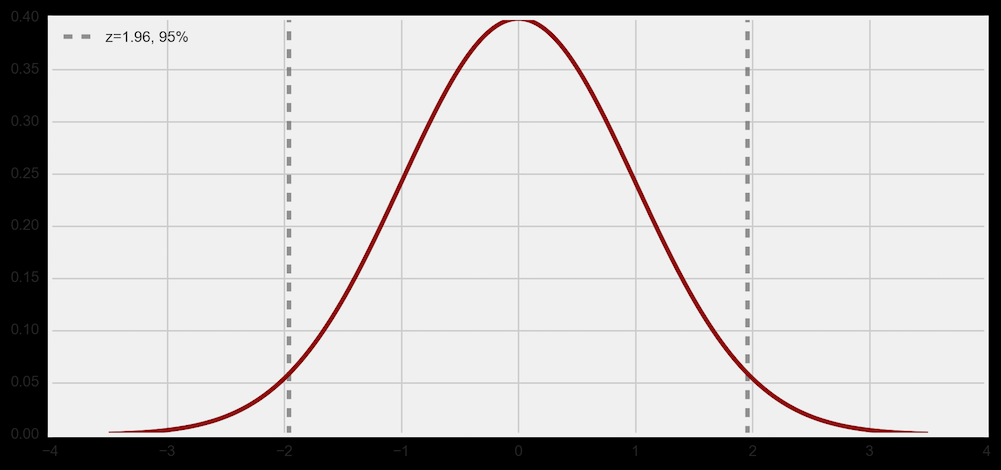
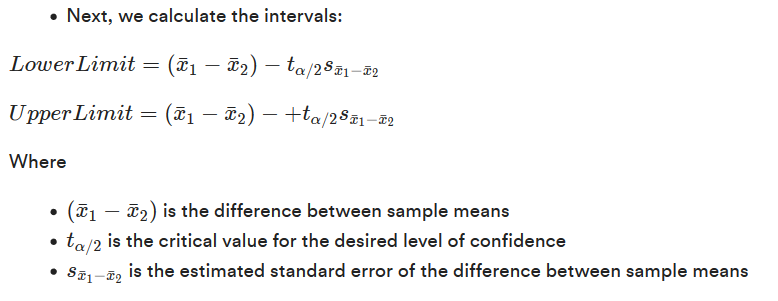
Unit 3-4 Confidence Intervals

* Drug Efficacy Example
  + We're interested in testing the efficacy of a new drug. The drug is supposed to lower blood pressure, but we need to test it.
  + We randomly select 50 people for the placebo control condition and 50 people to receive the treatment.
  + We measured a difference in blood pressure of -9.82 between the experimental and control groups.
* Uncertainty About the True Value
  + Working within the frequentist framework, we know that there's a true value for the difference between population means. If we had measured every single person in the population, we'd have this value. However, we're only able to measure a sample of the population.
  + The difference between the sample means is our point estimate of the difference between the population means. Depending on our sample size, we'll have some degree of uncertainty surrounding how far off the statistic of our random sample is from the true parameter. This distance is our measurement error.
  + As frequentists, we think of this sample as one of many hypothetical samples from the same overall population. It's important for us to frame our definitions in the context of repeated measurements.
* Confidence Intervals Defined
  + A **confidence interval** describes a set of possible values for a parameter based on a statistic. The confidence interval will be centered at the point estimate and include +/- a few standard errors.
  + **Standard error**: If you had a new random sample that was the same size as the original, the standard error would quantify our certainty in how far the new differences in sample means would be from the original difference in sample means.
  + We use confidence intervals to express the degree of uncertainty associated with a sample statistic. A confidence interval is an interval estimate combined with a probability statement.
  + Confidence intervals are preferred to point estimates and interval estimates, because only confidence intervals indicate both the precision and uncertainty of the estimate.
  + The general formula for a confidence interval is point estimate +/-(t or z)standard error.
  + For example, here is a 95-percent confidence interval on a z-distribution. The point estimate is centered on zero and the interval is indicated with the dashed lines:
  + 
* Confidence Intervals for T-Tests
  + Let's calculate the 95-percent confidence interval for our drug efficacy trial. We found that the difference in means between the treatment and experimental groups is -9.82. This is our point estimate. We'll use a confidence interval to investigate the accuracy of our estimate.
  + We'll make a few assumptions in order to do so:
    - The two populations have the same variance (homogeneity of variance).
    - The populations are normally distributed.
