Lecture 6: Multiple Linear Regression

Nancy R. Zhang

Statistics 203, Stanford University

January 25, 2010

Multiple Linear Regression

Design matrix:

$$X = \begin{pmatrix} X_{01} & X_{11} & \cdots & X_{p,1} \\ X_{02} & X_{12} & & X_{p,2} \\ \vdots & \ldots & \ddots & \vdots \\ X_{0n} & X_{1n} & & X_{p,n} \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Squared error loss function:

$$L(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} \beta_j X_{ji} \right)^2 .$$

In matrix notation:

$$L(\beta) = (y - X\beta)'(y - X\beta).$$

Multiple Linear Regression

Projection Matrices:

$$P_X = X(X'X)^{-1}X' \quad \text{Projects onto column space of } X.$$

$$P_{X^\perp} = I - P_X \quad \text{Projects onto null space of the column space of } X.$$

Least squares solution:

$$\text{Prediction:} \quad \hat{y} = P_X y.$$

$$\text{Parameters:} \quad \hat{\beta} = (X'X)^{-1}X'y.$$

$$\text{Residuals:} \quad r = y - \hat{y} = P_{X^\perp} y.$$

Goodness of fit

Sums of squares

$$SSE = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2$$

$$SSR = \sum_{i=1}^{n}(\overline{Y} - \widehat{Y}_i)^2 = \sum_{i=1}^{n}(\overline{Y} - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2$$

$$SST = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = SSE + SSR$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

$R = \sqrt{R^2}$ is called the multiple correlation coefficient.
R^2 is large: a lot of the variability in $\boldsymbol{Y}$ is explained by $\boldsymbol{X}$.

F-tests for R^2

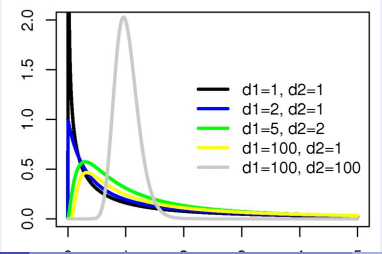
Assume model has intercept (design matrix has p columns).

$$F = \frac{SSR/(p)}{SSE/(n-p-1)}$$

F-distribution

If $W \sim \chi^2_q$ is independent of $Z \sim \chi^2_r$, then

$$\frac{W/q}{Z/r} \sim F_{q,r}.$$



F-Table

Source	Sum of Squares	d.f.	Mean Square	F
Regression	SSR	p	$MSR = \frac{SSR}{p}$	$F = \frac{MSR}{MSE}$
Residuals	SSE	$n - p - 1$	$MSE = \frac{SSE}{n-p-1}$	

Reject at level α if $F > F(p, n - p - 1, \alpha)$.

This tests the hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$.

Nested models

Test the hypothesis that a *subset* of β_i's are zero:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_r = 0.$$

That is, we have the model

$$RM : \quad Y = \beta_{r+1} X_{r+1} + \cdots + \beta_p X_p + \text{error}$$

nested within

$$FM : \quad Y = \beta_1 X_1 + \cdots + \beta_p X_p + \text{error}$$

.

Does $X_1, \ldots, X_r$ have a significant marginal effect, after adjusting for the other predictors?

Nested models

$$RM: \quad Y = \beta_{r+1}X_{r+1} + \cdots + \beta_p X_p + \text{error}$$

$$FM: \quad Y = \beta_1 X_1 + \cdots + \beta_p X_p + \text{error}$$

.

$$\Delta df = df(FM) - df(RM)$$

$$F = \frac{[SSE(RM) - SSE(FM)]/[\Delta df]}{SSE(FM)/[n - df(FM)]}.$$

$$F \sim F_{\Delta df, n - df(FM)}.$$

Testing Constraints

In some situations you want to test that your model parameters satisfy some constraint. Say you have model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error},$$

and want to test:

$$H_0 : \beta_1 = \beta_2.$$

This is equivalent to the model:

$$Y = \beta_0 + \beta_1 (X_1 + X_2) + \text{error}.$$

.

