Power I (2 sample)

Suppose we have two archaeological samples from two distinct, occupation areas at the same site at Lubbock Lake on the Southern High Plains, representing the mean and the standard deviations of flake artifact densities, measured as artifacts per square meter excavation unit. Both occupations are campsites and are well dated to two distinct time periods; the Pre-Ceramic, ca. 800 yr BP and the Proto-Apache, ca. 600 yr BP. We know historically that the Proto-Apache horizon represents the migration of Athapaskan populations into the southwestern United States, shortly before the arrival of the Spanish into the Southern High Plains, and that the Proto-Apache displaced the earlier indigenous population. Both populations are thought to have had a similar social structure, that is small, mobile, foraging bands that aggregated seasonally.

Even though the two campsites represent two distinct populations, and are clearly separated by time, it seems as though the campsites were used in similar ways, for similar amounts of time. One of the ways to look at this statistically is to run a 2 sample *t*-test on the average artifact densities of lithic flakes per area excavated. If we assume the presence of flakes represents tool manufacturing activities, then it would be very interesting to know if there is any statistical difference between the two samples.

So, we choose to run a 2 sample, 2-taied *t*-test on the data at hand, testing the hypothesis at the $a = 0.05$ (95%) level:

$$H_o : \overline{Y}_{PC} - \overline{Y}_{PA} = 0$$

$$H_a : \overline{Y}_{PC} - \overline{Y}_{PA} \neq 0$$

Let $\overline{Y}_{PC}$ = the mean number of flakes per excavation unit from the Pre-Ceramic sample, and let $\overline{Y}_{PA}$ = the mean number of flakes per excavation unit from the Proto-Apache sample.

The data is:

Pre-Ceramic Sample:

206.213	226.202	209.507	212.011
204.473	184.724	224.630	209.198
193.267	189.928	229.897	227.972
186.961	227.077	207.128	219.583
184.104	225.876	184.650	236.460
196.458	221.278	217.180	206.826
215.940	225.723	221.867	231.996
205.702	188.891	200.556	229.251
214.779	223.005	213.167	

Proto-Apache Sample:

202.985	213.267	213.302	195.612
167.167	197.708	207.976	212.937
217.418	233.415	192.396	194.263
229.483	202.645	213.069	233.028
209.328	202.033	167.526	220.342
219.739	227.662	212.586	218.784
185.481	222.624	185.951	203.811
218.811	218.708	187.435	203.506
182.863	217.459	216.800	200.470
200.856	227.290	235.030	200.024
196.696	199.129	196.059	188.917
226.850	204.813	223.946	

Running the descriptive stats we see:

Descriptive Statistics

```
Variable         N      Mean    Median  Tr Mean     StDev   SE Mean
PC              35    211.50    213.17   211.78     15.42      2.61
PA              47    207.41    207.98   208.02     16.23      2.37

Variable       Min       Max        Q1       Q3
PC          184.10    236.46    200.56   225.72
PA          167.17    235.03    196.70   218.81
```

Things look normal under the usual criteria, but we notice that the variances are roughly equal and the sample sizes are different. As a result we can pool the standard deviations:

$$se_p = s_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = 15.9\sqrt{\left(\frac{1}{35}+\frac{1}{47}\right)} = 3.5499$$

At this stage we can run our test in MINITAB following the usual set of options:

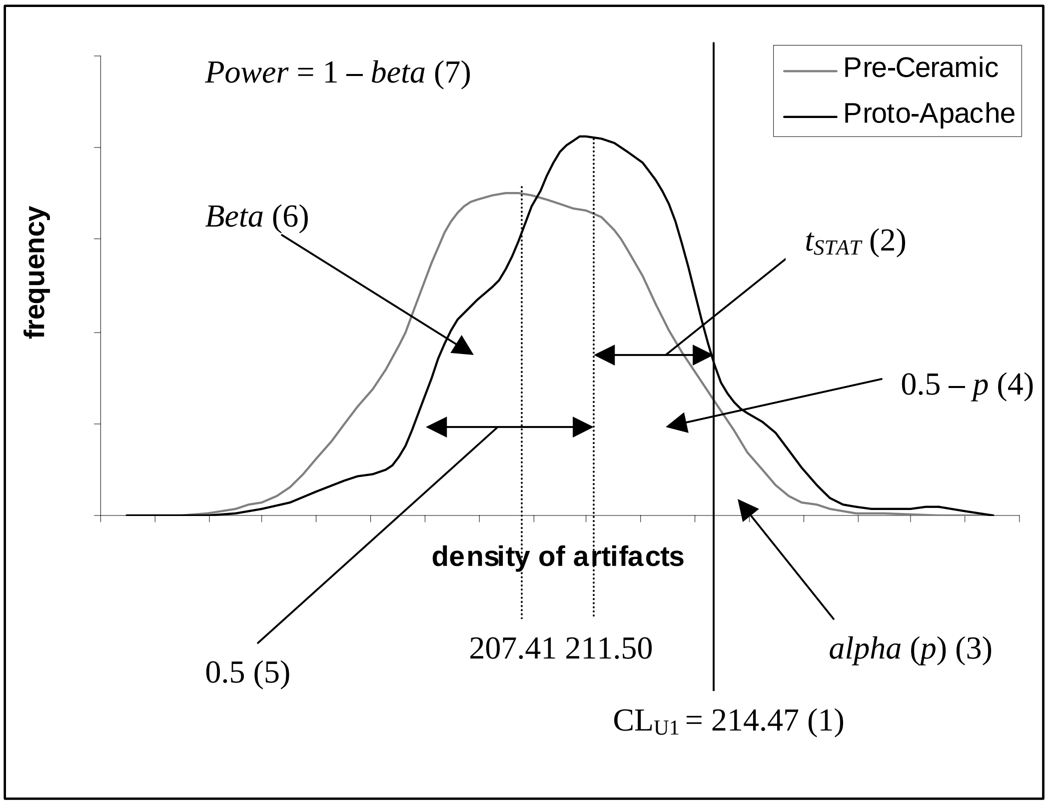
Two Sample T-Test and Confidence Interval

```
Two sample T for PC vs PA
     N      Mean     StDev   SE Mean
PC  35     211.5      15.4       2.6
PA  47     207.4      16.2       2.4

95% CI for mu PC - mu PA: ( -3.0,  11.2)
T-Test mu PC = mu PA (vs not =): T= 1.15  P=0.25  DF=  80
Both use Pooled StDev = 15.9
```

