Probability & Statistical Modelling

AQ077-3-2-PSMOD and Version VD1

Hypothesis Testing

Topic & Structure of The Lesson

- Introduction to hypothesis testing
- Hypothesis testing on population mean
- Hypothesis testing on population proportion
- Test of Goodness-of-fit
- Test of Independency

Learning Outcomes

- At the end of this section, You should be able to:
  - Explain the principles underlying hypothesis testing
  - Structure a business decision situation about means or proportions into the form of a test of a hypothesis
  - Apply systematic testing procedures.
  - Interpret hypothesis test results and draw conclusions

Key Terms You Must Be Able To Use

If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:
(Prepare your own list)

- **Hypothesis statement**
- **Significance level**
- **Test statistic**
- **Decision rule**
- **Critical value**

Introduction

- **Alternatively called significance testing.**
- **Is testing a belief or opinion by statistical methods.**
  - **In decision making, we make an assumption, called hypothesis, then we collect some sample data, produce sample statistics and use this information to decide how likely it is that our hypothesized population parameter is true.**
- **Commonly used for testing sample means & proportion**

- **In hypothesis testing, we must stated the assumed or hypothesized value of the population before we begin sampling.This assumption is called the** null hypothesis.
- **The Null hypothesis (H_o) usually assumes there is no difference between the observed and believed values.**
- **If our sample results fail to support the null hypothesis, then the conclusion that we do accept is called the** alternative hypothesis, H_1.

- There are only 4 possible results when we test a given hypothesis.
  - We accept a true hypothesis – a correct decision
  - We reject a false hypothesis – a correct decision.
  - We reject a true hypothesis – an incorrect decision. (Type I error)
  - We accept a false hypothesis – an incorrect decision. (Type II error)

Terminologies

- Significance level
  - Complementary concepts to confidence limits.
  - Probability of committing a TYPE 1 error, naming rejecting the null hypothesis when in reality it is true.
  - There is no single standard or universal level of significance for testing hypothesis.
  - The higher the significance level, the higher the probability of rejecting a null hypothesis when it is true.

- Type 1 error, α
  - Is the error of rejecting a null hypothesis when it is true.
- Type 11error, β
  - Is the error of accepting a null hypothesis when it is actually false.
- In order for any tests of hypothesis or rules of decision to be good, they must be designed so as to minimise errors of decision. The only way to reduce both types of errors is to increase the sample size.

- **One-tailed Test**
  - **Is a significance test in which the null hypothesis can be upset by values well above or below the mean but not both.**
- **Two-tailed test**
  - **Is a significance test in which it will reject the null hypothesis if the sample mean is significantly higher or lower then hypothesized population mean.(i.e. there are two rejection region)**

Procedures for Hypothesis Testing

1) State the null hypothesis, H_0, and the alternative hypothesis H_1.
2) Consider the appropriate distribution given by the null hypothesis.
3) Decide on the level of the test.
4) Decide on the rejection criteria.
5) Calculate the value of the test statistics.

When σ not given and n < 30

for mean: $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ or $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$ for proportion: $Z = \dfrac{p-\pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$

6) Make conclusion: If the value of test statistic lies in the critical region, reject H_0

If the value of test statistic does not lie in the critical region, do not reject H_0.

Steps in Hypothesis Testing

- Specify the population value of interest
- Formulate the appropriate null and alternative hypotheses
- Specify the desired level of significance
- Determine the rejection region
- Obtain sample evidence and compute the test statistic
- Reach a decision and interpret the result

Required to know:

- Hypothesis testing for one sample
  - Mean
    - One tailed test
    - Note that when sample size is 30 or less and the population standard deviation is not known, t-distribution is used.
  - Proportion
    - One tailed test

Example 1

- The manufacturer of 'Chummy Morsels' claims that at least 8 out of 10 dogs choose his product rather than that produced by a rival firm. In a random sample of 200 dogs, 152 chose 'Chummy Morsels' and the rest chose the rival brand. Comment on the manufacturer's claim. (Test at 5% level)

- *Conclusion*: We do not reject H_0 and conclude that there is not sufficient evidence, at the 5% level, to refute the manufacturer's claim.

Example 2

- A large college claims that it admits equal numbers of men and women. A random sample of 500 students at the college gave 267 males. Is there any evidence, at the 5% level, that the college population is not evenly divided into males and females?

- *Conclusion*: We do not reject H_0 and conclude that, at the 5% level, there is not sufficient evidence to refute the claim that the population is evenly divided into males and females.

