Chapter 7. Inference concerning the mean (B)

March 29, 2011

1 Nonstatistical Hypothesis Testing

In a trial, a jury must decide between two hypotheses. The null hypothesis is

H_0: The defendant is innocent

The alternative hypothesis or research hypothesis is

H_1: The defendant is guilty

The jury must make a decision on the basis of evidence presented.

In the language of statistics convicting the defendant is called **rejecting the null hypothesis in favor of the alternative hypothesis**. That is, the jury is saying that there is enough evidence to conclude that the defendant is guilty (i.e., there is enough evidence to support the alternative hypothesis).

If the jury acquits that there is not enough evidence to support the alternative hypothesis, the jury is not saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis. That is why we dont like to say that we accept the null hypothesis, although most people in industry will say "We accept the null hypothesis"

There are two possible errors.

- A **Type I error** occurs when we reject a true null hypothesis.

  That is, a Type I error occurs when the jury convicts an innocent person. We would want the probability of this type of error [maybe 0.001 —— beyond a reasonable doubt] to be very small for a criminal trial where a conviction results in the death penalty, whereas for a civil trial, where conviction might result in someone having to "pay for damages to a wrecked auto",we would be willing for the probability to be larger [0.49 —- preponderance of the evidence]

$$\text{P(Type I error)} = \alpha \text{ [usually 0.05 or 0.01]}$$

- A **Type II error** occurs when we don't reject a false null hypothesis [accept the null hypothesis].

That occurs when a guilty defendant is acquitted. In practice, this type of error is by far the most serious mistake we normally make. For example, if we test the hypothesis that the amount of medication in a heart pill is equal to a value which will cure your heart problem and "accept the hull hypothesis that the amount is ok". Later on we find out that the average amount is WAY too large and people die from "too much medication" [I wish we had rejected the hypothesis and threw the pills in the trash can], it's too late because we shipped the pills to the public.

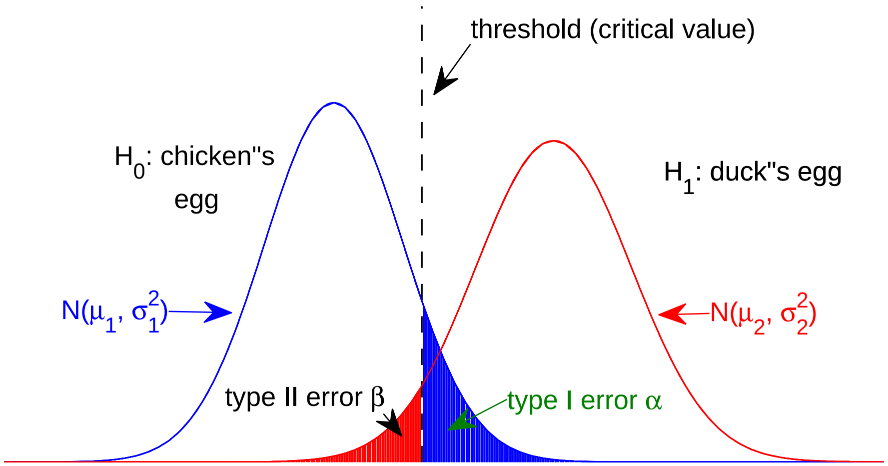
$$\text{P(Type II error)} = \beta \text{ [usually not being able to be controlled]}$$

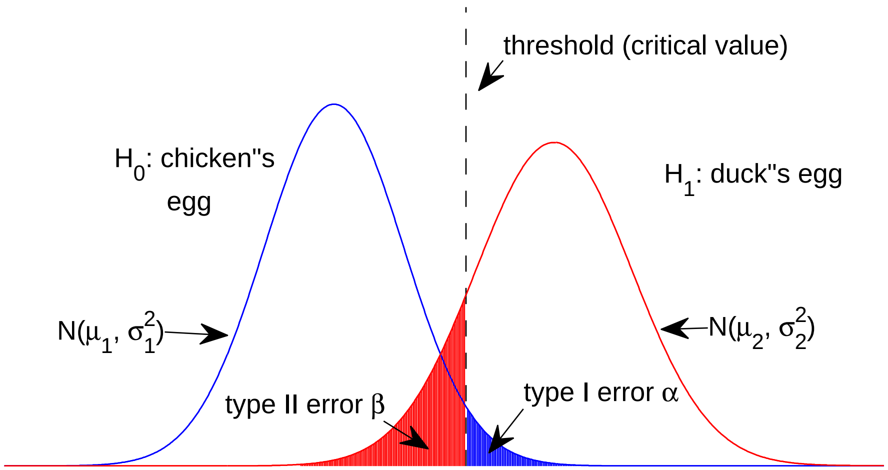
- The two probabilities are inversely related. Decreasing one increases the other, for a fixed sample size.

  In other words, you can't have α and β both real small for any old sample size. You may have to take a much larger sample size, or in the court example, you need much more evidence.