Fit these two models, apply F test with $\Delta df = 1$.

Testing Constraints – Another example

Another example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error},$$

and want to test:

$$H_0 : \beta_1 + \beta_2 = 1.$$

This is equivalent to the model:

$$Y = \beta_0 + \beta_1 X_1 + (1 - \beta_1) X_2 + \text{error}.$$

.
which can be simplified to:

$$Y - X_2 = \beta_0 + \beta_1 (X_1 - X_2) + \text{error}.$$

Fit these two models, apply F test with $\Delta df = 1$.
Usually Δdf equals the number of constraints.

Predictions

Suppose we get a new observation:

$$x'_{\text{new}} = (x_{\text{new},1}, \ldots, x_{\text{new},p}).$$

To predict the mean response, simply plug into model:

$$\hat{\mu}_{\text{new}} = x'_{\text{new}}\hat{\beta}.$$

The standard error of $\hat{\mu}_{\text{new}}$ is:

$$s.e.(\hat{\mu}_{\text{new}}) = \hat{\sigma}\sqrt{x'_{\text{new}}(X'X)^{-1}x_{\text{new}}}.$$

Confidence limits for the prediction is given by:

$$\hat{\mu}_{\text{new}} \pm t(n-p, \alpha/2)s.e.(\hat{\mu}_{\text{new}}).$$

Predictions - Subtleties

The actual response will be the predicted mean value plus a random error:

$$\hat{y}_{\text{new}} = \hat{\mu}_{\text{new}} + \text{error}.$$

The error has variance σ^2. So,

$$\hat{y}_{\text{new}} \pm t(n-p, \alpha/2) s.e.(\hat{y}_{\text{new}}).$$

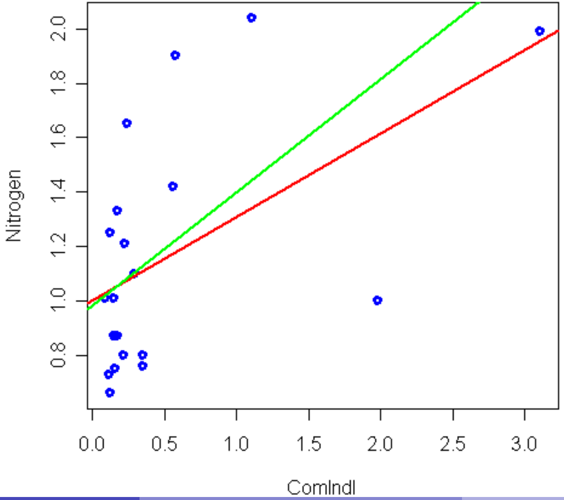
where

$$s.e.(\hat{y}_{\text{new}}) = \hat{\sigma}\sqrt{1 + x'_{\text{new}}(X'X)^{-1}x_{\text{new}}}.$$

Note that the prediction stays the same, only the width of the confidence interval changes.

Influence of Outliers

How much influence does the data point have on the model fit?



Different measures of Influence

1. How much influence does observation i have on its own fit?

$$(DFFITS)_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}}$$

   $DFFITS$ exceeding $2\sqrt{p/n}$ is considered large.

2. How much influence does observation i have on the fitted β's?

$$(DFBETAS)_i = \frac{\hat{\beta}_1 - \hat{\beta}_{1(i)}}{\sqrt{MSE_{(i)} c_{11}^{-1}}},$$

   where $c_{11} = \sum_i (x_i - \bar{x})^2$. $DFBETA$ exceeding $2/\sqrt{n}$ is considered large.

These conventional rules work for “reasonably sized” data sets.

