For Gaussian models,

**Exercise 1**

**(a).** AIC statistic: $AIC = -2\ln f(y|\hat{\theta}_k) + 2k$

Mallow's $C_p$: $C_p = \frac{RSS}{\hat{\sigma}_0^2} + 2d - n$     $\hat{\sigma}_0$: estimator of $\sigma_0^2$

$AIC = -2\ln f(y|\hat{\theta}_k) + 2k$

$= n\ln\hat{\sigma}^2 + n[\ln 2\pi + 1)] + 2(d+1)$,    $\hat{\sigma} = \frac{RSS}{n}$: MLE of $\sigma^2$

minimize AIC: $n\ln\hat{\sigma}^2 - n\ln\hat{\sigma}_0^2 + 2d = n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) + 2d$

$n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) \approx n\ln(1) + (\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} - 1)n$

$= \frac{n\hat{\sigma}^2}{\hat{\sigma}_0^2} - n$

$= \frac{RSS}{\hat{\sigma}_0^2} - n$.

$AIC = n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) + 2d \approx [\frac{RSS}{\hat{\sigma}_0^2} - n] + 2d = C_p$

**(b).** $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$    want $BIC = \frac{1}{n}(RSS + \ln(n) d\hat{\sigma}^2)$

$BIC = -2\ln f(y|\hat{\theta}_k) + d\ln n$

$= n + n\ln 2\pi + n\ln\hat{\sigma} + \ln n \cdot d$

minimize BIC: $n\ln\hat{\sigma}^2 - n\ln\hat{\sigma}_0^2 + \ln n = n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) + \ln n \cdot d$

$n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) = n\ln(1) + (\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} + 1)n$

$= \frac{n\hat{\sigma}^2}{\hat{\sigma}_0^2} - n$

$= \frac{RSS}{\hat{\sigma}_0^2} - n$

$BIC = n\ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}) + \ln(n)d = [\frac{RSS}{\hat{\sigma}_0^2} - n] + \ln(n)d$

Since $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$

$= [\frac{RSS}{\hat{\sigma}_0^2} - n] + 2d$,

$BIC = \frac{1}{n}(RSS + \ln(n)d\hat{\sigma}^2)$

**(c)** This representation provides a more general applicability, because it is easy to calculate RSS and $\hat{\sigma}_0$ for a Gaussian model instead of $\hat{\sigma}$. In addition, it can be used in linear and non linear least squares, which are more accurate than using the asymptotic normality