L12.slides

2020年4月23日 星期四　下午5:20

Lecture12_print-1

MATH 661

Lecture 12

Instructor: Padma Natarajan

1

Confidence interval on $(p_1 - p_2)$

p_1: proportion of successes in population 1 and
p_2: proportion of successes in population 2

The point estimate for the difference in proportions is given by

$$\hat{p}_1 - \hat{p}_2$$

If simple random samples from both independent binomial populations are sufficiently large, then the confidence interval formula shown below can be used to estimate $(p_1 - p_2)$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

point estimate $- m \le p_1 - p_2$ $\le$ point estimate $+ m$

2

Example

- A new treatment for asthma is administered in an inhaler and is compared to a standard treatment administered in the same way.
- A random sample of 375 asthmatic children are selected from a registry, of which 250 are randomized to the new treatment group and 125 are randomized to the comparison group (standard treatment).
- Both groups are provided instruction on the proper use of their inhalers and are followed for 6 months and monitored for ER use.

375　250 new　125 old

3

Example

- Of the children on the new treatment,
  - 60 used the ER during the 6 months for complications of asthma, and
- Of the children on the standard treatment,
  - 19 used the ER during the 6 months for complications of asthma during the same period
- Construct a 95% confidence interval for the **difference in the proportions** of asthmatic children on the **new** and **standard** treatments who *used ER during the 6-month* follow-up period.

4

$c = 95\%$　$\alpha/2 = 0.025$　$z^{-1}(0.975) = 1.96$

Solution

The data layout is as follows:

Asthma Registry: n = 250 New treatment, X = 60 use ER; n = 125 Standard treatment (control), X = 19 use ER

Study Start ... 6 Months

The sample proportions are $\hat{p}_1 = \frac{X_1}{n_1} = \frac{60}{250} = 0.24$

and $\hat{p}_2 = \frac{X_2}{n_2} = \frac{19}{125} = 0.15$

A point estimate for the difference in proportions is given by:

$$\hat{p}_1 - \hat{p}_2 = 0.24 - 0.15 = 0.09$$

5

Solution

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.09 \pm 1.96\sqrt{\frac{0.24(1-0.24)}{250} + \frac{0.15(1-0.15)}{125}}$$

$0.09 \pm 1.96(0.042)$
0.09 ± 0.082 m
$(0.008, 0.172)$

Thus, we are 95% confident that the true difference in the population proportions of asthmatic children on the new treatment as compared to children on standard treatment who used the ER during a 6-month period is between 0.008 and 0.172.

group　control　experiment　features, product.

6

Two-Sample z-Test for difference between Proportions

- Two sample z-test for proportions is used to test the difference between two population proportions, p_1 and p_2.
- Required conditions:
  - Random and independent samples
  - The samples must be large enough

7

Two-Sample z-Test for the Difference Between Proportions

- The **test statistic** is $\hat{p}_1 - \hat{p}_2$
- The **standardized test statistic** is

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad where$$

where $\hat{p}_1 = \frac{X_1}{n_1}$ and $\hat{p}_2 = \frac{X_2}{n_2}$

$\hat{p} = \frac{(X_1 + X_2)}{(n_1 + n_2)}$ Estimate of the common proportion　pool value

8

Hypothesis test on p_1-p_2

1. State the null and alternative hypotheses. State H_0 and H_1. H_0: $p_1 - p_2 = 0$ or $p_1 = p_2$; H_1: $p_1 - p_2 > 0$ or < 0 or $\neq 0$; H_1: $p_1 > p_2$ or $p_1 < p_2$ or $p_1 \neq p_2$
2. Identify the level of significance. Identify α.
3. Determine the critical value(s). Use Table A. Z_α　$\pm Z_{\alpha/2}$　$-Z_\alpha$
4. Determine the rejection region(s).
5. Find the estimate of the common proportion $\hat{p} = \frac{(X_1 + X_2)}{(n_1 + n_2)}$

9

Hypothesis test on p_1-p_2

6. Find the standardized test statistic. $Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\hat{p}_1 = \frac{X_1}{n_1}$ and $\hat{p}_2 = \frac{X_2}{n_2}$
7. Make a decision to reject or fail to reject the null hypothesis. If Z_0 is in the rejection region, reject H_0. Otherwise, fail to reject H_0.

