

General Linear Model → Equivalence of ANOVA & Regression

2-Group ANOVA oneway reptgood by reptdept (1,2).

Descriptives

rating of reptile quality - 1-10 scale

| | N | Mean | Std. Deviation |
|---------------|----|------|----------------|
| not separate | 6 | 4.00 | 1.90 |
| separate dept | 6 | 7.33 | 1.86 |
| Total | 12 | 5.67 | 2.50 |

ANOVA

rating of reptile quality - 1-10 scale

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|-------|------|
| Between Groups | 33.333 | 1 | 33.333 | 9.434 | .012 |
| Within Groups | 35.333 | 10 | 3.533 | | |
| Total | 68.667 | 11 | | | |

Regression Analysis with a Unit-coded (1-2) Binary Variable reg dep = reptgood / enter reptdept.

Model Summary

| Model | R | R Square |
|-------|-------------------|----------|
| 1 | .697 ^a | .485 |

a. Predictors: (Constant), type or reptile department

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|-------|-------------------|
| 1 | Regression | 33.333 | 1 | 33.333 | 9.434 | .012 ^a |
| | Residual | 35.333 | 10 | 3.533 | | |
| | Total | 68.667 | 11 | | | |

a. Predictors: (Constant), RDEPTEFF

b. Dependent Variable: rating of reptile quality - 1-10 scale

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|----------------------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .667 | 1.716 | | .389 | .706 |
| | type or reptile department | 3.333 | 1.085 | .697 | 3.071 | .012 |

a. Dependent Variable: rating of reptile quality - 1-10 scale

Since REPTDEPT is binary, we know that we can treat it as either a quantitative or a qualitative variable, in doing the ANOVA, we treated it as qualitative. If we treat it as quantitative, we could then perform a regression. Like this...

Interpreting regression weights from unit-coded binary predictors:

- the F and p-value are the same from the ANOVA and regression (as are the SS, MS and df)!!
- if we compute eta-squared from the ANOVA, we get $SS_{BG} / SS_{Tot} = 33.33 / 68.67 = .4854$, the R^2 !!
- **the constant** tells the expected value of y when x = 0
 - for unit coding (1-2), **a** does not have a meaningful interpretation (no one has an x value of 0)
- **the raw score regression weight** tells the expected change in y for a 1-unit increase in x.
 - for unit-coding (1-2), **b** tells the direction and size of the mean difference between the groups
 - +b → higher coded group (=2) has higher mean -b → lower coded group (=1) has higher mean
 - 3.33 is the difference between the mean = 4.0 for the "not separate departments" (coded 1) and mean = 7.33 for the "separate departments" (coded 2)
 - the **t-test & p-value** tell us that these two means are significantly different
 - the β (standardized weight) is the same for the original and effect-coded analyses -- as it should be -- but is difficult to interpret as "1-standard unit" doesn't apply well to the coded values

Keep in mind -- the point here is not that you should use regression instead of ANOVA. This is just the first in learning how to incorporate qualitative variables into regression analyses, so that you can use multiple regression as a truly all-purpose analytic tool!!

Regression Analysis of a Binary Variable Using Effect Coding

When using ANOVA to compare group means, we do not directly compare those group means to each other. Rather, each of the group means is compared to the grand mean or the overall mean of all the groups. Effects coding uses this same comparison procedure → the mean of the selected “target group” is compared to the grand mean. With only 2 groups, if the target group mean is different from the grand mean, then it must also be different from the mean of the other “comparison” group.

Effects Coding

- one group is selected at the “target group” and is coded “1”.
- the other group is the “comparison group”, and is coded “-1”

Use **File → New → Syntax to open a Syntax file & use the following code**

```
if (reptdept = 1) rdepteff = -1.  ← the "not separate reptile department" group is coded to "-1"
if (reptdept = 2) rdepteff = 1.  ← the "separate reptile department" group is coded to "1"
reg dep = reptgood / enter rdepteff.
```

Model Summary

| Model | R | R Square |
|-------|-------------------|----------|
| 1 | .697 ^a | .485 |

a. Predictors: (Constant), RDEPTEFF

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|-------|-------------------|
| 1 | Regression | 33.333 | 1 | 33.333 | 9.434 | .012 ^a |
| | Residual | 35.333 | 10 | 3.533 | | |
| | Total | 68.667 | 11 | | | |

a. Predictors: (Constant), RDEPTEFF

b. Dependent Variable: rating of reptile quality - 1-10 scale

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 5.667 | .543 | | 10.44 | .000 |
| | RDEPTEFF | 1.667 | .543 | .697 | 3.071 | .012 |

a. Dependent Variable: rating of reptile quality - 1-10 scale

Interpreting regression weights from effect-coded binary predictors:

- the F and p-value are the same as from the ANOVA and the regression with the unit-coded variable (as are the SS, MS and df)!!
- if we compute eta-squared from the ANOVA, we get $33.3 / 68.67 = .4854$, the R^2 !!
- **the constant** tells the expected value of y when $x = 0$
 - for effect coding (-1, 1) with equal-n, **a** tells the grand mean of the all the groups
 - 5.667 is the grand mean (check the ANOVA results above)
- **the raw score regression weight** tells the expected change in y for a 1-unit increase in x
 - for effect coding (-1, 1) with equal n, **b** tells the difference between the grand mean and the mean of the target group coded 1
 - $+b$ → target group has mean higher than grand mean $-b$ → target group has mean lower than the grand mean
 - 1.667 is the difference between the grand mean of 5.667 and the mean = 7.333 for the "separate departments" target group that was coded 1
 - the **t-test & p-value** tell us the mean of the "separate departments" target group coded 1 is significantly different from the grand mean (and so must be significantly different from the mean of the "not separate departments" comparison group coded -1
 - the β (standardized weight) is the same for the original and effect-coded analyses -- as it should be – but is difficult to interpret as “1-standard unit” doesn’t apply well to the coded values

Regression Analysis of a Binary Variable Using Dummy Coding

When using dummy coding, we directly compare group means to each other. Another advantage of using dummy coding is that the constant and the regression weight have meaningful interpretations. In fact, as we will see later, using dummy coding in complex multiple regression models increases the interpretability of the entire model!

Dummy Coding

- one group is the "target group," and is coded "1"
- one group is the "comparison group," and is coded "0"

For this analysis, I've chosen to make the "with separate reptile department" the target group and the "with no separate reptile department" the comparison group.

```
if (reptdept = 1) rdeptdum = 0.  ← the "not separate reptile department" group is coded to "0"
if (reptdept = 2) rdeptdum = 1.  ← the "separate reptile department" group is coded to "1"
reg dep = reptgood / enter = rdeptdum.
```

Model Summary

| Model | R | R Square |
|-------|-------------------|----------|
| 1 | .697 ^a | .485 |

a. Predictors: (Constant), RDEPTDUM

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|-------|-------------------|
| 1 | Regression | 33.333 | 1 | 33.333 | 9.434 | .012 ^a |
| | Residual | 35.333 | 10 | 3.533 | | |
| | Total | 68.667 | 11 | | | |

a. Predictors: (Constant), RDEPTDUM

b. Dependent Variable: rating of reptile quality - 1-10 scale

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 4.000 | .767 | | 5.212 | .000 |
| | RDEPTDUM | 3.333 | 1.085 | .697 | 3.071 | .012 |

a. Dependent Variable: rating of reptile quality - 1-10 scale

Interpreting weights from dummy coded binary predictors:

- the F and p-value are the same as from the ANOVA earlier regressions with the unit-coded variable, and with the effect-coded variable (as are the SS, MS and df)!!
- computing eta-squared from the ANOVA, we get $33.3 / 68.67 = .4854$, the R²!!
- **the constant** tells the expected value of y when x = 0
 - for dummy coding (0,1), **a** tells the mean of the "comparison" group that was coded 0
 - 4.00 is the mean of the comparison "not separate" department, that was coded 0
- **the raw score regression weight** tells the expected change in y for a 1-unit increase in x
 - for dummy coding (0, 1). **b** tells the direction and size of the mean difference between the groups
 - +b → higher coded group has higher mean -b → lower coded group has higher mean
 - +b → target group (coded 1) has higher mean -b → comparison group (coded 0) has higher mean
 - 3.33 is the difference between the mean = 4.0 of the "not separate departments" comparison group coded 0 and mean = 7.33 of the "separate departments" target group coded 1
- the **t-test & p-value** tell us the mean of the target group coded 1 is significantly different from the mean of the other comparison group coded 0
- the **β** (standardized weight) is the same for the original and effect-coded analyses -- as it should be -- but is difficult to interpret as "1-standard unit" doesn't apply well to the coded values

Just to be sure: We could "reverse" the dummy coding, with "not separate" = 1 and "separate" = 0. If we did...

- **a** would be the mean of the "separate" comparison group → 7.33
- **b** would be the mean difference between the groups → -3.33 (notice that the sign reverses)

Regression Analysis when the Qualitative Variable has Multiple Categories

So, ANOVA and Regression produce the same results when the qualitative variable has two values (binary)

But, what about when the qualitative variable has multiple categories? We can't just use the variable "as is"; just because we have the conditions of the qualitative variable as "1", "2" & "3", doesn't mean that they are "equally spaced" (interval data; consider -> 1 = no feedback, 2 = immediate feedback 3 = delayed feedback ?)!!

We will again use Dummy and Effects coding, but with a change -- we will use multiple codes to represent the qualitative variable. How many? The number of conditions minus 1 (k-1). Said differently, we will use one code for each degrees of freedom.

