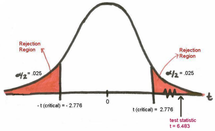
Pea-value versus *p*-value: An overview of teaching and learning of H2 Statistics

Ast/P Ho Weng Kin
MME, NIE

18 Oct 2012



Outline

Outline

Outline

1. Subtopics
2. Rare-event approach and empirical rule
3. Hypothesis testing

Outline

Subtopics to be covered

1. Summary of problems we face

Subtopics to be covered

1. Summary of problems we face
2. The theory of hypothesis testing

Subtopics to be covered

1. Summary of problems we face
2. The theory of hypothesis testing
3. Examples of statistical testing

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

1. Lack of content knowledge in statistics

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

1. Lack of content knowledge in statistics
2. Lack of coherence/justification in the syllabus

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

1. Lack of content knowledge in statistics
2. Lack of coherence/justification in the syllabus
3. Lack of time

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

1. Lack of content knowledge in statistics
2. Lack of coherence/justification in the syllabus
3. Lack of time
4. Lack of real statistic experience/sources

Summary of problems we face

In teaching H2 Level Statistics, we encounter the following problems:

1. Lack of content knowledge in statistics
2. Lack of coherence/justification in the syllabus
3. Lack of time
4. Lack of real statistic experience/sources
5. Lack of authentic examples

Pre-requisites

We assume certain facts from sampling and estimation theories in this workshop.

A lucky draw call

Lucky draw call

One day on your way back home, you received a phone call congratulating you for winning the first prize of a grand draw. To get the money deposited into your bank, you must first put $100,000 into a foreign account. What would you do?

A lucky draw call

Your motto in life ...

Where got so lucky?

A suspicious taxi-uncle

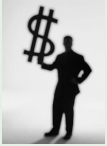
Sure your meter not spoiled?

A taxi-driver (a.k.a. taxi uncle) believes that his monthly income from driving his cab is low as a result of a mechanical fault in his meter.

A suspicious taxi-uncle

Sure your meter not spoiled?

To substantiate his belief, he collects information on the salaries of his counterparts, i.e., those taxi-driver friends who drive about the same duration per day, and with about the length of driving experience.



A suspicious taxi-uncle

Sure your meter not spoiled?

He finds that their monthly salaries have a mean of \$3,900 and a standard deviation of \$200. His monthly salary is about \$3,200. Does the information support his claim that his meter is spoiled?

A suspicious taxi-uncle

Sure your meter not spoiled?

He finds that their monthly salaries have a mean of \$3,900 and a standard deviation of \$200. His monthly salary is about \$3,200. Does the information support his claim that his meter is spoiled?

Discussion (5 min)

Please advise this taxi-uncle using your understanding of statistics.

Statistical inference principle

The preceding example exemplifies an approach to statistical inference called the *rare-event approach*.

Statistical inference principle

Definition (Rare-event approach)

1. An experimenter hypothesizes a specific frequency distribution to describe a population of measurements.

Statistical inference principle

Definition (Rare-event approach)

1. An experimenter hypothesizes a specific frequency distribution to describe a population of measurements.
2. Then a sample of measurements is drawn from the population.

Statistical inference principle

Definition (Rare-event approach)

1. An experimenter hypothesizes a specific frequency distribution to describe a population of measurements.
2. Then a sample of measurements is drawn from the population.
3. If the experimenter finds it unlikely that the sample came from the hypothesized distribution, the hypothesis is concluded to be false.

Rare-event approach

Solution.

First, we calculate the z-score for the taxi-driver's salary with respect to those of his counterparts.

Rare-event approach

Solution.

First, we calculate the z-score for the taxi-driver's salary with respect to those of his counterparts. Thus,

$$z = \frac{3,200 - 3,900}{200} = -3.5$$

Rare-event approach

Solution.

First, we calculate the z-score for the taxi-driver's salary with respect to those of his counterparts. Thus,

$$z = \frac{3,200 - 3,900}{200} = -3.5$$

This implies that his salary is 3.5 standard deviations below the mean of the salary distribution. $\square$

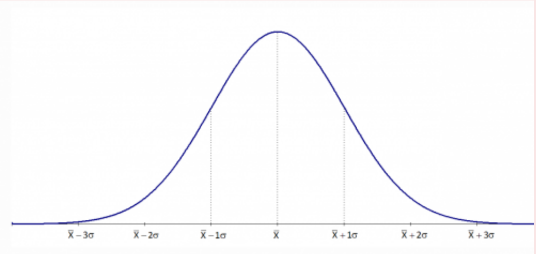
Distribution in question

To make further inference, we must know the nature of the distribution of monthly salaries for taxi-drivers.

Distribution in question

Assumption

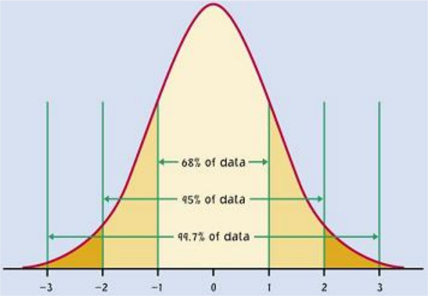
It is reasonable to assume that the frequency distribution of monthly salaries for taxi-drivers is mound-shape and symmetric, as shown below:



Empirical rule

Empirical rule

The *empirical rule* is a rule of the thumb that applies to data sets with frequency distributions that are mound-shaped and symmetrical, as shown below.



