

# Chapter 1

## Hypothesis Test

A report says the average yearly income of Chinese people is 74,318¥, but you suspect the average income is less than reported. Is your suspicion correct? In this chapter, we will learn how to solve this problem through **hypothesis test**.

## 1.1 Basics of Hypothesis Test

In order to test whether a given coin is fair or not, it is tossed 100 times, and resulted in 45 heads. Does this provides convincing evidence that this coin is not fair?

### • Hypotheses

In the above setting, we have two contending **hypotheses** about the parameter  $p$ , where  $p$  is the probability of getting a head when this coin is tossed. One hypothesis is  $p = \frac{1}{2}$ , which means the coin is fair. The other one is  $p \neq \frac{1}{2}$ , which means the coin is not fair. The first hypothesis is called **null hypothesis**, denoted as  $\mathbf{H}_0$ , and the second one is called **alternative hypothesis**, denoted as  $\mathbf{H}_a$ . They are written as

$$\begin{aligned}\mathbf{H}_0 : \quad p &= \frac{1}{2} \\ \mathbf{H}_a : \quad p &\neq \frac{1}{2}\end{aligned}$$

#### Remark:

- Usually, the  $\mathbf{H}_a$  is the hypothesis we want to seek evidence for.
- The hypotheses are about parameters. You can not set statements like  $\hat{p} = \frac{1}{2}$ ,  $\bar{x} = 0.6$  as hypotheses.

The **hypothesis test** is the procedure to make decisions between the two hypotheses.

In the above case, if  $\mathbf{H}_a$  is true, it can either be  $p < \frac{1}{2}$  or  $p > \frac{1}{2}$ . This type of *hypotheses test* is called **two-sided hypotheses test**.

If the hypotheses are set as either of the following

$$\begin{aligned}\mathbf{H}_0 : \quad p &= \frac{1}{2}, & \mathbf{H}_a : \quad p > \frac{1}{2} \\ \mathbf{H}_0 : \quad p &= \frac{1}{2}, & \mathbf{H}_a : \quad p < \frac{1}{2}\end{aligned}$$

the *hypotheses test* is called **one-sided hypotheses test**.

### Juicy Pineapples

At the Hawaii Pineapple Company, managers are interested in the size of the pineapples grown in the company's fields. Last year, the mean weight of the pineapples harvested from one large field was 31 ounces. A different irrigation system was installed in this field after the growing season. Managers wonder how this change will affect the mean weight of pineapples grown in the field this year.

- (a) State appropriate hypotheses for performing a significance test. Be sure to define the parameter of interest.
- (b) Is this a **one-sided** or **two-sided** hypotheses test?

- The reasoning of hypotheses test

Let's use the “fair coin” setting and the hypotheses are

$$\mathbf{H}_0 : \quad p = \frac{1}{2}$$

$$\mathbf{H}_a : \quad p \neq \frac{1}{2}$$

Among the 100 tosses, 45 are head.

How to choose between those two hypotheses? The reasoning is like: if we assume the *null hypotheses* is true, and we get a sample with the sample proportion  $\hat{p}$  far away from  $p = \frac{1}{2}$ , then we have convincing evidence for  $H_a$  and reject  $H_0$ .

**In the following,  $H_0$  is assumed to be true, with  $p = 1/2$ !!!**

How to tell whether  $\hat{p}$  is far away from  $p$  or not? We have to resort to the sampling distribution of  $\hat{p}$  when assuming  $H_0$  is true. According to what we learned before, if certain conditions are met,  $\hat{p}$  is approximately normally distributed.

$$\hat{p} \sim N(\mu_{\hat{p}}, \sigma_{\hat{p}}), \quad \mu_{\hat{p}} = p, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

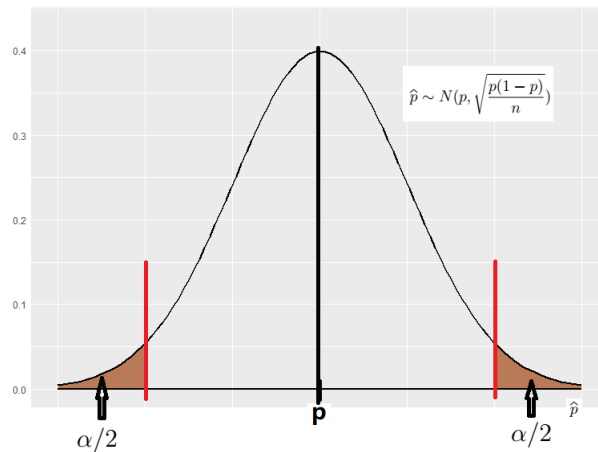


Figure 1.1:  $\hat{p}$  that is far away from  $p$

In figure1.1, we draw two lines away from  $p$ . If a  $\hat{p}$  falls to the outside of those two lines, indicated by the shaded areas, we say this  $\hat{p}$  is far away from  $p$ .

