Week 14

## Hypothesis Testing

### 2.Test of The Variance of a Population

Suppose that we have a random sample of n observations from a normal population with variance $\sigma^2$. If the observed sample variance $s^2$, then the following hypothesis testing could be constructed and test statistics are computed.

1.If the null hypothesis is simple,

$H_0: \sigma^2 = \sigma_0$ ($\sigma_0$ is a real number)

2.If the null hypothesis is compound,

$$H_0: \sigma^2 \geq (>/</\leq)\ \sigma_0$$

If Alternative hypothesis:

1.One-sided,

$1: \sigma^2 \geq (>/</\leq) \sigma_0$ ($\sigma_0$ is a real number)

2.Two-sided,

$1: \sigma^2 \neq \sigma_0$ ( $\sigma_0$ is a real number)

Ex: A company sells chemical substances and sends them in large quantities to its customers. The variance of the impurity percentage of the chemical substance is not allowed to be larger than 4. A sample size of 20 observations is drawn randomly and sample variance is found 5.62. The null hypothesis saying that the population variance is not larger than 4 is tested against the alternative hypothesis at 0.10 significance level and interpret the result.

$$H_0: \sigma^2 \leq 4$$

$$H_1: \sigma^2 > 4$$

$\chi^2_{(n-1),\alpha}$=$\chi^2_{19,0.10}$=27.20

Test statistics

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{19(5.62)}{4} = 26.69 < 27.20$$

Ho hypothesis is accepted.

Remark: at 0.10 significance level, the variance of the impurity in chemical substance is found to be less than 4

**3.The Test of Population Portion: Large Sample**

Suppose that we have a random sample of n observations from a population, a proportion p of whose members possess a particular attribute. Then, if the number of sample observations is large and the observed sample proportion is $\hat{p}$, the following tests have a significance level α.

1.If null hypothesis is simple,

$H_0: P = p_0$ ( $p_0$ is a ratio between 0 and 1)

2.If null hypothesis is compound,

$$H_0: P \geq (>/</\leq)\ p_0$$

If Alternative hypothesis,

1.One-sided,

$1: P \geq (>/</\leq) p_0$ ( $p_0$ is a ratio between 0 and 1)

2.Two-sided,

$1: P \neq p_0$ ( $p_0$ is a ratio between 0 and 1)

Test statistics

$$z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Ex: 378 out of 802 supermarket customers are asked to tell the price of a product after putting it into shopping cart. The null hypothesis saying that at least half of the customers tell the price of the product correctly is tested against the alternative hypothesis at 0.10 significance level and interpret the result.

$$H_0: P \geq 0.5$$

$$H_0: P < 0.5$$

$\hat{p} = \frac{378}{802} = 0.47$ and $1 - \hat{p} = 0.53$

$$z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.47 - 0.53}{\sqrt{\frac{(0.5)(0.5)}{802}}} = -1.64$$

$$z_{0.10} = 1.28$$

$$z_H = 1.64 > Z_T = 1.28$$

Ho is rejected

Remark: at 0.10 significance level, less than half the customers are found to remember the price of the product correctly

**Test of The Differences of the Means of Two Populations**

**1.Paired Samples**

Definition: We assume that a random sample of n matched pairs of observations is obtained from populations with means $\mu_x$ *and* $\mu_y$. The actual sample observations will be denoted $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Ex: In a study aimed at assessing the relationship between a subject's brain activity while watching a television commercial and subject's subsequent ability to recall the contents of the commercial, subjects were shown commercials for two brands of each of ten products. For each commercial, the ability to recall 24 hours later was measured, and each member of a pair of commercial was then designated high-recall or low-recall. The table below presents the data. Construct the hypothesis and test it at 0.05 significance level.

| Product No | High-Recall ($x_i$) | Low-Recall ($y_i$) | Difference=$d_i$ |
|---|---|---|---|
| 1 | 137 | 53 | 84 |
| 2 | 135 | 114 | 21 |
| 3 | 83 | 81 | 2 |
| 4 | 125 | 86 | 39 |
| 5 | 47 | 34 | 13 |
| 6 | 46 | 66 | -20 |
| 7 | 114 | 89 | 25 |
| 8 | 157 | 113 | 44 |
| 9 | 57 | 88 | -31 |
| 10 | 144 | 111 | 23 |

We can construct null and alternative hypothesis at any significance value of $\alpha$.

1.If null hypothesis is simple,

$$H_0: \mu_x - \mu_y = 0$$

2.If null hypothesis is compound,

$$H_0: \mu_x - \mu_y \geq 0 (>/</\leq)$$

If Alternative hypothesis,

1.One-sided,

$$H_1: \mu_x - \mu_y \geq 0 (>/</\leq)$$

2.Two-sided,

$1: \mu_x - \mu_y \neq 0$

Test statistics

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

$\bar{d} = \frac{\sum_{1=1}^{n} d_i}{n}$ ve $s_d^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}$

Table value

$t_{(n-1),\alpha}$ yada $t_{(n-1),\alpha/2}$

$$H_0: \mu_x - \mu_y = 0$$

$$H_1: \mu_x - \mu_y > 0$$

$$\bar{d} = \frac{84 + 21 + 2 + \cdots + (-31) + (23)}{10} = 21$$

$s_d^2 = \frac{(84-2\ \ )^2 + (21-21)^2 + \cdots + (-31-21)^2 + (23-31)^2}{10-1}$=1088

$$s_d = 32.98$$

$$t = \frac{21}{32.98/\sqrt{10}} = 2.014$$

$$t_{9,0.05} = 1.833$$

2.014>1.833. H0 hypothesis is rejected.