The suspicious taxi-driver revisited

Discussion (10 min)

Recall that our dear taxi-uncle's monthly salary is -3.5 standard deviations below the mean of the hypothesized salary distribution.

The suspicious taxi-driver revisited

Discussion (10 min)

Recall that our dear taxi-uncle's monthly salary is -3.5 standard deviations below the mean of the hypothesized salary distribution.

1. Decide what the hypothesized salary distribution in this scenario means.

The suspicious taxi-driver revisited

Discussion (10 min)

Recall that our dear taxi-uncle's monthly salary is -3.5 standard deviations below the mean of the hypothesized salary distribution.

1. Decide what the hypothesized salary distribution in this scenario means.
2. Using the empirical rule, comment on the likelihood that the monthly salary of a randomly selected taxi-driver having the salary of this particular taxi-uncle, i.e., \$3,200.

The suspicious taxi-driver revisited

Discussion (10 min)

Recall that our dear taxi-uncle's monthly salary is -3.5 standard deviations below the mean of the hypothesized salary distribution.

1. Decide what the hypothesized salary distribution in this scenario means.
2. Using the empirical rule, comment on the likelihood that the monthly salary of a randomly selected taxi-driver having the salary of this particular taxi-uncle, i.e., \$3,200.
3. Does the taxi-uncle have sufficient evidence to believe that his meter is spoiled?

Books check-out rate

Case study (5 min)

A city librarian claims that books have been checked out an average of 7 (or more) times in last year.

Books check-out rate

Case study (5 min)

You suspect that he has exaggerated the book check-out rate and that the mean number of checkouts per book per year is, in fact, less than 7. Using the computerized card catalog, you randomly select one book and find out that it has been checked out 4 times in last year. Assume that the standard deviation of the number of check-outs per book per year is approximately 1. Do you have reason to believe that the librarian's claim is incorrect?

Manufacture standards

Strength of sewer pipes

Suppose building specifications in a certain city require that the average breaking strength of residential sewer pipe be more than $3,600$ kg/m of length.

I used to play here as a kid, but of course, that was before Humans built that sewer pipe...

Manufacture standards

Strength of sewer pipes

Each manufacturer who wants to sell sewer pipes in this city must demonstrate that its product meets this specification.

Manufacture standards

Strength of sewer pipes

Each manufacturer who wants to sell sewer pipes in this city must demonstrate that its product meets this specification.
We are interested in making an inference about the mean μ of a population.

Manufacture standards

Strength of sewer pipes

Each manufacturer who wants to sell sewer pipes in this city must demonstrate that its product meets this specification.
We are interested in making an inference about the mean μ of a population.

Discussion (5 min)

What exactly is the population in this context?

Hypotheses

We are less interested in estimating the value of μ than we are in testing a *hypothesis*.

Hypotheses

We are less interested in estimating the value of μ than we are in testing a *hypothesis*.

We want to decide whether:

Hypothesis

The mean breaking strength of the pipe exceeds $3,600$ kg/m.

Rare-event concept

The method we use to reach a decision is based on the

Rare-event Concept

.

Null and alternative hypotheses

We define two hypotheses:

Null and alternative hypotheses

We define two hypotheses:

Definition (Null hypothesis)

The *null hypothesis* is that which represents status-quo to the party performing the sampling experiment – the hypothesis that will be accepted unless the data provide convincing evidence that it is false.

Null and alternative hypotheses

We define two hypotheses:

Null and alternative hypotheses

We define two hypotheses:

Definition (Alternative (or research) hypothesis)

The *alternative hypothesis* is that which will be accepted only if the data provide convincing evidence of its truth.

Null and alternative hypotheses

In this context, from the point of view of the city conducting the tests, the

Null and alternative hypotheses

In this context, from the point of view of the city conducting the tests, the

null hypothesis is ...

the manufacturer's pipe *does not* meet the specification

Null and alternative hypotheses

In this context, from the point of view of the city conducting the tests, the

null hypothesis is ...

the manufacturer's pipe *does not* meet the specification

and the

Null and alternative hypotheses

In this context, from the point of view of the city conducting the tests, the

null hypothesis is ...

the manufacturer's pipe *does not* meet the specification

and the

alternative hypothesis is ...

the manufacturer's pipe *does* meet the specification.

Null and alternative hypotheses

In this context, from the point of view of the city conducting the tests, the

null hypothesis is ...

the manufacturer's pipe *does not* meet the specification

and the

alternative hypothesis is ...

the manufacturer's pipe *does* meet the specification.

Discussion (5 min)

Why is it not the other way round?

Null and alternative hypotheses

Thus, the null and alternative hypotheses are therefore

Null and alternative hypotheses

Thus, the null and alternative hypotheses are therefore

- Null hypothesis (H_0): $\mu \leq 3,600$
  (i.e., manufacturer's pipe does not meet the specifications)

Null and alternative hypotheses

Thus, the null and alternative hypotheses are therefore

- Null hypothesis (H_0): $\mu \leq 3,600$
  (i.e., manufacturer's pipe does not meet the specifications)
- Alternative hypothesis (H_1): $\mu > 3,600$
  (i.e., manufacturer's pipe meets the specifications)

Null and alternative hypotheses

Thus, the null and alternative hypotheses are therefore

- Null hypothesis (H_0): $\mu \leq 3,600$
  (i.e., manufacturer's pipe does not meet the specifications)
- Alternative hypothesis (H_1): $\mu > 3,600$
  (i.e., manufacturer's pipe meets the specifications)

Discussion (5 min)

Now voice out your opinions on the above set-up.