10

Example

- A new drug is being compared to an existing one for its effectiveness in relieving headache pain.
- One hundred subjects who suffer from chronic headaches are randomly assigned to two groups
- Group 1: Existing drug therapy and Group 2: New drug therapy
- Among 50 assigned to the first group, 28 reported relief from headache pain within 60 minutes and
- Among the 50 in the second group, 34 reported relief from headache pain within 60 minutes.

11

Example

- Based on the data, is the proportion of subjects reporting relief from headache pain within 60 minutes under the new drug significantly different from the proportion of subjects reporting headache relief under the existing drug? ≠
- Test using α = 0.05.

12

Solution

1. Set up hypotheses.
   H_0: $p_1 = p_2$
   H_1: $p_1 \neq p_2$,　α=0.05 (Two tailed test)
   p_1 = proportion of patients who experience relief from headache pain using the existing drug and
   p_2 = proportion of patients who experience relief from headache pain using the new drug
2. Select the appropriate test statistic
   The sample proportions are

$$\hat{p}_1 = \frac{28}{50} = 0.56 \text{ and } \hat{p}_2 = \frac{34}{50} = 0.68$$

13

Solution

- The appropriate test statistic is $Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$　$\hat{p}_1 = \frac{28}{50} = 0.56$　$\hat{p}_2 = \frac{34}{50} = 0.68$

3. Decision rule:
   Reject H_0 if $Z_0 < -1.96$ or $Z_0 > 1.96$
   Do not reject H_0 if $-1.96 \le Z_0 \le 1.96$
4. Test statistic:
   First compute $\hat{p} = \frac{(X_1 + X_2)}{(n_1 + n_2)} = \frac{28+34}{50+50} = 0.62$ pool value (0.56, 0.68)

0.025　0.95　0.025　-1.96　1.96

14

Solution

$\hat{p}_1 = \frac{28}{50} = 0.56$　$\hat{p}_2 = \frac{34}{50} = 0.68$　$\hat{p} = \frac{(X_1+X_2)}{(n_1+n_2)} = \frac{28+34}{50+50} = 0.62$

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.56-0.68}{\sqrt{0.62(1-0.62)\left(\frac{1}{50}+\frac{1}{50}\right)}} = \frac{-0.12}{0.097} = -1.24$$

5. Conclusion: Do not Reject H_0 since $-1.96 \le -1.24 \le 1.96$
   We do not have significant evidence, α=0.05, to show a difference in proportions of subjects experiencing relief from headache pain with the New Drug within 60 minutes

15

Statistical inference procedures for two samples

- **For two independent samples**
  - Confidence intervals are concerned with estimating the difference in means (μ_1-μ_2)
  - In Hypothesis testing, the parameter of interest is difference in population means (μ_1-μ_2)
    H_0: $\mu_1 - \mu_2 = 0$;
    H_1: $\mu_1 - \mu_2 \neq 0$ or
    H_1: $\mu_1 - \mu_2 > 0$ or
    H_1: $\mu_1 - \mu_2 < 0$
- **For two dependent samples**
  - Confidence intervals are concerned with estimating the mean difference, μ_d
  - In Hypothesis testing, the parameter of interest is the mean difference, μ_d

16

Hypothesis tests on the difference of means, variances known, independent samples

Objective: To study the difference in means of two populations

Two populations:
Population 1: has mean μ_1 and variance σ_1^2
Population 2: has mean μ_2 and variance σ_2^2

Inferences will be based on two random samples of sizes n_1 and n_2 respectively from the two populations.