- Statistical understanding of the 2 types of errors. Consider an example: we have one egg, and we are not sure whether it a chicken's egg or duck's egg. Suppose weight of a chicken's egg $X \sim N(\mu_1, \sigma_1^2)$ and duck's egg $Y \sim N(\mu_2, \sigma_2^2)$. For the egg,

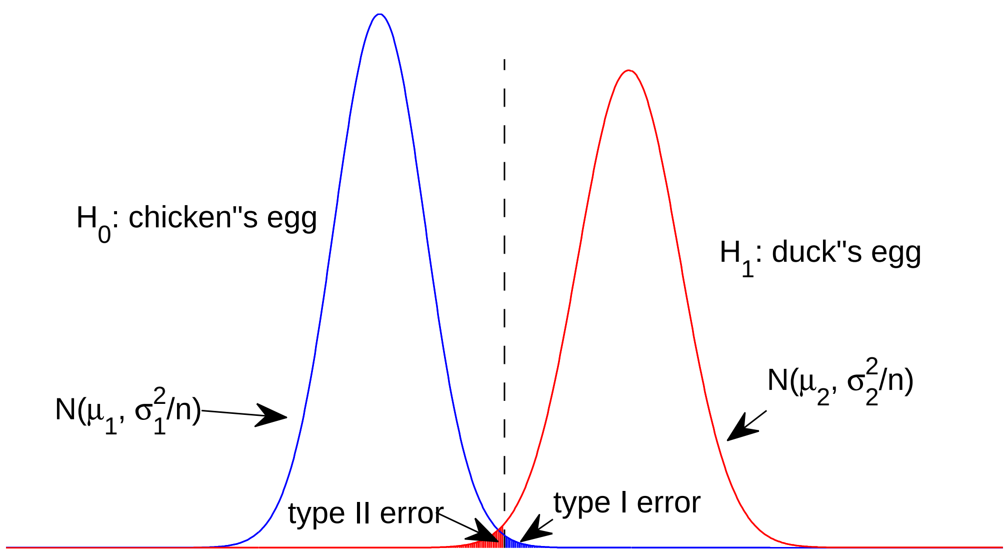
$$\mathrm{H}_0\text{: it is chicken's egg,} \qquad \mathrm{H}_1\text{: it is duck's egg}$$


threshold (critical value)
H0: chicken"s egg
H1: duck"s egg
N(μ1, σ1²)
N(μ2, σ2²)
type II error β
type I error α


threshold (critical value)
H0: chicken"s egg
H1: duck"s egg
N(μ1, σ1²)
N(μ2, σ2²)
type II error β
type I error α

How can we reduce the error

Now consider n eggs with all being chicken's eggs or all being duck's eggs. Consider $\bar{X}$.



As n increases, the errors can be reduced. In other words, as more evidences are available, we can make more accurate judgement.

The critical concepts

1. There are two hypotheses, the null and the alternative hypotheses.

2. The procedure begins with the assumption that the null hypothesis is true.

3. The goal is to determine whether there is enough evidence or contradiction to the null hypothesis to infer that the alternative hypothesis is true, or the null is not likely to be true.

4. enough evidence or contradiction to the null hypothesis can be obtained by statistical or non-statistical methods.

5. There are two possible decisions:

- Conclude that there is enough evidence or contradiction to the null hypothesis to support the alternative hypothesis. Reject the null hypothesis.
- Conclude that there is not enough evidence or contradiction to the null hypothesis to support the alternative hypothesis. Fail to reject the null hypothesis.

- The two hypotheses are called the null hypothesis and the other the alternative or research hypothesis. The usual notation is:

$$\mathrm{H}_0\text{: ....,} \qquad \mathrm{H}_1 \text{ (or } \mathrm{H}_a\text{): ....}$$ [1]

or

$$\mathrm{H}_0\text{: .... versus } \mathrm{H}_1 \text{ (or } \mathrm{H}_a\text{): ....}$$

- The null hypothesis (H_0) will always state that the parameter equals a value.

[1] H_0 is pronounced H "nought"

- Consider mean demand for computers during assembly lead time. Rather than estimate the mean demand, our operations manager wants to know whether the mean is different from 350 units. In other words, someone is claiming that the mean time is 350 units and we want to check this claim out to see if it appears reasonable. We can rephrase this request into a test of the hypothesis:

$$H_0 : \mu = 350$$

Thus, our alternative hypothesis becomes:

$$H_1 : \mu \neq 350$$

Intuitive understanding

Recall that the standard deviation σ was assumed to be 75, the sample size n was 25, and the sample mean $\bar{x}$ was calculated to be 370.16

For example, if we are trying to decide whether the mean is not equal to 350, a large value of $\bar{x}$ (say, 600) would provide "stronger" evidence.

If $\bar{x}$ is close to 350 (say, 355) we could not say that this provides a great deal of evidence to infer that the population mean is different than 350.

The bigger difference between the assumed and the observed, the stronger is the evidence

- The two possible decisions that can be made:
  - Conclude that there is enough evidence to support the alternative hypothesis (also stated as: reject the null hypothesis in favor of the alternative)
  - Conclude that there is not enough evidence to support the alternative hypothesis (also stated as: failing to reject the null hypothesis in favor of the alternative, or we accept the null hypothesis)

    NOTE: Some statisticians do not like to say that we accept the null hypothesis.