Example 3

Experience has shown that the scores obtained in a particular test are normally distributed with mean score 70 and variance 36. When the test is taken by a random sample of 36 students, the mean score is 68.5. Is there sufficient evidence, at the 3% level, that these students have not performed as well as expected?

Conclusion: We do not reject H_0 and conclude that there is not sufficient evidence, at the 3% level, that the students have not performed as well as expected.

Example 4

- A machine produces elastic bands with breaking tension normally distributed with mean 45N and s.d. 4.36N. On a certain day a sample of 50 was tested and found to have a mean breaking tension of 43.46N. Test at the 5% level of significance whether this indicates a change in the mean.

Conclusion: We reject H_0 and conclude that there is evidence, at the 5% level, of a change in the mean.

Example 5

A normal distribution is thought to have a mean of 50. A random sample of 100 gave a mean of 52.6 and a standard deviation of 14.5. Is there evidence that the population mean has increased at the 5% level?

Conclusion: We reject H_0 and conclude that there is evidence, at the 5% level, that the population mean has increased.

Example 6

- Five readings of the resistance, in ohms, of a piece of wire gave the following results:

  1.51 1.49 1.54 1.52 1.54

  and standard deviation = 0.019

  If the wire were pure silver, its resistance would be 1.50 ohms. If the wire were impure, the resistance would be increased. Test, at the 5% level, the hypothesis that the wire is pure silver.

Conclusion: We reject H_0 and conclude that at the 5% level, the wire is impure.

Non-parametric Tests

- **Introduction**
  - The significance tests covered so far depend, to greater or less extent, on the assumption, or presence of the normal distribution
  - They are also concerned with the parameters of the distribution e.g. mean, proportion. Hence given the mean of parametric tests.

- However, on occasions, the data are not normal, or contain extreme values or not enough is known to be able to make any assumption about the type of distribution. Then non-parametric or distribution free tests may be used.

Advantages of non-parametric tests

- No assumptions need to be made about the underlying distribution
- They can be used on data ranked in some order.
- Mathematic concepts are simpler than for parametric tests

- **Disadvantages of non-parametric tests**
  - **They are less discriminating than parametric tests. I.e. they are more prone to error and less powerful**
  - **Although simple, the arithmetic may take a long time**

Chi-square (χ^2) Distribution

- used when it is wished to compare an actual, observed distribution with a hypothesized, or expected distribution.
- Often referred to as a 'goodness of fit' test
- $\chi^2 = \sum \frac{(O-E)^2}{E}$
  - where O = the observed frequency of any value
    E = the expected frequency of any value

- The obtained value from the formula is compared with the value from χ^2 table for a given significance level and the number of degrees of freedom.
- Degrees of freedom = (Rows-1)(Columns –1)
- If χ^2 calculated is > χ^2 from table, the null hypothesis is rejected.

- Use broadly for
  - Test of goodness of fit (for one way classification or for one variable only)
    - Can also be used to determine how well empirical distributions I.e. those obtained from sample data fit theoretical distributions such as the Normal, Poisson and Binomial
  - Test of independence (for more than one row or column in the form of a contingency table covering several attributes.)

- **Note that :**
  - **When calculating, the expected cell values, the expected frequency is less than 5, the χ^2 test becomes inaccurate. In such circumstances the cell which is less than 5 is merged with an adjoining cell so that the expected frequencies in all resulting cells are at least 5.**

Chi-square Goodness-of-fit Test

- Does sample data conform to a hypothesized distribution?
  - **Examples:**
    - Are technical support calls equal across all days of the week? (i.e., do calls follow a uniform distribution?)
    - Do measurements from a production process follow a normal distribution?

Example 7

- **Example:**
  - **Are technical support calls equal across all days of the week? (i.e., do calls follow a uniform distribution?)**
    - **Sample data for 10 days per day of week:**

	Sum of calls for this day:
Monday	290
Tuesday	250
Wednesday	238
Thursday	257
Friday	265
Saturday	230
Sunday	192
	$\Sigma = 1722$

Logic of Goodness-of-Fit Test

- **If calls are uniformly distributed, the 1722 calls would be expected to be equally divided across the 7 days:**

$$\frac{1722}{7} = 246 \quad \text{expected calls per day if uniform}$$

- **Chi-Square Goodness-of-Fit Test: test to see if the sample results are consistent with the expected results**

Observed & Expected Frequencies

	Observed o_i	Expected e_i
Monday	290	246
Tuesday	250	246
Wednesday	238	246
Thursday	257	246
Friday	265	246
Saturday	230	246
Sunday	192	246
TOTAL	1722	1722

