ICAS Paper No. 68-17

THE STATISTICAL TREATMENT OF PILOT-OPINIONS ON FLYING QUALITIES

by

Jozsef Gedeon
Technical University
Chair for Aeronautics
Budapest, Hungary

The Sixth Congress of the International Council of the Aeronautical Sciences

DEUTSCHES MUSEUM, MÜNCHEN, GERMANY/SEPTEMBER 9-13, 1968

Preis: DM 2.00

STATISTICAL TREATMENT OF PILOT-OPINIONS ON FLYING QUALITIES

J. Gedeon
Technical University
Budapest, Hungary

Abstract

A public opinion research interviewing numerous experienced pilots and employing the Cooper numerical rating scale, has been conducted to check the reliability of pilot opinions on flying qualities. The distribution of the Cooper-ratings has been shown to converge with increasing number of votes on a binomial-like distribution, from wich follows the sequence of the means to be a regular sequence. Consequently, 90% confidence limits may be assigned to the mean ratings computed from a given number of votes, wich facilitates the design of more exact pilot opinion boundary charts.

Notation

j	number of votes
n	order of binomial distribution
p	parameter of binomial distribution
x	rating
$\bar{x}$	mean rating
A	constant
$B_{n,x}$	formula giving points of binomial distribution
D	deviation of binomial distribution
ϑ	error
ω_x	maximum steady roll rate at cruising speed

Subscripts

90	of 90% confidence

1. Introduction

Except in the worst cases of stability problems and spinning no definite criteria relating to the rating of handling characteristics can be given by theoretical or experimental flight mechanics. It seems that from the various methods proposed so far for the better understanding of what makes an airplane easy to fly and for establishing handling criteria it is only the systematic analysis of pilots assessments aided by simulation techniques and servo-analysis methods wich is sufficiently universal and has stood the proof of time. /See e.g.[1,2,3,4]/

For more uniformity and better definiton in the pilot-opinions considerable use is made of variants of the Cooper rating scale[5]. Nevertheless, pilot ratings more or less differ - if they were given independently and not after previous discussion /e.g. when working in a prototype jury/. The resulting scatter may be reduced by interviewing several pilots and computing the mean of their ratings, but nevertheless, some uncertainity remains. This fact presents a good opportunity for some opponents to deny - if not openly, tacitly - the real value of pilot opninions. On the contrary some pilots, not without local authority, are regarding their private opinion as the only possible correct one and are insisting on everybody agreing in this.

The purpose of the investigation reported herein is to make a thorough examination of the statistical rules of pilots ratings. Thereby pilot opinion boundary charts are hoped to be improved upon and individual disbeliefs or claims to absolute authority can be better tackled too.

2. Pilot-Opinion Poll and Preliminary Appraisal

Material for the investigation was taken from a pilot-opinion poll conducted among experienced sailplane pilots. Voting was on questionaries containing 31 questions on individual characteristics and 6 questions on general assesment of the type concerned. For rating a variant of the Cooper scale was employed, with 1 point assigned to the worst and lo points to the best classification.

In all, 532 questionaries rating 35 types were received. From this, 5 types assessed by thegreatest number of pilots /from 38 up to 7o each/ were selected for statistical examination.

Sampling from the ratings for individual characteristics showed that the actual distributions would most probably be of a shape similar to the binomial distribution.

According to our rating system the formula for binomial distributions of the order $n = 9$ with parameter p

should read

$$B_{n,x}=\binom{n}{x-1}p^{x-1}(1-p)^{n-x+1}=\binom{9}{x-1}p^{x-1}(1-p)^{10-x} \qquad (1)$$

The relationship between the parameter p and the mean of the binomial distribution is:

$$\bar{x}=1-np=1-9p \qquad (2)$$

It was thought to be the most convenient to select coordinate scales so that the cumulative distribution points of the ratings lay on a straight line and to make a combined graphical-numerical examination. No coordinate system to exactly meet this requirement for all binomial distributions is known to the author, but for practical purposes the Weibull probability coordinates will do quite well, giving even in the worst cases an approximation in the order of a few tenth per cent. /For details of Weibull distributions see In Figure 1 circles representing points of the binomial distributions of parameter p = 0.1, 0.2 ... 0.9 and the corresponding regression lines are shown.