- The testing procedure begins with the assumption that the null hypothesis is true.

  Thus, until we have further statistical evidence, we will assume:

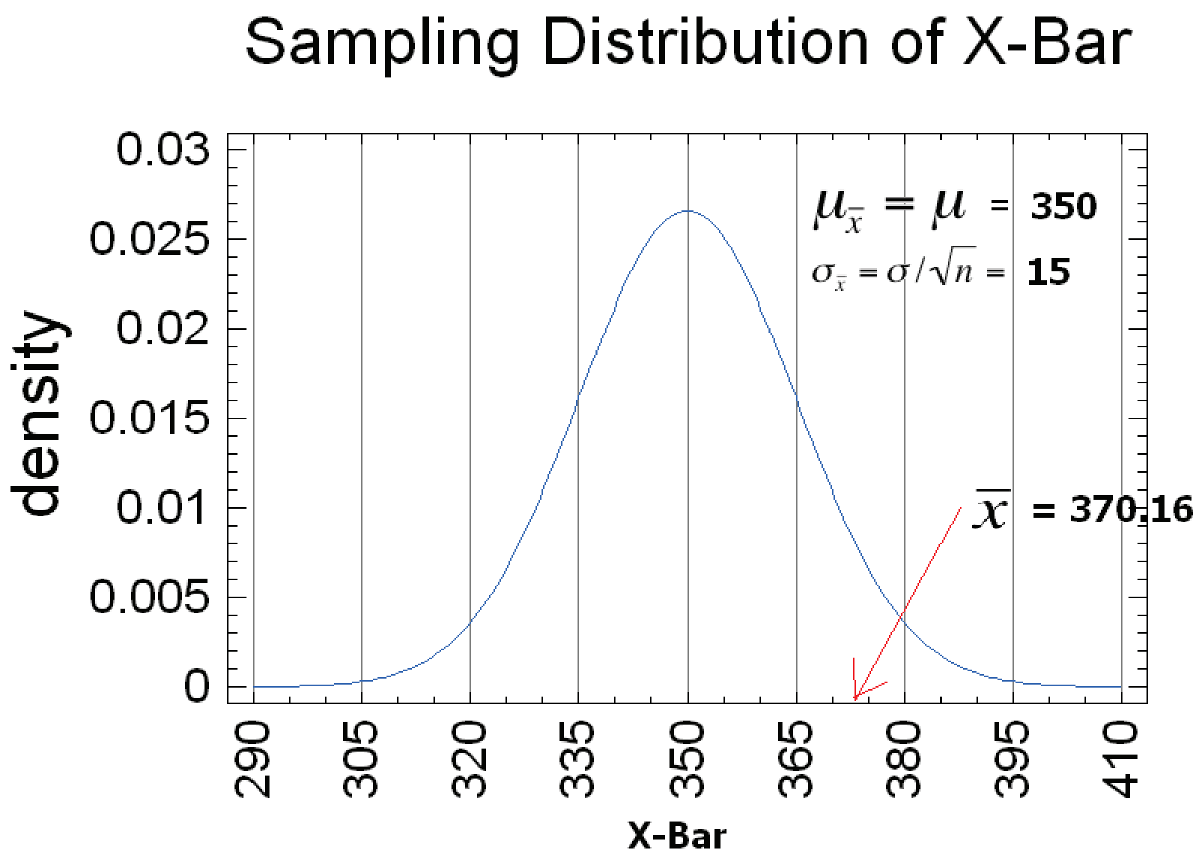
$$H_0 : \mu = 350 \qquad \text{(assumed to be TRUE)}$$

- The next step will be to determine the sampling distribution of the sample mean $\bar{X}$ assuming the true mean is 350. $\bar{X}$ is normal with

$$E\bar{X} = 350$$

and

$$Var(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{75}{\sqrt{25}} = 15.$$



where $\mu_{\bar{x}}$ stands for $E\bar{X}$ and $\sigma_{\bar{x}}$ for $\{Var(\bar{X})\}^{1/2}$.

