REGIONAL TRAINING COURSE ON APPLIED ECONOMETRIC ANALYSIS (SUMMER SCHOOL) FOR YOUNG ECONOMISTS / RESEARCHERS ORGANIZED BY WIUT AND IFPRI
JUNE 4-15, 2018

Introduction to statistical inference

Farrukh Ataev
Lecturer, WIUT
6 June 2018

Outline

- **Session 5 Introduction to statistical inference**
  - (Tue 6 June at 13:30-15:00)
- **Session 6 Introduction to statistical inference**
  - (Tue 6 June at 15:30-17:00)

Session 5 outline

- Hypothesis testing: one-tailed and two-tailed
- Hypothesis testing: one sample

Hypothesis testing

- Hypothesis testing is a statistical procedure that uses sample data to determine whether or not a statement about the value of a population parameter should be rejected
- Two cases for sample size: large ($n \geq 30$) or small ($n < 30$)
- The hypotheses must be two competing statements:

  a null hypothesis H_0 and an alternative hypothesis H_a
- H_0: the hypothesis tentatively assumed true
- H_a: the hypothesis opposite to the null hypothesis

Examples of hypothesis test

1. H_0: the defendant is innocent
   H_a: the defendant is guilty.
2. The manager of a hotel stated that the mean guest bill for a weekend is $400 or less.

$$H_0: \mu \leq 400$$
$$H_a: \mu > 400$$

3. The manager of an automobile store wants to introduce a new bonus plan that will increase sales volume. Currently, the mean sales volume is 14 cars per month. The manager conducts a research to see if the bonus plan enables to increase sales volume.

$$H_0: \mu \geq 14$$
$$H_a: \mu < 14$$

4. The label on a 450 gram can of sour cream states that the can contains 15gram of fat.

$$H_0: \mu = 15$$
$$H_a: \mu \neq 15.$$

To test these competing statements (hypotheses), a trial is held and based on the sample information the null hypothesis is either rejected or accepted.

Errors in hypothesis testing

- Since the hypothesis tests are based on sample information, errors are possible.
- There are two kinds of errors:

		State of Nature (reality)	
		H_0 is True	H_a is True
Conclusion	Accept H_0	Correct conclusion	Type II Error (β)
	Reject H_0	Type I Error (α)	Correct conclusion

H_0: the defendant is innocent
H_a: the defendant is guilty.

One-tailed hypothesis test about a population mean

Large-sample ($n \geq 30$) hypothesis test about a population mean of the form

$$H_0: \mu \geq \mu_0$$
$$H_a: \mu < \mu_0$$

Test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$. If σ is unknown, substitute s for σ in computing z.

Rejection rule at a level of significance of α: Reject H_0 if $z < -z_\alpha$.

(Rejection rule at a level of significance of α: Reject H_0 if $p < \alpha$).

Large-sample ($n \geq 30$) hypothesis test about a population mean of the form

$$H_0: \mu \leq \mu_0$$
$$H_a: \mu > \mu_0$$

Test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$. If σ is unknown, substitute s for σ in computing z.

Rejection rule at a level of significance of α: Reject H_0 if $z > z_\alpha$.

(Rejection rule at a level of significance of α: Reject H_0 if $p < \alpha$).

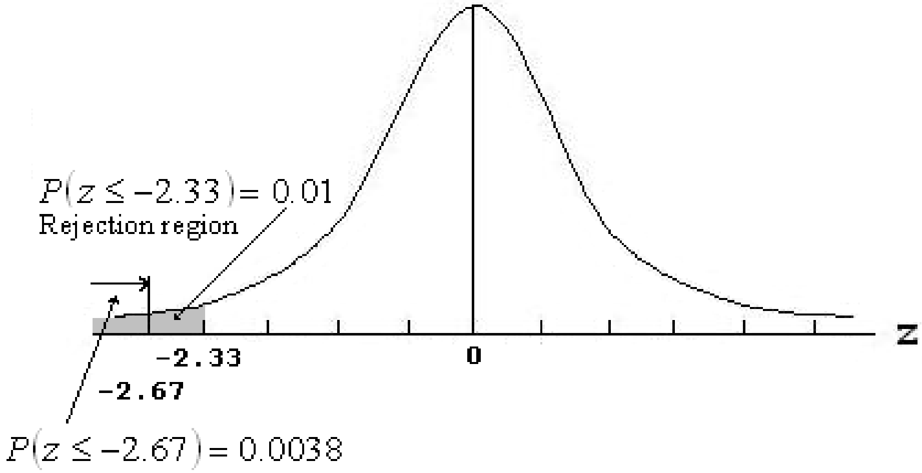
Note: The value of z that establishes the boundary of the rejection region is called a critical value.

