Stat509 Fall 2014 Homework 7

Instructor: Peijie Hou

11/11/2014

Instruction: There will be no quiz based on this homework, we will have the second midterm on 11/20.

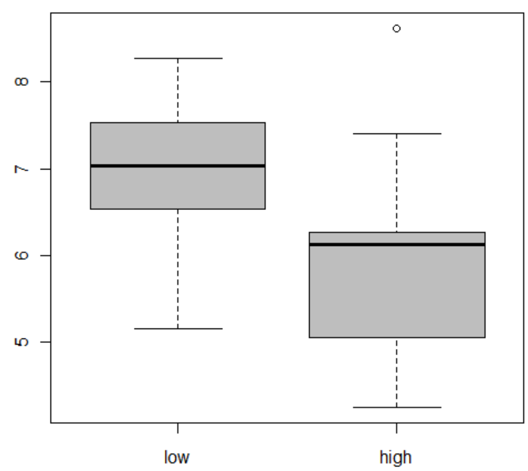
1. Data on pH for 16 random batches of low and high volt electrolyte were collected. The data are given by

   `Low volt:  7.78 5.77 7.08 6.75 7.09 8.27 6.5 5.16 6.81 7.28 7.88 7.87 7.2 5.95 6.58 6.99`
   and
   `high volt:  4.54 5.04 5.07 6.18 8.62 6.28 7.41 6.17 6.25 4.25 6.08 7.23 4.68 6.19 5.85 5.83`

   (a) Use `boxplot(sample1,sample2)` to draw the side-by-side boxplot in R. Do you think it is reasonable to assume $\sigma_1^2 = \sigma_2^2$ based on the plot?

   ```
   low<-c(7.78,5.77,7.08,6.75,7.09,8.27,6.5,5.16,6.81,7.28,7.88,7.87,7.2,5.95,6.58,6.99)
   high<-c(4.54,5.04,5.07,6.18,8.62,6.28,7.41,6.17,6.25,4.25,6.08,7.23,4.68,6.19,5.85,5.83)
   boxplot(low,high,names=c("low","high"),col="grey")
   ```

   Based on the plot, we do not have any overwhelming evidence to support $\sigma_1^2 = \sigma_2^2$.



   (b) In R, command `var.test(sample1, sample2)` can be used to test equal variance assumption, where `sample1` and `sample2` are the names of the data vector your give in R. Compare the R output with your calculation in part (a).

   `var.test(low, high)`

   You can draw the conclusion based on the p-value or you can check whether the confidence interval includes “1” or not.

   ```
   > var.test(low,high)

           F test to compare two variances
   ```

```
data:  low and high
F = 0.5338, num df = 15, denom df = 15, p-value = 0.2356
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1865103 1.5278126
sample estimates:
ratio of variances
         0.5338097
```

The p-value is 0.2356 (> 0.05) and the confidence interval includes "1", which suggest the equal variance assumption is satisfied. This is consistent with the solution in part (a)

(c) Assuming the two samples are independent. The engineer want to test that the low volt average pH is greater than the high volt average pH. Let μ_L be the average pH of low volt electrolyte and μ_H be the average pH of high volt electrolyte. State the null and alternative hypotheses.
We want to test $H_0 : \mu_L = \mu_H$ versus $H_a : \mu_L > \mu_H$.

(d) Calculate the appropriate test statistic for the test. The sample means and sample variances can be computed using R.
The summary statistics are

```
> mean(low);mean(high);var(low);var(high)
[1] 6.935
[1] 5.979375
[1] 0.6896
[1] 1.291846
```

The test statistic is

$$t_0 = \frac{\overline{y}_l - \overline{y}_h}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{6.935 - 5.979}{1.061\sqrt{\frac{1}{16} + \frac{1}{16}}} = 2.717,$$

where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \dfrac{15(0.690) + 15(1.292)}{16 + 16 - 2} = 0.991.$

(e) Use R to calculate the p-value of the test.
The p-value is very small.

```
> 1-pt(2.717,30)
[1] 0.005415681
```

(f) State your conclusion at a 0.05 level of significance.
We will reject H_0 at 0.05 level of significance. We have sufficient evidence to conclude that the low volt average pH is greater than the high volt average pH.

(g) Use `t.test` in `R` to check your work.

```
t.test(low,high,alternative="greater",paired = FALSE, var.equal = TRUE)
```

R output:

```
> t.test(low,high,alternative="greater",paired = FALSE, var.equal = TRUE)

        Two Sample t-test

data:  low and high
t = 2.7155, df = 30, p-value = 0.005435
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3583425       Inf
sample estimates:
mean of x mean of y
 6.935000  5.979375
```

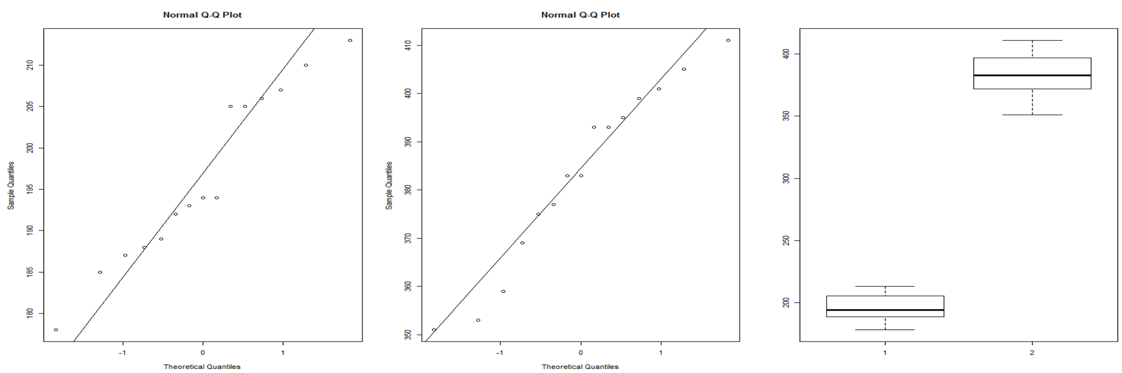
2. The deflection temperature under load for two different types of plastic pipe is being investigated. Two random samples of 15 pipe specimens are tested, and the deflection temperatures observed are as follows

(in Fahrenheit):

```
Type 1:  206, 188, 205, 187, 194, 193, 207, 185, 189, 213, 192, 210, 194, 178, 205
Type 2:  353, 393, 411, 401, 359, 351, 369, 399, 393, 383, 395, 375, 377, 405, 383
```

(a) Construct box plots and normal probability plots for the two samples. Do these plots provide support of the assumptions of normality and equal variances? Write a practical interpretation for these plots.

```
Type1<-c(206, 188, 205, 187, 194, 193, 207, 185, 189, 213,192, 210, 194, 178, 205)
Type2<-c(353, 393, 411, 401, 359, 351, 369, 399, 393, 383, 395, 375, 377, 405, 383)
qqnorm(Type1); qqline(Type1)
qqnorm(Type2); qqline(Type2)
boxplot(Type1, Type2)
```



Based on the plot, it is reasonable to assume the data came from normal populations. It is hard to tell about the equal variances assumption. Let's perform an formal variance test.

```
> var.test(Type1,Type2)

        F test to compare two variances

data:  Type1 and Type2
F = 0.3083, num df = 14, denom df = 14, p-value = 0.03523
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1034975 0.9182280
sample estimates:
ratio of variances
         0.3082763
```

The test rejects the equal variance assumption.

(b) Do the data support the claim that the mean deflection temperature under load for type 2 pipe exceeds that of type 1? Use $\alpha = 0.05$. Do the analysis in R.