Bypassing the hand calculations for the sake of this example, we can calculate our t_{CRIT} using the information from the MINITAB output; t_{CRIT} (v = 80, a = 0.05) = 1.99 (approx.). We need this for future use.

So, looking at the output, our p = 0.25, which is much larger than our a level, and the confidence limits encompass zero. Therefore, we would fail to reject the null hypothesis at our stated a level, and conclude that the artifact densities at the two campsites are statistically equivalent. However, how certain can we be of this result?



First we need to identify our *beta* area; MINITAB does not do this for us. To do this we first calculate the upper confidence limit of the Pre-Ceramic distribution, this is because we are interested in calculating how much the two distributions overlap.

$$CL_{PC,UPPER} = \overline{Y}_{PC} + t_{CRIT}\left(se_p\right) = 207.41 + 1.99(3.5499) = 214.47$$

So, our *alpha* area is the region to the right of the upper bound. Our *beta* area is therefore the region to *left* of the upper bound belonging to the Proto-Apache distribution. This is because our *beta* error is the probability of failing to reject our null hypothesis when we in fact should have, so, the area on the graph that represents this probability is the region to the left of the upper bound belonging to the alternative distribution.

To calculate this area we first have to work out how to solve for this region under the curve. In this case, what we need to do is work out the distance of the upper bound of the Pre-Ceramic distribution from the mean of the Proto-Apache distribution using the t_{STAT} method:

$$t_{STAT} = \frac{CL_{PC,UPPER} - \overline{Y_{PA}}}{se_P} = \frac{214.47 - 211.50}{3.55} = 0.83 \text{ se units}$$

To calculate how much of the curve 0.83 se units from the mean covers we need to find the corresponding *p* value. We can do this either by estimating it from the *t* table (which can be confusing) or making EXCEL do it for us. In EXCEL we type the following procedure into a cell:

=TDIST(*x*, degrees of freedom, number of tails)

=TDIST(0.83, 80, 1) = 0.202373

Now, notice that we chose a 1-tailed distribution. Why? Because we want all our *alpha* error put on the right hand tail of the distribution, we are not interested in the left tail. So, the *p* value above is the area to the right of the upper bound, our *alpha* area, **not the area between the upper confidence limit and the Proto-Apache mean.** So, to get the area we are interested in then we calculate 0.5 – *p* = 0.297627, and as we are interested in the entire left hand side of the distribution (another 0.5) we calculate 0.297627 + 0.5 = 0.797627. This is our *beta* error (notice, we could have just calculated 1 – *p*).

As *Power* = 1 – *beta*, then *Power* = 1 - 0.797627 = 0.202373.

In this case, our *Power* was actually the area under the curve to the right of the upper bound that we calculated with the EXCEL formula. This is often the case, but beware, because the area under the curve you have to solve for changes with each test.

Our *Power* = 0.20, that is to say we can only be 20% confident in our original result of failing to reject our null hypothesis of no difference; that is to say that there could very well be a difference between the two artifact densities at the two campsites, we just didn't have a large enough sample to tell either way.

Remember, a standard ball-park figure for establishing whether or not your test is powerful enough is a minimum *Power* = 0.80.

Finally, in this case we would go back and try to increase our sample size in order to give ourselves enough power to be confident in our results.

To check that we have got our *Power* correct go to the stats page at UCLA http://calculators.stat.ucla.edu/powercalc/ and choose the correct test etc.