Null and alternative hypotheses

Statistical idea

- Null hypotheses are usually specified as equalities, such as $\mu = 3,600$.

Null and alternative hypotheses

Statistical idea

- Null hypotheses are usually specified as equalities, such as $\mu = 3,600$.
- Even when the null hypothesis is an inequality, such as $\mu \leq 3,600$, we specify $H_0 : \mu = 3,600$, reasoning that ...

Null and alternative hypotheses

Statistical idea

If sufficient evidence exists to show that $H_1 : \mu > 3,600$ is true when tested against $H_0 : \mu = 3,600$, then surely sufficient evidence exists to reject the hypothesis that $\mu < 3,600$ as well.

Null and alternative hypotheses

Theorem

The null hypothesis H_0 is specified

- *as the value of μ closest to a one-sided alternative hypothesis; and*

Null and alternative hypotheses

Theorem

The null hypothesis H_0 is specified

- *as the value of μ closest to a one-sided alternative hypothesis; and*
- *as the only value not specified in a two-tailed alternative hypothesis.*

Null and alternative hypotheses

Theorem

The null hypothesis H_0 is specified

- *as the value of μ closest to a one-sided alternative hypothesis; and*
- *as the only value not specified in a two-tailed alternative hypothesis.*

Under H_0, this value of μ is denoted by μ_0.

Null and alternative hypotheses

The above theorem has an implication on our own mathematical pedagogical content knowledge concerning hypothesis testing.

Null and alternative hypotheses

The above theorem has an implication on our own mathematical pedagogical content knowledge concerning hypothesis testing.

Note

Henceforth, we should not just meaninglessly equate the term *null hypothesis* to be the one marked an equality.

The equality is a consequence of the preceding considerations.

Null and alternative hypotheses

Let us consolidate our learning here:

Null and alternative hypotheses

Let us consolidate our learning here:

Definition (Null hypothesis)

The *null hypothesis* is that will be accepted unless the data provide convincing evidence that it is false.

Null and alternative hypotheses

It can also be put in a more conservative tone:

Null and alternative hypotheses

It can also be put in a more conservative tone:

Definition (Null hypothesis)

The *null hypothesis* is that will not be rejected so long as the data does not provide convincing evidence that it is false.

Null and alternative hypotheses

Question

How does the city decide when enough evidence exists to conclude that the manufacturer's pipe meets the specification?

Null and alternative hypotheses

Question

How does the city decide when enough evidence exists to conclude that the manufacturer's pipe meets the specification?

Answer.

Since the hypothesis concerns the population mean μ, it is reasonable to use the sample mean $\overline{x}$ to make the inference. □

Test statistic

The city will conclude that the pipe meets specifications only when the sample mean $\overline{x}$ *convincingly* indicates that the population mean exceeds $3,600$ kg/m.

Test statistic

The city will conclude that the pipe meets specifications only when the sample mean $\overline{x}$ *convincingly* indicates that the population mean exceeds $3,600$ kg/m.

Statistical idea

Convincing evidence in favor of the alternative hypothesis will exist when the value of $\overline{x}$ exceeds $3,600$ by an amount that cannot be attributed to sampling variability.

Test statistic

To decide, we must compute a *test statistic*, i.e., the z-value that measures the distance between the value of $\overline{x}$ and the value of μ specified in the null hypothesis.

Test statistic

The test statistic is

$$z = \frac{\overline{x} - 3,600}{\frac{\sigma}{\sqrt{n}}}.$$

Test statistic

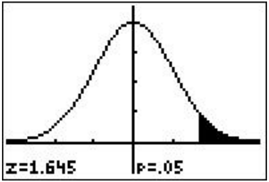
The test statistic is

$$z = \frac{\overline{x} - 3,600}{\frac{\sigma}{\sqrt{n}}}.$$

Implications

$z =$	implies that $\overline{x}$ is ... s.d. above $\mu = 3,600$
1	1
1.5	1.5

Rare-event concept



Assuming in fact that H_0 is true, i.e., that $\mu = 3,600$, the chance of observing 1.645 s.d. above $\mu = 3,600$ is only 0.05.

Rare-event concept

If the sample mean is more than 1.645 s.d. above $\mu = 3,600$, then either

- H_0 is true and a relatively rare event has occurred (with a 0.05 probability), or

Rare-event concept

If the sample mean is more than 1.645 s.d. above $\mu = 3,600$, then either

- H_0 is true and a relatively rare event has occurred (with a 0.05 probability), or
- H_1 is true and the population mean μ exceeds $3,600$.

Rare-event concept

If the sample mean is more than 1.645 s.d. above $\mu = 3,600$, then either

- H_0 is true and a relatively rare event has occurred (with a 0.05 probability), or
- H_1 is true and the population mean μ exceeds $3,600$.

Statistical idea: Where got so lucky?

The rare event approach bears upon us to reject the notion that a rare event has occurred, and thus we would reject H_0 in favour of H_1.

Rare-event concept

Discussion (5 min)

- Is there a chance that a decision making based on hypothesis testing can lead to wrong conclusions?

