2.4 Confidence intervals and prediction intervals in Linear Models

Confidence and prediction intervals

In this section, we will calculate the confidence and prediction intervals for the Linear Gaussian Models.

Confidence vs Prediction Interval

Given a certain vector of predictors x^*, we want to find a **confidence interval** for the conditional mean ${x^*}^\top\beta$ and a **prediction interval** for a future unobserved observation $Y^* = {x^*}^\top\beta + \epsilon^*$where ϵ^* is an error independent of ϵ_i, $i = 1, ..., n$, drawn from $N(0, \sigma^2)$. Below we give the statistical distributions from which the confidence and prediction intervals are derived from.

Confidence interval

First, note that Y^* is Gaussian. Its mean is

$$E({x^*}^\top\hat{\beta}) = {x^*}^\top\boldsymbol{E}[\hat{\boldsymbol{\beta}}] = {x^*}^\top\beta$$

and its variance

$$\mathrm{Var}({x^*}^\top\hat{\beta}) = \mathrm{Cov}({x^*}^\top\hat{\beta}) = {x^*}^\top\mathrm{Cov}(\hat{\beta})x^* = \sigma^2{x^*}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*$$

where $\boldsymbol{X}$ is the design matrix for the fitted linear model.

So we have

$${x^*}^\top\hat{\beta} \sim N({x^*}^\top\beta, \sigma^2{x^*}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*)$$

or

$$\frac{{x^*}^\top\hat{\beta} - {x^*}^\top\beta}{\sigma\sqrt{{x^*}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*}} \sim N(0, 1).$$

It can be shown that $\hat{\beta} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ and $\boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$ are independent (by showing that any row of $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is orthogonal to any row of $\boldsymbol{I} - \boldsymbol{H}$.

It then follows that ${x^*}^\top\hat{\boldsymbol{\beta}}$ and $(n-p)\hat{\sigma}^2/\sigma^2$ are independent. Since $(n-p)\hat{\sigma}^2/\sigma^2$ has a χ^2_{n-p}

distribution, the quotient of the two has a Student-t distribution with $(n-p)$ degrees of freedom:

$$\frac{x^{*\top}\hat{\beta} - x^{*\top}\beta}{\sigma\sqrt{x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*}} \Big/ \sqrt{\frac{\frac{(n-p)\hat{\sigma}^2}{\sigma^2}}{n-p}} \sim t_{n-p}$$

or

$$\frac{x^{*\top}\hat{\beta} - x^{*\top}\beta}{\hat{\sigma}\sqrt{x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*}} \sim t_{n-p}.$$

Prediction intervals

Define $\hat{Y}^* = x^{*\top}\hat{\beta}$ and note that $E(Y^* - \hat{Y}^*) = 0$. Since $x^{*\top}\hat{\beta}$ and ϵ^* are independent,

$$\begin{aligned}
\mathrm{Var}(Y^* - \hat{Y}^*) &= \mathrm{Var}(x^{*\top}\hat{\beta}) + \mathrm{Var}(\epsilon^*) \\
&= \sigma^2 x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^* + \sigma^2 \\
&= \sigma^2(1 + x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*).
\end{aligned}$$

