L18: α-level LRTs

1. LRT test

   (1) Model, hypothesis and SS table
   For linear model $y = X\beta + e \sim N(X\beta, \sigma^2 I_n)$ where $X \in R^{n\times p}$ has full column rank and consistent $H_0 : H\beta = b$ where $H \in R^{q\times p}$ has full row rank, we have

SS	DF	MS	F
SSH	q	MSH	MSH/MSE
SSE	n-p	MSE	
SSE$_r$	n-p+q		

   (2) A LRT

   $H_0 : H\beta = b$ versus $H_a : H\beta \neq b$
   Test Statistics: $F = \frac{MSH}{MSE}$
   Reject H_0 if $F > c$

   is a LRT scheme.
   Proof: Note that $\max[L(\beta, \sigma^2) : \beta, \sigma^2] = \left(\frac{n}{2\pi e}\right)^{n/2} \text{SSE}^{-n/2}$ and $\max[L(\beta, \sigma^2) : H_0] = \left(\frac{n}{2\pi e}\right)^{n/2} \text{SSE}_r^{-n/2}$. So

$$\begin{aligned} \text{LR} &= \frac{\max[L(\beta,\sigma^2):\beta,\sigma^2]}{\max[L(\beta,\beta):H_0]} = \left(\frac{SSE_r}{SSE}\right)^{n/2} = \left(\frac{SSH+SSE}{SSE}\right)^{n/2} = \left(1 + \frac{SSH}{SSE}\right)^{n/2} \\ &= \left(1 + \frac{MSH\cdot q}{MSE\cdot(n-p)}\right)^{n/2} = \left(1 + \frac{q}{n-p} F\right)^{n/2} \end{aligned}$$

   is an increasing function of $F = \frac{MSH}{MSE}$. Thus the given scheme is a LRT scheme.

   (3) Other forms of F
   $F = \frac{(SSE_r - SSE)/((DF\ of\ SSE_r) - (DF\ of\ SSE)}{MSE}$ is a general form.
   From the HW, $F = \frac{(H\widehat{\beta}-b)'[H(X'X)^{-1}H']^{-1}(H\widehat{\beta}-b)/q}{MSE}$. This F is for $H_0 : H\beta = b$ specifically.

2. Null distribution: $F = \frac{MSH}{MSE} \overset{H_0}{\sim} F(q, n-p)$

   (1) Recall: $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$ is independent to $\widehat{\beta}$.

   (2) $\frac{SSH}{\sigma^2} \overset{H_0}{\sim} \chi^2(q)$
   With $M_1 = X(I - H^+H)$, $\frac{SSH}{\sigma^2} = (y - XH^+b)'\frac{XX^+ - M_1M_1^+}{\sigma^2}(y - XH^+b)$
   where $y - XH^+b \overset{H_0}{\sim} N(M_1\gamma, \sigma^2 I_n)$.
   (i) $\frac{XX^+ - M_1M_1^+}{\sigma^2}(\sigma^2 I_n)\frac{XX^+ - M_1M_1^+}{\sigma^2} = \frac{XX^+ - M_1M_1^+}{\sigma^2}$
   (ii) $[E(y - XH^+b)]'\frac{XX^+ - MM^+}{\sigma^2}[E(y - XH^+b)] = 0$
   (iii) $\text{tr}\left(\frac{XX^+ - M_1M_1^+}{\sigma^2}\sigma^2 I_n\right) = p - (p-q) = q$.
   By the tool: $z \sim N(\mu, \Sigma)$ and $A\Sigma A = A = A' \Longrightarrow x'Ax \sim \chi^2(\mu' A\mu, \text{tr}(A\Sigma))$,
   $\frac{SSH}{\sigma^2} \overset{H_0}{\sim} \chi^2(q)$.

   (3) $F = \frac{MSH}{MSE} \overset{H_0}{\sim} F(q, n-p)$.
   SSE and $SSH = (H\widehat{\beta} - b)'[H(X'X)^{-1}H']^{-1}(H\widehat{\beta} - b)$ are independent.
   So $\frac{SSH/(\sigma^2\cdot q)}{SSE/(\sigma^2\cdot(n-p))} \overset{H_0}{\sim} F(q, n-p)$, i.e., $F = \frac{MSH}{MSE} \overset{H_0}{\sim} F(q, n-p)$.

3. α-level LRTs

(1) The following is an α-level LRT

$H_0: H\beta = b$ versus $H_a: H\beta \neq b$
Test statistic: $F = \frac{MSH}{MSE}$
Reject H_0 if $F > F_\alpha(q, n-p)$

Proof
$$\begin{aligned} P(\text{Type I Error}) &= P(\text{Rejecting } H_0 | H_0 \text{ is true}) \\ &= P(F > F_\alpha(q, n-p) | H_0) \\ &= P(F(q, n-p) > F_\alpha(q, n-p)) = \alpha. \end{aligned}$$

(2) p-value

$H_0: H\beta = b$ versus $H_a: H\beta \neq b$
Test statistic: $F = \frac{MSH}{MSE}$
p-value: $P(F(q, n-p) > F_{ob})$

Proof
$$\begin{aligned} & \alpha\text{-level test rejects } H_0 \Longleftrightarrow F_{ob} > F_\alpha(q, n-p) \\ \Longleftrightarrow\ & P(F(q, n-p) > F_{ob}) < P(F(q, p) > F_\alpha(q, n-p)) = \alpha \\ \Longleftrightarrow\ & P(F(q, n-p) > F_{ob}) \text{ is observed significance level.} \end{aligned}$$

(3) Implementation

In

```
proc reg;
  model y=x1 x2/noint noprint;
  test 2*x1-x2=1;
  run;
```

the components of β are referred to as x1 and x2.