17

Hypothesis Tests for a Difference in Means, Variances Known, Independent samples

Null Hypothesis:
H_0: $\mu_1 - \mu_2 = 0$ (no difference in means)
Alternative Hypothesis:
H_1: $\mu_1 - \mu_2 \neq 0$ (means are different) or
H_1: $\mu_1 > \mu_2$ (Mean of population 1 is larger than the Mean of population 2), or
H_1: $\mu_1 < \mu_2$ (Mean of population 1 is smaller than the Mean of population 2)

Test Statistic:

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$　$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

18

Confidence Interval on a Difference in Means, Variances Known

- If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2 from two populations with known variances σ_1^2 and σ_2^2, a $100(1-\alpha)$ % confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Point estimator of $\mu_1 - \mu_2$ is the difference in sample means $\bar{X}_1 - \bar{X}_2$

$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

19

Problem 1

- Following knee replacement surgery, we want to compare two physical therapy programs to see how far patients can walk independently: a 4-day and a 2-day program.
- 18 patients walked a mean of 47.2 feet in the 4-day program, while 16 patients walked a mean of 24.6 feet in the 2-day program.
- Assume that $\sigma_1^2 = 6.4$ and $\sigma_2^2 = 4.2$
- **Compute a 95% confidence interval.**

20

Solution 1

$n_1 = 18$, $\bar{X}_1 = 47.2$ feet, $\sigma_1^2 = 6.4$　$n_2 = 16$, $\bar{X}_2 = 24.6$ feet, $\sigma_2^2 = 4.2$

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$= (47.2 - 24.6) \pm 1.96\sqrt{\frac{6.4}{18} + \frac{4.2}{16}}$
$= 22.6 \pm 1.54$
$= (21.06, 24.14)$

$\mu_1 - \mu_2 = 0$. no different mean & reject H_0.

We are 95% confident that the difference in mean walking distances between patients undergoing physical therapy for 4 versus 2 days is between 21.06 and 24.14 feet

21

Solution 1

If we were to run a test of hypothesis, the hypotheses would be of the form H_0: $\mu_1 - \mu_2 = 0$, H_1: $\mu_1 - \mu_2 \neq 0$.

Since the confidence interval estimate does not include 0, we can conclude that there is a statistically significant difference between patients undergoing physical therapy for 4 versus 2 days.

22

Problem 2

- Hospital management hypothesizes that employees in nonsupervisory positions take significantly more sick days per calendar year than employees in supervisory positions.
- To test this claim, random samples of employees from each job classification is taken and the number of sick days that each took in the last calendar year is recorded.
- A random sample of 100 employees in nonsupervisory positions took a mean of 6.7 sick days in the last calendar year. The variance in the number of sick days among nonsupervisory employees is known to be 2.2 days.

23

Problem 2 continued

- A random sample of 100 employees in supervisory positions took a mean of 4.8 sick days in the last calendar year. The variance in the number of sick days among supervisory employees is known to be 3.1 days.
- Is there enough evidence to support the claim that nonsupervisory employees take off more sick days than supervisory employees? test, $\mu_1 > \mu_2$
- Test at 5% level of significance.

24

Solution 2

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$,　α=0.05　right tail　1.645　0.05

where,
μ_1 = the mean number of sick days per calendar year among employees in nonsupervisory positions
μ_2 = the mean number of sick days per calendar year among employees in supervisory positions

25

Solution 2

- Select the appropriate test statistic
- The test statistic is

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{6.7 - 4.8}{\sqrt{\frac{2.2}{100} + \frac{3.1}{100}}} = \frac{1.9}{0.23} = 8.25$$

Critical value:
Since it is a right tailed test $z_\alpha = z_{0.05} = 1.645$

26

Solution

Conclusion:
We Reject H_0 as $Z_0 > 1.645$
We have significant evidence, α=0.05, to show that the employees in nonsupervisory positions take more sick days per calendar year than employees in supervisory positions (i.e., $\mu_1 > \mu_2$).
P-value:
P-value < 0.05
P-value is less than α, so we reject H_0

$z_0 = 8.26$.
P-value $= 1 - \Phi(8.26) = 1 - 1 = 0$.
Pvalue < α reject H_0.

