ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 3, 109–123 (1968)

Intuitive Statistical Inferences about Variances[1]

LEE ROY BEACH
University of Washington

AND

THOMAS S. SCOPP
University of Michigan

Subjects saw samples from each of two populations of numbers and made intuitive inferences about which population had the larger variance. Then they either estimated the ratios of the variances or stated their confidence (subjective probability) in their inferences. The ratios were used to infer the subjective magnitudes of the sample variances; they were systematically inaccurate because of a tendency to underweight deviant sample data and because the subjects regard variance among large numbers as less variable than variance among small numbers. Then, confidence in the inferences about the population variances was compared to the probabilities that would have resulted if the ratios of sample variances had actually been the ratios that the subjects reported. Confidence was systematically related to these probabilities but it was always lower. The results are discussed in terms of the conservatism findings reported in other investigations of intuitive statistics. A Bayesian F-test is appended.

Information about the factors that govern occurrences of future events is a fundamental prerequisite for optimal decision making. When the domain of interest can be described in statistical terms, the required information consists of the parameters of the populations from which future samples will derive. In most decision experiments, particularly those that involve gambling, the subject is told the parameters of the relevant population, e.g., that the outcome of the bet will be determined by drawing from a binomial population with $P = .70$ and $1—P = .30$. The research then focuses on the optimality of subsequent choices and decisions. However, in their normal experience people seldom have direct access to envi-

[1] This research was done at the Engineering Psychology Laboratories, Institute of Science and Technology, The University of Michigan. Support was provided by USPHS Fellowship MF-12,744 and USPHS grant NIMH-11496, both from the National Institutes of Mental Health, and by the Air Force Office of Scientific Research under contract AF 49(638)-1731.

ronmental population parameters; usually they must infer the values from previously observed samples. Because of this, it is of interest to investigate how subjects use samples to make predecisional inferences about population parameters.

Previous experiments have examined how subjects use samples to infer specific values of a population parameter (e.g., Peterson and Phillips, 1966), others have studied how they infer which of two specified values of a parameter characterizes the population from which a sample was drawn (e. g., Phillips and Edwards, 1966) and others have studied subjects' use of samples to infer which of two populations had the larger value of a specified parameter (e. g., Irwin *et. al.*, 1956). A statistician could perform these tasks using the appropriate formal statistical inference procedures. Subjects, however, must perform them by relying upon what has been termed, intuitive statistical inference (Brunswik, 1956). The research has used the formal procedures as a logically consistent, normative framework within which to examine the process of intuitive statistical inference.

The majority of investigations of intuitive statistical inference have used binomial population proportions as the parameters of interest. Aside from three studies of means (Edwards, 1965; Irwin *et. al.*, 1956; Little and Lintz, 1965) and one of correlations (Beach and Scopp, 1966), normal populations have received less attention. Because a substantial part of subjects' environment is probably normally distributed, it is important to extend the investigation of intuitive statistical inference to include other parameters of these distributions. Because distribution variances are central to the precision of predictions about future events, it is important to examine subjects' abilities to evaluate them. For these reasons the experiment to be reported investigated intuitive inferences about the variances of normally distributed populations.

THE NORMATIVE MODEL

The experimental task required intuitive inferences about which of two populations had the larger variance. This is distinctly different from the question whether or not the two population variances merely differ. Therefore, the classical null-hypothesis test is inappropriate and the normative statistical model is provided by Bayesian statistics (Edwards *et. al.*, 1963). The Bayesian procedure focuses upon the probability that either of two populations has a larger variance than the other given samples from both.[2] In the face of no knowledge about which variance is larger the Bayesian statistician may begin with a rectangular prior probability distribution over the hypotheses that each of the two populations has the largest vari-

[2] A Bayesian procedure for inferring which of two populations has the larger variance is given in the Appendix.

