

Supplementary Handout for Dis 10: Hypothesis Testing

1 Motivation

- Last week, we discussed point and interval estimators, which are used to estimate the population parameter of interest.
- An interval estimator has the advantage of estimating a range for a population parameter.
 - For example, using a confidence interval constructed under certain confidence level, we can evaluate what's the likely true value of the population mean.
- What if we want to do the same for a point estimator?
 - For example, what if one wants to know whether the true population mean equals to a certain value / is within a certain range, given the sample mean estimate?
 - This is what **hypothesis testing** is all about.

2 Hypothesis

- When performing hypothesis testing, you can divide the outcome of your test into two hypotheses:
 1. The default outcome called **the null hypothesis**, denoted as H_0
 2. The alternative outcome called **the alternative hypothesis**, denoted as H_1
- How do we know which outcome is the null, and which one is the alternative?
 - By convention, the null hypothesis is where you write the outcome with equality.
 - For the outcome that doesn't have an equality sign (i.e. the outcome with \neq , $>$, or $<$ sign), it goes to the alternative hypothesis.

Exercise. Write down the null and alternative hypothesis for each proposed testing scenario:

1. Whether μ equals 5.
 $H_0 : \mu = 5$
 $H_1 : \mu \neq 5$
2. Whether μ is greater than 10.
 $H_0 : \mu = 10$
 $H_1 : \mu > 10$
3. If μ is less than 15.
 $H_0 : \mu = 15$
 $H_1 : \mu < 15$

- What do we do once we have the null and alternative hypothesis?
 - Test **whether we can reject the null**
 - If we **reject the null**, then we **accept the alternative** (under conditions that lead to the rejection).
 - If we **fail to reject the null**, then it **doesn't necessarily mean that the null is true**. It just happens to be that, compared to the given alternative hypothesis, the null is the more likely outcome.

3 Performing a Hypothesis Test

- So how do we test whether we can reject the null?
- In this section, we are going to use the example of testing null hypothesis related to the population mean μ , where the corresponding sample statistic is \bar{X} .
- To see whether the \bar{X} we get from a sample supports the rejection or failure of rejection of the null, it would be very helpful to know how \bar{X} is distributed; this way, we would know how \bar{X} compares to the proposed population mean level from the null hypothesis.

3.1 What is the distribution of \bar{X} ?

- From Dis 8, we learned about the sampling distribution of \bar{X} :
 1. If X is exactly normally distributed, then \bar{X} is also exactly normally distributed.
 2. If X is not exactly normally distributed, but the sample size $n \geq 30$, then we can apply central limit theorem (CLT), and claim that \bar{X} is approximately normally distributed.
 3. If X is not exactly normally distributed, and the sample size $n < 30$, then we're screwed.
- So in order to perform hypothesis testing, we need to be in case 1 or 2, and **we are going to assume for the rest of this handout that either case 1 or 2 hold.**
- Under case 1 or 2, we know that a normal distribution is needed to describe the relationship between \bar{X} and the hypothesized μ .
- Now begs the question: if \bar{X} follows a normal distribution (either exactly or approximately), what are the parameters of this normal distribution?
 - The variance of the normal distribution is pretty straight forward, since we are still assuming that the population standard deviation of X is known as σ , and the sample size n is given to you. Thus,

$$\text{Variance of the normal distribution} = \frac{\sigma^2}{n} = \left(\frac{\sigma}{\sqrt{n}} \right)^2$$

(Implicitly assuming that N is large enough so that finite population correction factor for the standard deviation of sample mean is not needed.)

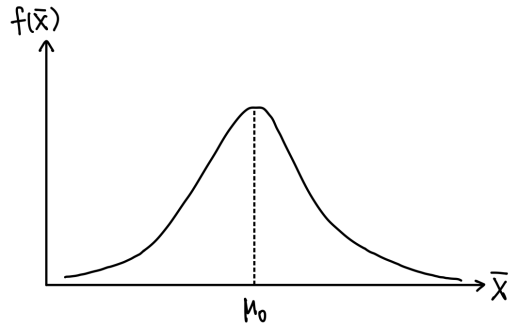
- What about the mean of the normal distribution then? This is the item that is unknown, but we have some hypothesized value about it. Specifically, the null hypothesis assumes that the mean of the normal distribution equals to a certain value. We are going to call the value of population mean given in H_0 as μ_0 .