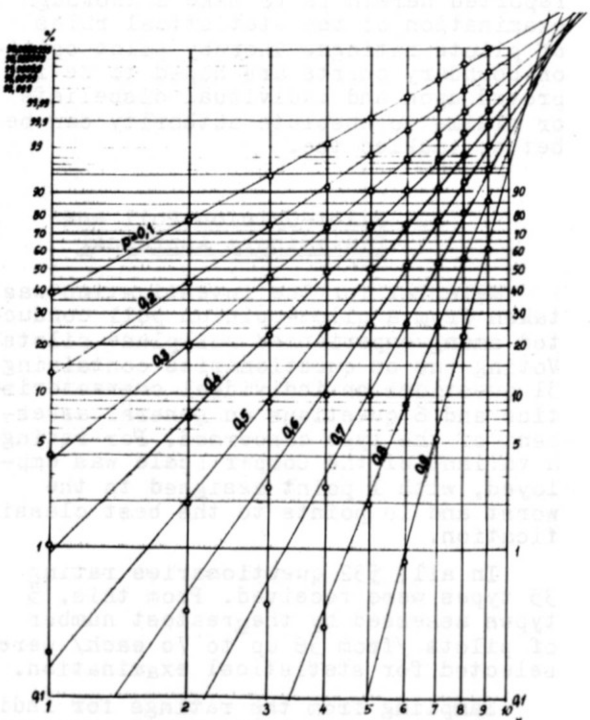

Figure 1. Approximation of binomial distributions with sraights in Weibull coordinates.

The only problem with this type of graphs is that the point for the rating 10 /100%/ is in infinity. The correct percentage of the ratings 10 should be controlled therefore by other means. Fortunately, there is a unique relationship between the gradient of regression-lines and the means of the corresponding distributions, which will be made recourse of later.

3. Rating Distribution

3.1 Classes of Rating Distributions

Traced on Weibull coordinates the distributions can be classified as:

Category 1.: all points lying on a straight line /Figure 2/. If the gradient of the straight of regression is the same as for a binomial distribution of equal mean, this category may be classified therefore as a binomial distribution.

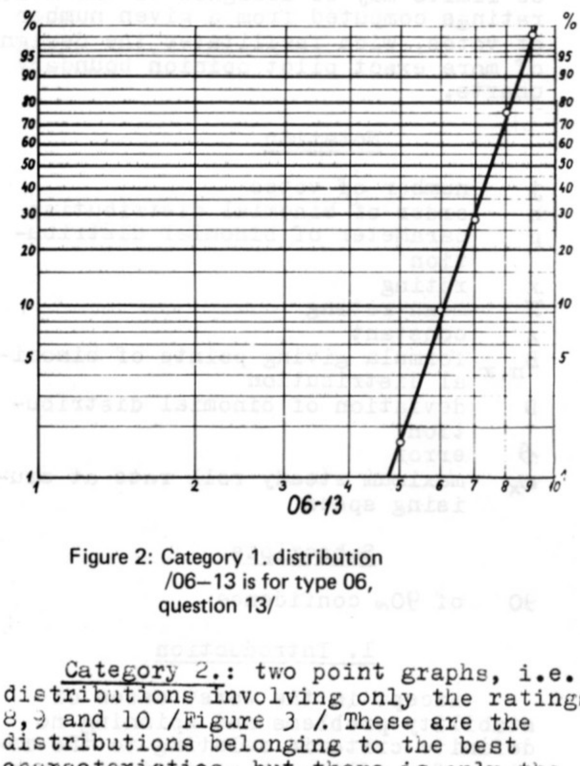

Figure 2: Category 1. distribution /06–13 is for type 06, question 13/

Category 2.: two point graphs, i.e. distributions involving only the ratings 8, 9 and 10 /Figure 3 /. These are the distributions belonging to the best characteristics, but there is only the relationship between gradient and mean to check their regularity.