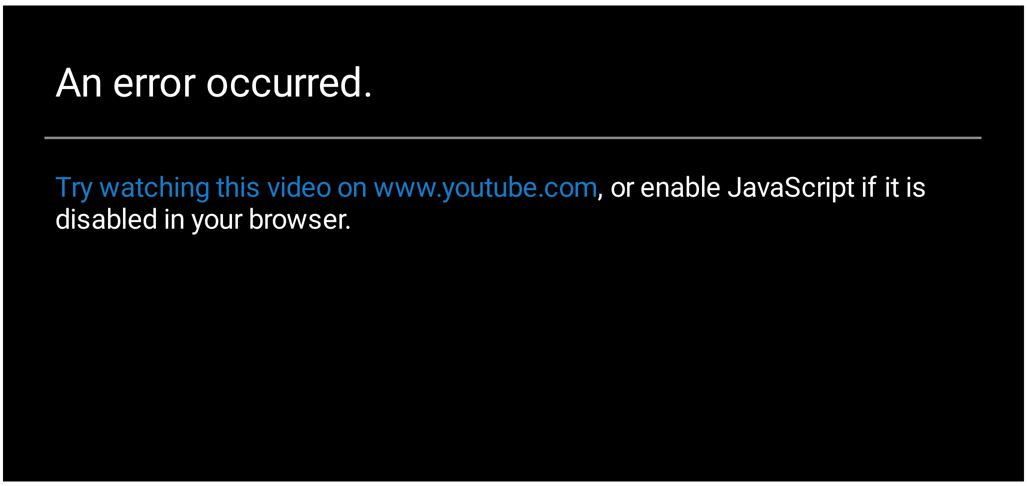
With a similar argument as above, it can be shown that $Y^* - \hat{Y}^*$ and $(n-p)\hat{\sigma}^2/\sigma^2$ are independent. Thus

$$\frac{Y^* - \hat{Y}^*}{\sigma\sqrt{1 + x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*}} \Big/ \sqrt{\frac{\frac{(n-p)\hat{\sigma}^2}{\sigma^2}}{n-p}} \sim t_{n-p}.$$

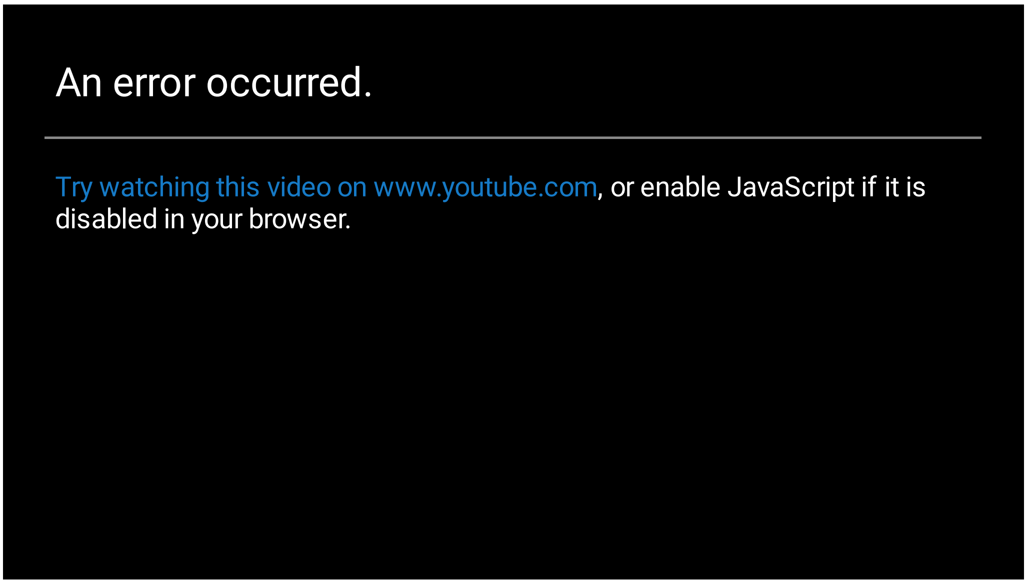
and upon simplifying

$$\frac{Y^* - \hat{Y}^*}{\hat{\sigma}\sqrt{1 + x^{*\top}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}x^*}} \sim t_{n-p}.$$

You may wish to watch the following example video below to help you reinforce your interpretation of this topic.



How to calculate confidence and prediction intervals for simple linear regression predictions in R



Confidence and prediction interval for Linear Gaussian Models

Example: Sydney maximum temperatures

In R, $predict()$ is a general function while $predict.lm()$ is specifically for prediction using a linear model object. The help file under $predict.lm()$ will give you the details of required and optional arguments.