Three ways to make a decision

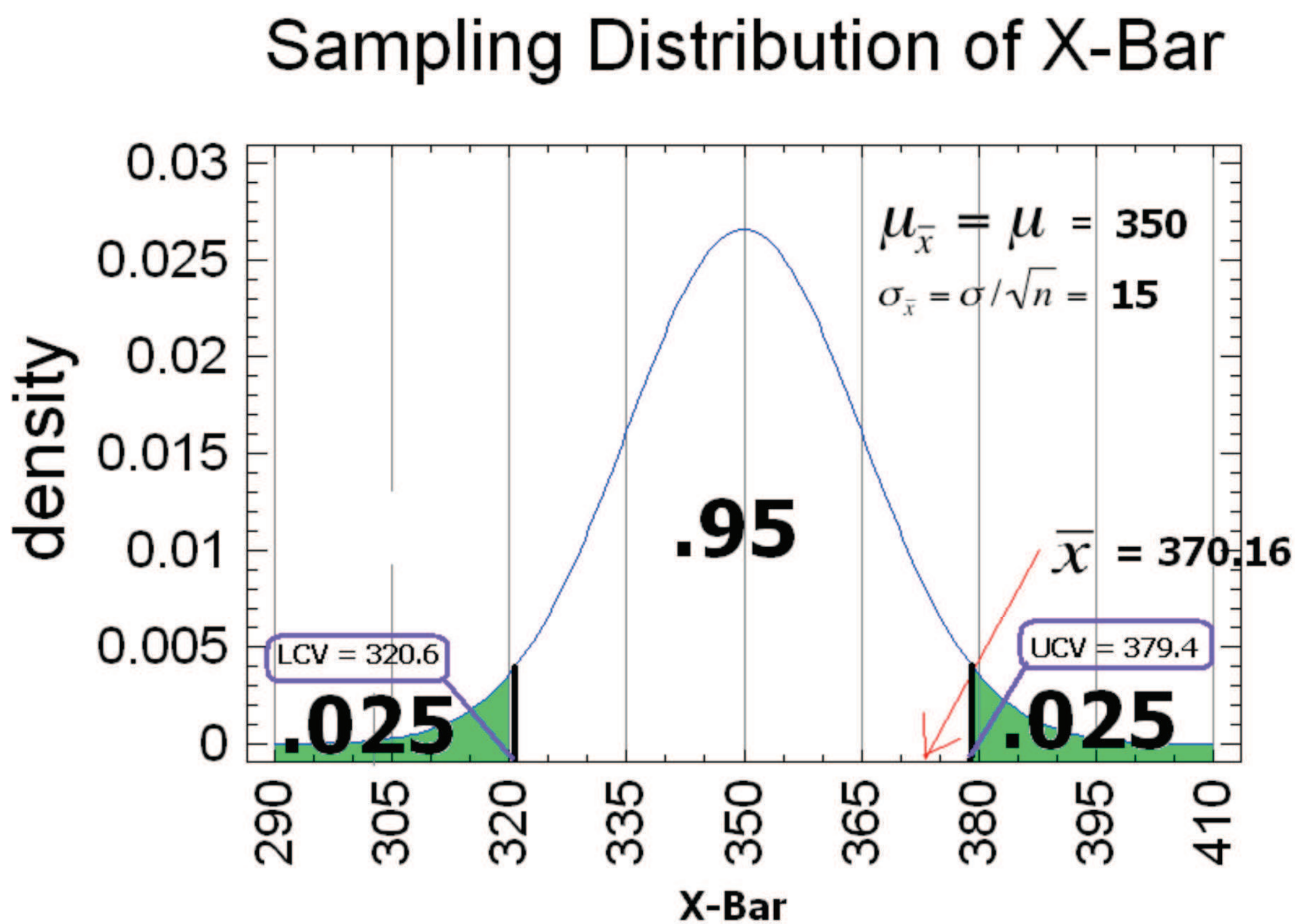
1. Unstandardized test statistic: Is $\bar{x}$ in the guts of the sampling distribution?

   Depends on what you define as the 'guts' of the sampling distribution.

   If we define the guts as the center 95% of the distribution, then the critical values that define the guts will be 1.96 standard deviations of X-Bar on either side of the mean of the sampling distribution 350, or

$$UCL = 350 + 1.96 * \frac{\sigma}{\sqrt{n}} = 350 + 29.4 = 379.4$$
$$LCL = 350 - 1.96 * \frac{\sigma}{\sqrt{n}} = 350 - 29.4 = 320.6$$

   Since $\bar{x}$ is in the region that is very likely, we have no evidence to reject H_0



Sampling Distribution of X-Bar
density
0.03
0.025
0.02
0.015
0.01
0.005
0
$\mu_{\bar{x}} = \mu$ = 350
$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ = 15
.95
$\bar{x}$ = 370.16
LCV = 320.6
UCV = 379.4
.025
.025
290
305
320
335
350
365
380
395
410
X-Bar

2. Standardized test statistic: Since we defined the "guts" of the sampling distribution to be the center 95%,

- If the z-value for the sample mean is greater than 1.96, we know that will be in the reject region on the right side or
- If the z-value for the sample mean is less than -1.96, we know that will be in the reject region on the left side.

$$Z = (\bar{X} - \mu)/(\sigma/\sqrt{n}) = (370.16 - 350)/15 = 1.344$$

Is this z-value in the guts of the sampling distribution???

