Nama : Muhammad Hisyam Nim : A11.2018.11106

1. **Business Understanding**

Apa hubungan antara hit (H) dan run batted in (RBI) pada pemain bisbol liga utama?

- Hipotesis: Jumlah H yang lebih tinggi akan menghasilkan RBI yang lebih besar dalam satu musim.

- Potensi keterbatasan: Variabel lain dapat mempengaruhi hubungan antara H dan RBI, memerlukan data tambahan untuk analisis masa depan. Misalnya, di mana dalam urutan pukulan pemukul memukul, jumlah peluang yang harus dipukul oleh pemukul dengan pelari dalam posisi mencetak gol, jenis lemparan yang dihadapi pemukul musim itu, dll.

1. **Data Understanding**

1) Jenis data apa yang dibutuhkan

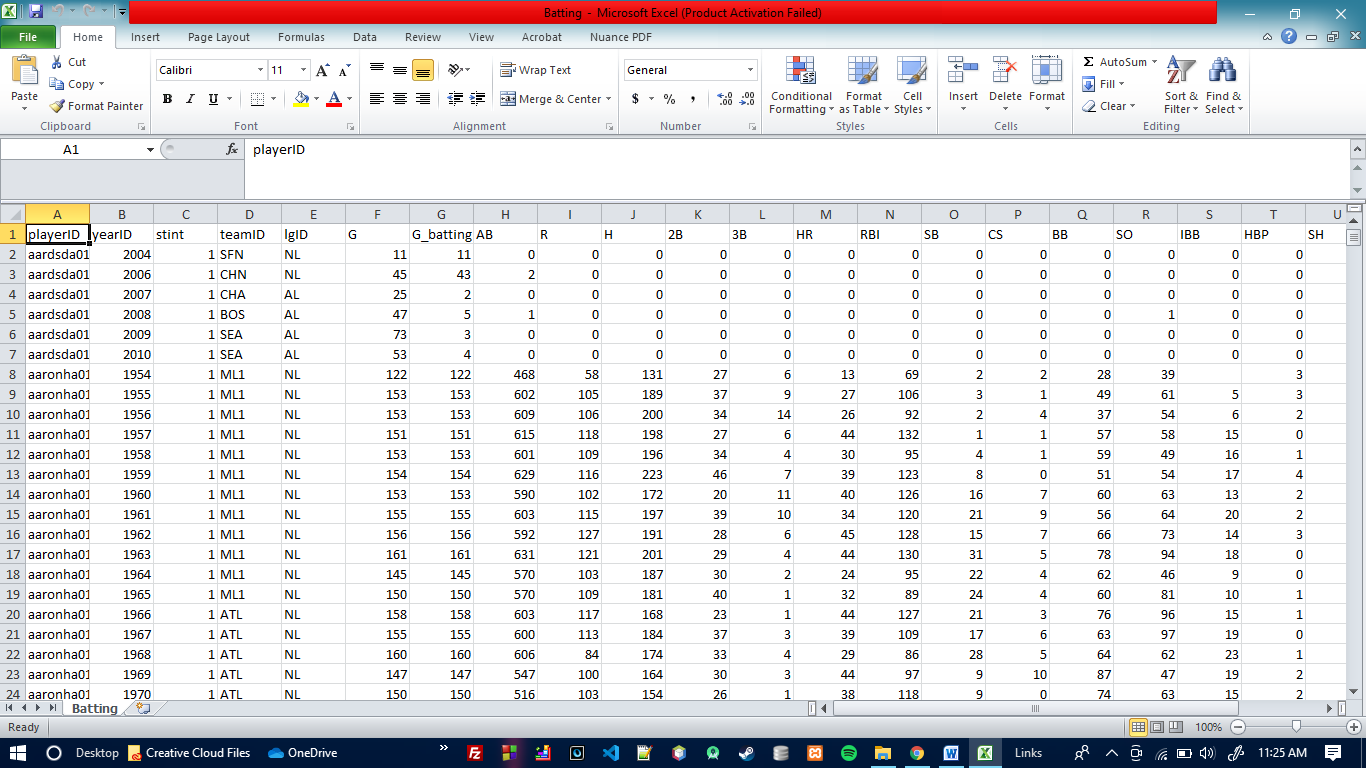
- Sumber data: Tabel Batting Database Lahman

- Data tersedia dalam paket 'Lahman' di dalam R

- Masalah data: Data batting disediakan dalam tabel dari tahun 1871 hingga 2011. Musim yang lebih lama mungkin kehilangan data.

