ECON20003 QM2

Tutorial 6 Semester 1, 2022

Chin Quek

Department of Economics

April 2022

Objectives

- Comparing Several Population Central Locations
  - Parametric tests (ANOVA F test based on independent samples, Welch F, ANOVA F test based on randomised blocks)
  - Nonparametric tests (Kruskal Wallis, Friedman test with ties and without ties F_{rc}, F_r)

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

Previously, we studied the comparison of two populations - independent samples and paired (repeated measures) samples.

- Generalisation of the two-independent-sample Z/t test: one-way analysis of variance (ANOVA) based on independent measures design
- Extension of the matched pair Z/t test: one-way ANOVA based on randomised blocks

Review Question 1

Complete the given one-way ANOVA (independent samples) table:

Source of Variation	**Degree of freedom**	**Sum of Squares**	**Mean Squares**	***F***
Between Treatments	2	a	27.25	b
Within Samples	c	d	e	-
Total	**11**	**272.25**	-	-

Review Question 1

Complete the given one-way ANOVA (independent samples) table:

Source of Variation	**Degree of freedom**	**Sum of Squares**	**Mean Squares**	***F***
Between Treatments	2	a	27.25	b
Within Samples	c	d	e	-
Total	**11**	**272.25**	-	-

a: $2 \times 27.25 = 54.5$

c: $11 - 2 = 9$

d: $272.25 - 54.5 = 217.75$

e: $217.75/9 = 24.19$

b: $27.25/24.19 = 1.13$

Exercise 1

The friendly folks at the Taxpayers Association are always looking for ways to improve the wording and format of their tax return forms. Three new forms have been developed recently. To determine which, if any, are superior to the current form, 120 individuals were asked to participate in an experiment. Each of the three new forms and the currently used form were filled out by 30 different people. The amount of time (in minutes) taken by each person to complete the task was recorded and stored in the *t6e1* Excel file.

Identify the statistical technique

Exercise 1

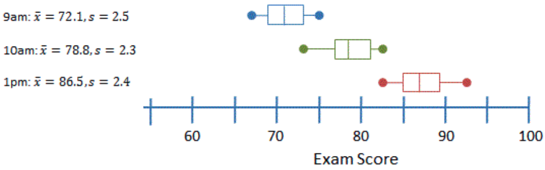
The friendly folks at the Taxpayers Association are always looking for ways to improve the wording and format of their tax return forms. Three new forms have been developed recently. To determine which, if any, are superior to the current form, 120 individuals were asked to participate in an experiment. Each of the three new forms and the currently used form were filled out by 30 different people. The amount of time (in minutes) taken by each person to complete the task was recorded and stored in the *t6e1* Excel file.

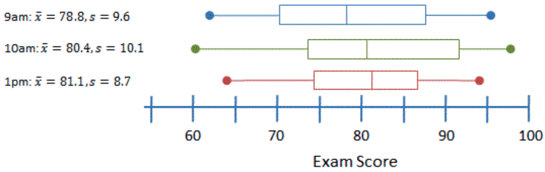
Identify the statistical technique

- One-Way ANOVA Based on the Independent Measures Design

One-Way ANOVA Based on the Independent Measures Design

Let's take a look at this example.





In this case, we can see that the means do appear to be different, but not by much. And within each sample, there is a lot of variation, so the difference in the means could just be due to the wide variation within each group.

> *The point here is that we can't just consider the differences in the means - whether those differences are significant or not depends on the standard deviations (and sample sizes, of course).*

This is the basic idea behind One-Way ANOVA. It's one-way, because we're focusing on a single characteristic (time of class period, in our example above). And the ANOVA stands for analysis of variance. While it may seem odd that the title refers to variance when we're actually comparing means, it's actually because of a **significant assumption** - to perform the statistical analysis.

This is the basic idea behind One-Way ANOVA. It's one-way, because we're focusing on a single characteristic (time of class period, in our example above). And the ANOVA stands for analysis of variance. While it may seem odd that the title refers to variance when we're actually comparing means, it's actually because of a **significant assumption** - to perform the statistical analysis.

Assumptions

Parametric one-way ANOVA based on independent samples has five conditions:

(i) The data set constitutes k independent random samples of independent observations drawn from k (sub-) populations.
(ii) The variable of interest is quantitative and continuous.
(iii) The measurement scale is interval or ratio.
(iv) Each (sub-) population is normally distributed, . . .
(v) . . . and has the same variance.

