MLR Inference Review

DS 301

Iowa State University

Today's Agenda

- HW 1 due by 11:59 pm today (on Canvas).
- HW 2 will be posted at midnight. Due next Wednesday.
- Hypothesis Testing Review / confidence intervals
- Multiple Testing Problem

Recap: How good are our least square estimates in the linear regression model?

$j = 0, 1, \dots, p.$

- Unbiasedness: $E(\hat{\beta}_j) = \beta_j$. Least square estimates are unbiased estimate of the true population parameters $\beta_0, \beta_1, \dots, \beta_p$.
  - Idea: Suppose we fit a linear regression line on a data set and we obtain $\hat{\beta}_1$. If we repeat this process for a huge number of datasets and average all of the $\hat{\beta}_1$'s we obtain, the average would exactly equal β_1.
  - An unbiased estimator does not systematically over or under-estimate the true parameter.
- Standard error: Quantifies the how far off a single estimate of $\hat{\beta}_j$ will be from β_j. We denote this as $se(\hat{\beta}_j)$. These depend on estimate of σ^2, which represents the variance of ϵ (and Y).

Correction from previous slides

$se(\hat{B_1})$, $se(\hat{B_2})$.

$\hat{\sigma}^2$

To keep things simple, the formula for $\hat{\sigma}^2$ we will be using is:

$$\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p + 1}$$

Review of Hypothesis Testing

Hypothesis tests provide a rigorous statistical framework for answering ‘yes-or-no’ questions about the data.

Our setting: Is the coefficient β_j in a linear regression of Y onto $X_1, \ldots, X_p$ equal to 0?

Framework:

1. Null/alternative hypothesis.
2. Test statistic.
3. Null distribution.
4. Compute the p-value.
5. Conclusion: decide whether or not to reject the null.

Define the null and alternative hypothesis

null $\qquad\qquad\qquad$ alternative

$Y = B_0 + B_1 X_1 + \mathcal{E}.$

(two-sided) $\qquad H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0 \qquad Y = B_0 + \mathcal{E}.$

(one-sided) $\qquad H_0 : B_j = 0$ vs. $H_1 : B_j > 0$

$H_0 : B_j = 0$ vs. $H_1 : B_j < 0.$

Treatment of H_0 and H_1 is asymmetric:

- H_0 is treated as the default state. We focus on using data to reject H_0. We can think of rejecting H_0 as making a **discovery** about our data.
- Rejecting the H_0 does not imply that the alternative hypothesis is true.

Test statistic

$Y = B_0 + B_1 X_1 + \varepsilon$

We assume that H_0 is true. The test statistic is a summary of our data. It provides evidence as to whether or not the H_0 holds.

$H_0: B_1 = 0$ vs. $H_1: B_1 \neq 0$.

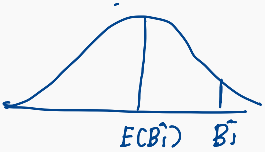
$$t_s = \frac{\hat{B}_1 - E(\hat{B}_1)}{se(\hat{B}_1)}$$

$$= \frac{\hat{B}_1 - B_1}{se(\hat{B}_1)}$$

holds b/c we assume H_0 is true

$$= \frac{\hat{B}_1 - 0}{se(\hat{B}_1)}$$

ex: $\hat{B}_1 = 5$



Null distribution

In order to decide whether or not our test statistic provides evidence in favor of H_0, we need to know the distribution of the test statistic.

Since we assume H_0 is true, we refer to this distribution as the null distribution.

$Y_i = B_0 + B_1 X_1 + \varepsilon_i$

(1) $\varepsilon_i \sim N(0, \sigma^2)$. $\Rightarrow$

(2) sample size is large enough for CLT to kick in.

OR

bootstrap (later)

$E(Y_i) = B_0 + B_1 X_1$

$Y_i \sim N(B_0 + B_1 X_1, \sigma^2)$

$\hat{B}$ are functions of Y

$\Rightarrow$ $\hat{B}$'s are also normal.

