Examples for 11/19/15, Part 1

Population 1
mean μ_1
std. dev. σ

$\Downarrow$

$y_{11}, y_{21}, \dots, y_{n_1 1}$

$\overline{y}_1,\ s_1^2$

Population 2
mean μ_2
std. dev. σ

$\Downarrow$

$y_{12}, y_{22}, \dots, y_{n_2 2}$

$\overline{y}_2,\ s_2^2$

Assumptions:

1) Two independent samples.
2) Both populations are normal.
3) The population standard deviations are equal.

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$$

A confidence interval for $\mu_1 - \mu_2$ is

$$\left(\overline{y}_1 - \overline{y}_2\right) \pm t_{\alpha/2} \cdot s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$n_1 + n_2 - 2$ degrees of freedom

To test $\quad \mathrm{H}_0: \mu_1 - \mu_2 = \delta_0$

Test Statistic: $$t = \frac{\left(\overline{y}_1 - \overline{y}_2\right) - \delta_0}{s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

$n_1 + n_2 - 2$ degrees of freedom

1. Assume that the distributions of Y_1 and Y_2 are $\mathrm{N}(\mu_1, \sigma^2)$ and $\mathrm{N}(\mu_2, \sigma^2)$, respectively. Given the $n_1 = 6$ observations of Y_1,

65, 68, 67, 66, 71, 68

and the $n_2 = 8$ observations of Y_2,

68, 70, 67, 71, 70, 72, 73, 69

test $\mathrm{H}_0: \mu_1 = \mu_2$ vs. $\mathrm{H}_1: \mu_1 \neq \mu_2$. Use a 5% level of significance. What can you say about the p-value of this test?

$\bar{y}_1 = \frac{405}{6} = 67.5$ $\bar{y}_2 = \frac{560}{8} = 70$

$s_1^2 = \frac{21.5}{5} = 4.3$ $s_2^2 = \frac{28}{7} = 4$

$$s_{\text{pooled}}^2 = \frac{(6-1)\cdot 4.3 + (8-1)\cdot 4}{6+8-2} = 4.125$$

Test Statistic: $$t = \frac{(67.5-70)-0}{\sqrt{4.125\cdot\left(\frac{1}{6}+\frac{1}{8}\right)}} = -2.279.$$

$n_1 + n_2 - 2 = 12$ degrees of freedom

The t Distribution

r	$t_{0.40}$	$t_{0.25}$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055

Rejection Region: Reject H_0 if $t < -t_{0.025}(12)$ or $t > t_{0.025}(12)$

$t_{0.025}(12) = 2.179.$ **Reject H_0 at $\alpha = 0.05$.**

OR

-2.681	$<$	-2.279	$<$	-2.179
$-t_{0.01}(12)$	$<$	t	$<$	$-t_{0.025}(12)$
0.01	$<$	one tail	$<$	0.025

p-value = two tails

0.02 < p-value < 0.05. **Reject H_0 at $\alpha = 0.05$.**

$J = 2.$ $\qquad$ $N = n_1 + n_2 = 6 + 8 = 14.$

$$\overline{y} = \frac{n_1 \cdot \overline{y}_1 + n_2 \cdot \overline{y}_2 + \ldots + n_J \cdot \overline{y}_J}{N} = \frac{6 \cdot 67.5 + 8 \cdot 70}{14} = \frac{965}{14} = 68.92857.$$

$$\text{SSB} = n_1 \cdot \left(\overline{y}_1 - \overline{y}\right)^2 + n_2 \cdot \left(\overline{y}_2 - \overline{y}\right)^2 + \ldots + n_J \cdot \left(\overline{y}_J - \overline{y}\right)^2$$

$$= 6 \cdot (67.5 - 68.92857)^2 + 8 \cdot (70 - 68.92857)^2 = 21.42857.$$

$$\text{MSB} = \frac{\text{SSB}}{J-1} = \frac{21.42857}{1} = 21.42857.$$

$$\text{SSW} = \left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2 + \ldots + \left(n_J - 1\right) \cdot s_J^2$$

$$= 5 \cdot 4.3 + 7 \cdot 4 = 49.5.$$

$$\text{MSW} = \frac{\text{SSW}}{N-J} = \frac{49.5}{12} = 4.125. \qquad [= s_{\text{pooled}}^2]$$

$$\text{SSTot} = \text{SSB} + \text{SSW} = 70.92857.$$

Test Statistic: $\qquad$ $\text{F} = \dfrac{\text{MSB}}{\text{MSW}} = \dfrac{21.42857}{4.125} = \mathbf{5.1948}.$ $\qquad$ $[= t^2]$

ANOVA table:

Source	**SS**	**DF**	**MS**	**F**
Between	21.42857	1	21.42857	**5.1948**
Within	49.5	12	4.125	
Total	70.92857	13		

Critical Value: $\qquad$ $\text{F}_{0.05}(1, 12) = 4.75.$ $\qquad$ $[= (t_{0.025}(12))^2]$

Reject $\mathbf{H_0}$ at $\alpha = 0.05$.

$F_{0.05}(1, 12) = 4.75 \quad < \quad 5.1948 \quad < \quad 9.33 = F_{0.01}(1, 12).$

0.05 > p-value > 0.01. **Reject H_0 at $\alpha = 0.05$.**

In general,

$$J = 2. \qquad\qquad N = n_1 + n_2.$$

$$\bar{y} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{N} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2}.$$

$$\text{MSB} = \frac{\text{SSB}}{J-1} = n_1 \cdot \left(\bar{y}_1 - \bar{y}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \bar{y}\right)^2$$

$$= n_1 \cdot \left(\bar{y}_1 - \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2} \right)^2 + n_2 \cdot \left(\bar{y}_2 - \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2}{n_1 + n_2} \right)^2$$

$$= n_1 \cdot \left(\frac{n_2 \cdot \left(\bar{y}_1 - \bar{y}_2\right)}{n_1 + n_2} \right)^2 + n_2 \cdot \left(\frac{n_1 \cdot \left(\bar{y}_2 - \bar{y}_1\right)}{n_1 + n_2} \right)^2$$

$$= n_1 \cdot \frac{n_2^2 \cdot \left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(n_1 + n_2\right)^2} + n_2 \cdot \frac{n_1^2 \cdot \left(\bar{y}_2 - \bar{y}_1\right)^2}{\left(n_1 + n_2\right)^2}$$

$$= n_1 n_2 \cdot \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(n_1 + n_2\right)} = \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\text{MSW} = \frac{\text{SSW}}{N - J} = \frac{\left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2}{n_1 + n_2 - 2} = s_{\text{pooled}}^2$$

$$\text{F} = \frac{\text{MSB}}{\text{MSW}} = \frac{\left(\bar{y}_1 - \bar{y}_2\right)^2}{s_{\text{pooled}}^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = t^2.$$

```r
y1 <- c(65,68,67,66,71,68)
y2 <- c(68,70,67,71,70,72,73,69)
t.test(y1, y2, alternative = c("two.sided"), var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  y1 and y2
## t = -2.2792, df = 12, p-value = 0.04174
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.8898756 -0.1101244
## sample estimates:
## mean of x mean of y
##      67.5      70.0
```

```r
y <- c(y1,y2)
pop <- c(rep(1,6),rep(2,8))
result = lm(y ~ factor(pop))
summary(aov(result))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## factor(pop)  1  21.43  21.429   5.195 0.0417 *
## Residuals   12  49.50   4.125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```