To perform the test, we focus on the between-sample variation (between the means) and the within-sample variation (i.e. the standard deviation). If the former is large in comparison to the latter, we can say that one of the means must be different.

The test statistic that we use is another F-statistic, and it's the ratio of these two variations:

$$F = \frac{s_0^2}{s_p^2} = \frac{MST}{MSE}$$

where $MST = \frac{SST}{k-1} = \frac{\sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{x}\right)^2}{k-1}$ and $MSE = \frac{SSE}{n-k} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} n_j\left(x_{ij} - \bar{x}_j\right)^2}{n-k}$

Note

This test is always a right-tail test!

Checking for equal population variances assumption

A generalisation of the equality of two population variances is the Levene's test:

$$H_0 : \sigma_0^2 = \sigma_2^2 = \cdots = \sigma_k^2 \qquad H_A : \text{ not all } \sigma_i^2 \text{ are equal}$$

So what if we reject the null of equal population variances?

For the Welch F-test, do remember to drop the `var.equal` argument in the `oneway.test()` R command

Exercise 2

It is common practice in the advertising business to create several different advertisements and then ask a random sample of potential customers to rate the ads on several different dimensions. Suppose that an advertising firm developed four different ads for a new breakfast cereal and asked a random sample of 400 shoppers to rate the believability of the advertisements. One hundred people viewed ad 1, another 100 viewed ad 2, another 100 saw ad 3, and another 100 saw ad 4. The ratings were: very believable (4), quite believable (3), somewhat believable (2) and not believable at all (1). The responses are stored in the *t6e2 Excel* file.

Based on this data, can the firm's management conclude at the 5% significance level that differences exist in believability among the four ads?

Identify the statistical technique

Exercise 2

It is common practice in the advertising business to create several different advertisements and then ask a random sample of potential customers to rate the ads on several different dimensions. Suppose that an advertising firm developed four different ads for a new breakfast cereal and asked a random sample of 400 shoppers to rate the believability of the advertisements. One hundred people viewed ad 1, another 100 viewed ad 2, another 100 saw ad 3, and another 100 saw ad 4. The ratings were: very believable (4), quite believable (3), somewhat believable (2) and not believable at all (1). The responses are stored in the *t6e2 Excel* file.

Based on this data, can the firm's management conclude at the 5% significance level that differences exist in believability among the four ads?

Identify the statistical technique

- Kruskal-Wallis Test

Kruskal-Wallis Test

When the mean does not exist or the sampled populations are clearly not normally distributed, we should use neither the ANOVA F-test nor the Welch F-test, but some nonparametric test instead. The nonparametric counterpart of these tests is the Kruskal-Wallis test, a generalisation of the Wilcoxon rank-sum test to two or more (sub-) populations.

The Kruskal-Wallis test is a one-way ANOVA test for the equality of $k \geq 2$ (sub-) population medians based on the ranks of the observations in the pooled set of k independent samples, one from each (sub-) population.

Kruskal-Wallis Test Assumptions

1. The data set constitutes k independent random samples of independent observations drawn from k (sub-) populations.
2. The variable of interest is quantitative and continuous.
3. The measurement scale is at least ordinal.
4. The sampled populations differ at most with respect to their central locations (i.e. medians).

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{T_j^2}{n_j} - 3(n+1) \sim \chi^2_{k-1} \qquad \text{for sample sizes } \geq 5$$

Question

WHY can we use Kruskal-Wallis for Ex 2? Doesn't it fail assumption ii.?

Question

WHY can we use Kruskal-Wallis for Ex 2? Doesn't it fail assumption ii.?

What is the underlying variable of interest?

The problem is that continuous variables can never be observed in practice with infinite precision due to the limitations of the measurement tools. No matter whether you think about time, length, weight, opinion, preference etc., which are all continuous variables, when we observe/measure them in practice, we do so with limited precision and the actual measurements are always rounded numbers. Hence, the possible values form a discrete set.

Question

WHY can we use Kruskal-Wallis for Ex 2? Doesn't it fail assumption ii.?

What is the underlying variable of interest?

The problem is that continuous variables can never be observed in practice with infinite precision due to the limitations of the measurement tools. No matter whether you think about time, length, weight, opinion, preference etc., which are all continuous variables, when we observe/measure them in practice, we do so with limited precision and the actual measurements are always rounded numbers. Hence, the possible values form a discrete set.

