Chapter 6, Linear Normal Models

Properties:

- As GLM
- Maximum Likelihood Estimate (MLE)
- Least Square Estimate
- Deviance
- Hypothesis testing
- Outlier detection (delta-beta, Cook's distance and leverage)
- Colinearity

Models:

- Multiple linear regression
  - Outlier detection / influential observation
  - Collinearity / multicollinearity
- Analysis of variance (ANOVA)
  - One factor ANOVA
  - Two factor ANOVA
- Analysis of covariance
- General linear model

GLM formulation

- Reponse: $Y_i \sim N(\mu_i, \sigma^2)$
- Expected value: $E(Y) = \mu_i$
- Identity link function: $\eta_i = \mu_i$
- Linear component: $\eta_i = x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \ldots \beta_{p-1} x_{ip-1}$

Deviance

Let β_{max} be the parameter vector for the *saturated* modeled, and β for the model of our interest. Let $l(\beta; y)$ be the log-likelihood function. The *deviance* of the model is

$$D = 2(l(b_{max}; y) - l(b; y)$$

where b and b_{max} are (ML) estimates.

Saturated model

The richest possible model. Each combination of (all possible known) explanatory variables have their own θ_i. $b = b_{max}$

Gaussian pdf

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-0.5\frac{(y-\mu)^2}{\sigma^2})$$

F-distribution ch 1.4

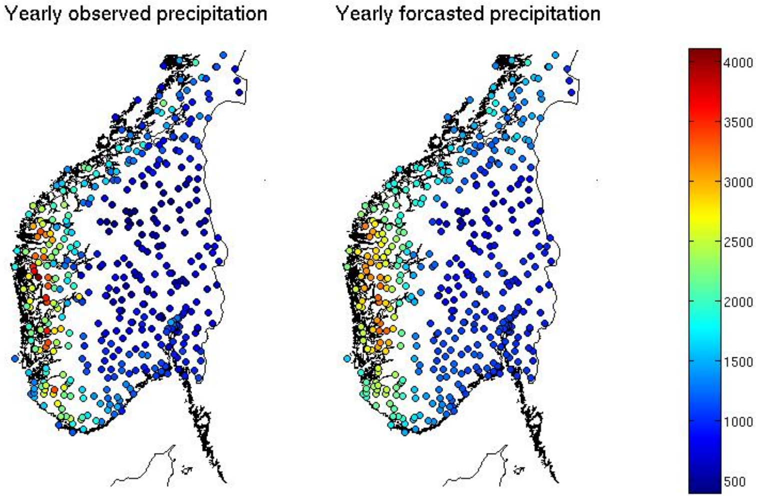
Definition central F-distribution

If $X_1^2 \sim \chi^2(n)$, $X_2^2 \sim \chi^2(m)$ and X_1^2 and X_2^2 are independent, then

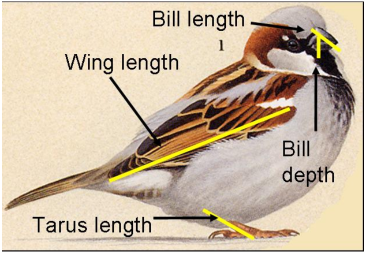
$$F = \frac{X_1^2/n}{X_2^2/m} \sim F(n, m)$$

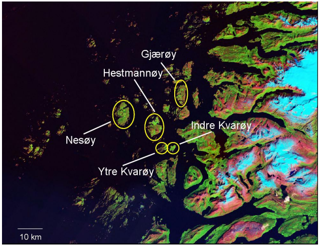
Precipitation

5 years of daily precipitation observation and forecast for 450 locations
$\Rightarrow$ 1.6 mill data.



House sparrows questions





1. Are birds heavier in summer then in winter?
2. Are birds relatively heavier on the outer islands in summer then on the inner islands?
3. Body mass modeled with tarsus length, wing length, bill length and bill depth.
4. Body mass modeled with tarsus length, wing length, bill length, bill depth and season.
5. Are birds heavier on the outer islands when we account for size (tarsus, wing length, etc.) ?