B
15409 fo

= B15075

LTR 1 WORKING PAPER 3

SWOV
BIBLIOTHEEK
VOORBURG
TEL. 070 - 69.41.21

CYCLING IN TOWNS:
A QUANTITATIVE INVESTIGATION

J. A. WALDMAN

B
15409 fo

DEPARTMENT OF TRANSPORT

DECEMBER 1977

BIBL. NR.
B
15409
80

Stichting Wetenschappelijk
Onderzoek Verkeersveiligheid SWOV
Ridder Snouckaertlaan 7 - Voorburg
Postbus 71 - 2270 AB Voorburg

NAAM	RETOUR DATUM
~~AW~~	~~STANDEX~~ ~~28 OKT. 1979~~

LTR1 WORKING PAPER 3

SWOV
BIBLIOTHEEK
VOORBURG
TEL. 070-69.41.21

CYCLING IN TOWNS: A QUANTITATIVE INVESTIGATION

J A WALDMAN

ABSTRACT

Variations in cycle use from town to town are related to measures of hilliness, traffic danger, rainfall, trip-length and socio-economic mix. It appears that hilliness and traffic danger are the main determinants of cycle use and estimates are provided of their effect. The analysis relied on readily available data and was correspondingly crude. Recommendations are made for further work.

LTR1 Working Papers are intended to facilitate limited dissemination of information about work in progress and work completed by LTR1 Division. Neither the Department of Transport nor the Division necessarily endorses any views or opinions contained in the Working Papers, they are the responsibility solely of the author(s). Short extracts may be reproduced without specific permission provided that the source is acknowledged.

(C) CROWN COPYRIGHT 1977

DEPARTMENT OF TRANSPORT

DECEMBER 1977

P-790713/06
S-790731/01

LIST OF CONTENTS

CYCLING IN TOWNS: A QUANTITATIVE INVESTIGATION

1. BACKGROUND

1.1 PURPOSE

1.1.1 The study set out to test the assertion that considerably more people would make journeys by bicycle if they could do so safely. To date the evidence for this has come only from attitude surveys, for example, one carried out by Camden Friends of the Earth. Their report, "Cycling in Camden", June 1975, has a table showing that 46% of respondents who do not own or ride a bicycle cited "danger" as their main reason. Similarly, at a conference - "Cycling Today", Cardington 1977 - Mr P Snelson, an officer of Bedfordshire County Council, reported that 45% of survey respondents in Bedford cited "danger" as the major disadvantage of cycling.

1.1.2 Whilst the results of such attitude surveys have striking implications they carry very little objective weight and one may rightly ask whether non-cyclists might not discover some other "main reason" (eg effort) for not cycling if safe cycling were made possible. It therefore seems necessary to find objective evidence that will support or refute the implications of the survey.

1.1.3 One approach is to observe actual behaviour. The avenue explored in this study was to construct a measure of "danger" related to the proportion of cyclists who had accidents, and to analyse statistically the extent to which differences in danger levels accounted for differences in the proportions of cyclists amongst the towns that were studied.

1.2 THE SCOPE OF THE STUDY

1.2.1 The study was confined to the use of readily available data requiring no new surveying effort. This meant that rough and ready measures had to be designed for factors thought to influence cycling. As a result the analysis had to contend with a large amount of measurement error.

1.2.2 The main tool of analysis was multiple linear regression. It soon became apparent that the factors which influence cycling do not do so in a simple additive manner, and it became necessary to employ variable transformations and to depart from the ideal conditions for the use of regression. Nevertheless, the least-squares approach involved in multiple linear regression seemed the most practical means of curve-fitting, the general shape of the curves having been decided upon by inspection of scatter-plots. It also made possible the use of R-square which measures the degree to which variation in the variable of interest (cycling) is accounted for by the model being used. This provided a means for judging whether changes to the model constituted improvements.

1.2.3 Bearing in mind the measurement and statistical limitations, the model that was developed is not being put forward for future use as a definitive, predictive planning-tool, but merely for its original purpose, ie a means of unravelling the factors which interact to determine the level of cycle usage in a town and to separate out the contribution made by "danger".

1.3 THE VARIABLES

1.3.1 The dependent variable, labelled CYCLE, was the proportion of people who live and work in an area who reported riding a bicycle as their major mode of travel to work. 1966 Census data were used as these were the latest available suitable for investigating country-wide and area - specific patterns of bicycle-use. The restriction "live and work in an area" was made in order to enable approximate estimation of trip-lengths for the journeys under consideration.

1.3.2 The sample consisted of 195 non-rural local authority areas. The sample and sampling procedure are discussed in appendix 1.

1.3.3 The factors thought to be the most influential in causing variation in bicycling levels were hilliness, rainfall, trip-lengths, accident-risk, availability of alternatives and life-style factors. Wind conditions may be important but a suitable measure was not available and its influence was therefore not modelled.

1.3.4 The following defines and labels the measures used:-

HILL = the number of 25 foot contour changes per road-mile in the built-up area, measured from Ordnance Survey maps.

RAIN = the number of days of rainfall in 1966, of more than 2.5mm of rain. The midpoints of the ranges quoted by the Meteorological Office were used.

R = the radius of the built-up area (see appendix 3).

TL = a trip length factor derived from R (see appendix 3).

RESTD = an estimate of the accident risk derived from TRRL data on the number of accidents in different areas (see appendix 2).

INCOME = average household income of the area (1966) according to the National Traffic Model.

SEGA = proportion of agricultural employees.

SEGM = proportion of non-agricultural manual workers.

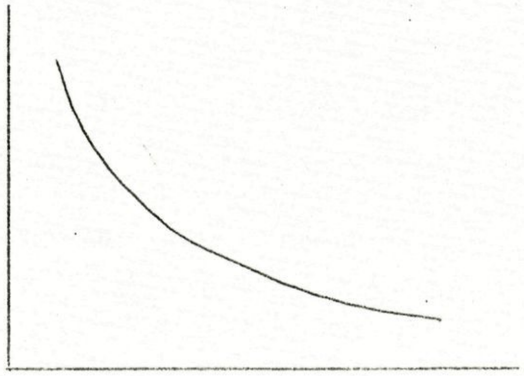
SEGN = proportion in non-manual occupations.

1.3.5 Some other variables were defined, but did not add to the explanatory powers of the model, so they will not be discussed here.

2. THE ANALYSIS

2.1 PRELIMINARY ANALYSIS

2.1.1 Some preliminary multiple linear regression was carried out which indicated that HILL, RESTD and RAIN were important influences. Figures 2.1 to 2.6 show CYCLE plotted against HILL, RESTD, RAIN, R, INCOME and SEGN. The first three graphs approximate monotonically decreasing functions as in the sketch.



2.1.2 The spread of CYCLE values decreases as HILL, RESTD or RAIN increase, suggesting that the effects combine in in a multiplicative manner rather than in an additive one.

2.1.3 It was felt that town size (R, which is related to trip lengths in the area) and socio-economic mix should have more to contribute to explaining CYCLE variation than is apparent from the graphs or linear regressions, and their role would be facilitated by explicit incorporation into a model. Appendix 3 explains how a trip length factor was derived, and the next section, the subsequent model development. For a less technical reading of this paper, one could skip to paragraph 3.3.3, where the developed model is used to unravel the influences of the contributory factors.

2.2 DEVELOPMENT OF THE MODEL

2.2.1 A simple model of CYCLE might be specified as follows: CYCLE (ie the proportion who cycle) is a function of the proportion of people in the different socio-economic categories. This is expressed in equation 2.1.

$$\text{CYCLE} = a\,\text{SEGA} + m\text{SEGM} + n\text{SEGN} + k \tag{2.1}$$

where a, m and n are parameters and k is a constant.

2.2.2 Such a model might well be adequate if the factors related to a person's occupation were the main determinants of propensity to cycle; the parameters a, m and n reflecting the different propensities. Other factors such as hilliness might increase or decrease this propensity. In concert, they would tend to weight the propensities as expressed in equation 2.2

$$CYCLE = waSEGA + wmSEGM + wnSEGN + k' \qquad (2.2)$$

where w is not a parameter, but a variable taking specific values according to a town's topographic and accident-risk factors.

2.2.3 Recalling paragraphs 2.1.1 and 2.1.2, a model of the form

$$w = e^{a_1X_1} . e^{a_2X_2} ... e^{a_nX_n} . e^{c} \qquad (2.3)$$

where the X's are topographic and accident-risk factors, the a's parameters and c is a constant, might be a satisfactory model for the non-socio-economic group factors. When calibrating the formula (2.3), CYCLE is used as the dependent variable. The formula can then be used to calculate a weighting factor, w, for each town, which is based on the non-socio-economic influences.

2.2.4 Other models were considered apart from the two-stage approach outlined above. The approaches included

a. using the total number of bicycle accidents for each area as a proxy danger value;

b. use of logs of variables, squares of variables and products of variables in the regressions;

c. using the transformation CYCLE/(1-CYCLE) as the dependent variable;

d. trying to fit hyperbolic curves;

e. greater differentiation of socio-economic categories;

f. separate calibrations for conurbations;

g. use of dummy variables for town characteristics, eg, free-standing, or conurbation.

2.2.5 All of these approaches, except (g), achieved worse or no better total R-square results than the two-stage approach which is reported below. In case (g) it was possible to add another 0.06 to the total R-squared of the two-stage approach, but this was at the expense of being able to separate out the effects of the different underlying factors. Having tried these various approaches, it seemed reasonable to conclude that any further research to improve the modelling of cycling would do better to use more accurate measures of danger and trip-lengths.

3. RESULTS

3.1 CALIBRATION OF THE EXPONENTIAL MODEL

3.1.1 To calibrate the exponential model expressed by equation 2.3, the logarithms of CYCLE were regressed on the non-socio-economic group variables in a stepwise, multiple linear regression. The best results, as far as the final model was concerned, were obtained when none of those cases for which the value of CYCLE was less than 0.04 were included in the calibration involving logarithms. This procedure overcomes some of the error distortion involved in the log-transformation. Four of the variables entered the log-regression, these being HILL, RESTD, RAIN and the log of TL, in that order of entry. The variables brought about R-square changes of 0.37, 0.20, 0.02 and 0.02 respectively. The calibrated formula is

$$w = \exp[-0.193\mathrm{HILL} - 0.00944\mathrm{RESTD} - 0.00623\mathrm{RAIN} - 0.104] \times \mathrm{TL}^{0.786} \qquad (2.4)$$

where HILL, RESTD, RAIN and log TL took the place of the X's of equation 2.3

3.2 SOCIO-ECONOMIC VARIABLES AND FINAL MODEL

3.2.1 The values of the variables SEGA, SEGM and SEGN were multiplied by w and CYCLE was regressed on these weighted variables. The inclusion of the variable INCOME amongst the independent variables improved the total R-square by 2%, so INCOME was included in the final model which is

$$\widehat{\text{CYCLE}} = w(3.88\text{SEGA} + 1.26\text{SEGM} + 0.618\text{SEGN}) + (5.37 \times 10^{-5}\text{INCOME}) - 0.0796 \qquad (2.5)$$

where w is the result of the calculation of equation 2.4. The R-square for CYCLE with w was 0.713 and with $\widehat{\text{CYCLE}}$ it was 0.745.