3. The p-value approach (which is generally used with a computer and statistical software): when μ is considered, p-value is defined as one of the following

   - (one-sided) If H_0 is true, p-value $= P(\bar{X} \geq \bar{x})$ or p-value $= P\{\bar{X} \leq \bar{x}\}$
   - (two-sided) If H_0 is true,

$$
\begin{aligned}
p\text{-value} &= P(|\bar{X} - \mu_0| > |\bar{x} - \mu_0|) \\
&= P(\bar{X} - \mu_0 > |\bar{x} - \mu_0|) + P(\bar{X} - \mu_0 < -|\bar{x} - \mu_0|) \\
&= 2P(\bar{X} - \mu_0 > |\bar{x} - \mu_0|)
\end{aligned}
$$

For this example, since $\bar{x} = 370.16$ and $\mu_0 = 350$, calculate

$$p\text{-value} = 2 * P(\bar{X} - \mu_0 > 370.16 - \mu_0) = 2 * P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > 1.344)$$

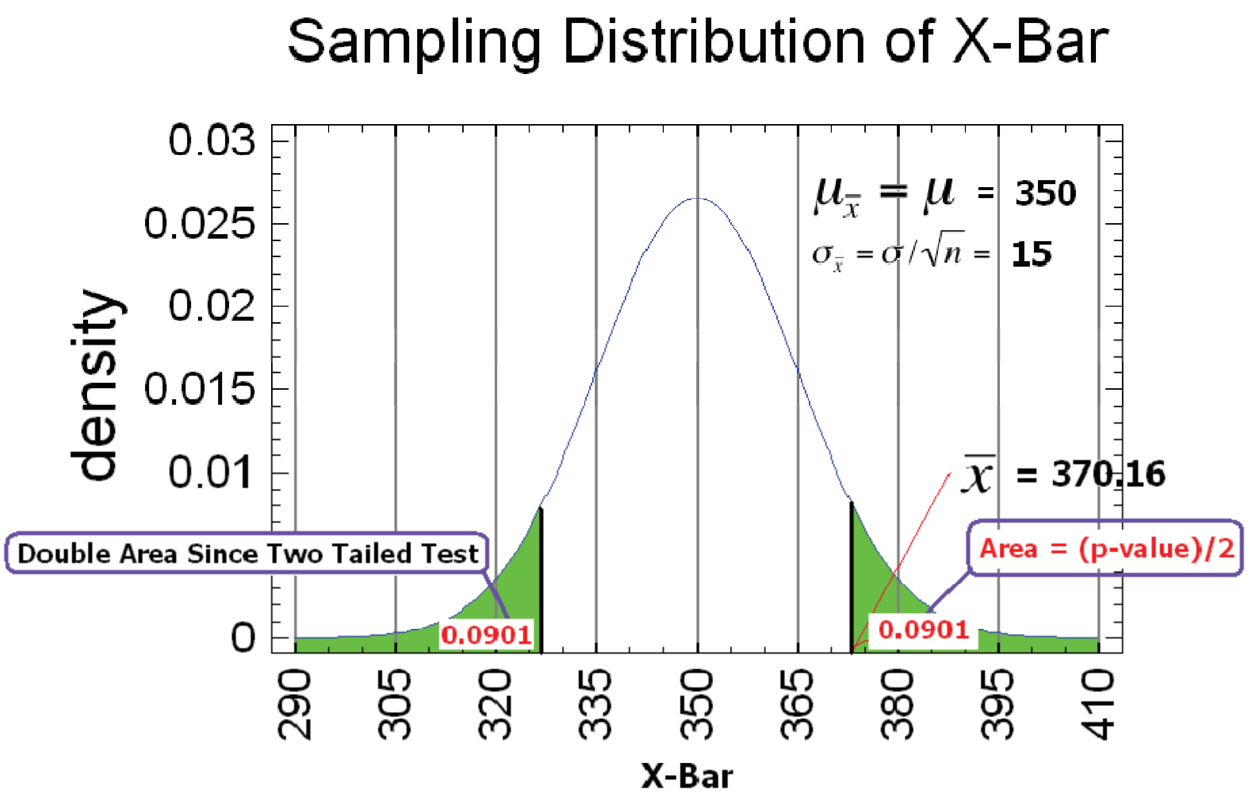
if indeed, $\mu = \mu_0$, then

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

we can find from the table $P(Z > 1.344) = 0.0901$. Thus

$$p\text{-value} = 2 * 0.0901 = 0.1902$$

"with H_0, $\mu = \mu_0$, our observed value is in a reasonable region", we have no evidence to reject H_0



Sampling Distribution of X-Bar
density
0.03
0.025
0.02
0.015
0.01
0
$\mu_{\bar{x}} = \mu$ = 350
$\sigma_{\bar{x}} = \sigma / \sqrt{n}$ = 15
$\bar{x}$ = 370.16
Double Area Since Two Tailed Test
Area = (p-value)/2
0.0901
0.0901
290
305
320
335
350
365
380
395
410
X-Bar

Steps for A Statistical Test of Hypothesis

- A statistical test of hypothesis consists of five parts:

  1. set up null hypothesis (H_0) & alternative hypothesis (H_1).
  2. Specify type I error probability (level of significance α, which is usually very small and that an event with probability α is very unlikely to happen. Commonly used α values are 0.05 and 0.01)
  3. Assuming H_0 is true, construct test statistic, its distribution and rejection criteria
  4. calculate the test statistic value based on the sampled data
  5. make conclusion

Step 1. Null & Alternative Hypotheses

- The null hypothesis (H_0) & alternative hypothesis (H_1) are two competing hypothesis.
- The alternative hypothesis (H_1): generally the claim we wish to establish.
- The null hypothesis (H_0): the claim from our opponent.
- We usually require H_0 and H_1 are complementing each other, i.e. if H_0 is true, then H_1 must be false, likewise, if H_1 is true, H_0 must be false.

