Homework 3 Computer Science Theory

Katherine Lee khl2145

Chapter 8: Hypothesis Testing

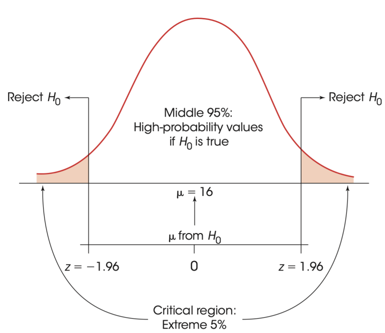
8.1 Hypothesis Testing

Hypothesis testing is the statistical method that uses sample data to evaluate a hypothesis about a population

Four Steps of a Hypothesis Test

- State hypothesis about population and select alpha level
  - NULL $H_0 = \mu_0 = \mu_1$
  - Alternative $H_1 = \mu_0 \neq \mu_1$
  - it might also have a direction (a one-tailed test) in this case you would use the operators $<, >$
- Set the criteria for a decision: when we should refute a null hypothesis (find critical region)
  - Alpha(α) level: aka level of significants for a hypothesis test
  - Critical region: extreme sample values that are very unlikely (defined by the alpha level) to test if the null is true
- Collect data nd compute sample statistics (find test statistic z-score)
  - this includes z-score for sample: this measures the difference between the sample and population
- Make a decision about the null
  - you the z-score to determine you decision
  - the SD for sample determines the amount of error you expect between the sample mean and population mean, when you compare the result you get from the α z-score, then you can see if the z-score is greater or less than the given score
  - for example $\alpha = 1.96$ if z-score from sample $= 2.40$, then it will reject the null

FIGURE 8.4
The critical region (very unlikely outcomes) for α = .05.



z-score for sample mean $z = \frac{M - \mu}{\sigma_M}$

standard error between M and μ: $\sigma = \frac{\sigma}{\sqrt{n}}$

8.2 Errors

Type I errors

occurs when a researcher rejects a null hypothesis that is actually true. In a typical research situation, a Type I error means the researcher concludes that there is evidence for a treatment effect when in fact the treatment has no effect

- alpha level for a hypothesis test is the probabilty the test will lead to a Type I error, so by selecting a small alpha level, you are reducing the probability for a Type I error

Type II errors β

A Type II error occurs when a researcher fails to reject a null hypothesis that is in fact false. In a typical research situation, a Type II error means that the hypothesis test has failed to detect a real treatment effect.

- its difficult to determine a single exact probability for a type II error: sample size and effect size can be factors.

		Actual Situation	
		No Effect, H_0 True	Effect Exists, H_0 False
Researcher's Decision	Reject H_0	Type I error	Decision correct
	Fail to Reject H_0	Decision correct	Type II error

Selecting α level

The largest permissible value is $\alpha = 0.05$ or 1 in 20 probability

Type II errors β

A result is said to be significant or statistically significant if it is very unlikely to occur when the null hypothesis is true. That is, the result is sufficient to reject the null hypothesis. Thus, a treatment has a significant effect if the decision from the hypothesis test is to reject H_0.

8.3 Assumptions with Hypothesis Tests with z-scores

Random Sampling It is assumed that the participants used in the study were selected randomly. Remember, we wish to generalize our findings from the sample to the popula- tion. Therefore, the sample must be representative of the population from which it has been drawn. Random sampling helps to ensure that it is representative.
Independent Observations The values in the sample must consist of independent observations. In everyday terms, two observations are independent if there is no consis- tent, predictable relationship between the first observation and the second. More precisely, two events (or observations) are independent if the occurrence of the first event has no effect on the probability of the second event. S
The Value of σ Is Unchanged by the Treatment A critical part of the z-score formula in a hypothesis test is the standard error, sM. To compute the value for the standard error, we must know the sample size (n) and the population standard deviation (s). In a hypothesis test, however, the sample comes from an unknown population (see Figure 8.2). If the population is really unknown, it would suggest that we do not know the standard deviation and, therefore, we cannot calculate the standard error. To solve this dilemma, we have made an assumption. Specifically, we assume that the standard deviation for the unknown population (after treatment) is the same as it was for the population before treatment. Actually, this assumption is the consequence of a more general assumption that is part of many statistical procedures. This general assumption states that the effect of the treat- ment is to add a constant amount to (or subtract a constant amount from) every score in the population. You should recall that adding (or subtracting) a constant changes the mean but has no effect on the standard deviation. You also should note that this assumption is a theoretical ideal. In actual experiments, a treatment generally does not show a perfect and consistent additive effect.