3.3 INTERPRETATION OF THE MODEL

3.3.1 Details of R-square changes and standard errors of coefficients are set out in appendix 4. Although the R-square changes due to RAIN and TL are small, these variables enter the regression at a level of significance better than 0.05. Since HILL correlates well with RAIN ($r = 0.43$), and RESTD with TL ($r = -0.32$), it is important to separate their effects by including them in the model (given that there is enough non-correlation to separate out the effects).

3.3.2 Rather than use R-square information, which indicates the relative power of the variables to account for variation in CYCLE levels around the country, we can turn to the calibrated model which indicates how sensitive CYCLE is to changes of the variables.

3.3.3 The model enables us to see how CYCLE varies as a function of any particular variable whilst all the others are held constant. We can set all "other" variables at their mean values and allow one to range from the minimum to the maximum of the sample. This procedure was carried out for each variable and the range of each variable was rescaled from zero to 10 so that they could all be graphed together for comparative purposes (figure 3.1). Income and socio-economic groups are not represented on the figure because it does not make sense to vary one and hold the others constant. The graph is slightly misleading for the HILL curve because its range incorporates an extreme outlier, Lyme Regis, which takes the value of 10 on the x-axis. Apart from this town, no other areas had a value greater than 6.1. Figure 3.2 reproduces figure 3.1 except for the re-scaling of the HILL curve after excluding the extreme outlier from the range of values considered.

3.4 DISCUSSION

3.4.1 The analysis of variation in cycling levels, as illustrated by the last figure, shows that low levels can be accounted for by places being either too hilly or too dangerous, whilst other factors have less influence. The discussion will deal firstly with the less influential variables.

3.4.2 SOCIO-ECONOMIC GROUPS AND INCOME

As a result of several regression runs it was found that these variables only made significant contributions when both were present in the analysis together. Even then they helped explain merely an extra 3% of the variation.

3.4.3 The coefficients of the socio-economic groups in equation 2.5 are consistent with the widely held view that, in any given area, manual workers are more likely to ride bicycles than are non-manual workers. However, the equation goes against commonsense by predicting higher cycle use in areas of higher income. INCOME and HILL had a correlation coefficient of -.13, suggesting that higher income areas are more likely to be flat than hilly (and be in the South East rather than the hillier North West). It may be that the HILL measure underestimates the true hilliness of several higher income areas and so the INCOME variable may be entering spuriously on this account. Perhaps the most reliable statement one can make is that socio-economic mix and income contribute little to the variation in cycling levels of one town from another, but that this leaves open the possibility that they are important factors within towns.

3.4.4 TOWN SIZE/TRIP LENGTH

Figure 3.2 indicates that in very small towns (of radius less than one kilometer), town size has a marked influence. One expects that people would prefer to walk the short distances involved. However, apart from the very few tiny towns, town-size per se has relatively little influence on the proportion who cycle to work. This suggests that, for most towns, the majority of internal work-journeys are not so lengthy as to be a deterrent to cycling. The scale of the radius variable should be treated with caution as the circularity,centrality and density assumptions may have led to systematic scale distortions.

3.4.5 RAINFALL

The effect of higher than average rainfall is certainly not negligible. Given a town of otherwise average characteristics, one would expect a level of cycling of 8½% for, say, the North West of England and 17½% in the dryer South East. However, rainfall alone is unlikely to be the factor determining whether a town will have many or few cyclists.

3.4.6 DANGER/ACCIDENT RISK

The measurement procedure for this variable was probably more prone to error than those for the others. However, it has proved to be useful for the analysis of variations between towns. The results indicate that high levels of danger are associated with low levels of cycling. Thus a highly dangerous town which is otherwise average with respect to other factors would have a 2% level of cycling as opposed to 20% if it were very safe.

3.4.7 HILLINESS

Hilliness emerges as an extremely strong influence, and, according to figure 3.2, flat places are likely to encourage people to cycle even more than will safe places.

3.4.8 DANGER AND HILLINESS

The model enables us to look at the effect on cycling levels when the two major determinants act together. Table 3.2 shows the cycling levels to be expected from combinations of low and high levels of danger and hilliness, all other factors being held at their mean sample values. The table indicates that cycling is low when an area is either hilly, or dangerous or both (4%, 6% and 0%

respectively), whilst if it is flat and safe, a large proportion will cycle to work (43%). The table is illustrated with some examples of actual towns with such combinations of characteristics.

3.4.9 RELIABILITY OF THE MODEL

Figure 3.3 presents a plot of CYCLE against model estimated CYCLE. One can detect a slight tendency for the model to overpredict places of low cycling and to underpredict places of high cycling. Although the standard error of the estimates is a bit too large (0.069) to allow us to have confidence in predictions for specific towns, the model is clearly capable of discriminating between towns in the sample with low, middle or high levels of cycle use, and is accurate enough to justify the general statement about the magnitudes of the effects of the variables.

4. CONCLUSIONS

4.1 The investigation separated out the effects of various influences on cycling and indicates that cycling levels could rise considerably, especially in non-hilly areas of high accident-risk if these were made safer for people to ride bicycles. Thus for areas of moderate urban hilliness and high accident-risk, eg Birmingham or Salford, up to 20% of residents could be expected to cycle to work if they could do so in safety, as opposed to about 3% of residents who were cycling to work in those areas when the 1966 Census surveys were carried out.

4.2 The sensitivity of cycling to hilliness suggests that there may also be a large demand for power-assisted bicycles if they could be used in safety.

4.3 Apart from danger and hilliness, cycling levels were not found to be markedly sensitive to rainfall, socio-economic mix or town size per se, except that in the smallest of towns (of radius less than one kilometer), people prefer to walk the short distances involved.

4.4 Although the predictions for particular places are bound to be suspect if indirect and crude measures are used, it would seem that purpose designed models could be developed for the evaluation of local bicycle schemes if more effort than was possible in this study is put into developing accurate and predictable measures, particularly of the accident-risk to riders of bicycles.

5. RECOMMENDATIONS

5.1 To make specific predictions for the effects of such schemes as cycle-routes, or specially designed junctions, would require further research the aims of which should be:

a. to develop measures of accident-risk which can be related to traffic conditions, and road-features such as junctions, in order to be able to assess improvements to cycle-safety and to predict subsequent levels of cycle-use;

b. to consider how cost-benefit analysis needs to be extended in order to be able to assess the appropriate costs and benefits of such schemes.

LTR1
Department of Transport
212 5385

December 1977

APPENDIX 1. THE SAMPLE

A.1.1. Rural districts were excluded from consideration, leaving Urban Districts, Municipal Boroughs, County Boroughs and London Boroughs (abbreviated to UD's, MB's, CB's and LB's), as the remainder.

A.1.2. In order to reduce the data collection and transcription, a sample of approximately one hundred of these districts and boroughs was considered to be large enough for modelling and statistical inference purposes.

A.1.3. The sample was selected by taking the first UD, MB, CB or LB that appeared on alternate pages of the Workplace and Transport Tables.

A.1.4. Early analysis of this data showed a range of cycling from 0% to 50% of the resident working populations of towns. However, the spread was uneven and so it was decided to fill in the gap for the range above 25% by selecting an additional 50 areas where the proportion of cyclists was at least 25%. This addition was achieved by incorporating all the remaining areas which exceeded 25% into two new subfiles.

A.1.5. This addition included no additional County Boroughs or London Boroughs and they seemed to be under-represented in the file (especially vis-a-vis the population that they contain). Also, it is the CB's and the LB's that one expects to be more dangerous for the cyclist.

A.1.6. Therefore it was decided to randomly select an additional 30 CB's and LB's from amongst the boroughs not yet in the sample.

These were then incorporated as two new subfiles.

A.1.7. A few areas were subsequently dropped from the file because of missing data for certain of the variables, eg. non-availability of maps, from the map library, for the areas.

A.1.8. The sub-file arrangements were thus as follows; the suffix '2' indicating the areas selected by the methods described in paragraphs A.1.4. to A.1.6. :

(a)	UD's	58	cases
(b)	MB's	37	"
(c)	CB's	18	"
(d)	LB's	4	"
(e)	UD2's	37	"
(f)	MB2's	14	"
(g)	CB2's	19	"
(h)	LB2's	8	"
Total	=	195	cases

APPENDIX 2. DETERMINATION OF ACCIDENT RISK

A2.1 Ideally we would use a measure of accident risk derived from knowledge of traffic conditions. It would then be possible to predict danger levels for changes in traffic conditions and to use the model to predict the new cycling levels. However, such knowledge is not yet at hand.

A2.2 The next best alternative seemed to be to use accident statistics in order to derive an accident-rate per cyclist. The number of accidents in each local authority area can be ascertained from TRRL records. The number of cyclists who were candidates for accidents is not known and has to be estimated from the proportions who cycle to work. The assumptions made were that the proportion of non-workers who cycle is identical to that of the workers, and that this applies to people cycling into the areas from outside and vice versa. The total cycle population was calculated as the number estimated to be cycling within the area, plus those cycling into it plus those starting off in it and cycling out of it.

A2.3 Because this approach will almost definitely lead to distortions and mis-representation of the relative danger in different areas, the ability of the proxy variable to explain variation in cycle-levels throughout the country will be diminished and the results of the investigation will not reflect the full extent to which accident-risk influences cycling.

A2.4 To proceed, we compute

PEOPLE = POPN + (INFLO x (POPN/EMPRES)) (A3.1)

where PEOPLE = the estimate of the number of people who make journeys in the area.

POPN = the resident population of the area.

INFLO = the number of people entering the area to work.

EMPRES = the number of residents who are in employment.

The proportion of cyclists among the PEOPLE can then be estimated as a weighted average:

PC = ((INFLO x CYCLEIN) + (OUTFLO x CYCLE OUT)
+ (RESWORK x CYCLE))/(INFLO + OUTFLO + RESWORK) (A3.2)

where PC = the proportion of cyclists among the people.

CYCLEIN = the proportion of the INFLO who cycle.

CYCLEOUT = the proportion of the OUTFLO who cycle.

RESWORK = the number of residents employed in the area itself.

CYCLE = the proportion of the RESWORK who cycle.

A2.5 The risk or danger variable is then defined as

D = BIKEAX/(PEOPLE x PC) (A3.3)

where BIKEAX is the recorded number of accidents involving cyclists in the local authority area. A similar measure, DP, or danger to pedestrians was also calculated where DP = PEDAX/PEOPLE, ie. the pedestrian accident rate where all people who reside in or enter the area are considered as potential accident victims.

A2.6 Use of the variable D has two drawbacks. Firstly, it involves the dependent variable CYCLE in its computation, which makes its use as an independent variable to predict CYCLE quite dubious. Secondly, for places with few cyclists, the figures cannot be expected to reflect accident risk with any accuracy, just as the observed proportion of heads for a few tosses of a coin has very little confidence attached to it as representing the true probability of heads for the coin.

