```{r, include=FALSE, results='hide'}
legend('topright', c('a', 'b', 'c'), pch=1:3)
boxplot(Y ~ somefactor, data = mydata)
options(contrasts=c('contr.sum', 'contr.sum'))
means = tapply(Y, somefactor, mean)
G = expand.grid(k=1:3, j=1:4, i=1:3)
leaps::regsubsets(y ~ ., data=d, ...)
MASS::stepAIC(model0, ...)
```

Binomial $X \sim B(n,p)$

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$$

$\mathrm{E}\, X = np, \quad \mathrm{Var}\, X = np(1-p)$

Poisson $X \sim P(\lambda)$

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 0, 1, \ldots$$

$\mathrm{E}\, X = \lambda, \quad \mathrm{Var}\, X = \lambda$

Normal $X \sim N(\mu, \Sigma)$, Σ positive definite

$$f(x) = \frac{\exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}}{(2\pi)^{\frac{k}{2}}\sqrt{\det(\Sigma)}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

mgf: $M_X(t) = \exp(\mu' t + \frac{1}{2} t' \Sigma t)$

Beta $X \sim \mathrm{Beta}(\alpha, \beta)$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad 0 \le x \le 1$$

$\mathrm{E}\, X = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\, X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Gamma $X \sim \mathrm{Gamma}(\alpha, \beta)$

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \qquad x > 0$$

$\mathrm{E}\, X = \frac{\alpha}{\beta}, \quad \mathrm{Var}\, X = \frac{\alpha}{\beta^2}$

$X \sim \mathrm{Gamma}(\alpha, \beta) \iff \beta X \sim \mathrm{Gamma}(\alpha, 1)$

X_i iid $\mathrm{Gamma}(\alpha_i, \beta)$, then

$$\sum X_i \sim \mathrm{Gamma}(\sum \alpha_i, \beta)$$

Exponential Special case: $X \sim \mathrm{Exp}(\lambda) \equiv \mathrm{Gamma}(1, \lambda)$

$f(x) = \lambda e^{-\lambda x}, \quad x > 0 \qquad \mathrm{E}\, X = \frac{1}{\lambda}, \quad \mathrm{Var}\, X = \frac{1}{\lambda^2}$

CDF $F(x) = 1 - e^{-\lambda x}$

Chi square Special case: $X \sim \chi^2_n \equiv \mathrm{Gamma}(\frac{n}{2}, \frac{1}{2})$

$f(x) \propto x^{\frac{n}{2}-1} e^{\frac{-x}{2}}, \quad x > 0 \qquad \mathrm{E}\, X = n, \quad \mathrm{Var}\, X = 2n$

Let Z_i be iid $N(0,1)$. $\sum_{i=1}^n Z_i^2 \sim \chi^2_n$

Noncentral χ^2. Let $Y \sim N(\mu, I)$ be an n vector. Then

$$\|Y\|^2 \sim \chi^2_n(\|\mu\|^2)$$

F if num. and den. independent then

$$F(m,n) \equiv \frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$

T if num. and den. independent then

$$t(n) = \frac{N(0,1)}{\sqrt{\frac{\chi^2_n}{n}}}$$

Conditional pdf:

$$f_{X|Y}(x|y) \equiv \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Iterated expectation:

$$E(Y) = E(E(Y|X))$$

Conditional variance formula:

$$\mathrm{Var}(Y) = \mathrm{Var}(E(Y|X)) + E(\mathrm{Var}(Y|X))$$

Singular Value Decompostion (SVD) Any matrix X can be written

$$X = UDV^T$$

with U, V orthogonal, and D diagonal.

Moore Penrose Psuedoinverse A^+ exists uniquely for every matrix A. Properties: $AA^+A = A$, $A^+AA^+ = A^+$ and AA^+, A^+A are symmetric.

Projection matrix P are symmetric and idempotent. They have eigenvalues either 0 or 1.

$$P = P^T \qquad P^2 = P$$

Covariance of linear transformations

$$Cov(Ay, Bx) = ACov(y,x)B^T$$

Invert 2×2 matrix: $A = [\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}]$

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Block matrix: $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} =$

$$\begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Sum identities:

$$\sum_{k=0}^{\infty} p^k = \frac{p}{1-p} \qquad \sum_{k=0}^{\infty} kp^k = \frac{p}{(1-p)^2} \qquad |p| < 1$$

Integration by parts:

$$\int uv' = uv - \int u'v$$

Matrix / Vector differentiation

$\frac{\partial A^T\beta}{\partial \beta} = A$, $\frac{\partial \beta^T A \beta}{\partial \beta} = (A + A^t)\beta = 2A\beta$ for A symmetric.

$\frac{\partial}{\partial \theta_i} \log(|A|) = tr(A^{-1}\frac{\partial A}{\partial \theta_i})$

232B

Linear mixed models Standard assumptions:

$$y = X\beta + Z\alpha + \epsilon$$

where $\alpha \sim N(0, G)$ independent of $\epsilon \sim N(0, R)$.

