L12 Estimation and prediction

1. Estimator and predictor

   (1) Unknown random variable and unknown parameter
   Consider model $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$.
   With $x_0' = (1, x_{01}, \cdots, x_{0k})$,
   $y(x_0) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} + \epsilon \sim N(\beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k}, \sigma^2)$,
   a future response, is an unknown random variable.
   But $E[y(x_0)] = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k}$ is an unknown parameter.
   $y(x_0) = x_0'\beta + \epsilon$ and $E(y(x_0)) = x_0'\beta$.

   (2) Estimator and predictor
   If $E[y(x_0)] = x_0'\beta$ is estimated by statistic T, then T is an estimator for $E[y(x_0)] = x_0'\beta$.
   If $y(x_0)$ is "estimated" by statistic T, then T is a predictor for $y(x_0)$.

   (3) Unbiased estimator and unbiased predictor
   T is an unbiased estimator (UE) for $E[y(x_0)]$ if $E(T) = E[y(x_0)]$.
   T is an unbiased predictor (UP) for $y(x_0)$ if $E[T - y(x_0)] = 0$.

   (4) Equivalency
   T is UE for $E[y(x_0)] \Longleftrightarrow T$ is UP for $y(x_0)$
   Proof. T is UE for $E[y(x_0)] \overset{def}{\Longleftrightarrow} E(T) = E[y(x_0)] \Longleftrightarrow E(T) - E[y(x_0)] = 0$
   $\Longleftrightarrow E[T - y(x_0)] = 0 \overset{def}{\Longleftrightarrow} T$ is UP for $y(x_0)$.

   Ex1: Let $\widehat{y}(x_0) = x_0'\widehat{\beta} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{0k}$.
   Then $E[\widehat{y}(x_0)] = E[x_0'\widehat{\beta}] = x_0'\beta = E[y(x_0)]$.
   So $\widehat{y}(x_0)$ is an UE for $E[y(x_0)]$. By (3) $\widehat{y}(x_0)$ is also UP for $y(x_0)$.
   Comment: $\widehat{y}_i$, $i = 1, ..., n$, and $\widehat{y}(x_0)$ are predictors for y_i, $i = 1, ..., n$, and $y(x_0)$.
   This is why SAS "proc reg; model y=x1 x2/p;" produces $\widehat{y}_i$,..., and $\widehat{y}(x_0)$.

2. Confidence interval and prediction interval

   (1) Confidence intervals
   Recall: If (L_1, U_1), $(-\infty, U_2)$ and (L_2, ∞) are random intervals such that
   $P(L < E[y(x_0)] < U) \geq 1 - \alpha$, $P(-\infty < E[y(x_0)] < U_2) \geq 1 - \alpha$ and
   $P(L_2 < E[y(x_0)] < \infty) \geq 1 - \alpha$, then they are two-sided, lower-sided and upper-sided confidence intervals for $E[y(x_0)]$ with confidence coefficient $1 - \alpha$.

   (2) Prediction intervals
   If (L_1, U_1), $(-\infty, U_2)$ and (L_2, ∞) are random intervals such that
   $P(L < y(x_0) < U) \geq 1 - \alpha$, $P(-\infty < y(x_0) < U_2) \geq 1 - \alpha$ and
   $P(L_2 < y(x_0) < \infty) \geq 1 - \alpha$, then they are two-sided, lower-sided and upper-sided prediction intervals for $y(x_0)$ with confidence coefficient $1 - \alpha$.

   (3) A framework
   If T and S are two statistics such that $\frac{T - y(x_0)}{S} \sim t(df)$, then $T \pm t_{\alpha/2}(df)S$, $(-\infty, T + t_\alpha(df)S)$ and $(T - t_\alpha(df)S, \infty)$ are PIs for $y(x_0)$ with confidence coefficient $1 - \alpha$.