Remark: At 0.05 significance level, there exists a difference between the mean values of the two populations.

**Independent Sample Tests**

**Test of The Differences of the Means of Two Populations: Independent Samples (Variances Known or Large Sample Sizes)**

Definition: Suppose that we have independent random samples of $n_x$ $and$ $n_y$ observations from normal distributions with means $\mu_x$ $and$ $\mu_y$ variances $\sigma_x^2$ $and$ $\sigma_y^2$. If the observed sample means are $\bar{x}$ $and$ $\bar{y}$, then the following hypothesis testing and test statistics are denoted as follows:

We can construct null and alternative hypothesis at any significance value of $\alpha$.

1.If null hypothesis is simple,

$$H_0: \mu_x - \mu_y = 0$$

2.If null hypothesis is compound,

$$H_0: \mu_x - \mu_y \geq 0 (>/</\leq)$$

If Alternative hypothesis,

1.One-sided,

$$H_1: \mu_x - \mu_y \geq 0 (>/</\leq)$$

2.Two-sided,

$1: \mu_x - \mu_y \neq 0$

1.The variance of the populations are known:

$$z = \frac{\bar{x}_1 - \bar{y}_1}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

2.Large Sample Size:

$$z = \frac{\bar{x}_1 - \bar{y}_1}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Ex: A test was designed to measure attitudes toward women as managers. High scores indicate negative attitudes and low scores indicate positive attitudes. Independent random samples were taken of 151 male MBA students 108 female MBA students. For the former group, the sample mean and standard deviation for the test scores were 85.8 and 19.3, while the corresponding figures for the latter group were 71.5 and 12.2. test the null hypothesis that the two population means are equal against the hypothesis that the true mean is higher at 0.05 significance level.

$$\bar{x} = 85.8, s_x = 19.3, n_x = 151$$

$$\bar{y} = 71.5.8, s_y = 12.2, n_x = 108$$

$$H_0: \mu_E - \mu_K = 0$$

$$H_1: \mu_E - \mu_K > 0$$

$$z = \frac{\bar{x}_1 - \bar{y}_1}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{85.8 - 71.5}{\sqrt{\frac{(19.3)^2}{151} + \frac{(12.2)^2}{108}}} = 7.29$$

$$z_{0.05} = 1.645$$

7.29>1.645. H0 is rejected.

Remark: At 0.05 significance level, the scores for male students are higher than the female students. On other words, the attitude toward women managers is more negative among the male than among female students.

**Tests for The Difference Between the Means of Two Normal Populations: Independent Samples, Population Variances Equal**

Definition: Suppose that we have independent random samples of $n_x$ *and* $n_y$ observations from normal distributions with means $\mu_x$ *and* $\mu_y$ variances $s_x^2$ *and* $s$. If the observed sample means are $\bar{x}$ *and* $\bar{y}$, then the following hypothesis testing and test statistics are denoted as follows:

We can construct null and alternative hypothesis at any significance value of $\alpha$.

$$H_0: \sigma_x^2 = \sigma_y^2$$

$$H_1: \sigma_x^2 \neq \sigma_y^2$$

$$F = \frac{s_x^2}{s_y^2}$$

Table value is denoted by $F_{(nx-1)(ny-1),\alpha/2}$

1.If null hypothesis is simple,

$$H_0: \mu_x - \mu_y = 0$$

2.If null hypothesis is compound,

$$H_0: \mu_x - \mu_y \geq 0 (>/</\leq)$$

If Alternative hypothesis,

1.One-sided,

$$H_1: \mu_x - \mu_y \geq 0 (>/</\leq)$$

2.Two-sided,

$1: \mu_x - \mu_y \neq 0$

Pooled variance is denoted by

$$s^2 = \frac{(n_x - 1) + s_x^2 (n_y - 1) s_y^2}{n_x + n_y - 2}$$

Test statistics is denoted by

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{n_x + n_y}{n_x n_y}}}$$

Table value is denoted by $t_{nx+ny-2,\alpha}$ ya da $t_{nx+ny \quad , \ \alpha/2}$.

