BS1a Applied Statistics
Lectures 1-2

Dr Geoff Nicholls

Week 1 MT10

The linear model: $i = 1, ..., n$ observations y_i,

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... \beta_p x_{i,p} + \epsilon_i,$$

with $\epsilon_i \sim \text{N}(0, \sigma^2)$ iid normal *errors*.

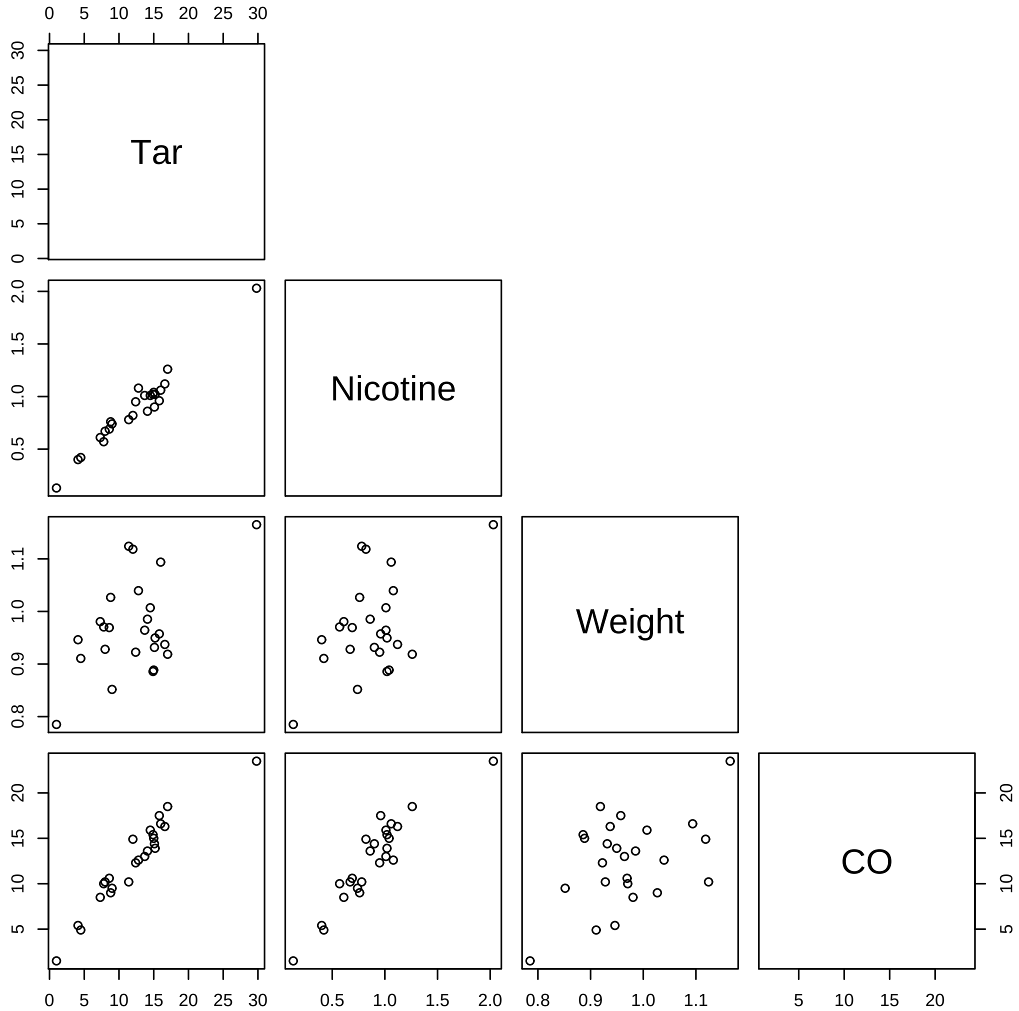
$n \times p$ *design* matrix $X = [x_{i,j}]$ or $X = (X_1, X_2, ..., X_p)$
$\beta = (\beta_1, \beta_2, ..., \beta_p)^T$

