PSY 513 Spring, 1985
Statistics R. Dawes

COURSE DESCRIPTION

The quarter is divided into three sections: (i) two-variable correlation and regression, (ii) multiple correlation and regression, and (iii) analysis of categorical variables. There will be an examination after each section; two hour examinations will deal only with the material presented in the previous section, and a final will be apportioned 66% to the third section, 17% to the first and 17% to the second. Each hour examination will have a maximal score of 50 points, the final 100. A cumulative total of 140 or more earns an A, of 100 or more earns a B.

Homework will be handed out at the beginning of each section and due at the last class meeting before the corresponding examination. Homework performance will positively affect the grade in the event of ambiguity; it will not negatively affect grades unless it is not handed in with "reasonable" effort demonstrated.

SECTION I: Two-variable correlation and regression.

April 4: The correlation coefficient: a descriptive interpretation in standard and raw score form. Readings Cohen and Chen (CC) pgs. 23-43, Dawes, Supplementary Material and Thought Questions (SMTQ) 28-33

April 9: Residuals, partials and covariances. Readings CC, 66-70, 83-91. SMTQ 32-33, 36-39. Dawes, Science letter (1972, 78, 229-230). Meehl, Journal of Abnormal Psychology article (1971, 77, 143-148).

April 11: Regression Effects. Readings CC, 43. SMTQ, 33-36. Dawes 410 handout. Furby Developmental Psychology article (1973, 8, 172-179). Lord, Psychological Bulletin article (1967, 68, 304-305).

April 16: Regression vs. Correlation: reference to fair use of tests. Readings SMTQ, 70-74. Thorndike Journal of Educational Measurement article (1971, 8, 63-67). Darlington, Journal of Educational Measurement article (1971, 8,71-82). Birnbaum American Psychologist article (1981, 36,1193-1195).

April 18: WPA, no class

April 23: Inferences about correlation and review. Readings CC, 48-66. SMTQ, 39-41. Dawes American Statistician letter (1971, 25, 56-57) Dawes memo--3/23/83.

April 25: EXAM

SECTION II: Multiple correlation and regression

April 30: Basic Concepts. Readings CC, 91-104. SMTQ, 50-52. Dawes, WAIS example, 4/23/82.

May 2: Hypothesis Testing. Readings CC, 104-117. Uldorf and Raymond American Psychologist comment (1973, 28, 1140).

May 7: Capitalization or chance: "Shrunken" R^2. Readings CC, 106-117 (again). SMTQ, 52-59. Dawes memo re Schmidt, Gayle and Rauschenberger's 1977 Psychological Bulletin article.

May 9: Ad hoc and unit weights. Readings Schmidt Educational and Psychological Measurement article (1971, 31, 699-714). Dawes "Citation Classic" comment. Dawes American Psychologist article (1979, 34, 571-582). Dawes memo 5/6/83.

May 14: EXAM

SECTION III: Analysis of categorical variables.

May 16: Definition and contingency between pairs of variables. Readings none.

May 21: 2 x 2 Relationships: phi and kappa. Readings CC, 37. Gilula's American Statistician article (1981, 35, 96-97). Dawes handout on phi, kappa, gamma, tau and r, 3/26/85.

May 23: Chi-square, log-likelihood chi-square and weighted kappa. Readings SMTQ, 2-17. Dawes critique of Stern, 5/12/80, and of Visintainer, Volpicelli, and Seigleman, 3/26/83, and of Hollingshead and Redlich 3/27/83.

May 28: Multiway contingency: the special 2 x 2 x 2 case. Readings Dawes, memo of 3/17/82. Gaetner and Bickman *Journal of Personality* article (1971, 20, 218-222). Dawes critique of Gaetner and Bickman, 6/3/83.

May 30: Decision: best guess and utility. Readings Paucker *Annals of Internal Medicine* article (1976, 85, 8-18), Dawes, Ch. 5 of *Rational Choice in an Uncertain World*

June 4: Signal detection and utility. Readings Swets *Science* article (1973, 182, 990-1000)

June 6: Review

June 14: 8:00 A.M. Final Examination

PSY 513 | Spring, 1985
Statistics | R. Dawes

Homework Section I

1. Algebraically derive the mean and variances of the $(x_i - ry_i)$'s, and the $(y_i - rx_i)$'s.

2. Consider the following four groups on a continuum, with 3 observations in each:

1	3	5	7
3	5	7	9
5	7	9	11

a. do a one-way ANOVA.

b. Perform a linear trend with coefficients -3, -1, +1, +3.

c. Compute the sum of squares for linear trend divided by the total sum of squares.

d. Correlate each observation with the linear trend coefficient for its group. That is:

Coeff.	Obs.
-3	1
-3	3
-3	5
-1	3
-1	5
.	.
.	.
.	.

etc.

e. square the correlation in d.

f. What do you conclude about the relationship between trend analysis and correlation coefficients?

