# Chi-Square & ANOVA Study Notes

Statistical Hypothesis Testing Guide

## Key Definitions

**Hypothesis Testing:** A statistical method to make decisions about population parameters based on sample data.

**Null Hypothesis ($H_0$):** A statement of no effect or no difference that we test against.

**Alternative Hypothesis ($H_1$):** A statement that contradicts the null hypothesis.

**Significance Level (α):** The probability of rejecting a true null hypothesis (commonly 0.05).

**Degrees of Freedom (df):** The number of independent values that can vary in a calculation.

**Categorical Data:** Data that can be divided into groups or categories (like gender, color, yes/no).

**Continuous Data:** Data that can take any numerical value within a range (like height, weight, test scores).

## Chi-Square Test

## What is it?

A statistical test used to determine if there's a significant relationship between two categorical variables.

**When to use:** When you have categorical data and want to test independence or goodness of fit.

## Formula

$$\chi^2 = \Sigma \left[(\text{Observed} - \text{Expected})^2 / \text{Expected}\right]$$

## Steps

1. **Set hypotheses:** $H_0$: Variables are independent, $H_1$: Variables are related
2. **Create contingency table:** Organize observed frequencies
3. **Calculate expected frequencies:** (Row Total × Column Total) / Grand Total
4. **Calculate $\chi^2$:** Use the formula above
5. **Find df:** (rows - 1) × (columns - 1)
6. **Compare with critical value:** If $\chi^2$ > critical value, reject $H_0$

### Example: Gender vs Product Preference

| | Product A | Product B | Total |
|---|---|---|---|
| **Male** | 20 | 30 | 50 |
| **Female** | 25 | 25 | 50 |
| **Total** | 45 | 55 | 100 |

**Result:** $\chi^2 = 1.01$, df = 1, Critical value = 3.84
**Conclusion:** Since 1.01 < 3.84, gender and product preference are independent.

**Important:** All expected frequencies should be ≥ 5 for reliable results.

## ANOVA (Analysis of Variance)

### What is it?

A statistical technique used to compare the means of three or more groups to determine if they are significantly different.

**When to use:** When you have continuous data and want to compare means of 3+ groups.

### Formula

$$F = MSB / MSW$$

**(Mean Square Between / Mean Square Within)**

### Types

- **One-Way ANOVA:** One independent variable
- **Two-Way ANOVA:** Two independent variables

### Steps (One-Way ANOVA)

1. **Set hypotheses:** $H_0$: All group means are equal, $H_1$: At least one mean differs
2. **Check assumptions:** Independence, normality, equal variances
3. **Calculate group means** and overall mean
4. **Calculate Sum of Squares:** Between groups (SSB) and Within groups (SSW)

**5. Calculate Mean Squares:** MSB = SSB/df_between, MSW = SSW/df_within

**6. Calculate F-statistic:** F = MSB/MSW

**7. Compare with critical value:** If F > critical value, reject $H_0$

## ANOVA Table Structure

| Source | SS | df | MS | F |
|---|---|---|---|---|
| **Between Groups** | $SS_B$ | k - 1 | $SS_B$ / (k-1) | $MS_B$ / $MS_W$ |
| **Within Groups (Error)** | $SS_W$ | N - k | $SS_W$ / (N-k) | - |
| **Total** | $SS_T$ | N - 1 | - | - |

**Where:**

- k = number of groups
- N = total sample size
- $SS_T = SS_B + SS_W$

## Two-Way ANOVA Table

| Source | SS | df | MS | F |
|---|---|---|---|---|
| **Factor A** | $SS_A$ | a - 1 | $MS_A$ | $MS_A$ / $MS_{Error}$ |
| **Factor B** | $SS_B$ | b - 1 | $MS_B$ | $MS_B$ / $MS_{Error}$ |
| **Interaction AB** | $SS_{AB}$ | (a-1)(b-1) | $MS_{AB}$ | $MS_{AB}$ / $MS_{Error}$ |
| **Error** | $SS_{Error}$ | ab(n-1) | $MS_{Error}$ | - |
| **Total** | $SS_{Total}$ | abn - 1 | - | - |

**Where:**

- a = levels of Factor A
- b = levels of Factor B
- n = observations per cell

**Example: Test Scores for Three Teaching Methods**

| Method A | Method B | Method C |
|---|---|---|
| 85, 88, 90, 87 | 78, 82, 80, 85 | 92, 89, 94, 91 |
| Mean = 87.5 | Mean = 81.25 | Mean = 91.5 |

**ANOVA Table for Example:**

| Source | SS | df | MS | F |
|---|---|---|---|---|
| **Between Groups** | 210.17 | 2 | 105.08 | 8.89 |
| **Within Groups** | 106.25 | 9 | 11.81 | - |
| **Total** | 316.42 | 11 | - | - |

**Result:** F = 8.89, df = (2,9), Critical value = 4.26
**Conclusion:** Since 8.89 > 4.26, teaching methods have significantly different effects.

**Post-hoc tests:** If ANOVA is significant, use additional tests (like Tukey's HSD) to determine which specific groups differ.

## Quick Comparison

| Feature | Chi-Square | ANOVA |
|---|---|---|

| Data Type | Categorical | Continuous |
|---|---|---|
| Purpose | Test independence | Compare means |
| Test Statistic | $\chi^2$ (Chi-square) | F (F-ratio) |
| Minimum Groups | 2 categories | 3 groups |

**Remember:**

- Chi-square = Categories and independence
- ANOVA = Numbers and mean comparison
- Both follow hypothesis testing steps
- Check assumptions before using either test