DS 3010

Spring 2026

Wenting Xu

IOWA STATE UNIVERSITY

Module 1: Multiple Linear Regression

Part 3: MLR Inference

Wenting Xu

Iowa State University

Copyright & Attribution

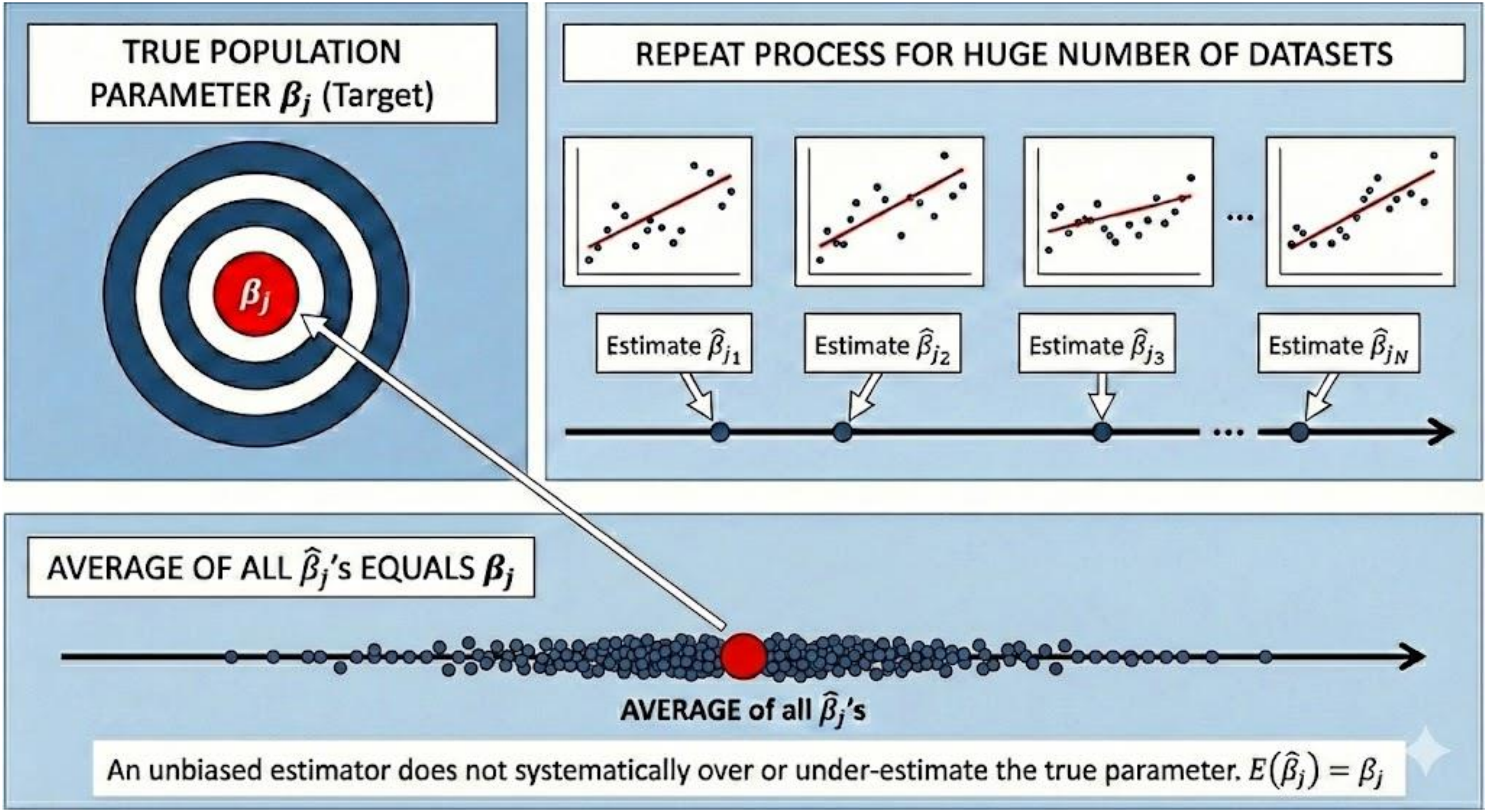
Some lecture slides and instructional materials in this course are adapted from the following sources:

•*An Introduction to Statistical Learning: With Applications in R (Second Edition)*
Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
Springer, 2021

