MATH 450: Mathematical statistics

Nov 5th, 2020

Lecture 18: Tests of Hypotheses

Overview

Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	Chapter 7: Point Estimation
Week 7	Chapter 8: Confidence Intervals
Week 10	**Chapter 9: Tests of Hypotheses**
Week 12	Chapter 10: Two-sample testing
Week 14	Regression

Overview

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

9.2 Tests about a population mean

9.4 P-values

9.3 Tests concerning a population proportion

9.5 Selecting a test procedure

Section 9.1: Hypotheses and test procedures

- null hypothesis
- alternative hypothesis
- test statistic
- rejection region
- type I error
- type II error

A statistical hypothesis

is a claim or assertion either about

- the value of a single parameter [Chapter 9]
- the values of several parameters [Chapter 10]
- the form of an entire probability distribution [Chapter 13]

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.
- The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false.
- If the sample does not strongly contradict H_0, we will continue to believe in the probability of the null hypothesis.

Hypothesis testing

The two possible conclusions from a hypothesis-testing analysis are then

- reject H_0 or
- fail to reject H_0

Hypothesis testing: an anology

In a criminal trial, there are to contradictory assertions

- the accused individual is innocent
- the accused individual is guilty

$\rightarrow$ the claim of innocence is the favored or protected hypothesis

Hypothesis testing: example

Setting: Suppose a company is considering putting a new additive in the dried fruit that it produces The true average shelf life with the current additive is known to be 200 days. With μ denoting the true average life for the new additive, the company would not want to make a change unless evidence strongly suggested that μ exceeds 200.

- Null hypothesis:

$$H_0 : \mu = 200$$

- Alternative hypothesis:

$$H_a : \mu > 200$$

Hypothesis testing: example

Before agreeing to purchase a large order of polyethylene sheaths for a particular type of high-pressure oil-filled submarine power cable, a company wants to see conclusive evidence that the true standard deviation of sheath thickness is less than .05 mm.

Hypothesis testing: example

- Before agreeing to purchase a large order of polyethylene sheaths for a particular type of high-pressure oil-filled submarine power cable, a company wants to see conclusive evidence that the true standard deviation of sheath thickness is less than .05 mm.
- Null hypothesis:

$$H_0 : \sigma = 0.05$$

- Alternative hypothesis:

$$H_a : \sigma < 0.05$$

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

$$H_0 : \theta = \theta_0$$

where θ_0 is a specified number called the *null value* of the parameter.

Implicit rules (of this chapter)

The alternative to the null hypothesis $H_0 : \theta = \theta_0$ will look like one of the following three assertions:

- $H_a : \theta > \theta_0$
- $H_a : \theta < \theta_0$
- $H_a : \theta \neq \theta_0$

Hypothesis testing: example

- The drying time of a certain type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation 9 min.
- Chemists have proposed a new additive designed to decrease average drying time.
- It is believed that drying times with this additive will remain normally distributed with $\sigma = 9$.
- Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted.
- Construct the null and alternative hypothesis.

Test procedures

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., $T \in \mathcal{R}$

Hypothesis testing: example

- The drying time of a certain type of paint follow $\mathcal{N}(75, 9^2)$
- Chemists have proposed a new additive designed to decrease average drying time. It is believed that drying times with this additive follows $\mathcal{N}(\mu, 9^2)$.

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

- Experimental data is to consist of drying times from $n = 25$ test specimens: $X_1, X_2, \ldots, X_{25}$.
- My rule:
  - Compute $\bar{x}$
  - If $\bar{x} \leq 70.8$, reject H_0. If not, fail to reject H_0

$\rightarrow$ this is a test procedure

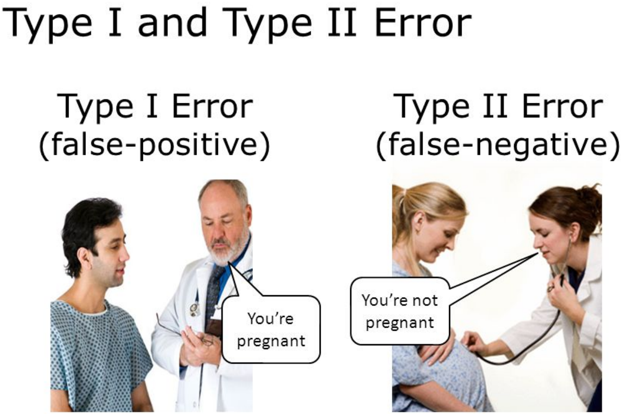
Question

Given a test procedure, how do we quantify how good the test is?