Chi-Square Test Statistic

H_0: The distribution of calls is uniform over days of the week

H_A: The distribution of calls is not uniform

- The test statistic is

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \qquad (\text{where } \ df = k - 1)$$

where:

k = number of categories
o_i = observed cell frequency for category i
e_i = expected cell frequency for category i

The Rejection Region

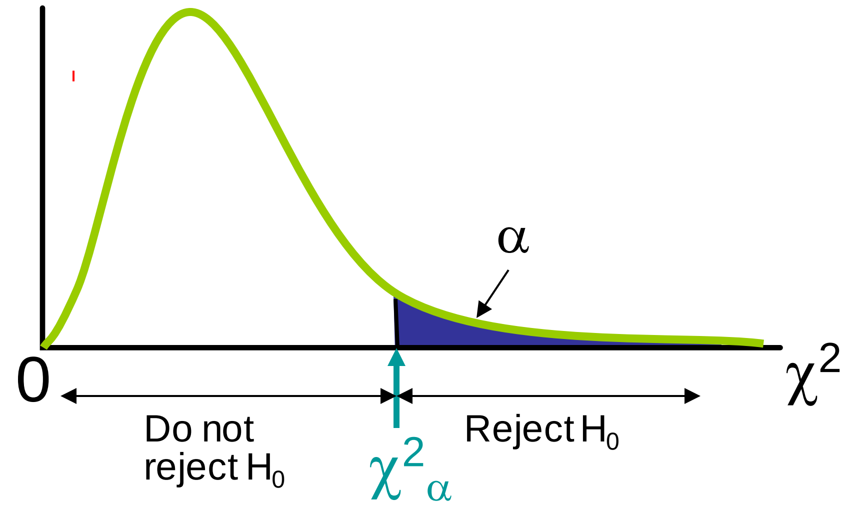
H_0: The distribution of calls is uniform over days of the week

H_A: The distribution of calls is not uniform

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

- Reject H_0 if $\chi^2 > \chi^2_\alpha$

(with k – 1 degrees of freedom)



Chi-Square Test Statistic

H_0: The distribution of calls is uniform over days of the week

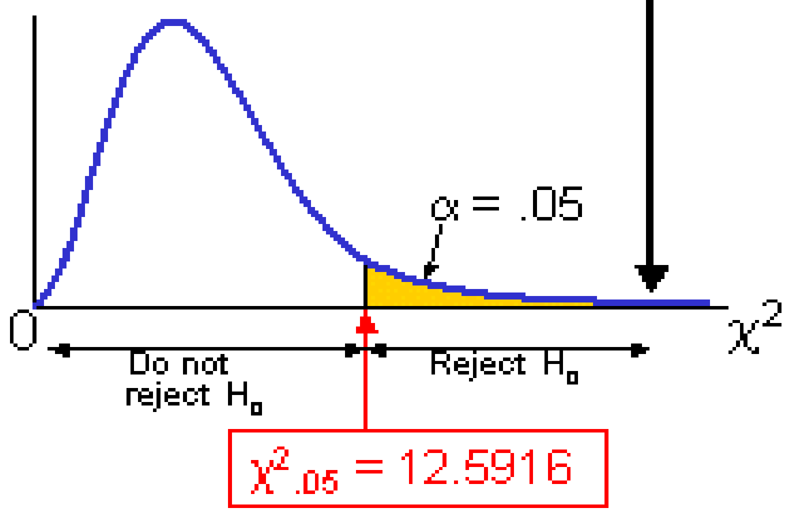
H_A: The distribution of calls is not uniform

$$\chi^2 = \frac{(290-246)^2}{246} + \frac{(250-246)^2}{246} + \ldots + \frac{(192-246)^2}{246} = \boxed{23.05}$$

k – 1 = 6 (7 days of the week) so use 6 degrees of freedom:

$\chi^2_{.05} = 12.5916$

Conclusion:
$\chi^2 = 23.05 > \chi^2_{\alpha} = 12.5916$ so **reject H_0** and conclude that the distribution is not uniform



Contingency Tables

- Situations involving multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called a cross-tabulation table.

Example 8:

- The following data concerning industrial accidents and absentees classified according to the types of employee.

Absence following accident	Type of employee		
	Men	Women	Juveniles
Up to one month	26	16	8
One month or longer	14	9	7

- Is there any evidence to suggest that the severity of accident is associated with type of employee ?