Category 3.: distributions with some irregularity at the lowest rating given. As there were 38 to 70 votes on each case, the worst of the ratings got usually 1-2. In a fair number of cases some irregularity was to be expected therefore on this side of the distributions

/Figure 4 /. This type of irregularity is likely to be inherent in the size of the samples.

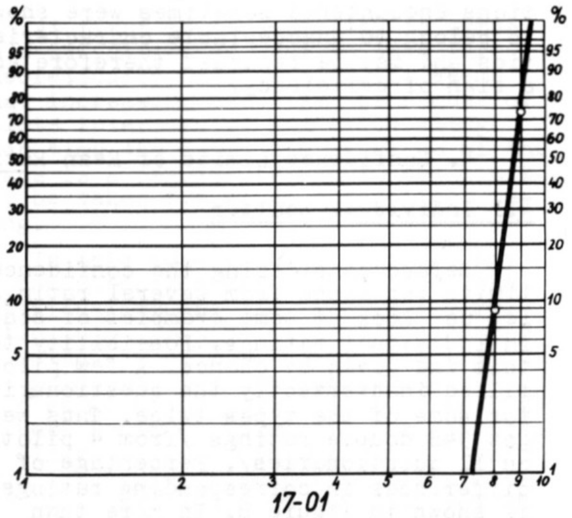

Figure 3. Category 2. distribution

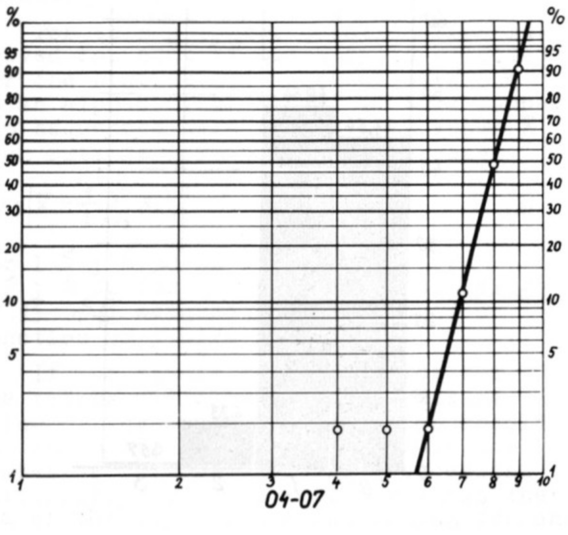

Figure 4. Category 3. distribution

Category 4.: distributions with double scatter. In some cases the points of the distributions could be fitted to two regression lines only /Figure 5 /. All such cases could be identified as belonging to bad characteristics. One possible explanation for this phenomenon is that some pilots are attaching less importance to the characteristic concerned while others are regarding it as essential.

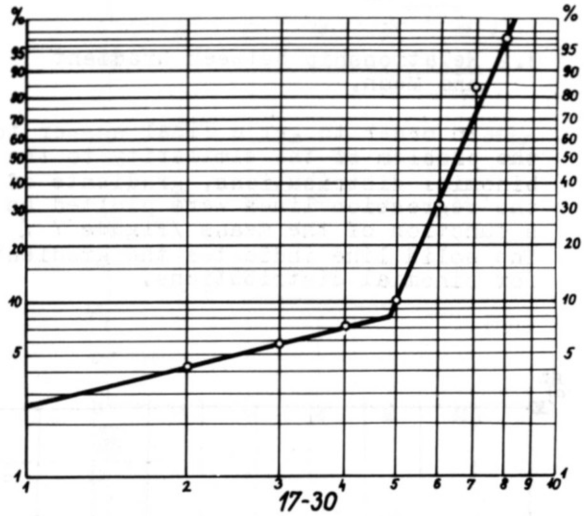

Figure 5. Category 4. distribution

Category 5.: irregular distributions. Some really bad characteristics showed scattered distribution points /Figure 6 /.