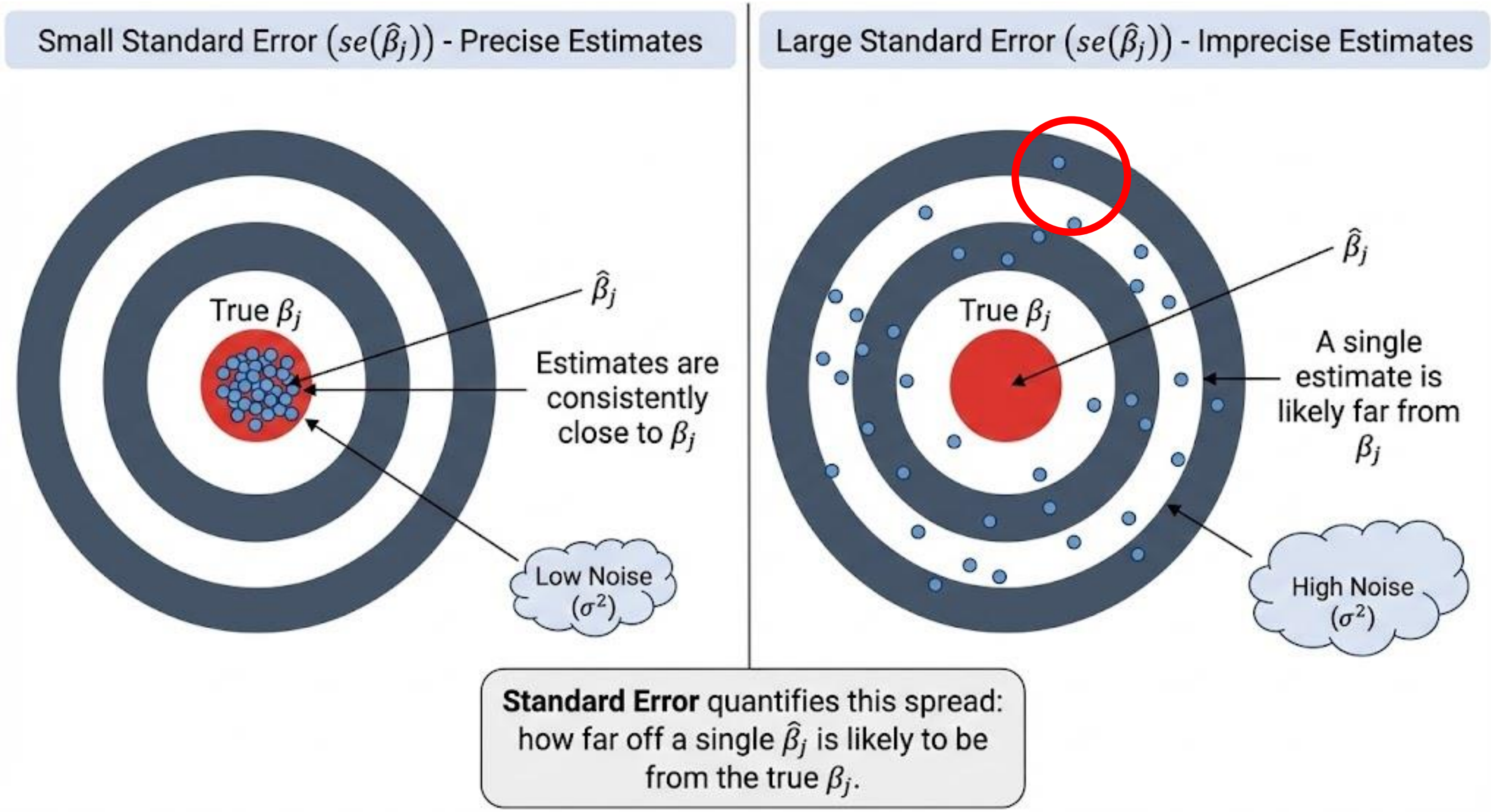
•Online course materials developed by Trevor Hastie, Robert Tibshirani, and collaborators.

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

Unbiasedness: $E(\hat{\beta}_j) = \beta_j$. Least square estimates are unbiased estimate of the true population parameters $\beta_0, \beta_1, \dots, \beta_p$.



Standard error: Quantifies the how far off a single estimate of $\hat{\beta}_j$ will be from β_j. We denote this as $\mathrm{se}(\hat{\beta}_j)$. These depend on estimate of σ^2, which represents the variance of ϵ (and therefore Y).



Estimate of σ

For simple linear regression, the standard error for $\hat{\beta}'$'s have simple forms:

$$SE(\hat{\beta}_0)^2 = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right], SE(\hat{\beta}_1)^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

For multiple linear regression, these formulas require linear algebra notation and formulas will not be shown here.
Our estimate $\hat{\sigma}^2$ is:

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

R gives you all of these estimates pain-free. → m1 ← lm(Y ~ X)
summary(m1)

Estimate of σ

We'll utilize these properties of our least square estimates to carry out inference on our model.