Proof. We show second one.

$$
\begin{aligned}
1-\alpha &= P(-t_\alpha(df) < t(df) < \infty) = P\left(-t_\alpha(df) < \tfrac{T-y(x_0)}{S} < \infty\right) \\
&= P\left(t_\alpha(df) > \tfrac{y(x_0)-T}{S} > -\infty\right) = P\left(T + t_\alpha(df)S > y(x_0) > -\infty\right) \\
&= P\left(-\infty < y(x_0) < T + t_\alpha(df)S\right).
\end{aligned}
$$

3. Formulas for PIs for $y(x_0)$

   (1) $\widehat{y}(x_0)$ and $S_{\widehat{y}(x_0)-y(x_0)}$
   $\widehat{y}(x_0) = x_0'\widehat{\beta} \sim N\left(x_0'\beta,\, \sigma^2 x_0'(X'X)^{-1}x_0\right)$ and $y(x_0) \sim N(x_0'\beta,\, \sigma^2)$ are independent.
   So $\widehat{y}(x_0) - y(x_0) \sim N\left(0,\, \sigma^2 + \sigma^2 x_0'(X'X)^{-1}x_0\right)$
   where $\mathrm{var}(\widehat{y}(x_0) - y(x_0)) = \mathrm{var}(\widehat{y}(x_0)) + \mathrm{var}(y(x_0)) = \sigma^2 + \sigma^2 x_0'(X'X)^{-1}x_0$
   has UE $S^2_{\widehat{y}(x_0)-y(x_0)} = MSE\left[1 + x_0'(X'X)^{-1}x_0\right]$.
   $S_{\widehat{y}(x_0)-y(x_0)}$ is called the standard error of $\widehat{y}(x_0) - y(x_0)$.

   (2) $\frac{\widehat{y}(x_0)-y(x_0)}{S_{\widehat{y}(x_0)-y(x_0)}} \sim t(n-p)$
   $\frac{\widehat{y}(x_0)-y(x_0)}{\sqrt{\sigma^2\left[1+x_0'(X'X)^{-1}x_0\right]}} \sim N(0,\, 1^2)$ is independent to $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.
   So $\frac{\widehat{y}(x_0)-y(x_0)}{\sqrt{\sigma^2\left[1+x_0'(X'X)^{-1}x_0\right]}} \frac{1}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t(n-p)$
   Thus $\frac{\widehat{y}(x_0)-y(x_0)}{S_{\widehat{y}(x_0)-y(x_0)}} \sim t(n-p)$.

   (3) Formulas for PIs for $y(x_0)$

$\widehat{y}(x_0) \pm t_{\alpha/2}(n-p)S_{\widehat{y}(x_0)-y(x_0)}$	is a $1-\alpha$ PI for $y(x_0)$
$(-\infty,\, \widehat{y}(x_0) + t_\alpha(n-p)S_{\widehat{y}(x_0)-y(x_0)})$	is a $1-\alpha$ lower-sided PI for $y(x_0)$
$(\widehat{y}(x_0) - t_\alpha(n-p)S_{\widehat{y}(x_0)-y(x_0)},\, \infty)$	is a $1-\alpha$ PI for $y(x_0)$

$$
\widehat{y}(x_0) = x_0'\widehat{\beta} = \begin{cases} \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{0k} & \text{With intercept} \\ \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{0k} & \text{Without intercept} \end{cases}
$$

$$
S_{\widehat{y}(x_0)-y(x_0)} = \sqrt{MSE\ \left[1 + x_0'(X'X)^{-1}x_0\right]}
$$

L13 Partial F-test

1. Intervals

   (1) $1-\alpha$ P.I. for $y(x_0)$
   $\widehat{y}(x_0) \pm t_{\alpha/2}(n-p)s_{\widehat{y}(x_0)-y(x_0)}$.
   Attach x_0 to data then

```
proc reg;
    model y=x1 x2 x3/cli alpha=0.10;
    run;
```

   Presentation:

   $\widehat{y}(x_0) \pm t_{0.05}(17)s_{\widehat{y}(x_0)-y(x_0)} = dd \pm dd \times \sqrt{MSE + dd^2} = (dd,\, dd)$
   is a 90% PI for $y(x_0)$.

   (2) $1-\alpha$ C.I. for $E(y(x_0))$
   $\widehat{y}(x_0) \pm t_{\alpha/2}(n-p)s_{\widehat{y}(x_0)}$. Attach x_0 to data then

```
proc reg;
    model y=x1 x2 x3/clm alpha=0.10;
    run;
```

   Presentation:

   $\widehat{y}(x_0) \pm t_{0.05}(17)s_{\widehat{y}(x_0)} = dd \pm dd \times dd = (dd,\, dd)$
   is a 90% CI for $E[y(x_0)]$.