- Guideline for selecting the null hypothesis: the term null hypothesis is used for any hypothesis set up primarily to see whether it can be rejected.

- Mathematically, H_0 is usually formulated as a form of "$=$", while H_1 is formulated as "$<$" or "$>$" or "$\neq$". Possible H_0&H_1's concerning μ are listed as follows

(a) H_0: $\mu = \mu_0$ v.s. H_1: $\mu > \mu_0$ — one sided test

(b) H_0: $\mu = \mu_0$ v.s. H_1: $\mu < \mu_0$ — one sided test

(c) H_0: $\mu = \mu_0$ v.s. H_1: $\mu \neq \mu_0$ — two sided test

where μ_0 is a given number.

Example: Mean Exam Score is 16?

The lecturer of ST2334 announced that the mean score of midterm is 16. A student doubt it, he thought that the mean score must be greater than 16. If we denote by μ the true mean score, then we have the following hypothesis

$$\mathrm{H}_0\colon\ \mu = 16, \qquad \mathrm{H}_1\colon\ \mu > 16$$

Step 2. Type I & Type II Error

- For any test of hypothesis problem, there are two possible conclusions:
  - Reject H_0 and therefore conclude H_1
  - Do not reject H_0 and therefore conclude H_0
- Whatever decision is made, there is a probability of making an error. Refer to the decision table below

	Not reject H_0	Reject H_0
H_0 is true	Correct Decision	Type I error
H_0 is false	Type II error	Correct Decision

- Type I error: Reject H_0 when H_0 is true

- Type II error: Do not reject H_0 when H_1 is true

- Denote by α the probability of making a type I error (also called the level of significance or significance level)

- Denote by β the probability of making a type II error

- Question: a decision is made, can one make type I and type II error simultaneously? NO; see the figures on Page 6.

Step 3. Test Statistic & Its Distribution

- Test statistic and its distribution are the main issue in the problem of tests of hypotheses
- They are derived by statisticians based upon the sample.
- Test statistic may vary. But a good statistic is the one that can make type I error get controlled and diminish the type II error probability β.

Step 4. Rejection Criteria (by finding an critical value or p-value)

- We introduce two types of rejection criteria: rejection region & p-value

- Rejection criteria should be established based on the test statistic, its distribution under H_0, and fixed α.

- Rejection region: an established region, such that if the statistic value after plugging in the data is located within the region, reject H_0, otherwise, do not reject H_0.

- Rejection region should be based on α: if H_0 is rejected, the type I error probability of the test should not exceed α.

- p-value: with H_0 being true, the probability of obtaining a value for the test statistic that is as extreme or more extreme than the value actually observed.
- Reject or not: compare p-value with the level of significance, α, if p-value$<$ α, reject H_0, otherwise do not reject H_0
- People need to be informed that both criteria are based on the philosophy that
  - type I error probability MUST be controlled by α, then...
  - try to decrease the probability of type II error, β.

Step 5. Test Statistic Value and Conclusion

- the test statistic value needs to be computed based on the observed sample.
- the conclusion is made
  - either by checking whether the computed statistic value is located within the rejection region
  - or by constructing the p-value based on the computed statistic value, and compare it with the level of significance α.

3 Hypotheses Concerning One Mean

Case I: σ Known, Data Normal

Steps for 2-sided test

- Suppose $X_1, X_2, ..., X_n$is a random sample. Now that we follow the steps from the last section to establish test of hypothesis for Case I.
- Step 1:
  - Null hypothesis is always H_0: $\mu = \mu_0$.
  - We use two-sided alternative hypothesis as an example: H_1: $\mu \neq \mu_0$.
- Step 2: set level of significance: α=0.05.

- Step 3 (main step in test of hypothesis)

  – Statistic & its distribution: with σ being known and population being Normal,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

  When H_0 is true, i.e. $\mu = \mu_0$, the above statistic becomes

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

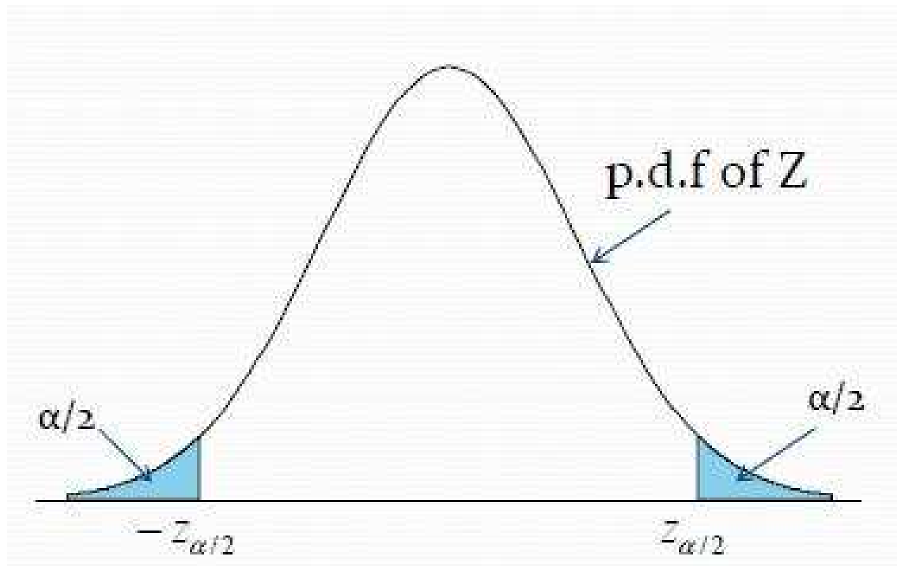
  which serves as our test statistic.