Logic of the test

H_0: Severity of accident is independent of type of employees

H_A: Severity of accident is not independent of type of employees

- If H_0 is true, then the proportion of severity of accidents should be the same as the proportion of type of employees

➢ Finding Expected Frequencies

40 men, 26 absence up to I month

25 women, 16 absence up to I month

15 Juveniles, 8 absence up to I month

Overall:

P(absence up to I month)

= 50/80 = .625

If independent, then

P(absence up to I month | men) = P(absence up to I month | women) = P(absence up to I month | Juveniles) = .625

So we would expect 62.5% of the 40 men, 62.5% of the 25 women and 62.5% of the 15 juveniles to be absent up to I month...

i.e., we would expect
- **(40)(.625) = 25 men to be absent up to I month**
- **(25)(.625) = 15.625 women to be absent up to I month**
- **(15)(.625) = 9.375 juveniles to be absent up to I month**

➢ Finding Expected Frequencies

Expected frequency

$$= \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

➤Observed vs. expected Frequencies

Type of employees	Absence		
	Up to 1 month	One month or longer	
Men	Observed = 26 Expected 25	Observed = 14 Expected = 15	40
Women	Observed = 16 Expected = 15.625	Observed = 9 Expected = 9.375	25
Juveniles	Observed = 8 Expected = 9.375	Observed = 7 Expected = 5.625	15
	50	30	80

- The Chi-square contingency test statistic is:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

with $\text{d.f.} = (r - 1)(c - 1)$

- **where:**

  ***O* = observed frequency**

  ***E* = expected frequency**

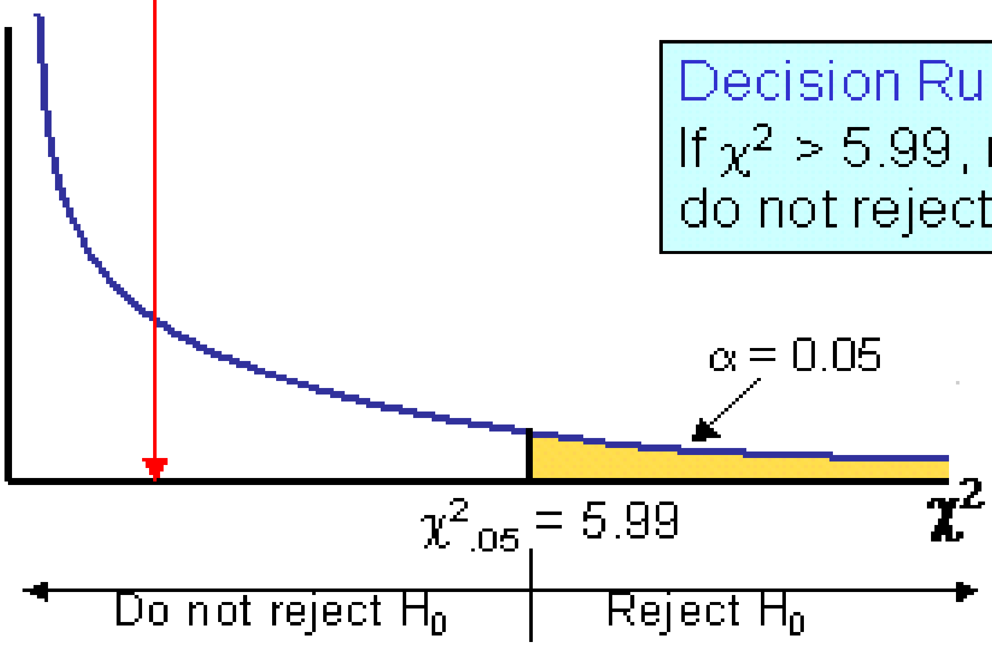
  r = number of rows

  c = number of columns

O	E	O-E	$\frac{(O-E)^2}{E}$
26	25	1	0.0400
14	15	-1	0.0667
16	15.625	0.375	0.0090
9	9.375	-0.375	0.0150
8	9.375	-1.375	0.2017
7	5.625	1.375	0.3361
			0.6685

Contingency Analysis

$$\chi^2 \boxed{= 0.6685} \quad \text{with} \quad d.f. = (r-1)(c-1) = (2)(1) = 2$$



Decision Rule:
If $\chi^2 > 5.99$, reject H_0, otherwise, do not reject H_0

Here, $\chi^2 = 0.6685 < 5.99$, so we do not reject H_0 and conclude that severity of accident and type of employee are independent

- **Yate's Correction**
  - **Used when there is only ONE degree of freedom.**

$$\chi^2 = \sum \frac{\left(|0 - E| - 0.5\right)^2}{E}$$

- **Example:**
  - Left-Handed vs. Gender
    - Dominant Hand: Left vs. Right
    - Gender: Male vs. Female

