

ANOVA Example

Below is the output for the SPSS ONEWAY procedure to compare the means of three learning strategies in an extension of our hypothetical studying example (data are the same as in the hand computation illustrated in prior ANOVA handout).¹ In this example, there is a third group, concept mapping, added to our earlier reading only and retrieval practice groups. Concept mapping involves drawing a diagram with nodes to illustrate links between concepts. This example has a pretty small sample size per group and such a small sample is not necessarily recommended. It is certainly legitimate to do an ANOVA with this size sample, but one should be particularly conscious of unequal variances.

Syntax²

```
means tables=recall by groups.
oneway recall by groups
/statistics=welch
/es=overall.
```

Menus

Analyze—Compare Means and Proportions—One-way ANOVA

Report

recall number of new terms recalled			
groups learning strategy experimental groups	Mean	N	Std. Deviation
1.00 reading only	6.0000	5	2.00000
2.00 retrieval practice	9.0000	5	1.00000
3.00 concept mapping	6.0000	5	.70711
Total	7.0000	15	1.92725

ANOVA

recall number of new terms recalled					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.000	2	15.000	8.182	.006
Within Groups	22.000	12	1.833		
Total	52.000	14			

ANOVA Effect Sizes^{a,b}

recall number of new terms recalled	Effect Size	Point Estimate	95% Confidence Interval	
			Lower	Upper
Eta-squared	.577	.087	.734	
Epsilon-squared	.506	-.066	.690	
Omega-squared Fixed-effect	.489	-.061	.675	
Omega-squared Random-effect	.324	-.030	.509	

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Robust Tests of Equality of Means

recall number of new terms recalled				
	Statistic ^a	df1	df2	Sig.
Welch	14.259	2	7.211	.003

a. Asymptotically F distributed.

Unequal group sizes, very different variances, and small sample sizes can increase Type I error rates (or sometimes increase Type II error rates; Clinch & Keselman, 1982; Tomarken & Serlin, 1986). If the ratio of variances in different groups is less than 4 to 1 and the group sizes are over 5, there is generally a very minor impact on Type I error rates. If data are very non-normal, variances are very unequal, and/or the group sample sizes are small and unequal, the Welch's robust test is an alternative that will generally have a more accurate Type I error rate.

¹ Although these are artificial data, they are based on actual research studies and the results do mirror findings in some learning strategy studies, such as results reported in Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772-775.

² It is also possible to do the same one-way ANOVA using several other procedures in SPSS, including the GLM through the menus or GLM, ANOVA, or MANOVA (syntax only) commands. Use of the GLM procedure, `glm recall by groups.`, will give the eta-squared (shown as R-squared in output), but the Welch's test is not an option.

Although the group sizes are equal, the small overall sample size and the relatively substantial differences in the standard deviations (and therefore larger differences in the variances) suggest that it may be wise to consider the Welch's test in this case (Welch, 1951). The statistical conclusion does not differ from the standard F -test, however (this could be noted in a write-up, where either or both is reported, if this comes up). With larger N s, as in most of the applications you will encounter, the correction would likely be unnecessary.

R

```
> library(lessR)
> ANOVA(recall ~ groups, brief=TRUE)
```

DESCRIPTIVE STATISTICS

	n	mean	sd	min	max
1	5	6.00	2.00	4.00	8.00
2	5	9.00	1.00	8.00	10.00
3	5	6.00	0.71	5.00	7.00

Grand Mean: 7

ANOVA

-- Summary Table for recall

	df	Sum Sq	Mean Sq	F-value	p-value
groups	2	30.00	15.00	8.18	0.0057
Residuals	12	22.00	1.83		

-- Association and Effect Size for recall

R Squared: 0.577
R Sq Adjusted: 0.506
Omega Squared: 0.489

Write-up

An analysis of variance (ANOVA) was conducted to compare the means of the three experimental groups. Results indicated student recall differed significantly among the groups, $F(2,12) = 8.18$, $p = .006$. Because of the substantial differences in variances among groups and the small sample size, the Welch's test was conducted. The results from the Welch's test also indicated a significant difference among the groups, $F_W(2,7.21) = 14.26$, $p = .003$.³ Students had the best recall in retrieval practice group ($M = 9.00$, $SD = 1.00$), and the lowest recall in the reading only ($M = 6.00$, $SD = 2.00$) and concept mapping groups ($M = 6.00$, $SD = .71$). Approximately 58% of the variance in recall was accounted for by the type of learning strategy, $\eta^2 = .58$.

References

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 330-336.

³ The Welch test sometimes may be referred to as W . It is not technically an F , but the F distribution is used as the referent distribution when determining significance.