A2.7 Just as one can calculate a confidence interval for a sample proportion, one can treat the observed accident-rate as a sample proportion and construct a confidence interval for it. The assumptions made are that each cyclist has the possibility of having one or no accidents in 1966, and that no accident involved more than one cyclist. Values labelled CI were computed such that D ± CI formed a 90% confidence interval for D. An analogous value, CIP was computed for the pedestrian accident rate, DP.

A2.8 Where the confidence interval on a value is small relative to that value, we can have a greater degree of confidence in that value representing accident-risk. In only four cases were the CI's smaller than 10% of the corresponding D values. In 42 cases the CI's were within 20%. Forty-two cases did not seem to be a large enough or representative enough sample on which to carry out further investigation. The largest danger estimates were close to a value of 0.024. The criterion adopted was that a value was considered acceptable if its CI value was not greater than 10% of 0.024. On this basis 144 areas were acceptable forming a good cross-section of sizes of town and of accident risk.

A2.9 A similar approach was used to select areas for the confidence in their pedestrian danger estimates.

A2.10 The correlation between D and DP was 0.63 (for the acceptable cases). This value was not felt to be high enough to warrant substituting pedestrian danger for cyclist danger in order to overcome the first objection to using D as an independent variable.

A2.11 A more reliable estimator for D came from a regression of D on DP, OUTFLO, CARPOP, INFLO, POPN, FLORATIO, ADMIN, PEOPLE and R where ADMIN = 1, 2, 3 or 4 for UD's MB's, CB's and LB's respectively, CARPOP = the ratio of cars to population, and FLORATIO = the ratio of INFLO + OUTFLO to RESWORK, these variables being defined as earlier on in this appendix.

A2.12 Only three of the variables entered the step-wise regression contributing more than 1% to the cumulative R-square.

STEP	VARIABLE ENTERED	SIGNIFICANCE	MULTIPLE R	R SQUARE	R SQUARE CHANGE
1.	OUTFLO	.000	.77	.59	.59
2.	DP	.000	.83	.70	.11
3.	CARPOP	.000	.86	.73	.04

Dependent variable = D

A.13 The coefficients were used to compute a new danger estimate, RESTD (re-estimate of D):-

$$\text{RESTD} = (0.127 \times 10^{-6} \times \text{OUTFLO}) + (4.56 \times \text{DP}) + (0.0205 \times \text{CARPOP}) - 0.00367 \qquad \text{(A3.4)}$$

This equation was used for all the 159 cases for which DP was acceptable. For the remaining 36, a value of RESTD was computed from the variables OUTFLO, ADMIN and R, the three significant variables in a regression not involving DP. Total R-square was equal to 0.64 and the formula was

$$\text{RESTD} = (0.801 \times 10^{-7} \times \text{OUTFLO}) + (0.757 \times 10^{-3} \times \text{R}) + (0.831 \times 10^{-3} \times \text{ADMIN}) + 0.00108 \qquad \text{(A3.5)}$$

A2.14 All the REST values were subsequently scales up by 10^4 so that they could be handled by the statistical package's printing and plotting routines. Thus, TABLE 3.1 shows the range of RESTD being (9,203). This corresponds to estimates of accident rates of 0.0009 and 0.0203 per annum, per cyclist.

A2.15 The use of two formulae for RESTD may introduce some distortion into the variable. On the whole they produce similar estimates for the 159 cases for which the DP value is acceptable (R = 0.88); a factor of 1.16 to be applied to the second formula to achieve the best fit. The 36 cases which did not pass the DP criterion were all small towns where the small population size and number of accidents did not allow for confident accident-risk estimation. The use of equation A3.3 has assigned them all low values of RESTD (RESTD = 20 to 40). When the two formulae were tested out on the 159 cases for which DP is acceptable, 49 cases were assigned values between 20 and 40 by the second formula. In 21 cases this underpredicted the first equation's estimate, in 19 it overpredicted whilst the remaining 9 cases were of equal magnitude. Thus the tests suggests that the two-formulae approach does not introduce major distortion.

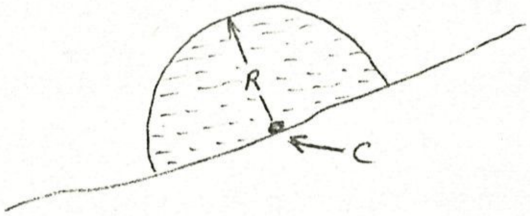
APPENDIX 3. DETERMINATION OF THE BUILT-UP AREA AND THE TRIP LENGTH FACTOR

A3.1 THE RADIUS OF THE BUILT-UP AREA, R

A3.1.1 The built-up areas were determined from Ordnance Survey maps, scale 1:25000. Bearing in mind that it was travel distances within the town which were of concern, the determination was as follows:

A3.1.2 The radii of roughly circular towns are measured directly.

A3.1.3 Semi-circular towns, whose town-centres are close to the edge, were treated as if the "missing half" really existed. Thus, if the shaded area of the sketch represents a town and "C" the town centre, the radius of the town would be "R".



A3.4.1 Oblong or elliptical towns would be treated by taking the average of its major and minor axes and then treating them as circular towns of that diameter.

A3.1.5 Approximations to these techniques were used for irregularly shaped towns.

A3.1.6 The arbitrariness and non-reproducibility of the values arrived at were not felt to be too important for the following reasons.

A3.1.7 Firstly, to arrive at an approximate distribution of journey lengths, it was decided to treat towns as circular with all trip attractors at the town centre and of uniform population density. Thus the methods used seemed a practical approach of sufficient accuracy for a variable which was bound to be an approximation.

A3.1.8 The magnitudes of the errors of estimating the radii in this manner are small relative to the between-town differences in radii. The largest cities have radii well over ten times the radii of the smaller urban districts.

A3.2 TRIP LENGTH FACTOR

A3.2.1 A simple model for CYCLE might be that CYCLE is a function of the proportion of people, living at certain distances from their work, multiplied by the probabilities of cycling such distances.

$$\text{CYCLE} = Wa_1PD_1+Wa_2PD_2+\ \ldots\ +Wa_nPD_n \qquad (A3.1)$$

where the PD's are the proportions of people living within certain bands of distances from their work (eg 1 to $1\frac{1}{2}$km), the a's are the probabilities of cycling those distances, and W is a different weighting factor for each town to take into account factors other than trip-length.

A3.2.2 The weighting factor employed was the best achievable estimate of CYCLE based on the variables HILL, RESTD and RAIN. An exponential model was used for this and was arrived at by the technique described in section 3 of the main text with the only difference that a trip-length factor is not incorporated into this weighting factor. The total R-square for this weighting factor was 0.66 (calculated when CYCLE was correlated with it in a separate test).

A3.2.3 To calculate the PD's, the towns were assumed to be circular with residents homogeneously spread within each. Work was assumed to be at the centre of the circle, and the proportion of the circle within the following distance-bands from the town centre was calculated for each area:

BAND LABEL	RANGE (MILES)	RANGE (KM's)
PD1	$0-\frac{1}{4}$	0 - 0.4
PD2	$\frac{1}{4}-\frac{1}{2}$	0.4 - 0.8
PD3	$\frac{1}{2}-1$	0.8 - 1.6
PD4	$1-1\frac{1}{2}$	1.6 - 2.4
PD5	$1\frac{1}{2}-2$	2.4 - 3.2
PD6	2-3	3.2 - 4.8
PD7	> 3	> 4.8

A3.2.4 For each area, the PD's were multiplied by the weights and CYCLE was regressed on these seven products. The regression was forced through the origin so as not to have a constant in the expression, corresponding to the form of equation A3.1 above.

A3.2.5 The regression enables us to calibrate equation A3.1. It is to be noted that the weighting factor employed might need to be multiplied by some constant in order that the calibration coefficients come out as true probabilities rather than multiples of them. However, we do not know what constant to employ, so we proceed on the understanding that the calibrated coefficients are some constant times the probabilities.

A3.2.6 The ranges of the distance bands were arrived at by trial and error so as to give a number of bands whose coefficients were significant and which were numerous enough to give a picture of how cycling varies with trip-lengths.

A3.2.7 The results of the regression are set out in the following table:

	Coefficient "a"	standard error of "a"	significance
a_1	0.183	0.208	0.381
a_2	1.477	0.138	0.000
a_3	1.116	0.077	0.000
a_4	0.802	0.160	0.000
a_5	0.772	0.314	0.015
a_6	0.531	0.344	0.125
a_7	0.361	1.263	0.775

The total R-square for the regression was 0.71.

A3.2.8 The results are displayed graphically in figure A3.1. The relative magnitudes of the coefficients above are to be taken as the relative probabilities of cycling for the respective distance bands.

A3.2.9 For any initial value of R, only one set of values (PD1, PD2,,, PD7) is determined, so it should be possible to encapsulate the trip-length effect in a unitary value.

3.2.10 Consider that people live at a distance between x and $x + \delta x$ from the centre of their circular town, A3.2.11 The proportion of people living in the annulus of width δx is approximated by its width times its length ie $2\pi x\delta x$, and the proportion of the circular town that it represents is $2x\delta x/R^2$. Each individual in that distance band has approximate probability, say, of $f(x)$ of cycling to work, so the proportion of the town who both live in such an annulus and cycle to work is given by the product $2xf(x)\delta x/R^2$. If we let δx tend to zero and integrate over the whole range of x values, we obtain the proportion of the town who cycle to work in the area, ie

$$\text{CYCLE} = \frac{2}{R^2} \int_0^R xf(x)dx \qquad (A3.2)$$

A3.2.12 This expression for estimating CYCLE ignores the effects of other factors such as hilliness. When such factors are controlled for, the expression above could be used as the unitary trip-length variable that we are seeking.

A3.2.13 The distance band data, graphically presented in figure A3.1, can help us to find a suitable function $f(x)$. One is tempted to fit a gamma-distribution to the points.

However, such a function is difficult to integrate. A simpler approach is shown in figure A3.2. Since the bulk of the built-up areas, of most of the places in the study, lies well within 4 km of the town-centre, we need not worry about the accuracy of the hyperbolic curve beyond this distance.

A3.2.14 If $g(x)$ is the curve graphed in figure A3.2, then $g(x) = cf(x)$ where c is some constant. If $g(x)$ is substituted for $f(x)$ in equation A3.2 it will merely change the scale of the trip length variable, but not the relative magnitudes of the variable.

A3.2.15 Substituting g(x) for f(x) in equation A3.2, we can compute a trip-length factor for each town as

$$TL = \frac{2}{R^2} \int_{0}^{\min(0.6,R)} x(3.23x - 0.46)dx, \text{ for } x \leq 0.6\text{km}$$

$$+ \frac{2}{R^2} \int_{0.6}^{R} x \left(\frac{0.62}{x} + 0.49\right) dx, \text{ for } x > 0.6\text{km}. \qquad (A3.3)$$

A3.2.16 This simplifies to TL = 2.15R -0.46 for towns whose radius is less than or equal to 0.6km, or $TL = (1.24/R) - (0.62/R^2) + 0.49$ for towns whose radius is greater than 0.6km. The curve of TL against R is graphed in figure A3.3. Although many sweeping assumptions have been made in order to arrive at TL, based upon idealised trip-length considerations, TL is merely R transformed into a form that can be more readily incorporated into the modelling/calibration procedure (it has a monotonic relation to CYCLE, which R did not have).

