Statistics for the Sciences

Hypothesis Testing of Population Means

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Components of a hypothesis test
- Alternative and Null Hypothesis
- Test Statistic
- p-value method
- Hypothesis Testing and Confidence Intervals
- Power and sample size
- Lab 1: Hypothesis test of μ
- Lab 2: Hypothesis test of $\mu_1 - \mu_2$

Components of a hypothesis test

- Components of a hypothesis test
  - Null Hypothesis (H_0) and Alternative Hypothesis (H_1 or H_a)
  - Test Statistic
  - Make a decision as to reject or not to reject the Null Hypothesis H_0
  - Conclusion: A statement that uses simple nontechnical wording that addresses the original claim

- Any statement regarding the value of a **population parameter** is called a hypothesis.

Alternative and Null Hypothesis

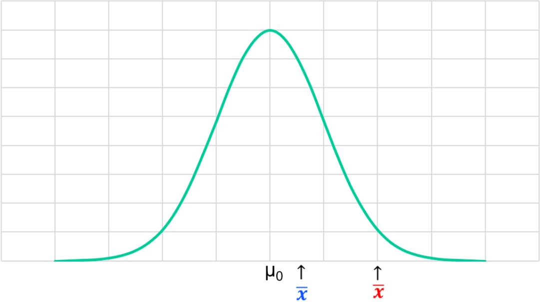
- The alternative hypothesis (denoted by H_1 or H_a) is the statement that a researcher wishes to support.
  - If we are conducting a study and want to use a hypothesis test to support our claim, the claim must be worded so that it becomes the alternative.
  - The symbolic form of the alternative hypothesis must use one of these symbols: $>$, $<$, $\neq$.
- The null hypothesis H_0 is contradictory to H_1.

Signs in H_0 and H_1 and Tails of a Test

	Two-sided Test	**Left-sided Test**	**Right-sided Test**
Sign in the null hypothesis H_0	$=$	$=$ or $\geq$	$=$ or $\leq$
Sign in the alternative hypothesis H_1	$\neq$	$<$	$>$

Alternative and Null Hypothesis

- Hypothesis testing main concept
  - We test the null hypothesis directly in the sense that we assume it is true and reach a conclusion to decide whether H_0 should be rejected in favor of H_1.
  - Hypothesis testing follows the Rare Event Rule: If, under a given assumption (H_0 is true in the hypothesis testing problems), the probability of a particular observed event is extremely small, we conclude that the assumption H_0 is probably not correct.
- For example, Test $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$

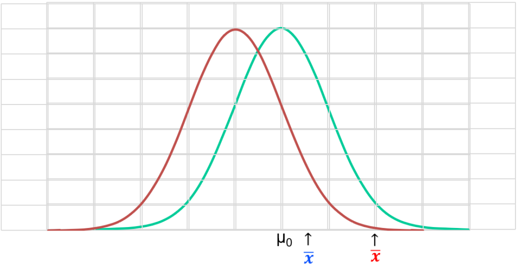


Alternative and Null Hypothesis

- Why we always use $=$ in the null hypothesis H_0?

$$H_0 : \textit{parameter} = \textit{hypothesized} \;\; \textit{value}$$

  - We always use a single fixed value in the construction of a test statistic to have a single sampling distribution.
  - If $H_0 : \mu = \mu_0$ (versus $H_1 : \mu > \mu_0$) is rejected, $H_0 : \mu \leq \mu_0$ will be rejected as well.



Alternative and Null Hypothesis

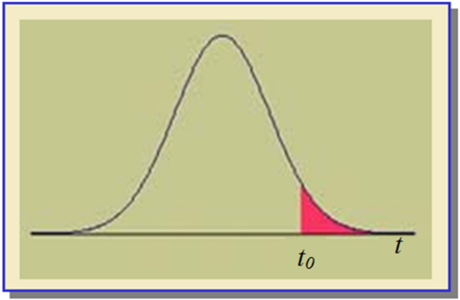
- Example: A researcher claims that a new fertilizer will result in an average plant height greater than 30 cm after 6 weeks of growth. Use μ as the true average height of the plants treated with the new Fertilizer. Which one of the following is the correct null and alternative?
  - **A** $H_0 : \mu < 30$ versus $H_1 : \mu \geq 30$
  - **B** $H_0 : \mu = 30$ versus $H_1 : \mu < 30$
  - **C** $H_0 : \mu > 30$ versus $H_1 : \mu \leq 30$
  - **D** $H_0 : \mu = 30$ versus $H_1 : \mu > 30$

Test Statistic

- Test Statistic is a single statistic calculated from the sample which will allow us to reject or not reject H_0. It is constructed by converting the point estimate (such as $\overline{X}$ and $\hat{p}$ to a standard score (such as t and z).
  - It is a random variable.
- After a test statistic is determined, we can then decide if H_0 can be rejected
  - The most popular method is the p-value method.