One-tailed hypothesis test about a population mean (large sample)

Case 1: The label on a large can of ABC coffee states that the can contains at least 3 pounds (1 pound = 453.6g) of coffee. A sample of 36 cans provided the mean of 2.92 pounds and it is known from previous studies that the population standard deviation is $\sigma = 0.18$. Test the ABC's claim at $\alpha = 0.01$.

Solution: $n = 36$, $\bar{x} = 2.92$, $\sigma = 0.18$, $\alpha = 0.01$.

$$H_0: \mu \geq 3$$
$$H_a: \mu < 3$$



Reject H_0 if $z < -z_\alpha$ (or $p < \alpha$).

Critical value: $z_\alpha = z_{0.01} = 2.33$.

Test statistic: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{2.92-3}{0.18/\sqrt{36}} = -2.67$.

$$p = P(z \leq -2.67) = 0.0038.$$

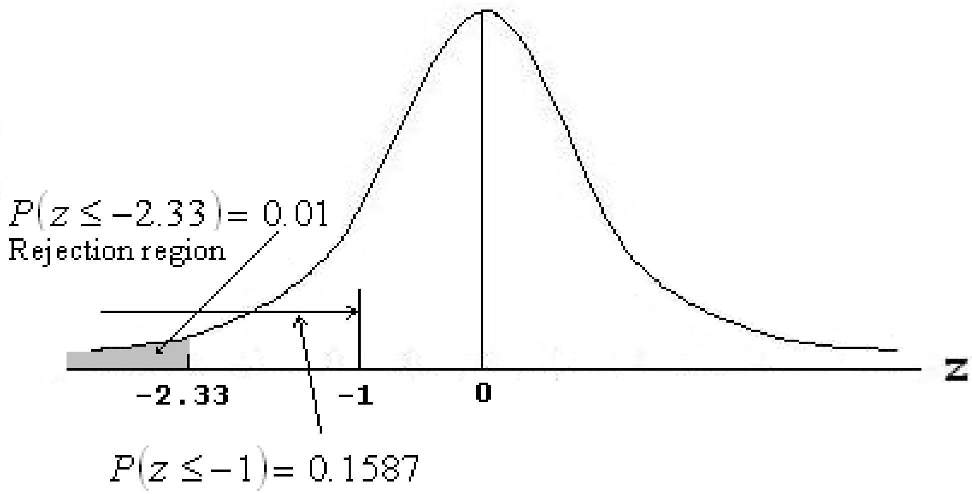
Reject H_0 because $-2.67 < -2.33$ (or $0.0038 < 0.01$).

One-tailed hypothesis test about a population mean (large sample)

Exercise: The label on a large can of ABC coffee states that the can contains at least 3 pounds (1 pound = 453.6g) of coffee. A sample of 36 cans provided the mean of 2.97 pounds and it is known from previous studies that the population standard deviation is $\sigma = 0.18$. Test the ABC's claim at $\alpha = 0.01$.

Solution: $n = 36, \bar{x} = 2.97, \sigma = 0.18, \alpha = 0.01.$

$$H_0: \mu \geq 3$$
$$H_a: \mu < 3$$



Reject H_0 if $z < -z_\alpha$ (or $p < \alpha$).

Critical value: $z_\alpha = z_{0.01} = 2.33$.

Test statistic: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{2.97-3}{0.18/\sqrt{36}} = -1$.

$$p = P(z \leq -1) = 0.1587.$$

Do not reject H_0 because $-1 \nless -2.33$ (or $0.1587 \nless 0.01$).

Two-tailed hypothesis test about a population mean

Large-sample ($n \geq 30$) hypothesis test about a population mean of the form

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

Test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$. If σ is unknown, substitute s for σ in computing z.

