Baltazar, Carl Joseph
Samson, Justine Aizel

FA # 8 - ANOVA

Introduction

If you want to determine whether there are any statistically significant differences between the means of two or more independent groups, you can use a one-way analysis of variance (ANOVA).

Assumptions

Assumption #1: You have one dependent variable that is measured at the continuous level.

Assumption #2: You have one independent variable that consists of three or more categorical, independent groups.

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group of the independent variable or among the groups themselves.

Assumption #4: There should be no significant outliers in the three or more groups of your independent variable in terms of the dependent variable.

Assumption #5: Your dependent variable should be approximately normally distributed for each group of the independent variable.

Assumption #6. You have homogeneity of variances (i.e., the variance of the dependent variable is equal in each group of your independent variable).

Null and alternative hypotheses

The null hypothesis for a one-way analysis of variance is

H_0: all group population means are equal (i.e., $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$)

where μ = population mean and k = a number of groups.

H_A: at least one group population mean is different (i.e., they are not all equal)

Baltazar, Carl Joseph
Samson, Justine Aizel

Dataset and Problem

A researcher believes that individuals who are more physically active are better able to cope with stress in the workplace. To test this theory, the researcher recruited 31 subjects and measured how many minutes of physical activity they performed per week and their ability to cope with workplace stress. The subjects were categorized into four groups based on the number of minutes of physical activity they performed: namely, "sedentary", "low", "moderate" and "high" physical activity groups. These groups (levels of physical activity) formed an independent variable called a group. The ability to cope with workplace stress was assessed as the average score of a series of Likert items on a questionnaire, which allowed an overall "coping with workplace stress" score to be calculated; higher scores indicated a greater ability to cope with workplace-related stress. This dependent variable was called coping_stress and "ability to cope with workplace-related stress" abbreviated as the "CWWS" score. The researcher would like to know if the CWWS score is dependent on physical activity level. In variable terms, is the mean coping_stress score different for different levels of a group?

Checking of Assumptions

Assumption #1: You have one dependent variable that is measured at the continuous level.

> **Remark.** This dependent variable was referred to as a group, and the "ability to cope with workplace-related stress" score was abbreviated as the "CWWS" score. This dependent variable is on a continuous scale.

Assumption #2: You have one independent variable that consists of two categorical, independent groups.

Remark. Physical activity is the independent factor, and it is classified into four distinct groups: sedentary, low, moderate, and high.

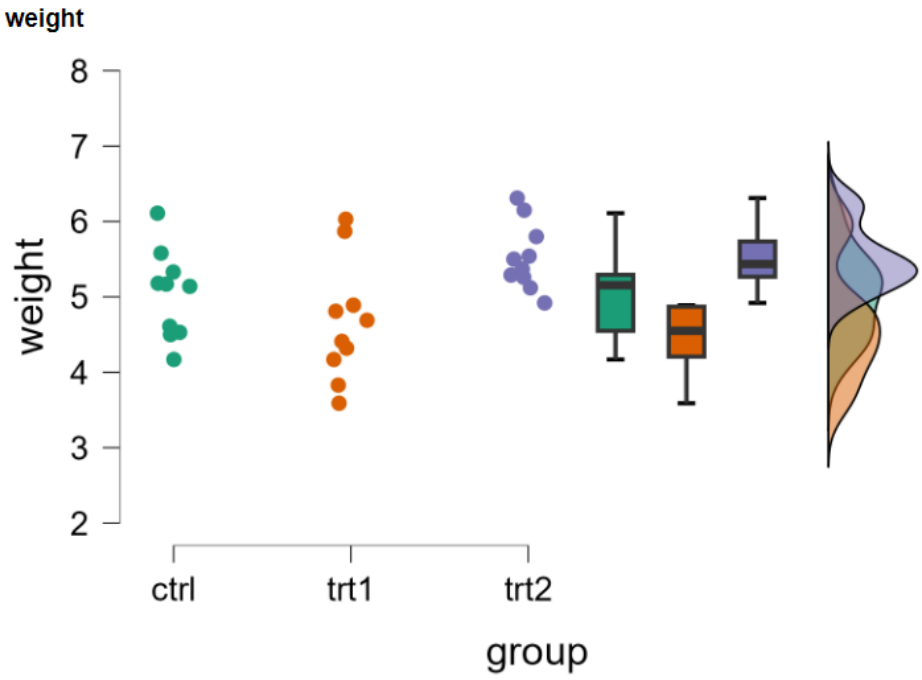
Assumption #3: You should have independence of observations.

> **Remark.** Given that there is no connection among the findings in each group of an independent variable or between the groups themselves, each observation is independent of the others.

Baltazar, Carl Joseph
Samson, Justine Aizel

Assumption #4: There should be no significant outliers in the three or more groups of your independent variable in terms of the dependent variable.

Raincloud plots



Remark. Visual inspection of boxplots revealed no significant outliers in the physical activity groups in terms of weight score.

Assumption #5: Your dependent variable should be approximately normally distributed for each group of the independent variable.

Descriptive Statistics

	weight
Valid	30
Missing	0
Mean	5.073
Std. Deviation	0.701
Skewness	−0.162
Std. Error of Skewness	0.427
Kurtosis	−0.553
Std. Error of Kurtosis	0.833
Shapiro-Wilk	0.983
P-value of Shapiro-Wilk	0.892
Minimum	3.590
Maximum	6.310

Remark. The group score is approximately normally distributed for each of the physical activity groups, as assessed by Shapiro-Wilk's test, $p > 0.983$.

Baltazar, Carl Joseph
Samson, Justine Aizel

Assumption #6. You have homogeneity of variances (i.e., the variance of the dependent variable is equal in each group of your independent variable).

Test for Equality of Variances (Levene's) ▼

F	df1	df2	p
1.237	2.000	27.000	0.306

Remark. There was homogeneity of variances of the dependent variable for all physical activity groups, as assessed by Levene's test of homogeneity of variances, $p = 0.306$.

Computation

ANOVA - weight

Cases	Sum of Squares	df	Mean Square	F	p	η^2	η^2_p
group	3.766	2	1.883	4.846	0.016	0.264	0.264
Residuals	10.492	27	0.389				

Note. Type III Sum of Squares

Descriptives – Group

Descriptives - weight

group	N	Mean	SD	SE	Coefficient of variation
ctrl	10	5.032	0.583	0.184	0.116
trt1	10	4.661	0.794	0.251	0.170
trt2	10	5.526	0.443	0.140	0.080

Post Hoc Tests ▼

Standard ▼

Bootstrapped Post Hoc Comparisons - group ▼

			95% bca† CI						95% CI for Cohen's d		
		Mean Difference	Lower	Upper	SE	bias	t	Cohen's d	Lower	Upper	p_{tukey}
ctrl	trt1	0.367	−0.207	0.930	0.300	9.338×10^{-4}	1.331	0.595	−0.565	1.755	0.391
	trt2	−0.503	−0.902	0.010	0.226	-2.879×10^{-4}	−1.772	−0.792	−1.967	0.382	0.198
trt1	trt2	−0.875	−1.387	−0.255	0.281	−0.001	−3.103	−1.388	−2.627	−0.149	0.012

† Bias corrected accelerated.
Note. Bootstrapping based on 1000 successful replicates.
Note. Mean Difference estimate is based on the median of the bootstrap distribution.
Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

Baltazar, Carl Joseph
Samson, Justine Aizel

Reporting

A one-way ANOVA was conducted to determine if the ability to cope with workplace-related stress (CWWS score) was different for groups with different physical activity levels. Participants were classified into four groups: sedentary (*n = 0.3*), low (*n = -0.2*), moderate (*n =9*), and high levels of physical activity (*n = 7*). There were no outliers, as assessed by visual inspection of the boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test (*p > 0.983*); and variances were homogeneous, as assessed by Levene's test of homogeneity of variances (*p = .120*). Data is presented as mean ± standard deviation. CWWS score was statistically significantly different between different physical activity groups, *F(3, 27) = 4.846, p < 0.016*, $\eta^2_p = 0.264$. CWWS score increased from the sedentary (*M = 5.032, SD = 0.583*) to the low (*M = 5.88, SD = 1.69*), moderate (*M = 4.661, SD = 0.794*), and high (M = 5.526, SD = 0.443) physical activity groups, in that order. Tukey post hoc analysis revealed that the mean increase from sedentary to moderate (*2.97, 95% CI [0.99, 4.96]*) was statistically significant (*p = .002*), as well as the increase from sedentary to high (*3.35, 95% CI [1.30, 5.40], p < .001*), but no other group differences were statistically significant.

Reference

Bundock, V. A. (n.d.). *Rdatasets/csv/datasets/PlantGrowth.csv at master · vincentarelbundock/Rdatasets*. GitHub. Retrieved November 16, 2023, from https://github.com/vincentarelbundock/Rdatasets/blob/master/csv/datasets/PlantGrowth.csv