Shooting balloons

- N trail subjects, $i = 1, 2, \ldots, N$
- Each shot n_i times, trying to hit balloons.
- Count hits y_i.
- Explanatory variables:
  - Experienced / non-experienced gunman
  - Wind speed

Shooting balloons, data

Trail person	1	2	3	...
Experiences	1	0	0	...
Wind speed	2.13	0.59	1.03	...
n_i	6	3	5	...
y_i	2	1	1	...

GLM

Generalized Linear Model

1. Likelihood; $f(y;\theta)$, member of the exponential family.
2. Link function; $g(\mu_i) = x_i\beta$, where $\mu_i = E(Y_i)$ and $g()$ is monotone and differentiable.
3. Linear component of explanatory variables; $g(\mu) = X\beta$

Shooting balloons, model

- $Y_i \sim bin(n_i, \pi_i)$, $i = 1, 2, \ldots, N$
- $\eta_i = logit(\pi_i) = \log(\frac{p_i}{1-p_i})$
-
  1. $\eta_i = \beta_0 \Rightarrow Y_i \sim bin(n_1, \pi)$
  2. $\eta_i = \beta_0 + \beta_1 x_1$
  3. $\eta_i = \beta_0 + \beta_2 x_2$
  4. $\eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Where $x_1 = 1$ for experienced gunman, otherwise $x_1 = 0$ and x_2 is wind speed.

Exponential family

$f(y;\theta)$ belongs to the exponential family if

$$f(y;\theta) = \exp[a(y)b(\theta)) + c(\theta) + d(y)]$$

Score statistics

Let $l(\theta; y_i)$ be log-likelihood function. Then the score statistic is:

$$U(\theta; y) = \frac{\partial l(\theta; y_i)}{\partial \theta}$$

Information

Let $U = U(\theta; Y)$ be the score statistic. Then the information is

$$\Im = Var(U)$$

If Y_i has pdf from exponential family:

- $\Im = E(U^2) = -E(\frac{\partial U}{\partial \theta}) = -E(\frac{\partial^2 l(\theta;y)}{\partial \theta^2})$

Hight of male students

- In population: $Y \sim (179.8, 6.5^2)$
- Mean of 91 male students: 181.9
- NTNU students higher then Norwegians?

$\chi^2()$ results ch 1.4 and 1.5

Definition χ^2

If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2(1)$.
If $Z_1, Z_2, \ldots Z_n$ are independent identical distributed $Z_i \sim N(0,1)$, the $\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$

Non iid

If $Y \sim MVN(\mu, \Sigma)$, then $(Y-\mu)^T \Sigma^{-1}(Y-\mu) \sim \chi^2(n)$

Subtraction

If $X_1^2 \sim \chi^2(m)$ and $X_2^2 \sim \chi^2(k)$, $m > k$, and X_1^2 and X_2^2 are independent, we have: $X^2 = X_1^2 - X_2^2 \sim \chi^2(m-k)$

5.3 Taylor series approximations for log-likelihood

Taylor approximations for $l(\beta)$ near estimate b:

$$l(\beta) = l(b) + (\beta - b)U(b) + \frac{1}{2}(\beta - b)^2 U'(b)$$

Approximate $U'(b)$ with $E(U') = -\Im(b)$:

$$l(\beta) = l(b) + (\beta - b)U(b) - \frac{1}{2}(\beta - b)^2 \Im$$

For a vector b

$$l(\beta) = l(b) + (\beta - b)U(b) - \frac{1}{2}(\beta - b)^T \Im(\beta - b)$$

5.3 Taylor series approximations for log likelihood

$$U(\beta) = U(b) - (\beta - b)U'(b)$$

Approximate $U'(b)$ with $E(U') = -\Im(b)$:

$$U(\beta) = U(b) - (\beta - b)\Im(b)$$