Semester 1 | Solution to Tutorial Set 10 | 2013

1. Answer: (b) Use t-table with df = 27 - 1 = 26 and upper area = 0.05.

2. Answer: (c) The pooled variance is $s_p^2 = \frac{(12-1)8.75^2+(15-1)10^2}{12+15-2} = \frac{2242.188}{25} = 89.6875$. Therefore, $s_p = \sqrt{89.6875} = 9.47$.

3. Answer: (d) Reason: df $= n_1 + n_2 - 2 = 12 + 15 - 2 = 25$.

4. Answer: The test statistic is $t_{\text{obs}} = \dfrac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{119.4 - 116.25}{9.47\sqrt{\frac{1}{12} + \frac{1}{15}}} = 0.86$.

   The P-value $= 2\text{P}(T_{25} > |0.86|) > 0.05$. Since this is large, we have evidence consistent with H_0.

   Assumptions : Each sample is from a normal population. The variances of the two populations are equal. The two samples are distributed independently of each other.

5. Let μ_1 = mean wear with sandpaper A and μ_2 = mean wear with sandpaper B.

   1. The hypotheses: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$.

   2. Test statistic:
   $$t_{\text{obs}} = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{27.4 - 24.1}{2.75\sqrt{\frac{1}{10} + \frac{1}{11}}} = 2.74,$$
   where $s_p^2 = \dfrac{9(2.3)^2 + 10(3.1)^2}{10 + 11 - 2} = 2.75^2$.

   3. The P-value: $2\text{P}(T_{19} > |2.74|) \in (0.01, 0.02)$.

   4. Conclusion: Since P-value < 0.05, we have strong evidence against H_0. That is, the mean abrasive wear is different for the two sandpapers.

6. Let p_1 = proportion of women who caught the flu and p_2 = proportion of men who caught the flu.

   1. Hypotheses: $H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$.

   2. Test statistic:
   $$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.38 - 0.51}{\sqrt{0.47(0.53)\left(\frac{1}{100} + \frac{1}{200}\right)}} = -2.13,$$

where $\hat{p} = \dfrac{100(0.38) + 200(0.51)}{100 + 200} = 0.47$.

3. The P-value: $2\text{P}(Z > |-2.13|) = 2(1 - 0.9834) = 0.0332$.

4. Conclusion: Since P-value < 0.05, we have strong evidence against H_0. That is, the percent of people with the flu differs between the genders.

7. Let p_1 = proportion of patients with reduced iron among the treated and p_2 = proportion of patients with reduced iron among the placebo group.

1. Hypotheses: $H_0 : p_1 = p_2$ vs. $H_1 : p_1 > p_2$.

2. Test statistic:

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{35}{120} - \frac{19}{120}}{\sqrt{0.225(1-0.225)\left(\frac{1}{120} + \frac{1}{120}\right)}} = 2.47,$$

where $\hat{p} = \dfrac{19 + 35}{120 + 120} = 0.225$.

3. The P-value: $\text{P}(Z > 2.47) = 1 - 0.9932 = 0.0068$.

4. Conclusion: Since P-value < 0.05, we have strong evidence against H_0. That is, the percent of people with reduced iron is higher among the treated patients.

8. For the P-value use `1- pt(3.45, 26)`. Answer is 0.0009627558. For the critical value use `qt(0.95, 26)`. Answer is 1.705618.

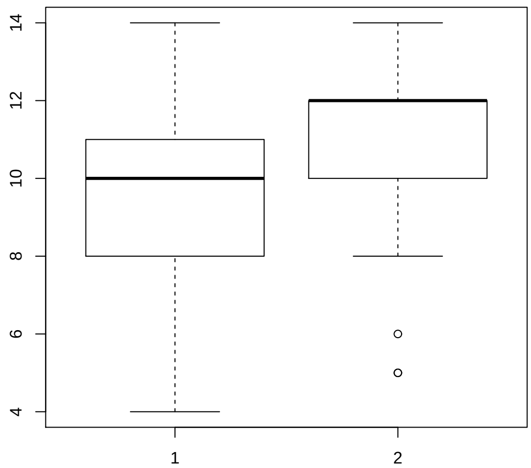
9.
```
> x=scan()
1: 6 7 11 11 8 12 12 11 13 10 10 12 4 8 9 10 8 7 11 8 10 14 11 5 8 7 8 12
29:
Read 28 items
> y=scan()
1: 12 5 10 12 9 13 12 9 12 13 13 14 12 12 6 11 12 8 12 12 10 11 11 14 10 12 5 12 14 11
31:
Read 30 items
>t.test(x, y, var.equal = T)

        Two Sample t-test

data:  x and y
t = -2.4611, df = 56, p-value = 0.01696
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.8548132 -0.2928059
sample estimates:
mean of x mean of y
 9.392857 10.966667

>boxplot(x,y)
```



Conclusion: Since the P-value $= 0.01696 < 0.05$, the data provide strong evidence against the null hypothesis. Therefore, we have evidence that the mean marks in the two tutorials are different.

