The General Linear Model for the One-Way ANOVA

Maxwell & Delaney

The General Linear Model

When we conduct research, we often measure subjects on dependent variables (e.g., IQ, anxiety, depression) and we try to explain the subjects scores on the dependent variables using certain factors (e.g., group membership, treatment condition, sex, income). However, we are not able to include all possible influences on the dependent variable.

Observed Score on Dependent Variable	=	*Effect of constant factors*	+	*Effect of factors considered*	+	*Randomly varying, uncontrolled influences*

For example, if we wanted to predict a child's IQ from their parent(s) IQ then we would use a model in the form of.....

Child's IQ	=	*Average IQ (e.g., 100)*	+	*Amount Parents' IQ differs from Average*	+	*Other Factors (including human variability) not controlled.*

What represents a "good" general linear model?

- Have a set of well-chosen variables (based on theory or reason).
- Include enough independent variables (factors) to predict the dependent variable.
- Use the simplest model possible (i.e., use as few variables as possible).

Using the General Linear Model to Analyze Data from a One Group Design

- This design is usually analzed with the **one sample *t* test**.
- The GLM approach is **mathematically equivalent** to the one sample *t* test.
- The principles of model comparisons apply to any number of samples.

Example. A professor wanted to measure the attitude level of students in statistics courses. The professor sampled 10 students about their attitudes levels. Students gave a number between 1=very negative and 7 = very positive. The responses are given below. Use an alpha of .05 to determine whether the attitude levels are significantly different than neutral.

Null Hypothesis	**Reduced (Null) Model**
Alternative Hypothesis	**Full (Alternative) Model**

	Reduced Model				**Full Model**		
Scores	**Predicted**	**Error**	**Squared Error**		**Predicted**	**Error**	**Squared Error**
7							
5							
3							
4							
7							
7							
5							
3							
2							
7							
Mean			E_R				E_F

Compare the prediction errors of the full and reduced models.

$df_R = N$

$df_F = N\text{-}1$

$E_R =$ $E_F =$

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/df_F}$$

Obtain the *Critical Value* from the *F* distribution to determine whether the *F* statistic is significantly greater than 1.

$df_{Numerator} = df_R - df_F = N - (N\text{-}1) = 1$

$df_{Denominator} = df_F = N - 1$

Critical Value =

Interpret the results of the test.

SPSS Instructions
→Analyze →Compare Means→One-Sample T Test
Test Variable(s): attitude
Test Value: 4

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
attitude	10	5.0000	1.94365	.61464

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
attitude	1.627	9	.138	1.00000	-.3904	2.3904

DEGREES OF FREEDOM FOR THE MODELS

df = number of independent components – number of parameters estimated

df_R =

df_F =

MINIMIZING PREDICTION ERRORS…

- The errors (e_i*'s*) reflect how much the scores differ from the predicted values.

- We want to minimize our prediction errors. Technically speaking, we seek to minimize the squared prediction errors. Minimizing the squared prediction errors is described as using the **least squares criterion**.

$$\min \sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n} (Y_i - \hat{\mu})^2 = \min \sum (Y_i - \hat{Y}_i)^2 \quad \textbf{where } \hat{Y}_i = \textit{predicted score}$$

- Using the least squares criterion ensures:
  - The average prediction error is zero.
  - The prediction errors are as small as possible.
  - The estimates are unbiased.
  - The estimates are minimum variance estimates (efficiency).

- Mathematicians have proven that the best estimate of the unknown population mean is the sample mean. Using the sample mean to estimate the population mean will minimize the squared prediction errors (least squares).

COMPARING PREDICTION ERRORS OF THE REDUCED AND FULL MODELS

- ❑ The total of the prediction errors for the reduced model (E_R) is never smaller than the total of the prediction errors for the full model (E_F). Thus, the reduced model is never more accurate than the full model.

- ❑ We really want to know whether the full model (i.e., alternative hypothesis) is **significantly more accurate** than the reduced model (i.e., null hypothesis).

- ❑ We estimate the <u>proportional increase in error</u> caused by using the reduced model. The increase in error is found by subtracting E_F from E_R ; the proportional increase in error is found by dividing the difference by the minimal error (E_F). $\boxed{\frac{E_R - E_F}{E_F}}$

- ❑ We adjust the <u>proportional increase in error</u> to reflect the difference in <u>degrees of freedom</u> of the Full and Restricted Models. This adjustment gives us an "average" amount of error per degree of freedom for each model and reflects the difference in the number of variables used in the Full and Restricted Models.

$$F = \frac{(E_R - E_F) / (df_R - df_F)}{E_F / df_F}$$

If the Null Hypothesis is Correct,

- The only difference between E_R and E_F is due to sampling variability. Thus, the numerator represents the sampling variability per degree of freedom.

