# Handout 5. One Sample Tests

## 1. One sample *t*-test

The t tests are based on an assumption that data come from the normal distribution. In the one-sample case we assume that data x1, . . . , xn are normal random variables with mean u and variance sigma. We wish to test the null hypothesis that H0:u=u0.

Consider an example concerning daily energy intake in kJ for 11 women (Altman, 1991, p. 183). First, the values are placed in a data vector:

```
> daily.intake <- c(5260,5470,5640,6180,6390,6515,6805,7515,7515,
8230,8770)
> mean(daily.intake)
> sd(daily.intake)
> quantile(daily.intake)
> res<-t.test(daily.intake,mu=7725)
> names(res)
[1] "statistic"   "parameter"   "p.value"   "conf.int"   "estimate"
[6] "null.value"  "alternative" "method"    "data.name"
> res$para
> res$conf.int
> res$statistic
> res$p.value
> res$method
```

**Example 1:**

The seven scores listed below are axial loads (in pounds) for a random sample of 7 12-oz aluminum cans manufactured by ALUMCO. An axial load of a can is the maximum weight supported by its sides, and it must be greater than 165 pounds, because that is the maximum pressure applied when the top lid is pressed into place.

*270, 273, 258, 204, 254, 228, 282*

**(1)** As the quality control manager, please test the claim of the engineering supervisor that the average axial load is greater than 165 pounds. Use $\alpha = 0.05$. What assumptions are needed for your test?

***Sol)***

**(1)** $\begin{cases} H_0 : \mu = 165 \\ H_a : \mu \succ 165 \end{cases}$

Assume the distribution is normal.

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$$

At the significance level of $\alpha = 0.05$, we reject $H_0$ in favor of $H_a$ if $T_0 \geq t_{n-1,\alpha}$

$$T_0 = \frac{87.7}{27.6/\sqrt{7}} \doteq 8.9 \succ 1.943$$ : We reject $H_0$

$$\text{CI}: \ \overline{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$$

```
data_cans<-c(270,273,258,204,254,228,282) #input the data#
data_cans<-data_cans-165;
data_cans
#by minusing 165, we just need to estimate if the new data greater than 0 significantly#
t.test(data_cans,alternative = "greater")
```

#Since we assume the samples comes from normal population, we will not test their normality and use t.test directly#

**Example 2:**

**Over then past 5 years, the mean time for a warehouse to fill a buyer's order has been 25 minutes. Officials of the company believe that the length of time has increased recently, either due to a change in the workforce or due to a change in customer purchasing policies. The processing time (in minutes) was recorded for a random sample of 15 orders processed over the past month.**

**28 25 27 31 10 26 30 15 55 12 24 32 28 42 38**

**Questions:**

**(a). Please check the normality of the data.**

**(b). Please test the research hypothesis at the significance level α = 0.05.**

(a)

```
Exercise<-c(28,25,27,31,10,26,30,15,55,12,24,32,28,42,38)
shapiro.test(Exercise)
```

#p-value = 0.4038, can NOT reject Ho (data are normal), so based on the data we claim the data do appear to follow the normal distribution#

(b)

```
t.test(Exercise, mu=25, alternative="greater")
```

#p-value = 0.1485, can NOT reject Ho (the mean time for a warehouse to fill a buyer's order has been 25 minutes or less), so based on the data we have, we can NOT claim the length of time has increased#

## 2. Wilcoxon signed-rank test

The t tests are fairly robust against departures from the normal distribution especially in larger samples, but sometimes you wish to avoid making that assumption. To this end, the distribution-free methods are convenient.

For the one-sample Wilcoxon test, the procedure is to subtract the theoretical u0 and rank the differences according to their numerical value, ignoring the sign, and then calculate the sum of the positive or negative ranks. The point is that, assuming only that the distribution is symmetric around u0, the test statistic corresponds to selecting each number from 1 to n with probability 1/2 and calculating the sum. The distribution of the test statistic can be calculated exactly, at least in principle. It becomes computationally excessive in large samples, but the distribution is then very well approximated by a normal distribution.

```
> res2<-wilcox.test(daily.intake,  mu=7725)
> names(res2)
[1] "statistic"  "parameter"  "p.value"  "null.value"  "alternative"
[6] "method"       "data.name"
```