In Exercise 2 of Tutorial 6, the variable of interest is opinion about **believability**. This is a continuous variable, although the actual ratings are measured on an ordinal scale that has only four possible categories. The numbers assigned to these categories (1: not believable at all, . . . , 4 very believable) are absolutely arbitrary. We could use any four different real numbers to distinguish and rank the four categories. For this reason, this data set is ranked but qualitative.

So, in short, it would help to focus on the variable of interest. It does not matter if the actual measurements are on an ordinal scale or not.

Review Question 2

Give the main reason to divide the experimental units into subgroups/blocks.

Review Question 2

Give the main reason to divide the experimental units into subgroups/blocks.

- If the experimental units are not homogeneous, then we divide it into subgroups such that within each block data is homogeneous.

Review Question 3

In a one-way ANOVA using randomised block design (with k treatments and b blocks), there are 4 treatments and 5 blocks. Complete the table below.

Source of Variation	Degree of freedom	Sum of Squares	Mean Squares	F
Treatments	a	147	b	c
Blocks	d	35	e	f
Error	g	43	h	-
Total	i	225	-	-

$$F_T = \frac{MST}{MSE} \sim F_{k-1,\, n-k-b+1}$$
$$F_B = \frac{MSB}{MSE} \sim F_{b-1,\, n-k-b+1}$$

Review Question 3

In a one-way ANOVA using randomised block design (with k treatments and b blocks), there are 4 treatments and 5 blocks. Complete the table below.

Source of Variation	Degree of freedom	Sum of Squares	Mean Squares	F
Treatments	a	147	b	c
Blocks	d	35	e	f
Error	g	43	h	-
Total	i	225	-	-

$$F_T = \frac{MST}{MSE} \sim F_{k-1,\, n-k-b+1}$$
$$F_B = \frac{MSB}{MSE} \sim F_{b-1,\, n-k-b+1}$$

a: $4 - 1 = 3$ d: $5 - 1 = 4$ i: $(4 \times 5) - 1 = 19$

g: $(19 - 3 - 4)$ or $(20 - 4 - 5 + 1) = 12$ b: $147/3 = 49$

e: $35/4 = 8.75$ h: $43/12 = 3.58$ c: $49/3.58 = 13.69$

f: $8.75/3.58 = 2.44$

Exercise 3

There is often concern among students as to the consistency of marking between lecturers. It is common that lecturers obtain reputations for being 'hard' or 'light' markers, but there is often little to substantiate these reputations. A group of students investigated the consistency of marking by submitting the same eight essays to four different lecturers. The mark given by each lecturer was recorded for each essay. It is important to emphasize that the same essays were used for all lecturers because this eliminates any individual differences in the standard of work that each lecturer marked. The data shown in the following table are on the *Wide* sheet of the *t6e3* Excel file.

	Lecturer_1	Lecturer_2	Lecturer_3	Lecturer_4
Essay_1	62	58	63	64
Essay_2	63	60	68	65
Essay_3	65	61	72	65
Essay_4	68	64	58	61
Essay_5	69	65	54	59
Essay_6	71	67	65	50
Essay_7	78	66	67	50
Essay_8	75	73	75	45

Conduct a one-way parametric ANOVA on these data to test the null hypotheses that

(i) the lecturers mark consistently;
(ii) the essays are marked consistently.

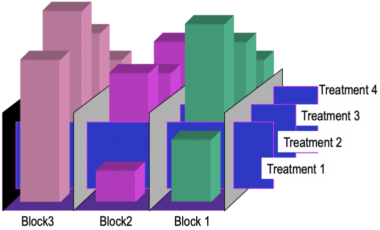
Use $\alpha = 0.05$. Perform the tests first manually and then in `R`.

One-Way ANOVA Based on the Randomised Block Design

The purpose of designing a randomised block experiment is to reduce the within-treatments variation, thus increasing the relative amount of between treatment variation.

This helps in detecting differences between the treatment means more easily.

Block all the observations with some commonality across treatments



	Treatment				
Block	1	2		k	**Block mean**
1	X11	X12	. . .	X1k	$\bar{x}[B]_1$
2	X21	X22		X2k	$\bar{x}[B]_2$
.					
.					
.					
b	Xb1	Xb2		Xbk	$\bar{x}[B]_b$
Treatment mean	$\bar{x}[T]_1$	$\bar{x}[T]_2$		$\bar{x}[T]_k$	

Another example of Randomised Block Designs

Are there differences in the effectiveness of blood pressure reduction drugs?