ance (i. e., $\sigma_1^2 > \sigma_2^2 < \sigma_2^2$). Then he observes the samples, and computes the ratio of the sample variances. Next, using the F distribution, he revises the prior probabilities to obtain a posterior distribution. The posterior-probability distribution favors the population that yielded the larger sample variance and the statistician infers that this population has the larger variance. His confidence in the accuracy of this inference depends upon the magnitude of the posterior probability for the associated hypothesis. The larger the posterior probability the smaller the risk of error in inferring that the selected population has the larger variance. Subsequent statements about the population variances and decisions related to them are then tempered in light of the magnitudes of the posterior probabilities. Comparisons of intuitive and Bayesian inference processes have yielded two major findings (Peterson and Beach, 1967). The first, called consistency, is that the procedures by which subjective prior probabilities are revised in light of data generally conform to the Bayesian logic and procedures (Beach, 1966; Peterson *et. al.*, 1965; Phillips and Edwards, 1966). The second, called conservatism, is that even though subjects follow the appropriate procedures, they tend to revise their subjective probability distributions less than the amount prescribed by the normative Bayesian model (Peterson and Miller, 1965; Phillips and Edwards, 1966; Phillips *et. al.*, 1966). That is, the data influence the subjective probabilities less than they do the Bayesian probabilities.

An important source of at least part of conservatism is subjects' misunderstanding of the implications of observed data for the possible inference. In more formal terms, subjects rely upon inaccurate subjective sampling distributions. This is demonstrated by the finding that if values from subjects' inaccurate estimates of sampling distributions are substituted for the veridical values in Bayesian equations it is possible to predict the suboptimal posterior subjective probability distributions (Peterson *et. al.*, 1967; Wheeler and Beach, 1967). Thus it appears that the suboptimality of inferences can have its roots in subjects' mistaken opinions about samples even though, within the constraints of these opinions, the intuitive and Bayesian inference processes may be quite similar.

The results of the Peterson *et. al.*, (1967) and the Wheeler and Beach (1967) experiments relating conservatism to inaccurate subjective sampling distributions illustrate the importance of taking subjects' opinions into account when investigating intuitive inference. When the samples are from normal distributions, and are not so easily summarized as binomial samples are, there are two kinds of erroneous opinions that can influence subsequent inferences. The first is inaccurate evaluation of the sample data, e. g., erroneous judgments of sample means and variances. The second is reliance upon inaccurate subjective sampling distributions. Investigations of the first of these show that intuitive judgments of sample means are fairly accurate (Beach and Swensson, 1966; Spencer, 1961, 1963), but that intuitive judgments of sample variances are influenced by the magnitude of the data in the sample; in general, judgments of sample variance decrease as means increase (Hofstatter, 1939; Lathrop, 1967). Moreover, some of Hofstatter's (1939) subsidiary findings suggest that in contrast to statistical practice, subjects may weight large deviations more or less heavily than they do small ones. This leads to their variance judgments being correspondingly larger or smaller than the objective variances.

There have been no investigations of the accuracy or inaccuracy of subjective-sampling distributuions for the parameters of normally distributed populations.

Because subjects' judgments of sample variance are inaccurate, the present experiment had two phases. The first examined the systematic inaccuracies of judgments of sample variances in light of the experiments just described, emphasizing the effects of data magnitudes (sample means) and how subjects weighted deviations. The second phase examined the correspondence between subjects' inferences and Bayesian inferences, taking the inaccurate judgments of the sample variances into account.

METHOD

Experimental Design

The subjects saw samples of equal size from two different populations and made inferences about which population had the larger variance. Then they either estimated the ratio of the population variances or stated their posterior subjective probabilities for the inferred population in the form of confidence in the inference.

A Bayesian statistician's posterior probabilities would favor the population that yielded the larger sample variance and that population would be the statistician's inference. The ratio of the sample variances is the best estimate of the ratio of the population variances and would be the statistician's ratio estimate. The statistician's confidence in his inference would be the posterior probability associated with the inferred population.

On the assumption that subjects behave as the statistician does, the ratio estimates for population variances were assumed equal to subjects' opinions about the ratios of the sample variances and the confidence estimates were assumed equal to their subjective posterior probabilities for the inferred population. Then the ratio estimates were used to infer subjects' judgments of each of the sample variances. The difference between these inferred judgments and the objective sample variances were accounted for in terms of how subjects treated deviations in making their judgments.