    - Each value is sampled independently from the other values.
  + These are reasonable assumptions for our example, but small violations of the first two don't make a huge difference.
  + 
  + We'll explain what each of these terms means as we calculate the confidence interval. Just remember that:
    - Lower limit = The difference between sample means - (t-statistic)(standard error)
    - Upper limit = The difference between sample means + (t-statistic)(standard error)
  + First, we'll estimate the standard error of difference between means. The formula for the difference in a population's sample means is: sx¯1−x¯2s\_{\bar{x}\_1 - \bar{x}\_2}sx¯1​−x¯2​​.
    - We can calculate this as:
      * MSE = (np.std(experimental)\*2 +np.std(control)\*2) / 2
      * print(MSE)
      * 51.1430935791
  + The number of subjects in each treatment (n) is 50, so we use the following:
    - # Standard error for difference in means is the square root of the MSE,
    - # divided by the number of subjects in each treatment.
    - se = np.sqrt(2\*MSE/50).
    - print(se)
    - 1.43028799308
  + Therefore, our standard error for the difference in means is 1.43.
  + Next, we'll find the t to use for the confidence interval. To calculate our t, we need to know the degrees of freedom. The degrees of freedom are the number of independent estimates of variance on which the MSA is based.
  + This is equal to *n*1 + *n*2 -2, where *n*1​ is the sample size of the first group and *n*2​ is the sample size of the second group.
  + So, for our study, we can solve this by performing the following:
    - # Calculate degrees of freedom by multiplying the sample sizes of
    - # each group and subtracting 2.
    - len(experimental)+ len(control)-2
    - 98
  + Therefore, our degrees of freedom are 98! Next, we can then use a t-table in Python to find the t in order to calculate a 95-percent confidence interval for 98 degrees of freedom.
  + Recall that we are looking for the critical value at tα/2 ​, so we pass 0.975 to the Python function to find the critical value.
  + stats.t.ppf(.975, 98)
  + 1.984467454426692
  + This means our critical value for t (with 98 degrees of freedom at a 95-percent confidence interval) is 1.98.
  + Therefore, the 95-percent confidence interval is:
    - Lower Limit=−9.82−(1.43)(1.98)=−12.65
    - Upper Limit=−9.82+(1.43)(1.98)=−6.99
  + We can write the confidence interval as: −12.65≤μe−μc≤−6.99
  + In other words, μe ​ is the population mean for the experimental group and μc ​ is the population mean for the control group.
  + This analysis provides evidence that the mean of systolic blood pressure for the experimental group is lower than the mean for the control, and that the difference between means in the population is likely to be between -12.65 and -6.99.
* Interpreting Confidence Intervals
  + The 95-percent confidence interval for the difference in means for our drug efficacy test is -12.65 to -6.99. So, how do we interpret these values?
  + Incorrect interpretation: There's a 95-percent probability that the true difference of the blood pressures between the treatment and control groups is between -12.65 and -6.99.
  + Correct interpretation: If we pulled 100 samples of the same size and constructed confidence intervals in the same manner, we expect that 95 of the intervals would contain the true difference in mean blood pressures between the groups.
  + The first interpretation is wrong because it assigns a probability to the true value. In frequentist statistics, the true value is fixed and the data are random. Confidence intervals make a statement of probability about the confidence interval range that could contain the true value.
  + We can also reframe this in terms of the random sampling procedure.
  + We are 95 percent confident that the difference in means is between -12.65 and -6.99.
  + While what's above is the common shorthand, we can also think about a supplementary statement:
    - We are also 5 percent confident that the difference in means does not fall in between -12.65 and -6.99.
* Interval Size
  + So far, we've only been dealing with the 95-percent confidence interval. We interpret this as being 95 percent confident that the population parameter of interest lies within this interval.
  + So, why don’t we always use a 99-percent confidence interval? If we want the confidence level to be as high as possible, why don't we just always use a higher interval?
  + As the confidence level increases, the margin of error increases as well. That means the interval itself is wider, but it may be so wide that it's useless.
  + For this reason, 95-percent confidence intervals are the most common.
  + In general, confidence intervals should be used in a way that leaves you comfortable with the uncertainty but should not be so strict as to lower the power of your study into irrelevance. For example, in a clinical trial for lipstick, you'd want to be very confident that your treatment isn't going to poison anyone — say, 99.9 percent. However, you'd be fine with a lower confidence interval when it comes to the lipstick staying on all day long.