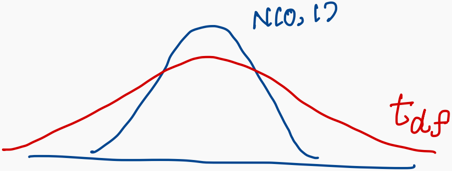
Null distribution

$\varepsilon_i \sim N(0, \sigma^2) \rightarrow Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

$\rightarrow t_s \sim N$

?

t-distribution.



σ^2 is unknown

↳ estimate σ^2 using $\hat{\sigma}^2$

↳ introduces additional uncertainty in your distribution

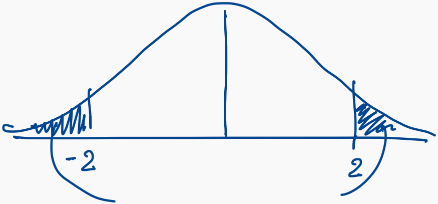
↳ to take into account this uncertainty, we use the t-distribution w/ n-p+1 df.

p-value

Given a value for our test statistic, does this provide strong evidence against H_0? $ts = 2 \longrightarrow$ prob?

p-value allows us to transform our test statistic into a probability that can answer this question.

p.value: probability of observing our test statistic or something more extreme, assuming H_0 is true.



$ts = 2$

$ts \overset{H_0}{\sim} t_{n-p+1}$

shaded area = p.value. $= \underline{0.01}$.

Conclusion

A small p-value indicates that such a large value of the test statistic is unlikely to occur under H_0, and thereby provides evidence against H_0. $\longrightarrow$ reject Ho

How small is small enough to reject H_0?

answer is in the eye of beholder.

p.value $<$ α $\Rightarrow$ reject Ho.

$\alpha = 0.05$

0.10

0.01

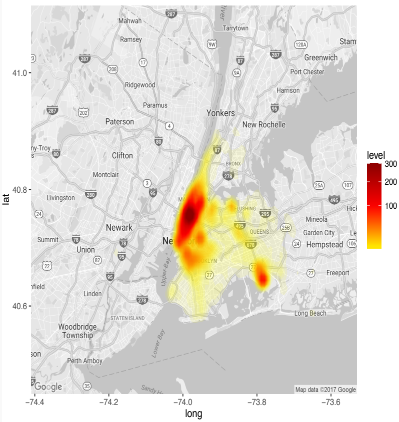
if we can reject Ho, we say our results are statistically significant.

Conclusion

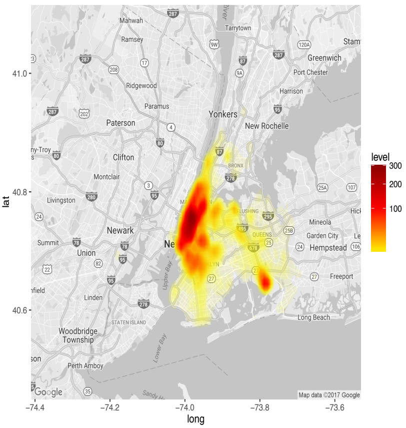
If we reject H_0, that means we have evidence that β_j is significantly different from 0, at significance level α.

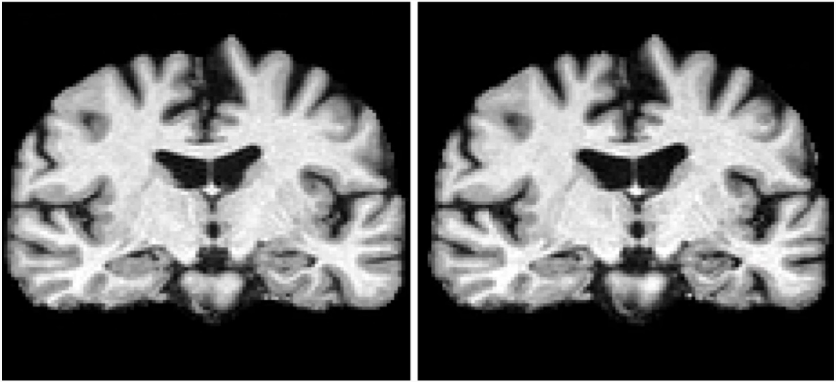
If we do not reject H_0, that means we do not have evidence that β_j is significantly different from 0, at significance level α.