  – Rejection region: intuitively, we should reject H_0 when $\bar{X}$ is too large or too small compare with μ_0, that is, when Z is too large or too small.

In theory, the probability $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$ is α i.e.

$$P(|Z| > z_{\alpha/2}) = \alpha$$

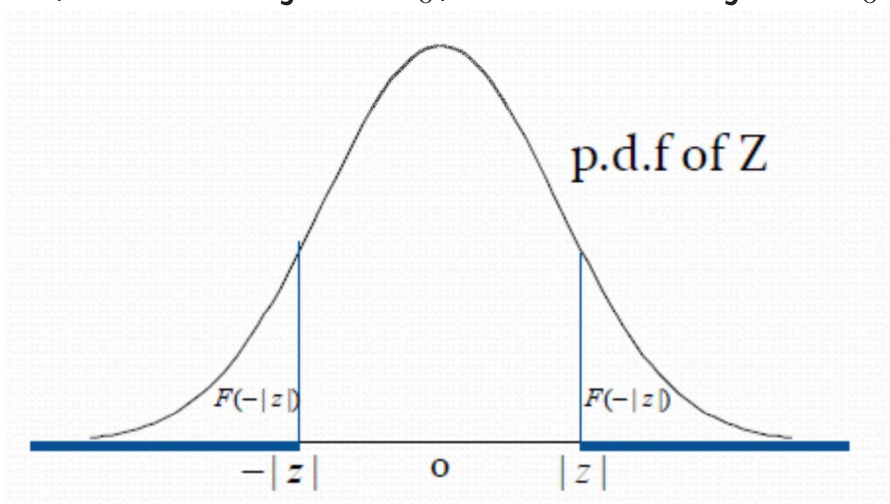
the rejection region is $|z| > z_{\alpha/2}$



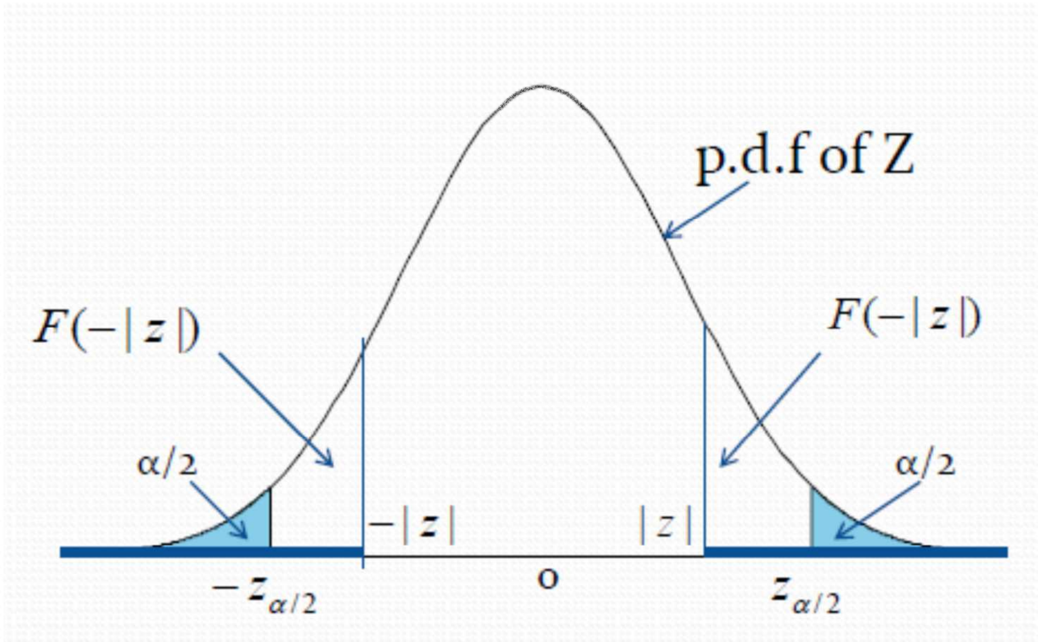
– p-value: suppose after plugging in the data into the statistic, the computed value is z. Then the p-value is given by

$$\text{p-value} = P(|Z| \geq |z|) = 2F(-|z|).$$

If p-value $> \alpha$, do not reject H_0, otherwise reject H_0.



– In a statistical test of hypothesis, either rejection region or p-value approach can be used.



- Step 4, z should be computed from the statistic above based upon the observed sample.

