\sigma_{e1} = 0$ if the quantity $-2\ln \lambda > \chi^2_{1-\alpha}(1)$, where $\chi^2_{1-\alpha}(1)$ is the upper α probability level of the chi-square distribution with 1 d. f.

To test whether $\sigma_{e2} = 0$, we simply interchange the r_i and s_i.

A t-test for the hypothesis that $\beta_1 = \beta_2$, or that the levels of measurement for instruments 1 and 2 agree, may be found by computing (from the z_i) the observed

(31) $$t_0 = \bar{z}\sqrt{n}/S(z) = \bar{v}_1\sqrt{n}/S(v_1)$$

Accept $\beta_1 < \beta_2$, $\beta_1 = \beta_2$ or $\beta_1 > \beta_2$ (in individual significance tests) at the α probability level, depending respectively on whether $t_0 < -t_{1-\alpha}$, $-t_{1-\alpha/2} \leq t_0 \leq t_{1-\alpha/2}$, or $t_0 > t_{1-\alpha}$

B. Three-Instrument Case

First, and as a routine for three simultaneous or common measurements, we might well use formulas (29) and (31) to check the performance of instruments 1 and 2 before proceeding with tests on instrument 3. Indeed, it may be desirable to calibrate instruments 1 or 2, or both. Otherwise, we proceed to test instrument 3 as follows:

B1. If it should happen that $\sigma_{e_1} = \sigma_{e_2}$, then use the quantities (5) and

(6), i. e. $t_i - (r_i + s_i)/2$ and $r_i - s_i$, to compute $S^2(u_1)$ and $S^2(v_1)$ from (9) and (10). Then find the quantity

(32) $$F_0(n-1, n-1) = 4S^2(u_1)/3S^2(v_1)$$

which allows for possible differences in levels of measurement for instruments 1 and 2, and refer F_0 to the F table. In individual tests we then accept $\sigma^2_{e3} < (\sigma^2_{e1} + \sigma^2_{e2})/2$, $\sigma^2_{e3} = (\sigma^2_{e1} + \sigma^2_{e2})/2$, or $\sigma^2_{e3} > (\sigma^2_{e1} + \sigma^2_{e2})/2$ at the α level of probability depending respectively on whether we find $F_0 < 1/F_{1-\alpha}(n-1, n-1)$, $1/F_{1-\alpha/2}(n-1, n-1) \leq F_0 \leq F_{1-\alpha/2}(n-1, n-1)$, or $F_0 < F_{1-\alpha}$ (n-1, n-1)

For a t-test when $\sigma_{e1} = \sigma_{e2}$ concerning the level of measurement of instrument 3, or the relative sizes of β_3 and $(\beta_1 + \beta_2)/2$, then compute the observed t from

(33) $$t_0 = \bar{u}_1\sqrt{n}/S(u_1)$$

and in individual tests accept $\beta_3 < (\beta_1 + \beta_2)/2$, $\beta_3 = (\beta_1 + \beta_2)/2$ or $\beta_3 > (\beta_1 + \beta_2)/2$ at the α probability level, depending respectively on whether $t_0 < t_{1-\alpha}$, $-t_{1-\alpha/2} \leq t_0 \leq t_{1-\alpha/2}$, or $t_0 > t_{1-\alpha}$. Thus, instrument 3, if not measuring at the proper level, could be calibrated if its imprecision of measurement, σ_{e3}, is suitably small.

We note that σ^2_{e3} may be estimated from

$$\hat{\sigma}^2_{e3} = S^2(u_1) - S^2(v_1)/4$$

if this quantity is positive and otherwise $\hat{\sigma}_{e3}$ is taken as equal to zero.

Following Hahn and Nelson (1970), one-sided lower and upper $1 - \alpha$ confidence bounds on

$$\sigma_{e3}/[(\sigma^2_{e1} + \sigma^2_{e2})/2]^{1/2}$$

may be obtained respectively from

(34)
$$\frac{2S^2(u_1)}{F_{1-\alpha}(n-1,\ n-1)\ S^2(v_1)} - 1/2$$

and

(35)
$$2F_{1-\alpha}(n-1,\ n-1)\ S^2(u_1)/S^2(v_1) - 1/2$$

B2. In case it is not known whether $\sigma_{e1} = \sigma_{e2}$, and the relative sizes of σ_{e1} and σ_{e2} are unknown, then use $t_i - (r_i + s_i)/2$ and $r_i - s_i$ to compute (9), (10) and (11). Then use the Pitman-Morgan (1939) test for correlated variances based on the observed value of $t = t_0$ for n - 2 d. f. and $\lambda = 3/4$, i. e.

