Hypothesis Testing with Two Samples

I. **Inferences About Two Proportions**

a. Requirements

i. We have proportions from two simple random samples that are independent, which means that the sample values selected from one population are not related to the sample values from the other population..

ii. For each of the two samples, the number of "successes" is at least five and the number of "failures" is at least five.

b. Notation

i. $p_1 =$ population proportion for population 1

ii. $p_2 =$ population proportion for population 2

iii. $n_1 =$ size of the sample taken from population 1

iv. $n_2 =$ size of the sample taken from population 2

v. $x_1 =$ number of successes in the sample taken from population 1

vi. $x_2 =$ number of successes in the sample taken from population 2

vii. $\hat{p}_1 = \frac{x_1}{n_1}$ (the sample proportion)

viii. $\hat{p}_2 = \frac{x_2}{n_2}$ (the sample proportion)

ix. $\hat{q}_1 = 1 - \hat{p}_1$

x. $\hat{q}_2 = 1 - \hat{p}_2$

c. Test statistic

i. $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\overline{q}}{n_1} + \frac{\overline{p}\overline{q}}{n_2}}}$

ii. where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

d. Significance values

i. Use the z distribution.

II. **Inferences About Two Means: Independent Samples**

a. Requirements

i. The samples are independent.

ii. Both samples are simple random samples.

iii. The two sample sizes are both large (with $n_1 > 30$ and $n_2 > 30$) or both samples come from populations having normal distributions.

b. Notation

i. $\bar{x}_1 =$ sample mean for population 1

ii. $\bar{x}_2 =$ sample mean for population 2

iii. $\mu_1 =$ mean for population 1

iv. $\mu_2 =$ mean for population 2

v. $s_1^2 =$ sample variance for population 1

vi. $s_2^2 =$ sample variance for population 2

vii. $n_1 =$ sample size for population 1

viii. $n_2 =$ sample size for population 2

c. Test statistic

i. $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

d. Significance values

i. Use the Student t distribution with appropriate degrees of freedom.

ii. $df = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

III. **Inferences About Two Means: Dependent Samples**

a. Requirements

i. The sample data consist of matched pairs.

ii. The samples are simple random samples.

iii. The two sample sizes are both large (with $n_1 > 30$ and $n_2 > 30$) or both samples come from populations having normal distributions.

b. Notation

i. $d =$ individual difference between the two values in a single matched pair

ii. $\mu_d =$ mean value of the differences for the population of all matched pairs

iii. $\bar{d} =$ mean value of the differences for teh paired sample data

iv. $s_d =$ standard deviation of teh differences for the paired sample data

v. $n =$ number of pairs of data

c. Test statistic

i. $t = \dfrac{\bar{d} - \mu_d}{\dfrac{s_d}{\sqrt{n}}}$

d. Significance values

i. Use the Student t distribution with appropriate degrees of freedom.

ii. $df = n - 1$

IV. Comparing Variation in Two Samples

a. Requirements
   i. The two populations are independent of each other.
   ii. The two populations are each normally distributed.

b. Notation
   i. $n_1 =$ size of the sample with the larger variance
   ii. $n_2 =$ size of the sample with the smaller variance
   iii. $s_1^2 =$ larger of the two sample variances
   iv. $s_2^2 =$ smaller of the two sample variances

c. Test statistic
   i. $F = \frac{s_1^2}{s_2^2}$

d. Significance values
   i. Use the F distribution with appropriate degrees of freedom.
   ii. $df_1 = n_1 - 1$ (numerator)
   iii. $df_2 = n_2 - 1$ (denominator)

Assumptions of Common Statistical Tests

I. Inferences about two independent means
   a. *z*-test
      i. The *z*-test is appropriate when the experiment involves two samples and population variances are known (i.e., σ_1^2 and σ_2^2 are known).
      ii. Assumption:
         1. The sampling distribution of the mean should be normally distributed.
   b. *t*-test
      i. The unpaired *t*-test is appropriate when the experiment involves two samples that are not correlated with one another and the population variances are unknown (i.e., σ_1^2 and σ_2^2 are unknown).
      ii. Assumptions:
         1. The sampling distributions of $\bar{Y}_1$ and $\bar{Y}_2$ should be normally distributed.
         2. The samples are drawn from populations of equal variances (i.e., homogeneity of variances).
   c. Mann-Whitney-Wilcoxon test
      i. The Mann-Whitney-Wilcoxon test is used in place of the two-sample *t*-test when the normality assumption is questionable or when the data are ordinal rather than mensural.
      ii. Assumptions:
         1. The data must be at least ordinal in nature.

II. Inferences about two dependent means
   a. *t*-test
      i. The paired *t*-test is appropriate when the experiment involves two groups that are correlated with one another
      ii. Assumptions:
         1. The sampling distribution of differences between groups ($\bar{Y}_1 - \bar{Y}_2$) should be normally distributed.
   b. Wilcoxon signed rank-test
      i. The Wilcoxon signed rank-test is used in place of the two-sample *t*-test when the normality assumption is questionable or when the data are ordinal rather than mensural.
      ii. Assumptions:
         1. The data must be at least ordinal in nature.
         2. The data distribution must be symmetric.