Multiple Regression Models

DASC 512

Multiple Regression Models

Multiple regression expands upon simple linear regression to include multiple independent variables to predict the response variable

The general form is now

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Multiple Regression Models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Interpretation remains consistent

- β_0 is still the y-intercept (when $x_1, x_2, \ldots, x_k = 0$)
- β_i is still the effect of independent variable x_i on response variable y

These models can also allow for "higher-order" relationships, e.g.,

$$x_2 = (x_1)^2$$
$$x_3 = \ln(x_1)$$

Multiple Regression Models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Deterministic portion of this model is

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

The random error component ϵ still has the same assumptions

1. The mean of ϵ is zero and constant
2. Variance of ϵ is equal to σ^2 and constant
3. The distribution of $\epsilon \sim N(0, \sigma^2)$
4. Each observed ϵ_i is independent (iid)

Six Steps to Multiple Regression Modeling

1. Hypothesize the deterministic component
2. Use sample data to estimate unknown parameters
3. Estimate the standard deviation of random error term
4. Check assumptions on error term and modify model as needed
5. Statistically evaluate the utility of the model
6. When satisfied that the model is useful, apply it

Method of Least Squares

Still using Ordinary Least Squares – we'll examine alternatives later

- Recall that OLS minimizes the squared errors in predictions

$$SSE = \sum(y - \hat{y})^2$$

Solving for each parameter estimate (β_i) is hard to compute manually

- For k predictor variables, we have to solve a system of $k - 1$ equations

We cannot do this effectively without linear algebra, so we'll rely on Python to estimate our parameters

Example – Patient Satisfaction

1. **Hypothesize the deterministic component**
2. Use sample data to estimate unknown parameters
3. Estimate the standard deviation of random error term
4. Check assumptions on error term and modify model as needed
5. Statistically evaluate the utility of the model
6. When satisfied that the model is useful, apply it

Example – Patient Satisfaction

1. Hypothesize the deterministic component
2. **Use sample data to estimate unknown parameters**
3. Estimate the standard deviation of random error term
4. Check assumptions on error term and modify model as needed
5. Statistically evaluate the utility of the model
6. When satisfied that the model is useful, apply it

Example – Patient Satisfaction

1. Hypothesize the deterministic component
2. Use sample data to estimate unknown parameters
3. **Estimate the standard deviation of random error term**
4. Check assumptions on error term and modify model as needed
5. Statistically evaluate the utility of the model
6. When satisfied that the model is useful, apply it

Example – Patient Satisfaction

1. Hypothesize the deterministic component
2. Use sample data to estimate unknown parameters
3. Estimate the standard deviation of random error term
4. **Check assumptions on error term and modify model as needed**
5. Statistically evaluate the utility of the model
6. When satisfied that the model is useful, apply it

Assessing Model Utility

Just like in simple regression, we can compute r and R^2

- r is not useful for multiple predictor variables
- R^2 is still the ratio of the explained sample variability to the total variability

$$R^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

- Interpretation of R^2 holds – it is **the proportion of total variability in y that is explained by the model**

Limitations of R^2

As predictors are added to a model, SSE can only decrease (no matter how bad those predictors are)

Thus, R^2 increases whenever a predictor is added.

$$R^2 = \frac{SS_{yy} - SSE}{SS_{yy}}$$

Limitations of R^2

We instead can use Adjusted R^2 for multiple regression to penalize the model for adding variables that don't explain much variability in y

$$R_a^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right)\frac{SSE}{SS_{yy}} = 1 - \left(\frac{n-1}{n-(k+1)}\right)(1 - R^2)$$

R_a^2 has the same interpretation as R^2

F-test

Hypotheses:

H_0: None of the independent variables explain any variability in y

(Alternatively, $\beta_1 = \beta_2 = \cdots = \beta_k = 0$)

H_a: At least one of the independent variables explain some variability in y

(Alternatively, at least one of $\beta_i \neq 0, i = 1, \dots, k$)

F-test

Test statistic

$$F = \frac{\frac{SS_{yy} - SSE}{k}}{\frac{SSE}{n-(k+1)}} = \frac{MSR}{MSE} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-(k+1)}}$$

Rejection Region:

$$F > F_{ISF}\left(q = \alpha, dfn = k, dfd = \left(n-(k+1)\right)\right)$$

P-value:

$$F_{SF}\left(F, dfn = k, dfd = \left(n-(k+1)\right)\right)$$

Other Assessments of Fit

When comparing models, one can use R_a^2 as a comparison, but there are others:

- Log Likelihood $(\ln(L))$: A measure of information lost. Should not be used for multiple regression for the same reason as R^2. Larger is better.
- Akaike Information Criterion (AIC $= 2k - \ln(L)$): Modifies log likelihood to penalize for extra parameters. Smaller is better.
- Bayesian Information Criterion (BIC $= k\ln(n) - 2\ln(L)$): Like AIC, but stronger penalty for extra parameters. Smaller is better.

Assessments of Fit

Typically they will all agree on the best model.

- R_a^2: Least punishment for extra variables
  - If they disagree, this will result in largest model of the three
- AIC: Punishes extra variables more than R_a^2 but less than BIC
  - If all data is normal, this is equivalent to R_a^2
  - This is a moderate approach, trading off between completeness and parsimony
- BIC: Punishes extra variables more than AIC and R_a^2
  - This will result in the most parsimonious model of the three

Example – Patient Satisfaction

1. Hypothesize the deterministic component
2. Use sample data to estimate unknown parameters
3. Estimate the standard deviation of random error term
4. Check assumptions on error term and modify model as needed
5. **Statistically evaluate the utility of the model**
6. When satisfied that the model is useful, apply it

Inference about β_i

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

For first order models, interpretation of β_i is the same

- β_0 is the predicted value when $x_1 = x_2 = \cdots = 0$. Likely has no meaning.
- β_i is the relative change in y for a unit change in x_i, all else held constant
- These relationships are still limited to the range of sample data, but now you have multiple dimensions to consider

Inference about β_i

Assumptions:

- All required assumptions for ϵ

Hypotheses:

$$H_0: \beta_i = 0, \qquad H_a: \beta_i < 0, \beta_i > 0, \text{or } \beta_i \neq 0$$

Test statistic:

$$t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$$

Rejection region, p-value, and confidence interval same as for simple regression except degrees of freedom $\nu = n - (k + 1)$

Inference about β_i

An additional consideration: If you are conducting **simultaneous inferences**, you must adjust your alpha level (Bonferroni)

This would affect both hypothesis tests and confidence intervals

If you are talking about each parameter independently, this is not required

Estimation/Prediction Intervals

These can no longer be easily plotted (plotting in 3+ dimensions is hard), but we can still calculate them in the same way.

Example – Patient Satisfaction

1. Hypothesize the deterministic component
2. Use sample data to estimate unknown parameters
3. Estimate the standard deviation of random error term
4. Check assumptions on error term and modify model as needed
5. Statistically evaluate the utility of the model
6. **When satisfied that the model is useful, apply it**

Next time…

Higher-Order Models