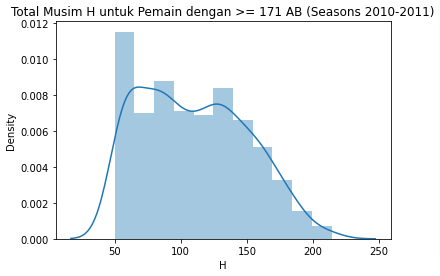
2) Pengumpulan/Pengukuran

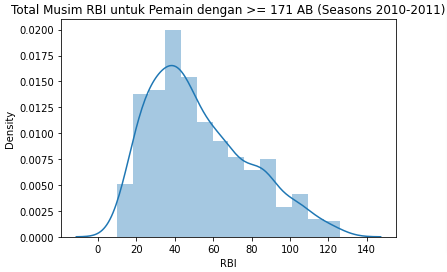
- Tidak ada data tambahan yang perlu dikumpulkan saat ini



1. **Data Preparation**

- Visual H dan RBI

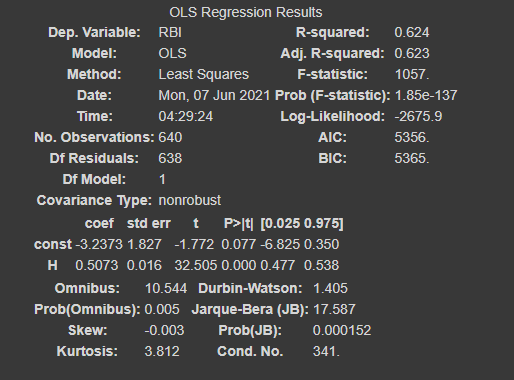




1. **Modeling**

Membangun model regresi sederhana

Identifikasi variabel dependen dan independen



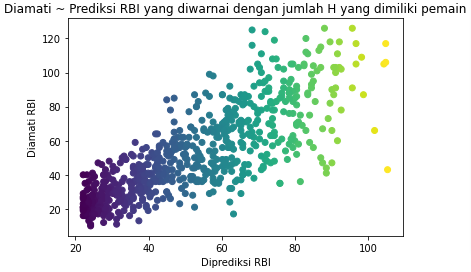
> Model menjelaskan sekitar 62% dari variabilitas yang diamati dalam RBI.

> Koefisien untuk H menunjukkan bahwa untuk peningkatan satu unit H, secara rata-rata kita akan

melihat peningkatan RBI sekitar 0,50. Jadi, 10 AB(At Bat), Kemunculan di plate ekstra dapat menyebabkan sekitar 5 RBI ekstra (10 \* 0,50).

1. **Evaluation**

Plot model diagnostic



> Sedikit menyebar dalam pola residual terhadap plot nilai prediksi (plot pertama) yang menunjukkan bahwa varian tidak konsisten di seluruh rentang nilai.

> Residu mengatur kira-kira terdistribusi normal sekitar 0.

> Plot terakhir menunjukkan RBI yang diprediksi vs yang diamati. lebih banyak variabilitas

dalam prediksi yang semakin ke kanan melangkah (karena titik-titik menjadi berwarna lebih terang,

menunjukkan bahwa pemain memiliki lebih banyak Hit tahun itu). Di sinilah kesalahan model lebih besar dalam prediksi.

Menghitung RMSE mengetahui nilai rata-rata kesalahan pada model dalam meprediksi nilai

Hasil RMSE : 15.83

1. **Deployment**

1) Pertanyaan penelitian adalah untuk menguji hubungan antara H dan RBI pada pemain bisbol liga utama. Tampaknya ada hubungan antara kedua variabel ini dengan lebih banyak hit menghasilkan jumlah Run yang lebih tinggi. Namun, dengan peningkatan hit, prediksi model menjadi lebih noisy.

2) Prediksi yang lebih baik dapat dibuat dengan lebih banyak data yang dapat memberikan konteks seputar hit (situasi, pitching, urutan line up, dll) dan harus dikumpulkan untuk analisis di masa mendatang.