

---

## Statistical Inference – Formula Sheet

---

### 1. Sampling Distribution of the Sample Mean

#### 1.1 Mean — Known Variance

(Normal population or large sample, CLT)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Standardized statistic:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

---

#### 1.2 Mean — Unknown Variance

(Normal population)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Degrees of freedom:

$$\nu = n - 1$$

---

### 2. Sampling Distribution of a Proportion

$$\hat{p} = \frac{X}{n}$$

For large samples ( $np \geq 5$ ,  $n(1-p) \geq 5$ ):

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

**Important:** even if population variance unknown, the sampling distribution is **Normal**, not t-Student.

---

### 3. Sampling Distribution of the Variance

(Normal population)

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \nu = n - 1$$

---

#### 4.1 Known Variances

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

---

#### 4.2 Unknown Variances

##### 4.2a Unequal Variances (Welch Test)

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\nu$$
$$\nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

##### 4.2b Equal Variances (Pooled t-test)

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

---

#### 5. Difference of Two Proportions

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

#### 6. Ratio of Two Variances

(Normal populations)

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1, n_2-1}$$

Under  $H_0 : \sigma_1^2 = \sigma_2^2$ , simplifies to  $F = S_1^2/S_2^2$ .

---

## Confidence Intervals

### 7. Mean

Known Variance

$$\bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Unknown Variance

$$\bar{X} \pm t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

---

### 8. Proportion

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

---

### 9. Difference of Means

Known Variances

$$(\bar{X}_1 - \bar{X}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Unknown Variances

Unequal (Welch)

$$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Equal (Pooled)

$$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

---

### 10. Difference of Proportions

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

---

### 11. Variance

$$\left( \frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \right)$$

---

~

## 12. Ratio of Variances

$$\left( \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\alpha/2, n_1-1, n_2-1}}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \right)$$

---

## Hypothesis Tests

### 13. Parametric Tests

- Mean: Z-test, t-test
  - Proportion: Z-test
  - Difference of means: Z-test, Welch t-test, Pooled t-test
  - Difference of proportions: Z-test
  - Variance: Chi-square test
  - Ratio of variances: F-test
- 

### 14. Chi-Square Test of Independence

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad E_{ij} = \frac{(\text{row total})(\text{column total})}{n}, \quad \text{df} = (r - 1)(c - 1)$$

---

### 15. Chi-Square Goodness-of-Fit Test

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}, \quad \text{df} = k - 1 - m$$

where  $m$  = number of estimated parameters.

---

### One-way ANOVA – F statistic

The ANOVA F-statistic is defined as:

$$F = \frac{MSB}{MSW}$$

where:

$$MSB = \frac{SSB}{k - 1} \quad (\text{Mean Square Between groups})$$

$$MSW = \frac{SSW}{n - k} \quad (\text{Mean Square Within groups})$$

---

## Sum of Squares

$$SSB = \sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2$$
$$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$
$$SST = SSB + SSW$$

---

### Notation

- $k$ : number of groups
  - $n_j$ : sample size of group  $j$
  - $\bar{X}_j$ : mean of group  $j$
  - $\bar{X}$ : overall mean
  - $n$ : total sample size
- 

## 16. Simple Linear Regression

Model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

---

Least Squares Estimators:

Slope:

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

Intercept:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

---

Correlation Coefficient:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$$

---

Coefficient of Determination:

$$R^2 = r^2$$

---

**Test for the Slope:**

**Hypotheses:**

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

**Test statistic:**

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$
$$SE(\hat{\beta}_1) = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$
$$s^2 = \frac{\sum(y_i - \hat{y}_i)^2}{n - 2}$$

**Distribution:**

$$t \sim t_{n-2}$$

---

**Confidence Interval for the Slope:**

$$\hat{\beta}_1 \pm t_{1-\alpha/2, n-2} SE(\hat{\beta}_1)$$

---