Statistics Ch 9 Notes

Ch 9.1, 9.3 and 9.4 Hypothesis Test basic

Hypothesis test: (or **test of significance**) is a procedure based on sample evidence and probability, used to test claim regarding a characteristic of one or more populations. Characteristic of population proportion (p) and mean (μ) will be covered in this course.

To test a hypothesis, you should state a pair of hypotheses, one that represents the claim and the other, its complement. When one of these hypotheses is false, the other must be true. The null hypothesis represents currently acceptable truth. The alternative hypothesis contains opposing viewpoint.

Basic process of a Hypothesis test (p-value method)

1) Identify the claim.
2) Translate the claim in algebraic symbolic form.
3) Identify the null and alternative hypothesis test and write in symbolic form. H0 and Ha (or H1) are the symbols for the two hypotheses.
4) Select a significant level α. (based on seriousness of making a type I error.)
5) Collect a sample and identify the "Type" of hypothesis test in Ha. Determine the sampling distribution (normal or t).
6) Use calculator to find test statistic and p-value.
7) Make conclusion on H0 based on p-value.
8) Rewrite the conclusion in simple non-technical term and address the original claim.

Part A: Step 1, 2, 3

The claim is about value of a population parameter, we can claim that the parameter is greater, less than, equal to or not equal to a value. (Note: claim of at most, at least is equivalent to claiming equal.)

a) Since the claim is about population parameter, the symbolic form can be:

claim: $\begin{pmatrix} p \\ \mu \end{pmatrix} \begin{matrix} = \\ > \\ < \\ \neq \end{matrix} (a\ value)$

b) The null hypothesis is the first assumption we use to calculate probability of the sample we obtained, so the null hypothesis must be of equality form.

c) The alternative hypothesis is the second assumption that must not overlap (or opposite) with the null hypothesis.

d) One of the two hypothesis must be exactly the claim.

H0: $\begin{pmatrix} p \\ \mu \end{pmatrix} = (a\ value)$; Ha: $\begin{pmatrix} p \\ \mu \end{pmatrix} \begin{matrix} \neq \\ > \\ < \end{matrix} (a\ value)$

" a value" is the number that shows up in the claim.

Ex1: Claim that mean body temperature is less than 98.6°F:
claim: μ < 98.6 H0: μ = 98.6 Ha: μ < 98.6

Ex2. Claim that proportion of red M&M is greater than 10%.
claim: p > 0.1; H0: p = 0.1 Ha: p > 0.1

Ex3. Claim that mean IQ scores of college professor is different from 100.
claim: μ = 100; H0: μ = 100 Ha: μ ≠ 100

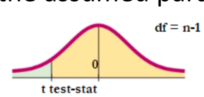
Ex4. Claim that mean IQ scores of college student is 105.
claim: μ = 105; H0: μ = 105 Ha: μ ≠ 105

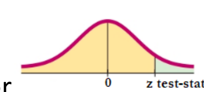
Part B: Step 4, 5 and 6:

Common significant level α is 0.05, but if Type 1 error is very undesirable, a lower significant level is better.

If the claim is about p, the sampling distribution is z normal because $\hat{p}$ is normally distributed.
If the claim is about μ, the sampling distribution is t or z normal depending on if σ is known. According to CLT, $\bar{x}$ is normally distributed.

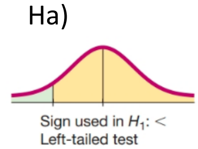
Obtain a sample and use calculator to find the
a "test statistic" and a "p-value"
Test statistic tells the number of SD the sample is from the assumed parameter value assumed in H0.



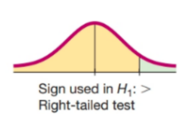
or


P-value tells the probability of getting the sample or worse if the assumption of H0 is true.

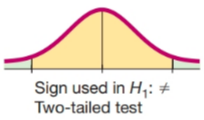
P-value are calculated based the "Type" of test:
Types of Hypothesis Test: (determined by looking at Ha)



Left-tail test: when Ha has the form of p < value. p-value is left area of test statistic.



right-tail test: when Ha has the form of p > value. P-value is right area of the test statistic.