   (3) $1-\alpha$ C.I. for β_i
   $\widehat{\beta}_i \pm t_{\alpha/2}(n-p)s_{\widehat{\beta}_i}$.

```
proc reg;
    model y=x1 x2 x3/clb alpha=0.10;
    run;
```

   Presentation:

   $\widehat{\beta}_i \pm t_{0.05}(17)s_{\widehat{\beta}_i} = dd \pm dd \times dd = (dd,\, dd)$
   is a 90% CI for β_i.

   (4) $1-\alpha$ C.I. for σ^2
   $\left(\frac{SSE}{\chi^2_{\alpha/2}(n-p)},\, \frac{SSE}{\chi^2_{1-\alpha/2}(n-p)}\right)$

   Proof. Note that $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$. So

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{1-\alpha/2}(n-p) < \chi^2(n-p) < \chi^2_{\alpha/2}(n-p)\right) \\ &= P\left(\chi^2_{1-\alpha/2}(n-p) < \frac{SSE}{\sigma^2} < \chi^2_{\alpha/2}(n-p)\right) \\ &= P\left(\frac{SSE}{\chi^2_{\alpha/2}(n-p)} < \sigma^2 < \frac{SSE}{\chi^2_{1-\alpha/2}(n-p)}\right) \end{aligned}$$

   Presentation:

   $\left(\frac{SSE}{\chi^2_{0.05}(24)},\, \frac{SSE}{\chi^2_{0.95}(24)}\right) = \left(\frac{dd}{dd},\, \frac{dd}{dd}\right) = (dd,\, dd)$
   is a 90% CI for σ^2.

2. ANOVA for $H_0 : \beta_{k-q+1} = \cdot = \beta_p = 0$

   (1) SS
   Model wih $x_1, ..., x_k$ has SSE, $SSE(x_1, ..., x_k) = y'(I - H)y$, for the variation in y unexplained by $x_1, ..., x_k$ with DF$= n - p$
   Model with $x_1, .., x_{k-q}$ has SSE$_r$, $SSE(x_1, .., x_{k-q}) = y'(I - H_r)y$, for the variation in y unexplained by $x_1, ..., x_{k-q+1}$ with DF$= n - (p - q)$.
   SSH $= SSE_r - SSE = y'(H - H_r)y$ is for the variation in y explained by $x_{k-q+1}, ..., x_k$ with $DF = [n - (p - q)] - (n - p_= q$, the number of β_i in H_0.

   (2) F-distribution
   It can be shown that under H_0, $F = \frac{MSH}{MSE} = \frac{SSH/q}{SSE/(n-p)} \sim F(q, n - p)$.

   (3) ANOVA

Source	SS	DF	MS	F	p
Hypothesis (N)	SSH	q	MSH	MSH/MSE	$P(F(q, n - p) > F_{ob})$
Error (D)	SSE	n-p	MSE		
Error (R)	SSE$_r$	n-(p-q)			

3. Partial F-test

   (1) α-level LRT
   For model with x_1, x_2, x_3, x_4

   $H_0 :\ \beta_2 = \beta_4 = 0$ vs $H_a :\ \beta_i \neq 0$ for some $i = 2, 4$
   Test statistic: $F = \frac{MSH}{MSE}$
   Reject H_0 if $F > F_\alpha(2, n - p)$.

   (2) p-value approach
   For model with x_1, x_2, x_3, x_4

   $H_0 :\ \beta_2 = \beta_4 = 0$ vs $H_a :\ \beta_i \neq 0$ for some $i = 2, 4$
   Test statistic: $F = \frac{MSH}{MSE}$
   p-value: $P(F(2, n - p) > F_{ob})$.

   (3) Get ANOVA for the test

```
proc reg;
   model y=x1 x2 x3 x4;
   run;
```

$\Longrightarrow SSE, MSE$

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

$\Longrightarrow SSE_r$

   $SSH = SSE_r - SSE\ MSH = SSH/2$.