ECON20003 QM2

Tutorial 5 Semester 1, 2022

Chin Quek

Department of Economics

March/April 2022

Admin

Mid Semester Test is scheduled next week, from 8am Monday 4th April till 10am Wednesday 6th April (Melbourne Time).

- Information about the actual test is provided in the Sample test.

Topics covered to date

Week 1 Lectures:

- Estimation and Hypothesis Testing

Week 2 Lectures:

- Desirable Properties of Point Estimators
- Parametric and Nonparametric Techniques
- The Assumption of Normality

Week 3 Lectures:

- Comparing Two Population Means or Central Locations with
  - Parametric, and
  - Non-parametric Techniques

Objectives

- Inferences about Population Variances
  - Single population
  - Two populations
- Inferences about Population Proportions
  - Single population
  - Two populations

Reminder: Always create an R Project and R script when starting on a new exercise.

Inferences about Population Variance(s) - Review Question 1

A process is in control if the maximum variance is 0.05. Assume that the population is normally distributed. A sample of size 26 showed a sample variance of 0.06. Is the process in control?

Inferences about Population Variance(s) - Review Question 1

A process is in control if the maximum variance is 0.05. Assume that the population is normally distributed. A sample of size 26 showed a sample variance of 0.06. Is the process in control?

$H_0 : \sigma_0^2 \leq 0.05 \qquad H_A : \sigma_0^2 > 0.05$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Test statistic

$$\chi^2_{obs} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{25 \times 0.06}{0.05} = 30$$

For $\alpha = 0.05$, $\chi^2_{0.05} = 37.7$

```
qchisq(0.05, 25, lower.tail = F)
```

```
## [1] 37.65248
```

Assumptions for the χ^2-test and corresponding CI

i. The data is a random sample of independent observations.
ii. The variable of interest is quantitative and continuous.
iii. The measurement scale is interval or ratio.
iv. The sampled population is normally distributed.

Review Quesion 2

Suppose we are interested in conducting a hypothesis test of the equality of two population variances at $\alpha = 5\%$ level of significance. What is/are the F-percentile values (use either the relevant statistical table or R) associated with this test. The sample sizes are $n_1 = 31$ and $n_2 = 23$.

$$F = \frac{s_1^2}{s_2^2} \sim F_{n_1 - 1, n_2 - 1}$$

Review Quesion 2

Suppose we are interested in conducting a hypothesis test of the equality of two population variances at $\alpha = 5\%$ level of significance. What is/are the F-percentile values (use either the relevant statistical table or R) associated with this test. The sample sizes are $n_1 = 31$ and $n_2 = 23$.

$$F = \frac{s_1^2}{s_2^2} \sim F_{n_1-1, n_2-1}$$

Critical values

$F_{0.025,30,22} = 2.2718$ and $F_{0.975,30,22} = 0.4623$

```
qf(0.025, df1=30, df2=22, lower.tail = FALSE)
```

```
## [1] 2.27184
```

```
qf(0.975, df1=30, df2=22, lower.tail = FALSE)
```

```
## [1] 0.4622932
```

If done manually, to determine the left-tail critical value, recall:

$$F_{0.975,30,22} = \frac{1}{F_{0.025,22,30}} = \frac{1}{2.163} = 0.4623$$

```r
qf(0.025, df1=22, df2=30, lower.tail = FALSE)
```

```
## [1] 2.16313
```

What if we define the test statistic to be $\frac{s_2^2}{s_1^2}$?

Assumptions for the F-test and corresponding CI

i. The data consists of two independent random samples of independent observations.
ii. The variable of interest is quantitative and continuous.
iii. The measurement scale is interval or ratio.
iv. The sampled populations are normally distributed.

The ability to taste the chemical Phenylthiocarbamide (PTC) is hereditary. Some people can taste it, while others cannot. Even though the ability to taste PTC was observed in all age, race, and sex groups, this does not address the issue about whether men or women are more likely to be able taste PTC.

Researchers want to know if the ability to taste PTC is a sex-linked trait and the following summarises results from a study:

Can Taste PTC	**Female**	**Male**	**Total**
No	15	14	**29**
Yes	51	38	**89**
Total	**66**	**52**	**118**

Conduct an appropriate hypothesis test using $\alpha = 0.05$.

- Also consider how the test change(s) (if any) should the researchers be interested in whether the proportion of males who can taste PTC is 10% more than the proportion of females.

Identify the statistical technique

- If the sample size is large enough, which can be determined by determining whether $np \geq 5$ **and** $nq \geq 5$ then we can use a standard normal distribution
- Must look at the null hypothesis when testing for differences between two population proportions,

$$H_0 : p_1 - p_2 = D_0.$$

Identify the statistical technique

- If the sample size is large enough, which can be determined by determining whether $np \geq 5$ **and** $nq \geq 5$ then we can use a standard normal distribution
- Must look at the null hypothesis when testing for differences between two population proportions,

$$H_0 : p_1 - p_2 = D_0.$$

If $D_0 = 0$ the the standard error is pooled e.g. $\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ while if $D_0 \neq 0$ then the standard error is $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$.

Identify the statistical technique

- If the sample size is large enough, which can be determined by determining whether $np \geq 5$ **and** $nq \geq 5$ then we can use a standard normal distribution
- Must look at the null hypothesis when testing for differences between two population proportions,

$$H_0 : p_1 - p_2 = D_0.$$

If $D_0 = 0$ the the standard error is pooled e.g. $\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ while if $D_0 \neq 0$ then the standard error is $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$.

From the question, the exact population proportion is unknown, so strictly speaking we cannot check whether the "large" requirements, $np \geq 5$ **and** $nq \geq 5$, are satisfied.

Identify the statistical technique

- If the sample size is large enough, which can be determined by determining whether $np \geq 5$ **and** $nq \geq 5$ then we can use a standard normal distribution
- Must look at the null hypothesis when testing for differences between two population proportions,
$$H_0 : p_1 - p_2 = D_0.$$

If $D_0 = 0$ the the standard error is pooled e.g. $\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ while if $D_0 \neq 0$ then the standard error is $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$.

From the question, the exact population proportion is unknown, so strictly speaking we cannot check whether the "large" requirements, $np \geq 5$ **and** $nq \geq 5$, are satisfied.

At best we can replace p with its estimate $\hat{p}$, and if $n\hat{p}$ and $n\hat{q}$ are both relatively large, then we can be willing to assume that $np \geq 5$ and $nq \geq 5$are also satisfied justifying the large-sample confidence interval estimation of the population proportion.

Let population 1 be females and population 2 be males.

$$H_0 : p_1 - p_2 = 0 \qquad\qquad H_A : p_1 - p_2 \neq 0$$

If H_0 is true, then

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

where $\hat{p} = \frac{f_1+f_2}{n_1+n_2}$

Calculations

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{51}{66} = 0.773 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{38}{52} = 0.731$$

$$\hat{p} = \frac{51+38}{66+52} = \frac{89}{118} \approx 0.7542$$

$$z_{obs} = \frac{0.773 - 0.731}{\sqrt{0.7542 \times (1 - 0.7542) \times \left(\frac{1}{66} + \frac{1}{52}\right)}} = 0.526$$

For a two-tailed test

- Reject H_0 if $z_{obs} > 1.96$ or $z_{obs} < -1.96$.

p-value:

$$\text{p-value} = 2 \times \Pr(Z > 0.526)$$

```
2*pnorm(0.526, lower.tail = F)
```

```
## [1] 0.5988882
```