However, where to draw those two lines? Referring to figure1.1, if the total area of the “shaded areas” are given, then the position of the two lines are decided. We call this given total area **significance level**  $\alpha$ . The most frequently used *significance level* is  $\alpha = 0.05$ .

Set significance level  $\alpha = 0.05$ .  
Is the  $\hat{p}$  far away from  $p$  in the case of “fair coin” with

$$H_0 = 1/2$$

$$H_a \neq 1/2$$

In the above *two-sided hypotheses*, when  $\hat{p}$  is either way larger or way smaller than  $p$ , we say  $\hat{p}$  is far away from  $p$  in the direction of  $H_a$ . Because, either “way larger” or “way smaller” is in favour of  $H_a$ .

How to define “far away” in the case of *one-sided hypotheses test*? Use the “fair coin” example, and change the hypotheses into

$$\mathbf{H}_0 : p = \frac{1}{2}$$

$$\mathbf{H}_a : p \leq \frac{1}{2}$$

Since we have to decide whether  $\hat{p}$  is far away from  $p$  in the direction of  $H_a$ , we have to draw a borderline to the left side of  $p$ , as shown in figure 1.2.

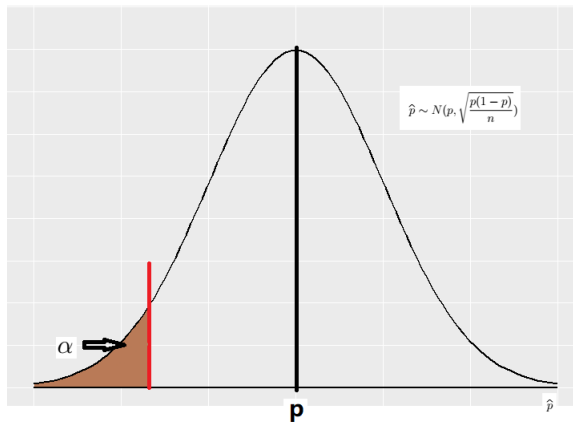


Figure 1.2: One-sided hypotheses test

Set significance level  $\alpha = 0.05$ . Is the  $\hat{p}$  far away from  $p$  in the case of “fair coin” with

$$H_0 = 1/2$$

$$H_a \leq 1/2$$

Here, the *significance level*  $\alpha$  is the area to the left side of the borderline demarcating whether  $\hat{p}$  is far away from  $p$  or not.

In the case of *one-sided hypotheses test* as

$$\mathbf{H}_0 : p = \frac{1}{2}$$

$$\mathbf{H}_a : p \leq \frac{1}{2}$$

everything is argued similarly.

**Remark:**

- Significance level  $\alpha$  is set before the sample is drawn.
- The hypotheses is set before the sample is drawn.
- All the above are based on the assumption that  $H_0$  is true.

### • The P-value

Still use the “fair coin” example, and the hypotheses are

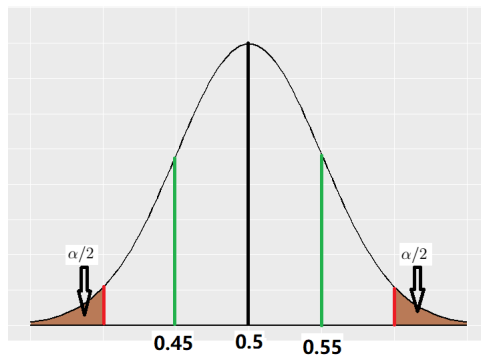
$$\mathbf{H}_0 : p = 0.5$$

$$\mathbf{H}_a : p \neq 0.5$$

The sample proportion  $\hat{p}_0 = 45/100 = 0.45$ .

Assume  $H_0$  is correct.

$$\mathbf{P}(|\hat{p} - 0.5| > |\hat{p}_0 - 0.5|) = \mathbf{P}(\hat{p} > 0.55) + \mathbf{P}(\hat{p} < 0.45)$$



Calculate the P-value of the “fair coin” example with hypotheses

$$\mathbf{H}_0 : p = 0.5$$

$$\mathbf{H}_a : p \neq 0.5.$$

Figure 1.3: P-value for two-sided hypotheses test

Reading figure1.3, we know that if  $\mathbf{P}(|\hat{p} - 0.5| \geq |\hat{p}_0 - 0.5|) \geq \alpha$ , then the sample proportion  $\hat{p}_0$  is not far away from  $H_0$  in the direction of  $H_a$  under the significance level  $\alpha$ . We fail to reject  $H_0$ , and don’t have convincing evidence for  $H_a$ . If  $\mathbf{P}(|\hat{p} - 0.5| \geq |\hat{p}_0 - 0.5|) < \alpha$ , then the sample proportion  $\hat{p}_0$  is far away from  $H_0$  in the direction of  $H_a$  under the significance level  $\alpha$ . We reject  $H_0$  and have convincing evidence for  $H_a$ .