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

y, X_i, ϵ each in R^n. Sometimes omit subscripts

$$y = \alpha + \gamma_1 z_1 + ... + \gamma_m z_m + \epsilon$$

is scalar with $\beta_1 = \alpha, x_1 = 1 \quad \beta_2 = \gamma_1, x_2 = z_1$ $etc.$

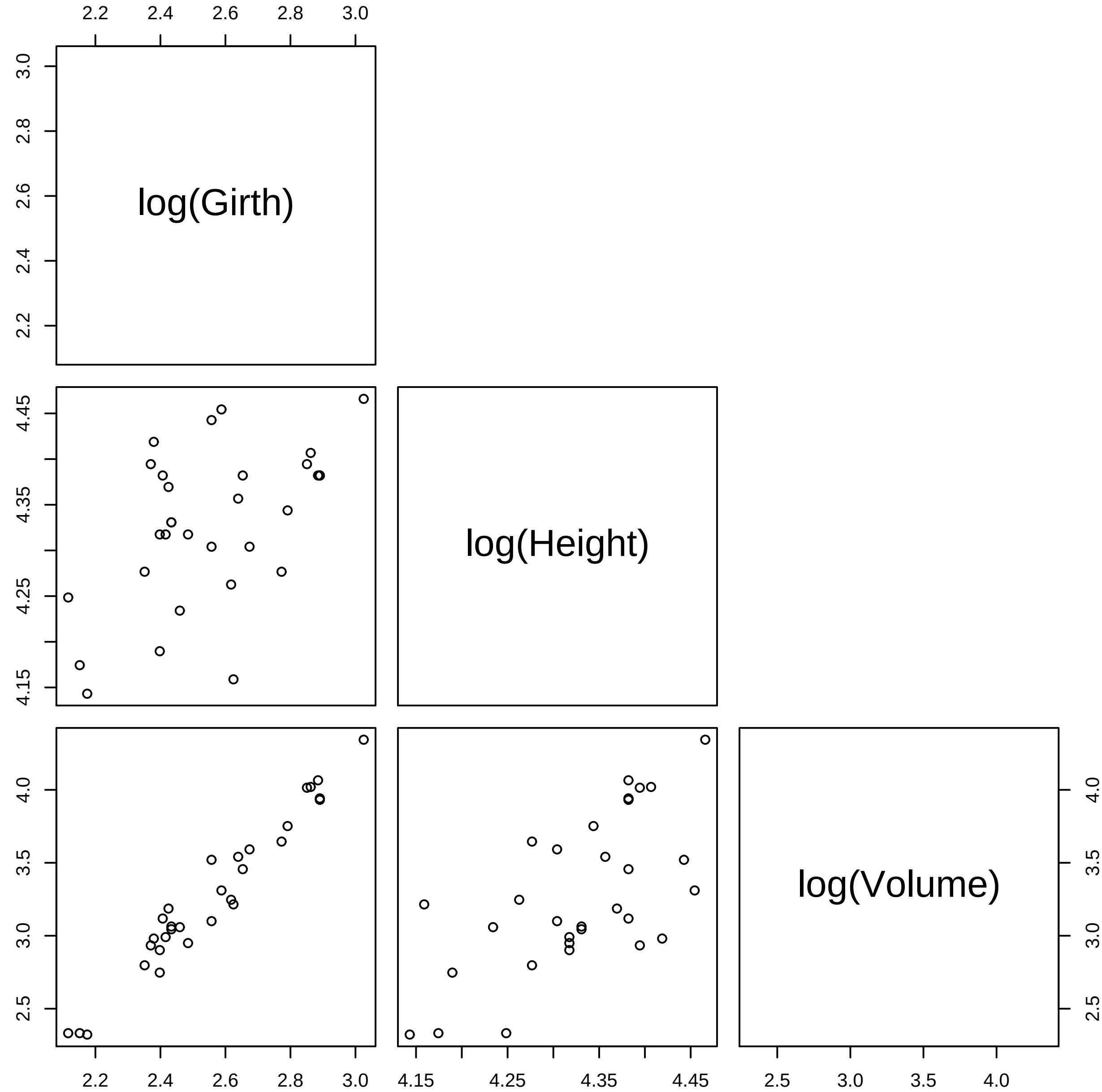


Cigarettes: x_2 =Nicotine, x_3 =Tar, x_4 =Weight.

