<https://github.com/boboppie/coursera-course-statistics_one>

**LECTURE 7– Intro to Regression**

**Segment 1: Intro to Regression**

Statistical analysis used to predict scores on an outcome variable(Y), based on scores on one or more predictor variables (X).

Regression equation **Y = m + bX + e 🡪 Y = a + BX +e 🡪**

**Y =B0 + B1X1 +e 🡪 ^Y = B0 + B1X1**

**Y -^Y = e (e is prediction error or residual)**

Regression model: used to model or predict the behaviour

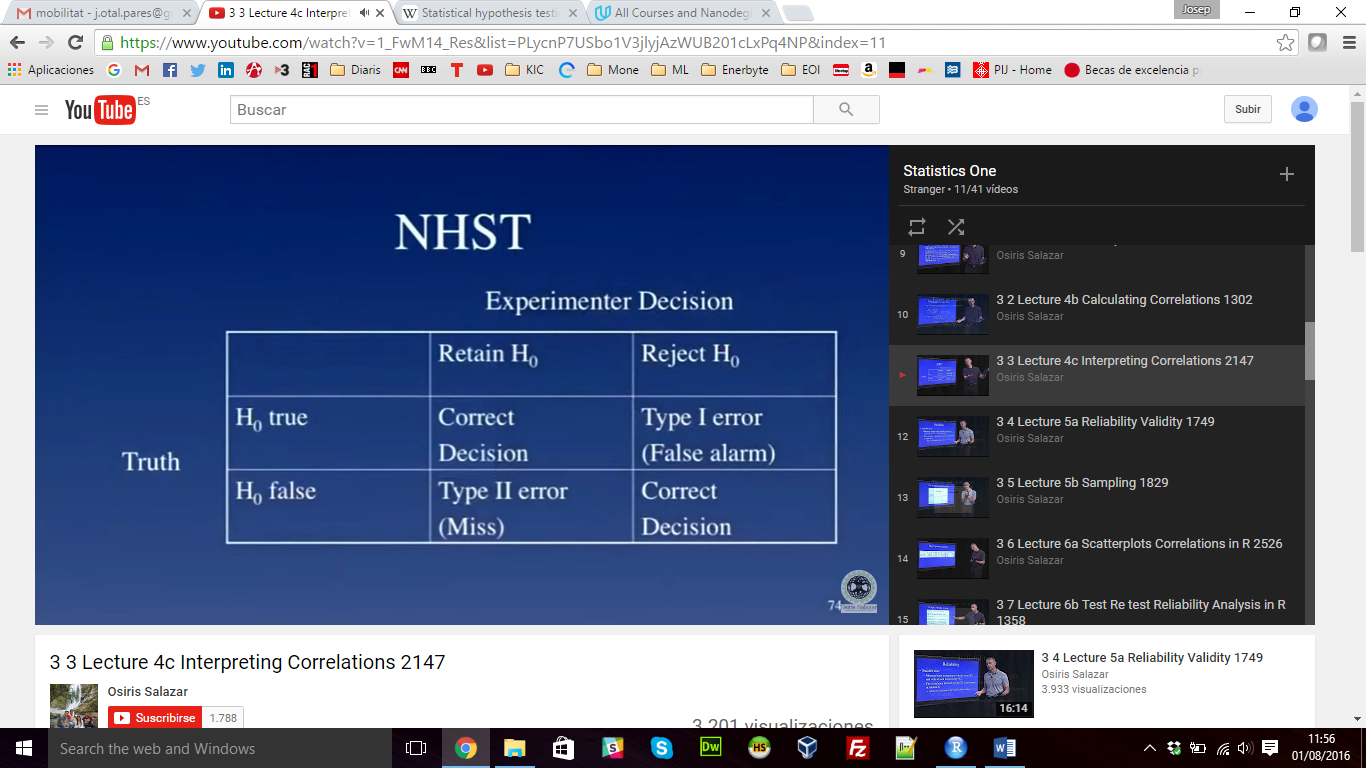
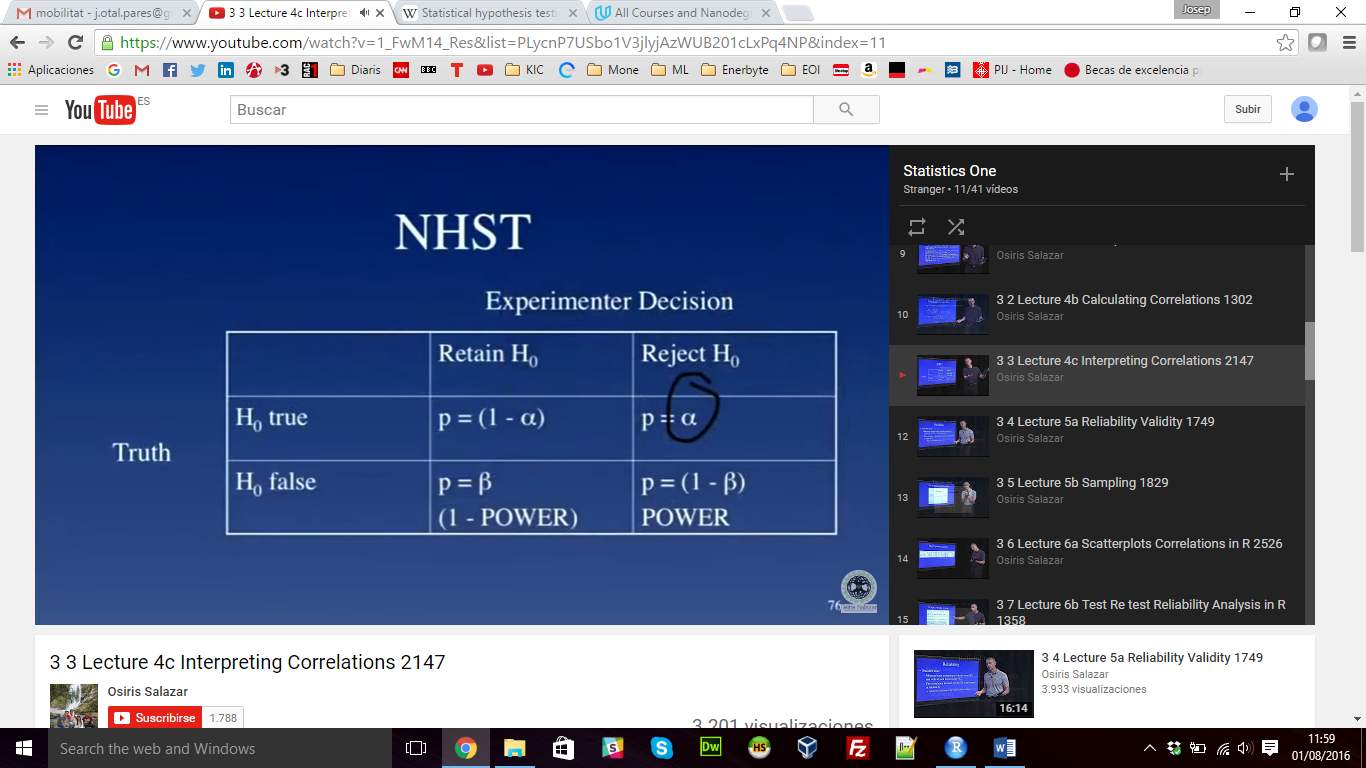
* Produce better models, more accurate

Estimation of coefficients:

* Minimize the residuals, so Sum of Squared Residuals is minimized
* SS.Residual = E(^Y-Y)2
* Unstandarized 🡪 B = r \*(SDy/SDx)
* Standarized regression coefficient 🡪 β =r
  + As SDy =SDx= 1

**Segment 2: NHST a closer look**

* Null Hypothesis Significance Testing (NHST)
  + H0 = Null hypothesis. E.g. no correlation r = 0
  + HA = alternative hypothesis e.g. r > 0
* **p = P(D|H0)** is the probability of getting the D value given that the Null hypothesis (H0) is true
* **α** is normally set to 0.05. If p<0.05, reject the null and have statistically significant results
* **t = B/SE 🡪** t-statistics, B(unstandardized regression coefficient),
  + SE standard error √SS.Residual / (N-2)

* Problems:
  + Biased by N
  + Binary outcome
  + Null model is a weak hypothesis
* Alternatives to NHST:
  + Effect size
  + Confidence intervals
  + Model comparison