Set 25, 26, 27 - Hypothesis Tests for a Single Population

March 22, 2023 1:23 PM

Stat 260 Lecture Notes

Sets 25, 26, 27 - Hypothesis Tests for a Single Population

Suppose a student is accused of cheating on the midterm. In our law system, the student is innocent until proven guilty. We need to gather enough evidence to show that they are guilty. To do this we will use *hypothesis testing*.

only trying to gather evidence for thing you're trying to prove

null hypothesis: H_0: the student did not cheat

↳ what we are assuming

alternative hypothesis: H_1: the student did cheat

↳ what we're trying to prove

We assume that the null hypothesis holds unless we have enough evidence to suggest the alternative hypothesis.

Two outcomes: Conclusion

1. There is enough evidence to reject H_0, so we accept H_1.
2. There is not enough evidence to reject H_0, so we keep believing it. (This is not the same as proving H_0 is true.)

We could be wrong.

Type I error: Conclude that we reject H_0, but we shouldn't have. This happens with probability α.

Type II error: Conclude there is not enough evidence to reject H_0, but we should have rejected it. This happens with probability β.

Type I: Conclude the student did cheat, but really did not.

Type II: Conclude that we keep believing student did not cheat, but really they did.

1

For us, making a Type I error is more serious, so we want α to be small.

α is also called the *significance level*.

Rule: As α (probability of Type I error) gets larger, β (probability of a Type II error) gets smaller, and vice versa. We can reduce the probability of both types of errors by increasing the sample size n.

↳ makes both smaller

Any time H_0 is rejected we could make a Type I error. Any time H_0 is not rejected we could make a Type II error.

population

To start, our hypothesis tests will be about the value of the true mean μ.

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

Three possible setups:

one-tailed test

1. Is μ greater than some amount μ_0?
   H_0 : $\mu \le \mu_0$ $\mu \le 3$ same as $H_0 : \mu = \mu_0$
   H_1 : $\mu > \mu_0$ ← what we're trying to prove! $\mu > 3$ $H_1 : \mu > \mu_0$
2. Is μ less than some amount μ_0?
   H_0 : $\mu \ge \mu_0$ same as : $H_0 : \mu = \mu_0$
   H_1 : $\mu < \mu_0$ $H_1 : \mu < \mu_0$

two-tailed test

3. Is μ different from some amount μ_0?
   H_0 : $\mu \neq 0$
   H_1 : $\mu \neq 0$

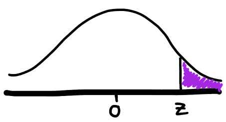
Will have actual numbers for μ_0

for all cases, all null are $H_0 : \mu = \#$

Here's the general outline of how hypothesis testing works:

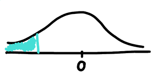
- Assume H_0 is true. (So assume $\mu = \mu_0$.)
- Measure $\bar{x}$. Standardizing $\overline{X}$ gives the *test statistic*. $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
- Find the probability of observing a mean value as extreme as the observed $\bar{x}$ while assuming the actual mean is $\mu = \mu_0$. This probability

same decision from set 23 for using z or T.



p-value = P(Z > z)

p-value = P(Z < z)



is called the p-value.

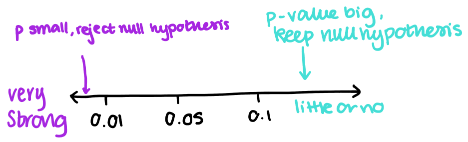
- If this probability is small, either we observed something extremely rare, or else our assumption that $\mu = \mu_0$ was wrong. It is more likely that $\mu = \mu_0$ was wrong, so reject the H_0 assumption. If the probability is not small, our $\mu = \mu_0$ assumption was probably OK and we keep believing H_0. p-value is small ⇒ reject H_0.
  p-value is large ⇒ keep H_0.

We only do hypothesis testing by the p-value approach. Another option is to use the rejection region method, but we will not use this in Stat 260.

The steps for hypothesis testing:

1. Define the parameter to be tested. (Are we doing a test on the average μ, or something else?)
2. Specify H_0 and H_1.
3. Specify the test statistic and identify its (approximate) distribution if H_0 is true.
4. Compute the observed value of the test statistic.
5. Compute the p-value.
   The *p-value* is the probability of observing a value as extreme as our test statistic while assuming H_0 is true.