The effects of sample means on intuitive judgments of sample variance were investigated by examining changes in subjects' agreement about inferences and by examining changes in their confidence as a function of different sample means, holding variances constant. Then, subjects' inferences and confidence were compared to the inferences and confidence a Bayesian statistician would report were he to observe variance ratios equal to those estimated by the subjects. This procedure provides information about how closely subjective inferences about variances conform to the normative Bayesian inferences, taking subjects' erroneous judgments of the sample variances into account.

Apparatus

The stimuli ostensibly were random samples of $n = 20$ index cards from decks of $N = 50$ cards that each had a two-digit number written on them. However, to control the samples that subjects saw, each deck actually consisted of the 20 sample cards on top of 30 blank cards. The first four samples each had a mean of 37, equal ranges, and approximately normal distributions. Each of 10 numbers in each sample occurred twice. Adding constants of 16 and 32 to all of the numbers in the samples produced two more sets of four decks with means of 53 and 69 and with variances equal to those in the first four decks. The variances were (1) 82.31, (2) 123.37, (3) 200.42, and (4) 250.95. Their ratios were $R_{4:3} = 1.25$, $R_{2:1} = 1.50$, $R_{3:2} = 1.62$, $R_{4:2} = 2.04$, $R_{3:1} = 2.43$, and $R_{4:1} = 3.05$.

Confidence responses were made on 100-point scales that were divided in the center with the two ends labeled Left and Right. The middle was labeled .50/.50 and the scales increased to 1.00, in units of .05, in both directions. For ratio estimates, the response sheets merely had two columns of blanks labeled Left and Right and subjects wrote numbers in the blanks.

Procedure

Inferences and confidence. The task was described to the subjects as one of making intuitive decisions about the variability of sets of numbers on the basis of samples from the sets. They would see the first 20 cards from each of two shuffled decks of 50 cards and then they were to make their inference about whether the right or left deck had the greater variability by marking the right or left end of the confidence scale. Their confidence that the inference would prove to be correct if they saw all 50 cards was to be indicated by the location of the mark upon the scale: .50/.50 if they were completely unsure about which deck was most variable, 1.00 if they were completely confident about the accuracy of the decision, and intermediate scale values corresponding to intermediate degrees of confidence. They were not informed of the correct answers. No written calculations were permitted.

Variance was explained as the degree to which numbers cluster about their mean and examples were given of distributions with different variances.

The subjects saw all six combinations of the four variances within all nine combinations of the three means in the high-variance deck and the low-variance deck, a total of 54 judgments. The decks in each pair were randomly presented; cards from each deck were simultaneously displayed for about 4 seconds, turned face down, the next pair presented, etc., on through the two samples. Then subjects had 6 to 8 seconds to make their responses. The samples were shuffled behind a screen between presentations.

Ratio estimates. Three weeks after the first experimental session about one-half (18) of the subjects returned for the second session. The procedure was essentially the same as in the first session except that instead of making confidence statements, they estimated ratios of the larger population variances to the smaller. Responses were of the form $1:x$ or $x:1$, where 1 represented the smaller variance, x was some number equal to or greater than 1, and the positions of the numbers represented whether the right or left deck had the larger variance.

To avoid complicating the ratio estimates by having different means for the pairs of samples, only the decks with means of 53 were used. A fifth deck with a mean of 53 and a variance of 273.47 was added to the original four to increase the number of comparisons. Each of the five decks was paired with the other four in both left and right positions for a total of 20 ratio estimates. The cards were shuffled between presentations. The subjects were not permitted to make written calculations.

Response scoring. Group posterior subjective probabilities for the inferences were obtained by calculating median confidence for that deck in each pair which the mean ratio estimates indicated to have the larger judged variance. For all pairs of decks this turned out to be the same inference a statistician would make. Therefore, the degree to which the group posterior probabilities were optimal could be obtained by comparing median confidence with the Bayesian posterior probabilities. Moreover, the majority of subjects always made the optimal inference prior to estimating ratios or stating confidence. This permitted computation of the percentage of the subjects who agreed that the normative (optimal) inference was the subjectively best choice. This percentage summarizes both the unanimity among subjects about the best inference and the degree of inferential optimality for the group as a whole. As an example of scoring, if a subject inferred that the left deck had the larger variance and was .60 confident, but the right deck was the optimal inference, the response was scored as one suboptimal inference and confidence of .40 in the optimal inference. Arbitrarily one-half of the .50/.50 responses were scored as optimal inferences and the other one-half as suboptimal. Ratio estimates were scored in the same way that confidence was scored and means were taken in logs.

