TECHNICAL NOTE

David A. Stoney,[1] *Ph.D. and John I. Thornton,*[2] *D.Crim.*

Multivariate Analysis of Typeface Damage Frequencies

REFERENCE: Stoney, D. A. and Thornton, J. I., **"Multivariate Analysis of Typeface Damage Frequencies,"** *Journal of Forensic Sciences,* JFSCA, Vol. 34, No. 3, May 1989, pp. 673-677.

ABSTRACT: To explore possible deterministic factors in typeface damage, multivariate analysis was performed on data developed by Dr. David Crown in his doctoral dissertation on typewriter individuality. Crown observed a slow increase in collective typeface damage with the duration of typewriter use, but found no apparent increase in damage correlated with individual letter use frequencies. These findings seemed peculiar and were the initial motivation for the present study.

For multivariate analysis, three groups of dependent variables were defined: those based on letter use frequencies; those based on letter form; and those based on basket or keyboard position. Both letter-use frequency and letter area showed statistically significant positive correlation with typeface damage frequencies. Together, these two variables accounted for 42% of the observed variation in damage frequencies. None of the other variables tested was able to explain any statistically significant portion of the remaining variation. These variables included use frequencies associated with difficult finger movements, vertical extension of letters, basket position, and keyboard row.

The 58% residual variation must be combined with additional variation introduced by the form and position of typeface damage. With this amount of variation left unexplained, the deterministic factors identified in this study are of little significance in the interpretation of typewriting comparisons.

KEYWORDS: questioned documents, typewriters, statistical analysis, typeface damage

To explore possible deterministic factors in typeface damage, multivariate analysis has been performed on data presented by Dr. David Crown in his doctoral dissertation on typewriter individuality. With the passing of type bar mechanisms into obsolescence, data of this type will not be easily generated, and a detailed analysis is therefore warranted.

In his doctoral dissertation [*1*], David Crown surveyed letter defects on 500 manual, segment shift, typewriters. The defects Crown noted were of 5 types: *physical damage* (further categorized by location: top, bottom, right, left), *tilt* (right or left), *rebound* (presence or absence), *heavy striking* (top, bottom, left, right), and *malalignment* (high, low, left, right). The specific types of defects observed were then examined with respect to 2 dependent variables: typewriter manufacturer and typewriter age.

Received for publication 2 Nov. 1987; accepted for publication 3 Oct. 1988.

[1]Assistant professor of criminalistics, Department of Criminal Justice, University of Illinois at Chicago, Chicago, IL.

[2]Professor of forensic science, Department of Biomedical and Environmental Health Sciences, University of California, Berkeley, CA.

Dr. Crown did not make use of inferential statistics, but did conclude that there was no correlation between frequencies of letter use and physical typeface damage and that this typeface damage was not affected by basket position. He also observed, however, that there was a slow increase in overall typeface defects with the duration of typewriter use. This overall increase in defects, without an increase in defects correlated with individual letter use, seemed peculiar and was the initial motivation for the present study.

Experimental Procedure

Independent Variable

The independent variable was the typeface damage frequency for the lowercase letters. Lowercase letters were used because Crown had observed that capital letters are damaged much less frequently ([*1*], p. 108). Standardization on form was also desired, as described below. The damage frequencies were calculated from Crown's tabulation of character damage on typewriters in different age groups ([*1*], pp. 40–42). The actual number of typewriters where damage was observed on a particular letter was calculated from the percentages given in the table, multiplied by the total number of typewriters in the particular age group ([*1*], pp. 13–14).

Definition of Dependent Variables

Given Crown's negative findings, a variety of dependent variables were considered for analysis. These fell into three groups: those based on letter use frequencies; those based on letter form; and those based on basket or keyboard position.

Four variables were defined based on letter-use frequencies. The first was the overall letter-use frequency [*2*]. The second was the frequency as the first letter in a word [*2*]. This variable was chosen because it seemed reasonable that when a letter was the first in a word, a collision or "pile-up" of keys would be less likely. It was hypothesized that this would contribute to a lower damage frequency. The third related variable was the frequency in two-letter combinations requiring difficult finger movements on the keyboard. Frequencies of typewriting errors can be related to finger movement difficulties [*3*], and it was hypothesized that type-bar collisions might also follow this pattern, leading to higher frequencies of damage. Difficult finger movements are described by Davis [*3*] as adjacent finger reaches (such as "aw" and "se"), adjacent finger hurdles (such as "be" and "ve"), remote finger hurdles (such as "xt" and "no"), single finger reaches (such as "ik" and "ju"), and single finger hurdles (such as "ec" and "rv"). For each letter, a relative measure of frequency of use in difficult letter combinations was determined by weighting each occurrence in a difficult pair by the pair's frequency of occurrence, taken from Pratt [*2*]. The fourth and final variable based on letter-use frequencies was similar to the one just described, except that only the letter pairs with the *most difficult* finger movements were considered. Adjacent finger reaches and remote hurdles were excluded from this variable because from a kinesthetic standpoint they appeared to be much easier movements than did the remainder.