6. If asked, report the strength of evidence against H_0 in favour of H_1.
   The smaller the p-value is, the less chance we make an error concluding that there is enough evidence to support H_1.
   - very strong evidence if $p\text{-value} \leq .01$. ⇒ reject H_0
   - strong evidence if $.01 < p\text{-value} \leq .05$.
   - moderate evidence if $.05 < p\text{-value} \leq .10$.
   - little or no evidence if $.10 < p\text{-value}$. ⇒ keep H_0



7. If asked, report the estimated value of the parameter along with our measured estimated standard error.

8. If we are asked to test H_0 at the significance level α, compare α with the p-value and reject H_0 exactly when the p-value $\leq \alpha$. If p-value$> \alpha$, we keep assuming that the null hypothesis H_0 is true. We didn't prove that H_0 is true, we just don't have enough evidence to say that it is false.

p-value $\leq \alpha \Rightarrow$ p-value small $\Rightarrow$ reject H_0

p-value $> \alpha \Rightarrow$ p-value big $\Rightarrow$ keep H_0

Example 1: It is recommended that stats students study at least 36 hours in the semester for the course. Suppose we want to test if the average student studies less than this amount. Suppose study times are normally distributed with $\sigma = 8$, and when we sample $n = 25$ students we observe that $\overline{x} = 34.25$. Test at a significance level of $\alpha = 0.10$.

$n=25$ (small), know σ, Xs are normal
$\Rightarrow \bar{x}$ is normal

* signs of hypothesis test

1. Define the parameter to be tested. averages or proportions?

testing μ - true mean study time

2. Specify H_0 and H_1.

$H_0: \mu = 36$

"claim" $H_1: \mu < 36$

steps together on assignment

3. Specify the test statistic and identify its (approximate) distribution.

using Z

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4. Compute the observed value of the test statistic.

from H_0

$$Z_{obs} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{34.25 - 36}{8/\sqrt{25}} = -1.09$$

5. Compute the p-value.

p-value $= P(Z < -1.09) = 0.1379$

6. Report the strength of evidence against H_0 in favour of H_1.

p-value = 0.1379

little or no evidence against H_0, since p-value > 0.10

7. Report the standard error of the parameter along with our measured estimated value of the parameter.

estimate → $\frac{\text{r.v.} - \text{expected value}}{\text{std. error}}$

Std. error $= \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{25}} = 1.6$

estimate $= \bar{x} = 34.25$

8. If we are asked to test H_0 at the significance level α, compare α with the p-value and reject H_0 exactly when the p-value $\leq \alpha$.

p-value = 0.1379 $\quad \alpha = 0.10$

p-value > $\alpha \Rightarrow$ p-value is big $\Rightarrow$ keep H_0.

Keeping null means not enough evidence for the alternative

Conclude that there is not enough evidence to say students study less than 36 hours on average

if we have α then use this for conclusion

Example 2: The drying time of paint is normally distributed with mean 75 minutes. Say we want to put an additive into the paint to make it dry faster. To test if the drying time is faster, we make 10 measurements and find that the average drying time is 70.8 minutes with a standard deviation of 9 minutes. Test the claim that using the additive has a faster drying time than if we don't use the additive. Use a significance level of $\alpha = 0.05$.

sample

coming from measurements is sample

$n = 10$ (small), using s $\Rightarrow$ $\bar{x}$ is T dist.

1. Define the parameter of interest using the correct notation. Then, state the null and alternative hypotheses for this study.

→ questions will be in similar format to this

testing μ - true mean drying time of paint with additive.

$H_0: \mu = 75$

$H_1: \mu < 75$ ← trying to prove/test

what we measured

2. Calculate the observed value of the test statistic. State the distribution (and degrees of freedom if needed) it follows. Z or T setup?

in H_0

$$T_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{70.8 - 75}{9/\sqrt{10}} = -1.476$$

$df = \nu = n - 1 = 10 - 1 = 9$

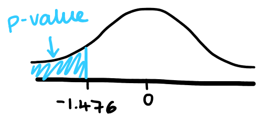
has to be same

3. Compute the p-value or provide a range of appropriate values for the p-value.

$p\text{-value} = P(T_9 < -1.476)$

$= P(T_9 > 1.476)$

$0.05 < p\text{-value} < 0.10$



4. Using the significance level $\alpha = 0.05$, state your conclusions about the paint drying time.

$0.05 < p\text{-value} < 0.10$ $\quad$ $\alpha = 0.05$ $\quad$ $p\text{-value} > \alpha \Rightarrow p\text{-value}$ big $\Rightarrow$ keep H_0

→ won't give α in p-value range
↳ will be bigger or less than

Not enough evidence to say the mean drying time is less than 75 minutes.