- Step 5, the conclusion

  – If rejection region approach is used, then check whether z is located within rejection region. If so, reject H_0, otherwise do not reject H_0.

  – If p-value approach is used, then compare the p-value with α. If p-value $\leq \alpha$, then reject H_0, otherwise, do not reject H_0.

Example: Midterm Exam Score

- Recall the midterm exam scores example in our Topic 6.
- The data he obtained are 20, 19, 24, 22, 25.
- If the lecturer announced that the variance of the exam score over the class is 5 (just believe that this is the truth).
- Remember that the exam scores can be well approximately by a normal distribution.
- The student used the five steps to test whether the average midterm score is not 16.

- Step 1: H_0: μ=16 versus H_1: $\mu \neq 16$
- Step 2: choose $\alpha = 0.01$.
- Step 3: Now that $\sigma = \sqrt{5}$ is known, data are normal, n=5. Therefore test statistic and its distribution is

$$Z = \frac{\bar{X} - 16}{\sqrt{5}/\sqrt{n}} \sim N(0,1)$$

rejection region: $|Z| > z_{\alpha/2}$, that is $|z| > 2.576$.

- Step 4: $z = (22 - 16)/(\sqrt{5}/\sqrt{5}) = 6$.
- Step 5: 6 is located within rejection region, therefore H_0 is rejected. (Or p-value $= 0.0000006 < \alpha$, reject H_0)

Steps for one-sided test

- If we are interest in whether μ is larger than μ_0 or not. i.e. we are interested in $\mu \leq \mu_0$ versus $\mu > \mu_0$. Therefore, we write H_0: $\mu = \mu_0$ versus H_1: $\mu > \mu_0$.

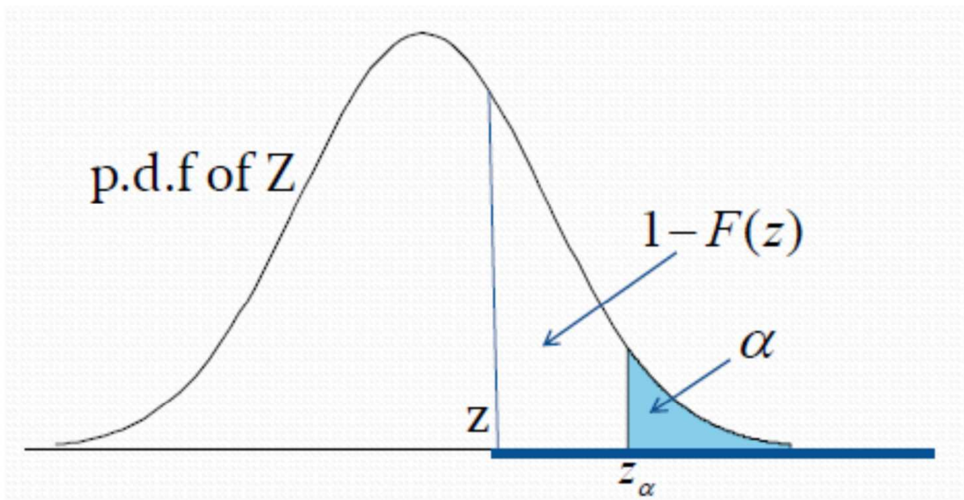
- Similar steps can be used to address this problem, we only need to do the following replacement:

  – Step 1, H_1 is replace with H_1: $\mu > \mu_0$.

  – Step 2, No need to change.

  – Step 3: test statistic and its distribution are kept the same.

- $*$ If rejection region is used, it should be replaced with $z > z_\alpha$, since now, we should reject only when $\bar{x}$ is large.
- $*$ If p-value is used, p-value $=$ 1-F(z), i.e. only the right-size tail should be used.



– Step 4: no need to change.

– Step 5: no need to change.

- There is another possibility in testing the mean: H_0: $\mu = \mu_0$ versus H_1: $\mu < \mu_0$. Similar steps can be established.

- A summary: for H_0: $\mu = \mu_0$ and different H_1. At the level of significance α. Test statistic and its distribution is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

H1	Rejection Region	p-value
$\mu > \mu_0$	$z > z_\alpha$	1-F(z)
$\mu < \mu_0$	$z < -z_\alpha$	F(z)
$\mu \neq \mu_0$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	2F(-|z|)

Example: Midterm Exam Score

- The student used the five steps to test whether the average midterm score is higher than 16.
- Step 1: H$_0$: μ=16 versus H$_1$: $\mu > 16$
- Step 2: choose $\alpha = 0.01$.
- Step 3: Now that $\sigma = \sqrt{5}$ is known, data are normal, n=5. Therefore test statistic and its distribution is

$$Z = \frac{\bar{X} - 16}{\sqrt{5}/\sqrt{n}} \sim N(0,1)$$

rejection region: $Z > z_\alpha$, that is $z > 2.33$.

- Step 4: $z = (22 - 16)/(\sqrt{5}/\sqrt{5}) = 6.$

- Step 5: 6 is located within rejection region, therefore H_0 is rejected. (Or p-value $= 0.0000003 < \alpha$, reject H_0)