BST 210
Applied Regression Analysis

"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Lecture 10
Plan for Today

1. Everyone please attend class next Tuesday (10/8)
2. Lab this week: quick scan prior to lab--Gaynor et al, 2015

3. (continued) Multivariable Model Selection
    a. Criteria for model selection
    b. Data reduction; collinearity

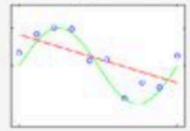
4. Multivariable hypothesis testing
    1. t-test
    2. F-test
    3. Nested models (including additive models)
    4. When LINE breaks down

Main Points of Model Selection?

(one demonstration of process)

1. Define models

Simple Model = Dog preference is predicted well by height.



$$\text{Model (S)} = m \cdot x + c$$

Augmented Model = Dog preference is predicted better by height and age

$$\text{Model (A)} = m \cdot x + n \cdot x^2 + c$$

2. Comparing Error reductions

$$F = \frac{(SSE - SSE) / (P - P)}{SSE / n - P}$$

SSE: the sum of squared error;
P: the number of parameters
n: the number of observations

3. Significance test

Criteria for Model Selection?

- Adjusted R^2
- F-test
- t-test
- AIC
- BIC
- Root MSE
- P-values

Akaike's Information Criteria (AIC)

- Pronounced 'ah-kai-ee-kay'

$$\text{AIC} = \{2\text{x}(\#\text{ of parameters}) - 2\log(L)\}$$

- where L is the maximized likelihood from the data

- Given a set of candidate models, the preferred model could be the one with the *minimum* AIC value
- Models don't need to be nested; don't need to have same covariates, but must have same outcome
- Dataset and number of observations must be the same
- (Same holds for adjusted R-squared)

Akaike's Information Criteria (AIC)

- Here $Y_i \sim N(\beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_p \cdot x_{ip}, \sigma^2)$

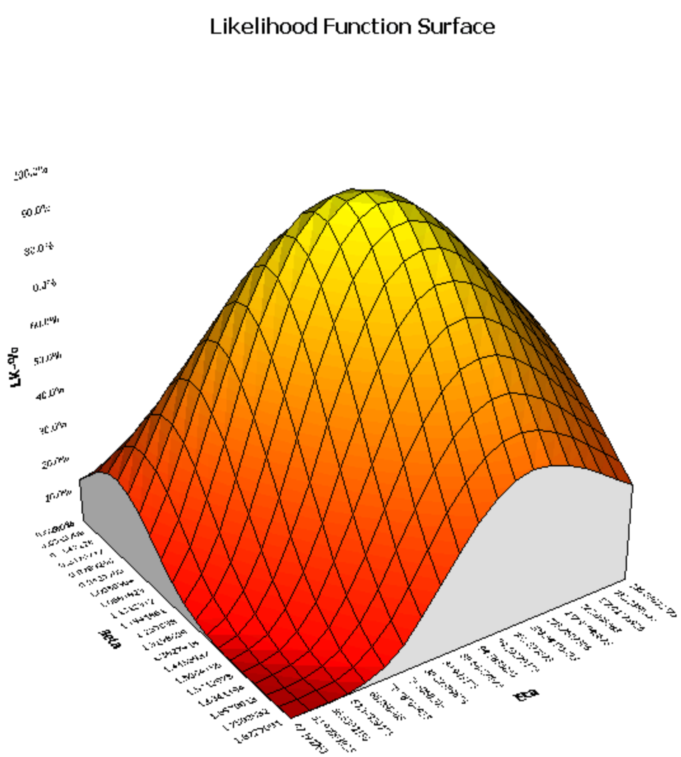
  or $Y_i \sim N(\mu_i, \sigma^2)$

- The likelihood function is given by

$$L(\beta_0, \beta_1, \ldots, \beta_p, \sigma^2) = \prod_{i=1}^{n} \exp\left(-(y_i - \mu_i)^2/2\sigma^2\right)/\sqrt{2\pi\sigma^2}$$

- The β vector (MLE) that maximizes this likelihood is also the least squares estimator (LSE) of β

Akaike's Information Criteria (AIC)



AIC = {2x(# of parameters) – 2 log(*L*)} where *L* is the maximized likelihood (MLE) from the data

Akaike's Information Criteria (AIC)

- A general rule of thumb is that models that differ by more than 2 AIC units are generally considered to "differ".