(36)
$$t_0(n-2) = \frac{[S^2(u_1)/S^2(v_1) - 0.75]\sqrt{n-2}}{[3(1-r^2)\{S^2(u_1)/S^2(v_1)\}]^{1/2}}$$

On the basis of this observed value, which is referred to a table of percentage points of t, we decide in individual tests whether $\sigma^2_{e3} < (\sigma^2_{e1} + \sigma^2_{e2})/2$, $\sigma^2_{e3} = (\sigma^2_{e1} + \sigma^2_{e2})/2$, or $\sigma^2_{e3} > (\sigma^2_{e1} + \sigma^2_{e2})/2$ at the α probability level, depending respectively on whether $t_0(n-2) < t_{1-\alpha}$, $-t_{1-\alpha/2} \leq t_0(n-2) \leq t_{1-\alpha/2}$, or $t_0(n-2) > t_{1-\alpha}$. We remark that (12) could be used to find confidence bounds on λ or even on the quantity $2\sigma^2_{e3}/(\sigma^2_{e1} + \sigma^2_{e2})$, or its square root.

For judging the level of measurement of instrument 3 as compared to the average level of instruments 1 and 2 for this case, i. e. the size of β_3 versus $(\beta_1 + \beta_2)/2$, then compute the observed value of t for n-1 d. f. given by

(37) $$t_0(n-1) = \bar{u}_1\sqrt{n}\ /S(u_1)$$

We conclude in individual tests at the α probability level that $\beta_3 < (\beta_1 + \beta_2)/2$, $\beta_3 = (\beta_1 + \beta_2)/2$, or $\beta_3 > (\beta_1 + \beta_2)/2$, according respectively to whether $t_0 < t_{1-\alpha}$, $t_{1-\alpha/2} \leq t_0(n-1)\ t_{1-\alpha/2}$, or $t_0(n-1) > t_{1-\alpha/2}$. We could then make corrections to or calibrate instrument 3 accordingly.

B3. Finally, if $\sigma_{e1} \neq \sigma_{e2}$, and we know their relative sizes, $\sigma_{e1} = k\sigma_{e2}$ say, from experience, then use (16), (17), (18) and (19) to compute the observed value of F in (26), i. e.

(38) $$F_0(n-1,\ n-1) = S^2(u_2)/S^2(v_2)[1/2 + \sigma^2_{e2}\ \sigma^2_{e1}/(\sigma^2_{e1} + \sigma^2_{e2})^2]$$

Refer this observed F_0 to percentage points of Snedecor's F-distribution to decide in individual tests if $\sigma^2_{e3} < (\sigma^2_{e1} + \sigma^2_{e2})/2$, $\sigma^2_{e3} = (\sigma^2_{e1} + \sigma^2_{e2})/2$, or $\sigma^2_{e3} > (\sigma^2_{e1} + \sigma^2_{e2})/2$, at the α probability level, depending respectively on whether $F_0 < 1/F_{1-\alpha}(n-1,\ n-1)$, $1/F_{1-\alpha/2}(n-1,\ n-1) \leq F_0 \leq F_{1-\alpha/2}(n-1, n-1)$, or $F_0 > F_{1-\alpha}(n-1,\ n-1)$.

To find if $\beta_3 < (\beta_1 + \beta_2)/2$, $\beta_3 = (\beta_1 + \beta_2)/2$, or $\beta_3 > (\beta_1 + \beta_2)/2$, then use (22) and compute the observed t from

(39) $$t_0(n-1) = \bar{u}_2\sqrt{n}/\hat{\sigma}_{u2}$$

and decide on the above order respectively according to whether in single tests $t_0 < t_{1-\alpha}$, $-t_{1-\alpha/2} \leq t_0 \leq t_{1-\alpha/2}$, or $t_0 > t_{1-\alpha}$.

We remark that confidence levels on the ratio $2\sigma^2_{e3}/(\sigma^2_{e1} + \sigma^2_{e2})$ or its square root may be obtained from (25).

For the three-instrument case, it should be noted that significance tests depend on only the errors of measurement e_{ij} being normally distributed, whereas the x_i represent the common or same true values measured by each of the three instruments.

REFERENCES

1. Grubbs, F. E., On Estimating Precision of Measuring Instruments and Product Variability, *Journal of the American Statistical Association, Vol. 43, pp 243-264, June 1948.*

2. Hahn, J. H. and Nelson, W., A Problem in the Statistical Comparison of Measuring Devices. *Technometrics, Vol. 12, No. 1, Feb. 1970, pp 95-102.*

3. Maloney, C. J. and Rastogi, S. C., Significance Test for Grubbs' Estimators. *Biometrika, Vol. 26, No. 4, pp 671-676, Dec. 1970.*

4. Morgan, W. A., A Test for the Significance of the Difference Between Two Variances in a Sample From a Normal Bivariate Population, *Biometrika, Vol. 31, pp 13-19 (1939).*

5. Pitman, E. J. G., A Note on Normal Correlation. *Biometrika, Vol. 31, pp 9-12 (1939).*