- first argument (required) is a fitted linear model;
- **`newdata`** argument: data frame, contains values at which predictions are to be made;
- If **`newdata`** is omitted, predictions are made for the original predictors;
- **`se.fit`** argument: if $\texttt{se.fit} = \mathrm{T}$ estimated standard errors of the predictions are returned;

The value returned by $predict()$ for a linear model object is a list containing predictions and standard errors if $\texttt{se.fit} = \mathrm{T}$, otherwise just the predictions are returned.

Consider now the Sydney maximum temperature dataset **`mos.df`**:

```
mos.df <- read.table("/course/data/mos.df.txt", header=TRUE, quote="\"")
head(mos.df)
```

```
  Maxtemp  Modst  Modsp  Modthik
1    14.3  288.6 1000.8   5591.2
2    15.2  284.6 1010.8   5438.6
3    18.7  285.7 1011.0   5505.2
```

```
4    18.4  285.8 1004.5   5560.3
5    20.9  286.5 1002.9   5530.8
6    23.4  287.6 1003.4   5558.3
```

Here the response is **`Maxtemp`** and the predictors are **`Modst`**, **`Modsp`** and **`Modthik`**.

Suppose we want predictions of maximum temperature when

- $\texttt{Modst} = \texttt{285.1}$, $\texttt{Modsp} = \texttt{1021.2}$ and **`Modthik`**=5380.4
- $\texttt{Modst} = \texttt{288.0}$, $\texttt{Modsp} = \texttt{1026.8}$ and **`Modthik`**=5388.4

To use the $predict.lm()$ function, we set up a data frame containing our new data (one row for each of our two predictions).

Then, using the $predict.lm()$ function we can obtain the confidence and prediction intervals as follows:

```
mosnew.df <-data.frame(Modst=c(285.1,288.0), Modsp=c(1021.2,1026.8), Modthik=c(5380.4,5388.4))

mos.df <- read.table("/course/data/mos.df.txt", header=TRUE, quote="\"")
mos.lm <- lm(Maxtemp ~ ., mos.df)

mos.pred <- predict.lm(mos.lm, newdata=mosnew.df, se.fit=T, interval="confidence", level=0.95)
mos.pred

mos.pred <- predict.lm(mos.lm, newdata=mosnew.df, se.fit=T, interval="prediction", level=0.95)
mos.pred
```

```
$fit
       fit      lwr      upr
1 15.50811 14.81804 16.19818
2 15.85861 15.09306 16.62416

$se.fit
        1         2
0.3509169 0.3892985

$df
[1] 365

$residual.scale
[1] 3.009404
```

```
$fit
       fit      lwr      upr
1 15.50811 9.550067 21.46615
2 15.85861 9.891352 21.82587

$se.fit
        1         2
0.3509169 0.3892985
```

```
$df
[1] 365

$residual.scale
[1] 3.009404
```

Activity in R: Plotting confidence and prediction intervals

Consider the following generated **x**-predictor and **y**-response vectors:

```
x<-rnorm(15)
x

## [1] -0.4723019  2.3502426 0.5491130  0.3013668 -1.7001447 -2.1002455
## [7]  1.0016725  1.1004800 0.5962411 -0.5250126  1.1277512  0.5301167
## [13] 0.1709285 -1.2492901 0.2313901

y<-x+rnorm(15)
y

## [1] -0.5248888  2.5948113 -1.5645586  0.9136845 -2.2149522 -2.8905325
## [7] -0.8397268  0.7740053  0.7110674 -0.8100348  0.6468322 -0.6134713
## [13] 0.5671229 -2.1764990  1.5912105
```

Next, define the new values of **x** for which your predictions will be calculated:

```
new <- data.frame(x = seq(-3, 3, 0.5))
```

In this activity

1. Predict the values of **y** given your new values of **x**
2. Find confidence and prediction intervals
3. Visualise your intervals using the **matplot**() function in R

Additional activity

Question *Submitted Mar 16th 2023 at 9:47:28 pm*

Which of the following is true of confidence and predication intervals:

- [ ] the prediction interval is narrower than the confidence interval;
- [x] the confidence interval refers to the confidence interval for the conditional mean, in our notation, ${x^*}^\top \beta$;
- [x] the confidence interval can be calculated in R using the $predict()$ function.