Marginal model $y \sim N(X\beta, V)$ where $V = R + ZGZ'$.

Restricted maximum likelihood (REML) uses a linear transformation Ay to remove the fixed effects, and then estimates the variance. We have $\text{rank}(A) = n - p$ where $p = \text{rank}(X)$ and $A'X = 0$. Then $z = A'y \sim N(0, A'VA)$ and we can maximize the restricted log likelihood to estimate the variance parameters. Let $P = A(A'VA)^{-1}A'$ and solve

$$0 = \frac{\partial l_R}{\partial \theta_i} = \frac{1}{2}\left(y'P\frac{\partial V}{\partial \theta_i}Py - tr(P\frac{\partial V}{\partial \theta_i})\right)$$

Predictions Let $\xi = b'\beta + a'\alpha$ be a mixed effect. These are what we're interested in estimating. We call them predictions rather than estimations because we're predicting a random component.

BLUE - Best linear unbiased estimator, $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$. This is the MLE of β.

BLUP - Best linear predictor, which plugs in $\tilde{\beta}$ into BP.

BP - Best predictor, $\text{E}(\xi|y) = b'\beta + a'GZ'V^{-1}(y - X\beta)$ a theoretical ideal that's usually difficult or impossible to derive.

EBLUP - Empirical best linear predictor, plugs in estimates for both fixed and variance components into the BP. This is typically the one we compute and use.

```
library(lme4)
library(HLMdiag)
fit1 = lmer(y ~ X1 + X2 + (1 | V), data=somedata)
s1 = summary(fit1)
# 2 ways to extract variance estimates:
s1$sigma^2
s1$varcar$V
HLMdiag::varcomp.mer(fit1)
getME(fit1, ...)  # Extract various components
predict(fit1, data=newdata)  # EBLUPs
```

2 ways to get standard errors for variance estimates are parametric bootstrap and asymptotic covariance matrix of the estimates.

232A

Least Squares Principle

$$\text{arg min}_\beta \left\|Y - X\beta\right\|^2$$

Normal Equations - Any b satistfying this solves the least squares

$$X^TXb = X^Ty$$

Gauss Markov Theorem - $\hat{\beta}$ is Best Linear Unbiased Estimator (BLUE) of β.

$$\hat{\beta} = (X^TX)^{-1}X^Ty \sim N(\beta, \sigma^2(X^TX)^{-1})$$

variance:

$$\frac{\left\|y - X\hat{\beta}\right\|^2}{\sigma^2} \sim \chi^2_{n-p}. \qquad \Longrightarrow \qquad \hat{\sigma}^2 = \frac{\left\|y - X\hat{\beta}\right\|^2}{n-p}$$

Use t test for hypothesis testing and confidence intervals for the value of a particular β_j coefficient. Let w_{ii} be the ith diagonal entry of $(X^TX)^{-1}$.

$$\frac{\beta_j - \beta_j^*}{\hat{\sigma}\sqrt{w_{ii}}} \sim t_{n-p}$$

$1 - \alpha$ Confidence intervals for new observation Y_h at x_h and $E[Y_h]$:

$$E[y_h] \approx \hat{y_h} \pm t(n-p, 1-\frac{\alpha}{2})\hat{\sigma}\sqrt{x_h^T(X^TX)^{-1}x_h}$$

$$y_h \approx \hat{y_h} \pm t(n-p, 1-\frac{\alpha}{2})\hat{\sigma}\sqrt{1 + x_h^T(X^TX)^{-1}x_h}$$

Simultaneous (Working-Hotelling) confidence interval for $\text{E}(y_h)$:

$$\hat{y}_h \pm \sqrt{pF_{p,n-p,1-\alpha}}se\{\hat{y}_h\}$$

$$\frac{(\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta)/p}{\hat{\sigma}^2} \sim F_{p,n-p}$$

General linear tests. Partition $\beta = (\beta_1, \beta_2)$ where β_1 is an r vector and β_2 is $p - r$. Null hypothesis $H_0 : \beta_2 = \beta_2^*$ (often 0), and $H_a : \beta_2 \neq \beta_2^*$. Then $SSE_r = \left\|y - X_2\beta_2^* - X_1\tilde{\beta}_1\right\|^2$ is the sum of squared error for the reduced model and $SSE_f = \left\|y - X\hat{\beta}\right\|^2$ is the squared sum of error for the full model. Under H_0:

$$\frac{\frac{SSE_r - SSE_f}{p-r}}{\frac{SSE_f}{n-p}} \sim F_{p-r,n-p}$$

Alternate forms of linear test, and testing a linear combination if $R\beta = r$, for R full rank $s \times p$ matrix.

$$\frac{(R\hat{\beta} - r)^T(R(X^TX)^{-1}R^T)^{-1}(R\beta - r)/s}{\hat{\sigma}^2} \sim F_{s,n-p}$$

$SSTO = \sum_{i=1}^n (y_i - \bar{y}) = \left\|y - \bar{y}1_n\right\|^2 = \left\|(I - J)y\right\|^2$

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y}) = \left\|\hat{y} - \bar{y}1_n\right\|^2 = \left\|(H - P)y\right\|^2$

$SSE = \sum_{i=1}^n (y_i - \hat{y}_i) = \left\|y - \hat{y}\right\|^2 = \left\|(I - H)y\right\|^2$

```
SSE = sum(residuals(fit)^2)
SSR = sum((fitted(fit) - mean(y))^2)
SSTO = sum((y - mean(y))^2)
```

If the model contains the intercept in the column space of X then $SSTO = SSR + SSE$.

