NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (16)

MLR: Inference with t and F [3.6]
MLR: Confidence and prediction interval (not in book)

Mette Langaas

To be lectured: February 25, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

Inference in MLR

- $H_0 : \beta_i = 0$ vs. $H_1 : \beta_i <> \neq 0$. Test based on t-distribution. Confidence intervals from same distribution.
- H_0: no regression, vs H_1: a regression. Test based on MSR and MSE ratio and F-distribution
- H_0: subset of βs zero, vs H_1: at least one different. Same as comparing two nested MLR models. Test based on difference in SSR between the two models compared with the SSE of the larger model. Partial F-test.
- Confidence intervals for the expected regression hyperplane and prediction interval for new observation. Based on t-distribution.

TMA4267 L16

MLR Inference — t-test / F-test / CI & PI

need

$\varepsilon \sim N_n(0, \sigma^2 I)$

$$\underset{n\times 1}{Y} = \underset{n\times p}{X}\,\underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \sim N_p(\beta, \underline{\sigma^2 (\overbrace{X^T X}^{C})^{-1}})$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = Y^T (I-H) Y \sim \sigma^2 \chi^2_{n-p}$$

$$E\left(\frac{SSE}{\sigma^2}\right) = n-p$$

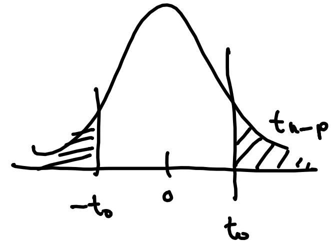
$\hat{\beta}$ and SSE are independent

$S^2 = \frac{SSE}{n-p}$ is an unbiased estimator for σ^2.

- Model assessment: Part 5 Ch 7
- DOE X? Part 6
- Model selection Part 7

1) $H_0: \beta_i = 0$ vs $H_1: \beta_i \neq 0$ $(\beta_i > 0, \beta_i < 0)$

$$T_i = \frac{\hat{\beta}_i - 0}{\sqrt{\widehat{Var}(\hat{\beta}_i)}} = \frac{\hat{\beta}_i - 0}{\sqrt{[C^{-1}]_{ii}\, s}} \sim t_{n-p}$$



This test is conditonal on the fitted model, $\beta_1, \dots, \beta_p$ fitted, it means the test is given that the remaining covariates are in the model. If we fit x_1, x_2, x_3 or x_1, x_2, x_3, x_4 then $\hat{\beta}_3$ will not be the same in the models. Why? $\hat{\beta}_3 = \left[(X^T X)^{-1} X^T Y\right]_{[3]}$ is not only dependent on x_3, but on all the x's in the model. If $(X^T X)$ is diagonal then $\hat{\beta}_3$ only depend on x_3.

R: lm, followed by summery

2) H_0: no regression vs H_1: at least one

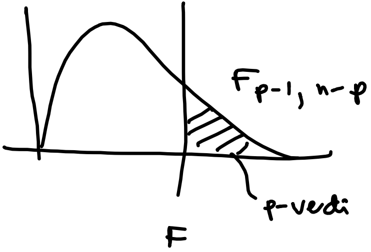
$\beta_2 = \beta_3 = \dots = \beta_p = 0$ $\qquad$ $\beta_i \neq 0 \quad i = 2, \dots, p$

L14: $SSR = Y^T(H - \frac{1}{n}J)Y \sim \sigma^2 \chi^2_{p-1}$ under H_0

and, SSR and SSE are independent.

$$F = \frac{\frac{SSR}{p-1}}{\frac{SSE}{n-p}} = \frac{MSR}{MSE} \sim F_{p-1, n-p}$$

· onesided



R: lm, followed by anova.

3) One last F-test, comparing two nested models → partial F-test.

$H_0: \beta_{r+1} = \beta_{r+2} = \cdots = \beta_p = 0$

H_1: at least one $\beta_i \neq 0 \quad i = r+1, \ldots, p$

This is the same as comparing

Model full : $\beta_2, \ldots, \beta_r, \beta_{r+1}, \ldots, \beta_p$

Model reduced : $\beta_2, \ldots, \beta_r$

Full: $\quad H = X(X^T X)^{-1} X^T \qquad X = n \times p$ full

$$SSR = SST - SSE = SST - Y^T (I - H) Y$$

Reduced model A: $\quad \underset{n \times r}{X_A} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1r} \\ 1 & \vdots & & \\ \vdots & \vdots & & x_{nr} \end{bmatrix}$

$$H_A = X_A (X_A^T X_A)^{-1} X_A^T$$

$$SSR_A = SST - SSE_A = SST - Y^T (I - H_A) Y$$

The contribution for the set $\beta_{r+1}, \ldots, \beta_p$ can be found as

$$R(\beta_{r+1}, \ldots, \beta_p \mid \beta_2, \ldots, \beta_r) = \underset{\text{full}}{SSR} - SSR_A$$

sequential sums-of-squares of regression

If we find the distribution of $R(\beta_{r+1},\dots,\beta_p \mid \beta_2,\dots,\beta_r) = SSR - SSR_A$ we may compare it to SSE (full model).