Rare-event concept

Discussion (5 min)

- Is there a chance that a decision making based on hypothesis testing can lead to wrong conclusions?
- What are the two kinds of wrong conclusions that can arise?

Type I decision error

Definition (Type I error)

An incorrect decision of rejecting the null hypothesis is false when in fact it is true is called a *Type I error*.

Type II decision error

Definition (Type II error)

An incorrect decision of accepting the null hypothesis as true when in fact it is false is called a *Type II error*.

Decision errors

The risk of making a

- Type I error is denoted by the symbol α; and

Decision errors

The risk of making a

- Type I error is denoted by the symbol α; and
- Type II error is denoted by the symbol β.

Decision errors

Question

How much risk I am ready to bear for rejecting the null hypothesis when in fact it is true?

Decision errors

Question

How much risk I am ready to bear for rejecting the null hypothesis when in fact it is true?

This risk is thus set before I begin the testing, i.e., even before I collect my data.

Type I decision error

Definition (Level of significance)

Note that

$$\begin{aligned} \alpha\% &= \text{P(Type I error)} \\ &= \text{P(Reject } H_0 \mid H_0 \text{ is in fact true)} \end{aligned}$$

is called the *level of significance* of the test.

Significance related to confidence

The smaller the level of significance, the higher the confidence we put in the test procedure.

Decision errors

In contrast, it is often difficult to determine precisely the risk of a Type II error. It depends on α as well as the particular value of μ proposed by an alternative hypothesis.

Decision errors

In contrast, it is often difficult to determine precisely the risk of a Type II error. It depends on α as well as the particular value of μ proposed by an alternative hypothesis.

Central idea about rejecting and acceptance

1. When the test statistic falls in the rejection region and we make a decision to reject the null hypothesis, we do so knowing the error rate for incorrect rejections of H_0.

Decision errors

In contrast, it is often difficult to determine precisely the risk of a Type II error. It depends on α as well as the particular value of μ proposed by an alternative hypothesis.

Central idea about rejecting and acceptance

1. When the test statistic falls in the rejection region and we make a decision to reject the null hypothesis, we do so knowing the error rate for incorrect rejections of H_0.
2. The situation corresponding to accepting the null hypothesis, and thereby risking a Type II error, is not as generally as controllable.

Policy in hypothesis testing

For the above reasons, we adopt the following policy:

Policy in hypothesis testing

For the above reasons, we adopt the following policy:

Policy

We do not reject H_0 when the test statistic does not fall in the rejection region, rather than risking an error of an unknown magnitude.

Policy in hypothesis testing

For the above reasons, we adopt the following policy:

Policy

We do not reject H_0 when the test statistic does not fall in the rejection region, rather than risking an error of an unknown magnitude.

The reason we can substantiate is that the data collected presents *insufficient evidence* for us to reject H_0.

Determining β

Discussion (10 min)

Let us begin with the set-up of the test as follows:

$H_0 : \mu = 3,600 \qquad H_1 : \mu > 3,600$

Test statistic: $z = \frac{\overline{x} - 3{,}600}{\sigma/\sqrt{n}}$,

where $\sigma \approx s = 200$ and $n = 50$.

Level of significance: 5%.

Rejection region: $z > 1.645$.

Given that the alternative distribution has the mean $\mu_a = 3,625$ and the same variance as before, find the corresponding value of β.

Determining β

Discussion (10 min)

Let us begin with the set-up of the test as follows:

$H_0 : \mu = 3,600 \qquad H_1 : \mu > 3,600$

Test statistic: $z = \frac{\overline{x} - 3{,}600}{\sigma/\sqrt{n}}$,

where $\sigma \approx s = 200$ and $n = 50$.

Level of significance: 5%.

Rejection region: $z > 1.645$.

Given that the alternative distribution has the mean $\mu_a = 3,625$ and the same variance as before, find the corresponding value of β.

Remember that

$$\begin{aligned} \beta\% &= \text{P(Type II error)} \\ &= \text{P(Accept } H_0 \mid H_0 \text{ is in fact false)} \end{aligned}$$

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Null hypothesis* (H_0): A theory about the values of one or more population parameters; it generally represents status-quo that we are not rejecting until proven false.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Null hypothesis* (H_0): A theory about the values of one or more population parameters; it generally represents status-quo that we are not rejecting until proven false.
2. *Alternative hypothesis* (H_1): A theory that contradicts the null hypothesis; it generally represents that which we will accept only when sufficient evidence exists to establish its truth.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Test statistic*: A sample statistic used to decide whether to reject the null hypothesis.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Test statistic*: A sample statistic used to decide whether to reject the null hypothesis.
2. *Rejection region*: The numerical values of the test statistic for which the null hypothesis will be rejected. The rejection region is chosen so that the probability is α that it will contain the test statistic when the null hypothesis is true, thereby leading to a Type I error. This value α is usually chosen to be small and is referred to as the *level of significance* of the test.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Assumptions*: Clear statement(s) of any assumptions made about the population(s) sampled.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Assumptions*: Clear statement(s) of any assumptions made about the population(s) sampled.
2. *Experiment and calculation of test statistic*: Performance of the sampling experience and determination of the numerical value of the test statistic.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Conclusion*:
   - If the numerical value of the test statistic falls in the rejection region, we reject the null hypothesis and conclude that the alternative hypothesis is true. We know that the hypothesis testing process will lead to this conclusion incorrectly (Type I error) only $\alpha\%$ of the time when in fact H_0 is true.

The elements of a test of hypothesis

We summarize all the elements of a test of hypothesis in point-form below.

Elements of hypothesis testing

1. *Conclusion*:
   - If the test statistic does not fall in the rejection region, we do not reject H_0. Thus, we reserve judgment about which hypothesis is true. We do not conclude that the null hypothesis is true because we do not (in general) know the probability $\beta\%$ that our test procedure will lead to an incorrect acceptance of H_0 (Type II error).

Sample parameter distribution

In any hypothesis testing, it is important to know the distribution function (or an approximation) of the sample parameter.

Sample parameter distribution

In any hypothesis testing, it is important to know the distribution function (or an approximation) of the sample parameter.

In what follows, we shall focus on the H2 Mathematics Syllabus, speaking about only the parameter of mean.

Large-sample test of hypothesis about μ

Drug and alcohol effect

The effect of drugs and alcohol on the nervous system has been the subject of considerable research recently. Suppose a research neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each to a neurological stimulus, and recording its response time.

Large-sample test of hypothesis about μ

Drug and alcohol effect

The neurologist knows that the mean response time for rats not injected with the drug (the "control" mean) is 1.2 seconds. She wishes to test whether the mean response time for drug-injected rats differs from 1.2 seconds. Set up the test of hypothesis for this experiment at a 1% level of significance. The sampling experiment is conducted with the following results:

$$\begin{aligned} n &= 100 \\ \overline{x} &= 1.05 \\ s &= 0.5 \end{aligned}$$

Large-sample test of hypothesis about μ

Since the neurologist wishes to detect whether the response time, μ, for drug-injected rats differs from the control mean of 1.2 seconds in *either* direction, i.e., $\mu < 1.2$ or $\mu > 1.2$, we conduct a *two-tailed statistical test*.

Large-sample test of hypothesis about μ

The parameter

Let X be the response time of a randomly selected rat.

Large-sample test of hypothesis about μ

The parameter

Let X be the response time of a randomly selected rat.
Denote by μ the mean of X.

Large-sample test of hypothesis about μ

The parameter

Let X be the response time of a randomly selected rat.
Denote by μ the mean of X.

The null hypothesis is the presumption that drug-injected rats have the same mean response as control rats unless the research indicates otherwise.

Large-sample test of hypothesis about μ

Thus,

Null and alternative hypotheses

H_0 : $\mu = 1.2$
H_1 : $\mu \neq 1.2$ (i.e., $\mu < 1.2$ or $\mu > 1.2$)

Large-sample test of hypothesis about μ

Test statistic

Under H_0, the test statistic is

$$Z = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0,1)$$

approximately by the **Central Limit Theorem** since

- the sample size $n > 100$ is large, and

Large-sample test of hypothesis about μ

Test statistic

Under H_0, the test statistic is

$$Z = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0,1)$$

approximately by the **Central Limit Theorem** since

- the sample size $n > 100$ is large, and
- the population variance σ^2 is unknown.

Large-sample test of hypothesis about μ

The rejection region must be designated to detect a departure from $\mu = 1.2$ in *either* direction, so we will reject H_0 for values of z that are either too small or too large.

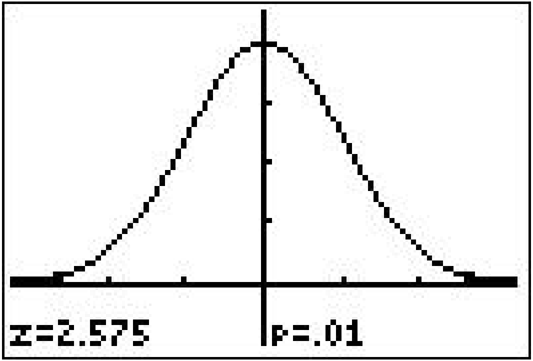
Large-sample test of hypothesis about μ

The rejection region must be designated to detect a departure from $\mu = 1.2$ in *either* direction, so we will reject H_0 for values of z that are either too small or too large.
For the selected level of 1%, one needs to divide this equally between the upper and the lower tails of the distribution of z, i.e., $\frac{\alpha}{2}\% = 2.5\%$ is to be placed at each tail.

Large-sample test of hypothesis about μ

Rejection region

$$z < -2.575 \quad \text{or} \quad z > 2.575$$



Large-sample test of hypothesis about μ

Assumption

- Since the sample size of the experiment is large enough ($n > 30$), the Central Limit Theorem will apply, and no assumption need be made about the population of response time measurements.

Large-sample test of hypothesis about μ

Assumption

- Since the sample size of the experiment is large enough ($n > 30$), the Central Limit Theorem will apply, and no assumption need be made about the population of response time measurements.
- The sampling distribution of the sample mean response of 100 rats will be approximately normal regardless of the distribution of the individual rats' response times.

