# I have continuous data broken into groups, how do I know what test to run?

* Are my groups independent or are they related (i.e. same people at different times)
* Is my data normal?
  + Could use, skewness + kurtosis, QQ plot, histogram, Shapiro test (be wary of this test with larger samples it is *very* sensitive)
  + Checks by test:
    - Independent t-test: check the normality of each sample’s data (i.e. check x1 and x2)
    - ANOVA: check the normality of each sample’s data (i.e. check x1, x2, …, xn)
    - Dependent t-test: check the normality of the differences between paired samples (i.e. check x1 - x2)
* How many groups of data am I comparing?
  + If 2 groups
    - Normal: t-test (independent and dependent)
    - Non-normal: Mann-Whitney U (independent) & Wilcoxon (dependent)
  + If 3 or more groups
    - Normal: ANOVA (independent) & Repeated measures ANOVA (dependent)
      * Follow-up with tukeyHSD
    - Non-normal: Kruskal-Wallis (independent) & Wilcoxon (dependent)
      * Follow-up with repeated testing of related 2 sample test
* How can I show the size of the difference between the groups
  + t-test: Cohen’s D (point estimate); confidence interval
  + 1-way ANOVA (independent): tukeyHSD
  + [Non-parametric](https://github.com/Thinkful-DSI-Cohort-6/class_notebooks/blob/master/week3/non_param_effect_size.py)
    - Common language effect size
    - Rank-biserial correlation coefficient

# I’ve performed a test, how do I present it

* Report the result (no diff or sig diff)
  + Interpret in light of the problem at hand
    - Don’t just say “reject the null the means are different”; instead say something like “we see evidence that having an AC unit significantly affects sale price”
* If significant difference, answer the question: “how big is the difference?”
  + Use appropriate effect size/confidence interval calcs
  + Interpret in light of the problem at hand
    - Don’t just say “a 95% CI is [1000, 5000]”; instead say something like “Having an AC unit can increase sale price anywhere from [$1000, $5000], as shown by a 95% CI.”