MLE $\widehat{\beta}$ for parameters, normal linear model

$$y = 3.2 - 2.63x_2 + 0.96x_3 - 0.13x_4 + \epsilon$$

Correlation $\Rightarrow$ cant attribute causal variable.



log(Girth)
log(Height)
log(Volume)
2.2
2.4
2.6
2.8
3.0
4.15
4.25
4.35
4.45
2.5
3.0
3.5
4.0

trees data.

v volume, h height and g girth, consider

$$v = \eta h^{1+\beta_2} g^{2+\beta_3} \gamma$$

η constant and γ a random variable.

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

with $y = \log(v/hg^2)$, $\beta_1 = \log(\eta)$, $x_2 = \log(h)$, $x_3 = \log(g)$ and $\epsilon = \log(\gamma)$.

MLE's were $\widehat{\beta}_1 = -6.6, \widehat{\beta}_2 = 0.12, \widehat{\beta}_3 = -0.02$ and $\widehat{\sigma} = 0.08$ so

$$v = \exp(-6.6) h^{1.12} g^{1.98} \gamma,$$

$\log(\gamma) \sim N(0, 0.08^2)$.

Test H0:$\beta_2 = \beta_3 = 0$ is a natural next step.

Estimators in the normal linear model.

Log-likelihood ($\mathbf{x}_i$ ith row of X)

$$\ell(\beta, \sigma^2; y) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i\beta)^2.$$

$\text{RSS}(\beta) = (y - X\beta)^T(y - X\beta)$.

MLE $\widehat{\beta}$ for β minimises the RSS at fixed σ.

$$\text{col}(X) = \{z \in R^n : z = X\beta, \beta \in R^p\}$$

Assume $X_i, i = 1, ..., p$ linearly independent col(X) is p-dim linear subspace of R^n.

Choose $\widehat{\beta}$ so $\widehat{y} = X\widehat{\beta}$ is the orthogonal projection of y into col(X). Then

$$X^T(y - X\widehat{\beta}) = 0$$

is p equations in p unknowns.

If $X^T X$ is invertible (eg X rank p)

$$\widehat{\beta} = (X^T X)^{-1} X^T y$$

$\widehat{\sigma}^2_{\text{MLE}}$ maximises

$$\ell(\widehat{\beta}, \sigma^2; y) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\widehat{\beta})^T(y - X\widehat{\beta})$$

at

$$\widehat{\sigma}^2_{\text{MLE}} = \frac{\text{RSS}}{n}$$

sub this in above to get

$$\ell(\widehat{\beta}, \sigma^2_{\text{MLE}}; y) = -\frac{n}{2}\log(\text{RSS}/n) - n/2.$$

Properties of Estimators

$n \times n$ hat matrix H

$$H = X(X^T X)^{-1} X^T$$

and estimated response $\widehat{y} = Hy$.

Residuals, $e = y - \widehat{y}$, with RSS($\widehat{\beta}$) $= e^T e$.

Main results (for the normal linear model)

- RSS$/\sigma^2 \sim \chi^2(n-p)$

- $\widehat{\beta}$ and RSS are independent ($\Rightarrow$ t-test)

- In fact (!) the residuals e and the estimated response $\widehat{y}$ are independent.

$U = (U_1, ..., U_p)^T$ and $W = (W_1, ..., W_n)^T$

$$\text{cov}(U, W) = \text{E}((U - \mu_U)(W - \mu_W)^T)$$

is $p \times n$ with $\text{cov}(U, W)_{i,j} = \text{cov}(U_i, W_j)$.

$$\text{var}(U) = \text{E}((U - \mu_U)(U - \mu_U)^T)$$

is symmetric $p \times p$ with $\text{var}(U)_{i,j} = \text{cov}(U_i, U_j)$.
$\text{var}(\epsilon) = \sigma^2 I_n$ so $\text{var}(Y) = \sigma^2 I_n$.

