# STATISTICS QUALIFYING EXAM – TAKE HOME

## Due 4 p.m. Thursday, August 17, 1995, Mathematics and Statistics Department Office

*Directions:* Answer both questions. The answers should be typed (or typeset), double spaced, and the answer to each question should be no more than 3 pages long. An Appendix is allowed but will be examined only at the discretion of the graders. (The better constructed the Appendix, the more likely it is to get examined.)

1. The table below contains pollution data. The data are from various years; they relate air pollution to mortality rates for various standard metropolitan statistical areas in the United States. The dependent variable $y$ is the total age adjusted mortality rate per 100,000 as computed for different metropolitan areas. The predictor variables are, in order, mean annual precipitation in inches, mean January temperature in degrees F, mean July temperature in degrees F, population per household, median school years completed by those over 25, percent of housing units that are sound and with all facilities, population per sq. mile in urbanized areas, percent nonwhite population in urbanized areas, relative pollution potential of sulphur dioxide, annual average of percent relative humidity at 1 pm. Analyze the data. Find a good predictive model for mortality. This should include variable selection and examination of residuals. Discuss the results of the analysis in plain English. Evaluate the potential for collinearity problems.

Pollution Data

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 36 | 27 | 71 | 3.34 | 11.4 | 81.5 | 3243 | 8.8 | 42.6 | 59 | 921.870 |
| 35 | 23 | 72 | 3.14 | 11.0 | 78.8 | 4281 | 3.5 | 50.7 | 57 | 997.875 |
| 44 | 29 | 74 | 3.21 | 9.8 | 81.6 | 4260 | .8 | 39.4 | 54 | 962.354 |
| 47 | 45 | 79 | 3.41 | 11.1 | 77.5 | 3125 | 27.1 | 50.2 | 56 | 982.291 |
| 43 | 35 | 77 | 3.44 | 9.6 | 84.6 | 6441 | 24.4 | 43.7 | 55 | 1071.289 |
| 53 | 45 | 80 | 3.45 | 10.2 | 66.8 | 3325 | 38.5 | 43.1 | 54 | 1030.380 |
| 43 | 30 | 74 | 3.23 | 12.1 | 83.9 | 4679 | 3.5 | 49.2 | 56 | 934.700 |
| 45 | 30 | 73 | 3.29 | 10.6 | 86.0 | 2140 | 5.3 | 40.4 | 56 | 899.529 |
| 36 | 24 | 70 | 3.31 | 10.5 | 83.2 | 6582 | 8.1 | 42.5 | 61 | 1001.902 |
| 36 | 27 | 72 | 3.36 | 10.7 | 79.3 | 4213 | 6.7 | 41.0 | 59 | 912.347 |
| 52 | 42 | 79 | 3.39 | 9.6 | 69.2 | 2302 | 22.2 | 41.3 | 56 | 1017.613 |
| 33 | 26 | 76 | 3.20 | 10.9 | 83.4 | 6122 | 16.3 | 44.9 | 58 | 1024.885 |
| 40 | 34 | 77 | 3.21 | 10.2 | 77.0 | 4101 | 13.0 | 45.7 | 57 | 970.467 |
| 35 | 28 | 71 | 3.29 | 11.1 | 86.3 | 3042 | 14.7 | 44.6 | 60 | 985.950 |
| 37 | 31 | 75 | 3.26 | 11.9 | 78.4 | 4259 | 13.1 | 49.6 | 58 | 958.839 |
| 35 | 46 | 85 | 3.22 | 11.8 | 79.9 | 1441 | 14.8 | 51.2 | 54 | 860.101 |
| 36 | 30 | 75 | 3.35 | 11.4 | 81.9 | 4029 | 12.4 | 44.0 | 58 | 936.234 |
| 15 | 30 | 73 | 3.15 | 12.2 | 84.2 | 4824 | 4.7 | 53.1 | 38 | 871.766 |
| 31 | 27 | 74 | 3.44 | 10.8 | 87.0 | 4834 | 15.8 | 43.5 | 59 | 959.221 |
| 30 | 24 | 72 | 3.53 | 10.8 | 79.5 | 3694 | 13.1 | 33.8 | 61 | 941.181 |
| 31 | 45 | 85 | 3.22 | 11.4 | 80.7 | 1844 | 11.5 | 48.1 | 53 | 891.708 |
| 31 | 24 | 72 | 3.37 | 10.9 | 82.8 | 3226 | 5.1 | 45.2 | 61 | 871.338 |
| 42 | 40 | 77 | 3.45 | 10.4 | 71.8 | 2269 | 22.7 | 41.4 | 53 | 971.122 |
| 43 | 27 | 72 | 3.25 | 11.5 | 87.1 | 2909 | 7.2 | 51.6 | 56 | 887.466 |
| 46 | 55 | 84 | 3.35 | 11.4 | 79.7 | 2647 | 21.0 | 46.9 | 59 | 952.529 |
| 39 | 29 | 75 | 3.23 | 11.4 | 78.6 | 4412 | 15.6 | 46.6 | 60 | 968.665 |
| 35 | 31 | 81 | 3.10 | 12.0 | 78.3 | 3262 | 12.6 | 48.6 | 55 | 919.729 |
| 43 | 32 | 74 | 3.38 | 9.5 | 79.2 | 3214 | 2.9 | 43.7 | 54 | 844.053 |
| 11 | 53 | 68 | 2.99 | 12.1 | 90.6 | 4700 | 7.8 | 48.9 | 47 | 861.833 |
| 30 | 35 | 71 | 3.37 | 9.9 | 77.4 | 4474 | 13.1 | 42.6 | 57 | 989.265 |
| 50 | 42 | 82 | 3.49 | 10.4 | 72.5 | 3497 | 36.7 | 43.3 | 59 | 1006.490 |
| 60 | 67 | 82 | 2.98 | 11.5 | 88.6 | 4657 | 13.5 | 47.3 | 60 | 861.439 |
| 30 | 20 | 69 | 3.26 | 11.1 | 85.4 | 2934 | 5.8 | 44.0 | 64 | 929.150 |
| 25 | 12 | 73 | 3.28 | 12.1 | 83.1 | 2095 | 2.0 | 51.9 | 58 | 857.622 |
| 45 | 40 | 80 | 3.32 | 10.1 | 70.3 | 2682 | 21.0 | 46.1 | 56 | 961.009 |
| 46 | 30 | 72 | 3.16 | 11.3 | 83.2 | 3327 | 8.8 | 45.3 | 58 | 923.234 |
| 54 | 54 | 81 | 3.36 | 9.7 | 72.8 | 3172 | 31.4 | 45.5 | 62 | 1113.156 |
| 42 | 33 | 77 | 3.03 | 10.7 | 83.5 | 7462 | 11.3 | 48.7 | 58 | 994.648 |
| 42 | 32 | 76 | 3.32 | 10.5 | 87.5 | 6092 | 17.5 | 45.3 | 54 | 1015.023 |
| 36 | 29 | 72 | 3.32 | 10.6 | 77.6 | 3437 | 8.1 | 45.5 | 56 | 991.290 |
| 37 | 38 | 67 | 2.99 | 12.0 | 81.5 | 3387 | 3.6 | 50.3 | 73 | 893.991 |
| 42 | 29 | 72 | 3.19 | 10.1 | 79.5 | 3508 | 2.2 | 38.8 | 56 | 938.500 |
| 41 | 33 | 77 | 3.08 | 9.6 | 79.9 | 4843 | 2.7 | 38.6 | 54 | 946.185 |
| 44 | 39 | 78 | 3.32 | 11.0 | 79.9 | 3768 | 28.6 | 49.5 | 53 | 1025.502 |
| 32 | 25 | 72 | 3.21 | 11.1 | 82.5 | 4355 | 5.0 | 46.4 | 60 | 874.281 |