Errors in Hypothesis Testing

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 and Type II Error
Type I Error
(false-positive)
You're
pregnant
Type II Error
(false-negative)
You're not
pregnant

Type I error: example

- It is believed that drying times of an additive follows $\mathcal{N}(\mu, 9^2)$.
- Test of hypotheses:

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

- Experimental data is to consist of drying times from $n = 25$ test specimens: $X_1, X_2, \ldots, X_{25}$.
- My rule:
  - Compute $\bar{x}$
  - If $\bar{x} \leq 70.8$, reject H_0.
- Question: What is the probability of making type I error when using this test procedure?

Type I error: example

- It is believed that drying times of an additive follows $\mathcal{N}(\mu, 9^2)$.
- Test of hypotheses:

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

- $n = 25$. My rule: If $\bar{x} \leq 70.8$, reject H_0.
- Question: What is the probability of type I error?

$$\begin{aligned}
\alpha &= P[\text{Type I error}] \\
&= P[H_0 \text{ is rejected while it is true}] \\
&= P[\bar{X} \leq 70.8 \text{ while } \mu = 75] \\
&= ?
\end{aligned}$$

Type II error: example

- It is believed that drying times of an additive follows $\mathcal{N}(\mu, 9^2)$.
- Test of hypotheses:

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

- $n = 25$. My rule: If $\bar{x} < 70.8$, reject H_0.
- Question: What is the probability of type II error if $\mu = 72$?

$$\begin{aligned}
\beta(72) &= P[\text{Type II error when } \mu = 72] \\
&= P[H_0 \text{ is not rejected while it is false because } \mu = 72] \\
&= P[\bar{X} > 70.8 \text{ while } \mu = 72] \\
&= P[\bar{X} < 70.8 \text{ while } \bar{X} \sim \mathcal{N}(72, 1.8^2)] \\
&= 0.7486
\end{aligned}$$

Type II error: example

- Test of hypotheses:

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

- $n = 25$. My rule: If $\bar{x} < 70.8$, reject H_0.
- Question: What is the probability of type II error if $\mu = 70$?

Type I and Type II errors

- $\beta(72) = 0.7486$
- $\beta(70) = 0.33$
- $\beta(67) = 0.0174$

Practice

Type I error

- Test of hypotheses:

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

- Experimental data is to consist of drying times from $n = 25$ test specimens: $X_1, X_2, \ldots, X_{25}$.
- New rule:
  - Compute $\bar{x}$
  - If $\bar{x} \leq 72$, reject H_0.
- Question: What is the probability of type I error?

$\Phi(z)$

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

Type I error: example

- Test of hypotheses:

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

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

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

Type II error

- Test of hypotheses:

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

- Experimental data is to consist of drying times from $n = 25$ test specimens: $X_1, X_2, \dots, X_{25}$.
- New rule:
  - Compute $\bar{x}$
  - If $\bar{x} \leq 72$, reject H_0.
- Question: What are $\beta(70)$?

$\Phi(z)$

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

Type II error

- Test of hypotheses:

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

- $n = 25$. New rule: If $\bar{x} \leq 72$, reject H_0.

$$\begin{aligned}
\beta(70) &= P[\text{Type II error when } \mu = 70] \\
&= P[H_0 \text{ is not rejected while it is false because } \mu = 70] \\
&= P[\bar{X} > 72 \text{ while } \mu = 70] \\
&= P[\bar{X} < 72 \text{ while } \bar{X} \sim \mathcal{N}(70, 1.8^2)] = 0.1335
\end{aligned}$$

$\alpha - \beta$ 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
- the resulting value of α is often referred to as the *significance level* of the test
- the corresponding test procedure is called a *level* α *test*

Significance level: example

- Test of hypotheses:

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

- $n = 25$. New rule: If $\bar{x} \leq c$, reject H_0.
- Find the value of c to make this a level 0.1 test