- AIC is based on information theory, dealing with the trade-off between the goodness of fit of the model and the complexity of the model, but tells us nothing about the quality of the model in an absolute sense

Akaike's Information Criteria (AIC)

- AIC is more useful in predictions of fitted values for future observations rather then estimation of β coefficients
- Usually more generous in allowing covariates than hypothesis testing (which recall is based on *p*-values)

Bayesian Information Criteria (BIC)

- (Schwartz) Bayesian information criterion is given by

$$\text{BIC} = \log(N)\times(\text{\# of parameters}) - 2\log(L)$$

- Here L is the maximized likelihood from the data and N is the sample size
- Chooses model that maximizes the 'posterior likelihood of the data, given the model' (Bayes)
- As log(N) gets bigger than 2 (most often), BIC becomes more restrictive in allowing covariates into model

Bayesian Information Criteria (BIC)

- Given a set of candidate models, the preferred model could be the one with the *minimum* BIC value
- Is more restrictive than AIC in including covariates, though again more useful in predicting fitted values for future observations

Bayesian Information Criteria (BIC)

- BIC:

$$\log(N) \times (\#\text{ of parameters}) - 2\log(L)$$

Versus

- AIC:

$$2 \times (\#\text{ of parameters}) - 2\log(L)$$

```
. regress pdi minutes vsd

      Source |       SS           df       MS      Number of obs   =       142
-------------+----------------------------------   F(2, 139)       =      5.02
       Model |  2266.77325         2  1133.38662   Prob > F        =    0.0079
    Residual |  31391.8465       139  225.840622   R-squared       =    0.0673
-------------+----------------------------------   Adj R-squared   =    0.0539
       Total |  33658.6197       141  238.713615   Root MSE        =    15.028

------------------------------------------------------------------------------
         pdi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     minutes |  -.1351676   .0587281    -2.30   0.023    -.2512836   -.0190516
         vsd |  -5.245585   3.112904    -1.69   0.094    -11.40035    .9091801
       _cons |   101.0308   2.413607    41.86   0.000     96.25865    105.8029
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        142 -589.7313  -584.7811       3    1175.562    1184.43
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. regress pdi dhca vsd

      Source |       SS           df       MS      Number of obs   =       142
-------------+----------------------------------   F(2, 139)       =      5.78
       Model |  2584.88318         2  1292.44159   Prob > F        =    0.0039
    Residual |  31073.7365       139  223.552061   R-squared       =    0.0768
-------------+----------------------------------   Adj R-squared   =    0.0635
       Total |  33658.6197       141  238.713615   Root MSE        =    14.952

------------------------------------------------------------------------------
         pdi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dhca |  -6.534115   2.510433    -2.60   0.010    -11.49769   -1.570544
         vsd |  -6.629234   3.037331    -2.18   0.031    -12.63458   -.6238895
       _cons |   99.90491   1.917255    52.11   0.000     96.11415    103.6957
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        142 -589.7313  -584.0579       3    1174.116   1182.983
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

```
. regress pdi minutes vsd birthwt

      Source |       SS           df       MS      Number of obs   =       142
-------------+----------------------------------   F(3, 138)       =      6.02
       Model |  3893.19946         3  1297.73315   Prob > F        =    0.0007
    Residual |  29765.4203       138  215.691451   R-squared       =    0.1157
-------------+----------------------------------   Adj R-squared   =    0.0964
       Total |  33658.6197       141  238.713615   Root MSE        =    14.686

------------------------------------------------------------------------------
         pdi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     minutes |  -.1471749    .0575597    -2.56   0.012    -.2609878   -.0333619
         vsd |   -4.71444    3.048297    -1.55   0.124    -10.74185    1.312968
     birthwt |   .0081227     .002958     2.75   0.007     .0022738    .0139715
       _cons |   72.63161    10.60759     6.85   0.000     51.65719    93.60603
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        142 -589.7313  -581.0038       4    1170.008   1181.831
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

. regress pdi dhca vsd birthwt

      Source |       SS           df       MS      Number of obs   =       142
-------------+----------------------------------   F(3, 138)       =      6.54
       Model |  4190.84279         3   1396.9476   Prob > F        =    0.0004
    Residual |  29467.7769       138  213.534615   R-squared       =    0.1245
-------------+----------------------------------   Adj R-squared   =    0.1055
       Total |  33658.6197       141  238.713615   Root MSE        =    14.613

------------------------------------------------------------------------------
         pdi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dhca |  -6.952038    2.458269    -2.83   0.005    -11.81278   -2.091293
         vsd |  -6.224385    2.972168    -2.09   0.038    -12.10126   -.3475067
     birthwt |   .0080636    .0029403     2.74   0.007     .0022497    .0138775
       _cons |   71.53173    10.51437     6.80   0.000     50.74162    92.32183
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        142 -589.7313  -580.2903       4    1168.581   1180.404
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

Comparisons of AIC and BIC

- AIC is relatively liberal, allowing for the inclusion of simple continuous or binary predictors
- In contrast, BIC imposes progressively more restrictive constraints as N increases, leading to a more parsimonious model (e.g., $p < 0.05$ in samples of about 50, $p < 0.01$ in samples of about 500, and $p < 0.009$ in samples of about size 1000)

Comparisons of AIC and BIC

- Because both depend on the number of parameters, they set the bar higher for multi-parameter situations, e.g., the inclusion of age categories (rather than continuous age) and the inclusion of restricted cubic splines (using multiple parameters to adjust for the covariate)

Bootstrapping Stepwise Selection

- Bootstrapping is based on random sampling with replacement from your original dataset, to make other samples (provided original sample is large enough)
- Stata allows bootstrapping of stepwise selection models to assess how often a covariate will be included (say 50% of the time)

```
ssc install swboot
swboot y x1 x2 x3, reps(1000)
```

- May be challenging to use with correlated covariates; SAS and R can do this similarly

Validation

- Comparing fits with an external validation dataset (try to replicate results in a comparable study)
- Breaking your dataset into test and validation datasets; fit model on test dataset and validate it on the validation dataset (need big enough N)
- Cross-validation (e.g., delete one methods; or break up data into ten groups, fit data on nine groups and validate on the tenth, repeat ten times)

Collinearity

- Highly correlated predictor variables can lead to inflated standard errors of regression coefficients or large *p*-values and competing similar covariates when making a model selection
- Collinearity can make it difficult to estimate and interpret a particular regression coefficient (as the effect on Y of one unit increase in x_i when x_i is highly correlated with other covariates)
- May not reach model convergence. Software isn't going to tell you what you're doing wrong.

Collinearity

- VIF: Some people assess variance inflation factors (VIF) which works with ‘sets’ of potentially collinear variables coming into the model, though VIF is not always informative (some prefer pre-model decisions versus calculation during modeling).

- VIF describes “the amount of an ‘X’ that is described by other ‘Xs’ in the model.” …..sound familiar?

$$\mathbf{VIF = 1/(1-R_j^2)}$$

where R_j^2 is the unadjusted coefficient of variation between variable j and the other variables.

Collinearity

- VIF > 2 raises concern; VIF > 10 prompts action
- Usually you know more about your covariates in advance and can plan for possible collinearity by data reduction methods (include some function of FEV and FVC; perhaps the avg of both parents' education level; BMI, etc)
- No reason why you should not know about collinearity well in advance of building your model (based on scatterplots, correlation coefficients, etc.)