Different measures of Influence: Cook's Distance

1. Cook's distance is defined as:

$$D_i = \frac{\sum_{j=1}^{n} (\hat{y}_j - \hat{y}_{j(i)})^2}{(p+1)MSE}$$

2. Considers the influence of Y_i on all of the fitted values, not just the i-th case.
3. It can be shown that D_i is equivalent to

$$\frac{\tilde{r}_i^2}{p+1} \frac{h_{ii}}{1-h_{ii}}.$$

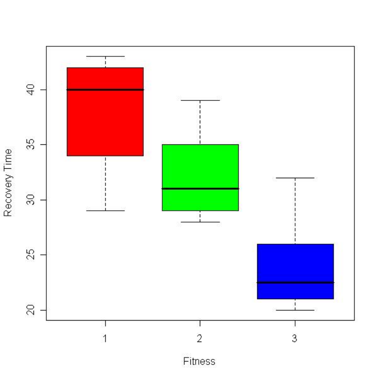
4. Compare D_i to the $F_{p+1,n-p-1}$ distribution.

One-way ANOVA

Can be viewed in two different ways:

1. Extension of “two-sample” *t*-test to more than two groups.
2. Extension of simple linear regression to case where X is qualitative.

Example: rehab surgery



How does prior fitness affect recovery from surgery? Observations: 24 subjects’ recovery time.

Three fitness levels: below average (8), average (10), above average (6).

One-way ANOVA model

$Y_{ij}, 1 \leq i \leq r, 1 \leq j \leq n_i :$ r groups and n_i samples in i-th group

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2).$$

Constraint $\sum_{i=1}^{r} \alpha_i = 0$ needed for "identifiability"

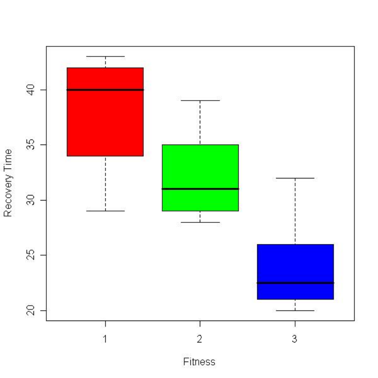
This is equivalent to:

$$Y_{ij} = \mu + \alpha_1 I_{fitness=1} + \alpha_2 I_{fitness=2} + \alpha_3 I_{fitness=3} + \varepsilon_{ij},$$

Can always phrase an ANOVA problem as a multiple linear regression problem using indicator variables.

$$Y_{ij} = \mu + \alpha_1 I_{fitness=1} + \alpha_2 I_{fitness=2} + \alpha_3 I_{fitness=3} + \varepsilon_{ij},$$

Example: rehab surgery



Design matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \vdots & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \vdots \\ 0 & 0 & 1 \end{pmatrix}$$

Fitting One-way ANOVA

Follow exactly the same principles as linear regression:

- Model is *easier* to fit:

$$\widehat{Y}_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \overline{Y}_{i\cdot\cdot}$$

  If observation is in i-th group: predicted mean is just the sample mean of the i-th group.

- Simplest question: is there any group (main) effect?

$$H_0 : \alpha_1 = \cdots = \alpha_r = 0?$$

- Test is based on F-test with full model vs. reduced model. Reduced model just has an intercept.
- Other questions: is the effect the same in groups 1 and 2?

$$H_0 : \alpha_1 = \alpha_2?$$

One-way ANOVA table

Source	SS	df	MS	E(MS)
Treatments	$SSTR = \sum_{i=1}^{r} n_i \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2$	$r-1$	SSTR/(r-1)	$\sigma^2 + \frac{\sum_{i=1}^{r} n_i \alpha_i^2}{r-1}$
Error	$SSE = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2$	$n-r$	SSE/(n-r)	σ^2

1. Notation: $\overline{Y}_{i.}$ is i-th group mean, $\overline{Y}_{..}$ is overall mean.
2. $MSTR = SSTR/(r-1)$ measures "variability" of the "cell" means. If there is a group effect we expect this to be large relative to MSE.
3. Under $H_0 : \alpha_1 = \cdots = \alpha_r = 0$, the expected value of $MSTR$ and MSE is σ^2.
4. Under H_0 the ratio of mean squares follow an F distribution.

