Recitation assignment due in recitation 3 - 23 - 10.
Coordinates with lectures 3 - 17 - 10 and 3 - 22 - 10.

1. An equal-probability with-replacement sample of n_1 = 200 trees is selected from a large tract of ~ 10,000 trees. Independently of this a second sample of n_2 = 400 trees is selected from a large tract of ~ 80,000 trees. We will observe the first sample trees, scoring each tree with x_1= number of growth rings (it is done using a core sampling device and does not seriously harm the tree). The second sample is likewise measured. The following sample data is found:

sample mean 1 = 87.44	**sample SD 1 = 24.77**
sample mean 2 = 99.05	**sample SD 2 = 31.21**
sample size 200	**sample size 400**
population size ~ 10,000	**population size ~ 80,000**

a. Determine a 95% z-based CI for the mean of the first population.

b. We actually have made sure to sample 200 without replacement. Modify your CI (a) accordingly.

c. It is of interest to know if there is any difference between the two population means. Determine a 95% z-based CI for the difference $\mu_1 - \mu_2$ without employing FPCs.

d. Re-do (c) using $FPC1^2$ and $FPC2^2$ under the square root for each of the two terms. Has it made much of a difference in the CI?

e. If the CI covers 0 (the value of $\mu_1 - \mu_2$ when they are the same) there is no reason to believe that the two populations have different means. On the other hand, if the CI lies entirely to one side of 0 there is some evidence that perhaps the population means differ. Which is the case or CI (d)?

f. If the two population means are actually the same what is the probability that a 95% CI for their difference **will cover 0** and what is the probability it would **fail to cover 0**?

g. **Suppose that it is known that each of populations 1 and 2 is normal distributed**. Since ring counts are integers this cannot possibly be true but may be approximately true. Consult pg. 619, particularly the footnote where you will find a formula applicable to the t-based CI for $\mu_1 - \mu_2$ when sample sizes are small. Suppose the same statistical data above except suppose that n1 = 2 and n2 = 4. Modify the CI (c) incorporating this different set of assumptions. (Note: FPC is not likely to be applicable in a setting in which the normal distribution is deemed appropriate because the normal is after all continuous, i.e. population essentially infinite). The "effective" DF for t is calculated for you below (always round **down** to be conservative, i.e. larger t, in this case for DF = 2). I would not ask you to calculate this "effective" DF on an exam.

$$\text{DF}[\text{n1\_},\ \text{n2\_},\ \text{s1\_},\ \text{s2\_}] := \frac{\left(\frac{\text{s1}^2}{\text{n1}} + \frac{\text{s2}^2}{\text{n2}}\right)^2}{\frac{\left(\frac{\text{s1}^2}{\text{n1}}\right)^2}{\text{n1}-1} + \frac{\left(\frac{\text{s2}^2}{\text{n2}}\right)^2}{\text{n2}-1}}$$

In[5]:=
```
DF[2, 4, 24.77, 31.21]
```

Out[5]= 2.65917

2. Ant colonies are observed to see whether they are in "very late" stage of development. Independent equal - probability with - replacement samples of colonies are selected from colonies in two types of soil conditions. The data:

soil type 1	**44 out of 120 are "very late development"**
soil type 2	**38 out of 160 are in "very late development"**

There are an estimated 1800 colonies of type 1 soil condition and an estimated 1400 colonies of type 2 soil condition.

a. Determine a 95% z-based CI for the population fraction p_1 of "very late development" colonies in type 1 soil conditions. Use not FPC.

b. Same as (a) but with FPC. Has it made much of a difference?

c. Determine a 95% z-based CI for the difference p_1 - p_2 between the fractions of "very late development" colonies between the two types of soil conditions. Do not use FPC.

d. Same as (c) but using FPC. Has it made much of a difference?

3. Here are independent samples selected from two normal populations.

sample 1	3.24	7.15	7.21
sample 2	4.13	5.02	5.82

a. Determine a 95% t-based CI for the difference between their respective population means. The formula for "effective" DF gives 2.54.

b. Was the location of this CI relative to 0 a 95% event or a 5% event if actually the two populations have the same means?