Normal Sampling Distribution To evaluate hypotheses with z-scores, we have used the unit normal table to identify the critical region. This table can be used only if the distribution of sample means is normal.

8.4 Directional (One-tailed) Hypothesis Tests

In a directional hypothesis test, or a one-tailed test, the statistical hypotheses (H_0 and H_1) specify either an increase or a decrease in the population mean. That is, they make a statement about the direction of the effect.

Use Symbols such as $<, \leq, \geq, >$ for your alternative hypothesis

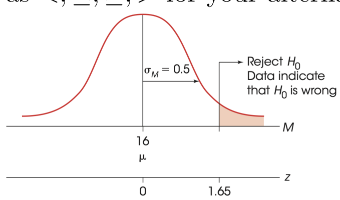


FIGURE 8.6
Critical region for Example 8.3.

8.5 Measuring Effect Size

A measure of effect size is intended to provide a measurement of the absolute magnitude of a treatment effect, independent of the size of the sample(s) being used.

Cohen's D

To Measure effect size we use Cohen's D, a smaller cohen's d = a smaller difference. Cohen's d measures stuff by standard devation. d = 1.00 = seffect of the treatment is equal to one standard deviation

$$\text{Cohen's D} = \frac{M - \mu}{\sigma}$$

Magnitude of d	Evaluation of Effect Size
$d = 0.2$	Small effect (mean difference around 0.2 standard deviation)
$d = 0.5$	Medium effect (mean difference around 0.5 standard deviation)
$d = 0.8$	Large effect (mean difference around 0.8 standard deviation)

Statistical significants tells us: whether the results observed in a study are likely to be real and not just due to random chance.

- It helps researchers determine whether the findings of a study can be generalized to the broader population or if they are simply the result of sampling variability.

* note the size of smaple has no effect on cohen d

8.6 Statistical Power

The power of a statistical test is the probability that the test will correctly reject a false null hypothesis.

- power is the probability that the test will identify a treatment effect if one really exists.

$$p\ (\text{Type II error}) = \beta.$$

Therefore, the power of the test (the second outcome) must be:

$$p\ (\text{rejecting a false } H_0) = 1 - \beta.$$

How to calculate power?

1. Sketch distributions of null and alternative hypotheses
2. Locate the critical region and compute $M_critical$

- use $\alpha = 0.05$, two-tailed test and look for the tail (C column) on your z-score

You find your z-score for the respective alpha score is looking at the unit normal and finding in C column where 0.05 percent chance
Now we have to idenity proporition of altnerative hypothesis which is in the 0.05 chance

$$M_{critical} = \mu_{null} + z(\sigma_M)$$

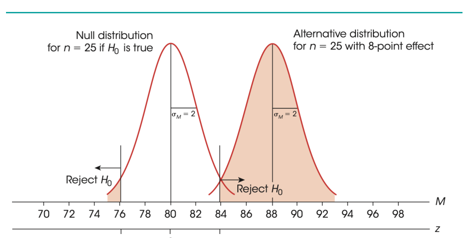
3. Compute Z-score for alternative distribution and find power