F-test for one-way ANOVA

Source	SS	df	MS	E(MS)
Treatments	$SSTR = \sum_{i=1}^{r} n_i \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$	$r-1$	SSTR/(r-1)	$\sigma^2 + \frac{\sum_{i=1}^{r} n_i \alpha_i^2}{r-1}$
Error	$SSE = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$	$n-r$	SSE/(n-r)	σ^2

$$H_0 : \alpha_1 = \cdots = \alpha_r = 0$$

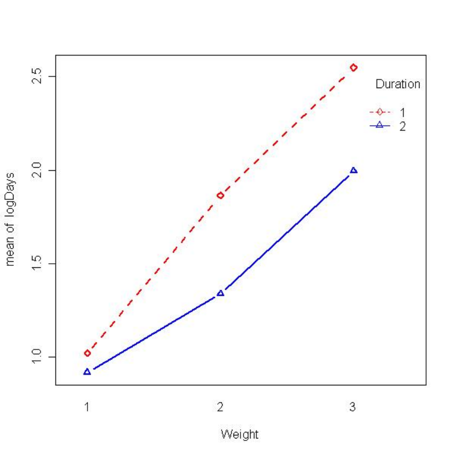
Under H_0,

$$F = \frac{MSTR}{MSE} = \frac{\frac{SSTR}{df_{TR}}}{\frac{SSE}{df_E}} \sim F_{df_{TR}, df_E}$$

Reject H_0 at level α if $F > F_{1-\alpha, df_{TR}, df_E}$.

Two-Way ANOVA

Example: rehab time from kidney failure



Recovery time depends on weight gain between treatments and duration of treatment.

Two levels of duration, three levels of weight gain.

Two-way ANOVA model: observations:

$$(Y_{ijk}), 1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq n_{ij}$$

a groups in first grouping variable (A),
b groups in second grouping variable (B),
n_{ij} samples in (i,j)-"cell".

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \sim N(0, \sigma^2).$$

In kidney example, $a = 3$ (weight gain), $b = 2$ (duration of treatment), $n_{ij} = 10$ for all (i,j).

Using indicator variables, this is still a multiple regression problem.

Two-way ANOVA: main questions of interest

- Are there main effects for the grouping variables?

$$H_0 : \alpha_1 = \cdots = \alpha_a = 0, \qquad H_0 : \beta_1 = \cdots = \beta_b = 0.$$

- Are there interaction effects?

$$H_0 : (\alpha\beta)_{ij} = 0, 1 \leq i \leq a, 1 \leq j \leq b.$$

Constraints needed for identifiability

- $\sum_{i=1}^{a} \alpha_i = 0$
- $\sum_{j=1}^{b} \beta_j = 0$
- $\sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, 1 \leq i \leq a$
- $\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, 1 \leq j \leq b.$

Term	SS
A	$SSA = nm\sum_{i=1}^{a}\left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot}\right)^2$
B	$SSB = nr\sum_{j=1}^{b}\left(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot\cdot\cdot}\right)^2$
AB	$SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdot\cdot\cdot}\right)^2$
Error	$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij\cdot})^2$

Two-way ANOVA table ($n_{ij} = n$)

SS	df	$E(MS)$
SSA	$a-1$	$\sigma^2 + nb\frac{\sum_{i=1}^{a}\alpha_i^2}{a-1}$
SSB	$b-1$	$\sigma^2 + na\frac{\sum_{j=1}^{b}\beta_j^2}{b-1}$
$SSAB$	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$
SSE	$(n-1)ab$	σ^2

F-tests for two-way ANOVA

SS	df	$E(MS)$
SSA	$a-1$	$\sigma^2 + nb\frac{\sum_{i=1}^{a} \alpha_i^2}{a-1}$
SSB	$b-1$	$\sigma^2 + na\frac{\sum_{j=1}^{b} \beta_j^2}{b-1}$
$SSAB$	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$
SSE	$(n-1)ab$	σ^2

$$MS = SS/df$$

F-tests:

$$F_{AB} = MSAB/SSE \sim F((a-1)(b-1), (n-1)ab)$$

$$F_A = MSA/SSE \sim F(a-1, (n-1)ab)$$

$$F_B = MSB/SSE \sim F(b-1, (n-1)ab)$$