GM/K

APPENDIX 4. REGRESSION/CALIBRATION DETAILS

A4.1 LOG CYCLE REGRESSION

A4.1.1 The following comes from the printout when log CYCLE was regressed on several variables. No case where CYCLE was less than 0.04 was used.

MULTIPLE R	R SQUARE	STD DEVIATION	COEFFICIENT OF VARIABILITY
.779	.607	.425	24.07%

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F	SIGNIFICANCE
REGRESSION	4	40.00	10.00	55.26	.000
RESIDUAL	143	25.88	0.181		

VARIABLES IN THE EQUATION

VARIABLE	B	STD ERROR B	F / SIGNIFICANCE	BETA / ELASTICITY
HILL	-.193	.0199	94.1 / 0	-.56 / .32
RESTD	-.00944	.00153	37.9 / 0	-.38 / .26
RAIN	-.00623	.00229	7.4 /.007	-.16/ .34
LN(TL)	.786	.327	5.8 /.017	.15 / .02
(CONSTANT)	-.104	.219	.2 /.636	

STEP	VARIABLE		R SQUARE
	ENTERED	REMOVED	CHANGE
1	HILL		.365
2	RESTD		.202
3	RAIN		.024
4	LN(TL)		.016

A4.1.2 The R square information above pertains to log CYCLE. The R square for CYCLE with the weighting factor W based upon the above regression/calibration is 0.713.

A4.2 SOCIO-ECONOMIC GROUP REGRESSION

A4.2.1 The following comes from the printout when CYCLE was regressed on the products of W with the socio-economic groups, where W is an estimate of CYCLE based on the previous calibration, together with unweighted INCOME.

MULTIPLE R	R SQUARE	STD DEVIATION	COEFFICIENT OF VARIABILITY
.863	.745	.068	41.66%

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F	SIGNIFICANCE
REGRESSION	4	2.55	.638	136.6	.000
RESIDUAL	187	.874	.00467		

VARIABLES IN THE EQUATION

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
W.SEGM	1.255	.117	115.3 / 0	.61 / .79
W.SEGA	3.882	.730	28.3 / 0	.23 / .09
INCOME	$.537x10^{-4}$	$.178x10^{-4}$	9.1 / .003	.13 / .40
W.SEGN	.618	.220	7.9 / .006	.16 / .21
(CONSTANT)	-.0796	.0231	11.8 / .001	

STEP	VARIABLE		R SQUARE
	ENTERED	REMOVED	CHANGE
1	W.SEGM		.665
2	W. SEGA		.039
3	INCOME		.030
4	W.SEGN		.011

A4.3 LIMITATIONS OF THE TABULATED STATISTICS

A4.3.1 The information on correlation, significances etc which is presented in this appendix was calculated by a computerised multiple linear regression program on assumptions that were broken by the data used. The range of CYCLE is 0 to 1, that of log CYCLE is – to 0, whilst, to satisfy the assumptions, they should be able to range from $-\infty$ to $+\infty$ and have their error term independent of the magnitude of CYCLE or log CYCLE, whichever is being used. Nevertheless the procedure calibrates the multiplicative model, as was required, minimising the squares of the errors.

4.3.2 Since a log CYCLE transformation stretches the scale as CYCLE tends to zero, the log CYCLE regression will have a greater tendency to account for variation at the lower end of the scale, than would a CYCLE regression. This bias seems acceptable as the aim of the study is to determine whether low cycling levels are low because of or despite the risk of accidents.

A.3.3 The calibrated model accounts for 74.5% of the variation in CYCLE levels for this sample. This figure requires no assumptions: it derives from the least-squares fit. However the tabulated significance level etc are suspect. A plot of actual against estimated values (figure 3.3) shows the error magnitudes as not particularly dependent on the magnitude of CYCLE. So the significance level (R not to be rejected at a level of 0.00001) is not likely to be seriously in error. However, because the variables cannot range between $-\infty$ and $+\infty$, extrapolations are more than ever suspect, and the conclusions that are drawn in the text are thus limited to the ranges of the sampled variables.

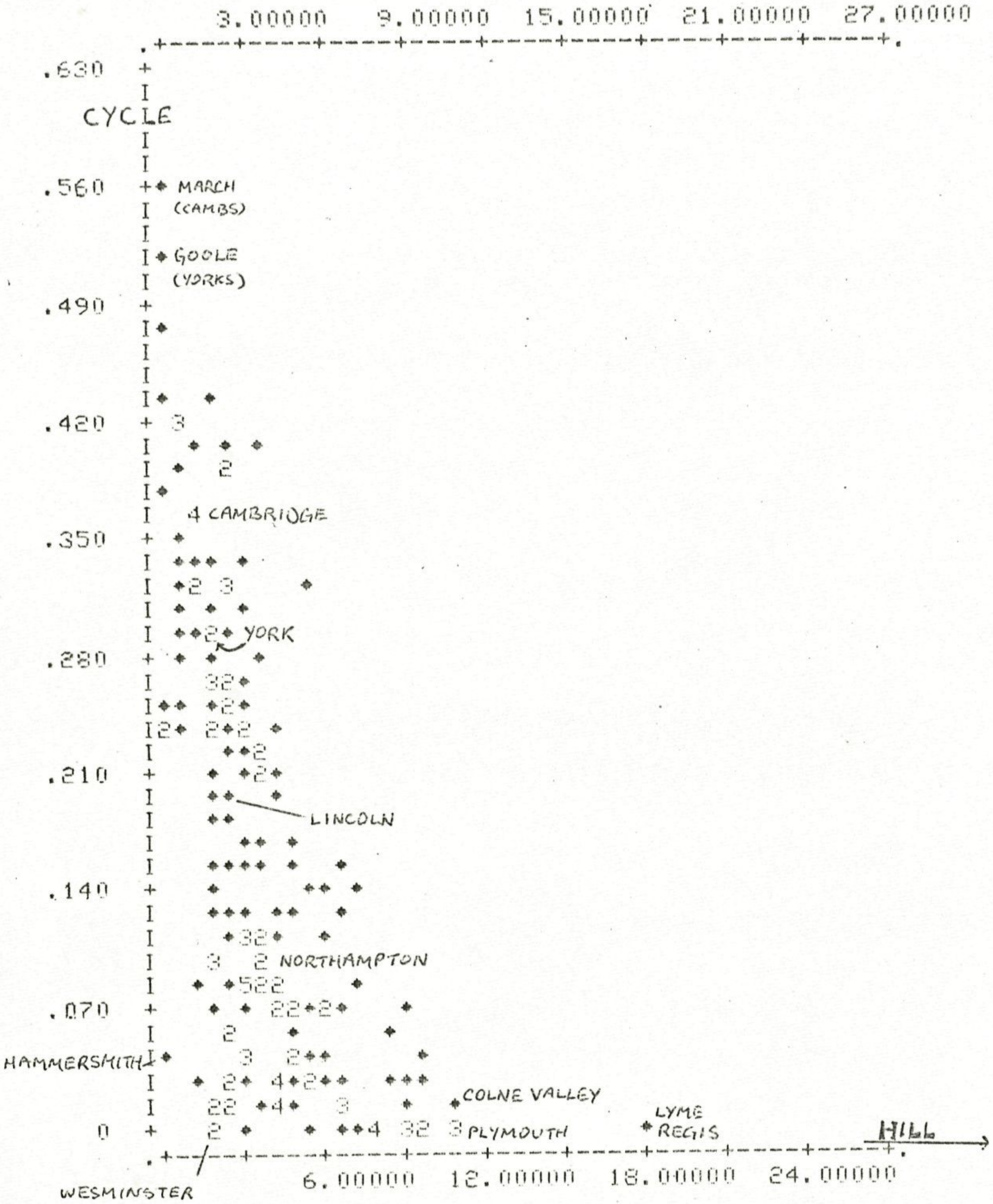


FIGURE 2.1 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "HILL" (THE HILLINESS OF THE BUILT-UP AREA).

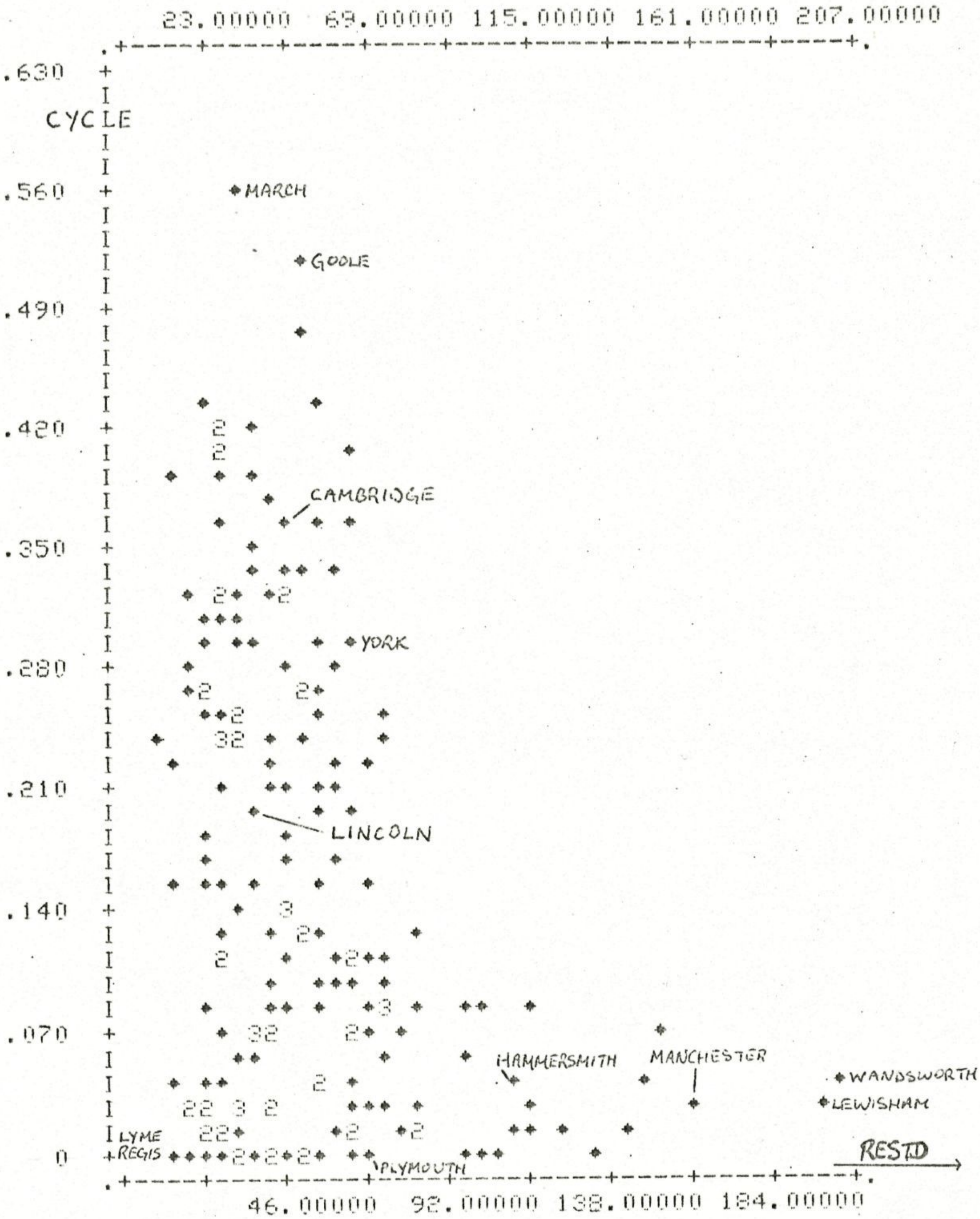


FIGURE 2.2 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "RESTD" (10^4x RE-ESTIMATE OF DANGER).