27

Comparing Two Means Variances unknown and not equal: The Two-Sample t

independent samples

When data come from two random samples or two groups in a randomized experiment, the statistic $\bar{x}_1 - \bar{x}_2$ is our best guess for the value of $\mu_1 - \mu_2$.

When the two samples are independent of each other, the standard deviation of the statistic $\bar{x}_1 - \bar{x}_2$ is:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Since we don't know the values of the parameters σ_1 and σ_2, we replace them in the standard deviation formula with the sample standard deviations. The result is the **standard error** of the statistic $\bar{x}_1 - \bar{x}_2$: $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

We standardize the observed difference to obtain a t statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Use the smaller of $n_1 - 1$ and $n_2 - 1$ for the degrees of freedom. *(conservative degrees of freedom)*

28

Two-Sample t Test

Suppose the Random, Normal, and Independent conditions are met. To test the hypothesis $H_0: \mu_1 - \mu_2 = 0$, compute the t statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Find the P-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_1. Use the t distribution with degrees of freedom given by the smaller of $n_1 - 1$ and $n_2 - 1$ (conservative df).

29

Example 7.15 text

Is there an improvement with directed reading activity? Does the treatment group (group 1) perform better than the control group (group 2)?
We want to perform a test of:
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$　right-tail
where μ_1 = True mean DRP scores from Group 1
and μ_2 = True mean DRP scores from Group 2
We will use $\alpha = 0.05$.

Given:
$n_1 = 21$　smaller
$n_2 = 23$
$\bar{x}_1 = 51.48$
$\bar{x}_2 = 41.52$
$s_1 = 11.01$
$s_2 = 17.15$
σ X know

30

Example 7.15 text

We perform a two-sample t-test for the difference $\mu_1 - \mu_2$.

Test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{[51.48 - 41.52] - 0}{\sqrt{\frac{11.01^2}{21} + \frac{17.15^2}{23}}} = 2.31$$

P-value Using the conservative df = 21 − 1 = 20, we can use Table D to show that the P-value is between 0.01 and 0.02.

Because the P-value is less than α = 0.05, we reject H_0. The data suggests that directed reading activity improves the DRP score at 5% significance level.

$t_{0.05, 20} = 1.725$　smaller one　reject H_0　2.31

31

Confidence Interval for μ_1 - μ_2

Two-Sample t Interval for a Difference Between Means

When the Random, Normal, and Independent conditions are met, a level C confidence interval for $(\mu_1 - \mu_2)$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t^*\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t^* is the critical value at confidence level C for the t distribution with degrees of freedom is equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

32

Example

The Wade Tract Preserve in Georgia is an old-growth forest of longleaf pines that has survived in a relatively undisturbed state for hundreds of years. One question of interest to foresters who study the area is "How do the sizes of longleaf pine trees in the northern and southern halves of the forest compare?" To find out, researchers took random samples of 30 trees from each half and measured the diameter at breast height (DBH) in centimeters. Construct and interpret a 90% confidence interval for the difference in the mean DBH for longleaf pines in the northern and southern halves of the Wade Tract Preserve.

Descriptive Statistics: North, South

From Minitab:

Variable	N	Mean	StDev
North	30	23.70	17.50
South	30	34.53	14.26

$df = 30 - 1 = 29$　$1 - \alpha = 0.9$　$\alpha = 0.1$　$\alpha/2 = 0.05$

33

Example

Our parameters of interest are
μ_1 = the true mean DBH of all trees in the southern half of the forest and
μ_2 = the true mean DBH of all trees in the northern half of the forest.
We want to estimate the difference $\mu_1 - \mu_2$ at a 90% confidence level.

34

Example

We can construct a two-sample t interval for the difference $\mu_1 - \mu_2$. We'll use the conservative df = 30 − 1 = 29.