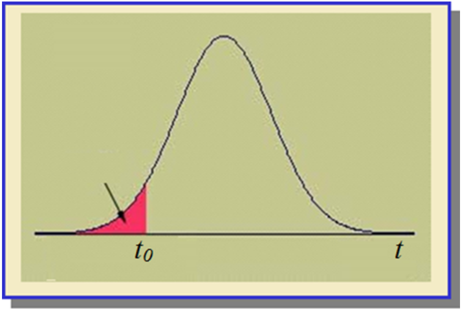
p-value method

- p-value is the probability of observing, just by chance, a test statistic as extreme as or more extreme than the one observed in the direction of H_1.
  - p-value tells us how surprised we would be to get these data given H_0 is true.
  - p-value measures the strength of the evidence against H_0. The smaller the p-value is, the stronger the evidence against H_0.
- Case 1: $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$



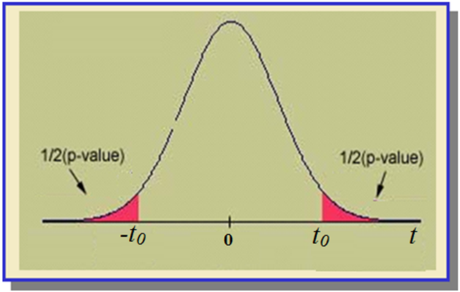
p-value method

- Case 2: $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$



p-value method

- Case 3: $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$



p-value method

Table 1: Summary of One-Sample Tests: Test statistic, Distribution of the test statistic under H_0, and p-value

Population	$N(\mu, \sigma^2)$ or n large $\theta = \mu$	$N(\mu, \sigma^2)$, σ unknown $\theta = \mu$	$b(1, p)$, p unknown $\theta = p$, $np_0, n(1 - p_0) > 5$
Test statistic.	$Z_0 = \frac{\overline{Y}_n - \mu_0}{\sigma/\sqrt{n}}$ or $\frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$t_0 = \frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$Z_0 = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(n-1)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta > \theta_0$	p-value=$P(Z \geq Z_0)$	p-value=$P(t \geq t_0)$	p-value=$P(Z \geq Z_0)$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta < \theta_0$	p-value=$P(Z \leq Z_0)$	p-value=$P(t \leq t_0)$	p-value=$P(Z \leq Z_0)$
Two-sided. $H_0 : \theta = \theta_0$ $H_a : \theta \neq \theta_0$	p-value=$2P(Z \geq \|Z_0\|)$	p-value=$2P(t \geq \|t_0\|)$	p-value=$2P(Z \geq \|Z_0\|)$

p-value method

Table 2: Summary of Two-Sample Tests at level α: Test statistic, Distribution of the test statistic under H_0, and p-value

Population	Normal or n large $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 \neq \sigma_2$ $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 = \sigma_2$ $\theta = \mu_1 - \mu_2$	p_1 and p_2 unknown $\theta = p_1 - p_2$ and $n_1\overline{p}$, $n_2\overline{p}$, $n_1(1-\overline{p})$, $n_2(1-\overline{p}) > 5$
Test statistic.	$Z_0 = \dfrac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ or $Z_0 = \dfrac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \dfrac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \dfrac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$Z_0 = \dfrac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(df = ?)$	$t_0 \sim t(n_1 + n_2 - 2)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 > \theta_2$	p-value$=P(Z \geq Z_0)$	p-value$=P(t \geq t_0)$	p-value$=P(t \geq t_0)$	p-value$=P(Z \geq Z_0)$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 < \theta_2$	p-value$=P(Z \leq Z_0)$	p-value$=P(t \leq t_0)$	p-value$=P(t \leq t_0)$	p-value$=P(Z \leq Z_0)$
Two-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 \neq \theta_2$	p-value$=2P(Z \geq \lvert Z_0\rvert)$	p-value$=2P(t \geq \lvert t_0\rvert)$	p-value$=2P(t \geq \lvert t_0\rvert)$	p-value$=2P(Z \geq \lvert Z_0\rvert)$

