**L20: Global $F$-tests in linear models**

1. A likelihood ratio test

   (1) Maximized likelihood functions
   Model M: $Y \sim N(X\beta, \sigma^2\Sigma)$ with rank$(X) = r$ and SSE has
   $$\max[L(\beta, \sigma^2) : \beta, \sigma^2] = \left(\frac{n}{2\pi e}\right)^{n/2} \cdot \frac{1}{|\Sigma|^{1/2}} \text{ SSE}^{-n/2}.$$
   If under $H_0$ M is reduced to $M_*$: $Y \sim N(X_*\beta_*, \sigma^2\Sigma)$ with rank$(X_*) = r_*$ and SSE$_*$,
   $$\max[L(\beta, \sigma^2) : H_0] = \left(\frac{1}{2\pi e}\right)^{n/2} \cdot \frac{1}{|\Sigma|^{1/2}} \text{ SSE}_*^{-n/2}$$
   Let $F = \frac{(SSE_* - SSE)/(r-r^*)}{MSE}$ where MSE $= \frac{SSE}{n-r}$.

   (2) Likelihood ratio and LRT
   LR$= \frac{\max[L(\beta,\sigma^2): \beta, \sigma^2]}{\max[L(\beta,\sigma^2): H_0]} = \left(\frac{SSE_*}{SSE}\right)^{n/2} = \left(1 + \frac{SSE_* - SSE}{SSE}\right)^{n/2} = \left(1 + F \cdot \frac{r-r_*}{n-r}\right)^{n/2}$
   is an increasing function of $F$. Therefore below is a LRT scheme.

   > $H_0$ : vs $H_a$ : $H_0$ is false
   > Test statistic: $F = \frac{(SSE_* - SSE)/(r-r_*)}{MSE}$
   > Reject $H_0$ if $F > c_0$

2. Two $\alpha$-level LRTs

   (1) Setting I
   For Model M with $1_n \in \mathcal{R}(X)$ there exists an hypothesis $H_0$, under $H_0$ M is reduced to $M_1 : Y \sim N(1_n\gamma, \sigma^2\Sigma)$. For M and this $H_0$ there is ANOVA table

   | Source | SS | DF | MS | F | $Pr > F$ |
   |---|---|---|---|---|---|
   | Model | SSM | $r-1$ | MSM | MSM/MSE | $P(F(r-1, n-r) > F_{ob})$ |
   | Error | SSE | $n-r$ | MSE | | |
   | C.Total | C.SSTO | $n-1$ | | | |

   (2) $\alpha$-level LRT in Setting I

   > $H_0$ : vs $H_a$ : $H_0$ is false
   > Test statistic: $F = \frac{MSM}{MSE}$
   > Reject $H_0$ if $F > F_\alpha(r-1, n-r)$.

   is an $\alpha$-level LRT

   **Proof.** Applying LRT scheme in 1 (2) to setting I,
   $F = \frac{(SSE_* - SSE)/(r-r_*)}{MSE} = \frac{(C.SSTO - SSE)/(r-1)}{MSE} = \frac{SSM/(r-1)}{MSE} = \frac{MSM}{MSE}$.
   $P(F > F_\alpha(r-1, n-r)|H_0) = P(F(r-1, n-r) > F_\alpha(r-1, n-r)) = \alpha$.
   **Comment:** $p$-value: $P(F(r-1, n-r) > F_{ob})$.

   (3) Setting II
   Consider Model M with $1_n \notin \mathcal{R}(X)$ and $H_0 : \beta = 0$. Under $H_0$, M is reduced to $M_0 : Y \sim N(0, \sigma^2\Sigma)$. For M and this $H_0$ there is ANOVA table

   | Source | SS | DF | MS | F | $Pr > F$ |
   |---|---|---|---|---|---|
   | Model | SSM | $r$ | MSM | MSM/MSE | $P(F(r, n-r) > F_{ob})$ |
   | Error | SSE | $n-r$ | MSE | | |
   | U.Total | U.SSTO | $n$ | | | |

(4) $\alpha$-level LRT in Setting II

$H_0: \beta = 0$ vs $H_a: \beta \neq 0$
Test statistic: $F = \frac{MSM}{MSE}$
Reject $H_0$ if $F > F_\alpha(r, n-r)$.

**Proof.** Applying LRT scheme in 1 (2) to setting II,
$F = \frac{(SSE_* - SSE)/(r-r_*)}{MSE} = \frac{(U.SSTO-SSE)/r}{MSE} = \frac{SSM/r)}{MSE} = \frac{MSM}{MSE}$.
$P(F > F_\alpha(r, n-r)|H_0) = P(F(r, n-r) > F_\alpha(r, n-r)) = \alpha$.
**Comment:** $p$-value: $P(F(r, n-r) > F_{ob})$.

3. Global $F$-tests
The first question in study a particular linear model with data is "Is this model useful?"

(1) Regression with intercept
For regression $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1} + \epsilon$,
$H_0$ : The model is useless $\Longleftrightarrow H_0: \beta_i = 0$ for all $i = 1, ..., p-1$.
Under $H_0$ the model is reduced to $M_1$ on data: $N(1_n\beta_0, \sigma^2 I_n)$. We have $\alpha$-LRT

$H_0: \beta_i = 0$ for all $i = 1, ..., p-1$ vs $H_a: \beta \neq 0$ for some $i = 1, ..., p-1$
Test statistic: $F = \frac{MSM}{MSE}$
$p$-value: $P(F(p-1, n-p) > F_{ob})$.

This test can be carried out by using ANOVA table created by SAS

```
proc reg;  model y=x1 x2 x3;  run;
```

(2) Regression without intercept
For regression $y = \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$,
$H_0$ : The model is useless $\Longleftrightarrow H_0: \beta_i = 0$ for all $i = 1, ..., p$.
Under $H_0$ the model is reduced to $M_0$ on data: $N(0, \sigma^2 I_n)$ by $H_0$. We have $\alpha$-LRT

$H_0: \beta_i = 0$ for all $i = 1, ..., p$ vs $H_a: \beta \neq 0$ for some $i = 1, ..., p$
Test statistic: $F = \frac{MSM}{MSE}$
$p$-value: $P(F(p, n-p) > F_{ob})$.

This test can be carried out by using ANOVA table created by SAS

```
proc reg;  model y=x1 x2 x3/noint;  run;
```

(3) One-way ANOVA
For one-way ANOVA $y = \mu_1 x_1 + \cdots + \mu_p \mu_p + \epsilon$,
$H_0$ : The model is useless $\Longleftrightarrow H_0: \mu_i = \mu_j$ for all $(i, j)$.
Under $H_0$ the model is reduced to $M_1$ on data: $N(1_n\mu_1, \sigma^2 I_n)$. We have $\alpha$-LRT

$H_0: \mu_i = \mu_j$ for all $(i, j)$ vs $H_a: \mu_i \neq \mu_j$ for some $(i, j)$
Test statistic: $F = \frac{MSM}{MSE}$
$p$-value: $P(F(p-1, n-p) > F_{ob})$.

This test can be carried out by using ANOVA table created by SAS

```
proc anova;  class sample; model y=sample;  run;
```