Upper-tail probability p

df	.10	.05	.025
28	1.313	1.701	2.048
29	1.311	1.699	2.045
30	1.310	1.697	2.042
	80%	90%	95%

Confidence level C

$(\bar{x}_1 - \bar{x}_2) \pm t^*\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (34.53 - 23.70) \pm 1.699\sqrt{\frac{14.26^2}{30} + \frac{17.50^2}{30}} = 10.83 \pm 7.00 = (3.83, 17.83)$

35

Hypothesis tests on the difference of means, variances unknown and equal

(independent sample)

Population variances unknown but assumed to be equal $\sigma_1^2 = \sigma_2^2$

Test statistic

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \quad df = n_1 + n_2 - 2$$

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

S_p is the pooled estimate of the common standard deviation

36 Two sample t for the Mean

Hypothesis tests on the difference of means, variances unknown and equal

- If the sample sizes are equal (i.e., $n_1 = n_2$), then

$$S_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

- Confidence Interval

point estimate $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ standard deviation

$(\bar{X}_1 - \bar{X}_2) - m \le \mu_1 - \mu_2 \le (\bar{X}_1 - \bar{X}_2) + m$

$m = t_{\alpha/2, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$\bar{x} - m \le \mu \le \bar{x} + m$　$m = t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$

37

Example

Fifteen subjects are assigned to the medication treatment and report a mean time to relief of 33.8 minutes with a variance of 2.85. (stdev 1.69)
A second random sample of 15 subjects is assigned to the relaxation treatment and report a mean time to relief of 22.4 minutes with a variance of 3.07.
Do the observed sample means provide evidence of a statistically significant difference in means in the populations?
Run the appropriate test to assess whether there is a significant *difference* in the mean time to relief under the two different treatments using a 5% level of significance.
Assume that population variances are equal.

38

Solution

$n_1 = 15$, $\bar{X}_1 = 33.8$ min, $s_1^2 = 2.85$　$n_2 = 15$, $\bar{X}_2 = 22.4$ min, $s_2^2 = 3.07$

1. Set up hypotheses.
   H_0: $\mu_1 = \mu_2$
   H_1: $\mu_1 \neq \mu_2$.　α=0.05　(two sided test)
   where, μ_1 = mean time to relief with medication treatment
   μ_2 = mean time to relief with relaxation treatment
2. Select the appropriate test statistic
   Assuming the population variances are equal the test statistic is $t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$t_{\alpha/2, n_1+n_2-2}$
$t_{0.025, 15+15-2} = t_{0.025, 28}$
2.048.

39

Solution

3. Decision rule
The appropriate critical values are found from Table.
We first compute degrees of freedom:
df=n_1+n_2-2=15+15-2=28.　α=0.05 (two-tailed)
The critical values are $t_{\alpha/2}$ = ±2.048.
Rejection Criteria:
Reject H_0 if $t_0 \ge 2.048$ or if $t_0 \le -2.048$
Do not reject H_0 if $-2.048 < t_0 < 2.048$

$-t_{\alpha/2, df}$　$t_{\alpha/2, df}$　-2.048　2.048　18.10　reject H_0

40

Solution

4. Computing Test statistic　sample size same $n_1 = n_2$

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(15-1)(2.85) + (15-1)(3.07)}{15+15-2}} = \sqrt{2.96} = 1.72$$

(Notice that S_p^2 (2.96) falls between the values of the two sample variances.) As sample sizes are equal here the following formula can also be used.

$$S_p = \sqrt{\frac{s_1^2 + s_2^2}{2}} = \sqrt{\frac{2.85 + 3.07}{2}} = 1.72$$

pooled values

41

Solution

Test Statistic

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{33.8 - 22.4}{1.72\sqrt{\frac{1}{15} + \frac{1}{15}}} = \frac{11.4}{0.63} = 18.10$$

5. Conclusion:
Reject H_0 because 18.10 ≥ 2.048.
We have significant evidence, α = 0.05, to show that the mean time to relief from headache pain is different under medication as compared to relaxation treatment.
P-value < 2(0.0005). (From Table) i.e., P-value < 0.001
P-value < α, Reject H_0

42

0.0005　Pvalue < 2(0.0005) < α. reject H_0.

df X in table. minitab. $t_{\alpha/2, 38}$

Example 2

- To investigate the number of hours that graduate students work in addition to full-time class load:
- A random sample of 20 male graduate students is selected who work a mean of 16.4 hours per week with a variance of 4.7. (stdev 2.17)
- A second random sample of 20 female graduate students is selected who work for a mean of 13.8 hours per week with a variance of 6.1.
- Construct a 95% confidence interval for the difference in mean number of hours worked between male and female graduate students.
- Assume that the variances are equal.