Large-sample test of hypothesis about μ

Experiment and calculation of test statistics

We now substitute the sample statistics into the test statistic:

Large-sample test of hypothesis about μ

Experiment and calculation of test statistics

We now substitute the sample statistics into the test statistic:

$$z = \frac{\overline{x} - 1.2}{\frac{\sigma}{\sqrt{n}}} \approx \frac{\overline{x} - 1.2}{\frac{s}{\sqrt{n}}} = \frac{\overline{1.05} - 1.2}{\frac{0.5}{\sqrt{100}}} = -3.0$$

Large-sample test of hypothesis about μ

Experiment and calculation of test statistics

We now substitute the sample statistics into the test statistic:

$$z = \frac{\overline{x} - 1.2}{\frac{\sigma}{\sqrt{n}}} \approx \frac{\overline{x} - 1.2}{\frac{s}{\sqrt{n}}} = \frac{\overline{1.05} - 1.2}{\frac{0.5}{\sqrt{100}}} = -3.0$$

This implies that the sample mean -1.05, which is 3 s.d. below the null-hypothesized value of 1.2, is in the lower-tail rejection region, i.e., $z < -2.575$.

Large-sample test of hypothesis about μ

Conclusion

This sampling experiment provides sufficient evidence to reject H_0 and conclude, at 1% level of significance, that the mean response time for drug-injected rats differs from the control mean of 1.2 seconds.

Large-sample test of hypothesis about μ

Conclusion

This sampling experiment provides sufficient evidence to reject H_0 and conclude, at 1% level of significance, that the mean response time for drug-injected rats differs from the control mean of 1.2 seconds.
It appears that the rats receiving an injection of this drug does have a mean response time that is less than 1.2 seconds.

Large-sample test of hypothesis about μ

Remarks

Large-sample test of hypothesis about μ

Remarks

1. Notice that the sample statistic z is less than -2.575, it is very tempting to state our conclusion at a significant level lower than 1%.

Large-sample test of hypothesis about μ

Remarks

1. Notice that the sample statistic z is less than -2.575, it is very tempting to state our conclusion at a significant level lower than 1%.
2. We must resist this temptation because the level of α is determined before the sampling experiment is performed.

Large-sample test of hypothesis about μ

Remarks

1. Notice that the sample statistic z is less than -2.575, it is very tempting to state our conclusion at a significant level lower than 1%.
2. We must resist this temptation because the level of α is determined before the sampling experiment is performed.
3. In general, the same data should not be used both to set up and to conduct the test.

p-value as another approach

According to the preceding statistical test procedure,

p-value as another approach

According to the preceding statistical test procedure,

1. the rejection region and, correspondingly, the value of α are selected prior to conducting the test, and

p-value as another approach

According to the preceding statistical test procedure,

1. the rejection region and, correspondingly, the value of α are selected prior to conducting the test, and
2. the conclusions are stated in terms of rejecting or not rejecting the null hypothesis.

p-value as another approach

A second method of presenting the results of a statistical test is one that reports the extent to which the test statistic disagrees with the null hypothesis.

p-value

Definition

The *p-value* for a specific statistical test is the (conditional) probability (assuming that H_0 is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the actual one computed from the sample data.

p-value

Definition

The *p-value* for a specific statistical test is the (conditional) probability (assuming that H_0 is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the actual one computed from the sample data.

The p-value is also known as the *observed level of significance*.

p-value

Sewer pipes revisited

Let X be the strength of a randomly selected sewer pipe (in kg/m), and denote by μ the mean of X.

p-value

Sewer pipes revisited

Let X be the strength of a randomly selected sewer pipe (in kg/m), and denote by μ the mean of X.
$H_0 : \mu = 3,600$
$H_1 : \mu > 3,600$
Under H_0, the test statistic is

$$Z = \frac{\overline{X} - 3,600}{\frac{s}{\sqrt{n}}} \sim N(0,1),$$

approximately by the Central Limit Theorem, where $\overline{x} = 3660$, $n = 50$ and $s = 200$.

p-value

Sewer pipes revisited

The sample statistics put into the test statistic yield:

$$z = 2.12.$$

p-value

Sewer pipes revisited

The sample statistics put into the test statistic yield:

$$z = 2.12.$$

Since the test is one-tailed, values of the test statistic even more contradictory to H_0 than the one observed ($z = 2.12$) would be values larger than $z = 2.12$.

p-value

Sewer pipes revisited

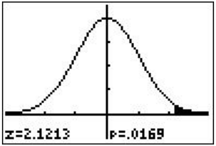
The sample statistics put into the test statistic yield:

$$z = 2.12.$$

Since the test is one-tailed, values of the test statistic even more contradictory to H_0 than the one observed ($z = 2.12$) would be values larger than $z = 2.12$. Thus, the p-value for this test is

$$p\text{-value} = \mathrm{P}(z \geq 2.12) = 0.0169.$$

p-value



Sewer pipes revisited

Because the *p*-value is 0.0169, it follows that

Significance level	Result
5%	Reject H_0
1%	Do not reject H_0

p-value

When publishing the results of a statistical test of hypothesis in journals, case studies, reports, etc., many researchers make use of p-values.

p-value

Instead of selecting the level of significance, $\alpha\%$ beforehand and then conducting a test, the researcher computes and reports the value of the appropriate test statistic and its associated p-value.

p-value

It is left to the reader of the report to judge the significance of the result, i.e., the reader must determine whether to reject the null hypothesis in favour of the alternative hypothesis, based on the reported p-value.

p-value

There are two advantages of the p-value approach:

p-value

There are two advantages of the *p*-value approach:

1. Readers are permitted to select the maximum value of α that they would be willing to tolerate if they actually carried out a standard test of hypothesis in the manner outlined earlier using the rejection region, and

p-value

There are two advantages of the *p*-value approach:

1. Readers are permitted to select the maximum value of α that they would be willing to tolerate if they actually carried out a standard test of hypothesis in the manner outlined earlier using the rejection region, and
2. a measure of the degree of significance of the result (i.e., the *p*-value) is provided.

p-value

Reporting test results as p-values:

p-value

Reporting test results as p-values:

1. Choose the maximum value of α that you are willing to tolerate.

p-value

Reporting test results as p-values:

1. Choose the maximum value of α that you are willing to tolerate.
2. If the p-value of the test is less than the chosen value of α, reject the null hypothesis. Otherwise, do not reject the null hypothesis.

p-value

If the probability of obtaining a test statistic which is at least as contradictory to H_0 as the one obtained from the sample data (got a lucky draw call) is lower than the risk of forgoing H_0 given that it is in fact true (how lucky you rate yourself), then you give up H_0 (hang up the call).

Examination question

N2006/II/25

The mass of vegetables in a randomly chosen bag has a normal distribution. The mass of the contents of a bag is supposed to be 10kg. A random sample of 80 bags is taken and the mass of the contents of each bag, x grams, is measured. The data are summarised by

$$\sum (x - 10,000) = -2510 \qquad \sum (x - 10,000)^2 = 2,010,203.$$

Test, at the 5% significance level, whether the mean mass of the contents of a bag is less than 10kg.
Explain, in the context of the question, the meaning of 'at the 5% significance level'.

Small-sample test of hypothesis about μ

We discuss this by using an example.

Small-sample test of hypothesis about μ

There are many reasons why we can only have small-samples, i.e., sample sizes $n < 30$.

Small-sample

The first issue is about the destructive nature of sampling.

Small-sample

The first issue is about the destructive nature of sampling.

Lifetime of components

It is of interest what the mean lifetime of a light bulb of a certain make is. To do so, we can light a sample of light bulbs, and burn these out, recording their lifetimes.

Small-sample

The first issue is about the destructive nature of sampling.

Lifetime of components

It is of interest what the mean lifetime of a light bulb of a certain make is. To do so, we can light a sample of light bulbs, and burn these out, recording their lifetimes. However, by doing so, one destroys the light bulbs sampled.

Small-sample

The first issue is about the destructive nature of sampling.

Lifetime of components

It is of interest what the mean lifetime of a light bulb of a certain make is. To do so, we can light a sample of light bulbs, and burn these out, recording their lifetimes. However, by doing so, one destroys the light bulbs sampled. Because of costing issues, one can only resort to small-sample testing.

Small-sample

The second issue is about the medical safety.

Small-sample

The second issue is about the medical safety.

Effects of new drug

A new cancer drug has been tested to be effective on mice. The next stage is to test it on human subjects.

Small-sample

The second issue is about the medical safety.

Effects of new drug

A new cancer drug has been tested to be effective on mice. The next stage is to test it on human subjects. For issues of medical ethics, it is impossible to test the new drug out on a large sample of human subjects.

Small-sample

The second issue is about the medical safety.

Effects of new drug

A new cancer drug has been tested to be effective on mice. The next stage is to test it on human subjects. For issues of medical ethics, it is impossible to test the new drug out on a large sample of human subjects. Thus, one is forced to use small-sample testing.

Small-sample

Generally, experiments that only can resort to small-sample testing are those of a special nature.

Small-sample

Generally, experiments that only can resort to small-sample testing are those of a special nature.

Carbon emission

Car manufacturers want to test a new engine to determine whether it meets pollution standards. It is destructive to the environment if a large number of cars is sampled for carbon emission since the very act of the test certainly pollutes the air.

Small-sample test of hypothesis of μ

Water treatment

Most water treatment facilities monitor the quality of their drinking water on an hourly basis. One variable monitored is pH, which measures the degree of alkalinity or acidity in the water. A pH below 7.0 is acidic, one above 7.0 is alkaline, and a pH of 7.0 is neutral.

Small-sample test of hypothesis of μ

Water treatment

The mean and standard deviation of 1 hour's test results, based on 17 water samples at this plant, are:

$$\overline{x} = 8.42, \quad s = 0.16.$$

Small-sample test of hypothesis of μ

Water treatment

The mean and standard deviation of 1 hour's test results, based on 17 water samples at this plant, are:

$$\overline{x} = 8.42, \quad s = 0.16.$$

Does this sample provide sufficient evidence that the mean pH level in the water differs from 8.5?

Small-sample test of hypothesis of μ

Let X be the pH value of a randomly selected water sample, and μ the mean of X.

Small-sample test of hypothesis of μ

Let X be the pH value of a randomly selected water sample, and μ the mean of X.

We establish the target pH as the null hypothesized value and then utilize a two-tailed alternative that the true mean pH differs from the target:

$$H_0 : \mu = 8.5$$
$$H_1 : \mu \neq 8.5$$

Small-sample test of hypothesis of μ

Problems

When we are faced with making inferences about a population mean using the information in a small sample.

1. The normality of the sample distribution for $\overline{x}$ does not follow from the Central Limit Theorem when the sample size is small.

Small-sample test of hypothesis of μ

Problems

When we are faced with making inferences about a population mean using the information in a small sample.

1. The normality of the sample distribution for $\overline{x}$ does not follow from the Central Limit Theorem when the sample size is small. We must assume that the distribution of measurements from which the sample was selected is approximately normally distributed in order to ensure the approximate normality of the sampling distribution of $\overline{x}$.

Small-sample test of hypothesis of μ

Problems

When we are faced with making inferences about a population mean using the information in a small sample.