- Factor: Blood pressure drugs
- Treatment (Response): Blood pressure
- Experimental Units: Patient
- Block: Same age, sex, overall condition

  Using this procedure eliminates the variability in BP reduction related to different combinations of age, sex, overall condition. This helps detect differences in the mean BP reduction attributed to the different drugs.

Another example of Randomised Block Designs

Are there differences in the effectiveness of blood pressure reduction drugs?

- Factor: Blood pressure drugs
- Treatment (Response): Blood pressure
- Experimental Units: Patient
- Block: Same age, sex, overall condition

  Using this procedure eliminates the variability in BP reduction related to different combinations of age, sex, overall condition. This helps detect differences in the mean BP reduction attributed to the different drugs.

Six assumptions of Parametric one-way ANOVA based on a randomised block design with k treatments and b blocks

(i) The data is a random sample of b independent blocks of k number of observations that are not independent of each other.
(ii) The variable of interest is quantitative and continuous.
(iii) The measurement scale is interval or ratio.
(iv) Each (sub-) population is normally distributed, . . .
(v) . . . and has the same variance.
(vi) The block and treatment effects are additive.

Decomposition of Total sum of squares SS

$$SS = SST + SSB + SSE$$

where

$$SS = \sum_{j=1}^{k}\sum_{i=1}^{b}(x_{ij} - \bar{x})^2 \qquad SST = b\sum_{j=1}^{k}(\bar{x}_{T,j} - \bar{x})^2$$

$$SSB = k\sum_{i=1}^{b}(\bar{x}_{B,i} - \bar{x})^2 \qquad SSE = \sum_{j=1}^{k}\sum_{i=1}^{b}(x_{ij} - \bar{x}_{T,j} - \bar{x}_{B,i} + \bar{x})^2$$

How is this different to independent samples design?

Hypothesis tests

To perform hypothesis tests for treatments and blocks, we need

- Mean square for treatments $MST = \frac{SST}{k-1}$
- Mean square for blocks $MSB = \frac{SSB}{b-1}$
- Mean square for error $MSE = \frac{SSE}{n-k-b+1}$

Test statistics

$$F_T = \frac{MST}{MSE} \sim F_{k-1,n-k-b+1} \qquad F_B = \frac{MSB}{MSE} \sim F_{b-1,n-k-b+1}$$

To perform hypothesis tests for treatments and blocks, we need

- Mean square for treatments $MST = \frac{SST}{k-1}$
- Mean square for blocks $MSB = \frac{SSB}{b-1}$
- Mean square for error $MSE = \frac{SSE}{n-k-b+1}$

Test statistics

$$F_T = \frac{MST}{MSE} \sim F_{k-1,n-k-b+1} \qquad F_B = \frac{MSB}{MSE} \sim F_{b-1,n-k-b+1}$$

F-test rejection regions

- Testing the mean responses for treatment: $F > F_{\alpha,k-1,n-k-b+1}$
- Testing the mean responses for blocks: $F > F_{\alpha,b-1,n-k-b+1}$

*When using R, it is important to take a look at whether you have the data in **long** or **wide** format!*

Exercise 4

The personnel manager of a national accounting firm has been receiving complaints from senior managers about the quality of recent hirings. All new accountants are hired through a process whereby four managers interview each candidate and rate him / her on several dimensions such as academic credentials, work experience, and personal suitability. Each manager then summarises the results and evaluates the candidates on a scale 1-5. The evaluations are finally combined to make the final decision.

The personnel manager believes that the quality problem is caused by the evaluation system. However, she needs to know whether there is general agreement among the interviewing managers in their evaluations. To test for differences, she takes a random sample of 8 applicants.

	Manager			
Applicant	1	2	3	4
1	2	1	2	2
2	4	2	3	2
3	2	2	2	3
4	3	1	3	2
5	3	2	3	5
6	2	2	3	4
7	4	1	5	5
8	3	2	5	3

What conclusion can the personnel manager draw from this experiment at the 5% significance level?

Friedman Test (Non-parametric)

The Friedman test is the non-parametric alternative to the one-way ANOVA with repeated measures.

> *In Exercise 4, The variable of interest is the managers' evaluation of the candidates. It is a discrete variable; thus, it is not normally distributed. Cannot use parametric ANOVA.*

Three assumptions of Friedman test with k treatments and b blocks

(i) The data is a random sample of b independent blocks of k number of observations that are not independent of each other.