It can be seen that (a–d)

$$R(\beta_{r+1},\dots,\beta_p \mid \beta_2,\dots,\beta_r) = SSR - SSR_A \sim \sigma^2 \chi^2_{\underline{p-r}}$$

when $H_0: \beta_{r+1} = \dots = \beta_p = 0$ is true.

a) $SSR - SSR_A = Y^T (H - H_A) Y$

b) $(H - H_A)$ is symmetric and idempotent

c) $\operatorname{rank}(H - H_A) = p - r$

d) $(H - H_A)(Y - X\beta) = (H - H_A)Y$ standard tricks to get $E(Y - X\beta) = 0$

and in addition:

e) $\underbrace{(H - H_A)}_{SSR-SSR_A}\underbrace{(I - H)}_{SSE} = 0$ so SSE and $(SSR - SSR_A)$ are independent

Test for H_0:

$$\frac{R(\beta_{r+1},\dots,\beta_p \mid \beta_2,\dots,\beta_r) / (p-r)}{\dfrac{SSE}{n-p}} \sim F_{p-r,\, n-p}$$

R: lm for full model
lm for reduced model } anova(lmRed, lmfull)

Inference about $E(Y_0)$ and Y_0

$\underset{p\times 1}{X_0}$ new observation — with potential new response Y_0.

$$\underbrace{x_0^T \beta}_{\mu_0} = E(Y_0) \qquad Cov(\hat{\beta}) = (X^TX)^{-1}\sigma^2$$

$$\underbrace{x_0^T \hat{\beta}}_{1\times 1} \sim N_1\big(x_0^T\beta,\ \underbrace{x_0^T (X^TX)^{-1} x_0 \sigma^2}\big)$$

$$Cov(x_0^T\hat{\beta}) = x_0^T Cov(\hat{\beta}) x_0$$

$$\frac{x_0^T\hat{\beta} - x_0^T\beta}{\sqrt{x_0^T (X^TX)^{-1} x_0}\, s} \sim t_{n-p}$$

$(1-\alpha)100\%$ CI (confidence interval) for $E(Y_0)$ at x_0

$$x_0^T\hat{\beta} \pm t_{n-p}\left(\tfrac{\alpha}{2}\right)\sqrt{x_0^T (X^TX)^{-1} x_0}\, s$$

What about a new observation Y_0?

$$\underbrace{(Y_0}_{N(X_0^T\beta,\,\sigma^2 I)} - \underbrace{x_0^T\hat\beta}_{\text{last page}}) \sim N_1\left(0, \left(1 + \underbrace{x_0^T (X^TX)^{-1} x_0}\right)\sigma^2\right)$$

$(1-\alpha)100\%$ prediction interval (PI) for Y_0

$$x_0^T\hat\beta \pm t_{n-p}\left(\frac{\alpha}{2}\right)\sqrt{1 + x_0^T (X^TX)^{-1} x_0}\; s$$

R: lm

predict

Prediction

Similar to the simple linear regression case.

- New value for the covariates, $\boldsymbol{x}_0$.
- $(1-\alpha)\cdot 100$ % CI for the regression line $\mathrm{E}(Y_0|\boldsymbol{x}_0)$:

$$[\boldsymbol{x}_0^T\hat{\beta} \pm t_{n-p}(\frac{\alpha}{2})\sqrt{\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0 s^2}]$$

- $(1-\alpha)\cdot 100$ % prediction interval for new observation Y_0:

$$[\boldsymbol{x}_0^T\hat{\beta} \pm t_{n-p}(\frac{\alpha}{2})\sqrt{(1+\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0)s^2}]$$

R:

`predict(fittedlm,newdata=thenewdata,interval=''prediction'')` or `interval=''confidence''`. New observations placed in `newdata`.

Gasoline example

- Aim: predict the gasoline consumption for a car driving 100 km at speed 80 km/h.
- Measurements:
  - y_i= gasoline consumption (liter)
  - x_{1i}=distance (km)
  - x_{2i}=speed (km/h)

$$X = \begin{bmatrix} 1 & 100 & 60 \\ 1 & 50 & 70 \\ 1 & 70 & 80 \\ 1 & 120 & 70 \\ 1 & 100 & 90 \end{bmatrix} \text{ and } Y = \begin{bmatrix} 12.5 \\ 8.0 \\ 11.5 \\ 14.5 \\ 13.5 \end{bmatrix}$$

Gasoline question in R

```
> gasex=data.frame(c(12.5,8,11.5,14.5,13.5),
                   c(100,50,70,120,100),
                   c(60,70,80,70,90))
> colnames(gasex)=c("gasoline","distance","speed")
> fit=lm(gasoline~.,data=gasex)> fit
Coefficients:
(Intercept)     distance        speed
    0.75657      0.08698      0.04850

> newobs=data.frame(distance=100,speed=80)
> predict(fit,newdata=newobs,interval="predict")
       fit      lwr      upr
1 13.33479 9.873419 16.79617
> predict(fit,newdata=newobs,interval="confidence")
       fit     lwr      upr
1 13.33479 11.6272 15.04238
```