- Solve the $M_{critical}$ value it should be a sample mean
- Then use that $M_{critical}$ and use it to calculate your z-score with the altenrate distrubtion

$$z = \frac{M_{critical} - \mu_{alternative}}{\sigma_M}$$

If we find the z-score being greater than 1.95 or less than -1.95 then we know that we can reject the hypothesis

The power refers to the c column in the z-score



Visual Notes

- high power = higher chance to detect erro
- effect size = small $\rightarrow$ statistical power = low
- power = low $\rightarrow$ LESS likely to detect an error
- 1 = type1 + type2β
- increasing sample size increase both likelyhood of rejecting NULL hypothesis and power of test
- larger sample and larger alpha will increase power

Power*	Medium effect size, d = 0.50 with α = .05, two tails	Small effect size, d = 0.20 with α = .05, two tails	Medium effect size, d = 0.50 with α = .01, two tails	Small effect size, d = 0.20 with α = .01, two tails
20%	5	32	13	76
30%	9	52	17	106
40%	12	73	22	135
50%	16	97	27	166
60%	20	123	33	201
70%	25	155	39	241
80%	32	197	47	292
90%	43	263	60	372

*Matlab[TM] was programmed to produce the values in this table. For each treatment effect size and alpha level in the table, power was computed for sample sizes ranging from 1 to 1,000, and the smallest sample size that exceeded the target level of power was recorded. A normal z distribution was assumed in the calculations.

Chapter 9: Intro to the T-statistic

Z-score cons: It requires population standard deviation to compute standard error σ and in most situations we do not know the σ for population

Chapter 9: Intro to the T-statistic

We can find the estimated standard error when we do not have a standard error known

calculating T statistic

$$\text{estimated standard error} = s_m = \frac{s}{\sqrt{n}}$$

$$\text{estimated standard error} = s_m = \sqrt{\frac{s^2}{n}}$$

$$\text{sample variance} = s^2 = \frac{SS}{n-1} = \frac{SS}{df}$$

$$\text{sample standard deviation} = s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}}$$

Estimated standard error s_M is the estimate of actual standard error σ_M when σ_M is unknown. It is found by the sample variance or deviation and gives the standard distance between sample mean M and population mean μ

$$\text{t statistic} = t = \frac{M - \mu}{s_M}$$

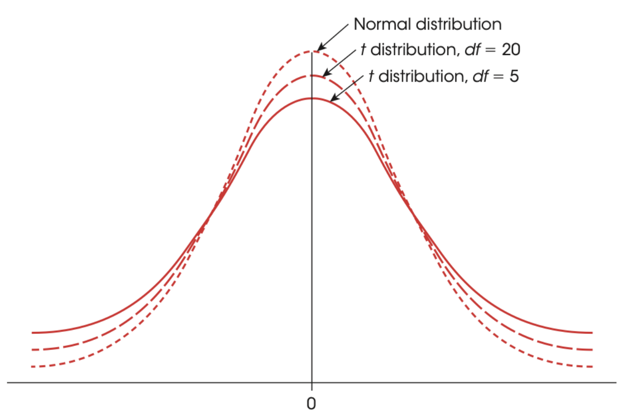
The t statistic tests a hypotheses about an unknown population mean, μ, when σ is unknown. The formula for the t statistic = z-score formula, but that the t-statistic uses the estimated standard error s_M.

$$\text{degrees of freedom} = df = n - 1$$

Degrees of freedom is the number of scores ina. sampel that are independent and free.

t-distribution

- if the sample size is n = 30, then it will be nearly equal to the normal distribution, but if n = 5, then it will start to look less and less like the curve of the normal distrubtion
- the shape of the distribution is dependent to the $df = n - 1$



- t statistics are more variable than z-scores

- t distributions are more flatter and more spread out
- as sample size and df increase, and variability in t-distribution decreases, it will look more like a normal distribution

Probabilities of t dstributions

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707

The orange color 0.05 directional t-value

- expect t statistic when null hypothesis is true = 0

9.2 Hypothesis Tests with the t Statistic

Using the t Statistic for Hypothesis Testing

- the t test does not require any prior knowledge about the population mean or the population variance.
- All you need to compute a t statistic is a null hypothesis and a sample from the unknown population

Hypothesis Testing Example SEE PG 300

Assumptions of a T test

- The values in the sample must consist of independent observations.
  - these two observations much be indepentent (the reslationship between the two events
- The population sampled must be normal.