Note. (1) $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ and $\overline{p} = \frac{n_1\widehat{p_1}+n_2\widehat{p_2}}{n_1+n_2}$. (2) For testing $H_0 : \mu_1 = \mu_2$ when the two population variances are not equal, the *df* of the t-distribution is determined by Welch-Scatterthwaite equation.

p-value method

- Statistical significance
  - If the p-value is less than or equal to 0.01, reject H_0. The results are highly significant.
  - If the p-value is between 0.01 and 0.05, reject H_0. The results are statistically significant.
  - If the p-value is between 0.05 and 0.10, do not reject H_0. But, the results are tending towards significance.
  - If the p-value is greater than 0.10, do not reject H_0. The results are not statistically significant.
- If p-value is not small, we do not say `accept $H_0$` since we did not prove that H_0 is true.

Hypothesis Testing and Confidence Intervals

- Both Confidence Intervals and Hypothesis Testing results will be produced by a software in general.
- Suppose that the confidence level is 100(1-α) where α is determined by users
  - H_0 is rejected if the p-value is less than or equal to α
  - H_0 is not rejected if the p-value is greater than α
  - If the value of the parameter under H_0 falls in the confidence interval, we do not reject H_0. Otherwise we reject H_0 in favor of H_a.

Power and sample size

Note that

- α = P(Type I error) = P(Rejecting $H_0 \mid H_0$) is called significance level
- β = P(Type II error) = P(Not Rejecting $H_0 \mid H_a$)
- Power of a test $= 1 - \beta$. High power means a high probability of detecting an effect or difference (parameter values between H_0 and H_a) when it exists.
- How do we calculate the power of a test?
  - $1 - \beta$ depends on the difference of the mean values μ_0 and μ_1 under H_0 and H_a, respectively.
  - $1 - \beta$ depends on the spread of the data.
  - $1 - \beta$ depends on α and n.

Power and sample size

- In general, as n becomes larger, $1 - \beta$ becomes larger for fixed H_0 and H_a.
- $\Delta = \dfrac{\mu_0 - \mu_1}{\sigma}$ is called **Effect Size** which needs to be estimated before calculating required sample size.
  - We can use software to calculate required sample size for given **spread of data**, **effect size** and **significance level** α.
- Rule of Thumb (for 1-sample test)

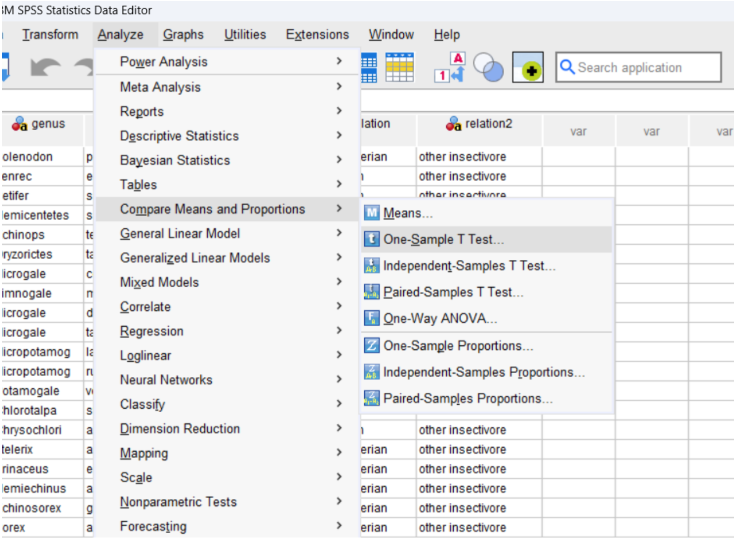
$$n = \frac{8}{\Delta^2}, \text{where } \Delta = \frac{\mu_0 - \mu_1}{\sigma}.$$

  - The formula is $n = \dfrac{16}{\Delta^2}$ for 2-sample tests.