Collinearity

Analysis of Variance

Source	DF	Seq SS	Seq MS	F-Value	P-Value
Regression	6	557.844	92.974	560.64	0.000
Age	1	243.266	243.266	1466.91	0.000
Weight	1	311.910	311.910	1880.84	0.000
BSA	1	1.768	1.768	10.66	0.006
Dur	1	0.335	0.335	2.02	0.179
Pulse	1	0.123	0.123	0.74	0.405
Stress	1	0.442	0.442	2.67	0.126
Error	13	2.156	0.166		
Total	19	560.000			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.407229	99.62%	99.44%	99.08%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-12.87	2.56	-5.03	0.000	
Age	0.7033	0.0496	14.18	0.000	1.76
Weight	0.9699	0.0631	15.37	0.000	8.42
BSA	3.78	1.58	2.39	0.033	5.33
Dur	0.0684	0.0484	1.41	0.182	1.24
Pulse	-0.0845	0.0516	-1.64	0.126	4.41
Stress	0.00557	0.00341	1.63	0.126	1.83

Data Reduction Methods

- Using some, but not all, of your covariates, based on subject matter knowledge (more off the start is not necessarily better)
- Eliminate some variables because of narrow distributions (ie only 5 ‘training time’ in N=488) or large numbers of missing values (careful here…)
- Averaging certain correlated covariates (perhaps after scaling) Ex: FEV, FVC type measures, or BMI

Data Reduction Methods

- Use principal components (PCA), which are weighted averages of your original data values in an attempt to summarize them with fewer components
- PCA utilizes eigenvalues and eigenvectors (and concept of orthogonality), thus breaking down the collinearity
- Typically only use the first 2-3 principal components
- Could be very difficult to interpret
- Common in data mining techniques

Summary

- Both a pro and con of various model selection techniques = ‘you’re not over-thinking it’
- Make sure you’re giving input, putting a lot of thought into the process, not handing entirely over to software (force lower order terms into model, awareness of collinearity, model makes clinical sense, etc)

Next

- Hypothesis testing and Confidence intervals for multiple coefficients (aka ‘multiple coefficient testing’)

- More examples!

Recall: Multiple Linear Regression

- Model: $E(Y_i) = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_p \cdot x_{ip}$
  - $E(Y_i)$ = expected value of Y for a given set of covariates, $x_{i1}, x_{i2}, \ldots, x_{ip}$
  - β_0 = intercept, or constant term, corresponding to the mean value of Y when <u>all</u> covariates = 0
  - β_j = slope, or the change in Y corresponding to a 1 unit increase in the j^{th} covariate, x_j, holding all the other covariates constant

Recall: Hypothesis Testing for a Single β Coefficient

- To test H_0: $\beta_i = 0$ vs. H_A: $\beta_i \neq 0$ we usually use the (Wald) test statistic

$$t = \frac{\hat{\beta}_i - 0}{s.e.(\hat{\beta}_i)}$$

using $n - (p + 1)$ degrees of freedom

- This test is *exact* if the LINE assumptions hold
- For large *n*, an approximate test could use 1.96 as the cutoff value

Recall: Hypothesis Testing for a Single β Coefficient

- To test H_0: $\beta_i = 10$ vs. H_A: $\beta_i \neq 10$ we use the test statistic

$$t = \frac{\hat{\beta}_i - 10}{s.e.(\hat{\beta}_i)}$$

using $n - (p + 1)$ degrees of freedom

Recall: Confidence Interval for a Single β Coefficient

- A 100(1– α)% confidence interval for β_i is given by

$$\hat{\beta}_i \pm\ t_{n-(p+1),1-\alpha/2} s.e.(\hat{\beta}_i)$$

- This confidence interval is *exact* if the LINE assumptions hold
- For large *n*, an *approximate* 95% confidence interval could use 1.96 as the cutoff value

New: Hypothesis Testing for Multiple β Coefficients

- To test H_0: $\beta_i = \beta_j$ or H_0: $\beta_i - \beta_j = 0$ vs. H_A: $\beta_i - \beta_j \neq 0$
- We use the test statistic

$$t = \frac{(\widehat{\beta}_i - \widehat{\beta}_j) - 0}{s.e.(\widehat{\beta}_i - \widehat{\beta}_j)}$$

  using $n - (p + 1)$ degrees of freedom
- But, what is the denominator?

Hypothesis Testing for Multiple β Coefficients

- By definition,

$$Var(\hat{\beta}_i - \hat{\beta}_j) = Var(\hat{\beta}_i) + Var(\hat{\beta}_j) - 2Cov(\hat{\beta}_i, \hat{\beta}_j)$$

- We need to get the variance/covariance matrix of the β coefficients in order to calculate this

Hypothesis Testing for Multiple β Coefficients

- Also,

$$s.e.(\hat{\beta}_i - \hat{\beta}_j) = \sqrt{Var(\hat{\beta}_i - \hat{\beta}_j)}$$

- From this one can calculate the test statistic by hand, although usually a computer package is used

Hypothesis Testing for Multiple β Coefficients (Nested Models)

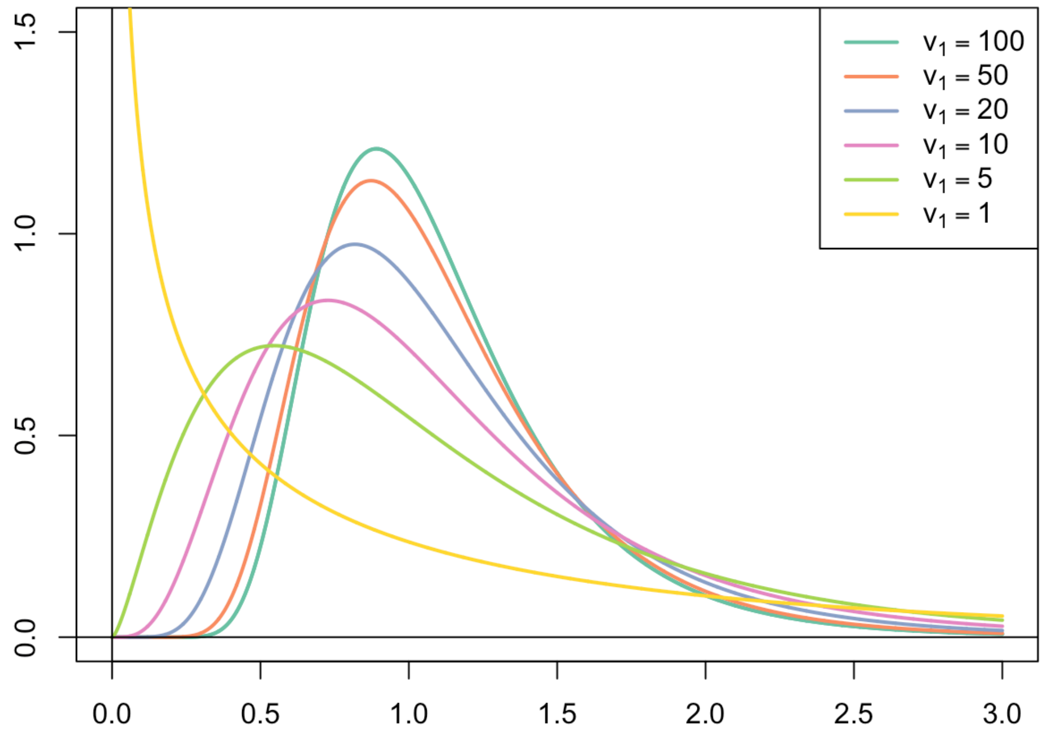
- To test H_0: $\beta_i = 0$, $\beta_j = 0$, $\beta_k = 0$ vs. H_A: At least of of β_i , β_j, and β_k are $\neq 0$
- Use an F test statistic, based on comparing the residual sums of square of the models including and excluding these coefficients

$$F = \frac{(RSS_{reduced} - RSS_{full})/3}{RSS_{full}/(n-(p+1))}$$

using 3 numerator and $n - (p + 1)$ denominator degrees of freedom

Quick review of F-distribution

$$F = \frac{Y_1/v_1}{Y_2/v_2}$$



Hypothesis Testing for Multiple β Coefficients (Nested Models)

