DegreeLive

B.Com Honours | Semester I | Calicut University

Essential Statistics for Business Analytics

Course Code: COM1MN109 • Module 2 Notes

1. Hypothesis Testing: Concepts and Large Sample Tests

In business analytics, managers must evaluate claims based on sample evidence—for instance, verifying if a new website layout increases conversions or if a manufacturing machine is outputting defective items. Hypothesis testing is a structured statistical decision-making procedure that evaluates sample evidence to determine whether to reject or fail to reject a claim. This module covers null and alternative hypotheses, Type I and Type II errors, large sample Z-tests, small sample t-tests, ANOVA, and Chi-Square tests.

Core Concepts of Hypothesis Testing

- **Null Hypothesis (H_0):** A statement of no effect, no change, or status quo. Assumed true until proven otherwise. E.g., H_0: $\mu = 50$.
- **Alternative Hypothesis (H_a or H_1):** The claim that contradicts H_0, indicating an effect or difference. E.g., H_a: $\mu \neq 50$ (two-tailed) or $\mu > 50$ (one-tailed).
- **Level of Significance (α):** The probability of rejecting H_0 when it is actually true (Type I error). Typically set at 5% (0.05) or 1% (0.01).

Type I and Type II Errors

Statistical Decision	H_0 is True	H_0 is False
Fail to Reject H_0	Correct Decision (Probability = 1 - α)	**Type II Error (β)** (Consumer's Risk)
Reject H_0	**Type I Error (α)** (Producer's Risk)	Correct Decision / Power of Test (1 - β)

Large Sample Tests (Z-Tests, n ≥ 30)

Z-tests are used when the sample size is large, assuming the normal distribution applies:

- **Z-Test for Single Mean:** Compares sample mean with population mean.
  [Z = \frac{\bar{x} - mu}{sigma / sqrt{n}}]
- **Z-Test for Single Proportion:** Checks if a sample proportion matches population proportion.
  [Z = \frac{p - P}{sqrt{\frac{P(1-P)}{n}}}]
- **Z-Test for Difference of Proportions:** Compares proportions of two independent samples.
  [Z = \frac{p_1 - p_2}{sqrt{hat{p}hat{q}left(\frac{1}{n_1} + \frac{1}{n_2} ight)}}]
  Where (hat{p} = \frac{x_1 + x_2}{n_1 + n_2}) is the pooled proportion, and (hat{q} = 1 - hat{p}).

2. Small Sample Tests and Categorical Analysis

When the sample size is small (n < 30) and the population standard deviation σ is unknown, standard Z-tests cannot be used. We apply specialized small-sample models.

Student's t-Test

Developed by William Sealy Gosset (under the pen name "Student"), the t-distribution is wider than the normal distribution, accounting for small sample uncertainty. The test statistic is:

[t = \frac{\bar{x} - mu}{s / sqrt{n}}]

Where *s* is the sample standard deviation, evaluated against **n - 1 degrees of freedom**. Types of t-tests include:

- **One-Sample t-Test:** Compares a single sample mean with a claimed population mean.
- **Independent Two-Sample t-Test:** Compares means of two separate groups (e.g., sales in region A vs. region B).
- **Paired t-Test:** Compares means of matched pairs (e.g., student test scores before and after a training module).

Chi-Square Test (χ^2)

A non-parametric test used to analyze categorical data:

- **Goodness of Fit:** Checks if observed categorical frequencies match expected theoretical frequencies.
  [chi^2 = sum \frac{(O - E)^2}{E}]
  Where **O** = Observed frequency, **E** = Expected frequency.

- **Test of Independence of Attributes:** Checks if two categorical variables (e.g., gender and brand preference) are independent.

Analysis of Variance (ANOVA)

Used to test the equality of means across three or more groups simultaneously by analyzing group variances.

- **One-Way ANOVA:** Evaluates one factor (e.g., impact of three different packaging designs on sales).
- **Two-Way Factorial ANOVA:** Evaluates two factors and their interaction (e.g., packaging design and price levels).
- **MANOVA (Multivariate ANOVA):** Analyzes multiple dependent variables simultaneously.