The Influence of Sample Size and Sample Variance

- larger variance = larger error = more likely H_0 is true
- large variance is bad for inferential statistics
- larger sample is smaller the error

9.3 Hypothesis Tests with the t Statistic

Estimated Cohen's d

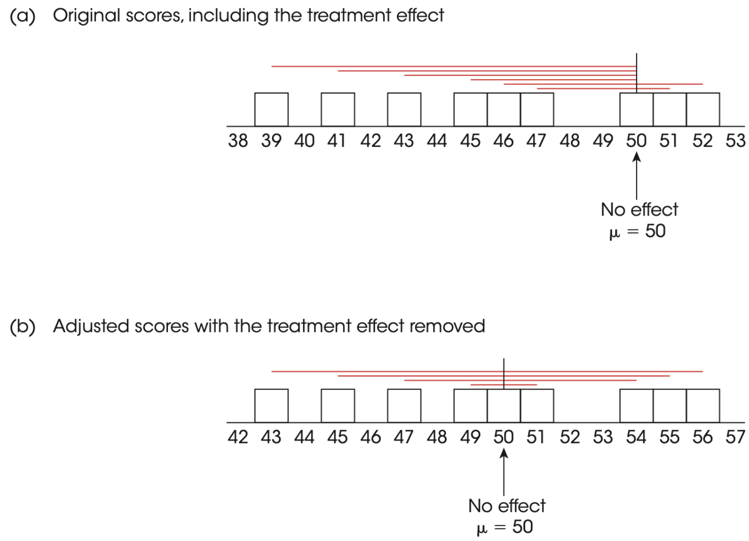
$$\text{Cohen's d} = \frac{\mu_{treatment} - \mu_{no\ treatment}}{\sigma}$$

$$\text{Estimated d} = \frac{M - \mu}{s}$$

Measuring Percent Variance r^2

Another way to measure effect size is to determine the varaibitly in teh scores is explained in the treatment effect.
We do this by adjusting the scores to by deviations to make it more visually appealing and understandable

FIGURE 9.6 Deviations from $\mu = 50$ (no treatment effect) for the scores in Example 9.2. The colored lines in part (a) show the deviations for the original scores, including the treatment effect. In part (b) the colored lines show the deviations for the adjusted scores after the treatment effect has been removed.



$$r^2 = \frac{\text{variability accounted for by the treatment effect}}{\text{total variabiilty}}$$

Using the t statistic

$$r^2 = \frac{t^2}{t^2 + df}$$

Percentage of Variance Explained, r^2	
$r^2 = 0.01$	Small effect
$r^2 = 0.09$	Medium effect
$r^2 = 0.25$	Large effect

Confidence Intervals for Estimating μ

A confidence interval is an interval, or range of values centered around a sample statistic. The logic behind a confidence interval is that a sample statistic, such as a sample mean, should be relatively near to the corresponding population parameter. Therefore, we can confidently estimate that the value of the parameter should be located in the interval near to the statistic.

$$\mu = M \pm t(s_M)$$

- Calculates upper and lower bound by the plus or minus
- M = sample mean
- t = t-static calulated by $t = \frac{M-\mu}{s_M}$
- s_M = estimated standard error $s_M = \frac{s}{\sqrt{n}}$ or $s_M = \sqrt{\frac{s^2}{n}}$

Also to find confidence inteval if a question asks 90 percent confidence you look at the t-distribution. examine the row (df) and the columns = percent confident (look for 10 in combined tails)

Directional Hypotheses and One-Tailed Tests

With directional tests, you use these symbols to showcase direction $\geq, >, \leq, <$

- For H_0 use $\leq, \geq$
- For H_1 use $<, >$

Locate the critical region, keep in mind it will only be on one side

Chapter 10: Independent-Measures Design

10.1 Indepentent Test

Independent test: A research design that uses a separate group of participants for each treatment condition (or for each population)
COMPARING 2 GROUPS, these two groups could be 2 different topics or repeating on same sample group but slightly different (perhaps A study comparing blood pressure before and after a workout.)

