HYPOTHESIS TESTING

In this session ….

- What is hypothesis testing?
- Interpreting and selecting significance level
- Type I and Type II errors
- One tailed and two tailed tests
- Hypothesis tests for population mean
- Hypothesis tests for population proportion
- Hypothesis tests for population standard deviation

What is Hypothesis Testing?

Hypothesis testing refers to

1. Making an assumption, called hypothesis, about a population parameter.
2. Collecting sample data.
3. Calculating a sample statistic.
4. Using the sample statistic to evaluate the hypothesis (how likely is it that our hypothesized parameter is correct. To test the validity of our assumption we determine the difference between the hypothesized parameter value and the sample value.)

HYPOTHESIS TESTING

Null hypothesis, H_0

- State the hypothesized value of the parameter before sampling.
- The assumption we wish to test (or the assumption we are trying to reject)
- E.g population mean $\mu = 20$
- There is no difference between coke and diet coke

Alternative hypothesis, H_A

All possible alternatives other than the null hypothesis.

E.g $\mu \neq 20$

$\mu > 20$

$\mu < 20$

There is a difference between coke and diet coke

Null Hypothesis

The null hypothesis H_0 represents a theory that has been put forward either because it is believed to be true or because it is used as a basis for an argument and has not been proven. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write

H_0: there is no difference between the two drugs on an average.

Alternative Hypothesis

The alternative hypothesis, H_A, is a statement of what a statistical hypothesis test is set up to establish. For example, in the clinical trial of a new drug, the alternative hypothesis might be that the new drug has a different effect, on average, compared to that of the current drug. We would write

H_A: the two drugs have different effects, on average.

or

H_A: the new drug is better than the current drug, on average.

The result of a hypothesis test:

'Reject H_0 in favour of H_A' OR 'Do not reject H_0'

Selecting and interpreting significance level

1. Deciding on a criterion for accepting or rejecting the null hypothesis.
2. Significance level refers to the percentage of sample means that is outside certain prescribed limits. E.g testing a hypothesis at 5% level of significance means

- that we reject the null hypothesis if it falls in the two regions of area 0.025.
- Do not reject the null hypothesis if it falls within the region of area 0.95.

3. The higher the level of significance, the higher is the probability of rejecting the null hypothesis when it is true. (acceptance region narrows)

Type I and Type II Errors

1. Type I error refers to the situation when we reject the null hypothesis when it is true (H_0 is wrongly rejected).
   e.g H_0: there is no difference between the two drugs on average.
   Type I error will occur if we conclude that the two drugs produce different effects when actually there isn't a difference.
   Prob(Type I error) = significance level = α
2. Type II error refers to the situation when we accept the null hypothesis when it is false.
   H_0: there is no difference between the two drugs on average.
   Type II error will occur if we conclude that the two drugs produce the same effect when actually there is a difference.
   Prob(Type II error) = ß

Type I and Type II Errors – Example

Your null hypothesis is that the battery for a heart pacemaker has an average life of 300 days, with the alternative hypothesis that the average life is more than 300 days. You are the quality control manager for the battery manufacturer.

(a) Would you rather make a Type I error or a Type II error?

(b) Based on your answer to part (a), should you use a high or low significance level?

Type I and Type II Errors – Example

Given H_0 : average life of pacemaker = 300 days, and H_A: Average life of pacemaker > 300 days

(a) It is better to make a Type II error (where H_0 is false i.e average life is actually more than 300 days but we accept H_0 and assume that the average life is equal to 300 days)

(b) As we increase the significance level (α) we increase the chances of making a type I error. Since here it is better to make a type II error we shall choose a low α.

Two Tail Test

Two tailed test will reject the null hypothesis if the sample mean is significantly higher or lower than the hypothesized mean.

Appropriate when $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$

e.g The manufacturer of light bulbs wants to produce light bulbs with a mean life of 1000 hours. If the lifetime is shorter he will lose customers to the competition and if it is longer then he will incur a high cost of production. He does not want to deviate significantly from 1000 hours in either direction. Thus he selects the hypotheses as

$H_0 : \mu = 1000$ hours and $H_A : \mu \neq 1000$ hours

and uses a two tail test.

One Tail Test

A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H_0 are located entirely in one tail of the probability distribution.

Lower tailed test will reject the null hypothesis if the sample mean is significantly lower than the hypothesized mean. Appropriate when $H_0 : \mu = \mu_0$ and $H_A : \mu < \mu_0$

e.g A wholesaler buys light bulbs from the manufacturer in large lots and decides not to accept a lot unless the mean life is at least 1000 hours.

$H_0 : \mu =$ 1000 hours and $H_A : \mu <$ 1000 hours

and uses a lower tail test.

i.e he rejects H_0 only if the mean life of sampled bulbs is significantly below 1000 hours. (he accepts H_A and rejects the lot)

One Tail Test

Upper tailed test will reject the null hypothesis if the sample mean is significantly higher than the hypothesized mean. Appropriate when $H_0 : \mu = \mu_0$ and $H_A: \mu > \mu_0$

e.g A highway safety engineer decides to test the load bearing capacity of a 20 year old bridge. The minimum load-bearing capacity of the bridge must be at least 10 tons.

$H_0 : \mu =$ 10 tons and $H_A: \mu >$ 10 tons

and uses an upper tail test.

i.e he rejects H_0 only if the mean load bearing capacity of the bridge is significantly higher than 10 tons.

Hypothesis test for population mean

$H_0 : \mu = \mu_0$ and Test statistic $\Delta = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s}$

For H_A: $\mu > \mu_0$, reject H_0 if $\Delta > t_{n-1,\alpha}$

For H_A: $\mu < \mu_0$, reject H_0 if $\Delta < -t_{n-1,\alpha}$

For H_A: $\mu \neq \mu_0$, reject H_0 if $|\Delta| > t_{n-1,\alpha/2}$

For $n \geq 30$, replace $t_{n-1,\alpha}$ *by* z_α

Hypothesis test for population mean

A weight reducing program that includes a strict diet and exercise claims on its online advertisement that it can help an average overweight person lose 10 pounds in three months. Following the program's method a group of twelve overweight persons have lost 8.1 5.7 11.6 12.9 3.8 5.9 7.8 9.1 7.0 8.2 9.3 and 8.0 pounds in three months. Test at 5% level of significance whether the program's advertisement is overstating the reality.

Hypothesis test for population mean

Solution:

H_0: μ = 10 (μ_0) H_A: μ < 10 (μ_0)

n = 12, x(bar) = 8.027, s = 2.536, α = 0.05

$$\Delta = \frac{\sqrt{12}(8.075 - 10)}{2.536} = \frac{3.46 \times -1.925}{2.536} = -2.62$$

Critical t-value = $-t_{n-1,\alpha}$ = $- t_{11,0.05}$ = -2. 201 (TINV)

Since $\Delta < -t_{n-1,\alpha}$ we reject H_0 and conclude that the program is overstating the reality.

(What happens if we take α = 0.01? Is the program overstating the reality at 1% significance level?)

Hypothesis test for population proportion

$H_0 : p = p_0$ and Test statistic $\Delta = \dfrac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1-p_0)}}$

For H_A: $p > p_0$ reject H_0 if $\Delta > z_\alpha$

For H_A: $p < p_0$ reject H_0 if $\Delta < -z_\alpha$

For H_A: $p \neq p_0$ reject H_0 if $|\Delta| > z_{\alpha/2}$

Hypothesis test for population proportion

A ketchup manufacturer is in the process of deciding whether to produce an extra spicy brand. The company's marketing research department used a national telephone survey of 6000 households and found the extra spicy ketchup would be purchased by 335 of them. A much more extensive study made two years ago showed that 5% of the households would purchase the brand then. At a 2% significance level, should the company conclude that there is an increased interest in the extra-spicy flavor?

Hypothesis test for population proportion

$$n = 6000, \quad \hat{p} = \frac{335}{6000} = 0.05583$$

$$H_0 : p = 0.05\,(p_0) \qquad H_A : p > 0.05$$

$$\Delta = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1-p_0)}} = \frac{\sqrt{6000} \times 0.00583}{\sqrt{0.05 \times 0.95}}$$

$$= \frac{77.459 \times 0.00583}{0.218} = 2.072$$

$$\alpha = 0.02$$

$Z_\alpha (\textit{the critical value of } Z) = 2.05$ (NORMSINV)

$\because \Delta > Z_\alpha$ *we reject* H_0 i.e the current interest is significantly greater than the interest of two years ago.

Hypothesis test for population standard deviation

$H_0 : \sigma = \sigma_0$ and Test statistic $\Delta = \dfrac{(n-1)s^2}{\sigma_0{}^2}$

For H_A: $\sigma > \sigma_0$ reject H_0 if $\Delta > \chi^{2(R)}_{(n-1),\alpha}$

For H_A: $\sigma < \sigma_0$ reject H_0 if $\Delta < \chi^{2(R)}_{(n-1),1-\alpha}$

For H_A: $\sigma \neq \sigma_0$ reject H_0 if $\Delta < \chi^{2(R)}_{(n-1),1-\alpha/2}$ or $\Delta > \chi^{2(R)}_{(n-1),\alpha/2}$

Hypothesis test for comparing two population means

Consider two populations with means μ_1, μ_2 and standard deviations σ_1 and σ_2. $\mu_{\bar{x}_1} = \mu_1$ *and* $\mu_{\bar{x}_2} = \mu_2$ are the means of the sampling distributions of population1 and population2 respectively. $\sigma_{\bar{x}_1}$ *and* $\sigma_{\bar{x}_2}$ denote the standard errors of the sampling distributions of the means.

$\mu_{\bar{x}_1 - \bar{x}_2}$ is the mean of the difference between sample means and $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ is the corresponding standard error.

$H_0 : \mu_1 = \mu_2$ and test statistic, $\Delta = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}}$

For H_A: $\mu_1 > \mu_2$ reject H_0 if $\Delta > Z_\alpha$

For H_A: $\mu_1 < \mu_2$ reject H_0 if $\Delta <- Z_\alpha$

For H_A: $\mu_1 \neq \mu_2$ reject H_0 if $|\Delta| > Z_{\alpha/2}$

Here Δ denotes the standardized difference of sample means

(decision makers may be concerned with parameters of two populations e.g do female employees receive lower salary than their male counterparts for the same job)

Hypothesis test for comparing population means

A sample of 32 money market mutual funds was chosen on January 1, 1996 and the average annual rate of return over the past 30 days was found to be 3.23% and the sample standard deviation was 0.51%. A year earlier a sample of 38 money-market funds showed an average rate of return of 4.36% and the sample standard deviation was 0.84%. Is it reasonable to conclude (at $\alpha = 0.05$) that money-market interest rates declined during 1995?

Hypothesis test for comparing population means

$$n_1 = 32, \bar{x}_1 = 3.23,\ \sigma_1 = 0.51 \quad n_2 = 38, \bar{x}_2 = 4.36,\ \sigma_2 = 0.84$$

$$H_0 : \mu_1 = \mu_2 \qquad H_A : \mu_1 < \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}} = \sqrt{\frac{0.26}{32} + \frac{0.71}{38}} = \sqrt{0.026} = 0.163$$

$$\Delta = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{-1.13 - 0}{0.163} = -6.92$$

$$\alpha = 0.05$$

$\textit{Critical value of } Z = -Z_\alpha = -1.64$

$\because \Delta < -Z_\alpha$ *we reject* H_0 *and conclude that there has been a decline.*

Hypothesis test for comparing population proportions

Consider two samples of sizes n_1 and n_2 with $\bar{p}_1$ and $\bar{p}_2$ as the respective proportions of successes. Then

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$ is the estimated overall proportion of successes in the two populations.

$$\hat{\sigma}_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$$ is the estimated standard error of the difference between the two proportions.

$H_0 : p_1 = p_2$ and test statistic, $\Delta = \dfrac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}$

For H_A: $p_1 > p_2$ reject H_0 if $\Delta > Z_\alpha$

For H_A: $p_1 < p_2$ reject H_0 if $\Delta <- Z_\alpha$

For H_A: $p_1 \neq p_2$ reject H_0 if $|\Delta| > Z_{\alpha/2}$

A training director may wish to determine if the proportion of promotable employees at one office is different from that of another.

Hypothesis test for comparing population proportions

A large hotel chain is trying to decide whether to convert more of its rooms into non-smoking rooms. In a random sample of 400 guests last year, 166 had requested non-smoking rooms. This year 205 guests in a sample of 380 preferred the non-smoking rooms. Would you recommend that the hotel chain convert more rooms to non-smoking? Support your recommendation by testing the appropriate hypotheses at a 0.01 level of significance.

Hypothesis test for comparing population proportions

$$n_1 = 400, \bar{p}_1 = \frac{166}{400} = 0.415, \qquad n_2 = 380, \bar{p}_2 = \frac{205}{380} = 0.5395$$

$$H_0 : p_1 = p_2 \qquad H_A : p_1 < p_2$$

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{400 \times 0.415 + 380 \times 0.5395}{400 + 380} = 0.4757$$

(Proportion of success in the two populations)

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.4757 \times 0.5243\left(\frac{1}{400} + \frac{1}{380}\right)} = 0.0358$$

$$\alpha = 0.01$$

$$\textit{Critical value of } Z = -Z_\alpha = -2.32$$

$$\Delta = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)_{H_0}}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}} = \frac{-0.1245 - 0}{0.0358} = -3.48$$

$$\because \Delta < -Z_\alpha \textit{ we reject } H_0$$

The hotel chain should convert more rooms to non-smoking rooms as there has been a significant increase in the number of guests seeking non-smoking rooms.