Rejection rule at a level of significance of α: Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$.

(Rejection rule at a level of significance of α: Reject H_0 if $p < \alpha/2$).

Two-tailed hypothesis test about a population mean (large sample)

Case 2: In the XYZ company it is assumed that the average weight of a chocolate product is 400 grams. The standard deviation of these weights is 20 grams. A sample of 100 items is taken and found to have a mean of 402 grams. Test the hypothesis at $\alpha = 0.05$.

Solution: $\sigma = 20$, $n = 100$, $\bar{x} = 402$.

$$H_0: \mu = 400$$
$$H_a: \mu \neq 400$$

Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ (or $p < \alpha/2$).

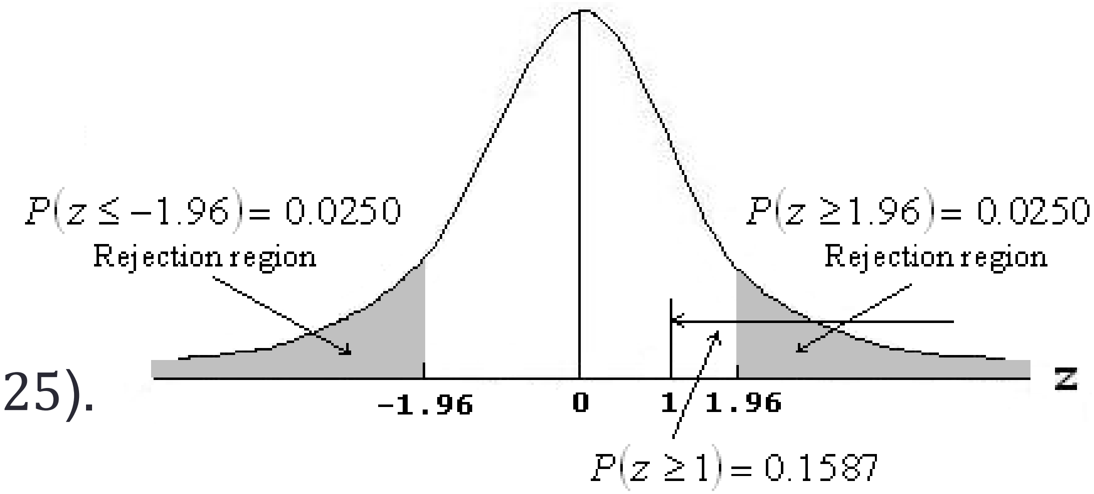
Critical value: $z_{\alpha/2} = z_{0.025} = 1.96$.

Test statistic: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{402-400}{20/\sqrt{100}} = 1.$