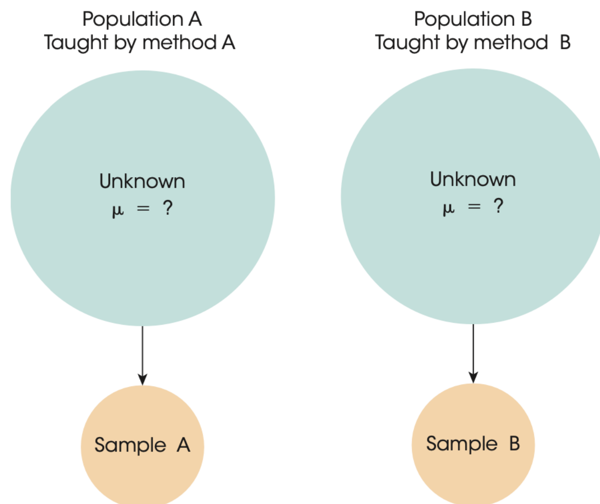


FIGURE 10.1
The structure of an independent-measures research study. Two separate samples are used to obtain information about two unknown populations or treatment conditions.

10.2 The Hypotheses and the Independent-Measures t Statistic

The null is when there is no difference between the two population means

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Lets use two tests so it follows similar to independent- measures t

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{M_1 - M_2}}$$

Numerator is the mean difference and teh denominator is the difference that is expected

Interpreting estimated standard error

To estimate the standard error of $M_1 - M_2$

- measures the standard distance between $(M_1 - M_2)$ and $(\mu_1 - \mu_2)$
- when teh null is true that measure the standard of average size of the difference between the means

Calculating the estimated standard error

$$S_{M_1 - M_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Tihis combines the errors for the first and second sample mean.
This represnts pooled variance only when $n_1 = n_2$
recall

$$s^2 = \frac{SS}{df}$$

When there are two SS values and two df values then we should cominte to create a pooled varaince

$$\text{pooled variance} = s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

This is the average of the two sample variances, you would like to know the pooled variance to calculate the estimated standard error (the pool)

$$\text{estimated standard error of } M_1 - M_2 = s_{(M_1 - M_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Now to find hte t statistic

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}}$$

$$df = df_1 + df_2 = (n_1 - 1) + (n_2 - 1)$$

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample t statistic	M	μ	$\sqrt{\frac{s^2}{n}}$	$s^2 = \frac{SS}{df}$
Independent-measures	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$

10.3 Hypothesis Tests with the Independent-Measures t Statistic

Three assumptions before you use indepnet measure t formula

- The observations within each sample must be independen
- The two populations from which the samples are selected must be normal.
- The two populations from which the samples ares elected must have equal variances.

10.4 Effect Size and Confidence Intervals for the Independent-Measures t

Estimated cohen's d which is the standard measure of mean differnce. It caluculates the difference between the two sample means to estimate the mean diffference

$$\text{estimated d} = \frac{M_1 - M_2}{\sqrt{s_p^2}}$$

To find the variance and r^2 this measures effect size by computing the percentage variance for r^2

$$r^2 = \frac{t^2}{t^2 + df}$$

To calculate confidence intervals then we see the difference

$$\mu_2$$

10.5 The Role of Sample Variance and Sample Size in the Independent-Measures t Test

- I

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707