- The denominator also represents the sampling variability per degree of freedom.
- The F statistic has an expected value of $\frac{\sigma_e^2}{\sigma_e^2}$ which **equals 1**.

If the Alternative Hypothesis is Correct,

- The difference between E_R and E_F represents the <u>increase in error</u> per degree of freedom caused by collapsing across an important factor.

- The denominator represents the sampling variability per degree of freedom.

- The F statistic has an expected value of $\frac{\sigma_e^2 + \frac{\sum_j n_j \alpha_j^2}{a-1}}{\sigma_e^2}$ which is **larger than 1.**

Using the GLM to Analyze Data from a Two Group Design

- ❑ This design is usually analyzed with the *t* test for two independent samples.
- ❑ The GLM is mathematically equivalent to the *t* test for two independent samples.

Example. A researcher wanted to determine whether gender is a significant predictor of preferred number of children. The responses are given below. Use an alpha of .05 to determine whether gender is a significant predictor of preferred number of children.

Null Hypothesis	**Reduced (Null) Model**
Alternative Hypothesis	**Full (Alternative) Model**

Note. 5 observations were used per group, but we should have had at least 15 per group!

		Reduced Model			**Full Model**		
Gender	**Scores**	**Predicted**	**Error**	**Squared Error**	**Predicted**	**Error**	**Squared Error**
Female	0						
Female	2						
Female	1						
Female	5						
Female	2						
Male	4						
Male	3						
Male	4						
Male	5						
Male	4						
Mean			E_R				E_F

Comparing the Prediction Errors of the Full and Reduced Models

$df_R = N - 1 =$

$df_F = N - a =$

$E_R =$ $E_F =$

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F / df_F}$$

Obtain the critical value from the *F* distribution to determine whether the *F* statistics is significantly greater than 1.

$df_{Numerator} = df_R - df_F = (N - 1) - (N - a) = a - 1$

$df_{Denominator} = df_F = N - a$

Critical Value =

Interpret the results of the test.

SPSS Instructions
→Analyze→Compare Means→Independent Samples T Test
Test Variable(s): Kids
Grouping Variable: Gender
→Define Groups
Group 1: 1
Group 2: 2

T-Test

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
KIDS	Female	5	2.0000	1.87083	.83666
	Male	5	4.0000	.70711	.31623

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
KIDS	Equal variances assumed	1.600	.242	-2.236	8	.056	-2.0000	.89443	-4.06255	.06255
	Equal variances not assumed			-2.236	5.120	.074	-2.0000	.89443	-4.28308	.28308

SPSS Instructions
→Analyze→Compare Means→One-Way ANOVA
Dependent List: Kids
Factor: Gender
→Options
√ Descriptives

Oneway

Descriptives

KIDS

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Female	5	2.0000	1.87083	.83666	-.3229	4.3229	.00	5.00
Male	5	4.0000	.70711	.31623	3.1220	4.8780	3.00	5.00
Total	10	3.0000	1.69967	.53748	1.7841	4.2159	.00	5.00

ANOVA

KIDS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.000	1	10.000	5.000	.056
Within Groups	16.000	8	2.000		
Total	26.000	9			

Using the GLM to Analyze Data from Three or More Groups

- This design is usually analyzed with the GLM (one-way ANOVA)

Example. A researcher wanted to determine whether marital status (never married, divorced, married) is a significant predictor of the hours per week of television viewing (tvhours). Use an alpha of .05 to determine whether marital status is related to tvhours.

Null Hypothesis	**Reduced (Null) Model**
Alternative Hypothesis	**Full (Alternative) Model**

		Reduced Model			**Full Model**		
Marital	**Scores**	**Predicted**	**Error**	**Squared Error**	**Predicted**	**Error**	**Squared Error**
Never	7						
Never	12						
Never	5						
Never	10						
Never	16						
Divorced	15						
Divorced	10						
Divorced	12						
Divorced	10						
Divorced	13						
Married	11						
Married	26						
Married	22						
Married	13						
Married	13						
Mean			E_R				E_F

Comparing the Prediction Errors of the Full and Reduced Models

$df_R = N - 1 =$

$df_F = N - a =$

$E_R =$ $E_F =$

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F / df_F}$$

Obtain the critical value from the *F* distribution to determine whether the *F* statistics is significantly greater than 1.

