Hypothesis Testing – Means and Proportions

8.1, 8.3

Today’s topics

Hypothesis Testing

- Definitions
- Testing for one mean
- p-value
- Testing for one proportion

Statistics overview

One goal in Statistics is to make *inferences* about populations based on samples taken from the population.

Previously, we estimated population parameters:

- Point estimates (MLE, MOM)
- Interval estimates (Confidence Intervals)

notes

Testing

Another way to do inference, is to **make a decision** about a parameter.

Examples:

my Mazda 3 manual claims that it gets 35 highway mpg

Dustin’s pudding packs actually contain 3.25 oz

Terms

- Null hypothesis, H_0
- Alternative hypothesis, H_A or H_1
- Type I error:
- Type II error
- Simple hypothesis
- Compound hypothesis

Null and Alternative Hypotheses

Say an experimenter wants to test the plausibility of the statement $\mu = \mu_0$. a fixed value e.g. 50

We can formally describe this as a **null hypothesis.**

- **H_0 :** $\mu = \mu_0$
- The word "hypothesis" indicates that we will be testing this statement (with data).

We will associate the null hypothesis with a different one that we are testing 'for', called the "alternative hypothesis".

- **H_A :** $\mu \neq \mu_0$ or **H_A :** $\mu > \mu_0$ or **H_A :** $\mu < \mu_0$

notes

Type 1 and Type 2 Error

	Do not reject H_0	Reject H_0
H_0 true	Correct	Type I Error
H_0 not true	Type II Error	Correct

Hypothesis Test Example – “Compound” H_A

Perdont university claims that students at their school are above average intelligence. A random sample of thirty students IQ scores have a mean score of **102.5**.

Suppose the mean population IQ score is 100 with a standard deviation of 15. Is there sufficient evidence to support this claim?

H_0 :

H_A :

notes

Hypothesis Test Example – "Simple" H_A

Example 8.1-1 Let X equal the breaking strength of a steel bar. If the bar is manufactured by process I, X is $N(50, 36)$, i.e., X is normally distributed with $\mu = 50$ and $\sigma^2 = 36$. It is hoped that if process II (a new process) is used, X will be $N(55, 36)$. Given a large number of steel bars manufactured by process II, how could we test whether the five-unit increase in the mean breaking strength was realized?

H_0: $\mu = 50$ H_1: $\mu = 55$

We want to set up a "rule" to determine whether to stick with H_0 or not. This rule will lead to a decision about what to do with H_0.

Partition sample space into 2 parts: C and C'.

If $(x_1, x_2, \ldots x_n) \in C$, reject H_0

If $(x_1, x_2, \ldots x_n) \in C'$, do not reject H_0

e.g. $\mathbf{x} = \{64.4, 54.7, 57.2, 61.6, 51.3\}$ or $\mathbf{x} = \{51.2, 54.7, 47.2, 51.6, 46.3\}$

We often partition the sample space in terms of values of a statistic called a **test statistic.**

Test statistic

Often, we partition the sample space based on the value of a statistic called the **test statistic**.

One common example is $\bar{X}$ (for testing the mean).

We might want to reject the null hypothesis if the sample average is larger or smaller than a certain number. E.g. $\bar{X} > 53$.

i.e. $C = \{(x_1, x_2, \ldots x_n) : \bar{x} > 53\}$

C is referred to as the **rejection region**, or the **critical region.**

p value

The plausibility of a null hypothesis can be measured with a **p-value,** which is a number between 0 and 1.

- A p-value is sometimes referred to as the *observed level of significance*
- The smaller the p-value, the less plausible H_0 is.