Tools we will develop to carry out inference:

- **Hypothesis testing**
- **Confidence intervals**
- **Prediction intervals**

Review of Hypothesis Testing

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

slope

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

not related

<u>Framework:</u>

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

$$H_0: \beta_j = 0 \text{ versus } H_1: \beta_j \neq 0$$

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 H_1 is true.

Test statistic

Rejecting the H_0 does not imply that the alternative hypothesis is true.

H_0 : The suspect is innocent.
H_1 : The suspect is guilty.

If strong evidence suggests the suspect was at the crime scene, we might reject H_0 (innocence).

However, this does not 100% prove guilt-there could be another explanation (e.g., wrong place, wrong time).

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.

Test statistic

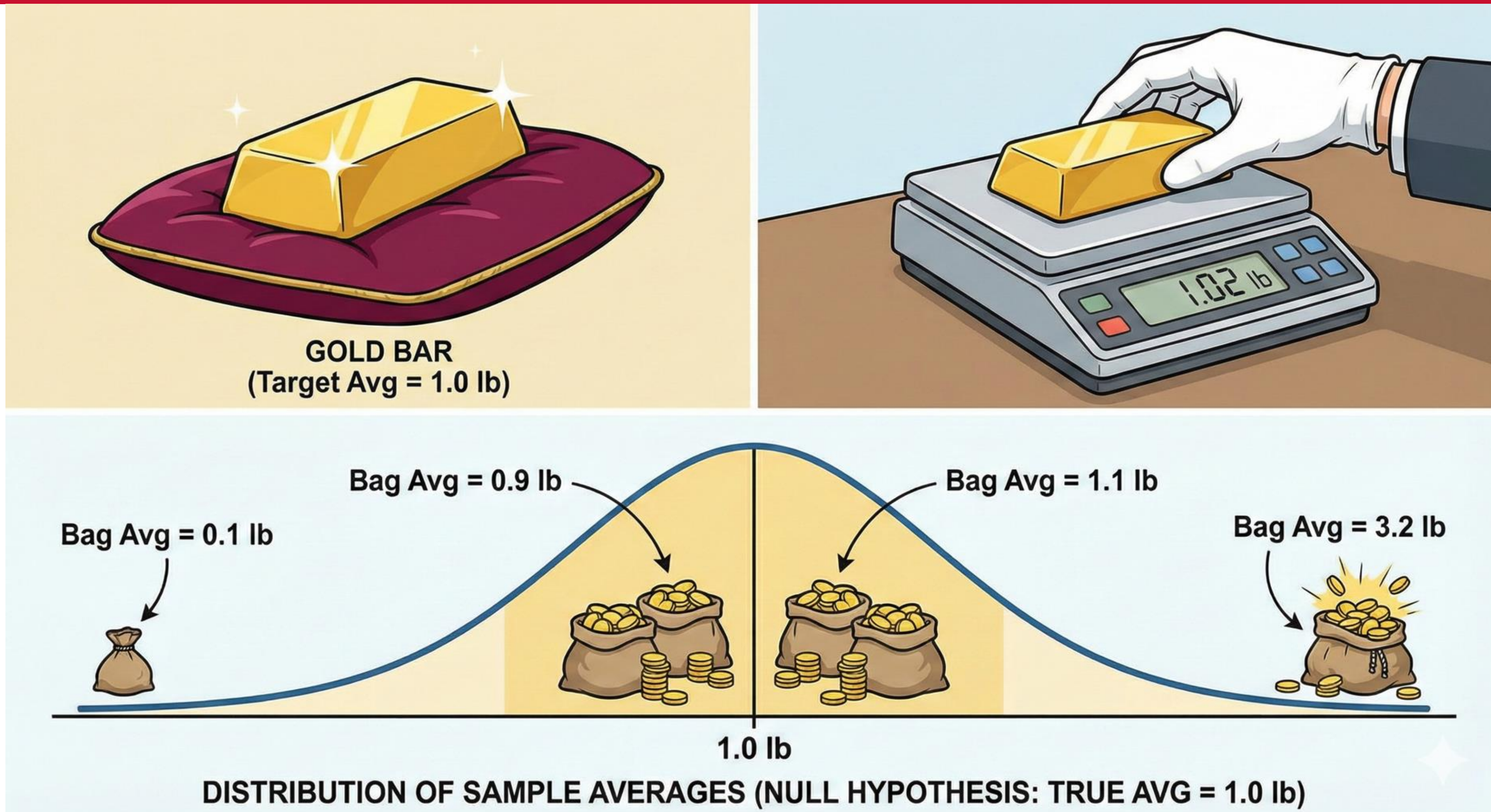
Can we just use $\hat{\beta}_j$ as our test statistic in this setting?

```
> 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
```

Null distribution

- In a criminal trial, the suspect is **presumed innocent** until proven guilty.
- In statistics, we presume the variable is **"innocent"** (has zero effect, $H_0: \beta_j = 0$)
- Since we assume H_0 is true, we refer to this distribution as the 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.


