9 One sample tests

This week we look at two tests based on one sample problems.

9.1 One sample test for a binomial proportion (P.129-132)

A test for a binomial proportion will be very useful in many life science problems. For example, an agency decides to make a decision based on the current unemployment rate or interest rate for future planning.

Motivating example: Suppose that the agency is interested to check whether the current unemployment rate in 0.05 (or 5%). In this case the null hypothesis of interest is $H_0 : p = 0.05$ (a pre-specified rate) for the population proportion p. Many authors prefer to denote this known proportion as p_0 or write $p_0 = 0.05$. Therefore, in general, $H_0 : p = p_0$ and a suitable test is related to the binomial distribution.

In practice, suppose that we take a random sample of size n. Let X be the count of a certain event of interest (e.g. unemployment rate) in n independent and identical trials. Clearly, the distribution of X is binomial or $X \sim B(n, p)$.

Recall that the mean and variance of X are, $E(X) = np$ and $Var(X) = np(1 - p)$. By the CLT, we have

$$\hat{p} = \frac{X}{n} \overset{\text{approx.}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

provided n is large.

Under the null hypothesis, $H_0 : p = p_0$, we have the following standardized distribution:

$$Z_{\text{obs}} = Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1).$$

Note: The standard error (SE) of $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}}$$

and the hypothesized proportion p_0 under H_0 is used in the calculation of SE.

Example: (One-sided test) *The Big Egg* manufacturer claims that *no more than* 3% of the eggs it sends out are less than 60g. The supermarket *WWW.com* impressed with this claim, has purchased a large quantity of such egg cartons. To determine if the manufacturer's claim is correct, a sample of 300 eggs was selected and found that 12 of these eggs weigh less than 60g. Should the manufacturer's claim be rejected?

Solution: Let X be the number of underweight eggs in a sample of 300 and p be the true proportion of underweight eggs.

1. Hypotheses: The null hypothesis is

____

An estimate of p is

$$\hat{p} = \frac{X}{n} = \_\_\_\_\_\_\_\_\_\_$$

Hence the supermarket wants to make a claim of whether the true proportion of underweight eggs is actually *more than 0.03*, i.e. against the manufacturer's claim of *no more than 0.03*. Therefore the test is one-sided and the alternative hypothesis is

____

2. Test statistic: Under H_0,

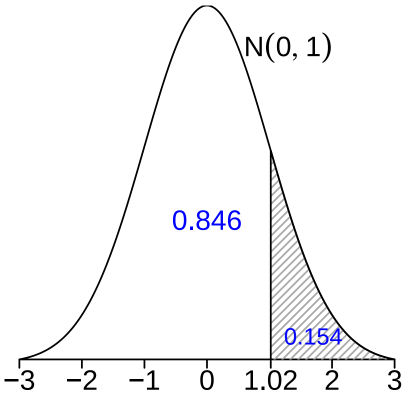
$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \_\_\_\_\_\_\_\_\_\_\_\_$$

Note: (a) The hypothesized proportion $p_0 = 0.03$ under H_0 is used in the calculation of SE.

(b) Assumption: n is large so that Z is normal by CLT.

3. P-value: The P-value is

$P(X \geq 12) =$ ____



4. Conclusion: ____
__________ In other words the manufacturer's claim is reasonable.

An Alternative Solution

2. Test statistic: Under H_0, ________________

3. P-value (exact):

$$P(X \geq 12) =$$

____

which is difficult to evaluate manually but can be easily obtained by the R command:

```
1-pbinom(11,300,0.03)
```

The answer is 0.1940 and the conclusion is the same.

Remark: This P-value is *exact* because CLT is not applied to approximate the distribution of sample proportion $\hat{p}$ by normal.

Instead, the binomial distribution for the sample total X is used. Hence this method can be applied even if the sample size n is less than 30.

Example: (2-sided test) An antibiotic producer claims that 8% of patients taking the drug have mild side effects. A hospital administered the antibiotic treatment to 200 patients and found 24 of them have mild side effects. Does this finding indicate that the side effects rate *has changed* at the 5% significance level?

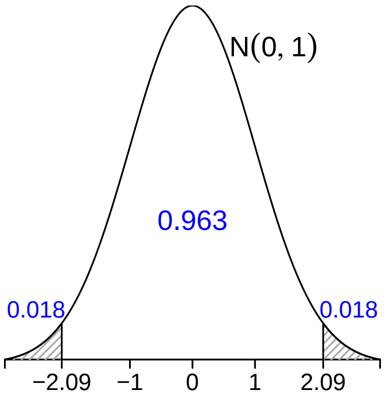
Solution: Let p be the true proportion of patients who have mild side effect. To test whether there is a change of percentage, the alternative hypothesis should be 2-sided. The one-sample Z-test for the population proportion is

1. **Hypotheses:** ______________________________
2. **Test statistic:** $Z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _______________

   where an estimate of p is $\hat{p} = \dfrac{X}{n} =$ ___________
3. **P-value:** $2P(\hat{p} > 0.12) =$ ___________________________

   ____
4. **Conclusion:** _____________________________________

   __________________________________. Therefore, the side effects rate is significantly different from 8%.



Exercise: Test whether the side effects rate has *increased* using the above data. That is, test $H_0 : p = 0.08$ against $H_1 : p > 0.08$.

Ans: Since the p-value $= P(Z \geq |2.09|) = 0.0183 < 0.05$, there is very strong evidence in the data against H_0. The side effects rate has significantly increased from 8%.

Note: Since the one-sided p-value is even smaller, the test indicates even *stronger evidence* against H_0.

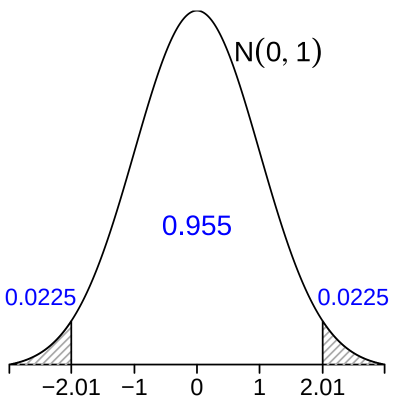
Example: (Cardiovascular Disease; Q7.10-7.11 (P.137)) Suppose that the incidence rate of MI (myocardial infection) per year was 5 per 1000 among 45 to 54 year old men in 1990. To study changes in incidence over time, 5000 men between 45 to 54 year old were followed for one year starting in 2000. Fifteen new cases of MI were found. Test the hypothesis that incidence rates of MI changed from 1990 to 2000.

Solution: The two-sided Z-test for the population proportion is

1. **Hypotheses:** ______________________________
2. **Test statistic:** $Z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ ____________________

   where an estimate of p is $\hat{p} = \dfrac{X}{n} =$ ____________
3. P**-value:** $2P(\hat{p} < 0.003) =$ ______________________
4. **Conclusion:** ____

   ____________________________. The incidence rates of MI changed from 1990 to 2000.



Exercises: Book P.134 Q6 and Q10-11.

Now we look at another important hypothesis testing problem that arise in life sciences.

9.2 Matched pairs t-test (P.139-143)

Previous hypothesis tests for the population mean μ use the information provided by a single sample. This section considers an extension by comparing pairs.

Motivating example: A medical researcher wants to compare the effects of two treatments. The best way to get satisfactory results from this experiment is to give both treatments to the same group of people with a long time gap (e.g. Monday treatment I and Tuesday treatment II), and repeat this for a number of individuals. In this way any effects other than treatment, e.g. personal characteristic, can be eliminated.

Definition: Data consisting of two observations from each (or identical) individual are called *paired data* or *matched pairs*. An experiment with this kind of design is called *a matched pairs design*.

We develop a t-test called *the paired t-test for paired observations*. Note that the *paired t-test* can be reduced to a one sample t-test.

Remark: By taking two measurements on a single (or two similar) individual(s), we can reduce any extra variations in sampling due to other factors not being studied or controlled.

Example: A large pharmaceutical company wants to compare the effect of a new drug (A) for chronic insomnia against an existing drug (B). The best way to do this test is as follows:

Week 1:
Select a group of individuals suffering from chronic insomnia.

Give the old drug and record the number of hours they sleep.

Week 2:
Give the new drug and record the number of hours they sleep.

Suppose that the above study consists of 6 people with chronic insomnia and are given the new medication (A) in week 1 and the old drug (B) in week 2. Their hours of sleep each night are recorded. The data are:

Person	1	2	3	4	5	6
New drug A	4.8	4.1	5.8	4.9	5.3	7.4
Old drug B	3.9	4.2	5.0	4.9	5.4	7.1

The company wants to test whether the *new drug is more effective.*

Solution: Let μ_A and μ_B be the mean sleeping times with drug A and B respectively.

If the two drugs are equally effective, then there is no difference in sleeping times. Therefore,

$H_0 : \mu_A = \mu_B$, or equivalently, $H_0 : \mu_d = 0$ where $\mu_d = \mu_A - \mu_B$.

If drug A is more effective than drug B, then A gives a longer sleeping time than the drug B and therefore,

$H_1 : \mu_A > \mu_B$, or equivalently, $H_1 : \mu_d > 0$.

Then, to compare the treatments, we compute the differences between new-old (or before-after) treatments, $A - B$ in hours of sleep. Suppose that $d_1, d_2, \cdots, d_6$ are the differences. Then

$d_1 = 0.9,\ d_2 = -0.1,\ d_3 = 0.8,\ d_4 = 0.0,\ d_5 = -0.1,\ d_6 = 0.3$

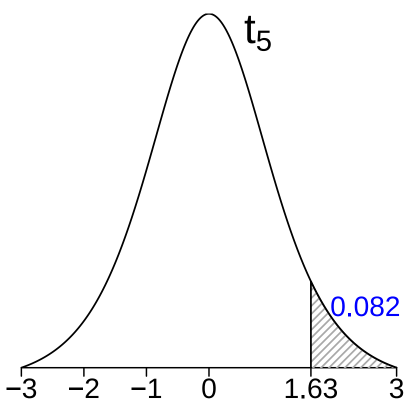
and the mean and sd of these differences d_i are

$$\bar{d} = 0.3 \quad \text{and} \quad s_d = 0.452.$$

With these, we can conduct a *one-sample* t-test on the mean difference μ_d using the sample of differences d_i:

1. **Hypotheses:** ____________________
2. **Test statistic:** $t_0 = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}} =$ ______________
3. P**-value:** $P(\bar{d} > 0.3) =$ ________________
4. **Conclusion:** ________________________________ ________________. In other words, the new drug has no significant effect on sleeping.

Note: The paired t-test is equivalent to the one sample t-test on d_i. It requires the assumption of normally distributed d_i and the population sd σ_d being unknown.



Example: An experiment is conducted involving 20 Year 11 students divided into 10 pairs based their Year 10 Mathematics marks. One member of each pair is randomly chosen and taught calculus using method A. The other member is taught using method B. Test scores for each pair are recorded and the following differences are calculated:

$d = A - B$: 9, -2, -5, 4, -1, 8, 0, -6, 10, -5

Do these data suggest that the mean scores for the two methods are *significantly different*?

Solution: Let μ_d be the potential mean difference in the population of interest.

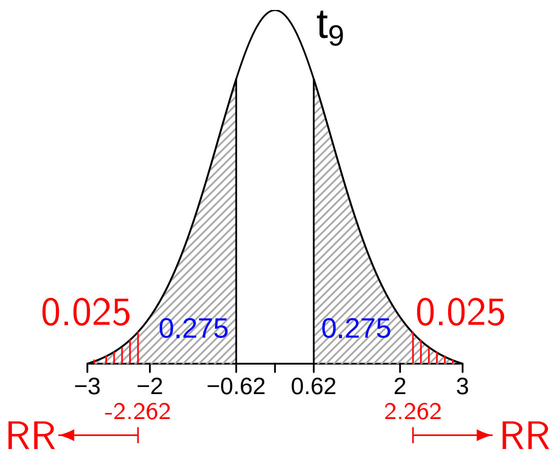
Why a *paired t-test* should be considered?
Each pair has similar Year 10 Mathematics marks. Moreover, the sample size $n = 10$ is small and true sd σ is unknown.

Why a *two-sided* test should be conducted?
The objective is to detect whether there is significant difference in mean scores.

We calculate $\bar{d} = 1.2$ and $s_d = 6.125$ (check these values using your calculator). Then the one sample t-test for the mean difference μ_d is

1. **Hypotheses:** ______________________________
2. **Test statistic:** $t_0 = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}} =$ __________________
3. P**-value:** $2P(\bar{d} \geq 1.2) =$ ______________________________

4. **Conclusion:** ____ ______________________________. There is no significant difference in the mean scores for the two methods.



9.3 Relationship between P-value and critical value

In the previous example, the data are consistent with H_0 because

$$\text{P-value} > 0.05.$$

This is equivalent to the test statistic

$$t_{\text{obs}} = 0.62 < 2.262 = T_{9,0.025}$$

where $T_{9,0.025} = 2.262$ is called the *critical value* which is the 0.975 quantile of t_9. Since

the larger is the absolute value of the test statistic t_{obs},
the smaller is its upper area or P-value which implies
the stronger is the evidence against H_0.

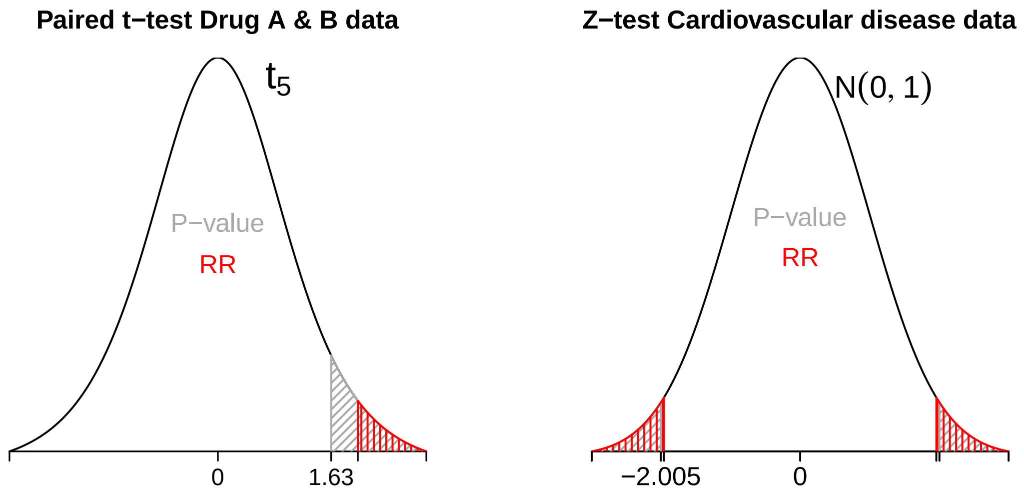
Hence decision on H_0 can be made alternatively by comparing the test statistic with critical value as summarized below:

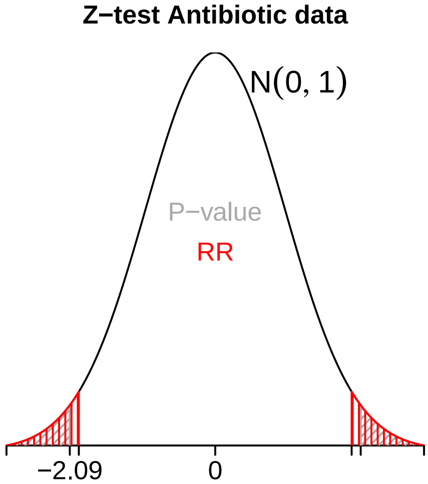
For	$H_1 : \mu < \mu_0$	Evidence against H_0 if P-value $< \alpha \Leftrightarrow t_{\text{obs}} < -T_{n-1,\alpha}$
For	$H_1 : \mu > \mu_0$	Evidence against H_0 if P-value $< \alpha \Leftrightarrow t_{\text{obs}} > T_{n-1,\alpha}$
For	$H_1 : \mu \neq \mu_0$	Evidence against H_0 if P-value $< \alpha \Leftrightarrow \lvert t_{\text{obs}} \rvert > T_{n-1,\alpha/2}$

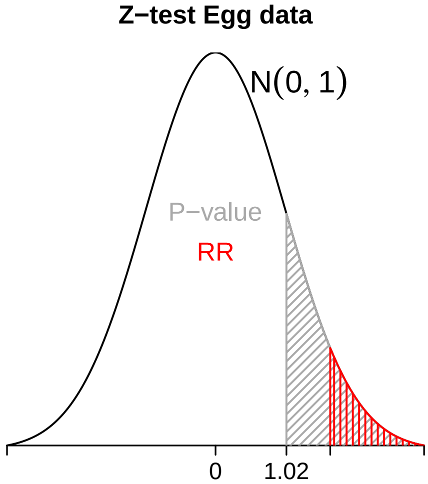
Note: Regions such as $t_{obs} < -T_{n-1,\alpha}$ are called *rejection regions* (RR).

Previous examples

Paired t-test: *(Drug A&B)*	$t_o = 1.63 < t_{5,0.05} = 2.015$	$\Leftrightarrow$ Consistent with H_0
Z-test: *(Cardi. dis.)*	$Z_o = \lvert -2.005 \rvert > Z_{0.975} = 1.96$	$\Leftrightarrow$ Against H_0
Z-test: *(Antibiotic)*	$Z_o = 2.09 > Z_{0.975} = 1.96$	$\Leftrightarrow$ Against H_0
Z-test: *(Egg)*	$Z_o = 1.02 < Z_{0.95} = 1.645$	$\Leftrightarrow$ Consistent with H_0







Read P.142 to 148 and example 8.6 on P.148.

Remarks: The same idea of constructing the $(1-\alpha)\%$ CI can be applied to μ_d based on sample mean $\bar{d}$ and sd s_d.

Exercises: Q8.1-3 (P.152) and Q8.31-36 (P.153)

Short answers:

Q8.1: $\bar{d} = 20$, $s_d = 35$, $t_{19} = 2.556$, P-value is < 0.05. Evidence against H_0.

Q8.2: $\bar{d} = 3$, $s_d = 12$, $t_{19} = 1.118$, P-value is > 0.05. Data are consistent with H_0.

Q8.3: $\bar{d} = 5.2$, $s_d = 8$, $t_{19} = 2.907$, P-value is < 0.05. Evidence against H_0.

Q8.31-36: $t_9 = 3.674$, P-value is < 0.05. Evidence against H_0. 95% CI for μ_d: (1.3835, 5.8165). Since 0 lies outside the CI, there is evidence against $H_0 : \mu_d = 0$.