Definition of a p-value:

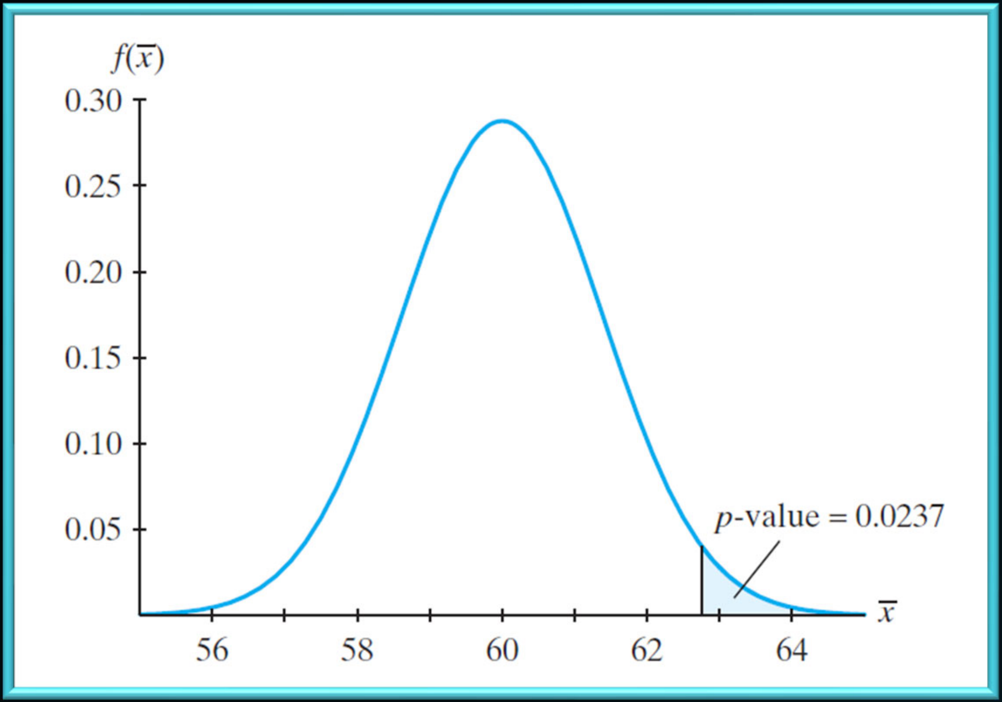
"Probability of observing data at least as extreme as the observed sample given that H_0 is true."

p-value illustration

H_0: $\mu = 60$ vs

H_A: $\mu > 60$

Suppose $\bar{x}$ = 62.75



Hypothesis Testing (Steps)

1. Formulate H_0 and H_A (based on the scenario)
2. Identify a test statistic to use and its distribution under H_0
3. Evaluate the test statistic
4. Calculate a p-value, compare to α.
5. Make a decision
   - if $p < \alpha$, reject H_0. Otherwise, (if $p > \alpha$), do not reject H_0.
6. State conclusion **in the context of the original question.**
   - "There is/isn't enough evidence to show that…"

Hypothesis Test Example – "Compound" H_A

Perdont university claims that students at their school are above average intelligence. A random sample of thirty students IQ scores have a mean score of **102.5**.

Suppose the mean population IQ score is 100 with a standard deviation of 15. Is there sufficient evidence to support this claim?

H_0 :

H_A :

Two ways to perform a hypothesis test

1. Calculate a p-value and compare to significance level

2. Define a rejection rejection (RR) and see if sample falls in RR. (also known a critical region)

Table 8.1-1 Tests of hypotheses about one mean, variance known

H_0	H_1	Critical Region
$\mu = \mu_0$	$\mu > \mu_0$	$z \geq z_\alpha$ or $\overline{x} \geq \mu_0 + z_\alpha \sigma/\sqrt{n}$
$\mu = \mu_0$	$\mu < \mu_0$	$z \leq -z_\alpha$ or $\overline{x} \leq \mu_0 - z_\alpha \sigma/\sqrt{n}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z\rvert \geq z_{\alpha/2}$ or $\lvert\overline{x} - \mu_0\rvert \geq z_{\alpha/2} \sigma/\sqrt{n}$

$$Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Table 8.1-2 Tests of hypotheses for one mean, variance unknown