- In general, if we have a <u>full</u> model with p covariates and a <u>reduced</u> model with q covariates that is <u>nested</u> within the more complex full model, then

$$F = \frac{(RSS_{reduced} - RSS_{full})/(p-q)}{RSS_{full}/(n-(p+1))}$$

using $p - q$ numerator and $n - (p + 1)$ denominator degrees of freedom

Hypothesis Testing for Multiple β Coefficients (Nested Models)

- Caveat: You need to make sure that the samples used in both regression models is the same (so the sample sizes used in both models should be the same)
- This could be a problem if you have some missing covariates
- This can be ameliorated by limiting the regressions to exactly the same sample

Hypothesis Testing for Multiple β Coefficients (Full Model)

- As a special case, most software packages perform a test of whether or not all of the p (non-intercept) covariates are equal to zero or not using a F test with p numerator and $n - (p + 1)$ denominator df

```
. regress pdi minutes vsd

      Source |       SS           df       MS      Number of obs   =       142
-------------+----------------------------------   F(2, 139)       =      5.02
       Model |  2266.77325         2  1133.38662   Prob > F        =    0.0079
    Residual |  31391.8465       139  225.840622   R-squared       =    0.0673
-------------+----------------------------------   Adj R-squared   =    0.0539
       Total |  33658.6197       141  238.713615   Root MSE        =    15.028

------------------------------------------------------------------------------
         pdi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     minutes |  -.1351676    .0587281    -2.30   0.023    -.2512836   -.0190516
         vsd |  -5.245585    3.112904    -1.69   0.094    -11.40035    .9091801
       _cons |   101.0308    2.413607    41.86   0.000     96.25865    105.8029
------------------------------------------------------------------------------
```

Hypothesis Testing for Multiple β Coefficients (Nested Models)

Example:

$$\text{Model 9: } E[Y] = \beta_0 + \beta_1 x$$
$$\text{Model 10: } E[Y] = \beta_0 + \beta_1 x + Spline(x)/GAM(x)$$

Write out the model!

```
mod11=lm(y~x)
mod12=lm(y~x+bSpline(x, df=6, degree = 3))
## or use GAM mod10=gam(y~x+s(x,4))
anova(mod11,mod12)
```

Hypothesis Testing for Multiple β Coefficients (Nested Models)

Example continued:

```
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ x + bSpline(x, df = 6, degree = 3)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     98 6036003
## 2     93   80004  5   5955999 1384.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation?

Highly significant F test, suggesting nonlinear effect of x on y.

Hypothesis Testing for Multiple β Coefficients (Nested Models)

Example: What does the model represent?

$$E[Y] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x-a)_+^3 + \beta_5 (x-b)_+^3 + \beta_6 (x-c)_+^3$$

What does this hypothesis test do?

$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$
H_A: At least one is not equal to zero

Why does the F statistic have df(5,93)?

Hypothesis Testing for Multiple β Coefficients (Nested Models)

Example: What does the following model test?

```
mod13=lm(y~x+x^2)
mod14=gam(y~x+x^2+s(x))
anova(mod13,mod14)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x + x^2
## Model 2: y ~ x + x^2 + s(x)
##   Res.Df     RSS     Df Sum of Sq      F    Pr(>F)
## 1     98 6036003
## 2     94  192375 3.9998   5843628 713.88 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(for reference from Lecture 7) Piecewise Cubic Splines

- For knot points at $a < b < c$ for a continuous covariate x, the cubic spline function is given by

$$
\begin{aligned}
E(Y) = {} & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \\
& + \beta_4 (x - a)_+^3 + \beta_5 (x - b)_+^3 \\
& + \beta_6 (x - c)_+^3
\end{aligned}
$$

- Thus, k knot points would lead to a model with $k + 4$ parameters (including the intercept)

Linear Regression Assumptions and Extensions

- What if LINE assumptions are not exactly satisfied for linear regression modeling?
- Some extensions of linear regression