In

```
proc reg;
  model y=x1 x2/noprint;
  test intercept-x1=0, x2=1;
  run;
```

the components of β are referred to as intercept, x1 and x2.

L19: Tests on the usefulness of the model

1. Regression with intercept

   (1) Model, hypothesis and ANOVA table
   For $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1} + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$, $H_0:$ The model is useless is equivalent to $H_0: \beta_i = 0$ for all $i = 1, ..., p-1$. Under H_0 the reduced model $y = \beta_0 + \epsilon$ has $SSE_r = \mathbf{y}'\left(I_n - \frac{\mathbf{1}_n\mathbf{1}_n'}{n}\right)\mathbf{y} = \sum_{i=1}^n (y_i - \overline{y})^2 \stackrel{def}{=\!=}$ C.SSTO for Corrected Sum of Squares for TOtal variation in $\mathbf{y}$. $SSH = C.SSTO - SSE \stackrel{def}{=\!=} SSM$ is the variation explained by the model. Thus we have ANOVA table

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

   (2) Test on the usefulness of the model and SAS
   With the ANOVA table in (1), test on the usefulness of the model can be carried out.

   $H_0: \beta_i = 0\,\forall\, i = 1, .., p-1$ versus $H_a: \beta_i \neq 0\,\exists\, i = 1, .., p-1$
   Test statistic: $F = \frac{MSM}{MSE}$
   p-value: $P(F(p-1, n-p) > F_{ob})$

   (3) SAS
   The ANOVA table is one of basic output of SAS

```
data a;
    infile "D:\ex.txt";
    input y x1 x2 @@;
proc reg;
    model y=x1 x2;
    run;
```

2. Regression without intercept

   (1) Model, hypothesis and ANOVA table
   For $y = \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$, $H_0:$ The model is useless is equivalent to $H_0: \beta_i = 0$ for all $i = 1, ..., p$. Under H_0 the reduced model $y = \epsilon$ has $SSE_r = \mathbf{y}'I_n\mathbf{y} = \sum_{i=1}^n (y_i - 0)^2 \stackrel{def}{=\!=}$ U.SSTO for Un-corrected Sum of Squares for TOtal variation in $\mathbf{y}$. $SSH = U.SSTO - SSE \stackrel{def}{=\!=} SSM$ is the variation explained by the model. We have ANOVA table

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

   (2) Test on the usefulness of the model
   With ANOVA table in (1), the test on the usefulness of the model can be carried out.

$H_0:\ \beta_i = 0\,\forall\, i = 1,..,p$ versus $H_a:\ \beta_i \neq 0\,\exists\, i = 1,..,p$
Test statistic: $F = \frac{MSM}{MSE}$
p-value: $P(F(p,\, n-p) > F_{ob})$

(3) SAS
The ANOVA table is one of standard output of SAS.

```
data a;
    infile "D:\ex2.txt";
    input y x1 x2 @@;
proc reg;
    model y=x1 x2/noint;
    run;
```

3. One-way ANOVA

(1) Model, hypothesis and ANOVA table
In one-way ANOVA $y = \mu_1 x_1 + \cdots + \mu_p x_p + \epsilon$, y is the response to a factor with p levels, μ_i is the mean response to the ith level of the factor and x_i is the indicator for the ith level, i.e., $x_i = \begin{cases} 1 & y \text{ is the response to the } i \text{ th level of factor} \\ 0 & \text{Otherwise} \end{cases}$.

$H_0:$ The model is useless is equivalent to $H_0:\ \mu_1 = \cdots = \mu_p \stackrel{def}{=\!=} \mu_.$. Under H_0, the model becomes $y = \mu_. 1 + \epsilon$ with $SSE_r = \sum_i (y_i - \overline{y})^2 =$ C.SSTO. SSE is also denoted as SSW for the total variation of **y** within each response and SSH=SSM is denoted as SSB for the total variation of **y** between responses. Thus

Source	SS	DF	MS	F	p
Model, Between	SSM=SSB	p-1	MSM	MSM/MSE	$P(F(p-1,\, n-p) > F_{ob})$
Erorr, Within	SSE=SSW	n-p	MSE		
C.Total	C.SSTO	n-1			

(2) SAS
With ANOVA table in (1), one can test the usefulness of the model. The ANOVA table is one of standard output for SAS.

```
data a;
    infile "D:\ex.txt";
    input y Tt $ @@;
proc anova;
    class Tt;
    model y=Tt;
    run;
```