**Unit 8 Notes**

* Make sure you understand the ecdf function
* Pearson\_r for correlation coefficient
* a, b = Np.polyfit(x, y, 1) for linear regression, 1 for exponent, slope, intercept
* Graphical EDA **PRIOR** to regression
* Bootstrapping – a resampled array of data with replacement and then compute a summary statistic again
  + Can plot an ECDF

Sampling Distribution Notes

*Central Limit Theorem*

* As long as a distribution has a well-defined mean and standard deviation, we can take sample means which will produce an approximately normal distribution. It doesn’t take that larges of samples (n = 10+) to get very close to a normal distribution.
* As n approaches infinity, the resulting distribution approaches
* Positive kurtosis has a pointier center, negative has a rounder less distributed shape (half circle)

Significances Tests Notes

Your p-value is the likelihood that the outcome is true given the null. The probably of a type I error is your significance level. Type II error is failing to reject a false null.

What is Power? P (rejecting null | null is false)

= 1 – Type II error

You can increase the power by increasing your alpha (significance level), but type I error will increase

A/B Testing Notes

Shopify Article

* The longer an A/B out there, the longer the results can get polluted
  + They recommend a maximum of four weeks to conduct a test, and if you can’t get the necessary traffic in 2 – 4 weeks, it probably isn’t worth your time.
* [Craig Sullivan Hypothesis Kit](https://medium.com/@optimiseordie/hypothesis-kit-2-eff0446e09fc)