ORIGIN ≡ 0

Two Sample t-Test for Populations with Equal Variances

This test is employed where two sets of measurements are derived from samples with approximately equal variances.

ZAR := READPRN("c:/DATA/Biostatistics/ZarEX8.1M.txt") **< Zar Example 8.1**

$i := 0..5$

$X1_i := \left(ZAR^{\langle 1 \rangle}\right)_i$ $X2 := ZAR^{\langle 2 \rangle}$

$n_1 := \text{length}(X1)$ $n_1 = 6$ $n_2 := \text{length}(X2)$ $n_2 = 7$

$X1_{bar} := \text{mean}(X1)$ $X1_{bar} = 8.75$ $X2_{bar} := \text{mean}(X2)$ $X2_{bar} = 9.742857$

$s_1 := \sqrt{\text{Var}(X1)}$ $s_1 = 0.582237$ $s_2 := \sqrt{\text{Var}(X2)}$ $s_2 = 0.818244$

ZAR =

	0	1	2
0	1	8.8	9.9
1	2	8.4	9
2	3	7.9	11.1
3	4	8.7	9.6
4	5	9.1	8.7
5	6	9.6	10.4
6	7	0	9.5

^ Note: since MathCad doesn't allow empty or 'NA' data values in its data arrays, I had to put 0.0 as a value for $X1_6$ in the original data table ZAR8.1.txt. I then read data for X1 by indexing the values using variable *i*. Most statistical programs, including R, does a much better job with missing values.

Assumptions:

- **Observed values $X_{1,1}$, $X_{1,2}$, $X_{1,3}$, ... $X_{1,n1}$ are a random sample from $\sim N(\mu_1,\sigma_1^2)$**
- **Observed values $X_{2,1}$, $X_{2,2}$, $X_{2,3}$, ... $X_{2,n2}$ are a random sample from $\sim N(\mu_2,\sigma_2^2)$**
- **Variances σ_1^2 & σ_2^2 are approximately equal but *unknown.***
- **Samples $X_{1,n1}$ and $X_{2,n2}$ are *independent.***

^ Note: this test is reasonably robust especially if sample sizes are nearly equal and two tailed hypotheses are entertained.

Hypotheses:

H_0: $\mu_1 = \mu_2$ **< No difference in mean between populations $X_{1.}$ & $X_{2.}$**

H_1: $\mu_1 \neq \mu_2$ **< Two Sided Case**

H_1: $\mu_1 < \mu_2$ **< One Sided Case Lower Tail)**

H_1: $\mu_1 > \mu_2$ **< One Sided Case (Upper Tail)**

Pooled Sample Variance:

$$s_p := \sqrt{\frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}}$$

$s_p^2 = 0.5192857$ **< variance is pooled from the two samples and adjusted for each sample's size n_1 & n_2.**

Test Statistic:

$$t := \frac{X1_{bar} - X2_{bar}}{\sqrt{s_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$t = -2.4765$ **< t is the normalized mean X_2bar - X_2bar**

< s_p^2 is the pooled sample variance defined above

$n_1 + n_2 - 2 = 11$ **< degrees of freedom of pooled variance**

Sampling Distribution:

If Assumptions hold and H_0 is true, then $t \sim t_{(n1+n2-2)}$

Critical Values of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$C_1 := qt\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)$ $C_1 = -2.201$ **< Two sided lower Critical Value**

$C_2 := qt\left(1 - \frac{\alpha}{2}, n_1 + n_2 - 2\right)$ $C_2 = 2.201$ **< Two sided upper Critical Value**

$C := |C_1|$ $C = 2.201$ **< Critical value used for two sided test (to simplify)**

$C_3 := qt(\alpha, n_1 + n_2 - 2)$ $C_3 = -1.7959$ **< One sided lower Critical Value**

$C_4 := qt(1 - \alpha, n_1 + n_2 - 2)$ $C_4 = 1.7959$ **< One sided upper Critical Value**

Decision Rules:

IF |t| > C THEN REJECT H_0, OTHERWISE ACCEPT H_0 **< Two sided case**
IF t < C_3, THEN REJECT H_0, OTHERWISE ACCEPT H_0 **< One sided case lower tail**
IF t > C_4, THEN REJECT H_0, OTHERWISE ACCEPT H_0 **< One sided case upper tail**

$t = -2.4765$

Probability Values:

$P := 2 \cdot pt(t, n_1 + n_2 - 2)$ $P = 0.03076$ **< if** $t \le 0$ **< Two sided case**

$P := 2 \cdot (1 - pt(t, n_1 + n_2 - 2))$ $P = 1.969235$ **< if t > 0**

$P := pt(t, n_1 + n_2 - 2)$ $P = 0.015382$ **< One sided case lower tail**

$P := 1 - pt(t, n_1 + n_2 - 2)$ $P = 0.984618$ **< One sided case upper tail**

Confidence Intervals for the difference in mean:

$$CI := \left[X1_{bar} - X2_{bar} - C \cdot \sqrt{s_p^{\,2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad X1_{bar} - X2_{bar} + C \cdot \sqrt{s_p^{\,2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \right]$$

$X1_{bar} - X2_{bar} = -0.9929$

$CI = (-1.8752609 \quad -0.1104534)$ **< Two sided case**

$$CIL := X1_{bar} - X2_{bar} - C_3 \cdot \sqrt{s_p^{\,2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

minus infinity to $CIL = -0.2728634$

^ One sided case lower tail

$$CIU := X1_{bar} - X2_{bar} - C_4 \cdot \sqrt{s_p^{\,2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$CIU = -1.712851$ **to infinity** **< One sided case upper tail**

Prototype in R:

```
#ZAR EXAMPLE 8.1
ZAR=read.table("c:/DATA/Biostatistics/ZarEX8.1.txt", na.strings="NA")
ZAR
attach(ZAR)
X1=na.omit(gpB)
X1
X2=gpG
X2
n1=length(X1)
n1
n2=length(X2)
n2
X1bar=mean(X1)
X1bar
X2bar=mean(X2)
X2bar
s1=sqrt(var(X1,na.rm=TRUE))
s1
s2=sqrt(var(X2))
s2
sp=((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2)
sp

#PERFORMING TWO-SAMPLE TWO SIDED t-TEST:
t.test(X1,X2,alternative="two.sided",var.equal=TRUE, conf.level=0.95)
detach(ZAR)
```

```
        Two Sample t-test

data:  X1 and X2
t = -2.4765, df = 11, p-value = 0.03076
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.8752609 -0.1104534
sample estimates:
mean of x mean of y
 8.750000  9.742857
```

```
#CHECK THE FOLLOWING DATASET FOR A MORE
#EFFICIENT WAY TO INPUT DATA USING read.table()

ZAR2=read.table("c:/DATA/Biostatistics/ZarEX8.1R.txt")
ZAR2
attach(ZAR2)
X1=data[group=="gpB"]
X1
X2=data[group=="gpG"]
X2
```

```
> ZAR2
  data group
1  8.8   gpB
2  8.4   gpB
3  7.9   gpB
4  8.7   gpB
5  9.1   gpB
6  9.6   gpB
7  9.9   gpG
8  9.0   gpG
9  11.1  gpG
10  9.6  gpG
11  8.7  gpG
12 10.4  gpG
13  9.5  gpG
```

#PERFORMING TWO-SAMPLE ONE SIDED LOWER TAIL t-TEST:
t.test(X1,X2,alternative="less",var.equal=TRUE, conf.level=0.95)

```
        Two Sample t-test

data:  X1 and X2
t = -2.4765, df = 11, p-value = 0.01538
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.2728634
sample estimates:
mean of x mean of y
 8.750000  9.742857
```

#PERFORMING TWO-SAMPLE ONE SIDED UPPER TAIL t-TEST:
t.test(X1,X2,alternative="greater",var.equal=TRUE, conf.level=0.95)

```
        Two Sample t-test

data:  X1 and X2
t = -2.4765, df = 11, p-value = 0.9846
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.712851       Inf
sample estimates:
mean of x mean of y
 8.750000  9.742857
```