3, 4 on brightspace

Example 3: Aspirin bottles are filled by weight, so if the pills are larger than they should be, fewer pills will be put in a bottle. The production process is designed so that pills should have an average weight of 5 grams. Say 100 pills are randomly selected and we find that the average weight in this group is 5.13 g and the standard deviation is 0.35 g. Does this give us enough evidence to say that the production process is creating pills that weigh more than they should? Test at a significance level of $\alpha = 0.01$.

1. Define the parameter of interest using the correct notation. Then, state the null and alternative hypotheses for this study.

2. Calculate the observed value of the test statistic. State the distribution (and degrees of freedom if needed) it follows.

3. Compute the p-value or provide a range of appropriate values for the p-value.

4. Using the significance level $\alpha = 0.01$, state your conclusions about the pills created by the production process.

Example 4: Suppose students in the US average 494 point on the SAT-I math exam with a standard deviation of 124 points. At one school, 86 students are registered in a special math program and when they took the SAT-I math exam they achieved an average of 517 points. The administration wants to know if the special program had an effect (either positive or negative) on test scores. Test at the $\alpha = 0.05$ significance level.

1. Define the parameter of interest using the correct notation. Then, state the null and alternative hypotheses for this study.

2. Calculate the observed value of the test statistic. State the distribution (and degrees of freedom if needed) it follows.

3. Compute the p-value or provide a range of appropriate values for the p-value.

4. Using the significance level $\alpha = 0.05$, state your conclusions about the math program.

Example 5: Suppose we have a population that follows the normal distribution. From 20 observations we find that the mean is 52 and the standard deviation is 15. Suppose we want to test if the average is different from 50. State the strength of evidence for this claim.

$n = 20$ (small), using $s \Rightarrow \bar{x}$ is T dist.

1. Define the parameter of interest using the correct notation. Then, state the null and alternative hypotheses for this study.

testing μ - population mean

$H_0: \mu = 50$

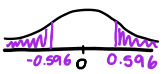
$H_1: \mu \neq 50$

2. Calculate the observed value of the test statistic. State the distribution (and degrees of freedom if needed) it follows.

$$T_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{52-50}{15/\sqrt{20}} = 0.596 \qquad df = n-1 = 20-1 = 19$$

3. Compute the p-value or provide a range of appropriate values for the p-value. * pick tail that's easier on stat table

$$\text{p-value} = P(T_{19} < -0.596) + P(T_{19} > 0.596)$$
$$= 2 \cdot P(T_{19} > 0.596)$$



$2 \cdot 0.20 < \text{p-value} < 2 \cdot 0.30$

$0.40 < \text{p-value} < 0.60$

4. State the strength of evidence against H_0 (and make any possible conclusions).

$0.40 < \text{p-value} < 0.60 \quad p > 0.10$

So little or no evidence against H_0.

CUTOFF FOR TEST 3!

We can also do hypothesis tests about the population proportion, p.

Three possible setups for hypotheses:

1. Is p greater than some amount p_0?
   H_0 :
   H_1 :
2. Is p less than some amount p_0?
   H_0 :
   H_1 :
3. Is p different from some amount p_0?
   H_0 :
   H_1 :

The procedure for hypothesis tests on p is the same as the procedure for hypothesis tests on μ. The only difference is that here our test statistic will be

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

where $\hat{p}$ is the measured proportion, and p is the proportion value we use in the null hypothesis. Note that we always use the normal distribution for hypothesis tests on p.

Example 6: Suppose we want to test the claim that the true proportion of people with Type O blood is greater than 40%. When we test the blood of 1000 people we find that 429 of them have Type O blood. Perform a hypothesis test on our claim using the significance level of $\alpha = 0.05$.