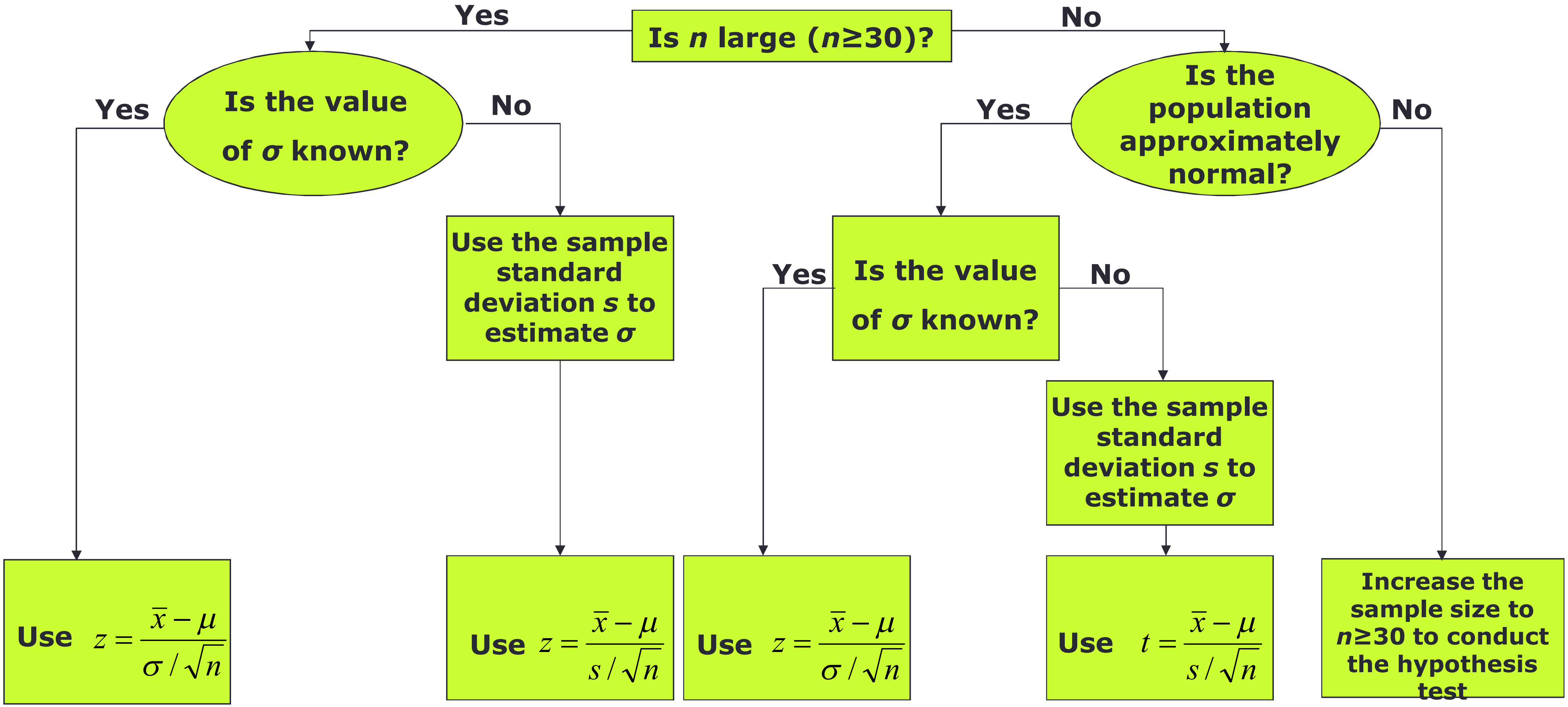
$$p = P(z \geq 1) = 0.1587.$$

Do not reject H_0 because

$1 \nless -1.96$ or $1 \ngtr 1.96$ or (or $0.1587 \nless 0.025$).



Summary of the test statistics to be used for hypothesis tests about a population mean



Two-tailed hypothesis test about a population mean (small sample)

Case 3: The average daily sales income in a company is assumed to be \$2000. Over a sample of 20 days the sales incomes average \$1800 per day, with a standard deviation of \$300 per day. Use a suitable hypothesis test to examine the assumption at $\alpha = 0.05$.

Solution: n = 20, $\bar{x} = 1800$, s = 300.

H_0: μ = 2000 H_a: μ ≠ 2000

Reject H_0 if $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$ (or $p < \alpha/2$)

Critical value: $t_{\alpha/2;19} = t_{0.025;19} = 2.093$.

Test statistic: $t = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1800-2000}{300/\sqrt{20}} = -2.98$.

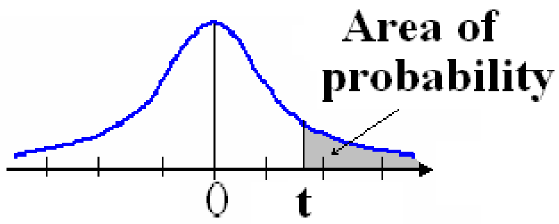
Reject H_0 because $-2.98 < -2.093$.

Note: To find the values of t one should refer to t-distribution with $n-1$ degrees of freedom, where n is a sample size.

t-distribution table

Deg. of freedom	Area in upper tail 0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947

Deg. of freedom	Area in upper tail 0.10	0.05	0.025	0.01	0.005
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.046	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750



The entries in this table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and 0.05 area in the upper tail $t_{0.05;\,10} = 1.182$.

Reading

1) Murray R. Spiegel, *Schaum's outline of Theory and Problems of Probability and Statistics*, McGraw-Hill, 23 edition, 1998.

2) Nitis Mukhopadhyay, Probability and Statistical Inference, Marcel Dekker, Inc. 2000.