$df_{Numerator} = df_R - df_F = (N - 1) - (N - a) = a - 1$

$df_{Denominator} = df_F = N - a$

Critical Value =

Interpret the results of the test.

Using SPSS to Analyze Data from the *a* Group Design: One-Way ANOVA

SPSS Instructions
→Analyze→Compare Means→One-Way ANOVA
Dependent List: TVHOURS
Factor: MARITALSTATUS
→Options
√ Descriptives

Oneway

Descriptives

TVHOURS

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Never	5	10.0000	4.30116	1.92354	4.6594	15.3406	5.00	16.00
Divorced	5	12.0000	2.12132	.94868	9.3660	14.6340	10.00	15.00
Married	5	17.0000	6.59545	2.94958	8.8107	25.1893	11.00	26.00
Total	15	13.0000	5.31843	1.37321	10.0548	15.9452	5.00	26.00

ANOVA

TVHOURS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	130.000	2	65.000	2.932	.092
Within Groups	266.000	12	22.167		
Total	396.000	14			

Conducting a One-Way ANOVA from Descriptive Statistics

Marital Status	Sample Size (n)	*df* (*n* – 1)	Mean (*M*)	*SD* (*s*)	Variance *SS*/*df*	Sum of Squares (*SS*) *df**Variance
Never	5		10	4.30		
Divorced	5		12	2.12		
Married	5		17	6.60		
Total	**15**		**13**	**5.32**		

$\boldsymbol{E_R = SS_{Total} =}$

$\boldsymbol{df_R = df_{Total} =}$

$\boldsymbol{E_F = SS_{Within} =}$

$\boldsymbol{df_{Full} = df_{Within} =}$

$\boldsymbol{SS_{Between} = E_R - E_F = SS_{Total} - SS_{Within}}$

$\boldsymbol{df_{Between} = df_R - df_F = df_{Total} - df_{Within} =}$

ANOVA Summary Table

Source	*SS*	*df*	$MS = SS/df$	$F = MS_{Between}/MS_{Within}$
Between				
Within				
Total				

$$\eta^2 = \frac{E_R - E_F}{E_R} = \frac{SS_{Between}}{SS_{Total}}$$

MEASURES OF EFFECT SIZE

Standardized Difference Between Means:

Two Means

$$d = \frac{(\mu_1 - \mu_2)}{\sigma_e}\text{; the sample-based equivalent is } \hat{d} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{MSE}}$$

a Means

(1) Use the maximum difference between the *a* means as the estimate of the population effect sizes.

$$\hat{d} = \frac{(\bar{Y}_{\max} - \bar{Y}_{\min})}{\sqrt{MSE}}$$

(2) Use the standard deviation of the means divided by the within-group standard deviation...this yields an overall standardized effect size. **Cohen (1988) suggests .10, .25, and .40 corresponds to small, medium, and large effect sizes, respectively.**

$$f = \frac{\sigma_{means}}{\sigma_{within}}\text{ ; the sample-based equivalent is…}$$

$$f = \frac{S_{means}}{\sqrt{MSE}} \text{ where } S_{means} = \sqrt{\frac{\sum(\bar{Y}_j - \bar{Y})^2}{a}} = \sqrt{\frac{\sum \hat{\alpha}_j^2}{a}}$$

The *f* statistic for the standardized effect size is frequently used in conjunction with **power** calculations. **NOTE:** The standard deviation of the means is NOT equivalent to $MS_{between}$.

MEASURES OF ASSOCIATION

Cohen (1988) suggests values of .01, .06, and .14 may be viewed as small, medium, and large, respectively, for the three measures of association.

Coefficient of Multiple Determination (ranges between 0 and 1, really!)

$$R^2 = \frac{SS_A}{SS_{Total}} = \frac{E_R - E_F}{E_R}$$

Adjusted *R*^2^ (ranges between 0 and 1; if less than 0, report 0).

$$R^2_{Adj} = 1 - \left(\frac{N-1}{N-a}\right)\left(1 - R^2\right)$$

Omega-Squared (ranges between 0 and 1; if less than 0, report 0).

$$\omega^2 = \frac{(a-1)(F-1)}{(a-1)(F-1) + N}$$

Assumptions and Robustness of the One-Way ANOVA

Three assumptions must be met for *F* statistic to follow an *F* distribution.

1. The residuals of the dependent variable must be normal within each group.

- ANOVA generally controls the *p(type I)* when the residuals are non-normal.