Check $\hat{y}$ and residuals e are uncorrelated

$$\text{cov}(\hat{Y}, Y - \hat{Y}) = 0_{n,n}$$

[Note: *jointly normal*, with zero covariance, implies independent...]

If $W \sim N(\mu_W, \Sigma)$ and L is $p \times n$ rank $p \leq n$ then $LW \sim N(L\mu_W, L\Sigma L^T)$.

$$\hat{\beta} \sim \text{N}(\beta, \sigma^2 (X^T X)^{-1}).$$

Here $L = (X^T X)^{-1} X^T$ is rank p and $W = Y$.
[End L1]

Claim:
e and $\hat{Y}$ (and $\hat{\beta}$ and RSS) are independent.

Let $\varepsilon_1, ..., \varepsilon_p$ be orthonormal basis for col(X).

Extend to $\varepsilon_1, ..., \varepsilon_p, \varepsilon_{p+1}, ..., \varepsilon_n$, orthonormal basis for R^n.

Expand Y

$$Y = Z_1\varepsilon_1 + ... + Z_n\varepsilon_n.$$

with $Z_i = \varepsilon_i^T Y$ random weights. Now

$$\hat{Y} = Z_1\varepsilon_1 + ... + Z_p\varepsilon_p,$$

since $H\varepsilon_i = 0, i > p$. Next $e = Y - \hat{Y}$ so

$$e = Z_{p+1}\varepsilon_{p+1} + ... + Z_n\varepsilon_n.$$

Now claim Z_i are jointly independent. If so, $\hat{Y}$ and e are independent.

Weights

$$Z_i = \varepsilon_i^T Y \qquad Z_i \sim \mathsf{N}(\varepsilon_i^T \mathsf{E}(Y), \sigma^2)$$

or $Z = (Z_1, ..., z_n)^T$, $E = (\varepsilon_1, ..., \varepsilon_n)$

$$Z = E^T Y \qquad Z \sim \mathsf{N}(E^T \mathsf{E}(Y), \sigma^2 I_n)$$

by LW-property above and

$$\mathsf{cov}(Z_i, Z_j) = \varepsilon_i^T \mathsf{cov}(Y, Y) \varepsilon_j$$

so uncorrelated $i \neq j$ and $\mathsf{var}(Z_i) = \sigma^2$.

Since they are jointly normal and uncorrelated, they are independent.

Claim:
RSS$/\sigma^2 \sim \chi^2(n-p)$

Since E$(Y) = X\beta$, and $\varepsilon_i^T X = 0_{1,p}$ for $i > p$

$$Z_i \sim \mathrm{N}(0, \sigma^2), \quad i = p+1, ..., n$$

Using RSS $= e^T e$,

$$\mathrm{RSS} = Z_{p+1}^2 + ... + Z_n^2.$$

Now $(Z_i/\sigma)^2 \sim \chi^2(1)$ so

$$\mathrm{RSS}/\sigma^2 \sim \chi^2(n-p),$$

under $H0$. Also $A \sim \chi^2(r)$ then E$(A) = r$ so E(RSS$/\sigma^2) = n - p$ and

$$s^2 = \frac{\mathrm{RSS}}{n-p}$$

is an unbiased estimator for σ^2.

Tests: R-output for `trees` data was

```
    Coefficients:
                Estimate Std. Err t val Pr(>|t|)
(Intercept) -6.63162  0.79979 -8.292 5.06e-09 ***
log(Height)  0.11712  0.20444  0.573 0.571
log(Girth)  -0.01735  0.07501 -0.231 0.819
```

Have seen

$$\widehat{\beta} \sim \mathrm{N}(\beta, \sigma^2 (X^T X)^{-1})$$

and

$$\frac{\mathrm{RSS}}{\sigma^2} \sim \chi^2(n-p)$$

independently, in the NLM so

$$\frac{\widehat{\beta}_k}{\texttt{Std.Error}(\widehat{\beta}_k)} \sim t(n-p)$$

under the null model with $\beta_k = 0$.