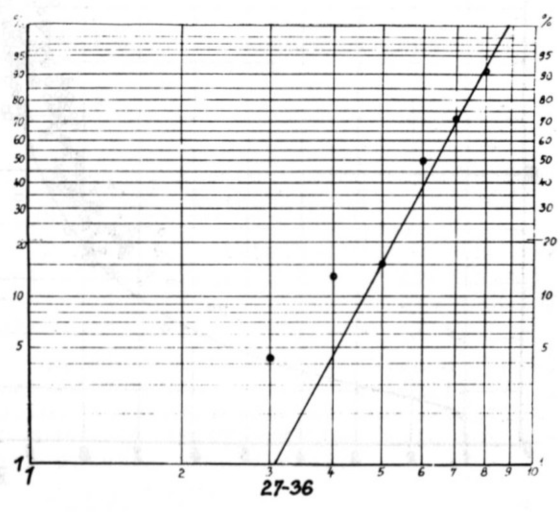

Figure 6. Category 5. distribution

Counting the percentage of these categories among the 175 distributions investigated gave the following results:

Category	1	2	3	4	5
%	34.28	15.43	38.86	8.00	3.43

Table 1: Percentage distribution of categories

It appers that perfect distributions or those with minor defects are accounting for 88.57% of the cases, while from the remaining only 3.43% may be classified as irregular.

3.2 Relationship between Gradient and Mean.

In order to get a final answer to the problem of the similarity to the binomial distributions, gradients of the regression lines were plotted as a function of the means /Figure 7 /. The solid line indicates the gradient for binomial distributions.

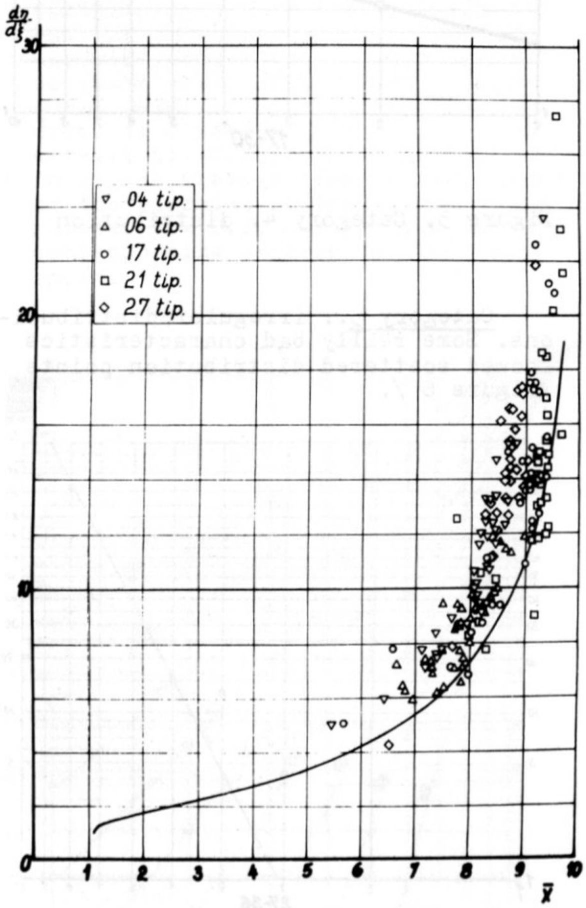

Figure 7. Gradient of the distributions as function of the means

Points are scattered most of them being above the line indicating that on the average there are less 10 point ratings than in equivalent binomial distributions.

On the basis of all these the distribution of the Cooper-ratings may be said usually to converge with increasing number of votes on a binomial - like distribution. Dual and irregular distributions encountered sometimes were shown to belong to unacceptable characteristics and may be regarded therefore as a sign of deficiency.

4. Confidence Limits of Mean Ratings

4.1 Individual Scatter

Before considering the confidence limits for means from several ratings, let us look at some examples of scatter in individual ratings. Possibility to this was given by chance. A few pilots filled inadvertently the questionaries for some of the types twice. Thus we got 348 double ratings /from 4 pilots on 11 questionaries/. Percentage of differences in corresponding ratings is shown in Figure 8. In more than 93% of double votes the difference of ratings did not exceed 1 point.