H_0	H_1	Critical Region
$\mu = \mu_0$	$\mu > \mu_0$	$t \geq t_\alpha(n-1)$ or $\overline{x} \geq \mu_0 + t_\alpha(n-1)s/\sqrt{n}$
$\mu = \mu_0$	$\mu < \mu_0$	$t \leq -t_\alpha(n-1)$ or $\overline{x} \leq \mu_0 - t_\alpha(n-1)s/\sqrt{n}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t\rvert \geq t_{\alpha/2}(n-1)$ or $\lvert\overline{x} - \mu_0\rvert \geq t_{\alpha/2}(n-1)s/\sqrt{n}$

$$T = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

Example

A machine that pours beer is supposed to fill buckets with a mean of 3.0 gallons. A sample of 13 buckets gives a sample mean of 2.879 gallons and s = 0.325. Perform a hypothesis test at α= 0.05 to see if this machine is **accurately doing its job**.

H_0:

H_A:

t = ~t_{12}

p-value =

Decision:

Conclusion:

Example

A HAL-8000 machine that pours beer is supposed to fill buckets with a mean of 3.0 gallons. Dave takes a sample of 13 buckets and finds a sample mean of 2.879. Suppose the true standard deviation of these machines is 0.2 gallons. Perform a hypothesis test at α= 0.05 to see if this machine is **underfilling**.

H_0:

H_A:

z =

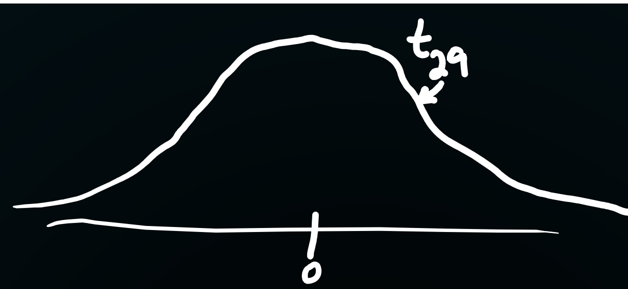
p-value =

Decision:

Conclusion:

$$T = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

Example



Nick Fury claims that the (true) mean number of push-ups his superheroes can do is at least 40.0. A random sample of 30 superheroes gives $\bar{x} = 38.518$ pushups, s = 2.299. Perform a hypothesis test at α = 0.01 to determine if this is true (or if they can't really make it to 40 pushups).

H_0:

H_A:

t = $\sim t_{29}$

p-value:

Decision:

Conclusion:

Table 8.3-1 Tests of hypotheses for one proportion

H_0	H_1	Critical Region
$p = p_0$	$p > p_0$	$z = \dfrac{y/n - p_0}{\sqrt{p_0(1-p_0)/n}} \ge z_\alpha$
$p = p_0$	$p < p_0$	$z = \dfrac{y/n - p_0}{\sqrt{p_0(1-p_0)/n}} \le -z_\alpha$
$p = p_0$	$p \ne p_0$	$\lvert z \rvert = \dfrac{\lvert y/n - p_0 \rvert}{\sqrt{p_0(1-p_0)/n}} \ge z_{\alpha/2}$

$$\boxed{Z = \frac{Y/n - p_0}{\sqrt{p_0(1-p_0)/n}}}$$

Hypothesis test for proportions

Aladdin is a frequent flyer. He thinks he gets security-screened more than normal at the magic-carpet-port. Assume security randomly screens 10% of all people (so he should be screened 10% of the time). In the past few years, he has been (randomly?) selected 16 out of 100 times.

Perform a hypothesis test at 0.05 significance to see if the screening process is random, or biased towards screening him more.

$$Z = \frac{Y/n - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

$$Z = \frac{Y/n - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Hypothesis test for proportions

Assume security randomly screens 10% of all people (so he should be screened 10% of the time). Aladdin has been selected 16 out of 100 times. Is process biased towards screening him more? Test at α=0.05.

H_0:

H_A:

Test statistic:

p-value:

Decision:

Conclusion: