Mathematical statistics

April 26^{th}, 2019

Lecture 25: Tests about a population mean

Overview

9.1 Hypotheses and test procedures
- test procedures
- errors in hypothesis testing
- significance level

9.2 Tests about a population mean
- normal population with known σ
- large-sample tests
- a normal population with unknown σ

9.4 P-values

9.3 Tests concerning a population proportion

9.5 Selecting a test procedure

Hypothesis testing

Hypothesis testing

In any hypothesis-testing problem, there are two contradictory hypotheses under consideration

- The null hypothesis, denoted by H_0, is the claim that is initially assumed to be true
- The alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0.

Implicit rules (of this chapter)

- H_0 will always be stated as an equality claim.
- If θ denotes the parameter of interest, the null hypothesis will have the form
$$H_0 : \theta = \theta_0$$
- θ_0 is a specified number called the *null value*
- The alternative hypothesis will be either:
  - $H_a : \theta > \theta_0$
  - $H_a : \theta < \theta_0$
  - $H_a : \theta \neq \theta_0$

Test procedures

A test procedure is specified by the following:

- A test statistic T: a function of the sample data on which the decision (reject H_0 or do not reject H_0) is to be based
- A rejection region $\mathcal{R}$: the set of all test statistic values for which H_0 will be rejected

The null hypothesis will then be rejected if and only if the observed or computed test statistic value falls in the rejection region, i.e., $T \in \mathcal{R}$

Type I and Type II errors

- A type I error consists of rejecting the null hypothesis H_0 when it is true
- A type II error involves not rejecting H_0 when H_0 is false.

Type I error

- Test of hypotheses:

$$H_0 : \mu = 75$$
$$H_a : \mu < 75$$

- $n = 25, \sigma = 9$. Rule: If $\bar{x} \leq 72$, reject H_0.
- Question: What is the probability of type I error?

$$\begin{aligned}\alpha &= P[\text{Type I error}] \\ &= P[H_0 \text{ is rejected while it is true}] \\ &= P[\bar{X} \leq 72 \text{ while } \mu = 75] \\ &= P[\bar{X} \leq 72 \text{ while } \bar{X} \sim \mathcal{N}(75, 1.8^2)] = 0.0475\end{aligned}$$

$\alpha - \beta$ compromise

Proposition

Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then decreasing the size of the rejection region to obtain a smaller value of α results in a larger value of β for any particular parameter value consistent with H_a.

Significance level

The approach adhered to by most statistical practitioners is

- specify the largest value of α that can be tolerated
- find a rejection region having that value of α rather than anything smaller
- α: the *significance level* of the test
- the corresponding test procedure is called a *level* α *test*

Significance level: example

- Test of hypotheses:

$$H_0 : \mu = 75$$
$$H_a : \mu < 75$$

- $n = 25, \sigma = 9$. Rule: If $\bar{x} \leq c$, reject H_0.
- Find the value of c to make this a level α test

$$\begin{aligned}\alpha &= P[\text{Type I error}] \\ &= P[H_0 \text{ is rejected while it is true}] \\ &= P[\bar{X} \leq c \text{ while } \bar{X} \sim \mathcal{N}(75, 1.8^2)] \\ &= P\left[\frac{\bar{X} - 75}{1.8} \leq \frac{c - 75}{1.8}\right]\end{aligned}$$

Hypothesis testing for one parameter

1. Identify the parameter of interest
2. Determine the null value and state the null hypothesis
3. State the appropriate alternative hypothesis
4. Give the formula for the test statistic
5. State the rejection region for the selected significance level α
6. Compute statistic value from data
7. Decide whether H_0 should be rejected and state this conclusion in the problem context

Normal population with known σ

Test about a population mean

- Null hypothesis

$$H_0 : \mu = \mu_0$$

- The alternative hypothesis will be either:
  - $H_a : \mu > \mu_0$
  - $H_a : \mu < \mu_0$
  - $H_a : \mu \neq \mu_0$

Rejection region

$$\begin{aligned}\alpha &= P[\text{Type I error}] \\ &= P[H_0 \text{ is rejected while it is true}] \\ &= P[\bar{X} \le c \text{ while } \bar{X} \sim \mathcal{N}(75, 1.8^2)] \\ &= P\left[\frac{\bar{X} - 75}{1.8} \le \frac{c - 75}{1.8}\right]\end{aligned}$$

- Rejection rule: $\bar{x} \le 75 - 1.8 z_\alpha$
- To make it simpler, define $z = (\bar{x} - 75)/(1.8)$, then the rule is

$$z \le -z_\alpha$$

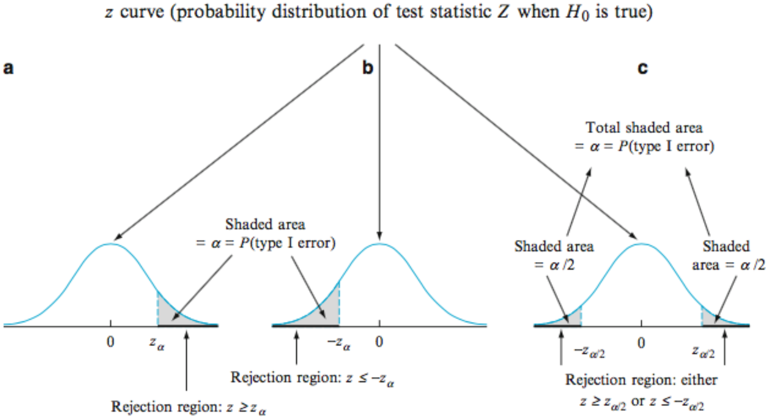
Normal population with known σ

Null hypothesis: $\mu = \mu_0$
Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Alternative Hypothesis	Rejection Region for Level α Test
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
H_a: $\mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