In the above *two-sided hypotheses test*,  $\mathbf{P}(|\hat{p} - 0.5| \geq |\hat{p}_0 - 0.5|)$  plays an important role in making conclusions. It is called the **P-value** of the test.

$\mathbf{P}(|\hat{p} - 0.5| \geq |\hat{p}_0 - 0.5|)$  is the probability of getting a  $\hat{p}$  as extreme as or more extreme than the observed  $\hat{p} - 0$  in the direction of  $H_a$ . A more formal definition of *P-value* is given below.

**P-value** is the probability, computed assuming  $H_0$  is true, that the statistic(such as  $\hat{p}$  or  $\bar{x}$ ) would be as extreme as or more extreme than the observed value, in the direction of  $H_a$ .

Sometimes you are asked to interpret the *P-value*. You have to interpret in the context. In the above example we should interpret the *P-value* as “Assuming the probability of getting a head is 0.5, the probability of tossing 100 times with the relative frequency of heads less than or equal to 0.45, or larger than or equal to 0.55 is ‘*P-value*’ ”.

In the “fair coin” example, if the hypotheses test is a *one-sided hypotheses test* with hypotheses

$$H_0 : p = 0.5$$

$$H_a : p \leq 0.5$$

The *P-value* is indicated in figure 1.4.

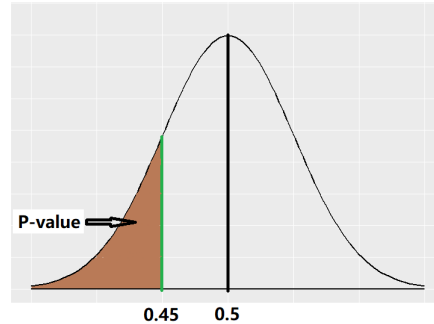


Figure 1.4: P-value for *one-sided hypotheses test*

The characteristics and the definition of *P-value* of *one-sided hypotheses test* is the same as of *two-sided hypotheses test*.

### Healthy Bones

Calcium is a vital nutrient for healthy bones and teeth. The National Institutes of Health (NIH) recommends a calcium intake of 1300 mg per day for teenagers. The NIH is concerned that teenagers aren't getting enough calcium. Is this true?

Researchers want to perform a test of

$$H_0 : \mu = 1300; \quad H_a : \mu < 1300$$

where  $\mu$  is the true mean daily calcium intake in the population of teenagers. They ask a random sample of 20 teens to record their food and drink consumption for 1 day. The researchers then compute the calcium intake for each student. Data analysis reveals that  $\bar{x} = 1198\text{mg}$  and  $s_x = 441\text{ mg}$ . After checking that conditions were met, researchers performed a significance test and obtained a P-value of 0.1404.

- Explain what it would mean for the null hypothesis to be true in this setting.
- Interpret the P-value in context.

- **P-value and conclusions**

If the *P-value* is smaller than the *significance level*  $\alpha$ , we say the results of the study are **statistically significant at level**  $\alpha$ . In this case, we reject the null hypotheses  $H_0$  and conclude that there is convincing evidence for  $H_a$ .

In nutshell, the relation between P-value and conclusions comes down to

P-value  $< \alpha \implies$  reject  $H_0 \implies$  convincing evidence for  $H_a$

P-value  $> \alpha \implies$  fail to reject  $H_0 \implies$  no convincing evidence for  $H_a$

- **Test statistic**

In the “fair coin” example, the P-value of the *two-sided hypotheses test* is given by

$$\mathbf{P}(|\hat{p} - 0.5| > |\hat{p}_0 - 0.5|), \quad \hat{p}_0 = 0.45$$

If conditions are met,  $\hat{p}$  follows a normal distribution.

$$\hat{p} \sim N(\mu_{\hat{p}}, \sigma_{\hat{p}}), \quad \mu_{\hat{p}} = p = 0.5, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = 0.05$$

$$\mathbf{P}(|\hat{p} - 0.5| > |\hat{p}_0 - 0.5|) = \mathbf{P}\left(\left|\frac{\hat{p} - 0.5}{\sigma_{\hat{p}}}\right| > \left|\frac{\hat{p}_0 - 0.5}{\sigma_{\hat{p}}}\right|\right) = \mathbf{P}(|z_{\hat{p}}| > |z_{\hat{p}_0}|)$$

where  $z_{\hat{p}} \sim N(0, 1)$  and  $z_{\hat{p}_0} = -1$ .

$$\mathbf{P}\text{-value} = \mathbf{P}(|z_{\hat{p}}| > 1)$$

It is very easy to calculate the P-value by referring to the standard normal distribution.