While this will allow us to represent the k conditions in the analysis it will not provide us with "all the information" in the k conditions. More on this later!

Multiple Group ANOVA

As an example, let's look at the mean "goodness" of the fish in different types of pet stores. First the ANOVA.

| | N | Mean | Std. Deviation |
|-----------------|----|------|----------------|
| chain store | 4 | 8.50 | .577 |
| privately owned | 4 | 7.25 | 1.258 |
| coop | 4 | 4.25 | 1.500 |
| Total | 12 | 6.67 | 2.146 |

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|--------|------|
| Between Groups | 38.167 | 2 | 19.083 | 13.740 | .002 |
| Within Groups | 12.500 | 9 | 1.389 | | |
| Total | 50.667 | 11 | | | |

Regression Analysis of Multiple-Category IV using Effects Codes

Effect Coding

- we need k-1 Effect Codes for the Store variable ($3 - 1 = 2$)
- select one of the conditions of Store as the "comparison condition"
 - that condition will receive a code of "-1" for each effect code
 - I've chose "private" as the comparison condition
- the other conditions of Store will receive a code of "1" on one effect code and of "0" on all others

```
* first effect code for Store (Chain = 1, Private (comparison) = -1, Co-op = 0).  
if (store = 1) storec1 = 1.  
if (store = 2) storec1 = -1.  
if (store = 3) storec1 = 0.
```

```
* second effect code for Store (Chain = 0, Private (comparison) = -1, Co-op = 1).  
if (store = 1) storec2 = 0.  
if (store = 2) storec2 = -1.  
if (store = 3) storec2 = 1.
```

Model Summary

| Model | R | R Square |
|-------|-------------------|----------|
| 1 | .868 ^a | .753 |

a. Predictors: (Constant), storec2, storec1

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 38.167 | 2 | 19.083 | 13.740 | .002 ^a |
| | Residual | 12.500 | 9 | 1.389 | | |
| | Total | 50.667 | 11 | | | |

a. Predictors: (Constant), storec2, storec1
b. Dependent Variable: rating of fish quality - 1-10 scale

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 6.667 | .340 | | 19.596 | .000 |
| | storec1 | 1.833 | .481 | .728 | 3.811 | .004 |
| | storec2 | -2.417 | .481 | -.960 | -5.023 | .001 |

a. Dependent Variable: rating of fish quality - 1-10 scale

Interpreting regression weights from effect-coded binary predictors:

- the F and p-value are the same as from the ANOVA and the regression with the unit-coded variable (as are the SS, MS and df)!!
- if we compute eta-squared from the ANOVA, we get $36.45 / 50.667 = .719$, the R^2 !!
- the constant** tells the expected value of y when $x = 0$
 - for effect coding with equal-n, **a** tells the grand mean of the all the groups
 - 6.667 is the grand mean (check the ANOVA results above)
- the raw score regression weight** tells the expected change in y for a 1-unit increase in x
 - for effect coding (-1, 1) with equal n, **b** tells the difference between the grand mean and the mean of the target group coded 1 for that effect code
 - +b → target group has mean higher than grand mean -b → target group has mean lower than the grand mean
 - for storec1, 1.833 is the difference between the grand mean of 6.667 and the higher mean = 8.50 for the "chain store" target group that was coded 1
 - for storec2, -2.417 is the difference between the grand mean of 6.667 and the lower mean = 4.25 for the "co-op" target group that was coded 1
- the **t-test & p-value** for each effect code regression weight tells us if the mean of the target group coded 1 for that code is significantly different from the grand mean – while we can deduce pairwise differences, we do not get direct tests of them
- the β (standardized weight) is difficult to interpret as “1-standard unit” doesn’t apply well to the coded values

Regression Analysis of Multiple-Category IV using Dummy Codes

Dummy Coding

- we need $k-1$ Dummy Codes ($df, 3 - 1 = 2$) for the Store variable
- select one of the conditions of Store as the "comparison condition"
 - that condition will receive a code of "0" for each dummy code
 - I've chose "private" as the comparison condition
- the other conditions of Store will receive a code of "1" on one dummy code and of "0" on all others

```
* first dummy code for Store (Chain = 1, Private (comparison) = 0, Co-op = 0).
if (store = 1) storedc1 = 1.
if (store = 2) storedc1 = 0.
if (store = 3) storedc1 = 0.

* second dummy code for Store (Chain = 0, Private (comparison) = 0, Co-op = 1).
if (store = 1) storedc2 = 0.
if (store = 2) storedc2 = 0.
if (store = 3) storedc2 = 1.
```