GOLD BAR
(Target Avg = 1.0 lb)
1.02 lb
Bag Avg = 0.1 lb
Bag Avg = 0.9 lb
Bag Avg = 1.1 lb
Bag Avg = 3.2 lb
1.0 lb
DISTRIBUTION OF SAMPLE AVERAGES (NULL HYPOTHESIS: TRUE AVG = 1.0 lb)

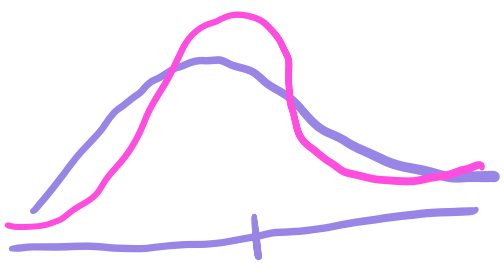
Null distribution

$N \sim (\mu, \sigma^2)$

t-distribution vs. normal distribution.

σ^2 is unknown

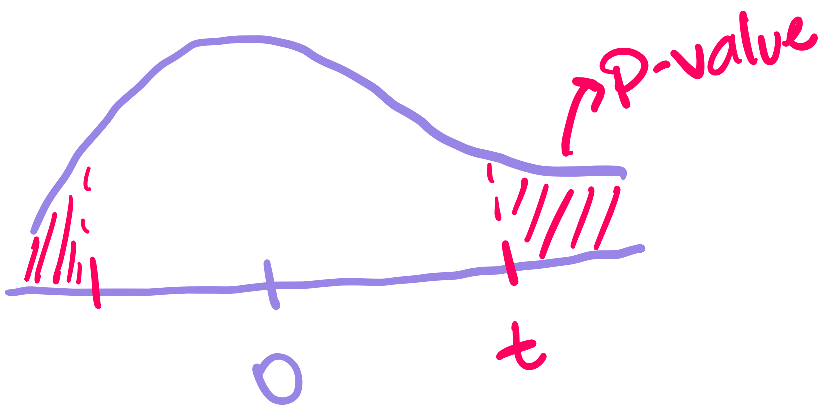
$$t = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)}$$



p-value

Given a value for our test statistic, does this provide strong evidence against H_0 ?

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



the probabilty of observing the test statistic or (more extreme) if the null hypothesis (H_0) was true

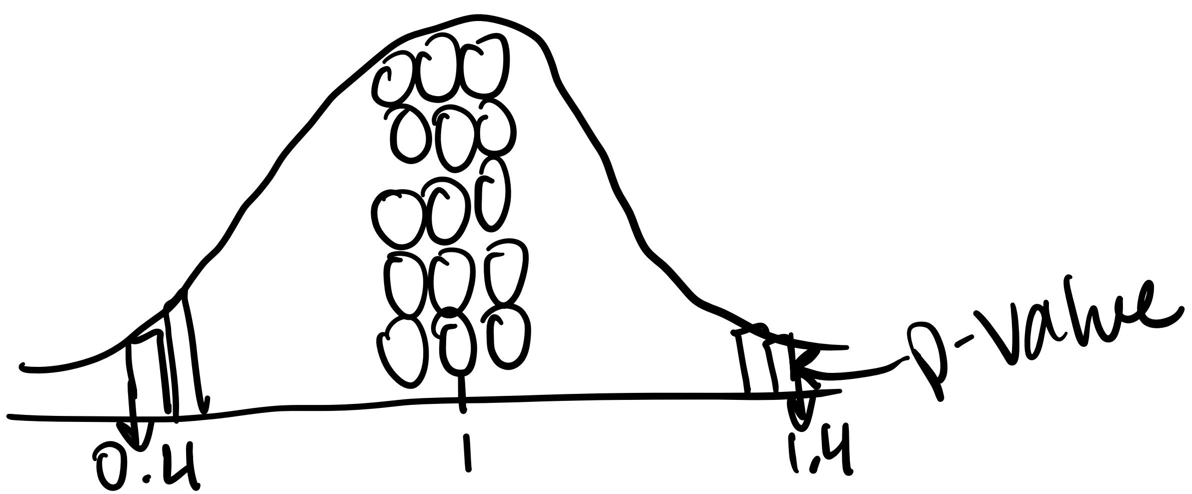
Conclusion

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.