In the above calculation,  $z_{\hat{p}_0}$  is called the **test statistic**.

The general formula for *test statistics* is

$$\mathbf{test\ statistic} = \frac{\mathbf{statistic-parameter}}{\mathbf{standard\ deviation\ of\ the\ statistic}}$$

*Test statistic* plays an important role in calculating the *P-value*. In this book, we learn three type of test statistics: **z**, **t**  $\chi^2$ . They are corresponding to normal distribution, t distribution, and  $\chi^2$  distribution. Figure1.5 gives the P-value for statistic  $z$  in different tests.

## 1.2 One sample z test for population proportion

The setting is: The alleged population proportion of a population is  $p$ . You suspect the real population proportion might be different from, or less than, or more than  $p$ . You draw an SRS, or design a study, with sample size  $n$  and sample proportion  $\hat{p}$ , and perform a hypotheses test on basis of the results.

- **The test statistic  $z$  and the P-value**

Suppose all conditions are met and the null hypotheses is  $H_0 = p_0$ , and the sample proportion is  $\hat{p}$ , and the sample size is  $n$ , then the test statistic  $z$  is given by

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

The P-values are shown by the shaded area in figure 1.5.

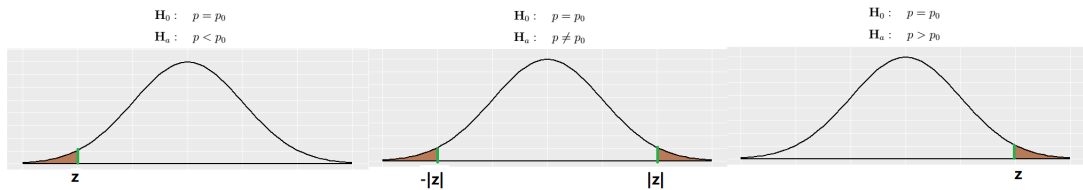


Figure 1.5: P-value in different cases for statistic  $z$

- **General steps**

- State correct hypotheses and set the significance level
- Identify correct test and check conditions
- Calculate P-value: shall report test statistic, degree of freedom, etc.
- Conclude

**Remark:**

All the previous formulas about hypotheses test is built on the prerequisite that the sampling distribution is a normal distribution, thus the conditions to be checked is exactly the same as the corresponding confidence intervals.



### • Example

A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have blemishes, the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips.

A supervisor selects a random sample of 500 potatoes from the truck. An inspection reveals that 47 of the potatoes have blemishes. Carry out a significance test at the  $\alpha = 0.05$  significance level. What should the chi producer conclude?

#### Solution:

##### – Set up hypotheses

$$\mathbf{H}_0 : p = 0.08$$

$$\mathbf{H}_0 : p > 0.08$$

where  $p$  is the percentage of potatoes in the shipment have blemishes.

##### – We will perform a one-sample z test for $p$ if conditions are met.

\* **Random:** the sample is a random sample by the stem of the problem.

· **10% condition** The sample size 500 is less than 10% of the total truck load of potatoes

\* **Large counts condition:**

$$np = 500 \cdot 0.08 = 40 \geq 10, \quad n(1 - p) = 500 \cdot (1 - 0.08) = 460 \geq 10$$

Large counts condition is met.

##### – Calculate P-value

$$\text{The test statistic } z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \left( \frac{47}{500} - 0.08 \right) \bigg/ \sqrt{\frac{0.08(1 - 0.08)}{500}} = 1.15$$

The P-value  $\mathbf{P}(z \geq 1.15)$  given by command `normalcdf(lower:1.15, upper:  $\infty$ ,  $\mu = 0$ ,  $\sigma = 1$ )` is 0.1251.

##### – Conclusion

Since the P-value is 0.1251 larger than the significance level 0.05, we fail to reject the null hypotheses, and don't have convincing evidence that the percentage of potatoes in the shipment have blemishes if larger than 8%. The producer should use the potatoes to make chips.

When checking the *large counts condition*, why do we use  $np$  and  $n(1 - p)$  instead of  $n\hat{p}$  and  $n(1 - \hat{p})$ ?

### Teen Drivers

A states Division of Motor Vehicles (DMV) claims that 60% of teens pass their driving test on the first attempt. An investigative reporter examines an SRS of the DMV records for 125 teens; 86 of them passed the test on their first try. Is there convincing evidence at the  $\alpha = 0.05$  significance level that the DMVs claim is incorrect?

### 1.3 Type I error and type II error

Let take a look at a *one-sided hypotheses test* with hypotheses

$$\mathbf{H}_0 : p = p_0$$

$$\mathbf{H}_a : p < p_0$$

The significance level is  $\alpha$ . The basic idea of hypotheses test is, assuming  $\mathbf{H}_0$  is true, if we draw a sample and the sample statistic is too far away from the parameter given by  $\mathbf{H}_0$  in the direction of  $\mathbf{H}_a$ , then we reject  $\mathbf{H}_0$ , have convincing evidence for  $\mathbf{H}_a$ .