For example, if $H_0 : \mu = 4$, then $\mu_0 = 4$. Anyways,

$$\text{Mean of the normal distribution} = \mu_0$$

- Thus, \bar{X} follows (exactly or approximately)

$$N\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

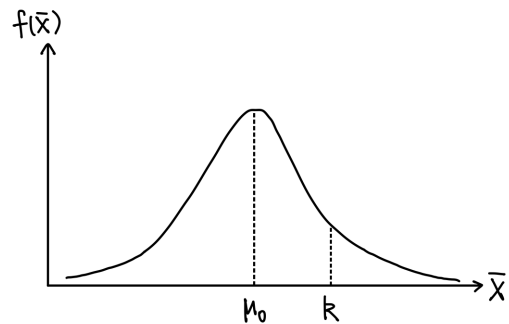


3.2 Test method 1: test statistic and rejection region

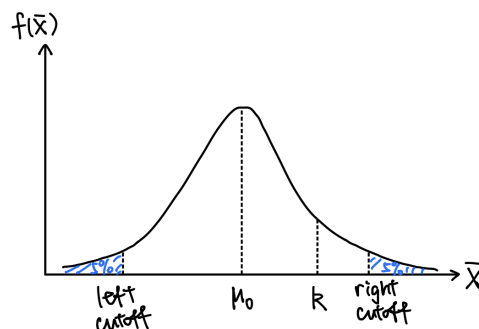
- Say that the sample mean calculated is $k > 0$, and our hypotheses are the following:

$$H_0 : \mu = \mu_0$$

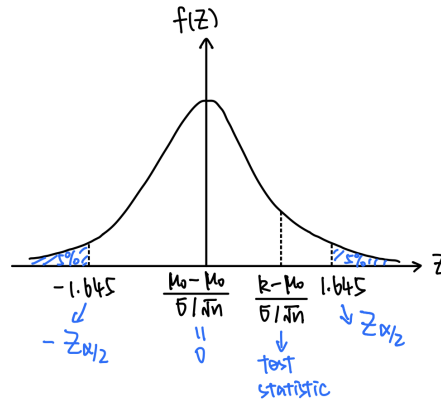
$$H_1 : \mu \neq \mu_0$$



- To reject the null and accept the alternative, we need k to be **as far away from the hypothesized mean μ_0 as possible** – either to on the far left side of μ_0 , or on the far right side of μ_0 .
 - Exactly how far away from μ_0 is a choice made by the researcher, and this is where **significance level (α)** comes into play.
- Say that one sets the significance level $\alpha = 10\%$. Then we'd like the probability of being in the far left side and the far right side of the distribution to sum up to be 10% – pretty unlikely outcome, so if we reach these far sides, we know that the null of $\mu = \mu_0$ is unlikely to hold.
 - Since the far left and far right side probability sum up to be 10%, people tend to just equally divide them for each tail:



- So if $k < \text{left cutoff}$, or if $k > \text{right cutoff}$, then we reject the null at significance level $\alpha = 10\%$.
- There is just one drawback to this whole plan so far: \bar{X} doesn't follow a standard normal distribution, so the left and right cutoff value will change all the time. To make it easier, we can standardize \bar{X} so that a standard normal distribution is followed. This means that

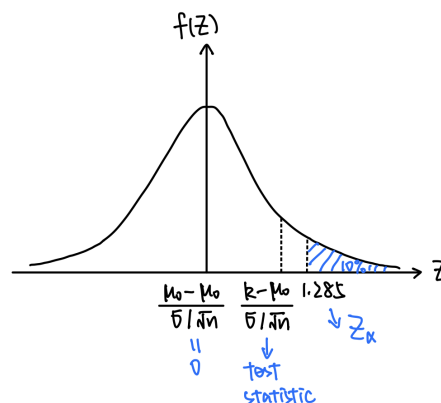


- Now, if $\frac{k - \mu_0}{\sigma / \sqrt{n}}$ is less than $-Z_{\alpha/2} = -1.645$, or if $\frac{k - \mu_0}{\sigma / \sqrt{n}}$ is greater than $Z_{\alpha/2} = 1.645$, then we reject the null at $\alpha = 10\%$ confidence level.
 - **Test statistic** = $\frac{k - \mu_0}{\sigma / \sqrt{n}}$ (standardized sample mean)
 - **Rejection region:** $|\text{test statistic}| > Z_{\alpha/2}$

The aforementioned method considers an alternative hypothesis $H_1 : \mu \neq \mu_0$, which is why we reject when test statistic is either too far to the left, or too far to the right of the standardized sample mean distribution. Such test is called a **two-tailed test**.