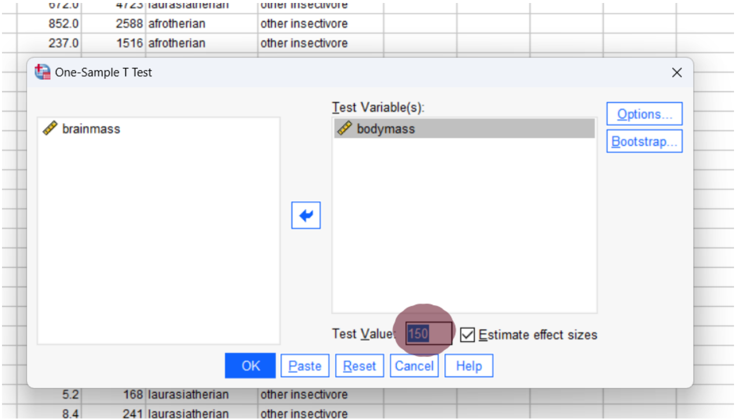
Lab 1: Hypothesis test of μ

- Consider the variable `bodymass` in data `kaufman.csv`
- Suppose $H_a : \mu > 150$
- Write the four steps of hypothesis tesing with the help of SPSS.

Lab 1: Hypothesis test of μ



Lab 1: Hypothesis test of μ



Lab 1: Hypothesis test of μ

T-Test

One-Sample Statistics

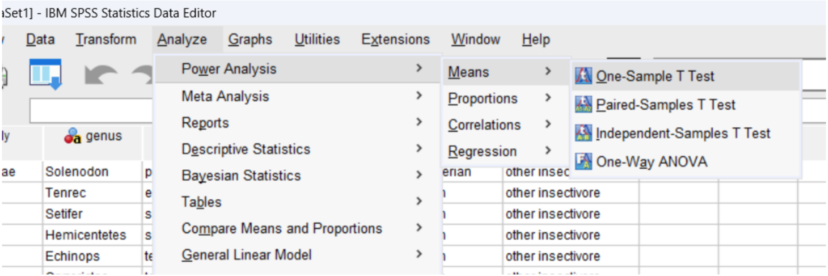
	N	Mean	Std. Deviation	Std. Error Mean
bodymass	56	154.086	253.4390	33.8672

One-Sample Test

	Test Value = 150						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
bodymass	.121	55	.452	.904	4.0857	-63.786	71.957