2. If the population standard deviation σ is unknown, as is usually the case, then we cannot assume that s will provide a good approximation for σ when the sample size is small.

Small-sample test of hypothesis of μ

Problems

When we are faced with making inferences about a population mean using the information in a small sample.

2. If the population standard deviation σ is unknown, as is usually the case, then we cannot assume that s will provide a good approximation for σ when the sample size is small. Instead, we must use the t-distribution rather than the standard normal z-distribution to make inferences about the population mean μ.

Student t-distribution

Figure: William Sealy Gossett (1876–1937)

Student t-distribution

Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$.

Student t-distribution

Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$. Then, the quantity

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim T_{\nu=n-1},$$

where $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$.

Student t-distribution

For the degree of freedom ν, the distribution function of T_ν is given by

$$f_\nu(t) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}}.$$

Student t-distribution

For $\nu = 1, 2$, the distribution curves of T_ν are given below:

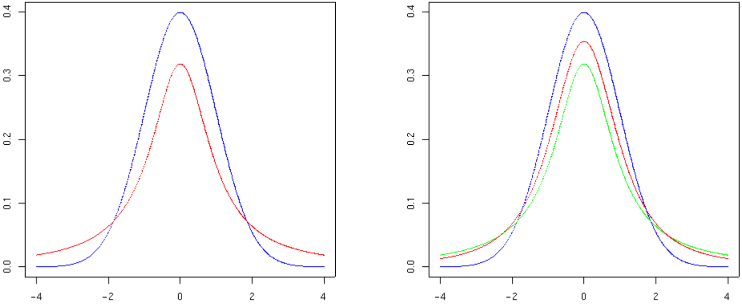


Figure: Blue: normal curve

Student t-distribution

For $\nu = 5, 30$, the distribution curves of T_ν are given below:

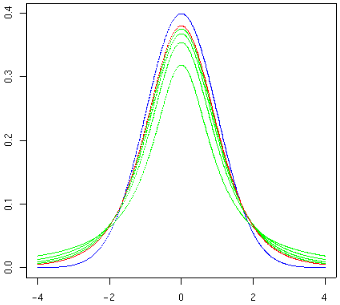


Figure: Blue: normal curve

Student t-distribution

Notice that when $n = 30$, i.e., $nu = n - 1 = 29$,

$$T_{29} \approx N(0, 1).$$

Student t-distribution

Notice that when $n = 30$, i.e., $nu = n - 1 = 29$,

$$T_{29} \approx N(0, 1).$$

Discussion (5 min)

What is the implication of the above observation?

Student t-distribution

Notice that when $n = 30$, i.e., $nu = n - 1 = 29$,

$$T_{29} \approx N(0, 1).$$

Discussion (5 min)

What is the implication of the above observation?

In fact,

$$\lim_{\nu \to \infty} T_\nu = N(0, 1).$$

Small-sample test of hypothesis about μ

Test statistic

Under H_0,

$$T = \frac{\overline{X} - 8.5}{S/\sqrt{n}} \sim T_{n-1},$$

where $n = 17$.

Small-sample test of hypothesis about μ

Level of significance

$\alpha\% = 5\%$.

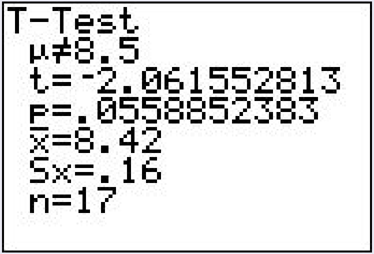
Small-sample test of hypothesis about μ

Level of significance

$\alpha\% = 5\%$.

Sample statistic

Given that $\overline{x} = 8.42$, and $s = 0.16$,



Small-sample test of hypothesis about μ

Conclusion

Since p-value ($p = 0.0559$) is more than the level of significance of 0.05, we do not have sufficient evidence to reject H_0.
Thus, the water treatment plant should not conclude that the mean pH differs from the 8.5 target based on the sample evidence.

Small-sample test of hypothesis about μ

Discussion (5 min)

Carry out, instead, a Z-test for the above experiment.

Small-sample test of hypothesis about μ

Discussion (5 min)

Carry out, instead, a Z-test for the above experiment.
What is your conclusion now?
Comment about this *incorrect* use of Z-test.

Examination question

N2009/II/10

A company supplis sugar in small packets. The mass of sugar in one packet is denoted by X grams. The masses of a random sample of 9 packets are summarised by

$$\sum x = 86.4 \qquad \sum x^2 = 835.92.$$

(i) Calculate the unbiased estimates of the mean and variance of X.

Examination question

N2009/II/10

The mean mass of sugar in a packet is claimed to be 10 grams. The company directors want to know whether the sample indicates that this claim is incorrect.

(ii) Stating a necessary assumption, carry out a t-test at 5% significance level. Explain why the Central Limit Theorem does not apply in this context.

(iii) Suppose now that the population variance of X is known, and the assumption made in part (ii) is still valid. What change would there be in carrying out the test?

References

1. McClave, Dietrich & Sincich, *Statistics: Annotated Instructor's Edition*, 7th Ed, Prentice Hall, 1997.
2. Johnson & Bhattacharyya, *Statistics: Principles and methods*, John Wiley & Sons, 1985.
3. Larsen & Marx, *Statistics*, Prentice Hall, 1990.