Comment: It seems that the midmorning tutorial has marks that are not normally distributed - the distribution is heavily left skewed and has two outliers. Also the equality of variance assumption might not be met because the spread of the two distribution are not similar. The midmorning tutorial has lower spread (variance) after excluding the two outliers.

10. `prop.test(c(35, 19), c(120, 120), alt = "g", correct = F)`

```
        2-sample test for equality of proportions without continuity
        correction

data:  c(35, 19) out of c(120, 120)
X-squared = 6.1171, df = 1, p-value = 0.006694
alternative hypothesis: greater
95 percent confidence interval:
```

```
 0.0457972 1.0000000
sample estimates:
   prop 1    prop 2
0.2916667 0.1583333
```

```
2.47^2
[1] 6.1009
```

Result almost the same as `X-squared = 6.1171` (if we keep 3 dp for $z_0 = 2.473$, then the result is more accurate).

```
1-pnorm(2.47)
[1] 0.006755653
```

Again, result is almost identical to the P-value calculated from the Z-table, except that `R` provides more precision.

Additional Problems to Week 10 - Solutions

1. Two sample t-test:

Year	Size	Mean	St. Dev
1975	30	12.78	0.43
1985	40	12.42	0.67

Let μ_1 = mean age at menarche of girls in the 1975 class and μ_2 = mean age at menarche of girls in the 1985 class.

1. The hypotheses: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$.

2. Test statistic:

$$t_0 = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{12.78 - 12.42}{0.5799239\sqrt{\frac{1}{30} + \frac{1}{40}}} = 2.570237,$$

where $s_p^2 = \dfrac{29(0.43)^2 + 39(0.67)^2}{30 + 40 - 2} = 0.5799239^2$.

3. The P-value: $2\text{P}(T_{68} > 2.570237) \in (0.005, 0.01)$.

4. Conclusion: Since P-value < 0.05, we have strong evidence against H_0. That is, the mean age at menarche of girls in the 1975 class is different from those from the 1985 class.

If the comparison is made on classes from the same school, the school effect is controlled or eliminated.

2. Paired sample t-test:

Person i	1	2	3	4	5	6	7	8	9	10	mean	s.d.
Aspirin A x_i	15	26	13	28	17	20	7	36	12	18		
Aspirin B y_i	13	20	10	21	17	22	5	30	7	11		
Difference d_i	2	6	3	7	0	-2	2	6	5	7	3.6	3.0984

Let μ_1 =1-hour concentration for Aspirin A and μ_2 =1-hour concentration for Aspirin B.

1. The hypotheses: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$.

2. Test statistic:

$$t_0 = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{3.6 - 0}{3.0984/\sqrt{10}} = 3.674219,$$

3. The P-value: $2\text{P}(T_9 > 3.674219) \in (0.001, 0.005)$.

4. Conclusion: Since P-value < 0.05, we have strong evidence against H_0. That is, the 1-hour concentration for Aspirin A is different from those from Aspirin B.

3. Two sample Z-test:

Group	Size	Count
Current smoker	2000	40
Ex-smoker	1000	10

Let p_1 = proportion of current woman smoker who developed MI and p_2 = proportion of woman ex-smoker who developed MI.

1. Hypotheses: $H_0 : p_1 = p_2$ vs. $H_1 : p_1 > p_2$.

2. Test statistic:

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{40}{2000} - \frac{10}{1000}}{\sqrt{\frac{50}{3000}(1 - \frac{50}{3000})\left(\frac{1}{2000} + \frac{1}{1000}\right)}} = 2.016878,$$

where $\hat{p} = \dfrac{40 + 10}{2000 + 1000} = \dfrac{50}{3000}$.

3. The P-value: $2\text{P}(Z > 2.016878) = 2(1 - 0.9783) = 0.0434$.

4. Conclusion: Since P-value < 0.05, we have sufficient evidence against H_0. That is, the incidence rate of MI among current woman smoker is different from those among woman ex-smoker.