See that $\beta_2 = 0$ and $\beta_3 = 0$ allowed separately but would like to check this holds jointly...

The F-Test (background): drop k variables from a NLM

$$H0 : y = \sum_{i=1}^{p-k} \beta_i^{(0)} x_i + \epsilon.$$

$X^{(0)} = [X_1, ..., X_{p-k}]$, $H^{(0)} = \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T, ...$

$$H1 : y = \sum_{i=1}^{p} \beta_i^{(1)} x_i + \epsilon.$$

$X^{(1)} = [X_1, ..., X_p]$, so $X^{(1)} = X$, $H^{(1)} = H, ...$

LRT:

$$\begin{aligned}\Lambda(Y) &= -2(\ell(\widehat{\beta}^{(0)}, \widehat{\sigma}^2_{MLE,0}; Y) - \ell(\widehat{\beta}^{(1)}, \widehat{\sigma}^2_{MLE,1}; Y)) \\ &= n \log\left(1 + \frac{\mathrm{RSS}^{(0)} - \mathrm{RSS}^{(1)}}{\mathrm{RSS}^{(1)}}\right).\end{aligned}$$

Reject $H0$ if $\Lambda(y) > \chi^2_{1-\alpha}(k)$ (approximate)

F-Test:

$$F(y) = \frac{(\mathsf{RSS}^{(0)} - \mathsf{RSS}^{(1)})/k}{\mathsf{RSS}^{(1)}/(n-p)}.$$

$\mathsf{N}(\mu, \sigma^2)$, $\chi^2(k)$, $t(k)$, ... $F(a, b)$

Independent r.v. $A \sim \chi^2(a)$ and $B \sim \chi^2(b)$

$$F = \frac{(A/a)}{(B/b)} \Rightarrow F \sim F(a, b)$$

for such F, $\mathsf{E}(F) = b/(b-2)$ so for example

$$\mathsf{E}(F(Y)) = \frac{(n-p)}{(n-p-2)}$$

about 1 when $n \gg p$.

Reject $H0$ if $F(y) > F_{1-\alpha}(k, n-p)$.

Example (trees again)

Compare $v = \eta h g^2 \gamma$ with $v = \eta h^{1+\beta_2} g^{2+\beta_3} \gamma$

$$H0 : y = \beta_1^{(0)} + \epsilon^{(1)}$$

$$H1 : y = \beta_1^{(1)} + \beta_2^{(1)} x_2 + \beta_3^{(1)} x_2 + \epsilon^{(1)}$$

```
> tr.lm0<-lm(log(Volume/(Height*Girth^2))~1)
> (rss0<-sum(tr.lm0$residuals^2))
[1] 0.1876858

> tr.lm1<-lm(log(Volume/(Height*Girth^2))~
             1+log(Height)+log(Girth),data=trees)
> (rss1<-sum(tr.lm1$residuals^2))
[1] 0.1854634

> k<-2; n.minus.p<-31-3
> (F<-(rss0-rss1)*n.minus.p/(k*rss1))
[1] 0.1677617
> (p<-1-pf(F,k,n.minus.p))
[1] 0.8463989
```

The p-value is $p = 0.85$
Accept $H0 : \beta_2 = \beta_3 = 0$.

Claim:
The F-test is an exact LRT for the NLM.

$$\Lambda = n \log \left(1 + \frac{k}{(n-p)} F \right)$$

Λ is a strictly increasing function of F

Set C', C so

$$\Pr(F(Y) > C' | H0) = 1 - \alpha$$

and

$$\Pr(\Lambda(Y) > C | H0) = 1 - \alpha.$$

$$\left.\begin{array}{r} F\text{-Test} \\ \text{LRT (exact)} \end{array}\right\} \text{ reject H0 if } \left\{\begin{array}{l} F(y) > C' \\ \Lambda(y) > C \end{array}\right.$$

but $F(y) > C'$ iff $\Lambda(y) > C$ since monotone. [end L2]