How small is small enough to reject H_0 ?

significance level = α (e.g. 0.01, 0.05, 0.1)

if p-value < α → reject H_0



Conclusion

If we **reject $\boldsymbol{H_0}$**, that means we have evidence that β_j is significantly different from 0 , at significance level α.

If we **do not reject $\boldsymbol{H_0}$**, that means we do not have evidence that β_j is significantly different from 0 , at significance level α.

R output

```
> 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
```

Predictions

Once we have fit the multiple regression model, it is straightforward to apply the model in order to predict Y for given values of $X_1, \ldots, X_p$.

How accurate are our predictions?

- We can use $\hat{Y}$ as our predicted value.
- But this is just a single value based on our dataset. If we had a slightly different set of data, then our $\hat{Y}$ would also be different.
- So there is uncertainty around our predicted value. How do we quantify that uncertainty?

There are 2 sources of uncertainty associated with a prediction.

1. Reducible error (estimation error and model bias)
2. Irreducible error

2 sources of uncertainty

Reducible error

- Coefficient estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ are estimates for the true population parameters $\beta_0, \beta_1, \dots, \beta_p$.
- Inaccuracy in the coefficient estimates is related to the reducible error.
- Assuming a linear model for $f(x)$ is almost always an approximately of reality.
- This is an additional source of potentially reducible error which we call model bias.

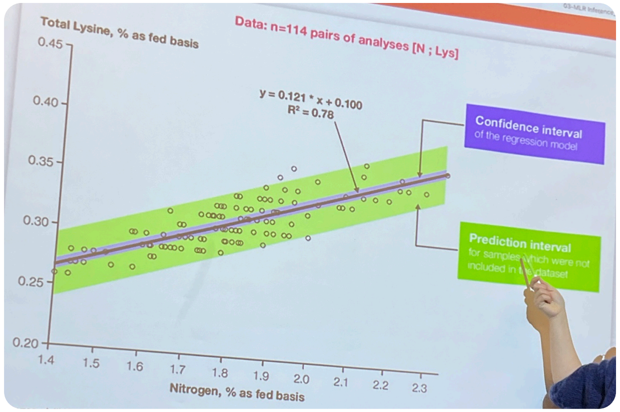
2 sources of uncertainty

Irreducible error

- Even if we knew the true values of $\beta_0, \beta_1, \ldots, \beta_p$, the response value Y cannot be predicted perfectly because of the random error ϵ in the model.
- This is what we refer to as the irreducible error.
- Prediction intervals are always wider than confidence intervals.

CI versus PI

- We can compute a confidence interval in order to quantify our uncertainty around estimating $f(X)$.
  - Only takes into account reducible error (estimation error & model bias).



confidence interval

Prediction Interval: for samples not included in the dataset

- We can compute a prediction interval in order to quantify our uncertainty around predicting Y.
  - Takes into account irreducible error **and** reducible error.

CI versus PI

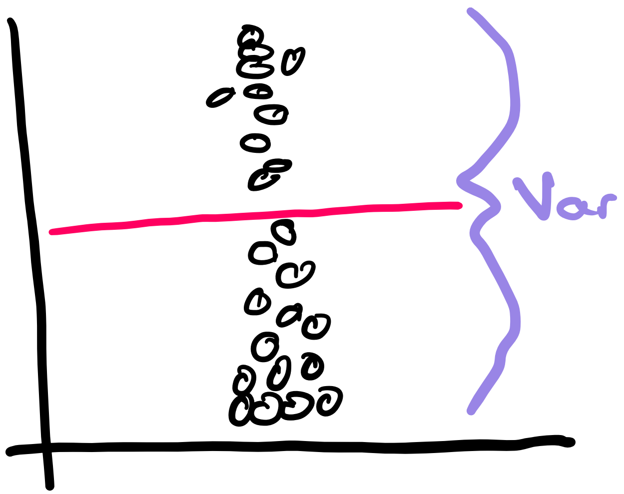
Example: patient data

Y = patient satisfaction score

X_1 = age

X_2 = severity

X_3 = anxiety



We use a confidence interval to quantify the uncertainty surrounding the average patient score given a set of predictors.

We use a prediction interval to quantify the uncertainty surrounding the satisfaction score for a particular patient given a set of predictors.

See R script `MLR Inference.R`