

2 Sample t -Test (unequal variances, equal sample size)

From an early Upper Paleolithic cave site in Israel we have two samples of hominid skeletons from the same stratigraphic layer; a sample of archaic *Homo sapiens* ($n = 65$) and a sample of *H. sapiens neandertalis* ($n = 40$). We are interested in establishing whether there is any significant difference between the cranial capacities of these two hominid species from this cave site.

Let \bar{Y}_{HS} = the sample mean of cranial capacities (cc) for the archaic *H. sapiens* specimens, and \bar{Y}_N = the sample mean of the cranial capacities (cc) for the *H. sapiens neandertalis* specimens.

Our null hypothesis states that at the $\alpha = 0.05$ level (95%):

$$H_O : \bar{Y}_{HS} - \bar{Y}_N = 0$$

$$H_A : \bar{Y}_{HS} - \bar{Y}_N \neq 0$$

Sample 1 (*H. sapiens*)

1515.76	1449.28	1562.24	1554.26	1557.96	1510.08
1543.60	1544.76	1601.12	1451.81	1426.20	1483.16
1514.12	1642.25	1615.58	1341.99	1528.13	1493.79
1487.02	1544.99	1666.85	1385.76	1450.12	1704.03
1486.32	1383.82	1592.90	1593.60	1490.65	1436.54
1657.99	1431.63	1498.54	1493.67	1610.25	1469.66
1376.44	1524.47	1525.90	1570.23	1455.96	1496.65
1379.01	1478.99	1555.18	1356.47	1501.22	1467.37
1451.92	1443.61	1439.34	1581.18	1551.29	1583.46
1493.35	1629.15	1579.33	1512.91	1583.87	1435.96
1614.69	1512.78	1509.27	1602.42	1513.34	

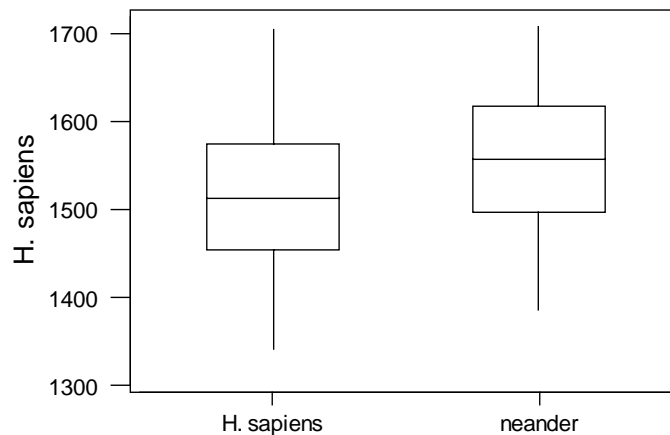
Sample 2 (*H. sapiens neandertalis*)

1484.87	1513.47	1585.60	1629.96	1556.09
1611.27	1431.65	1631.36	1452.82	1574.92
1603.14	1480.14	1512.62	1644.48	1515.71
1512.63	1619.96	1538.36	1493.08	1438.90
1618.47	1596.34	1558.54	1628.90	1707.60
1598.62	1466.86	1385.80	1501.70	1569.05
1679.39	1518.82	1622.09	1579.56	1659.30
1535.89	1605.96	1525.57	1495.84	1452.51

Assuming the above data sets are normally distributed, we can test the above null hypothesis with a 2 sample t -test, taking into account sample sizes and variances.

The descriptive output and boxplots from MINITAB are:

Descriptive Statistics						
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
H. sapie	65	1514.6	1512.8	1514.3	78.4	9.7
neander	40	1553.4	1557.3	1553.7	74.7	11.8
Variable	Min	Max	Q1	Q3		
H. sapie	1342.0	1704.0	1453.9	1574.8		
neander	1385.8	1707.6	1497.3	1616.7		



Looking at the means and medians for both samples, we see that they are pretty similar. The boxplot shows that the medians fall roughly in the middle of the body of the data, and the distributions seem to be symmetrical. We conclude that these data are close enough to being normally distributed to go ahead with a t -test. We note that the sample sizes are unequal, but the standard deviations are similar enough to assume equal variances.

Therefore, we can pool our standard errors, using the familiar equations:

$$se_p = s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \text{ where } s_p = \sqrt{\frac{(s_1^2 + s_2^2)}{2}}, \text{ or equivalently } s_p = \frac{s_1 + s_2}{2},$$

So, in our case:

$$s_p = \frac{s_1 + s_2}{2} = \frac{78.4 + 74.7}{2} = 76.55,$$

And:

$$se_p = 76.55 \sqrt{\left(\frac{1}{65} + \frac{1}{40}\right)} = 76.55(0.20093) = 15.38$$

To calculate our degrees of freedom (ν) we do not have to bother with the Satterthwaite approximation as we have assumed our variances are equal, so we can use the more familiar equation $\nu = n_1 + n_2 - 2 = 103$. So, our t_{CRIT} ($\nu = 103$, $\alpha = 0.05$) = 1.98.

To calculate our confidence limits at the 95% level:

$$L_L = (\bar{Y}_{HS} - \bar{Y}_N) - 1.98(15.38) = -38.8 - 30.45 = -69.25$$

$$L_U = (\bar{Y}_{HS} - \bar{Y}_N) + 1.98(15.38) = -38.8 + 30.45 = -8.35$$

Given these bounds, we would reject the null hypothesis that there is no statistical difference between the two samples, in favor of the alternative, as the bounds do not encompass zero.

To calculate our actual t -test:

$$t_{STAT} = \frac{\bar{Y}_{HS} - \bar{Y}_{NS}}{se_p} = \frac{38.8}{15.38} = 2.52$$

As our $t_{CRIT} = 1.98$, $t_{STAT} > t_{CRIT}$ indicating that we should reject our null hypothesis of no difference in favor of the alternative.

We would conclude from this test that there is a statistical difference at the 95% level in mean cranial capacity between a sample of archaic *H. sapiens* and *H. sapiens neandertalis*.

To run this test in MINITAB we follow these procedures:

```

>STAT
  >BASIC STATISTICS
    >2 SAMPLE t
      >Click SAMPLES IN DIFFERENT COLUMNS
        >Put the samples in the appropriate places
          >ALTERNATIVE is NOT EQUAL
            >Click ASSUME EQUAL VARIANCE
              >OKAY
  
```

And the resulting output is:

Two Sample T-Test and Confidence Interval

Two sample T for H. sapiens vs neander

	N	Mean	StDev	SE Mean
H. sapie	65	1514.6	78.4	9.7
neander	40	1553.4	74.7	12

95% CI for mu H. sapie - mu neander: (-69.6, -8)

T-Test mu H. sapie = mu neander (vs not =): T= -2.51 P=0.014 DF= 103

Both use Pooled StDev = 77.0

So, to check ourselves we first look at the *Pooled StDev* in the output, and MINITAB gives us a value of 77, whereas we calculated 76.55. The confidence limits at the 95% level are given as -69.6, and -8, whereas we calculated -69.25, and -8.35. The $T = -2.51$ in the output, and we calculated it as (-)2.52, with $DF = 103$. So, all our hand computed values came out pretty similar to the MINITAB output, and we see that MINITAB gives us the final $p = 0.014$; a value less than our α , so we reject the null hypothesis.