If H_1 only considers one side (i.e. a **left-tailed / right-tailed test**; e.g. $H_1 : \mu > \mu_0$ is a right-tailed test), then we only need to see if the test statistic is too far to one side of the sample mean distribution.

If we continue to consider the sample mean calculated as k , the significance level chosen as 10%, and a right-tailed test is performed, then



Rejection region: test statistic = $\frac{k - \mu_0}{\sigma / \sqrt{n}} > Z_{\alpha}$

3.3 Test method 2: confidence interval test (for a two-tailed test only)

- Last week, we learned how to construct a confidence interval to estimate population mean. Recall that when confidence level is $(1 - \alpha)$, the confidence interval constructed is

$$\left[\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- If you remember how we constructed the confidence interval, you'll see that this is the reversed process compared with performing a two-tailed test using test statistic and rejection region – confidence interval gives you the region to **not reject** the null under significance level α .
- That is, for

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

- If $\mu_0 \in \text{CI}$ constructed at $(1 - \alpha)$ confidence level, then we fail to reject the null at α significance level
- If $\mu_0 \notin \text{CI}$ constructed at $(1 - \alpha)$ confidence level, then we reject the null at α significance level

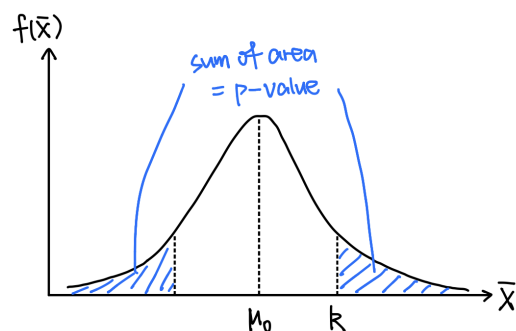
3.4 Test method 3: p-value

- Say that we are still performing a two-tailed test:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

- Instead of looking at which region to reject given a significance level, we can think about if the current sample mean is used as the cutoff value, then what's the probability that we can reject the null hypothesis – this is the **p-value**:



$$\begin{aligned} \text{p-value} &= P(\bar{X} > k) + P(\bar{X} < -k) = P(\bar{X} > k) \times 2 \\ &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) \times 2 = P\left(Z > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) \times 2 \end{aligned}$$

(Note: for example, if a right-tailed test is performed, then $\text{p-value} = P\left(Z > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right)$)

- Now, since p-value represents the probability that we can reject the null given the sample mean k ,
 - If p-value > significance level α , we fail to reject the null
 - If p-value < significance level α , we reject the null

Essentially, **p-value is the smallest significance level needed to reject the null.**

Note:

- If you have a two-tailed test, you can use any of the three methods.
- If you have a one-tailed test, you can only use method 1 (test statistic and rejection region) or method 3 (p-value).
- No matter which test method is used, the same conclusion should be reached**

4 Errors When Performing a Hypothesis Test

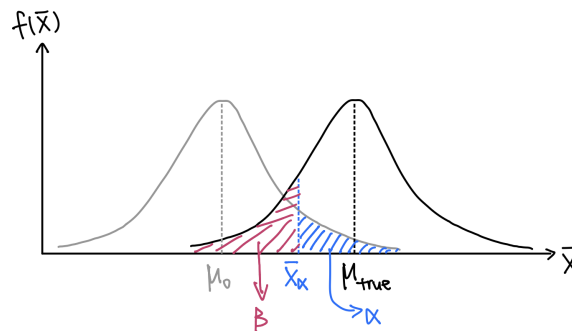
| | H_0 is True | H_0 is False |
|------------------|---|---|
| Reject H_0 | Type I Error Probability = α | Power Probability = $1 - \beta$ |
| Not Reject H_0 | GOOD JOB Probability = $1 - \alpha$ | Type II Error Probability = β |

- The way we construct the aforementioned tests is by setting the level of Type I Error (i.e. setting what α is), since we think rejecting the null when it is true is a lot more costly than not rejecting the null when it is false (recall the criminal justice analogy from lecture).
- Sometimes, if you have information about what the true population mean is, it is helpful to quantify the probability of committing Type II Error as well; this way, we can also evaluate how the tests perform when the null should be rejected.
- Say that we have a right-tailed test

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

Let the true population mean be μ_{true} , and let \bar{X}_α be the point where significance level is set as α :



Thus, the probability of committing Type II error = $\beta = P(\bar{X} < \bar{X}_\alpha | \mu = \mu_{true}) = P(Z < \frac{\bar{X}_\alpha - \mu_{true}}{\sigma/\sqrt{n}})$