The p-value is less than 0.05, we reject H_0 : the mean deflection temperature under load for type 2 equals that of type 1, there is strong evidence to support the mean deflection temperature for type 2 exceeds that of type 1. Note that you need to specify `var.equal=F` to assume unequal variances.

```
> t.test(Type1,Type2,alternative="less",var.equal=F)

        Welch Two Sample t-test

data:  Type1 and Type2
t = -33.4989, df = 21.883, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
```

```
      -Inf -177.1592
sample estimates:
mean of x mean of y
 196.4000  383.1333
```

3. According to research published in Science (Feb 20,2004), the mere belief that you are receiving an effective treatment for pain can reduce the pain you actually feel. Researchers tested the placebo effect on 24 volunteers. Each volunteer was put inside an MRI for two consecutive sessions. During the first session electric shocks were applied to their arms and the blood oxygen level-dependent (BOLD) signal (a measure related to neural activity in the brain)was recorded during pain. The second session was identical to the first, except that prior to applying the electric shocks the researchers smeared a cream on the volunteer's arms. The volunteers were informed that the cream would block the pain, when, in fact, it was just a regular skin lotion (ie, placebo). If the placebo is effective in reducing pain, the BOLD measurements should be higher on average, in the first MRI session than in the second MRI session. The differences are calculated by subtracting second MRI measurement from first MRI measurement.

   (a) State the null and alternative hypotheses.
   Let $\mu_D = \mu_1 - \mu_2$, where μ_1 is the population mean bold measurements without placebo, and μ_2 is the population mean bold measurements with placebo. We want to test $H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$.

   (b) The differences between the first BOLD measurements and the second were computed and the summarized results is as follows:

Variable	n	$\overline{y}_D$	s_D
size	24	0.21	0.47

   Calculate the test statistic.
   The test statistic is

$$t_0 = \frac{\overline{y}_D - 0}{s_D/\sqrt{n}} = \frac{0.21}{0.47/\sqrt{24}} = 2.189.$$

   (c) Calculate the p-value.
   The p-value of the test is 0.020.

```
> 1-pt(2.189,23)
[1] 0.01950389
```

   (d) State your conclusion.
   We reject $H_0 : \mu_D = 0$ at 0.05 significance level, since p-value is less than 0.05. We conclude that there exists the placebo effect.

4. A programmable lighting control system is being designed. The purpose of the system is to reduce electricity consumption costs in buildings. The system eventually will entail the use of a large number of transceivers (a device comprised of both a transmitter and a receiver). Two types of transceivers are being considered. In life testing, 200 transceivers (randomly selected) were tested for each type. Transceiver 1: 20 failures were observed (out of 200) Transceiver 2: 14 failures were observed (out of 200). The engineers want to test for the equality of the proportions. Define p_1 (p_2) to be the population proportion of Transceiver 1 (Transceiver 2) failures.

   (a) State the null and alternative hypotheses.
   Let p_1 denote the proportion of failures for Transceiver 1 and p_2 denote the proportion of failures for Transceiver 2, we want to test $H_0 : p_1 = p_2$ against $H_a : p_1 \neq p_2$.

   (b) Calculate the test statistic to 4 decimal places.
   First note that the common proportion estimate under H_0 is $\hat{p}_0 = (14 + 20)/(200 + 200)$. The test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{20/200 - 14/200}{\sqrt{34/400(1-34/400)\left(\frac{1}{200} + \frac{1}{200}\right)}} = 1.076$$

(c) Calculate the p-value.
The p-value is:

$$2P(Z < -1.076) = 0.2819$$

(d) What is your decision and conclusion at $\alpha = 0.05$?
Since p-value is greater than α, we do not reject H_0. We do not have sufficient evidence to conclude that the two proportions are different.

(e) Use `prop.test` in `R` to check your work.

```
>  prop.test(c(20,14),c(200,200),correct=F)

        2-sample test for equality of proportions without continuity
        correction

data:  c(20, 14) out of c(200, 200)
X-squared = 1.1572, df = 1, p-value = 0.2821
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02458069  0.08458069
sample estimates:
prop 1 prop 2
  0.10   0.07
```