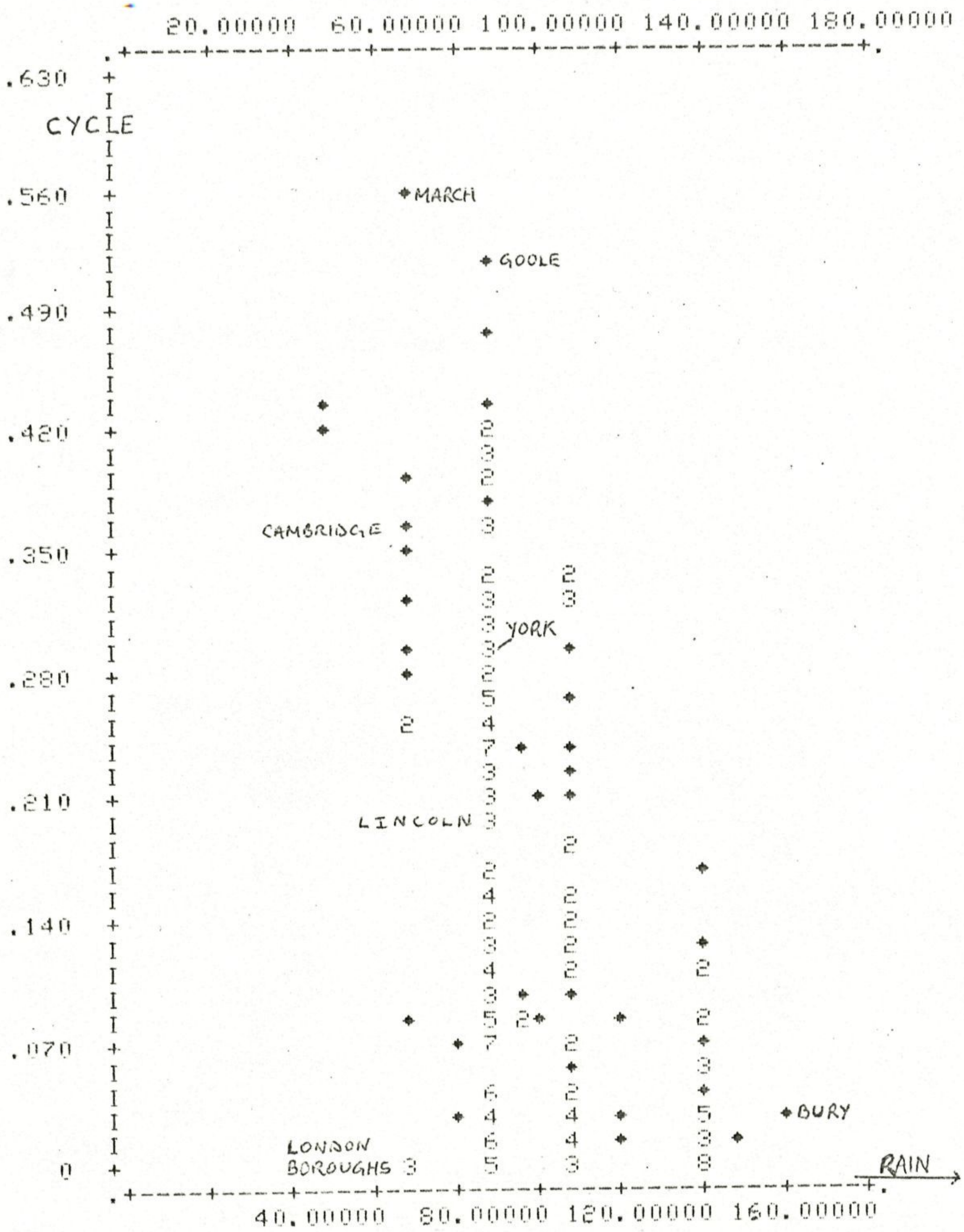


FIGURE 2.3, "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "RAIN" (THE NUMBER OF DAYS, IN 1966, HAVING RAINFALL GREATER THAN 2·5 m.m.).

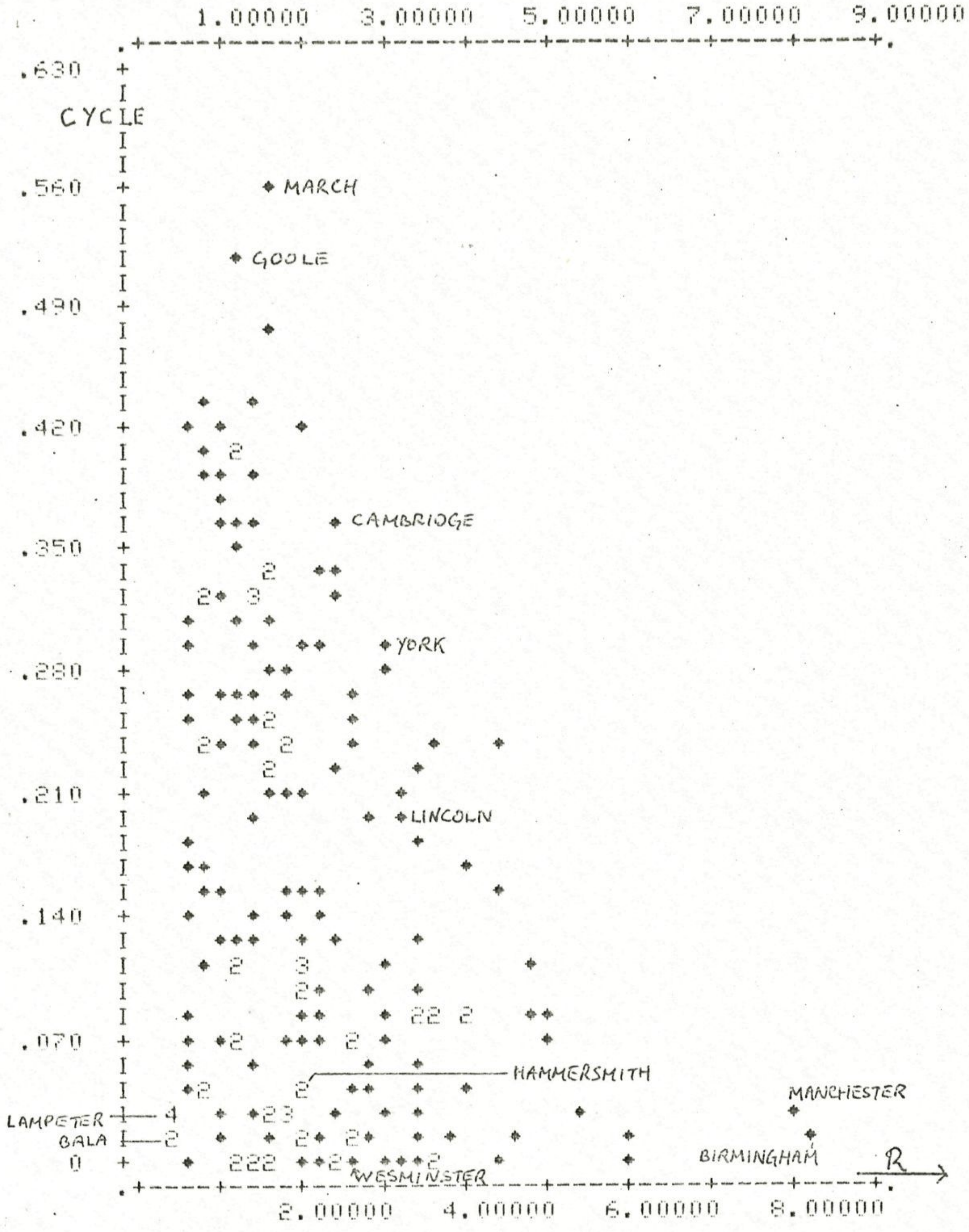


FIGURE 2.4 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "R" (THE APPROXIMATE RADIUS, IN KILOMETERS, OF THE TOWN).

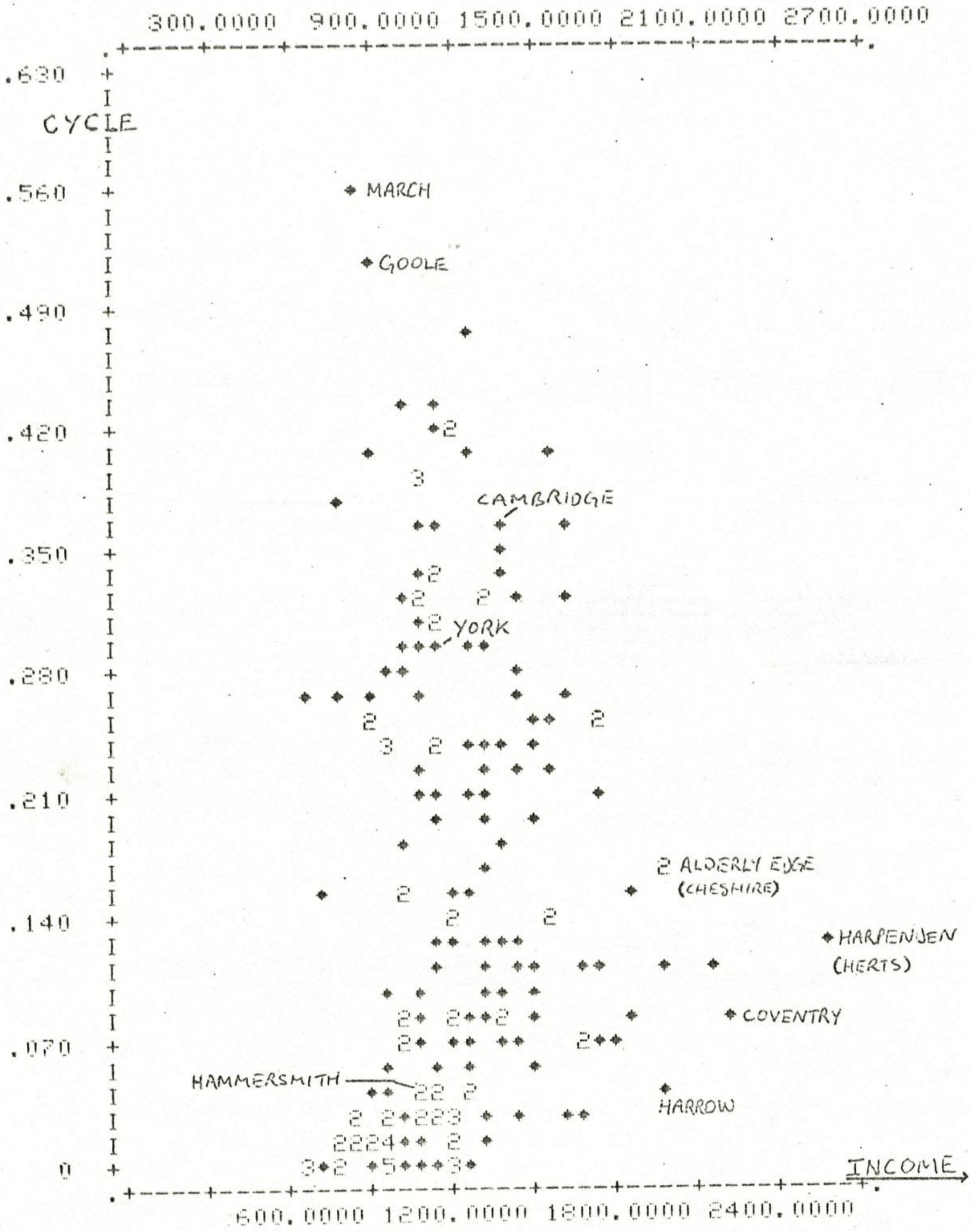


FIGURE 2.5 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "INCOME" (THE ESTIMATED AVERAGE HOUSEHOLD INCOME OF THE AREA FOR 1966).