- ANOVA is generally robust in terms of *power* when the residuals are non-normal. An exception is when the data come from a mixed-normal distribution. The power of the ANOVA test would be greatly reduced if the data are mixed-normal.

2. The population variances of the residuals for each group should be the same: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2$

Equal n

- With *moderate cell sizes*, the ANOVA is robust to fairly large differences in population variances.
- With *small sample sizes, the Welch's ANOVA F test should be used instead of the one-way ANOVA.*

Unequal n

- If the **small sample sizes** are associated with the **smaller variances** then the actual *p(type I)* is drastically **less** than the alpha set by the researcher.

- If the **small sample sizes** are associated with the **larger variances** then the actual *p(type I)* may be drastically **greater** than the alpha set by the researcher.

- With unequal sample sizes, the *Welch's ANOVA F test should be used instead of the one-way ANOVA.*

3. The residuals must be statistically independent of each other.

- Positively correlated residuals would lead to an unacceptably **high *p(type I)***.
- Negativley correlated residuals would lead to an unacceptably **low *p(type I)***.
- If the residuals are correlated, a different statistical procedure must be used.

Welch's *F* Test—An alternative to the one-way ANOVA

- ❑ Use when the sample sizes are unequal.
- ❑ Use when the sample sizes are equal but small.

$$w_j = \frac{n_j}{S_j^2}$$

$$u = \sum w_j$$

$$\bar{Y} = \frac{\sum w_j \bar{Y}_j}{u}$$

$$A = \frac{1}{a-1} \sum w_j \left(\bar{Y}_j - \bar{Y}\right)^2$$

$$B = 1 + \left[\frac{2(a-2)}{a^2 - 1}\right] \sum \frac{\left[1 - \left(\frac{w_j}{u}\right)\right]^2}{n_j - 1}$$

$$WelchF = \frac{A}{B}$$

$$df_{Numerator} = a - 1$$

$$df_{Denominator} = \left[\left(\frac{3}{a^2 - 1}\right) \sum \frac{\left[1 - \left(\frac{w_j}{u}\right)\right]^2}{n_j - 1}\right]^{-1}$$

Using SPSS to Analyze Data From the *a* Group Design: Welch *F* Test

SPSS Instructions
→Analyze→Compare Means→One-Way ANOVA
Dependent List: GAZE
Factor: PICTURE
→Options
√ Descriptives
√ Welch

Oneway

Descriptives

GAZE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
newspaper	5	7.0000	1.58114	.70711	5.0368	8.9632	5.00	9.00
white screen	3	5.0000	1.00000	.57735	2.5159	7.4841	4.00	6.00
face	4	9.7500	1.70783	.85391	7.0325	12.4675	8.00	12.00
Total	12	7.4167	2.35327	.67933	5.9215	8.9119	4.00	12.00

ANOVA

GAZE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	40.167	2	20.083	8.711	.008
Within Groups	20.750	9	2.306		
Total	60.917	11			

Robust Tests of Equality of Means

GAZE

	Statistic[a]	df1	df2	Sig.
Welch	9.676	2	5.760	.014

a. Asymptotically F distributed.

Power

Power: The probability of rejecting the Null Hypothesis given the Null Hypothesis is False.

RECOMMENDED! **A priori Sample Size Calculations.** The information required to determine the recommended sample size for a specific power level includes:

1. The desired power. Common desired power levels are .80, .95, and .99. Higher desired power levels require larger sample sizes.
2. The number of groups. More groups require larger sample sizes for a given power.
3. The Type I error rate. Higher type I error rates require smaller sample sizes for a given power.
4. An estimate of the effect-size for the population differences among the groups. Smaller group differences require larger sample sizes for a given power.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

The effect size used by Cohen for the one-way ANOVA is the same effect size estimate we discussed above.

$$f = \frac{S_{means}}{\sqrt{MSE}} \text{ where } S_{means} = \sqrt{\frac{\sum (Y_j - Y)^2}{a}} = \sqrt{\frac{\sum \hat{\alpha}_j^2}{a}}$$

Sample Size Needed for .80 Power

	Alpha = .01			**Alpha = .05**		
ANOVA	**Sm**	**Med**	**Lg**	**Sm**	**Med**	**Lg**
2g	586	95	38	393	64	26
3g	464	76	30	322	52	21
4g	388	63	25	274	45	18
5g	336	55	22	240	39	16
6g	299	49	20	215	35	14
7g	271	44	18	195	32	13

Maxwell and Delaney (2004) present a similar table to Cohen, but they use the maximum difference between two means as the estimate of the effect size....less accurate than Cohen (1992).