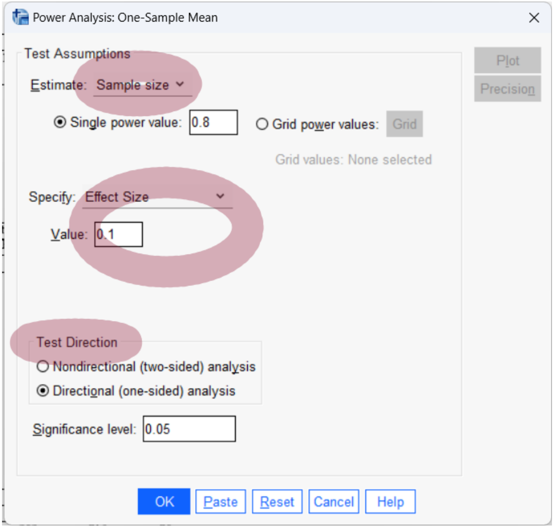
Lab 1: Hypothesis test of μ

- Sample size calculation



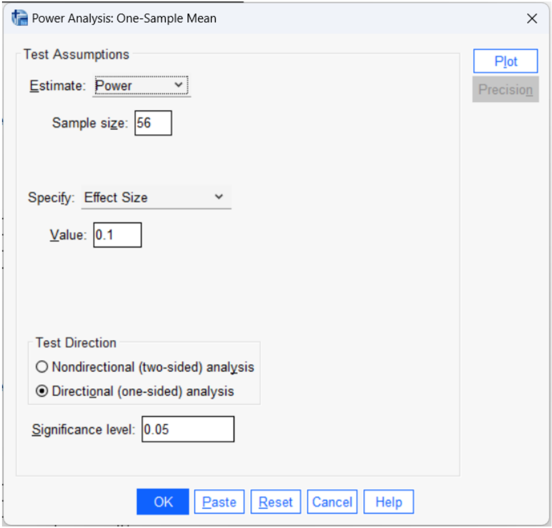
Lab 1: Hypothesis test of μ

- Sample size calculation



Lab 1: Hypothesis test of μ

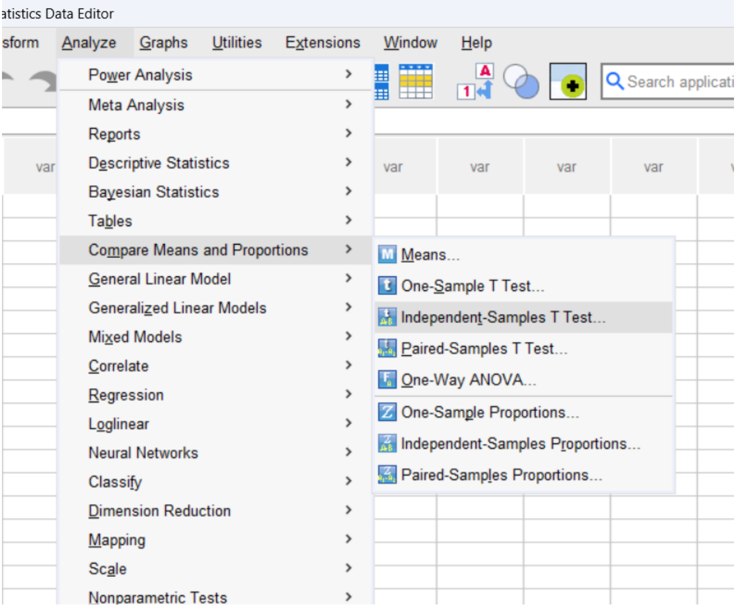
- Power of the test



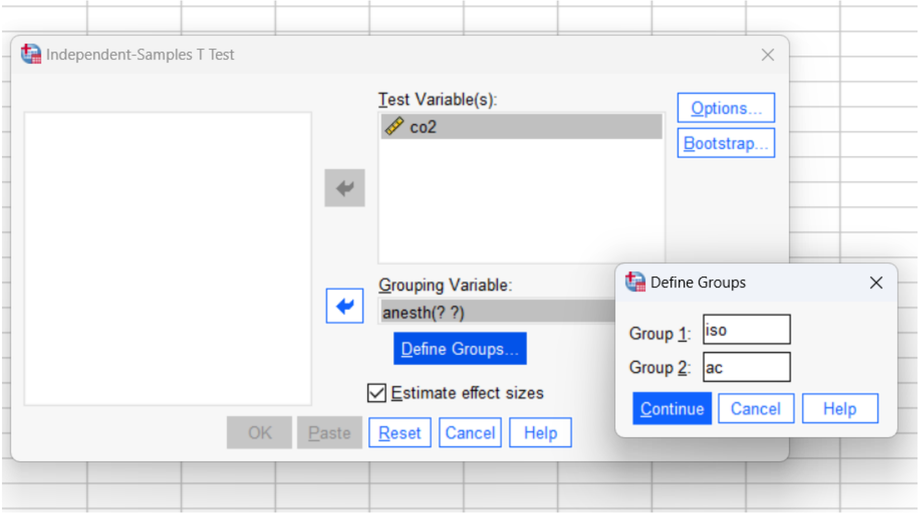
Lab 2: Hypothesis test of $\mu_1 - \mu_2$

- Consider `lowco2.csv` data. Low et al (2016) examined the effects of two different anesthetics on aspects of the physiology of the mouse. Twelve mice were anesthetized with isoflurane and eleven mice were anesthetized with alpha chloralose and blood CO_2 levels were recorded after 120 minutes. The H_0 was that there is no difference between the anesthetics in the mean blood CO_2 level. This is an independent comparison because individual mice were only given one of the two anesthetics.
- Write the four steps of hypothesis tesing with the help of SPSS.

Lab 2: Hypothesis test of $\mu_1 - \mu_2$



Lab 2: Hypothesis test of $\mu_1 - \mu_2$



Lab 2: Hypothesis test of $\mu_1 - \mu_2$

[DataSet1]

Group Statistics

	anesth	N	Mean	Std. Deviation	Std. Error Mean
co2	iso	12	50.00	11.394	3.289
	ac	11	70.91	20.201	6.091

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
co2	Equal variances assumed	4.144	.055	-3.093	21	.003	.006	-20.909	6.761	-34.969	-6.849
	Equal variances not assumed			-3.021	15.485	.004	.008	-20.909	6.922	-35.623	-6.195

License