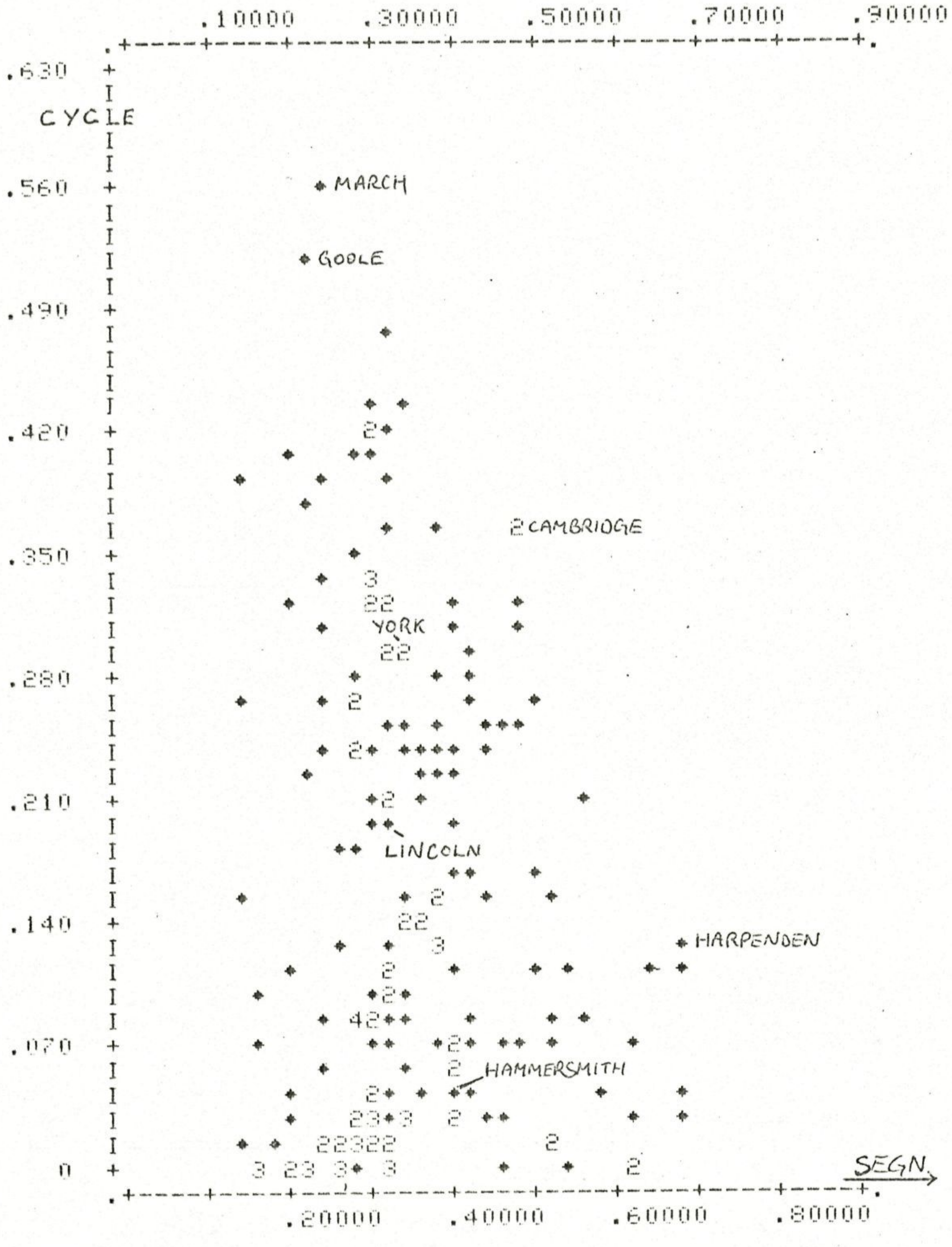


FIGURE 2.6 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "SEGN" (THE PROPORTION OF THE RESIDENT WORKFORCE IN NON-MANUAL OCCUPATIONS).

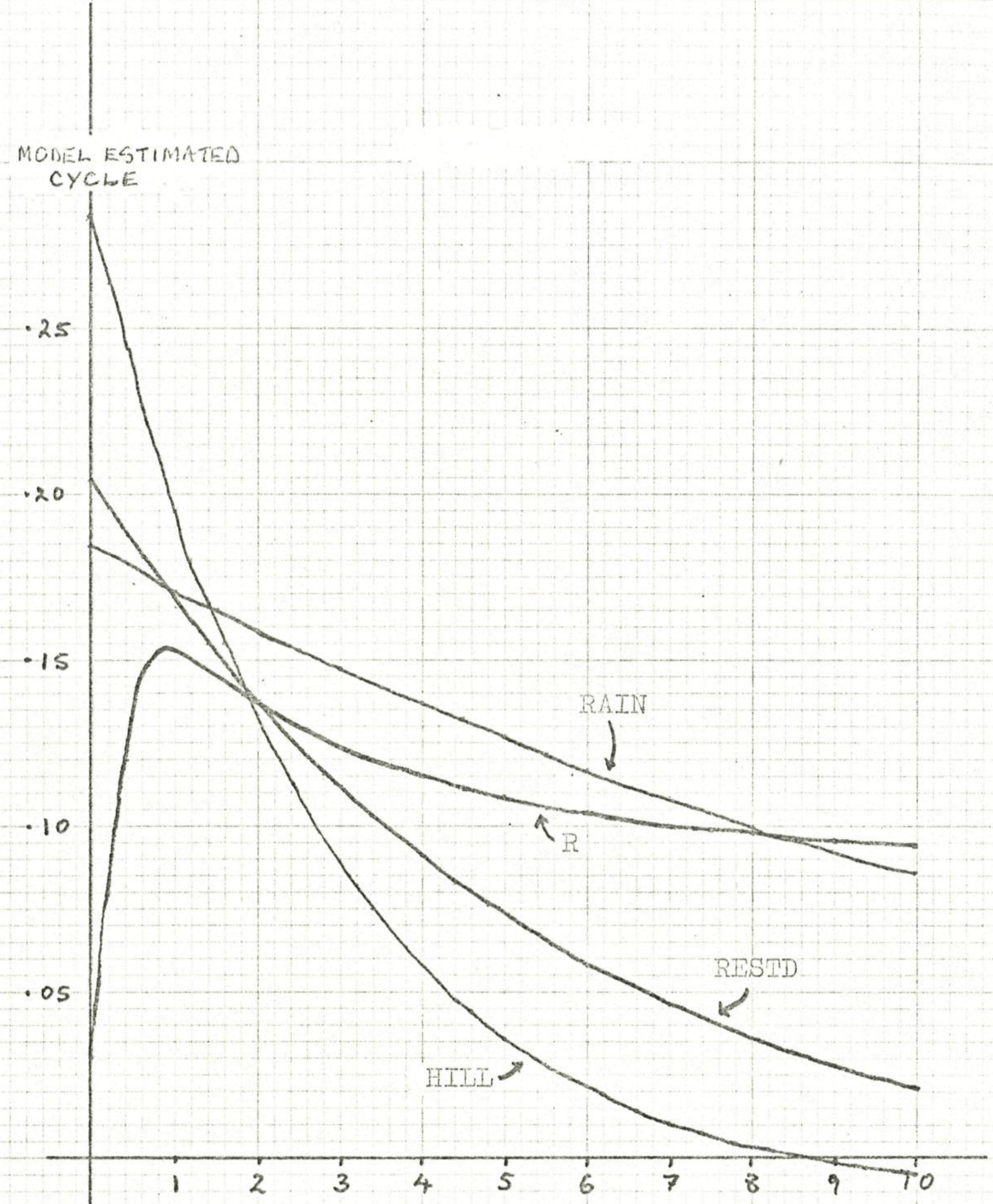


FIGURE 3.1

Graph of model-estimated CYCLE against each of the variables HILL, RESTD, RAIN and R (representing hilliness, danger, rainfall and radius of the built-up area). All variables but for the one which is allowed to vary are held at their mean values. The range of values of each variable has been rescaled from zero to ten.