Multiple Linear Regression

- Model: $E(Y_i) = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_p \cdot x_{ip}$
  - $E(Y_i)$ = expected value of Y for a given set of covariates, x_{i1} , x_{i2} ,…, x_{ip}
  - β_0 = intercept, or constant term, corresponding to the mean value of Y when all covariates = 0
  - β_j = slope, or the change in Y corresponding to a 1 unit increase in the j^{th} covariate, x_j, holding all the other covariates constant

Multiple Linear Regression

- LINE for $E(Y_i) = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_p \cdot x_{ip}$
  – The mean of *Y* is a linear function of the covariates (quadratic terms, regression spline, etc. are okay)
  – All responses are independent
  – The residuals are normally distributed
  – The residuals have equal variance (homoscedasticity)

What if LINE Not Satisfied?

- L: Specification of the model for E(Y) needs to be correct – otherwise all bets are off and there is no reason to get consistency of estimates of β, even for large samples, for an incorrect specification of the regression model

What if LINE Not Satisfied?

- I: Independence is required in order for standard errors to be estimated appropriately – methods need to be extended if you have correlated or longitudinal responses
- However, generalized estimating equations (GEEs) for correlated responses with working independence would give the same β estimates (but need to "fix" the s.e. estimates so they are appropriate)

What if LINE Not Satisfied?

- N and E: Here is where we can relax the assumptions a bit. In fact, provided that the residuals have mean zero and finite variance, we don’t have to have that they are normally distributed or have equal variance.
- Well, the residuals will have mean zero with an intercept in the model, one can show

What if LINE Not Satisfied?

- N and E: The scope is for inference class, but one can show that $E\left(\hat{\beta}\right) = \beta$ for the least squares estimator if the residuals have mean 0, and even when the residuals are not normal or homoscedastic
- Great for estimating β, but what about s.e. estimates, p-values, etc.? Surely the inferences are affected by the nonnormality

What if LINE Not Satisfied?

- N and E: Because N and E aren't satisfied, we need to "adjust" the usual s.e. estimates of the β coefficients
- These are called robust standard errors, or Huber-White standard errors, or sandwich standard errors, or GEE standard errors (with independent responses)
- Given these, one can calculate CI's and P-values

What if LINE Not Satisfied?

- N and E: These standard errors can be calculated easily in most packages
- And, since these s.e. estimates will get smaller as the sample size gets larger, we will still get consistency of the β estimates

What if LINE Not Satisfied?

- Some might recommend using the robust s.e. estimates all the time (rather than the usual s.e. estimates), but:
  - these don't lead to exact *t* tests or *F* tests
  - they are asymptotic (large sample) s.e.'s, not "exact" for small samples, as the *t*-based CIs are for normal, homoscedastic errors
  - there are no $\hat{\sigma}^2$, adjusted R^2, MSE, AIC, BIC, leverages, studentized residuals, Cook's distances, etc.

Alternate Approaches

- *Weighted least squares* could be used when the residuals are normally distributed but have unequal variance
- But, you need to develop a way of setting $Var(Y_i)$ as a function of covariates

Alternate Approaches

- Various *robust regression methods* have been proposed that down-weight high residual observations, so as to give them less effect on the estimation of β

Bottom Line

- If you can verify the LINE assumptions through model assessment (histograms of residuals, QQ plots, etc.), then would usually prefer to stick with “ordinary” least squares and multiple linear regression analysis
- But, it can be useful to use these robust methods when the N and E assumptions are not exactly satisfied, or to compare them at the end to your usual s.e. estimates

Many Extensions

- *Nonlinear least squares*, when the model for the mean is a nonlinear function of the β coefficients, perhaps still with normally distributed errors

Many Extensions: Longitudinal

- *Random or mixed effects models*, which allow some β coefficients to vary by subject when you have repeated measures on subjects
- *Generalized estimating equations* for adjusting the s.e. estimates, even if the model for the covariances are incorrect
- *Conditional models, time series models*, etc.
- (Focus of BST 226)

Coming Up

- Start of logistic regression! – odds ratios, logits, interpretation of logistic regression coefficients