Subjects. Forty-four university students, both males and females served as subjects. None of the subjects were at all statistically sophisticated; about one-fourth had heard of variance in a brief discussion in introductory psychology courses.

RESULTS

Weighting of Deviations

The mean ratio estimates showed that while the objectively larger sample variance in each pair was always judged as being the largest, the ratios did not correspond to the objective ratios. Using the Baker and Dudek (1955) method for ratio estimates, a scale of judged variance was constructed in which the smallest judged variance was the unit and the other judged variances were multiples of the unit. This scale is shown in Fig. 1.

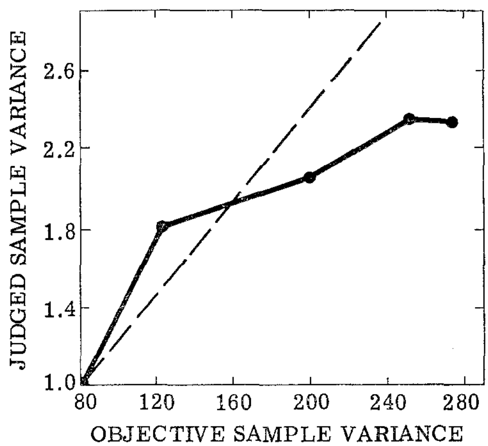

Fig. 1. The relation between judged and objective sample variances.

Hofstatter's (1939) data suggest that the discrepancy between judged and objective variance results from differences in how subjects and statisticians weight deviations. On the basis of this, the relation in Fig. 1 can be made nearly proportional ($r = .966$, slope $= 1.028$) by assuming that the equation for intuitive judgment of variance is a modified form of the usual variance equation;

$$\text{Judged Variance} = \sum_{i=1}^{n} |\bar{X} - \bar{X}_i|^{.39} \Big/ n$$

The effect of weighting deviations by raising them to a power of less than 1.0 is the opposite of squaring; it emphasizes the small deviations relative to the large ones and results in smaller judgments of variance than the statistician would give.

Effects of Sample Means

Although subjective judgments of the sample variances are inaccurate, the judged magnitudes increased as a function of the objective variances. This relation permits examination of the effects of sample means on the judgments. In general, as the ratio of the two objective sample variances increases, i.e., becomes more discriminable, the percentage of subjects who agree in giving the optimal inference should increase, and the group's confidence in the inferences also should increase. As a result, any variable that systematically increases or decreases the judged sizes of the variances should influence both the percent agreement among subjects and the group's confidence. Therefore, the influence of sample means upon judgments of any two samples' variances should be reflected in increases or decreases in percent agreement and in group confidence as a function of various combinations of the two samples' means.

The reasoning is as follows: Hofstatter's (1939) and Lathrop's (1967) results show that judgments of sample variances decrease as the associated mean increases. This implies that an increase in the mean associated with the large sample variance, i.e., the numerator of the F-ratio, should reduce the judged size of that variance and lead to a corresponding decrease in the judged size of the ratio. Therefore the ratio should have a smaller effect on subjects' rectangular prior subjective probabilities and should lead to smaller posterior probabilities for the favored population. Because of this, percent agreement should be low and group confidence should be small.

By the same token, increases in the size of the sample mean associated with the smaller sample variance, i.e., the denominator of the F-ratio should decrease the judged size of of the smaller variance still further, and thereby increase the judged size of the ratio. This larger ratio should have a large effect upon the prior probabilities and result in large posterior subjective probabilities for the favored deck. Therefore, percent agreement should be high and group confidence should be large.