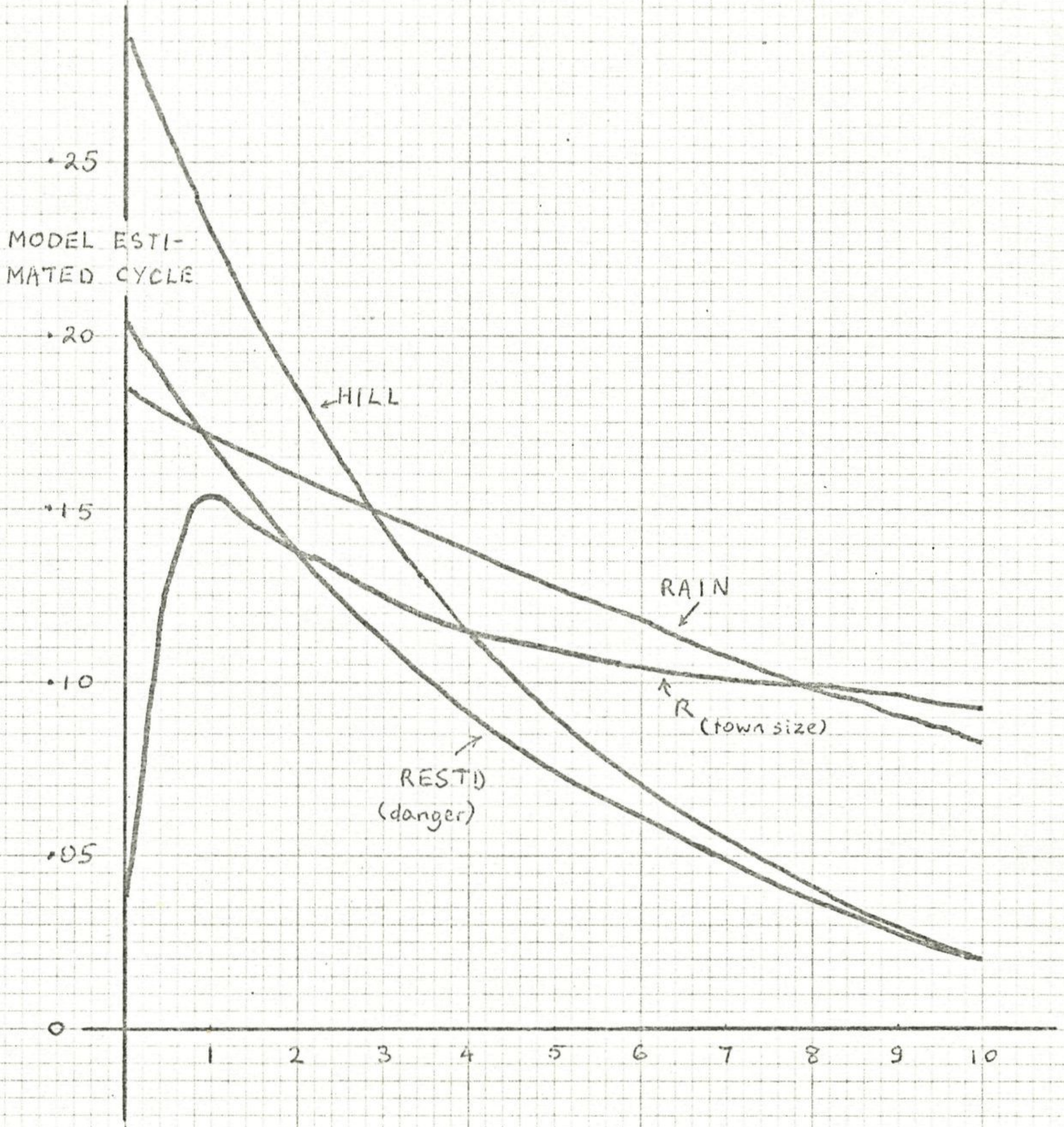


FIGURE 3.2 GRAPHS OF MODEL ESTIMATED "CYCLE" AS A FUNCTION OF VARIOUS FACTORS.

R:	RADIUS OF THE BUILT-UP AREA
RESTD	RE-ESTIMATE OF DANGER BASED UPON BICYCLE ACCIDENT RATES
HILL	NUMBER OF 25 FOOT CONTOUR LINES PER ROAD MILE ON 1:25000 O.S. MAPS
RAIN	NUMBER OF RAINY DAYS
CYCLE	PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK

THE PLOTS ARE AGAINST EACH VARIABLE WHILST ALL OTHER MODELLED VARIABLES ARE HELD CONSTANT AT THEIR MEAN SAMPLE VALUES. THE RANGE OF EACH VARIABLE HAS BEEN RESCALED, ZERO REPRESENTING THE SAMPLE MINIMUM AND TEN THE SAMPLE MAXIMUM (AN EXTREME OUTLIER FOR THE HILL-CURVE HAS BEEN EXCLUDED).

E.R.

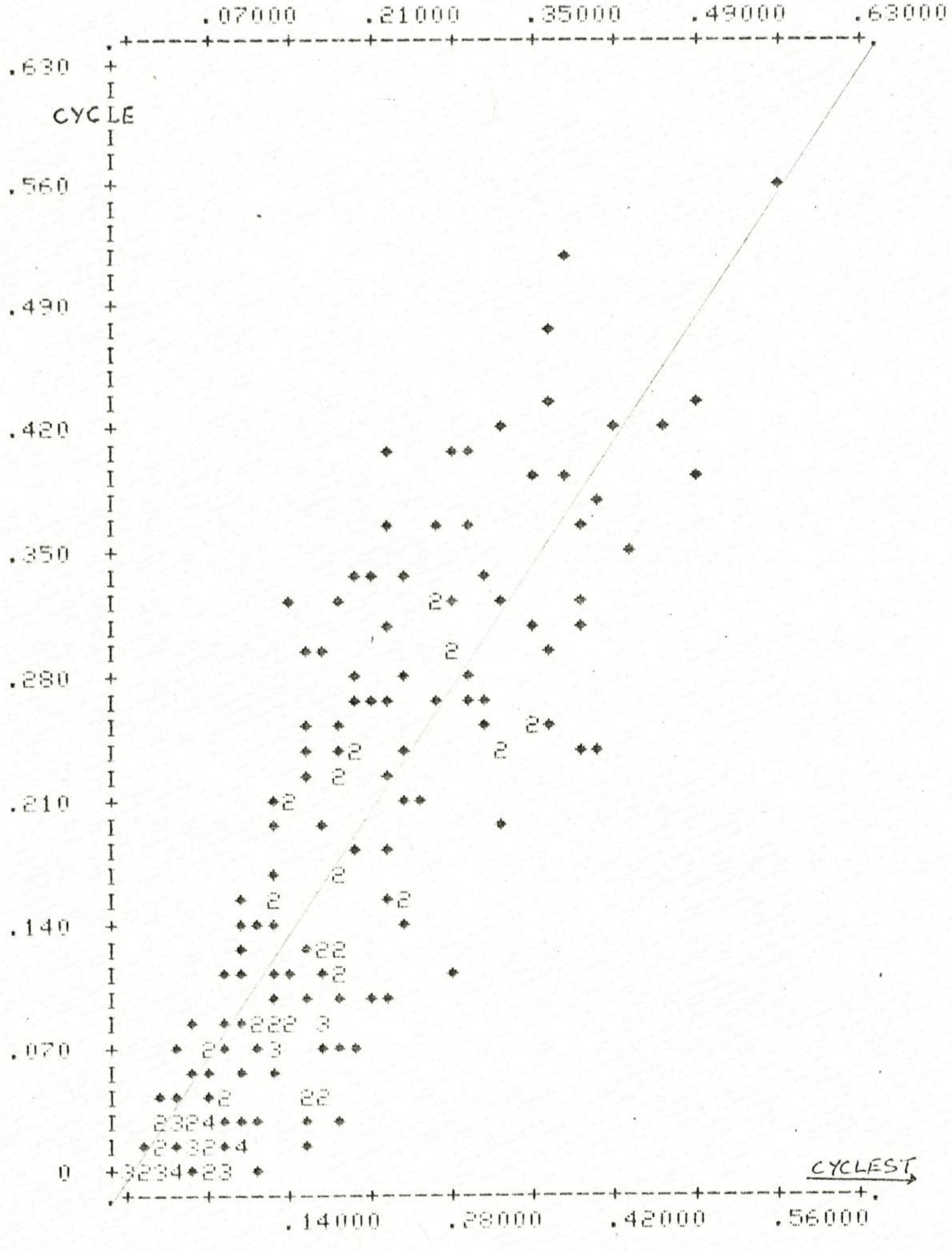
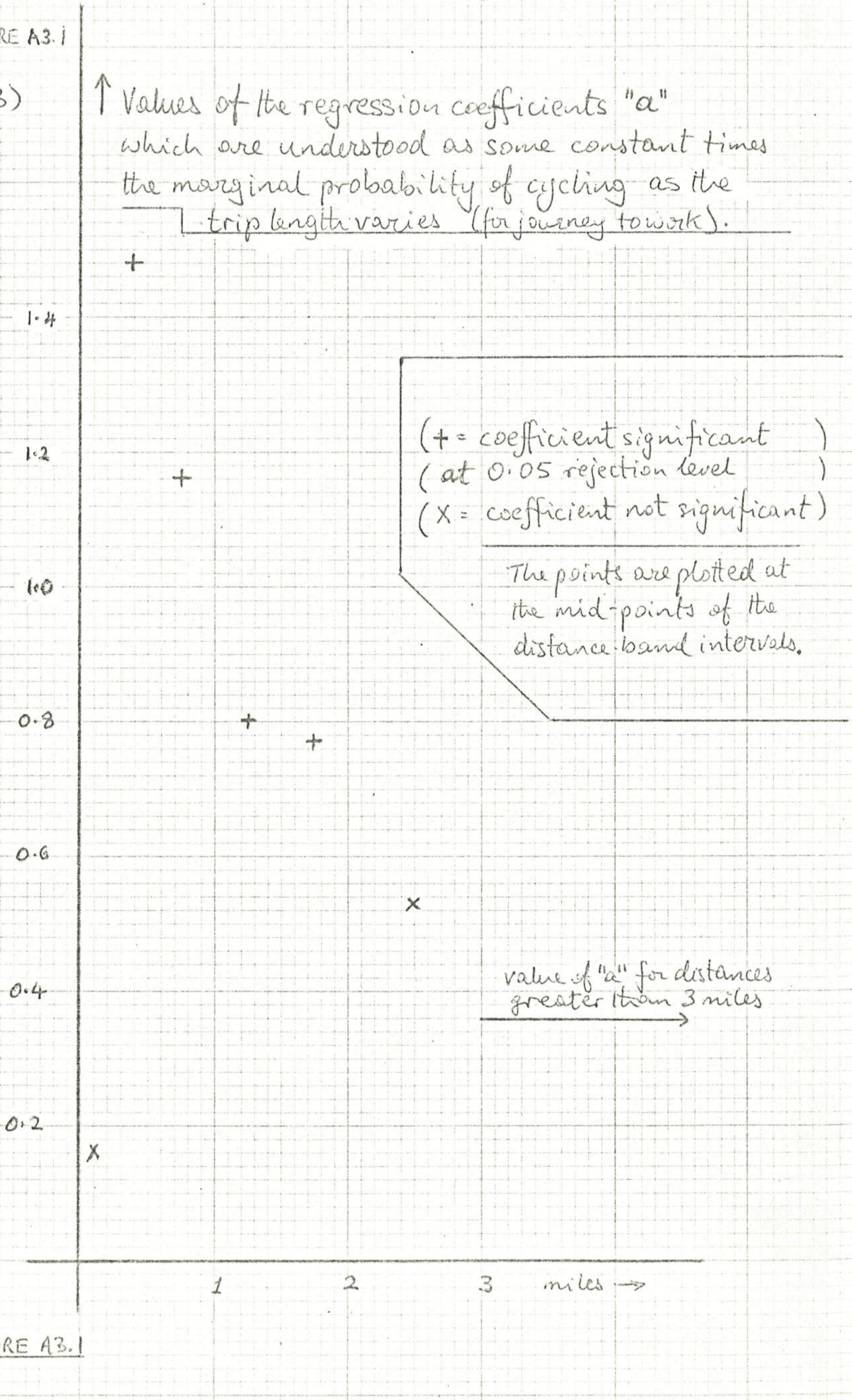


FIGURE 3.3 "CYCLE" (THE PROPORTION OF THE RESIDENT WORKFORCE WHO CYCLE TO WORK) AGAINST "CYCLEST" (MODEL ESTIMATED "CYCLE").