Two-tail test: when Ha has the form of p ≠ value
p-value is twice of area in left or right area from the test statistic.

The online calculator will calculate test statistic and p-value. https://www.statdisk.com/#
Analysis/Hypothesis testing/
More detail will be covered in Ch 9.5.

Part C: Step 7 and 8:

The conclusion of a hypothesis test is based on the Rare Event Rule:
"Based on an assumption, if the observed event is very rare (lower than significant level α), we conclude the assumption is properly not true."
In a Hypothesis Test, the assumption is H0, the null hypothesis. We determine if the sample is rare based on the assumed claim population characteristic in H0.

Step 7: If p-value ≤ α, reject null hypothesis.
Result is Significant!
If p-value > α, fail to reject the null hypothesis.

[If p-value is low, the null must go.
If p-value is high, the null will fly.]

Step 8: Make conclusion about the claim.
If the p-value ≤ α: use "there is sufficient evidence…"
If the p-value > α, use "there is not sufficient evidence…"
If the claim is H0: Reject H0 implies reject the claim.
If the claim is Ha: Reject H0 implies support the claim.
Or

Condition	Conclusion
Original claim does not include equality, and you reject H_0.	"There is sufficient evidence to *support* the claim that . . . (original claim)."
Original claim does not include equality, and you fail to reject H_0.	"There is not sufficient evidence to support the claim that . . . (original claim)."
Original claim includes equality, and you reject H_0.	"There is sufficient evidence to warrant *rejection* of the claim that . . . (original claim)."
Original claim includes equality, and you fail to reject H_0.	"There is not sufficient evidence to warrant rejection of the claim that . . . (original claim)."

or Use the conclusion flow chart to make the final claim statement.

Note:
"Not sufficient evidence to reject the claim" implies it is plausible that the claim is true.
"Not sufficient evidence to support the claim" implies the claim may not be true.

Ex1. If significant level = 0.05 and p-value = 0.04, what is the conclusion on H0?

Since 0.04 < 0.05, sample is significant, so reject H0.

Ex2. If significant level = 0.05 and p-value = 0.006, what is the conclusion on H0?
Since 0.006 < 0.05, the sample is significant, so reject H0.

Ex3. If significant level = 0.01 and p-value = 0.03, what is the conclusion on H0
Since 0.03 > 0.01, the sample is not significant, so fail to reject H0.

Ex4. If the claim is Ha and you fail to reject H0, what is the claim conclusion?

There is not enough evidence to support the claim.

Ex5. If the claim is H0 and you reject H0, what is the claim conclusion?

There is sufficient evidence to reject the claim.

All steps practice:

Ex1. Test a claim that body temperature of adult is less than 98.6°F. Use α = 0.05. A random sample of 38 body temperatures (with $\bar{x} = 98.1, s = 1.2$) gives a test statistic of t = –2.56 and p-value of 0.0072.
a) Write the claim and hypothesis,
b) Determine the type of distribution used and "type of hypothesis test." and significant level.
c) Interpret the meaning of test statistic and p-value.
d) Use p-value to make conclusion about H0 and about the claim.

Answer:
a) Claim: μ < 98.6 H0: μ = 98.6 Ha: μ < 98.6
b) Since μ is the parameter, σ is not known, so use t distribution. Since Ha is "<", Type of test is left tail. α = 0.05.
c) Test statistic of t = -2.56 means the sample data is 2.56 standard deviation below μ = 98.6.
p-value of 0.0071 means the probability of having the sample or worse is 0.71% if the real mean is 98.6.
d) Conclusion about H0:
Since p-value < 0.05, reject H0, sample is significant
p-value ≤ α -> use " there is sufficient evidence"..

Slaim is Ha -> use " support the claim."
"There is sufficient evidence to support the claim that mean body temperature is less than 98.6° F.
Flow chart: claim is Ha, reject H0, Box 3.