(ii) The variable of interest is quantitative and continuous.

(iii) The measurement scale is at least ordinal.

Friedman Test (Non-parametric)

The Friedman test is the non-parametric alternative to the one-way ANOVA with repeated measures.

In Exercise 4, The variable of interest is the managers' evaluation of the candidates. It is a discrete variable; thus, it is not normally distributed. Cannot use parametric ANOVA.

Three assumptions of Friedman test with k treatments and b blocks

(i) The data is a random sample of b independent blocks of k number of observations that are not independent of each other.
(ii) The variable of interest is quantitative and continuous.
(iii) The measurement scale is at least ordinal.

Recall how the one-way ANOVA F-test on randomised blocks require the subpopulations to be normally distributed and of equal variance?

The Friedman test can be used to test for differences between (sub-) populations using arandomised block design when the normality and homoskedasticity (i.e. equal variance) requirements of parametric ANOVA are not satisfied

OR when the dependent variable being measured is ordinal.

Hypotheses

$H_0 : \eta_1 = \eta_2 = \cdots = \eta_k$ and H_A : not all population medians are equal

Rank the observations in each block from smallest to largest (k), and sum the ranks across each treatment

Correction factor

Test statistic MUST BE divided by a correction factor if there are tied observations

$$F_{rc} = \frac{F_r}{C} \qquad \text{where } F_r = \frac{12}{b \times k \times (k+1)} \sum_{j=1}^{k} T_j^2 - 3b(k+1)$$

and

$$C = 1 - \frac{\sum_{i-1}^{b} (t_i^3 - t_i)}{b(k^3 - k)}$$

and t_i is the number of tied scores in the i^{th} block

Sampling distribution of F_r and F_{rc} is non-standard, but when k or b is sufficiently large ($k > 6$ or $b > 24$), it is approximated with a χ^2_{k-1}.

Calculations of the ranks

	Manager							
Applicant	1	Rank	2	Rank	3	Rank	4	Rank
1	2	3.0	1	1.0	2	3.0	2	3.0
2	4	4.0	2	1.5	3	3.0	2	1.5
3	2	2.0	2	2.0	2	2.0	3	4.0
4	3	3.5	1	1.0	3	3.5	2	2.0
5	3	2.5	2	1.0	3	2.5	5	4.0
6	2	1.5	2	1.5	3	3.0	4	4.0
7	4	2.0	1	1.0	5	3.5	5	3.5
8	3	2.5	2	1.0	5	4.0	3	2.5
Sum		21.0		10.0		24.5		24.5

Do we need to use the correction factor?

$$C = 1 - \frac{\sum_{i-1}^{b}(t_i^3 - t_i)}{b(k^3 - k)} = 1 - \frac{2(3^3 - 3) + 6(2^3 - 2)}{8(4^3 - 4)} = 0.825$$

and the corrected test statistic is

$$F_{rc} = \frac{F_r}{C} = \frac{10.61}{0.825} = 12.86$$

What is the distribution of the test statistic? Would it be χ^2*?*

Small sample critical values for the Friedman test[1]

	$k=3$		$k=4$		$k=5$		$k=6$	
b	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
2	-	-	6.000	-	7.600	8.000	9.143	9.714
3	6.000	-	7.400	9.000	8.533	10.13	9.857	11.76
4	6.500	8.000	7.800	9.600	8.800	11.20	10.29	12.71
5	6.400	8.400	7.800	9.960	8.960	11.68	10.49	13.23
6	7.000	9.000	7.600	10.20	9.067	11.87	10.57	13.62
7	7.143	8.857	7.800	10.54	9.143	12.11		
8	6.250	9.000	7.650	10.50	9.200	12.30		
9	6.222	9.556	7.667	10.73	9.244	12.44		
10	6.200	9.600	7.680	10.68				
11	6.545	9.455	7.691	10.75				
12	6.500	9.500	7.700	10.80				
13	6.615	9.385	7.800	10.85				
14	6.143	9.143	7.714	10.89				
15	6.400	8.933	7.720	10.92				
16	6.500	9.375	7.800	10.95				
17	6.118	9.294	7.800	11.05				
18	6.333	9.000	7.733	10.93				
19	6.421	9.579	7.863	11.02				
20	6.300	9.300	7.800	11.10				
21	6.095	9.238	7.800	11.06				
22	6.091	9.091	7.800	11.07				
23	6.348	9.391						
24	6.250	9.250						