

The Basics

Multiple Regression (MR)

 Predicting one DV from a set of predictors, the DV should be interval/ratio or at least assumed I/R if using Likert scale for instance

Assumptions

- Y must be normally distributed (no skewness or outliers)
- X's
 - do not need to be normally distributed, but if they are it makes for a stronger interpretation
 - o linear relationship w/ Y
- no multivariate outliers among Xs predicting Y

MV Outliers

- Leverage distance of score from distribution
- Discrepancy misalignment of score and the rest of the data
- Influence change in regression equation with and without value
- Mahalanobis distance and χ2

Homoskedasticity

 Homoskedasticity – variance on Y is the same at all values of X

Multicollinearity/Singularity

- Tolerance = 1 SMC (Squared Multiple Correlation)
 - Low tolerance values <.01 indicate problems
- Condition Index and Variance
 - Condition Index > 30
 - Plus variance on two separate variables
 > .50.

Regression Equation

$$y' = a + b_1 x_1 + b_2 x_2 + b_i x_i$$

- What is a?
- What are the Bs?
- Why y predicted and not y?

Questions asked by Multiple Regression

Can you predict Y given the set of Xs?

- Anova summary table significance test
- R² (multiple correlation squared) variation in Y accounted for by the set of predictors
- Adjusted R² sample variation around R² can only lead to inflation of the value. The adjustment takes into account the size of the sample and number of predictors to adjust the value to be a better estimate of the population value.
- R² is similar to η² value but will be a little smaller because R² only looks at linear relationship while η² will account for non-linear relationships.

Is each X contributing to the prediction of Y?

- Test if each regression coefficient is significantly different than zero given the variables standard error.
 - o T-test for each regression coefficient,

Which X is contributing the most to the prediction of Y?

 Cannot interpret relative size of Bs because each are relative to the variables scale but Betas (standardized Bs) can be interpreted.

 $y' = a + b_1 x_1 + b_2 x_2 + b_3 x_3$

$$Zy' = beta_1(Zx_1) + beta_2(Zx_2) + beta_3(Zx_3)$$

 a being the grand mean on y a is zero when y is standardized

Can you predict future scores?

- Can the regression be generalized to other data
- Can be done by randomly separating a data set into two halves,
- Estimate regression equation with one half
- Apply it to the other half and see if it predicts

Sample size for MR

There is no exact number to consider

- Need enough subjects for a "stable" correlation matrix
- T and F recommend 50 + 8m, where m is the number of predictors
 - If there is a lot of "noise" in the data you may need more than that
 - If little noise you can get by with less
- If interested generalizing prediction than you need at least double the recommended subjects.

GLM and Matrix Algebra Ch. 17 and Appendix A (T and F)

General Linear Model

$$y = a + b_1 x_{1i} + b_2 x_{2i} + b_j x_{ji} + \varepsilon$$

Think of it as a generalized form of regression

Regular old linear regression

• With four observations $\hat{y} = b_1 x_{1i} + a + \varepsilon_i$ is:

 $\begin{array}{rcl} \hat{y}_{1} &=& bx_{11} \;+\; a + \varepsilon_{1} \\ \hat{y}_{2} &=& bx_{12} \;+\; a + \varepsilon_{2} \\ \hat{y}_{3} &=& bx_{13} \;+\; a + \varepsilon_{3} \\ \hat{y}_{4} &=& bx_{14} \;+\; a + \varepsilon_{4} \end{array}$

Regular old linear regression

- We know y and we know the predictor values, the unknown is the weights so we need to solve for it
- This is where matrix algebra comes in

Regression and Matrix Algebra

The above can also be conceptualized as:

$$\begin{vmatrix} y_{1} \\ y_{2} \\ y_{3} \\ y_{4} \end{vmatrix} = \begin{vmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \end{vmatrix} * |b| + \begin{vmatrix} \varepsilon_{1} \\ \varepsilon_{2} \\ \varepsilon_{3} \\ \varepsilon_{4} \end{vmatrix}$$
$$|4x1| = |4x1| * |1x1|$$
$$|data| = |Design| * |Coefficients|$$

But what's missing?

Regression and Matrix

 We can bring the intercept in by simply altering the design matrix with a column of 1s:

 $|y_1|$ $|1 x_{11}|$ \mathcal{E}_1 y_2 $1 x_{12}$ |a|ε2 $* \begin{vmatrix} a \\ b \end{vmatrix} +$ = note: a is often called b₀ y_3 $1 x_{13}$ ε_3 ε_4 y_4 $1 x_{14}$ |4x1| = |4x2| * |2x1| + |4x1||data| = |Design| * |Coefficients| + |Error|V - (X * B) + F

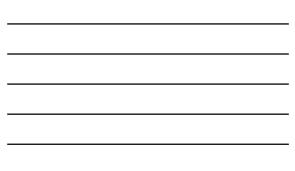
$$\begin{array}{c} \mathbf{I} - (\mathbf{A} \quad \mathbf{D}) + \mathbf{L} \\ 4x1 \quad 4x2 \quad 2x1 \quad 4x1 \end{array}$$

Column of 1s???

