Test Procedures

Test Procedures

- **Test procedure** is a rule, based on **sample data**, for deciding whether to reject H_0.
- Test of H_0: $p = 0.10$ versus H_a: $p < 0.10$ in circuit board problem might be based on examining random sample of $n = 200$ boards.

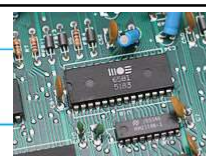
Test Procedures

H_0: $p = 0.10$

H_a: $p < 0.10$

23

Test Procedures



This procedure has two constituents:

(1) ***Test Statistic***, or function of sample data used to make a decision, and

(2) ***Rejection Region*** consisting of those x values for which H_0 will be rejected in favor of H_a.

○ For rule just suggested, the rejection region consists of

$x = 0, 1, 2, \ldots,$ and 15.

○ H_0 will not be rejected if $x = 16, 17, \ldots, 199,$ or 200.

25

Test Procedures

A test procedure is specified by the following:

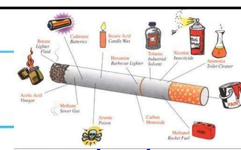
1. Test Statistic, a function of the sample data on which the decision (reject H_0 or do not reject H_0) is to be based

2. Rejection Region, the set of all test statistic values for which H_0 will be rejected

- Null hypothesis will then be rejected if and only if the observed or computed test statistic value falls in the rejection region.

26

Example :



- Suppose cigarette manufacture claims that average nicotine content μ of brand B cigarettes in (at most) 1.5 mg.
- It would be unwise to reject manufacturer's claim without strong contradictory evidence, so an appropriate problem formulation is to test

  $H_0 : \mu = 1.5$ versus $H_a : \mu > 1.5$.
- Consider decision rule based on analyzing a random sample of 32 cigarettes.
- Let $\overline{X}$ denote the sample average nicotine content.
- If H_0 is true, $E(\overline{X}) = \mu = 1.5$ whereas

  If H_0 is false, we expect $\overline{X}$ to exceed 1.5
- Thus we might use $\overline{X}$ as a **test statistic** along with **rejection region** $\overline{x} \geq 1.6$

27

Errors in Hypothesis Testing

Errors in Hypothesis Testing

- The basis for choosing particular rejection region lies in consideration of errors that one might be faced with in drawing conclusion.
- Consider **rejection region $x \leq 15$** in circuit board problem.

$\boxed{E(X) = np = 200(0.10) = 20}$

- Even when H_0: $p = 0.10$ is true, it might happen that unusual sample results in $x = 13$, so that H_0 is erroneously rejected.

$\boxed{E(X) = np = 200(0.10) = 20}$

- On the other hand, even when H_a: $p < 0.10$ is true, unusual sample might yield $x = 20$, in which case H_0 would not be rejected—again an incorrect conclusion.

Errors in Hypothesis Testing

- Thus it is possible that

  H_0 may be rejected when it is true

  or that

  H_0 may not be rejected when it is false.

- These possible errors are not consequences of a foolishly chosen rejection region.

30

Errors in Hypothesis Testing

Definition

- **Type I error** consists of rejecting null hypothesis H_0 when H_0 is true.

- **Type II error** involves not rejecting H_0 when H_0 is false.

32

Errors in Hypothesis Testing

$H_0 : \mu = 1.5$, $H_a : \mu > 1.5$

- In nicotine scenario, a type I error consists of rejecting manufacturer's claim that $\mu = 1.5$ when it is actually true.
- If rejection region $\bar{x} \geq 1.6$ is employed, it might happen that $\bar{x} = 1.6$ even when $\mu = 1.5$, resulting in type I error.
- Alternatively, it may be that H_0 is false and yet $\bar{x} = 1.52$ is observed, leading to H_0 not being rejected (a type II error)

Type I error consists of rejecting null hypothesis H_0 when H_0 is true.
Type II error involves not rejecting H_0 when H_0 is false.

33

Errors in Hypothesis Testing

- In the best of all possible worlds, test procedures for which neither type of error is possible could be developed.
- However, this ideal can be achieved only by basing a decision on an examination of the entire population.
- The difficulty with using a procedure based on sample data is that because of sampling variability, an unrepresentative sample may result.
- Even though $E(\bar{X}) = \mu$, the observed value $\bar{x}$ may differ substantially from μ

34

Errors in Hypothesis Testing

$H_0 : \mu = 1.5$, $H_a : \mu > 1.5$

- Thus when $\mu = 1.5$ in the nicotine situation, $\bar{x}$ may be larger than 1.5, resulting in erroneous rejection of H_o

$H_0 : \mu = 1.6$, $H_a : \mu > 1.6$

- Alternatively, it may be that $\mu = 1.6$ yet $\bar{x}$ much smaller than this is observed, leading to type II error.

Type I error consists of rejecting null hypothesis H_0 when H_0 is true.
Type II error involves not rejecting H_0 when H_0 is false.

35

Errors in Hypothesis Testing

- Instead of demanding error-free procedures, we must seek procedures for which either type of error is unlikely to occur.
- That is, good procedure is one for which probability of making either type of error is small.
- The choice of particular rejection region cutoff value fixes probabilities of type I and type II errors.

36

Errors in Hypothesis Testing

- These error probabilities are traditionally denoted by α and β, respectively.
- Because H_0 specifies a unique value of parameter, there is a single value of α.
- However, there is a different value of β for each value of parameter consistent with H_a.

37

Example

- A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph (16.09 km/h) crash tests.
- A modified bumper design has been proposed in an effort to increase this percentage.

- Let p denote the proportion of all 10-mph crashes with this new bumper that result in no visible damage.
- The hypotheses to be tested are

  H_0: $p = 0.25$ (no improvement) versus

  H_a: $p > 0.25$.
- The test will be based on an experiment involving $n = 20$ independent crashes with prototypes of the new design.

38

Example

cont'd

- Intuitively, H_0 should be rejected if substantial number of crashes show no damage.

- Consider the following test procedure:

Test statistic: X = the number of crashes with no visible damage

Rejection region: $R_8 = \{8, 9, 10, \ldots, 19, 20\}$; that is, reject H_0 if $x \geq 8$, where x is the observed value of test statistic.

39

Example

cont'd

- This rejection region is called *upper-tailed* because it consists only of large values of test statistic.

- When H_0 is true, X has binomial probability distribution with $n = 20$ and $p = 0.25$. Then

$$\alpha = P(\text{type I error}) = P(H_0 \text{ is rejected when } H_0 \text{ is true})$$

$$= P(X \geq 8 \text{ when } X \sim \text{Bin}(20, 0.25)) = 1 - B(7; 20, 0.25)$$

$$= 1 - 0.898$$

$$= 0.102$$

10.2%

n = 20

	p														
	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	.818	.358	.122	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.983	.736	.392	.069	.024	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
2	.999	.925	.677	.206	.091	.035	.004	.000	.000	.000	.000	.000	.000	.000	.000
3	1.000	.984	.867	.411	.225	.107	.016	.001	.000	.000	.000	.000	.000	.000	.000
4	1.000	.997	.957	.630	.415	.238	.051	.006	.000	.000	.000	.000	.000	.000	.000
5	1.000	1.000	.989	.804	.617	.416	.126	.021	.002	.000	.000	.000	.000	.000	.000
6	1.000	1.000	.998	.913	.786	.608	.250	.058	.006	.000	.000	.000	.000	.000	.000
7	1.000	1.000	1.000	.968	.898	.772	.416	.132	.021	.001	.000	.000	.000	.000	.000
8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

Example

cont'd

- That is, when H_0 is actually true,
  roughly 10% of all experiments consisting of 20 crashes
  would result in ***H_0* being incorrectly rejected**.

Type I error

41

Example

cont'd

- In contrast to α, there is not single β.
- Instead, there is different β for each different p that exceeds 0.25.

- Thus there is value of β for $p = 0.3$ (in which case $X \sim \text{Bin}(20, 0.3)$),
  another value of β for $p = 0.5$, and so on.

42

Example

cont'd

For example,

$\beta(0.3) = P(\text{type II error when } p = 0.3)$
$= P(H_0 \text{ is not rejected when } H_0 \text{ is false because } p = 0.3)$
$= P(X \leq 7 \text{ when } X \sim \text{Bin}(20, 0.3)) = B(7; 20, 0.3) = 0.772$

n = 20

	p														
	0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	.818	.358	.122	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.983	.736	.392	.069	.024	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
2	.999	.925	.677	.206	.091	.035	.004	.000	.000	.000	.000	.000	.000	.000	.000
3	1.000	.984	.867	.411	.225	.107	.016	.001	.000	.000	.000	.000	.000	.000	.000
4	1.000	.997	.957	.630	.415	.238	.051	.006	.000	.000	.000	.000	.000	.000	.000
5	1.000	1.000	.989	.804	.617	.416	.126	.021	.002	.000	.000	.000	.000	.000	.000
6	1.000	1.000	.998	.913	.786	.608	.250	.058	.006	.000	.000	.000	.000	.000	.000
7	1.000	1.000	1.000	.968	.898	.772	.416	.132	.021	.001	.000	.000	.000	.000	.000
8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

- When p is actually 0.3 rather than 0.25 (a "small" departure from H_0), roughly 77% of all experiments of this type would result in H_0 being incorrectly not rejected!

Example

cont'd

- The accompanying table displays β for selected values of p (each calculated for rejection region R_8).

p	0.3	0.4	0.5	0.6	0.7	0.8
$\beta(p)$	0.772	0.416	0.132	0.021	0.001	0.000

- Clearly, β decreases as value of p moves farther to the right of the null value 0.25.

- Intuitively, the greater the departure from H_0, the less likely it is that such departure will not be detected.

Example … cont'd

- Let us use the same experiment and test statistic X as previously described in automobile bumper problem but now consider

rejection region $R_9 = \{9, 10, ..., 20\}$

- Since X still has a binomial distribution with parameter *n=20 and p.*

47

Example … cont'd

Type I error

$\alpha = P(H_0$ is rejected when $p = 0.25)$

$= P(X \geq 9$ when $X \sim \text{Bin}(20, 0.25)) = 1 - B(8; 20, 0.25) = 0.041$

$n = 20$

		0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
x	8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

- Type I error probability has been decreased by using new rejection region.

48

Example … cont'd

- However, a price has been paid for this decrease :

$\beta(0.3) = P(\text{H}_0$ is not rejected when $p = 0.3)$

$= P(X \leq 8$ when $X \sim \text{Bin}(20, 0.3)) = B(8; 20, 0.3) = 0.887$

$n = 20$

x	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000

$\beta(0.5) = P(\text{H}_0$ is not rejected when $p = 0.5)$

$= P(X \leq 8$ when $X \sim \text{Bin}(20, 0.5)) = B(8; 20, 0.5) = 0.252$

Example … cont'd

	p	0.3	0.5
$\alpha = 0.102$	$R8 \rightarrow \beta(p)$	0.772	0.132
$\alpha = 0.041$	$R9 \rightarrow \beta(p)$	0.887	0.252

- Both these βs are larger than corresponding error probabilities 0.772 and 0.132 for region R_8
- This is not surprising; α is computed by summing over probabilities of test statistic values in rejection region, where as β is probability that X falls in complement of rejection region.
- Making rejection region smaller must therefore decrease α while increasing β for any fixed alternative value of parameter.

Errors in Hypothesis Testing

Proposition

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

51

Errors in Hypothesis Testing

- This proposition says that once test statistic and n are fixed, there is no rejection region that will simultaneously make both α and all β's small.
- Region must be chosen to effect a compromise between α and β.
- Because of suggested guidelines for specifying H_0 and H_a, a type I error is usually more serious than a type II error (this can always be achieved by proper choice of hypotheses).

52

Errors in Hypothesis Testing

- Approach adhered to by most statistical practitioners is then to specify the largest value of that can be tolerated and find rejection region having that value of α rather than anything smaller.

This makes β as small as possible subject to the bound on α.

Resulting value of α is often referred to as **significance level** of test.

- Traditional levels of significance are 0.10, 0.05, and 0.01, though level in any particular problem will depend on the seriousness of type I error.

The more serious this error, the smaller should be significance level.

Errors in Hypothesis Testing

- The corresponding test procedure is called a **level α test** e.g.,
  - level 0.05 test or
  - level 0.01 test.

- A test with significance level α is one for which the type I error probability is controlled at specified level.

54

Example

- Again let μ denote true average nicotine content of brand B cigarettes.
- The objective is to test

$$H_o: \mu = 1.5 \text{ versus } H_a: \mu > 1.5$$

based on a random sample $X_1, X_2, \ldots, X_{32}$ of nicotine content.

- Suppose distribution of nicotine content is known to be normal with $\sigma = 0.20$.
- Then $\overline{X}$ is normally distributed with
  mean value $\mu_{\overline{X}} = \mu$ and
  standard deviation $\sigma_{\overline{X}} = \frac{0.20}{\sqrt{32}} = 0.0354$

55

Distribution of Sample Mean (from Chapter 5)

Proposition

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$
2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

56

Example

$Z = \frac{X - \mu}{\sigma}$ ⇨ $Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$ ⇨ $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$

cont'd

- Rather than use $\overline{X}$ itself as the test statistic, let's standardize $\overline{X}$, assuming that H_0 is true.

Test statistic : $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ⇨ $Z = \frac{\overline{X} - 1.5}{\frac{0.20}{\sqrt{32}}}$ $= \frac{\overline{X} - 1.5}{0.0354}$

- Z expresses the distance between $\overline{X}$ and its expected value (μ) when H_0 is true as some number of standard deviations.

- For example, $z = 3$ results from $\overline{x}$ that is 3 standard deviations larger than we would have expected it to be were H_0 true.

57

Example

cont'd

That is, the form of the rejection region is $z \geq c$.

Let's now determine c so that $\alpha = 0.05$.

When H_0 is true, Z has a standard normal distribution. Thus

$\alpha = P(\text{type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$
$= P(Z \geq c \text{ when } Z \sim N(0, 1))$

- Value c must capture upper-tail area 0.05 under the z curve.
- So, directly from Appendix Table A.3,

$\Phi(z) = P(Z \leq z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$\boxed{C = z_{0.05} = 1.645.}$

59

Example

$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$ $\Rightarrow$ $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ cont'd

- Notice that

$Z = \frac{\overline{X} - 1.5}{0.0354}$

$z \geq 1.645$ is equivalent to $\overline{x} - 1.5 \geq (0.0354)(1.645)$

$\Downarrow$

$\overline{x} \geq 1.56$

- Then β involves the probability that $\overline{X} < 1.56$ and can be calculated for any μ greater than 1.5.

End of Section 8.1