Ex2: Claim that proportion of red M&M is greater than 10%. Use $\alpha = 0.05$. A sample (15 red out of 102 M&M candies) gives a test statistic of z = 1.91, p-value = 0.0566.

a) Write the claim and hypothesis,

b) Determine the type of distribution used and "type of hypothesis test" and significant level.

c) Interpret the meaning of test statistic and p-value.

d) Use p-value to make conclusion about H0 and about the claim.

Answer:

a) claim: p > 0.1; H0: p = 0.1; Ha: p > 0.1

b) $\alpha = 0.01$. Since p is the claim parameter, sampling distribution $\hat{p}$ is normal, z distribution is used.
Since Ha use p > 0.1, Type of test is right tail test.
$\alpha = 0.05$.

c) A test statistic of z = 1.91 means the sample proportion is 1.91 times of SD above the mean of p = 0.1.
P-value of 0.0566 means there is 5.66% probability to obtain such as sample if p = 0.1 is true.

d) p-value of 0.0566 > 0.05: fail to reject H0 (sample is not significant because the sample is not a rare event.)
Since p-value > α so use "there is not sufficient evidence"
claim is Ha, → use " support the claim"
Final conclusion: there is not sufficient evidence to support the claim that proportion of red M&M is greater than 10%.
Flow chart: Ha is claim , fail to reject H0, Box4.

Ex3. Claim that IQ scores of college student has a mean equal to 104. Given that σ for IQ score is 15. A sample of IQ scores from 40 college students ($\bar{x}$ = 106, s = 16) give a test statistic z = 0.84 and a p-value = 0.1995. Use $\alpha = 0.05$ to test the claim.

a) Write the claim and hypothesis.

b) Determine the type of distribution used and type of hypothesis test and significant level.

c) Interpret the meaning of test statistic and p-value.

d) Use p-value to make conclusion about H0 and about the claim.

Ans:

a) claim: μ = 104; H0: σ = 104; Ha: μ ≠ 104

b) Since claim parameter is μ but σ is known, so use z-Normal distribution. Since Ha has the form of ≠, the Hypothesis test is a "Two-tail Test". $\alpha = 0.05$.

c) Test statistic z = 0.8432 means the sample is 0.84 times of standard deviation from μ = 104.
p-value of 0.1995 means there is 19.95% chance of getting such sample if μ = 104 is true.

d) p-value 0.1995 > 0.05, the sample is not a rare event, fail to reject H0, sample is not significant.
P-value > α so, "there is not sufficient evidence."
Claim is in H0 → Use "reject the claim" statement.
Conclusion: There is not sufficient evidence to reject the claim that the college student's mean IQ is equal to 104. (Conclude that mean IQ for college student could be 104.)
Flowchart: Claim is H0, Fail to reject H0: use Box2.

Ch 9.5- part 1 Full Hypothesis Test for population proportion

Notations:

x = number of success

n = sample size (number of observations)

$\hat{p} = \frac{x}{n}$ = sample proportion

p = claim proportion mentioned in claim, use in H0

q = 1 – p

α = significant level. (probability of unlikely)

STEPS:

1) Write claim, H0 and Ha in symbolic form.
   Identify n, x.
2) Determine significant level α, type of test (left-tail, right-tail or two-tail test based on Ha and sampling distribution. Sampling distribution of $\hat{p}$ is Normal.
3) Use Statdisk.com to find test-statistic and p-value:
   Analysis/Hypothesis testing/Proportion One sample.
   Select Population proportion "not equal" or " >" or "< "claimed proportion according to Ha.
   Input Significance α; claimed proportion (in H0); sample size n; number of successes x. Evaluate.
   Output: test stat z and p-value p.
4) Make conclusion about H0.
   If p-value ≤ α, reject H0. Sample is significant.
   If p-value > α, fail to reject H0
5) Conclusion about the claim:
   There is (sufficient or not sufficient) evidence to (support / reject) the claim that ……
   Use table or flowchart for wordings.
6) Make additional inference from the conclusion.