- In regression, if you regress an outcome onto a constant of 1s you get back the mean of that variable.
- This only works in the matrix approach, if you try the "by hand" approach to regressing an outcome onto a constant you'll get zero
- Computer programs like SPSS, Excel, etc. use the matrix approach

| E | xample |) | | |
|------|--------|--------------|-------|----------------|
| Case | Y | х | CONST | |
| А | 46 | 90 | 1 | Here is some |
| В | 40 | 78 | 1 | example data |
| С | 32 | 57 | 1 | taken from |
| D | 42 | 65 | 1 | |
| E | 61 | 80 | 1 | some of |
| F | 46 | 75 | 1 | Kline's slides |
| G | 32 | 45 | 1 | |
| н | 51 | 91 | 1 | |
| 1 | 55 | 67 | 1 | |
| | | | | |
| Mean | 45 | 72 | 1 | |
| SD | 9.785 | 15.091 | 0 | |

| [E> | kam | ole - | norr | nal | | | | |
|-------------------|--------------|----------------|-------------|-------------|----------------|-------------|--------------|-------------|
| SUMMARY OUTPU | г | | | | | | | |
| Regression S | tatistics | | | | | | | |
| Multiple R | 0.634852293 | | | | | | | |
| R Square | 0.403037433 | | | | | | | |
| Adjusted R Square | 0.317757067 | | | | | | | |
| Standard Error | 8.082373467 | | | | | | | |
| Observations | 9 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 1 | 308.726674 | 308.726674 | 4.726028384 | 0.066230335 | | | |
| Residual | 7 | 457.273326 | 65.32476086 | | | | | |
| Total | 8 | 766 | | | | | | |
| | | | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 15.3622393 | 13.89683242 | 1.105448985 | 0.305505467 | -17.49854766 | 48.22302625 | -17.49854766 | 48.22302625 |
| х | 0.411635565 | 0.189349746 | 2.17394305 | 0.066230335 | -0.036105437 | 0.859376568 | -0.036105437 | 0.859376568 |



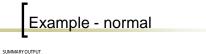
Example – constant only

| SUMMARY | OUTPUT |
|---------|--------|

| Regression Statistics | | | | | | | | |
|-----------------------|-------------|--|--|--|--|--|--|--|
| Multiple R | 0.979624982 | | | | | | | |
| R Square | 0.959665105 | | | | | | | |
| Adjusted R Square | 0.834665105 | | | | | | | |
| Standard Error | 9.785192895 | | | | | | | |
| Observations | 9 | | | | | | | |

| | df | SS | MS | F | Significance F |
|------------|----|-------|-------|-------------|----------------|
| Regression | 1 | 18225 | 18225 | 190.3394256 | 2.4809E-06 |
| Residual | 8 | 766 | 95.75 | | |
| Total | 9 | 18991 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|-----------|--------------|----------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Intercept | 0 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| CONST | 45 | 3.261730965 | 13.79635552 | 7.35719E-07 | 37.47843491 | 52.52156509 | 37.47843491 | 52.52156509 |



| Regression Statistics | | | | | | | |
|-----------------------|-------------|--|--|--|--|--|--|
| Multiple R | 0.987887431 | | | | | | |
| R Square | 0.975921577 | | | | | | |
| Adjusted R Square | 0.82962466 | | | | | | |
| Standard Error | 8.082373467 | | | | | | |
| Observations | 9 | | | | | | |

| ANUVA | | | | | | |
|------------|----|---|-------------|-------------|-------------|----------------|
| | df | | SS | MS | F | Significance F |
| Regression | | 2 | 18533.72667 | 9266.863337 | 141.8583584 | 8.88246E-06 |
| Residual | | 7 | 457.273326 | 65.32476086 | | |
| Total | | 9 | 18991 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|-----------|--------------|----------------|-------------|-------------|--------------|-------------|--------------|-------------|
| Intercept | 0 | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A | #N/A |
| х | 0.411635565 | 0.189349746 | 2.17394305 | 0.066230335 | -0.036105437 | 0.859376568 | -0.036105437 | 0.859376568 |
| CONST | 15.3622393 | 13.89683242 | 1.105448985 | 0.305505467 | -17.49854766 | 48.22302625 | -17.49854766 | 48.22302625 |

Solving for Bs Matrix Style

- In regular algebra you can solve for b
 y = x * b -> y/x = b
- Another way to divide is to multiply by 1/x (inverse)
- Can't just divide in matrix algebra you must multiply by inverse of the X matrix:

B = Y * X⁻¹

 But in order to take the inverse the X matrix must be a square matrix (which rarely, if ever, happens)

Pseudoinverse

- The pseudoinverse A pseudoinverse is a matrix that acts like an inverse matrix for non-square matrices:
- Example: The Moore-Penrose pseudoinverse
- Apseudo-1=(A'A)-1 *A'
- The inverse of (A'A) is defined because it is square (so long as A is full rank, in other words no multicollinearity/singularity)

Estimation

- The goal of an estimator is to provide an estimate of a particular statistic based on the data
- There are several ways to characterize estimators

B Estimators

- bias
 - an unbiased estimator converges to the true value with large enough sample size
 - Each parameter is neither consistently over or under estimated

B Estimators

- likelihood
 - the maximum likelihood (ML) estimator is the one that makes the observed data most likely
 - ML estimators are not always unbiased for small N

B Estimators

- variance
 - an estimator with lower variance is more efficient, in the sense that it is likely to be closer to the true value over samples
 - the "best" estimator is the one with minimum variance of all estimators