E.R.

FIGURE A3.1

(Appendix 3)



↑ Values of the regression coefficients "a" which are understood as some constant times the marginal probability of cycling as the trip length varies (for journey to work).

(+ = coefficient significant)
(at 0·05 rejection level)
(X = coefficient not significant)

The points are plotted at the mid-points of the distance-band intervals.

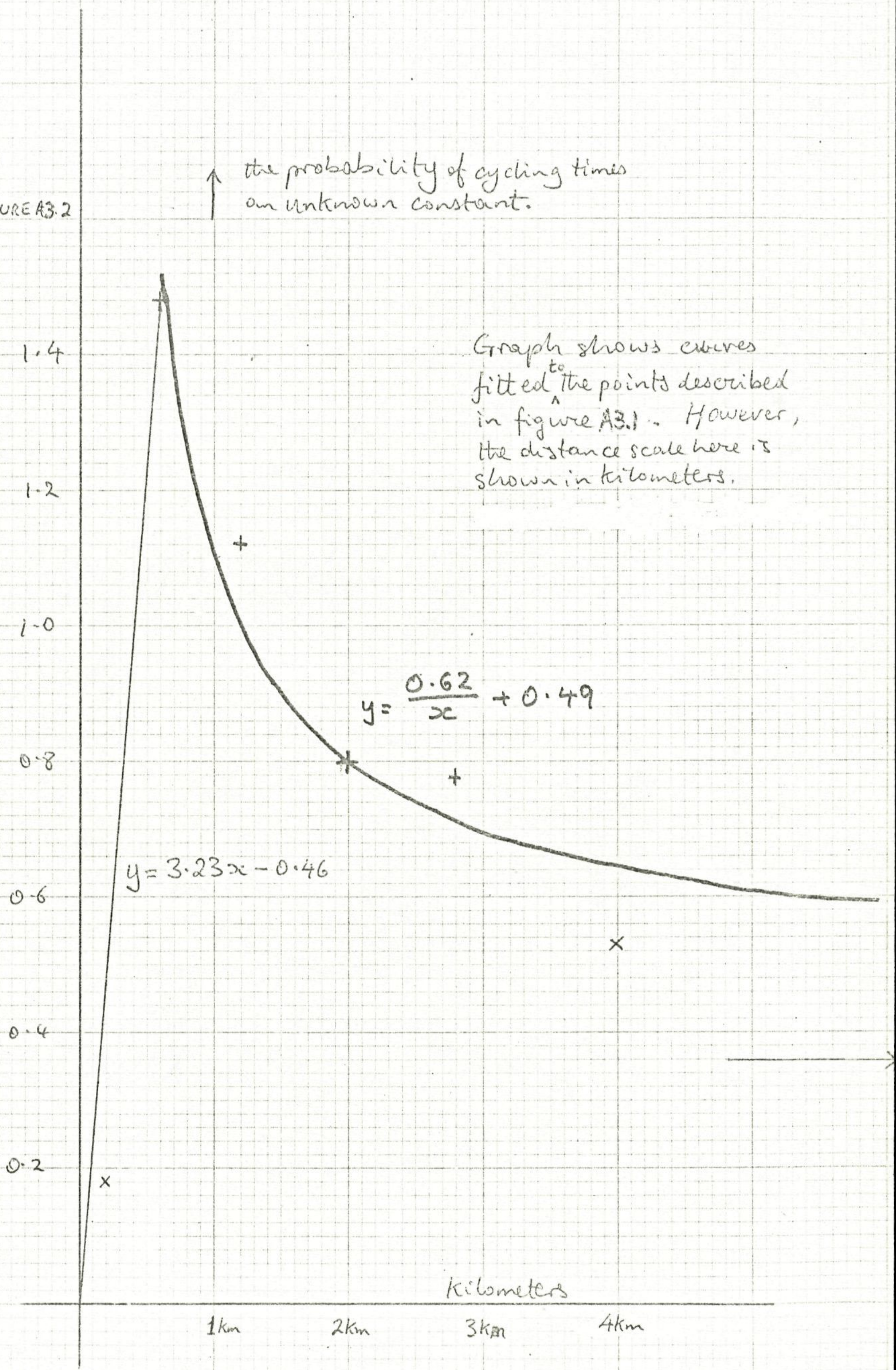
value of "a" for distances greater than 3 miles →

FIGURE A3.1

E.R..

(Appendix 3)

FIGURE A3.2

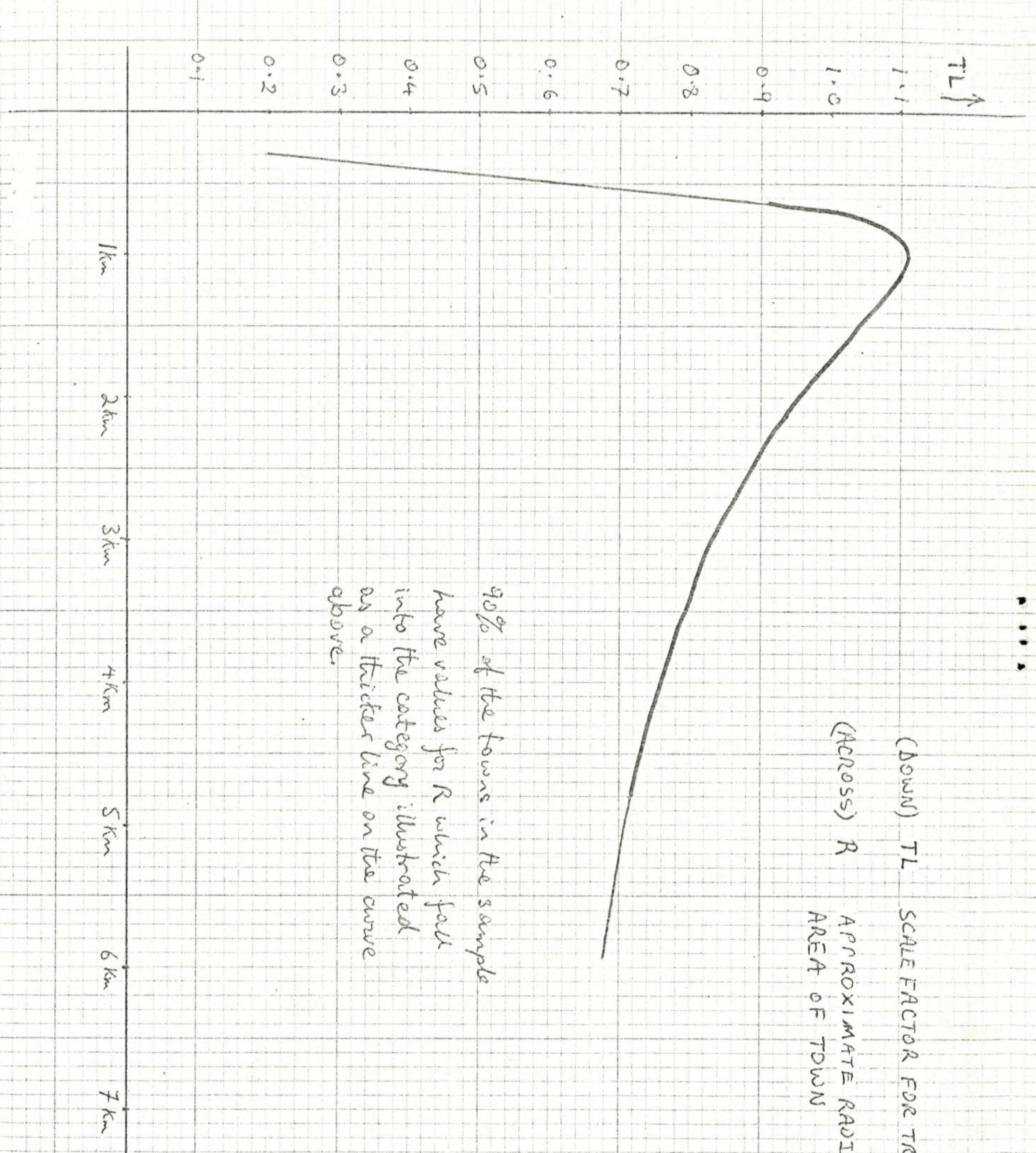


Graph shows curves fitted to the points described in figure A3.1. However, the distance scale here is shown in kilometers.

FIGURE A3.2

FIGURE A3.3

(Appendix 3)



(DOWN) TL SCALE FACTOR FOR TRIP-LENGTH EFFECT

(ACROSS) R APPROXIMATE RADIUS OF BUILT-UP AREA OF TOWN

90% of the towns in the sample have values for R which fall into the category illustrated as a thicker line on the curve above.

FIGURE A3.3

TABLE 3.1

RANGES OF VARIABLES

VARIABLE	MINIMUM	MEAN	MAXIMUM
CYCLE	0	0.16	0.57
HILL	0	3.7	18.2
RAIN	50	100	160
RESTD	9	52	203
TL	0.204	0.929	1.110
R	0.309	2.093	8.118
INCOME	649	1218	2586
SEGN	0.133	0.345	0.682
SEGM	0.295	0.629	0.865
SEGA	0	0.017	0.152

CYCLE is the proportion of people who reside and work in an area who cycle to work. HILL is the number of 25 foot contour changes per road-mile in the built-up area. RESTD is an estimated danger value and is approximately 10^4 times the cyclist accident rate. TL is a trip-length factor derived form R which is an approximate radius measured in kilometers for the built-up area under consideration. INCOME is the National Traffic Model estimate of average income in pounds sterling (1966). The SEG groups are proportions of the population in non-manual, manual and agricultural occupations.

TABLE 3.2.

ILLUSTRATION OF THE JOINT EFFECTS OF HILLINESS AND DANGER ON PEOPLE'S USE OF BICYCLES

TYPE OF TOWN OR BOROUGH	PREDICTED CYCLE LEVEL	EXAMPLES	ACTUAL SHARE OF JOURNEYS TO WORK
HILLY BUT SAFE	4%	Matlock, Derbyshire	4%
		Worsley, Lancashire	4%
		Bodmin, Cornwall	6%
FLAT BUT DANGEROUS	6%	Hammersmith, London	5%
		Liverpool, Lancashire	3%
		Barking, London	9%
HILLY AND DANGEROUS	0%	Sheffield, Yorkshire	1%
		Plymouth, Devon	2%
		Burnley, Lancashire	2%
FLAT AND SAFE	43%	Goole, Yorkshire	52%
		Newark, Nottinghamshire	42%
		Cambridge, Cambridgeshire	36%

Notes: Extreme sample values for HILL and RESTD (hilliness and danger) were used for the four predictions, though the exceptionally hilly Lyme Regis value of 18·2 was not used, the next most hilly value, i.e. 11·0, being used instead. Values of other factors, e.g. RAIN, are at their mean sample values.

The predicted CYCLE level is the percentage of residents who work in the town or borough who would be expected to cycle to work.

The last column shows the actual CYCLE levels, as measured by the 1966 Census, for some towns and boroughs with relatively high or low levels of danger and hilliness.