2. Below are 12 observations from an agricultural experiment with 3 variables and a corresponding yield. The variables are:
   $X_1 =$ fertilizer $A$ (0-150 lbs. per plot)
   $X_2 =$ fertilizer $B$ (0-50 lbs. per plot)
   $X_3 =$ farm

| | Farm 1 | | | Farm 2 | | |
|---|---|---|---|---|---|---|
| fert A\fert B | 15 | 20 | 25 | 15 | 20 | 25 |
| 50 | 5.63 | 11.57 | 7.62 | 6.42 | 12.16 | 8.32 |
| 100 | 1.38 | 5.72 | 8.28 | 1.94 | 4.69 | 7.51 |

a) Give an appropriate ANOVA model for these data and justify your choice.
b) Fit your ANOVA model to the data, examine orthogonal polynomial contrasts and residuals. Find a reasonably fitting ANOVA model.
c) Describe what you have learned about the fertilizers from part b.
d) Using only effects that were important in part b, fit a corresponding regression model to the data. This will be a polynomial in two variables, perhaps with cross-product terms.
e) Describe the results of the regression analysis. How do they differ from the results of the ANOVA. Should the assumptions be rechecked?
f) Give your best estimate of the amounts of fertilizers that maximize yield.
g) Ignoring the fact that the amounts of fertilizers in part f are random, i.e., treating them as fixed, give a 95% confidence interval for the mean yield at these levels. Give a 95% prediction interval.