3-7. SMTQ pg. 41-48: Numbers 1 (you either get it or you don't; if you don't, don't spend time on it), 9, 11, 16, 17.

8. Suppose you make a prediction that $y_i'' = x_i$ rather than $y_i' = rx_i$.
What is the MSE of prediction? What is the correlation between y" and y? Between y" and y'? How can the correlations of y' and y" with y be the same while the MSE's are different?

9. Suppose patients are tested on variable X prior to therapy and then retested afterwards. Consider and compare three methods of defining improvement.

$X_{time\ 2} - X_{time\ 1}$, $x_{time\ 2} - x_{time\ 1}$, and $x_{time\ 2} - rx_{time\ 1}$ where r is the correlation between time 1 and time 2. Discuss the statement (which I've heard): "our therapy works best with the most disturbed patients." (Lower case x's refer to standard scores.)

10. Under what circumstances can one obtain regression to the mean when predicting X from Y but not when predicting Y from X?

11. The part correlation between x and y partialling out z is the correlation between the x_i's and the $(y_i - r_{yz}z_i)$'s. Derive the formula for this part correlation. What is the relationship of its numerator to that of the partial correlation between x and y partialling out z? What about the denominators?

PSY 513 Spring, 1985

Statistics R. Dawes

Homework Section II

1. Under H_0: $R^2_{pop} = 0$, $E(R^2_{sample}) = \frac{k}{n - 1}$. Show that if $R^2 = \frac{k}{n - 1}$. $\hat{R}^2_{pop}$ estimated by Wherry-Lord equals 0.

2. What does $F = \frac{R^2 (n - k - 1)}{(1 - R^2)k}$ turn out to be? (show work)

3. Let k = 1; why does $E(R^2_{sample}) = 1/(n - 1)$ rather than 0?

Pick <u>two</u> studies in your field that use multiple regression/correlation. Show how the author(s) of each deal(s) with:

4, 7. Capitalization on chance.

5, 8. The significance of weights and R^2.

6, 9. The sensitivity of the weights.

10, 11. Pick two contexts in which you would like to predict values on a variable Y from values on two or more variables X_1, X_2, ...-- or see if such prediction is possible. Which variables would you choose? Why? Be concrete (e.g., not "I would choose the two variables that predict best") concise and thorough in your answer. Avoid platitudes, even mine.

PSY 513
Statistics

Spring, 1985
R. Dawes

Homework: Section III

SMTQ

pgs. 15-17: numbers 3, 4, 5, 7, 8

pgs. 23-27: numbers 2, 3, 4, 5, 6

Consider a study involving small N's by Doob, A.N. (1970). Carthesis and aggression: The effects of hurting one's enemy. Journal of Experimental Research in Personality, 4, 291-296.

Dichotomies: confederate annoys or he doesn't annoy

confederate is punished by losing money to experimenter or he isn't punished

subject zaps confederate with a shock of greater than median length or he doesn't

Pg. 295: "The basic hypothesis was that losing money to the experimenter would have essentially no effect in the no annoy condition but would lower the amount of shock in the annoy condition."

If the "theory" is essentially that I don't have to avenge myself when God (the experimenter) does it for me, I would prefer

"being annoyed in the no punishment (loss) condition will raise the amount of shock but it will not do so in the punishment (loss) condition.

Doob's analysis

A:

	Z	~Z
P	6	14
~P	14	6

~A:

	Z	~Z
P	9	11
~P	11	9

He concludes that since the first chi-square is significant and the second isn't, his hypothesis is supported. But suppose he had constructed his tables as:

Doob's analysis

P:

	Z	~Z
A	6	14
~A	9	11

~P:

	2	~Z
A	14	6
~A	11	9

N.S.

Point is: you can't conclude a three-way interaction from the existance of a "significant" chi-square on one slice and a non-significant one on another slice.

Reanalyze using Goodman's z test.