Ex: A study attempted to assess the effect of the presence of a moderator on the number of ideas generated by a group. Groups of four members, with or without moderators, were observed. For a random sample of four groups with a moderator, the mean number of ideas generated per group was 78.0 and the sample standard deviation 24.4. for an independent random sample of four groups without a moderator, the mean number of ideas generated was 63.5, and the sample standard deviation was 20.2. Assuming that the population distributions are normal with equal variances, test the null hypothesis that the population means are equal against the alternative that the true mean is higher for groups with moderator at 0.10 significance value.

$$H_0: \sigma_x^2 = \sigma_y^2$$

$$H_1: \sigma_x^2 \neq \sigma_y^2$$

$F = \frac{(24.4)^2}{(20.2)^2} = \frac{595.36}{408.04} = 1.46$ ve$F_{3,30.05=}$9.2766.

1.46<9.2766, H0 is accepted, the it is assumed that variances are homogenous.

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x > \mu_y$$

$$s^2 = \frac{(3)(24.4)^2 + (3)(20.2)^2}{4+4-2} = 501.7$$

S=22.4

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{n_x + n_y}{n_x n_y}}} = \frac{78 - 63.5}{22.4\sqrt{\frac{8}{16}}} = 0.915$$

$$t_{nx+ny-2,\ \alpha} = 1.44$$

0.915<1.44, H0 is accepted.

**Tests for The Difference Between Two Population Proportions: Large Samples**

Definition: Let $\widehat{p_x}$ denote the proportion of successes in a random sample of $n_x$ observations from a population with proportion $P_x$ $successes$ and $\hat{p}_y$ denote the proportion of successes in random sample of $n_y$, observations from a population with proportion of $P_y$ successes.

We can construct null and alternative hypothesis at any significance value of $\alpha$.

1.If null hypothesis is simple,

$$H_0: P_x - P_y = 0$$

2.If null hypothesis is compound,

$$H_0: P_x - P_y \geq 0 (>/</\leq)$$

If Alternative hypothesis is,

1.One-sided,

$$H_1: P_x - P_y \geq 0 (>/</\leq)$$

2.Two-sided,

$1: P_x - P_y \neq 0$

Test statistics

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{(1-P_x)P_x}{n_x} + \frac{(1-P_y)P_y}{n_y}}}$$

Table value $z_\alpha$ $veya$ $z_{\alpha/2}$

Ex: Employees of a supermarket chain, facing shutdown, were surveyed on a prospective employee ownership plan. Some employees pledged 5.000 TL to this plan, putting up 200 TL immediately, while others indicated that they did not intend to pledge. Of a random sample of 175 pledgers, 44.6 percent had already been laid off, while 34.5 percent of an independent random sample of 604 non-pledgers had already been laid off. Test against two-sided alternative the null hypothesis that the population proportions already laid off were the same for pledgers as for non-pledgers at 0.05 significance level.

$$H_0: P_x = P_y$$

$$H_0: P_x \neq P_y$$

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{(1-P_x)P_x}{n_x} + \frac{(1-P_y)P_y}{n_y}}} = \frac{0.446 - 0.345}{\sqrt{\frac{(0.446)(0.544)}{175} + \frac{(0.345)(0.655)}{604}}} = 2.44$$

$$z_{0.025} = 1.96$$

2.44>1.96. H0 is rejected.

## Correlation

Definition**:** Let X and Y be a pair of random variables, with means $\mu_x$ $and$ $\mu_y$ and variances $\sigma_x^2$ $and$ $\sigma_y^2$. A measure of the strength of linear associations is provided by the correlation coefficient, $\rho$ $(ro)$. The sample correlation coefficient is denoted by

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

It can be shown that the correlation must lie between -1 and 1, that is,

$$-1 \leq \rho \leq 1$$

With the following interpretations:

1.A correlation of -1 implies perfect negative linear association

2. A correlation of 1 implies perfect positive linear association

3.A correlation of 0 implies no linear association.

4.The larger in absolute value the correlation, the stronger the linear association between the random variables.

5.For the correlation coefficients taking values in between, the table given below is used.

| Correlation Values | Interpretations |
|---|---|
| 0.0-0.19 | Very Weak |
| 0.20-0.39 | Weak |
| 0.40-0.59 | Medium |
| 0.60-0.79 | Strong |
| 0.80-0.99 | Very Strong |

Ex: For X and Y random variables, the data is given as follows:

| X | Y |
|---|---|
| 12 | 32 |
| 14 | 36 |
| 18 | 49 |
| 22 | 28 |
| 27 | 29 |

| $(x_i - \bar{x})(y_i - \bar{y})$ | $(x_i - \bar{x})^2$ | $(y_i - \bar{y})^2$ |
|---|---|---|
| (12-18.6)(32-34.80) | $(12 - 18.6)^2$ | $(32 - 34.8)^2$ |
| (14-18.6)(36-34.80) | $(14 - 18.6)^2$ | $(36 - 34.8)^2$ |
| (18-18.6)(49-34.80) | $(18 - 18.6)^2$ | $(49 - 34.8)^2$ |
| (22-18.6)(28-34.80) | $(22 - 18.6)^2$ | $(28 - 34.8)^2$ |
| (27-18.6)(29-34.80) | $(27 - 18.6)^2$ | $(29 - 34.8)^2$ |

r=-.0.326.