The three conditions for using the hypothesis test are:
a) Number of success and failures are at least 5.
$np \geq 5$, $nq \geq 5$.
b) Fixed number of samples with two outcomes and independent sample. (binomial requirement)
c) Sample are randomly collected.

Ex1: A Pitney Bowers survey of 1009 customers shows that 545 of customers are uncomfortable with Drone deliveries. Test the claim that majority of customers are uncomfortable with Drone deliveries. Use α=0.05. (Majority means greater than 50%.)

Answer:
1) Claim: $p > 0.50$ H0: $p = 0.50$ Ha: $p > 0.5$
$n = 1009$, $x = 545$,
2) $\alpha = 0.05$, right- tail test, use z distribution
3) Use Statdisk/Analysis/Hyp Test/proportion one sample. Select ">" for Alternative Hypothesis.
Enter Significance = 0.05, claimed proportion 0.5.
n = 1009, x = 545 , calculate.
Output: z-stat = 2.55 and p-value =0.0054
This means the sample is 2.55 sd from the mean of p = 0.5 and the probability of getting the sample or worse is 0.0054 when H0 is true.
4) Since $0.0054 < 0.05$ Reject H0, significant sample.
5) Use "sufficient evidence" because p-value < α,
Use "support the claim" because claim is Ha.
"There is sufficient evidence to support the claim that majority of customers are uncomfortable with drone deliveries."
Flowchart: Claim is Ha, reject H0, Box 3.

b) Should a company start to invest in drone deliveries now?
No, because majority of customers are uncomfortable with drone deliveries.

Ex2: Test the claim that less than 30% of adults have sleep-walked. Use significant level α = 0.05. A random sample shows 29.7% of 1913 adults have sleep-walked.

Answer:
1) claim: $p < 0.30$ H0: $p = 0.30$ Ha: $p < 0.30$
$n = 1913$, $x = 1913(0.297) = 568$
2) α=0.05, left-tail test, use z distribution
3) Use Statdisk/Analysis/Hyp Test/proportion one sample. Select "<" for Alternative Hypothesis.
Enter Significance = 0.05, claimed proportion 0.3,
n = 1913, x = 568 , calculate.
Output: Z = – 0.29, p-value = 0.384
This means the sample is – 0.29 of s.d. below the assumed true H0 value. The probability of getting the sample is 0.384 when H0 is true.
4) $0.384 > 0.05$, fail to reject H0. (not significant)
5) Fail to reject H0: use "not sufficient evidence.."
claim is in Ha: use "…support the claim."
There is not sufficient evidence to support the claim that less than 30% of adults have sleep-walked.
Flow chart: Claim is Ha, fail to reject H0, Box4.

b) Is the claim true? No, not sufficient evidence to support the claim so the claim is not true.

Ex3: In a USA today survey of 510 people, 53% said that we should replace passwords with biometric security such as fingerprints.
a) Use a significant level of 0.1 to test the claim that exactly 50% of all adults like the idea of replacing passwords with biometric security.

Answer:
1) Claim: $p = 0.5$, H0: $p = 0.5$, Ha: $p \neq 0.5$
$n = 510$, $x = 510(0.53) = 270$,
2) $\alpha = 0.1$, two-tail test, use z- distribution.
3) Use Statdisk/Analysis/Hyp Test/proportion one sample. Select "not =" for Alternative hypothesis.
Enter Significance = 0.1, claimed proportion 0.5,
n = 510, x = 270, calculate.
output: Test statistic z = 1.33, p-value = 0.184
4) $0.184 > 0.1$, fail to reject H0.
Since we fail to reject H0: use " not sufficient evidence…"
The claim is in H0, use "… reject the claim." statement.

There is not sufficient evidence to reject the claim that half of all adults like the idea of replacing password with biometrics.

b) Is the claim true? yes, not sufficient evidence to reject the 50% proportion, conclude the only 50% of the population like the idea.

c) Discuss if the conditions for hypothesis test of proportion is satisfied?
The success-failure condition: $np = 510(0.5) > 5$
$nq = 510(1-0.5) > 5$. Assuming the sample is a simple random sample, so conditions are satisfied.