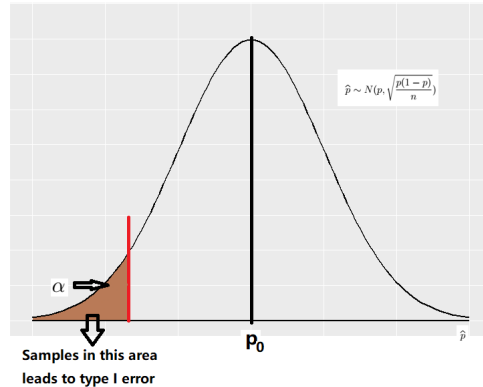


Figure 1.6: Type I error equals significance level  $\alpha$

As shown in figure1.6, if  $\mathbf{H}_0$  is really true, there are still some samples in the shaded area with P-value less than significance level  $\alpha$ , and lead to a wrong conclusion by rejecting  $\mathbf{H}_0$ . The probability of making this type of wrong conclusion equals significance level  $\alpha$ .

**Type I error** is the error that we reject  $\mathbf{H}_0$  when  $\mathbf{H}_0$  is true.

The probability of making type I error equals the significance level  $\alpha$ .

If we set  $\alpha$  at a low level, such as 0.05, and we come to the conclusion that we have convincing evidence for  $\mathbf{H}_a$ , then we are “pretty sure” about this conclusion, because the probability of making mistake is 0.05. This explains why we set the conclusion we are interested in as the  $\mathbf{H}_a$ . Why?

Since the probability of *type I error* is totally controlled by the significance level  $\alpha$ , why don't we set  $\alpha$  as 0, or some value extremely small? Let's take a look at the other type of error before answering this question.

Consider the same one-sided hypotheses test as above

$$\mathbf{H}_0 : p = p_0$$

$$\mathbf{H}_a : p < p_0$$

The P-value is calculated while assuming  $\mathbf{H}_0$  is correct. As show in the first sampling distribution in figure1.7. Any sample with sample proportion  $\hat{p}$  falls to the right side of the vertical line can leads to the conclusion of “fail to reject  $\mathbf{H}_0$ ”. If the truth is  $p = p_1 < p_0$ , then we will make a mistake, this type of mistake is called **type II error**.

Refer to the “potato chip” example in section 2. State the **type I error** and it's consequence.

**Type II error** is the error of failing to reject  $H_0$  while  $H_0$  is not correct.

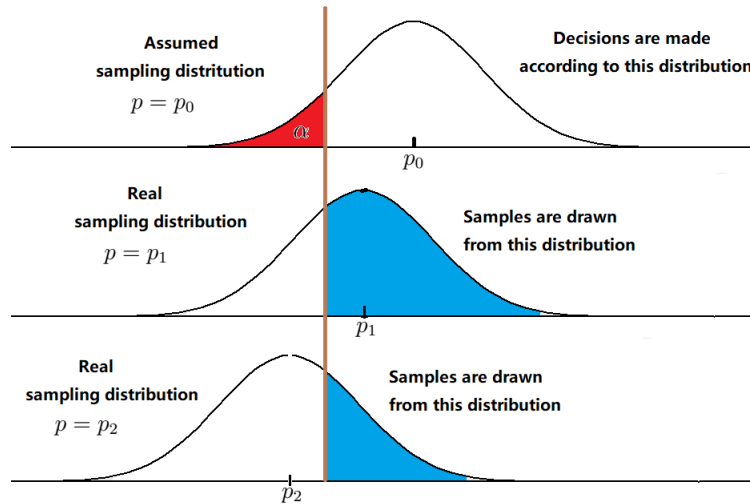


Figure 1.7: Type II error

The standard of whether to reject or fail to reject  $H_0$  comes from the distribution while assuming  $p = p_0$ . However, the samples are drawn from the true sampling distribution with  $p = p_1$ . As shown in the second sampling distribution in figure 1.7, any sample falls in the shaded area would have P-value larger than the significance level  $\alpha$ , thus leads to *type II error*.

Can you find out the areas that represent  $\beta$  and the power in figure 1.7?

The probability of making *type II error* is denoted as  $\beta$ .  $1 - \beta$  is called the **power** of the test.

How to increase the *power* of a test?

- **Increase the significance level  $\alpha$ .** This is indicated in figure 1.7. It means there is a trade off between the probability of type I error and the probability of type II error.
- **Increase the difference between the null and alternative parameter values that is important to detect.** This is shown in figure 1.7, when the true value of  $p$  takes  $p_1$  and  $p_2$  separately.
- **Increase the sample size.** Suppose the true value of  $p$  is  $p_1$ .

$$\begin{aligned} \text{Power} &= \mathbf{P}\left(\hat{p} < p_0 + z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}\right) \\ &= \mathbf{P}\left(\frac{\hat{p} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} < \frac{p_0 + z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) \\ &= \mathbf{P}\left(z_{\hat{p}} < \frac{p_0 - p_1}{\sqrt{p_1(1-p_1)}} \sqrt{n} + z_\alpha \sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\right) \end{aligned}$$

Can you explain in detail the equations for the power?

where  $z_\alpha$  is the z-score with  $\mathbf{P}(z < z_\alpha) = \alpha$

From the equation, we know that as sample size  $n$  increases, the *power* of the test increase.

**Protect your own interest**

Refer to previous example about the potato chip producer and its main supplier. Answer the following questions.

- (a) State the *type II error* and its consequence.
- (b) If you are the potato chip producer, would you prefer a larger significance level  $\alpha$  or a smaller one? What if you are the main supplier?
- (c) According to the result of the example, what type of mistake may be made?

## 1.4 Two-sample z test for $p_1 - p_2$

Take a one-sided hypotheses test for example.

Population 1 is with proportion  $p_1$  and population size  $N_1$ . Population 2 is with proportion  $p_2$  and population size  $N_2$ . Is  $p_1 > p_2$ ? In order to test this hypotheses, we draw an SRS with size  $n_1$  from population 1, and an SRS with size  $n_2$  from population 2. The sample proportions for those two samples are  $\hat{p}_1$  and  $\hat{p}_2$  respectively.

The general steps are the same as *one-sample z test for population proportion* except for a few things we should pay attention to.

The hypotheses are

$$\mathbf{H}_0 : p_1 - p_2 = 0$$

$$\mathbf{H}_0 : p_1 - p_2 > 0$$

- **Large counts condition**

When assuming  $\mathbf{H}_0$  is true, we know  $p_1 = p_2$ , but don't know the exact value. When checking the large counts condition we check the following

$$n_1\hat{p}_1 \geq 10, \quad n_1(1 - \hat{p}_1) \geq 10; \quad n_2\hat{p}_2 \geq 10, \quad n_2(1 - \hat{p}_2) \geq 10$$

- **The test statistic  $z$**

$$\text{Test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of the statistic}}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Since we don't know  $p_1$  and  $p_2$ , we can replace them by  $\hat{p}_1$  and  $\hat{p}_2$ . But, the test statistic  $z$  is calculated when assuming  $\mathbf{H}_0$  is true, that is  $p_1 - p_2 = 0$ . Therefore  $p_1$  and  $p_2$  should be replaced by the same value instead of  $\hat{p}_1$  and  $\hat{p}_2$  separately. We replace  $p_1$  and  $p_2$  by the following value.

$$\hat{p}_C = \frac{X_1 + X_2}{n_1 + n_2}$$

where  $X_1$  and  $X_2$  are the numbers of "success" in the two samples.

$$X_1 = n_1\hat{p}_1, \quad X_2 = n_2\hat{p}_2$$

$\hat{p}_C$  is called the **pooled(combined) sampled proportion**.

The formula for test statistic  $z$  is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_2}}}.$$

### **Cholesterol and Hear Attacks**

High levels of cholesterol in the blood are associated with higher risk of heart attacks. Will using a drug to lower blood cholesterol reduce heart attacks? The Helsinki Heart Study recruited middle-aged men with high cholesterol but no history of other serious medical problems to investigate this question. The volunteer subjects were assigned at random to one of two treatments: 2051 men took the drug gemfibrozil to reduce their cholesterol levels, and a control group of 2030 men took a placebo. During the next five years, 56 men in the gemfibrozil group and 84 men in the placebo group had heart attacks.

- (a) Is this difference statistically significant at the  $\alpha = 0.01$  level?
- (b) Construct a 1% confidence interval for the difference of the ratios of heart attacks between those who take drugs and those who take placebos.
- (c) Compare the above two results.

## 1.5 One-sample z test for $\mu$

### One-sample t test for $\mu$

- One-sample z test for  $\mu$

A two-sided hypotheses test is: there is a population with size  $N$ , population mean  $\mu_0$  and standard deviation  $\sigma$ . Now we are suspect that the population mean might be different from  $\mu_0$ . In order to test this suspicion, an SRS with size  $n$  is drawn from the population, and the sample mean is  $\bar{x}$ . What is the conclusion about the suspicion?

We follow the sample procedure as before: set up hypotheses and significance level, identify correct test and check conditions, calculate P-value and conclude.

In this setting the hypotheses are

$$\mathbf{H}_0 : \mu = \mu_0$$

$$\mathbf{H}_a : \mu \neq \mu_0$$

The conditions to be checked is exactly the same as the conditions to be checked when constructing *one-sample z interval for  $\mu$* .

The test statistic is given by

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- One-sample t test for  $\mu$

In the above setting, if we don't know the standard deviation of the population, we have to use *one-sample t test for  $\mu$* . This situation is the same as constructing *one-sample t interval for  $\mu$* .

Everything is the same as *one-sample z test for  $\mu$* , except the test statistic.

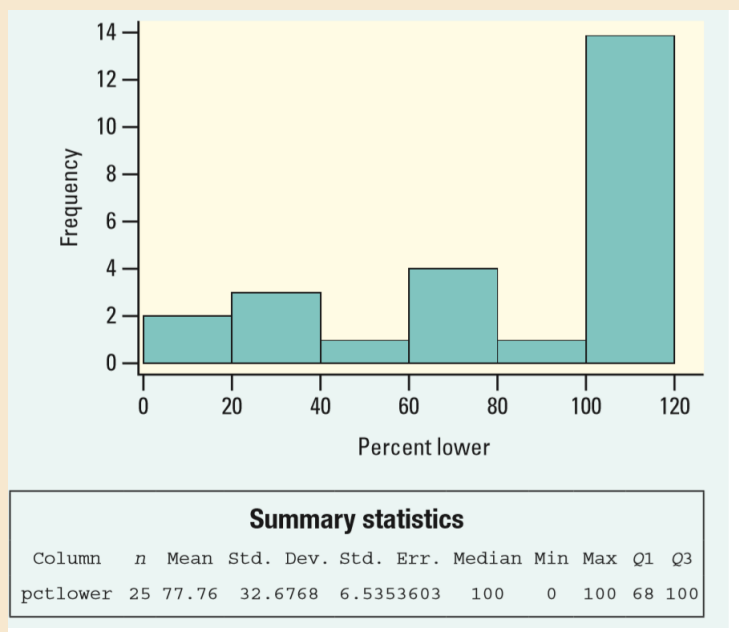
$$t = \frac{\bar{x} - \mu_0}{\frac{s_X}{\sqrt{n}}}, \quad df = n - 1$$

Be sure to report the *degree of freedom* in the exam.



### Paying High Prices?

A retailer entered into an exclusive agreement with a supplier who guaranteed to provide all products at competitive prices. The retailer eventually began to purchase supplies from other vendors who offered better prices. The original supplier filed a lawsuit claiming violation of the agreement. In defense, the retailer had an audit performed on a random sample of 25 invoices. For each audited invoice, all purchases made from other suppliers were examined and compared with those offered by the original supplier. The percent of purchases on each invoice for which an alternative supplier offered a lower price than the original supplier was recorded.<sup>21</sup> For example, a data value of 38 means that the price would be lower with a different supplier for 38% of the items on the invoice. A histogram and some computer output for these data are shown below. Explain why we should not carry out a one-sample t test in this setting.



**Interpreting the output**

The recommended daily allowance (RDA) of calcium for women between the ages of 18 and 24 years is 1200 milligrams (mg). Researchers who were involved in a large-scale study of women's bone health suspected that their participants had significantly lower calcium intakes than the RDA. To test this suspicion, the researchers measured the daily calcium intake of a random sample of 36 women from the study who fell in the desired age range. The Minitab output below displays the results of a significance test.

Summary statistics						
Test of $\mu = 1200$ vs $< 1200$						
Variable	N	Mean	StDev	SE Mean	T	P
Calcium	36	856.2	306.7	51.1	6.73	0.000

- (a) Do these data give convincing evidence to support the researchers suspicion? Justify your answer.
- (b) Interpret the P-value in context.

### Flu Vaccine

A drug company has developed a new vaccine for preventing the flu. The company claims fewer than 5% of the adults who use its vaccine will get the flu. To test the claim, researchers give the vaccine to a random sample of 1000 adults. Of these, 43 get the flu.

- (a) Do these data provide convincing evidence to support the companys claim?
- (b) Which kind of mistake, Type I error or Type II error, could you have made in (a)? Explain.
- (c) From the companys point of view, would a Type I error or Type II error be more serious? Why?

## 1.6 Two-sample z test for $\mu_1 - \mu_2$

### Two-sample t test for $\mu_1 - \mu_2$

- **Two-sample z test for  $\mu_1 - \mu_2$**

A one-sided hypotheses test might be setted as this. There are population 1 and population 2 with the standard deviation  $\sigma_1$  and  $\sigma_2$  respectively, and the alleged difference of the population means is  $\mu_1 - \mu_2 = D$ . We suspect the difference might be smaller. To test our suspicion, an SRS is drawn from population 1, with sample size  $n_1$ , sample mean  $\bar{x}_1$  and sample standard deviation  $s_1$ , the other SRS is drawn from population 2 with sample size  $n_2$ , sample mean  $\bar{x}_2$  and sample standard deviation  $s_2$ . What can we conclude from this.