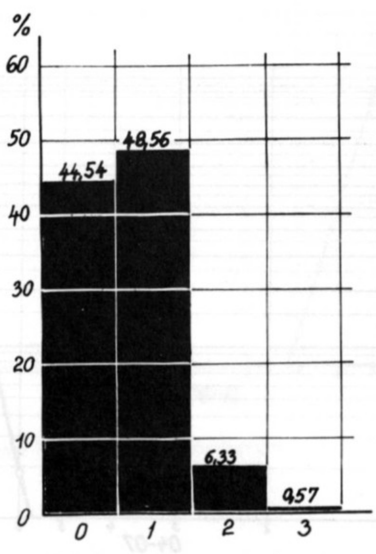

Figure 8. Scatter in individual ratings

Probable scatter for the mean ratings will be calculated by other methods, but this adds nevertheless something to our knowledge about the reliability of pilot-opinions.

4.2 Relation between Confidence Limits and Number of Votes

Practically the only possibility to

reduce, probable errors in mean ratings is to interview more pilots. It is therefore of prime importance to know exactly the degree of resulting improvement.

At first it is necessary to agree on a definition of what to take for theoretically exact rating. It seems that increasing stepvise the number of pilots interrogated and calculating each time the mean, the sequence of the means is a regular sequence, the limit may be taken for exact theoretical rating.

In this case probable errors may be computed from the deviation of the distributions and they are in inverse ratio to $\sqrt{j}$, square root of the number of votes. To be more exact, the $1/\sqrt{j}$ law holds exactly for deviations as counted from the actual limit /i.e.: if it were known in advance/. For confidence limits assigned to means computed from j ratings, a proportionality factor of $1/\sqrt{j-1}$ would be correct, but in practice this makes little difference.

At first, this law was checked by a Galton experiment of 2x1500 casts. After getting affirmative results here, the distributions of ratings were also investigated.

For this purpose, questionaries received for each of the 5 types selected for the statistical research were put in an alphabetical order. In this manner 175 sequences were formed with known $\bar{x}$ mean ratings. Forming the difference

$$\vartheta = \left|\bar{x} - \frac{1}{j}\sum_{k=1}^{j} x(k)\right| \qquad (3)$$

for $1 \leqq j \leqq 50$ on 3 types and for $1 \leqq j \leqq 30$ for 2 types and multiplying by $\sqrt{j}$ $\vartheta\sqrt{j}$ weighed differences as a function of j were formed.

If the square root law holds for the rating distributions, the average value of the weighed differences should be independent of j. More exactly the number of ratings yielding $\bar{x}$ exceeding but slightly 50 or 30 resp., towards the end of the sequences a decreasing tendency is expectable.

$\vartheta\sqrt{j}$ values were grouped by fives /j = 1-5, 6-10, etc./ and rating on all characteristics in each group of five were averaged. Trends in these mean weighed differences are to be seen in Figure 9.

Trends in weighed differences were roughly as expected, it seems therefore permissible to regard $\vartheta\sqrt{j}$ as a constant, independent of j.

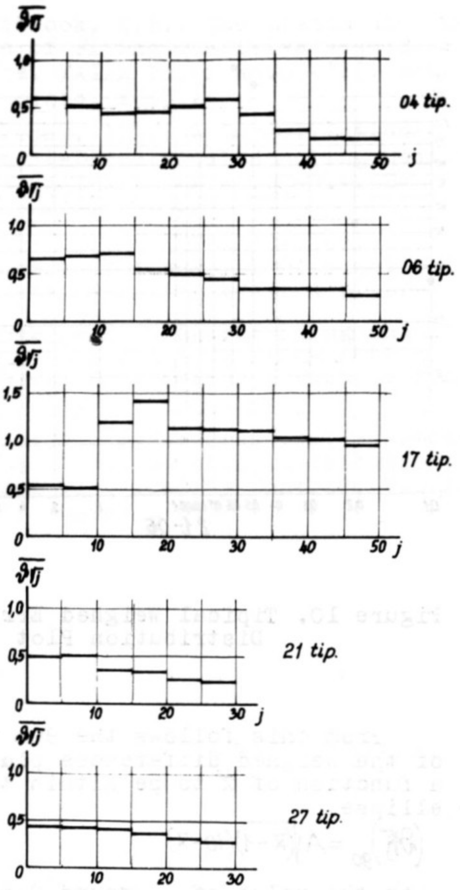

Figure 9. Trends in Weighed Differences

4.3 Check of the Deviation Formula for Binomial Distributions

Percentage distributions of the values of $\vartheta\sqrt{j}$ - calculated as previously described - were counted separately for each question, on each type /Figure 10 /.