Ch 9.5-part 2 Full Hypothesis Test for Mean

Terms: Population mean: μ

Sample mean: $\bar{x}$

sample standard deviation: s

sample size: n

Pop. standard deviation: σ (given or unknown)

significant Level: α (probability of unlikely)

1) Write claim, H0 and Ha in symbolic form.
   Identify $\sigma, n, \bar{x}, s$, or input sample data in one column of statdisk sample editor.
2) Determine significant level and type of test (left-ail, right tail or two-tail test based on Ha), the sampling distribution for mean is z if σ is known, the sampling distribution for mean is t if σ is not given.
3) Use Statdisk/Analysis/Hypothesis Testing/ Mean one sample:
   - If $n, \bar{x}, s$ is available, use summary statistic tab,
   - If sample data is available, use "Data" tab.
   - Select Alternative Hypothesis: "not equal" for two-tail test, "<" for left tail test, ">" for right tail test.
   - Enter significance, claimed mean, population SD if known. Select data column or enter $n, \bar{x}, s$. Evaluate
   - Output: Test stat z or t, p-value

4) Make conclusion about H0.
   If p-value ≤ α, reject H0. Sample is significant.
   If p-value > α, fail to reject H0
5) Conclusion about the claim:
   There is (sufficient or not sufficient) evidence to (support /reject) the claim that ……
   Use table or flow chart to make conclusion.
6) Make inference from the conclusion.

Conditions for Hypothesis Testing of mean:
1) Sample is SRS.
2) The population is normally distributed or n >30.
   Use Normal quantile plot to check for normality if data is given.

Ex1: Test the claim that mean amount of adult sleep is less than 7 hours. A sample 12 adults gives $\bar{x} = 6.82\ hr, s = 1.99\ hours$. Use a significant level of 0.05. Given that hours of sleep are normally distributed.

Answer:
1) Claim: μ < 7, H0: μ = 7 , Ha: μ < 7
   $n = 12\ , \bar{x} = 6.82\ hr, s = 1.99\ hours$
2) α = 0.05, Left-tail test and use t-distribution
3) Use Statdisk/Analysis/Hypothesis Test/ Mean one sample/ Use summary statistics tab.
   Select Alternative Hypothesis Test: select "<"
   Enter significance = 0.05, $n, \bar{x}, s$. calculate.
   Output: Test statistic t = - 0.31, p-value = 0.3799.
4) Since p-value > 0.05, Fail to reject H0
5) There is no sufficient evidence to support the claim that mean sleep hour is less than 7 hours.
   Flowchart: Claim is Ha, fail to reject H0, Box4

b) Public Health guideline for hours of sleep per night is 7 hours or more. Does the public follow the guideline from Public Health?
Yes, the result concludes that the mean is not less than 7 hour so it is plausible to be 7 hours or more.

Ex2: Given below are the measured radiation emissions (in W/kg) corresponding to a sample of 11 most popular brand of cell phones. Use a 0.05 significant level to test the claim that cell phones have a mean radiation level that is greater than 0.7 W/kg.
0.38 0.55 1.54 1.55 0.50 0.60 0.92 0.96 1.00 0.86 1.46

Answer:
1) Claim: μ > 0.7, H0: μ = 0.7 , Ha: μ > 0.7
   Input data to a column in statdisk.
2) α = 0.05, Right-tail test and use t-distribution.
3) Use Statdisk/Analysis/Hypothesis Test/ Mean one sample/ Use data tab.
   Select Alternative Hypothesis Test for ">"
   Enter significance = 0.05, select data column, Evaluate.
   Output: Test statistic = 1.868, p-value = 0.0456.
4) Since p-value = 0.0456 < 0.05 Reject H0
5) Because we reject H0, use sufficient evidence statement. The claim is in Ha, use "support the claim" statement.
   There is sufficient evidence to support the claim that the mean radiation for cell phones is greater than 0.7 W/kg.
   Flowchart: claim is Ha, reject the claim: Box 3.

b) Is this result useful for all cell phones in use?
- No, the sample is not a random sample. The sample are from each of the top selling cell phone so it is not useful for all cell phones in use.

c) Discuss if the condition for hypothesis test for mean is satisfied or not.

