

Chapter 12 Exam Prep: Inference on Linear Regression

Statypus Insight: The “Critical Commander” Approach

In modern data science, running a regression model takes exactly one line of code. The true test of a statistician is not whether they can build the engine, but whether they can critically evaluate the architecture that comes out of it.

Raw Exam Question & Output:

A university admissions office uses the Statypus `RegressionInference()` function to predict a student’s First-Year College GPA (y) based on their High School GPA (x). They test the model for an incoming student with a 3.5 High School GPA ($x_0 = 3.5$).

```

$'Confidence Level'
[1] 0.95

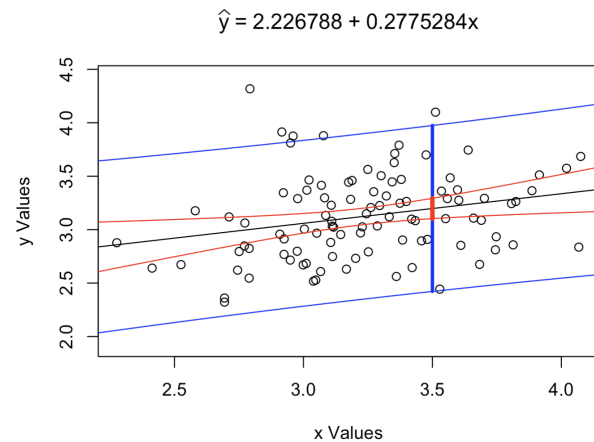
$'Conf. Int. for yHat'
[1] 3.102919 3.293355

$'Pred. Int. for y'
[1] 2.421832 3.974442

$'Tested Point'
[1] 3.5

$'yHat at x0'
[1] 3.198137

```



Exam Question 1:

An admissions counselor looks at this output and says, “*We have an incoming student named Sarah with a 3.5 High School GPA. We can be 95% confident that Sarah’s first-year college GPA will be between 3.10 and 3.29.*” Explain the specific statistical error the counselor is making and then give the correct interpretation of what you can predict for Sarah.

Sally the Statypus says: Don’t get trapped by the lines! Remember that the red lines (CI) are for the **average** behavior of a group, while the blue lines (PI) are for the behavior of **individuals**. Sarah is an individual!

Bill the Statypus says: Variation Alert! It is much easier to predict where a group average will fall than it is to predict where a single data point will land. Think about the standard errors we discussed: individuals have much more “noise.”

Your Turn (Answers for Part 1):**Exam Question 2: CI vs. PI**

For each scenario, state whether the researcher should look at the Confidence Interval (the red lines or CI) or the Prediction Interval (the blue lines or PI), and briefly justify your choice.

Scenario A: A landscaper orders a bulk shipment of 1000 ferns which are 12 weeks old. To ensure that his personal pickup truck is not overloaded, the manager needs to estimate the total weight of the shipment. They use a model where Age (x) predicts the total Weight (y) of ferns.

Scenario B: Once the shipment of the 12 week old ferns arrives, they receive a call asking for a range of heights they have in the order. Before they are able to unload the shipment, they use a model where Age (x) predicts the Height (y) of ferns.

Bill the Statypus says: I know this can be a bit tough to get straight, so try this. CI has a C and is for the **C**ommunity or average whereas PI has a P and is for the **P**erson or each individual.

Your Turn (Record your answers for Question 2 here):