Oct. 20


level
300
200
100
lat
41.0
40.8
40.6
long
-74.4
-74.2
-74.0
-73.8
-73.6
Mahwah
Tarrytown
Ramsey
Greenwich
Port Chester
Ridgewood
Harrison
Paramus
Yonkers
Paterson
New Rochelle
Clifton
BRONX
Montclair
Livingston
FLUSHING
Mineola
Newark
QUEENS
Garden City
Summit
Hempstead
Union
Freeport
Linden
Long Beach
STATEN ISLAND
Woodbridge
Township
Perth Amboy
Upper Bay
Lower Bay
Google
Map data ©2017 Google

Nov. 27


level
300
200
100
lat
41.0
40.8
40.6
long
-74.4
-74.2
-74.0
-73.8
-73.6
Mahwah
Tarrytown
Ramsey
Greenwich
Port Chester
Ridgewood
Harrison
Paramus
Yonkers
Paterson
New Rochelle
Clifton
BRONX
Montclair
Livingston
FLUSHING
Mineola
Newark
QUEENS
Garden City
Summit
Hempstead
Union
Freeport
Linden
Long Beach
STATEN ISLAND
Woodbridge
Township
Perth Amboy
Upper Bay
Lower Bay
Google
Map data ©2017 Google



Confidence Intervals

$(10, 20) \rightarrow$ reject H_0

Confidence intervals are close cousins to hypothesis tests. Duality between 2-sided hypothesis tests and confidence intervals.

$1-\alpha$ CI for B_1:

$H_0: B_1 = 0$ vs. $H_1: B_1 \neq 0$.

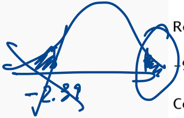
$\boxed{\text{estimate:}} \pm \boxed{\text{critical value; quantile}} \times \boxed{\text{standard error of your estimate}}$

$\Rightarrow \hat{B_1} \pm t_{n-p+1; 1-\alpha/2} \times se(\hat{B_1})$

95% CI for B_1: $(-3, 5)$.

95% confident the true B_1 falls btwn $(-3, 5)$.

$\Rightarrow$ if I repeated this many, many, many, many times, 95% of intervals would contain B_1.




```
> summary(lm(crim~.,data=Boston))

Call:
lm(formula = crim ~ ., data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-9.924 -2.120 -0.353  1.019 75.051

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.033228   7.234903   2.354 0.018949 *
zn            0.044855   0.018734   2.394 0.017025 *
indus        -0.063855   0.083407  -0.766 0.444294
chas         -0.749134   1.180147  -0.635 0.525867
nox         -10.313535   5.275536  -1.955 0.051152 .
rm            0.430131   0.612830   0.702 0.483089
age           0.001452   0.017925   0.081 0.935488
dis          -0.987176   0.281817  -3.503 0.000502 ***
rad           0.588209   0.088049   6.680 6.46e-11 ***
tax          -0.003780   0.005156  -0.733 0.463793
ptratio      -0.271081   0.186450  -1.454 0.146611
black        -0.007538   0.003673  -2.052 0.040702 *
lstat         0.126211   0.075725   1.667 0.096208 .
medv         -0.198887   0.060516  -3.287 0.001087 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.439 on 492 degrees of freedom
Multiple R-squared:  0.454,	Adjusted R-squared:  0.4396
F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

X_1: zn

X_2: indus

$\alpha = 0.05$

$H_0: B_1 = 0$ vs. $H_1: B_1 \neq 0$.

$H_0: B_2 = 0$ vs. $H_1: B_2 \neq 0$

$ts = 2.394$

Null distr: t_{492}

p.value $= \frac{0.017025}{2}$

conclusion: we reject Ho at sig. level $\alpha = 0.05$

See R script `MLR_Inference.R`