Three variables were based on letter form. The first was the *total letter area.* Greater letter area was presumed to give greater susceptibility to damage. Letter area was measured using a single typeface from the Interpol standard references set. Standard fiche No. 55 (Smith-Corona Pica No. 1) was projected onto graph paper, polygons were drawn around each of the lower case letters, and the areas were determined by counting the number of squares within each polygon. The second letter form variable was a three-way classification based on the presence of ascending or descending vertical shafts. Letters ascending above the central letter position (b, d, f, h, k, l, and t) were coded $+1$, letters with descenders (j, p, q, y) were coded -1, and the remainder were coded as zero. The third letter variable, closely related to

TABLE 1—*Raw data for multivariate analysis.*

Letter	Damage Frequency (per 500)	Frequency of Use	Frequency as First Letter	Occurrence in Difficult Combinations	Occurrence in Most Difficult Combinations	Area	Extension Classification	Carriage Position	Row
a	50	81.5	123.6	72	0	200	0	20	2
b	11	14.4	42.4	38	29	214	+1	2	3
c	1	27.6	47.0	69	62	169	0	10	3
d	36	37.9	20.6	108	78	256	+1	12	2
e	9	131.0	27.7	592	195	178	0	13	2
f	5	29.2	41.9	33	15	231	+1	8	2
g	15	19.9	12.3	55	35	279	−1	4	2
h	92	52.6	47.5	52	9	281	+1	0	2
i	40	63.4	59.0	262	141	199	0	7	1
j	0	1.3	7.0	2	2	174	−1	4	2
k	45	4.2	1.8	11	8	296	+1	8	2
l	27	33.9	22.4	80	28	210	+1	12	2
m	53	25.4	39.2	69	41	240	0	6	3
n	60	71.0	21.6	214	132	237	0	2	3
o	31	80.0	70.8	232	28	162	0	11	1
p	29	19.8	40.9	65	0	267	−1	15	1
q	16	1.2	2.2	1	1	239	−1	21	1
r	26	68.3	31.2	285	91	206	0	9	1
s	3	61.0	69.1	148	3	184	0	16	2
t	11	104.7	181.3	130	67	183	+1	5	1
u	42	24.6	14.1	40	37	205	0	3	1
v	18	9.2	4.1	43	43	136	0	6	3
w	24	15.4	63.3	60	2	186	0	17	1
x	11	1.7	0.4	16	12	224	0	14	3
y	39	19.8	7.5	13	12	250	−1	1	1
z	0	0.8	0.4	1	0	181	0	18	3

the second, was a two-way classification based on the presence or absence of vertical extension (ascenders and descenders combined into one classification).

Two variables were based on key and typebar positions. The first was the position relative to the central key bar (the letter h). This variable was chosen because the more distant letters strike at a more oblique angle, which could possibly be correlated with letter damage in some way. This was also one of Crown's variables. The second variable was the row on the keyboard where the letter appeared. Conceivably, the row position could have an effect on errors and collisions. Home keys, for example, might have higher collision rates because they are struck more quickly. Raw data for both dependent and independent variables are presented in Table 1.

Statistical Methods

Standard Statistical Package for the Social Sciences (SPSS) library routines were used for the multivariate statistical analysis [*4*]. Principal components analysis, multiple regression, and covariance analysis were used in the sequential analysis.

Results

The letter-use frequency variables were examined by principal components analysis and found to be highly correlated. Using all four variables gave poorer results ($p = 0.169$, R square $= 0.077$) than did using the overall letter-use frequency by itself ($p = 0.076$, R square $= 0.1251$). The other three variables, being more speculative, were dropped from further consideration.

Multiple regression analysis was conducted using the letter-use frequency and the two other quantitative variables, letter area and basket position. Basket position was found to have no statistically significant effect on typeface damage frequencies (p [to remove] $=$ 0.347). Both letter-use frequency and letter area, however, were found to show statistically significant positive correlation with typeface damage frequencies (p [to remove] $= 0.044$ and 0.005, respectively). Together, these two variables accounted for 42% of the observed variation in damage frequencies ($p = 0.002$).

Covariance analysis demonstrated that each of the remaining three classification variables showed no statistically significant relationship to typeface damage frequencies ($p = 0.183$ for three-way classification of letter extensions, $p = 0.08$ for two-way classification of letter extensions, and $p = 0.738$ for keyboard row).

Conclusion

In Dr. Crown's words, "The statistical data . . . is not a goal in itself, but (serves) as a guide for decisions regarding the probative value of typewriting individuality" [*1*, p. 11].

Although the present work has identified two deterministic factors with high statistical significance (letter-use frequency and letter area), there is still 58% residual variation. With this amount of variation left unexplained, the factors identified need not be given much weight in the interpretation of typewriting comparisons. The remaining variables studied show no evidence of affecting typeface damage.

Although general letter area was identified as a significant variable in this study, the obvious omission was any consideration of serifs, which are generally recognized as being particularly susceptable to damage. The actual typefaces used on the typewriters in Crown's survey were unknown and prevented study of this variable.

References

[1] Crown, D. A., "A Statistical Evaluation of Typewriting Individuality," D. Crim. dissertation, University of California, Berkeley, CA, 1969, pp. 13-14, 40-42, 108.
[2] Pratt, F., *Secret and Urgent*. Bobbs-Merrill, New York, 1939, pp. 252-259.
[3] Davis, D. D. W., "A Comparison of Word Errors on the Universal and on the Simplified Typewriter Keyboards. An Evaluation of the Simplified Typewriter Keyboard, Part IV," *Journal of Business Education*, Vol. 11, No. 1, Sept. 1935, pp. 21-22, 38, and Vol. 11, No. 2, Oct. 1935, pp. 19-21.
[4] Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H., *SPSS—Statistical Package for the Social Sciences*, 2nd ed., McGraw-Hill, New York, 1975.

Address requests for reprints and additional information to
David A. Stoney
Department of Criminal Justice
University of Illinois at Chicago
Box 4348
Chicago, IL 60680