The hypotheses are

$$\mathbf{H}_0 : \mu_1 - \mu_2 = D$$

$$\mathbf{H}_a : \mu_1 - \mu_2 < D$$

The conditions to be checked is exactly the same as the conditions checked when constructing two-sample z interval for  $\mu_1 - \mu_2$ .

The test statistic z is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2 - D}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- **Two-sample t test for  $\mu_1 - \mu_2$**

The setting is the same as above, except that we don't know the standard deviations of the populations

The conditions to be checked is exactly the same as the conditions checked in constructing the two-sample t interval for  $\mu_1 - \mu_2$

The test statistic if given by

$$t = \frac{\bar{x}_1 - \bar{x}_2 - D}{\mathbf{SE}_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - D}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The degree of freedom of  $t$  are given by either a conservative method,  $\min(n_1 - 1, n_2 - 1)$ , or by *technology*. This is the same as when we construct the two-sample t interval for  $\mu_1 - \mu_2$ .

Most of the time, we just report the degree of freedom given by the calculator.

**Calcium and Blood Pressure**

Does increasing the amount of calcium in our diet reduce blood pressure? Examination of a large sample of people revealed a relationship between calcium intake and blood pressure. The relationship was strongest for black men. Such observational studies do not establish causation. Researchers therefore designed a randomized comparative experiment.

The subjects were 21 healthy black men who volunteered to take part in the experiment. They were randomly assigned to two groups: 10 of the men received a calcium supplement for 12 weeks, while the control group of 11 men received a placebo pill that looked identical. The experiment was doubleblind. The response variable is the decrease in systolic (top number) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase appears as a negative number. Here are the data:

<b>Group 1 (calcium):</b>	7	4	18	17	3	5	1	10	11	2	
<b>Group 2 (placebo):</b>	1	12	1	3	3	5	5	2	11	1	3

In the matched pairs design, it looks like we have got two set of data, and we should perform a two-sample hypotheses test or construct a two-sample confidence interval. In reality, we only have one set of data, which is the difference within each pair.

Refer to the exercise “**Nitrogen in tires —a lot of hot air** ” in the chapter of “**Designing studies**”. Answer the following questions.

- (a) Perform a hypotheses test under the significance level  $\alpha = 0.05$  to test the claim that filling automobile tires with nitrogen reduces pressure loss.
- (b) Construct a 95% confidence interval for the difference of the mean pressure loss between tires filled with air and tires filled with nitrogen.

## Investigation problems with simulated sampling distribution

## The accuracy of EKG

An exercise electrocardiogram (EKG) checks for changes in your heart during exercise and is useful in diagnosing coronary artery disease. An EKG has fewer potential side effects but is much less precise than thallium tomography. In one EKG study, 500 volunteers with known coronary artery disease and 500 volunteers with healthy arteries underwent EKG checks. The physicians administering and evaluating the tests did not know the physical condition of any volunteer. The following table gives the numbers of volunteers whom the physicians evaluated as “positive” for coronary disease.

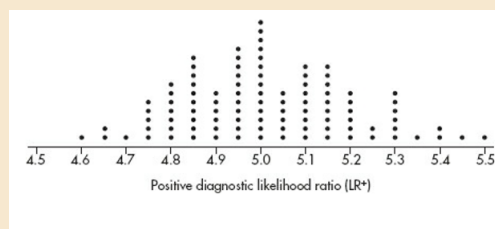
	Test for coronary disease	
	Positive	Negative
Healthy volunteers	100	400
Volunteers with disease	305	195

- (a) *Sensitivity* is defined as the probability of a positive test given that the subject has disease. What was the sensitivity of this study?
- (b) *Specificity* is defined as the probability of a negative test given that the subject is healthy. What was the specificity of this study?
- (c) A valuable tool for assessing the accuracy of such studies is the *positive diagnostic likelihood ratio* ( $LR^+$ ) which gives the ratio of the probability a positive test result will be observed in a diseased person compared to the probability that the same result will be observed in a healthy person.

$$LR^+ = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

What was  $LR^+$  in this study, and explain why the larger the value of  $LR^+$ , the more useful the test.

- (d) Suppose in one such sample study,  $LR^+ = 4.7$ . To determine whether or not this is sufficient evidence that the population  $LR^+$  is below the desired value of 5.0, 100 samples from a population with a known  $LR^+$  of 5.0 are generated, and the resulting simulated values of  $LR^+$  are shown in the dotplot:



Based on this dotplot and the sample  $LR^+ = 4.7$ , is there evidence that the population  $LR^+$  is below the desired value of 5.0? Explain.