Table 1 contains percent agreement and group confidence, computed across all combinations of the sample variances, for each combination of the sample means. Reading across the rows or down the columns of the

TABLE 1
PERCENT AGREEMENT AMONG SUBJECTS' INFERENCES AND MEDIAN CONFIDENCE FOR EACH COMBINATION OF SAMPLE MEANS

Mean of low-variance sample	Mean of high-variance sample			
	37	53	69	Over-all
37	75% .65	71% .65	64% .60	70% .65
53	73% .70	82% .70	66% .64	74% .70
69	82% .75	69% .75	74% .60	75% .70
Over-all	77% .70	74% .70	68% .60	

table shows that, in general, the sample means had the expected effects on agreement and confidence. As the means associated with the variances in numerators of the ratios increased, both percent agreement and group confidence tended to decrease. As the means associated with the variances in the denominators increased, both percent agreement and group confidence tended to increase. Appropriately, the two most extreme cases are the upper-right cell in the table and the lower-left cell. In the first, the small mean does not change the judgment of the denominator variances very much, so they stay relatively large while the large mean reduces the numerator variances a good deal. The result is a small-judged ratio, less agreement, and lower group confidence. The opposite conditions hold for the lower-left cell, and the agreement and confidence are high.

The diagonal cells in the table show that the effects of the means do not cancel for ratios in which both samples have the same mean. The 53-53 combination yields both higher percent agreement and higher group confidence than do the 37-37 and 69-69 combinations. Examination of responses for individual pairs of variances revealed that this discrepancy held consistently throughout the data, but provided no obvious explanation of why it did so.

INFERENCES ABOUT POPULATION VARIANCES

If the inaccurate ratio estimates represent subjects' judgments of the ratios of the sample variances, and if these ratios served as the basis for the revisions of the subjective prior probabilities, then percent agreement for optimal inferences and the group's median confidence should be systematically related to the estimates. Figure 2 shows the relation. The percent of optimal inferences made in both the first (confidence estimation) and the second (ratio estimation) experimental sessions are listed at the top of the figure for each ratio of sample variances. Group confidence is plotted against objective ratios and against mean estimated ratios.[3] The brackets around the data points indicate the range of the middle 50% of the response distributions. The smooth function represents the Bayesian posterior probabilities for the corresponding ratios of variances for samples of $n = 20$; these probabilities derive from Equation 6 in the Appendix.

On the left side of Fig. 2, neither the percent of optimal inferences nor

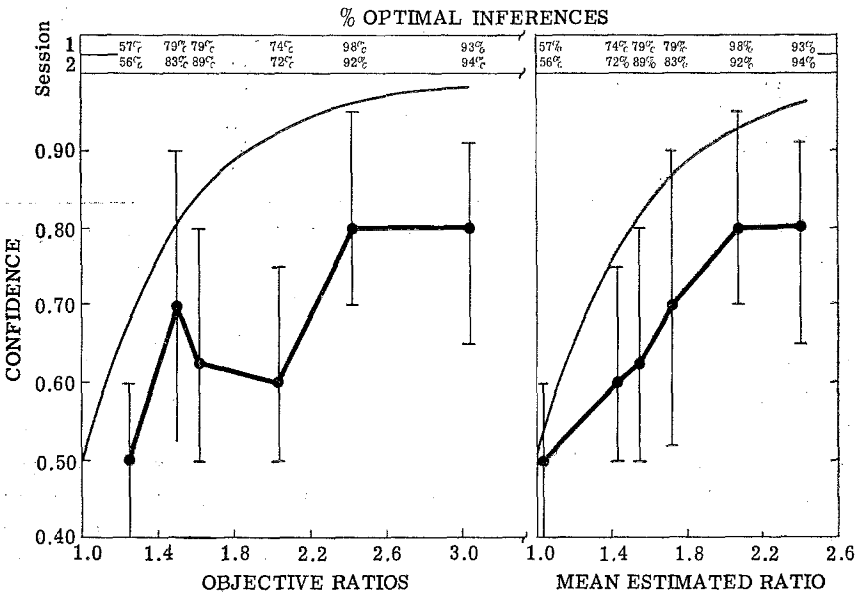

FIG. 2. Median confidence, percent agreement on optimal inference, and Bayesian posterior probabilities as functions of objective and estimated ratios of the sample variances.

group confidence is systematically related to the objective ratios of the sample variances, and confidence looks nothing like the smooth Bayesian function. On the right, however, the relations are systematic and both the percent of optimal inferences and group confidence increase as the mean estimated ratios increase.

The Bayesian curve in the right graph represents the posterior probabilities that a statistician would obtain from Equation 6 in the Appendix if the samples he observed had the variance ratios that subjects estimated.

[3] The ratio estimates were made for decks with means of 53; the other means have been shown to result in different judgments of the variances (Table 1). Therefore the data in Fig. 2 for percent of optimal inferences and for confidence derive only from the pairs of samples in the first experimental session that both had means of 53. Corresponding curves for the other two equal mean conditions (37–37 and 69–69) are parallel and somewhat lower than the curves in Fig. 1.

The group confidence curve is roughly the same shape as the Bayesian curve, but it is a good deal lower. As in previous investigations of intuitive statistical inference, the sample data, in this case subjects' incorrect judgments of the sample variances, resulted in less revision of the rectangular subjective prior probabilities than they would have for Bayesian probabilities. The degree of difference in amount of revision for each data point can be quantified by computing the ratio of $\log c{:}1 - c$ to $\log p{:}1-p$, where c is the group confidence for the optimal inference and p is the Bayesian posterior probability for that inference. This ratio is formally the same as the Accuracy Ratio used in previous Bayesian studies of intuitive statistics. A ratio of 1.0 indicates that subjects' revision of their probabilities was equal to the Bayesian revision and, as a result, that the group confidence and the Bayesian posterior probability were equal. A ratio of less than 1.0 indicates the degree to which the median revision was conservative relative to the Bayesian revision; i.e., it indicates the proportion of Bayesian revision represented by the median revision. Because the ratio for $c = .50$ is undefined, only the remaining five points on the right side of Fig. 2 are reported. From bottom to top they are .34, .32, .46, .53, and .44; values which correspond to the range reported in previous experiments (e.g., Peterson and Miller, 1965).

DISCUSSION

These results show that while subjects' judgments of sample variances are inaccurate and their posterior subjective probabilities are conservative, their inferences are based on the ratios of their judgments of the sample variances; both the percent of optimal inferences and group confidence show a clear relation to estimates of the ratios. For group confidence this relation is similar in form to the Bayesian posterior probability function that a statistician would obtain from observation of sample variance ratios equal to those that the subjects reported.

Intuitive judgments of sample variance, and hence of ratios, appear to be influenced both by the tendency to weight small deviations more heavily than large ones and by the magnitudes of the sample numbers, as reflected by the sample means. Care is necessary in interpreting the weighting results. The exponent of .39 may be determined more by the experimental conditions than by any general weighting practice used by subjects. With normally distributed samples, the majority of deviations are small. And, because the ranges were held constant across all samples that had the same mean, the most extreme data were not informative about sample variance. These two characteristics of the stimuli may have encouraged subjects to rely upon the small deviations to make their judgments.

While the possible dependency of weighting upon the specific experimental stimuli was not crucial to this experiment, the results point up the need to complement the Hofstatter (1939) and Lathrop (1967) investigations of the effects of sample means on judgments of sample variances with a detailed examination of the conditions that influence weighting of deviations. Squaring of deviations is a convention adopted by statisticians for its mathematical convenience and there is no reason why subjects follow it; the conventions they follow may be systematic and more reasonable in light of their opinions about what variance is and what role it fills in the prediction of future events.

The results showing the influence of sample means on percent agreement and confidence (Table 1) are congruent with those reported by Hofstatter (1939) and Lathrop (1967). This was so in spite of the fact that confidence was a fairly insensitive dependent variable. That is, confidence was conservative, subjects tended to make their marks on the .05-unit boundaries of the confidence scale, and both inferences and confidence were based on subjectively computed ratios of variances from two simultaneously presented samples. Any of these factors might have been sufficient to prevent the effects of means from being observed. As Lathrop (1967) has shown, however, the effect of means is robust and quite compelling; even sophistication about the statistical concept of variance does little to overcome it.

The suboptimality of group confidence that remains after the inaccuracies of judged sample variance are taken into account, is similar in form and amount to that obtained in previous Bayesian investigations of intuitive inference. It is not unlikely that this conservatism in revising subjective probabilities results from the second source of error suggested in the introduction; the subjects' F-distributions may be inaccurate. In other words, in addition to errors in judging the sample variances that go into the ratios, the subjects may incorrectly evaluate the implications of their inaccurate ratios for the revision of their subjective prior probabilities. If the results reported by Peterson *et. al.* (1967) can be generalized to this experiment, it might be possible to account for the conservatism in the right half of Fig. 2 by ascertaining the subjective F-distributuins. This experiment has accounted for some of the variance in intuitive inferences about population variance in terms of errors in judging sample variances. The investigation of the role of inaccurate subjective F-distributions remains to be done.

APPENDIX[4]

The task is one of inferring which of two normal populations has the greater variance given two equal-sized samples from each of the populations.

[4] This section and the Bayesian test in it are by Dr. W. M. Kincaid, Department of Mathematics, University of Michigan.

Let the sample variances be s_1^2 and s_2^2 and the corresponding population variances be σ_1^2 and σ_2^2. For given σ_1^2,σ_2^2 the ratio $s_1^2/\sigma_1^2(s_2^2/\sigma_2^2)^{-1}$ has an F-distribution with $n - 1$ *df* in both numerator and denominator. This fact is the basis for the usual statistical test for the equality of variances.

With a Bayesian approach, the problem of interest is the reverse one of determining the (subjective) distribution of σ_1^2/σ_2^2, given an observed value of s_1^2/s_2^2 and prior distribution of σ_1^2/σ_2^2. While this approach seems natural, it should be pointed out that it is somewhat incomplete from a strictly Bayesian point of view. For ideally there is a joint distribution of four parameters, namely the population means and variances, from which the distribution of σ_1^2/σ_2^2 is derived. A discussion of the problem in these terms would be very laborious at best. These considerations are relevant, however only to the extent that two different sets of data having the same value of s_1^2/s_2^2 could lead to different posterior distributions of σ_1^2/σ_2^2. The interest here is only in the situation where such differences would be negligible, that is, in the case of a very flat prior distribution. Under these circumstances, it seems reasonable to avoid the difficulty by making the simplifying assumption that the ratio s_1^2/s_2^2 is the only information available.

At this point it is convenient to introduce the ratios

$$\tau = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \text{ and } \omega = \frac{s_1^2}{s_1^2 + s_2^2}$$

These ratios are limited to the range [0,1], and they are effectively symmetric in the two variances, since

$$1 - \tau = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

The task can now be restated as one of determining the posterior distribution of τ given ω.

A prior distribution of τ is needed; making the simplest choice, suppose that it is uniform on [0, 1]. The posterior density of $f(\tau|\omega)$ is then proportional to the conditional density $f(\omega|\tau)$.

It was noted above that $s_1^2\sigma_2^2/s_2^2\sigma_1^2$ has an F-distribution for given σ_1^2,σ_2^2; that is, the conditional density $f(x|\tau)$ has the form

$$f(x|\tau) = Cx^{(n-3)/2}(1 - x)^{-(n-1)}, \qquad x > 0 \tag{1}$$

where $x = s_1^2\sigma_2^2/s_2^2\sigma_1^2$ and C is a constant. The density $f(\omega|\tau)$ can be derived from (1) and the relation

$$\omega = \frac{\tau x}{1 - \tau + \tau x}; \tag{2}$$

it is the form

$$f(\omega|\tau) = C' \frac{\tau^{(n-1)/2}(1 - \tau)^{(n-1)/2}\omega^{(n-3)/2}1 - \omega^{(n-3)/2}}{(\tau + \omega - 2\tau\omega)^{n-1}}. \tag{3}$$

Consequently, $f(\tau|\omega)$ is of the form (3) also, or, for any fixed ω,

$$f(\tau|\omega) = \frac{k\tau^{(n-1)/2}(1-\tau)^{(n-1)/2}}{(\tau+\omega-2\tau\omega)^{n-1}} \tag{4}$$

where k can be determined from the condition

$$\int_0^1 f(\tau|\omega)\,d\tau = 1 \tag{5}$$

In our case we are interested in the probability $P(\sigma_1^2 < \sigma_2^2) = P(\tau < \frac{1}{2})$ which in view of (4) and (5) is given by

$$\int_0^{1/2} \frac{\tau^{(n-1)/2}(1-\tau)^{(n-1)/2}}{(\tau+\omega-2\tau\omega)^{n-1}}\,d\tau \times \left[\int_0^1 \frac{\tau^{(n-1)/2}(1-\tau)^{(n-1)/2}}{(\tau+\omega-2\tau\omega)^{n-1}}\,d\tau\right]^{-1} \tag{6}$$

The integral (6) is most readily evaluated numerically.

REFERENCES

BAKER, K. E., AND DUDEK, F. J. Weight scales from ratio judgments and comparisons of existent weight scales. *Journal of Experimental Psychology,* 1955, **50,** 293–308.

BEACH, L. R. Accuracy and consistency in the revision of subjective probabilities. *I.E.E.E. Transactions on Human Factors in Electronics,* 1966, HFE-7, 29–37.

BEACH, L. R., AND SCOOP, T. S. Inferences about correlations. *Psychonomic Science,* 1966, **6,** 253–254.

BEACH, L. R., AND SWENSSON, R. G. Intuitive estimation of means. *Psychonomic Science,* 1966, **5,** 161–162.

BRUNSWIK, E. *Perception and the representative design of experiments.* Berkeley: University of California Press, 1956.

EDWARDS, W. Human processing of equivocal information. Technical Documentary Report No. ESD-TRD-64-601, Air Force Systems Command, Hanscom Field, Bedford, Mass., 1965.

EDWARDS, W., LINDMAN, H., AND SAVAGE, L. J. Bayesian statistical inference for psychological research. *Psychological Review,* 1963, **70,** 193–242.

HOFSTATTER, P. R. Ueber die schatzung von gruppeneigenschaften. *Zeitschrift fur Psychologie,* 1939, **145,** 1–44.

IRWIN, F. W., SMITH, W. A. S., AND MAYFIELD, J. F. Tests of two theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology,* 1956, **51,** 261–268.

LATHROP, R. G. Perceived variability. *Journal of Experimental Psychology,* 1967, **73,** 498–502.

LITTLE, K. B., AND LINTZ, L. M. Information and certainty. *Journal of Experimental Psychology,* 1965, **70,** 428–432.

PETERSON, C. R., AND BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin,* 1967, **68,** 29–46.

PETERSON, C. R., DUCHARME, W. M., AND EDWARDS, W. Sampling distributions and probability revisions. *Journal of Experimental Psychology,* in press.

PETERSON, C. R., AND MILLER, A. J. Sensitivity of subjective probability revision. *Journal of Experimental Psychology,* 1965, **70,** 117–121.

PETERSON, C. R., AND PHILLIPS, L. D. Revision of continuous subjective probability distributions. *I.E.E.E. Transactions on Human Factors in Electronics,* 1966, HFE-7, 19–22.

PETERSON, C. R., ULEHLA, Z. J., MILLER, A. J., BOURNE, L. E., AND STILSON, D. W. Internal consistency of subjective probabilities. *Journal of Experimental Psychology,* 1965, **70**, 526–533.

PHILLIPS, L. D., AND EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology,* 1966, **72**, 346–354.

PHILLIPS, L. D., HAYS, W. L., AND EDWARDS, W. Conservatism in complex probabilistic inference. *I.E.E.E. Transactions on Human Factors in Electronics,* 1966, HFE-7, 7–18.

SPENCER, J. Estimating averages. *Ergonomics,* 1961, **4**, 317–328.

SPENCER, J. A further study of estimating averages. *Ergonomics,* 1963, **6**, 255–265.

WHEELER, G., AND BEACH, L. R. Subjective sampling distributions and conservatism. *Organizational Behavior and Human Performance,* 1968, **3**, 36–46.

RECEIVED: June 29, 1967