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

Adjusted $R_a^2 = 1 - \frac{SSE/(n-p)}{SSTO/(n-1)}$

$AIC = n \log SSE + 2p$

$BIC = n \log SSE + p \log n$

$Cp = \frac{SSE}{MSE} - (n - 2p)$

`influence.measures, cooks.distance`

Residuals: $\hat{\epsilon}_i = y_i - \hat{y}_i$

Studentized residuals (`rstandard` in R):

$$\gamma_i = \frac{\hat{\epsilon}_i}{s\{\hat{\epsilon}_i\}} = \frac{\hat{\epsilon}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

Studentized deleted residuals (`rstudent` in R):

$$t_i = \frac{\hat{\epsilon}_i}{\sqrt{MSE_{(-i)}(1 - h_{ii})}} \sim t_{n-p-1}$$

Where $MSE_{(-i)} = SSE_{(-i)}/(n-1-p)$ and $SSE_{(-i)} = SSE - \frac{\hat{\epsilon}_i}{1-h_{ii}}$ can be used to calculate without refitting model.

Prediction sum of squares (PRESS) is the same as leave one out cross validation (LOOCV). Prediction error on ith observation is called deleted residuals:

$$y_i - \hat{y}_{i(-i)} = \frac{y_i - \hat{y}_i}{1 - H_{ii}}$$

Works for ridge regression also, letting $H = X(X^T X + \lambda I)^{-1} X^T$.

Ridge Regression for $\lambda > 0$ solves

$$\min_{\beta} ||Y - X\beta||^2 + \lambda ||\beta||^2$$

```
W = solve(XtX + lambda * Ip)
betahat = W %*% t(X) %*% y
```

ANOVA

Three principles of experimental design: 1) Replication 2) Randomization 3) Blocking

One way ANOVA with n total observations, K groups:

SS		DF
SSTR	$\sum_{j=1}^{K} n_j(\bar{y_{j\cdot}} - \bar{y_{\cdot\cdot}})^2$	K - 1
SSE	$\sum_{i=1}^{n}(y_{ij} - \bar{y_{j\cdot}})^2$	n - K
SSTO	$\sum_{i=1}^{n}(y_{ij} - \bar{y_{\cdot\cdot}})^2$	n - 1

Contrasts are sums of the form $\Phi = \sum_{i=1}^{K} c_i \mu_i$ with $\sum_{i=1}^{K} c_i = 0$. Tukey's works for all pairwise contrasts. Scheffe's and extended Tukey works for all contrasts. Bonferroni's is for a limited number of pre specified contrasts.

Example of contrast L with standard error $se(\hat{L}) = \hat{\sigma}\sqrt{\sum_{i=1}^{I} c_i^2/n_i}$. Each way to compute a confidence interval for L can be expressed in the form:

$$\hat{L} \pm se(\hat{L}) \times \text{multiplier}$$

```
m = c(simple = qt(1 - alpha / 2, fit$df.residual)
, bonferroni = qt(1 - alpha / (2 * 20), fit$df.residual)
, ex_tukey = qtukey(1 - alpha, I, n - I)
, scheffe = sqrt((I - 1) *
                qf(1 - alpha, I - 1, fit$df.residual)))
width = seL * multipliers
TukeyHSD(aov(y ~ x, data))  # pairwise intervals
```

Multivariate Normal

log likelihood for k vector $x \sim N(\mu, \Sigma)$

$$l_x = -\frac{k}{2}\log 2\pi - \frac{1}{2}\{\log\det\Sigma + (x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

Stein's formula: $X \sim N(\mu, \sigma)$

$$\mathrm{E}(g(X)(X-\mu)) = \sigma^2\,\mathrm{E}(g'(X))$$

assuming these expectations are finite.

$X \sim N(\mu, \Sigma)$, A an $m \times n$ matrix, then

$$AX \sim N(A\mu, A\Sigma A^t)$$

For Σ full rank it's possible to transform between $Z \sim N(0, I)$ and X:

$$X = \Sigma^{1/2}Z + \mu \qquad Z = \Sigma^{-1/2}(X-\mu)$$

In block matrix form:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

Assuming Σ_{11} is positive definite then the conditional distribution

$$X_2|X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$