This was intended to check a formula, derived for 90% confidence limits of binomial distributions. As the deviation of binomial distributions /with our rating points/ should read:

$$D = \frac{1}{3}\sqrt{(\bar{x}-1)(10-\bar{x})} \qquad (4)$$

90% confidence limits may be expressed with fair approximation:

$$\vartheta_{90} \cong \frac{0.548}{\sqrt{j}}\sqrt{(\bar{x}-1)(10-\bar{x})} \qquad (5)$$

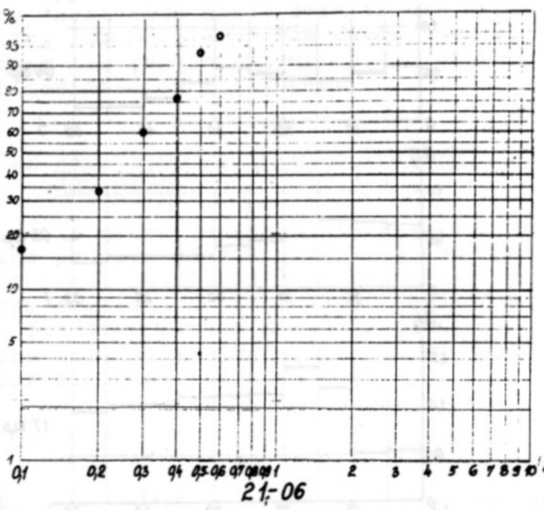

Figure 10. Tipical Weighed Error Distribution Plot

From this follows the 90% values of the weighed differences plotted as a function of $\overline{x}$ to be within the semi-ellipse

$$\left(\vartheta\sqrt{j}\right)_{90} = A\sqrt{(\overline{x}-1)(10-\overline{x})} \qquad (6)$$

with the value of A around 0.548.

This check was done in Figure 11, where values of $(\vartheta\sqrt{j})_{90}$ for all error distributions were plotted as a function of $\overline{x}$. Ellipse arcs are for A = 0.1, 0.2 ... 0.6.

All points except one are within the theoretical A = 0.548 value. Counting the points between the ellipses and calculating their percentage distribution as a function of A gives the fairly regular Weibull plot shown in Figure 12.

It may be concluded therefore that 90% confidence limits may be calculated as:

$$\vartheta_{90} = \frac{0.5}{\sqrt{j}}\sqrt{(\overline{x}-1)(10-\overline{x})} \qquad (7)$$

or, if the theoretical value of the constant is prefered:

$$\vartheta_{90} = \frac{0.548}{\sqrt{j}}\sqrt{(\overline{x}-1)(10-\overline{x})} \qquad (7a)$$