- A normal quantile plot and boxplot shows the points are close to a straight line and there is no outliers. So the requirement for Normal distribution is satisfied.

Ex3: Output from a Minitab software shows the following after inputting a sample that have $\bar{x}$ = 9.81 km , s = 5.01km and n=50.

MINITAB

```
Test of mu = 10 vs not = 10

 N   Mean  StDev  SE Mean       95% CI          T      P
50  9.810  5.010    0.709  (8.386, 11.234)  -0.27  0.790
```

Use the output to test the claim that mean depth of all earthquakes is equal to 10 km at α = 0.05.

Answer:

1) Claim: μ = 10, H0: μ = 10, Ha: μ ≠ 10.
2) test statistic = t = -0.27, p value = 0.790 from above.
3) Since p-value 0.790 > 0.05, fail to reject H0
4) Use "not sufficient evidence" statement and
   Use "reject the claim" statement (because claim is in H0)

Not sufficient evidence to reject the claim that mean depth is 10 km, so claim is true the mean depth of all earthquakes is 10km.

Flowchart: Claim is H0, fail to reject H0, Box 2

Claim is true.

Ch 9.2 Outcomes and Type1 and Type2 Errors

Conclusion from Hypothesis Test are based on sample observations which are determined by chance. Hypothesis conclusion may not always reflect the actual true population parameters.

	H0 is true	H0 is false
Reject H0	Type 1 error	correct
Fail to reject H0	Correct	Type 2 error

Type I error: Rejecting a null hypothesis but actually the hypothesis is true. Probability of making a type 1 error is α. If result of type 1 error is serious, we will want to use a low α instead of the default of 0.05.

Type II error: Failing to reject a null hypothesis when you should have rejected it because the null hypothesis is actually false. The probability of making a type II error is β. The power of the Test is 1 – β.

Note: we may not know if we have made hypothesis error, but we can plan for it by adjusting the significant level α.

Ex1. Claim that a medical procedure will increase likelihood (more than 50%) of a baby girl.

a) Discuss the type 1 and type II error in the context of the problem.

Answer:

Claim: p > 0.5, H0: p = 0.5

Type 1 error is concluding the procedure increase percentage of girl to more than 50% but actually the true percentage of girl is only 50%. The sample evidence leads us to believe that the medical procedure is effective but actually it is not.

Type II error is concluding that the percent of girl after using the procedure is only 50% but the real percent of girl using the procedure is more than 50%. Concluding the procedure is not effective, but actually it is effective.

b) Evaluate which error is more serious and advise on the level of significance.

Type I error is more serious from the public perspective, a lower α is advisable because probability of making a Type I error is α

Ex2:

A company manufacturing computer chips finds that 8% of all chips manufactured are defective. Management is concerned that employee inattention is partially responsible for the high defect rate. In an effort to decrease the percentage of defective chips, management decides to offer incentives to employees who have lower defect rates on their shifts. The incentive program is instituted for one month. If successful, the company will continue with the incentive program.

Answer:

claim: $p < 0.08$ H0: $p = 0.08$

Type I error: Concluding that the incentive program can lower defective rate to less than 8% but actually the defective rate is still 8%. Concluding that the incentive program is useful but actually it is not useful.

Type II error:

Concluding that the defective rate is 8% but actually the defective rate after the incentive program is less than 8%. Conclude that the incentive program is not useful but actually it is useful.

b) Discuss if type I or type II error is more serious.
Should management use a high or low significant level?

Ans:

Type I error means company will spend money on something that is not useful. So to avoid unnecessary expense, Type I error should be avoided.

Hence a low significant level may be better.

If cost is a priority, Type I error should be avoided so a lower α is better.

Type II error means you bypass a good program that can decrease defective rate or increase profit. Type II error should be avoided if profit is a priority, hence a higher α is better.