General rule



Example

Problem

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is $130^\circ F$*. A sample of* $n = 9$ *systems, when tested, yields a sample average activation temperature of* $131.08^\circ F$*.*

If the distribution of activation times is normal with standard deviation $1.5^\circ F$*, does the data contradict the manufacturer's claim at significance level* $\alpha = 0.01$*?*

$\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997

Solution

- Parameter of interest: $\mu =$ true average activation temperature
- Hypotheses

$$H_0 : \mu = 130$$
$$H_a : \mu \neq 130$$

- Test statistic:

$$z = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

- Rejection region: either $z \leq -z_{0.005}$ or $z \geq z_{0.005} = 2.58$
- Substituting $\bar{x} = 131.08$, $n = 25 \to z = 2.16$.
- Note that $-2.58 < 2.16 < 2.58$. We fail to reject H_0 at significance level 0.01.
- The data does not give strong support to the claim that the true average differs from the design value.

Large-sample tests

Large-sample tests

Null hypothesis: $\mu = \mu_0$
Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Alternative Hypothesis	**Rejection Region for Level α Test**
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
H_a: $\mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

[Does not need the normal assumption]

Test about a normal population with unknown σ

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic value: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Alternative Hypothesis	**Rejection Region for a Level α Test**
H_a: $\mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (upper-tailed)
H_a: $\mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (lower-tailed)
H_a: $\mu \neq \mu_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$ (two-tailed)

[Require normal assumption]

Example

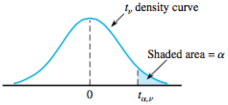
Problem

The amount of shaft wear (.0001 in.) after a fixed mileage was determined for each of $n = 8$ *internal combustion engines having copper lead as a bearing material, resulting in* $\bar{x} = 3.72$ *and* $s = 1.25$.

Assuming that the distribution of shaft wear is normal with mean μ, *use the t-test at level* 0.05 *to test* $H_0 : \mu = 3.5$ *versus* $H_a : \mu > 3.5$.

t-table

Table A.5 Critical Values for *t* Distributions



	α						
ν	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745

Practice

Problem

The standard thickness for silicon wafers used in a certain type of integrated circuit is 245 μm. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18 μm and a sample standard deviation of 3.60 μm.
Does this data suggest that true average wafer thickness is larger than the target value? Carry out a test of significance at level .05.

Type II error and sample size determination

Hypothesis testing for one parameter

1. Identify the parameter of interest
2. Determine the null value and state the null hypothesis
3. State the appropriate alternative hypothesis
4. Give the formula for the test statistic
5. State the rejection region for the selected significance level α
6. Compute statistic value from data
7. Decide whether H_0 should be rejected and state this conclusion in the problem context

Type II error and sample size determination

- A level α test is a test with $P[\text{type I error}] = \alpha$
- Question: given α and n, can we compute β (the probabilities of type II error)?
- This is a very difficult question.
- We have a solution for the cases when: the distribution is normal **and** σ is known

Practice problem

Problem

The drying time of a certain type of paint under specified test conditions is known to be normally distributed with standard deviation 9 min. Assuming that we are testing

$$H_0 : \mu = 75$$
$$H_a : \mu < 75$$

from a dataset with $n = 25$.

- *What is the rejection region of the test with significance level* $\alpha = 0.05$.
- *What is* $\beta(70)$ *in this case?*

General cases

- Test of hypotheses:

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

- Rejection region: $z \leq -z_\alpha$
- This is equivalent to $\bar{x} \leq \mu_0 - z_\alpha \sigma/\sqrt{n}$
- Let $\mu' < \mu_0$

$$\begin{aligned}
\beta(\mu') &= P[\text{Type II error when } \mu = \mu'] \\
&= P[H_0 \text{ is not rejected while it is false because } \mu = \mu'] \\
&= P[\bar{X} > \mu_0 - z_\alpha \sigma/\sqrt{n} \text{ while } \mu = \mu'] \\
&= P\left[\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_\alpha \text{ while } \mu = \mu'\right] \\
&= 1 - \Phi\left(\frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_\alpha\right)
\end{aligned}$$

Remark

- For $\mu' < \mu_0$:

$$\beta(\mu') = 1 - \Phi\left(\frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_\alpha\right)$$

- If n, μ', μ_0, σ is fixed, then

$$\begin{aligned}
&\beta(\mu') \text{ is small} \\
&\leftrightarrow \Phi\left(\frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_\alpha\right) \text{ is large} \\
&\leftrightarrow \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_\alpha \text{ is large} \\
&\leftrightarrow \alpha \text{ is large}
\end{aligned}$$

$\alpha - \beta$ compromise

Proposition

Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then decreasing the size of the rejection region to obtain a smaller value of α *results in a larger value of* β *for any particular parameter value consistent with* H_a*.*

General formulas

Alternative Hypothesis	Type II Error Probability $\beta(\mu')$ for a Level α Test
H_a: $\mu > \mu_0$	$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

where $\Phi(z)$ = the standard normal cdf.
The sample size n for which a level α test also has $\beta(\mu') = \beta$ at the alternative value μ' is

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a one - tailed (upper or lower) test} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a two - tailed test (an approximate solution)} \end{cases}$$