43

Solution

male $n_1 = 20$　$\bar{X}_1 = 16.4$　$s_1^2 = 4.7$　df=38
female $n_2 = 20$　$\bar{X}_2 = 13.8$　$s_2^2 = 6.1$
Assume $\sigma_1^2 = \sigma_2^2$.

$(\bar{X}_1 - \bar{X}_2) \pm t_\alpha s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$　$n_1 = n_2$　$S_p = \sqrt{\frac{s_1^2 + s_2^2}{2}} = \sqrt{\frac{4.7 + 6.1}{2}} = \sqrt{5.4} = 2.32$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{19(4.7) + 19(6.1)}{20+20-2}} = \sqrt{5.4} = 2.32$$

95% Confidence interval
$(16.4 - 13.8) \pm (2.024)(2.32)\sqrt{\frac{1}{20} + \frac{1}{20}}$
2.6 ± 1.50
$(1.1, 4.1)$　$t_{\alpha/2, 38}$

44

Pooled Two-Sample Procedure

Using Minitab:

Stat > Basic Statistics > Two sample t

45

Dependent Samples: data are matched or paired

Here, the observations from the two populations of interest are collected in **pairs.**

To perform a two-sample hypothesis test with dependent samples, the difference between each data pair is first found:

$d_j = X_{1j} - X_{2j}$, j=1,2,…n　Difference between entries for a data pair.

46

Data are matched or paired

Suppose we are interested in testing
H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$

The test statistic is

$$t_0 = \frac{\bar{X}_d - \mu_d}{s_d/\sqrt{n}}, \text{ where } \bar{X}_d = \frac{\sum d}{n}$$

$\bar{X}_d$ is the sample average of the n differences d_1, d_2, …,d_n and
s_d is the sample standard deviation of these differences.

47

Data are matched or paired

Test Statistic: $t_0 = \frac{\bar{X}_d - \mu_d}{s_d/\sqrt{n}}$　Confidence Interval: $\bar{X}_d \pm t_\alpha\frac{s_d}{\sqrt{n}}$

$\bar{X}_d$ & s_d are the mean and standard deviation of the difference scores

$$\bar{X}_d = \frac{\sum d}{n} \quad s_d^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}, \quad s_d = \sqrt{s_d^2}$$

48

Paired t-test

Null Hypothesis	Test Statistic	Alternative Hypothesis	Rejection Criteria
H_0: $\mu_d = 0$	$t_0 = \frac{\bar{X}_d - \mu_d}{s_d/\sqrt{n}} = \frac{\bar{X}_d}{s_d/\sqrt{n}}$	H_1: $\mu_d \neq 0$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
		H_1: $\mu_d > 0$	$t_0 > t_{\alpha,n-1}$
		H_1: $\mu_d < 0$	$t_0 < -t_{\alpha,n-1}$

49

Example

- A nutrition expert is examining a weight-loss program to examine its effectiveness (i.e., if participants lose weight on the program). reduction
- Ten subjects are randomly selected for the investigation.
- The subjects' initial weights are recorded, they follow the program for 6 weeks, and they are weighed again.

X mention 2-tail

50

Example

Subject	Initial Weight	Final Weight
1	180	165
2	142	138
3	126	128
4	138	136
5	175	170
6	205	197
7	116	115
8	142	128
9	157	144
10	136	130

The samples of initial and final weights are matched by participants.

That is, for each participant an initial and a final weight are recorded and these two measurements are specific to each individual.