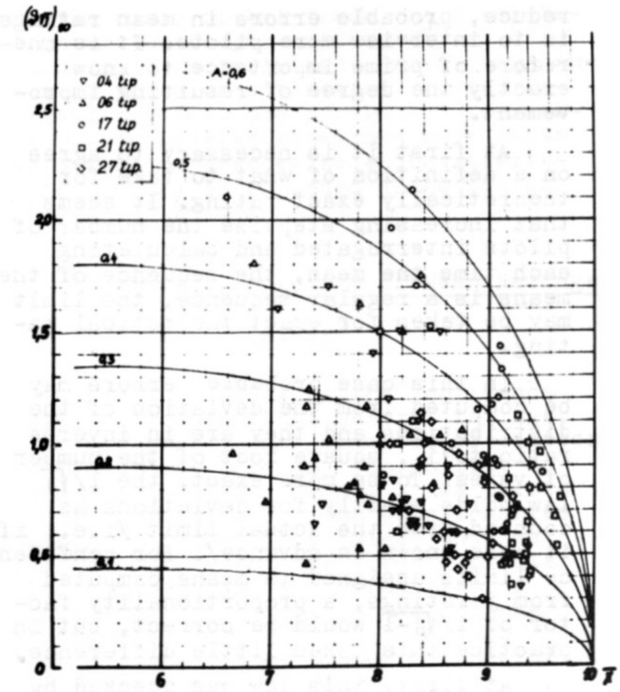

Figure 11. 90% Values of Weighed Error Distributions

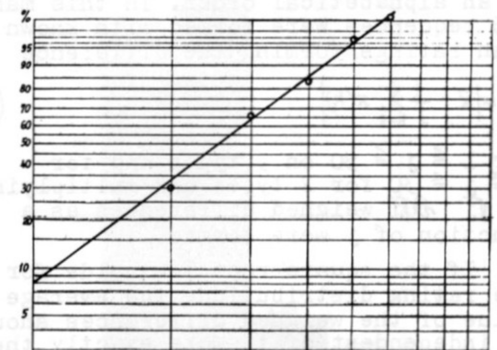

Figure 12. Percentage Distribution of 90% Confidence Points Within the Ellipse with Parameter A.

An example of how this works out in practice may be seen in Figure 13.

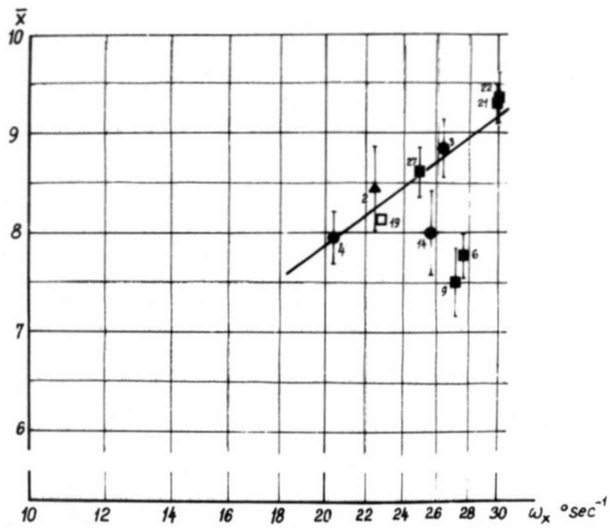

Figure 13. Trends in mean ratings for roll response of sailplanes as a function of steady roll rate.

Types 6, 9 and 14 were downgraded by the pilots because of insufficient aileron power on the ground and excessive aileron forces respectively.

5. Summary

The calculation of 90% confidence limits may be expected to facilitate considerably the sorting out of significant differences between the mean values of pilot ratings and the design of more exact pilot opinion boundary charts.

Furthermore, the statistical characteristics of the pilot opinion poll results indicate that:

a/ Even experienced pilots give frequently and regularly different ratings, hence individual opinions may be accepted unconditionnally in extreme cases only.

b/ As pilot ratings given independently and in sufficient number are showing fairly regular distribution, their means may be counted upen as giving a fair qualification on flying characteristics.

References

1. Seckel, E.: Stability and Control of Airplanes and Helicopters. /New York, London, 1964./

2. Westbrook, C.B.: The Status and Future of Flying Qualities Requirements. /AIAA Paper No.65-313. San Francisco, 1965./

3. Ashkenas, I.L.: A Consolidation of Lateral-Directional Handling Qualities. /AIAA Paper No.65-314. San Francisco, 1965./

4. O'Hara, F.: Handling Criteria. /Journal of the Royal Aeronautical Society. Vol.71., 1967. April./

5. Cooper, G.E.: Understanding and Interpreting Pilot Opinion. /Aeronautical Engineering Review, March, 1957./

6. Berettoni, J.N.: Practical Applications of the Weibull Distribution. /Industrial Quality Control, Vol.21. No.2; August 1964./