H_0: Hand preference is independent of gender

H_A: Hand preference is not independent of gender

Quick Review Question

- **Example:**

Sample results organized in a contingency table:

sample size = n = 300:

120 Females, 12 were left handed

180 Males, 24 were left handed

Gender	Hand Preference		
	Left	Right	
Female	12	108	120
Male	24	156	180
	36	264	300

Logic of the test

H_0: Hand preference is independent of gender

H_A: Hand preference is not independent of gender

- If H_0 is true, then the proportion of left-handed females should be the same as the proportion of left-handed males
- The two proportions above should be the same as the proportion of left-handed people overall

➢Finding Expected Frequencies

120 Females, 12 were left handed

180 Males, 24 were left handed

Overall:

P(Left Handed) = 36/300 = .12

If independent, then

P(Left Handed | Female) = P(Left Handed | Male) = .12

So we would expect 12% of the 120 females and 12% of the 180 males to be left handed…

i.e., we would expect (120)(.12) = 14.4 females to be left handed
(180)(.12) = 21.6 males to be left handed

➢Observed vs. expected Frequencies

Gender	Hand Preference		
	Left	Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
	36	264	300

- The Chi-square contingency test statistic is:

$$\chi^2 = \sum \frac{\left(\left|O - E\right| - 0.5\right)^2}{\text{E}}$$

- **where:**

  O **= observed frequency**

  E **= expected frequency**

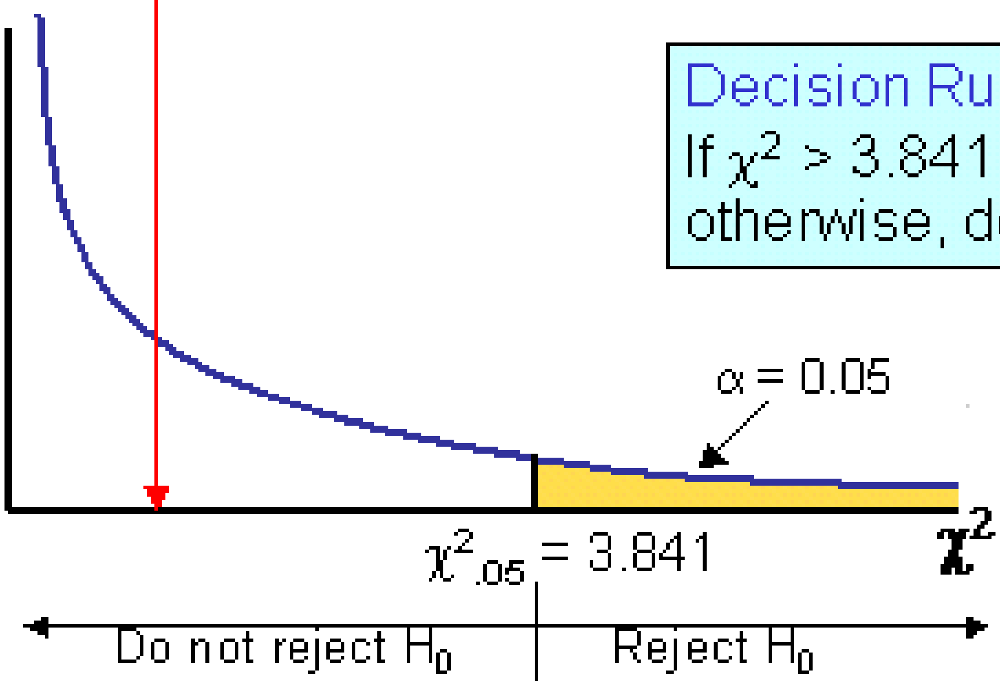
  r = number of rows

  c = number of columns

O	E	O-E	$(\lvert O-E\rvert-0.5)^2$	$\dfrac{(\lvert O-E\rvert-0.5)^2}{E}$
12	14.4	-2.4	3.61	0.2507
108	105.6	2.4	3.61	0.0342
24	21.6	-2.4	3.61	0.1671
156	158.4	2.4	3.61	0.0228
				0.4748

➢ Contingency Analysis

$$\chi^2 = 0.4748 \quad \text{with} \quad \text{d.f.} = (r-1)(c-1) = (1)(1) = 1$$



Decision Rule:

If $\chi^2 > 3.841$, reject H_0, otherwise, do not reject H_0

Here, $\chi^2 = 0.6848 < 3.841$, so we do not reject H_0 and conclude that gender and hand preference are independent

Summary of Main Teaching Points

- **Hypothesis testing on population mean and proportion**

Question and Answer Session

Q & A

What we will cover next

- Decision Making Techniques