51

Solution

The test of interest is:
H_0: $\mu_{initial} = \mu_{final}$
H_1: $\mu_{initial} > \mu_{final}$
where $\mu_{initial}$ = mean initial weight
μ_{final} = mean final weight

Set up hypotheses
H_0: $\mu_d = 0$
H_1: $\mu_d > 0$

52

Solution

Subject	Initial Weight	Final Weight	Difference (d)	d^2
1	180	165	15	225
2	142	138	4	16
3	126	128	-2	4
4	138	136	2	4
5	175	170	5	25
6	205	197	8	64
7	116	115	1	1
8	142	128	14	196
9	157	144	13	169
10	136	130	6	36
			Sum: 66	Sum : 740

$$\bar{X}_d = \frac{\sum d}{n} = \frac{66}{10} = 6.6$$

53

Solution

$$s_d^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1} = \frac{740 - \frac{(66)^2}{10}}{9} = \frac{740 - 435.6}{9} = \frac{304.4}{9} = 33.82$$

$s_d = \sqrt{33.82} = 5.82$

1) Set up hypotheses
H_0: $\mu_d = 0$
H_1: $\mu_d > 0$ right tail.
Where μ_d is the mean difference in weights or mean weight loss
2) Select the appropriate test statistic (The test statistic is t)

54

textbook problem

Solution

$\bar{X}_d = \frac{\sum d}{n} = \frac{66}{10} = 6.6$　$s_d = \sqrt{33.82} = 5.82$

$$t_0 = \frac{\bar{X}_d - \mu_d}{s_d/\sqrt{n}} = \frac{6.6 - 0}{5.82/\sqrt{10}} = 3.59$$

df=n-1=10 -1 =9
3) Critical value (Right tailed test)
$t_{\alpha,n-1} = t_{0.05,9} = 1.833$
Since $t > t_{\alpha,n-1}$, i.e., 3.59 > 1.833 we would reject H_0.
We have significant evidence, α =0.05, to show that the mean weight loss following 6 weeks of the program is greater than zero. P-value: ?
0.0025 < P-value < 0.005　P-value < α

1.833　3.59　reject H_0

	0.005	0.0025
9	3.250	3.690

3.59　reduction　Pvalue ∈ (0.0025, 0.005) < 0.005 = α

55

Solution Problem

Subject	d_j	d_j^2
1	19.2	368.64
2	5.6	31.36
3	-0.6	0.36
4	-17.2	295.84
5	0.6	0.36
6	-5	25
7	7	49
8	26	676
9	5.8	33.64
10	1.2	1.44
11	-6.2	38.44
12	0.6	0.36
13	4	16
14	-24	576
	Σd_j=17	Σd_j^2=2112.44

$$s_d^2 = \frac{2112.44 - \frac{(17)^2}{14}}{13} = \frac{2091.797}{13} = 160.9075$$

$s_d = \sqrt{160.9075} = 12.68$

$\bar{X}_d = \frac{\sum d_j}{n} = \frac{17}{14} = 1.21$

56

Solution

$$\bar{X}_d \pm t\frac{s_d}{\sqrt{n}}$$

$1.21 \pm 1.771\frac{(12.68)}{\sqrt{14}}$
$(-4.79, 7.21)$

$t_{0.05, 13}$　$1 - \alpha = 0.9$　$\alpha = 0.1$　$\alpha/2 = 0.05$　$\mu_d = 0$

Notice that the confidence interval on μ_D includes zero. This implies that, at the 90% level of confidence, the data do not support the claim that the two cars have different mean parking times μ_1 and μ_2. That is, the value $\mu_D = \mu_1 - \mu_2 = 0$ is not inconsistent with the observed data.

57

Two sample Hypothesis Test on μ

Dependent or paired samples → Paired t test: $t_0 = \frac{\bar{X}_d}{S_d/\sqrt{n}}$

Independent Samples →
- Variance Known: $z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
- Variance unknown and not equal: $t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$; Degrees of freedom given by $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$ or smaller of $(n_1 - 1)$ and $(n_2 - 1)$ (if computed by software) conservative df
- Variance unknown and equal: $t_0 = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

58