Model Summary

| Model | R | R Square |
|-------|-------------------|----------|
| 1 | .868 ^a | .753 |

a. Predictors: (Constant), storedc2, storedc1

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 38.167 | 2 | 19.083 | 13.740 | .002 ^a |
| | Residual | 12.500 | 9 | 1.389 | | |
| | Total | 50.667 | 11 | | | |

a. Predictors: (Constant), storedc2, storedc1

b. Dependent Variable: rating of fish quality - 1-10 scale

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 7.250 | .589 | | 12.304 | .000 |
| | storedc1 | 1.250 | .833 | .287 | 1.500 | .168 |
| | storedc2 | -3.000 | .833 | -.688 | -3.600 | .006 |

a. Dependent Variable: rating of fish quality - 1-10 scale

Interpreting weights from dummy coded binary predictors:

- the F and p-value are the same as from the ANOVA and the regression with the unit-coded variable (as are the SS, MS and df)!!
- if we compute eta-squared from the ANOVA, we get $36.45 / 50.667 = .719$, the R²!!
- **the constant** tells the expected value of y when x = 0
 - for dummy coding (0,1), **a** tells the mean of the "comparison" group that was coded 0 for all the dummy codes
 - 7.25 is the mean of the "Private" comparison group, that was coded 0 for all the dummy codes
- **the raw score regression weight** tells the expected change in y for a 1-unit increase in x
 - for dummy coding (0, 1). **b** tells the direction and size of the mean difference between the comparison group coded 0 and the target group for that dummy code that was coded 1
 - +b → higher coded group has higher mean -b → lower coded group has higher mean
 - +b → target group (coded 1) has higher mean -b → comparison group (coded 0) has higher mean
 - for storedc1, 1.25 is the difference between the mean = 7.25 of the "Private" comparison group coded 0 and mean = 8.50 of the "Chain" target group coded 1
 - for storedc2, -3.00 is the difference between the mean = 7.25 of the "Private" comparison group coded 0 and mean = 4.25 of the "Co-op" target group coded 1
- the **t-test & p-value** tell us the mean of the target group for that dummy code that is coded 1 is significantly different from the mean of the other comparison group coded 0
 - for storedc1, the non-significant result tells us that the means of the "Private" and "Chain" groups are not significantly different
 - for storedc2, the significant result tells us that the means of the "Private" and "Co-op" groups are significantly different
- the βs (standardized weight) are the same for the effect-coded analyses -- as it should be – but is difficult to interpret as "1-standard unit" doesn't apply well to the coded values

But what about the pairwise comparison of Chain and Co-op, you ask??? One disadvantage of only doing df or k-1 of the pairwise comparisons, is that there's lots of information we don't get. And the proportion of information we don't get increases quickly as we add conditions! 2 of 3 pairwise comparisons for 3 groups, 3 of 6 for 4 groups and 4 of 10 for 5 groups, for starters. What to do?

With 3 groups, one strategy is to pick the group with the middle-most mean as the comparison group. Then the omnibus model test tells you if the 2 most-different groups are significantly different, and the 2 dummy codes provide the pairwise comparisons of those groups to the middle-most group.

Another approach, which will be necessary with more than 3 groups, is to simply run multiple analyses with different groups as the comparison group in each analysis. For example, in the analysis above, we didn't get the pairwise comparison of Chain & Co-op. We will get that pairwise comparison from the following dummy codes. We will also duplicate the pairwise comparison of the Chain and Private scores.

```
* first dummy code for Store (Chain (comparison) = 0, Private = 1, Co-op = 0).
if (store = 1) storedc1 = 0.
if (store = 2) storedc1 = 1.
if (store = 3) storedc1 = 0.

* second dummy code for Store (Chain (comparison) = 0, Private = 0, Co-op = 1).
if (store = 1) storedc2 = 0.
if (store = 2) storedc2 = 0.
if (store = 3) storedc2 = 1.
```

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 8.500 | .589 | | 14.425 | .000 |
| | storedc1 | -1.250 | .833 | -.287 | -1.500 | .168 |
| | storedc2 | -4.250 | .833 | -.975 | -5.100 | .001 |

a. Dependent Variable: rating of fish quality - 1-10 scale

Interpreting these Dummy Codes

- constant of 8.5, tells the mean of the comparison "Chain" group (coded 0)
- for storedc1, -1.25 is the difference between mean = 8.50 of the "Chain" comparison group coded 0 the mean = 7.25 of the "Private" target group coded 1 – this difference is not significant
- for storedc2, -4.25 is the difference between the mean = 8.50 of the "Chain" comparison group coded 0 and mean = 4.25 of the "Co-op" target group coded 1 – this difference is significant

So from these two runs we got all the pairwise comparisons:

Chain (8.50) = Private (7.25) Chain (8.50) > Co-op (4.25) Private (7.25) > Co-op (4.25)

Obviously you need to be careful when running alternative dummy codes in different analyses. It is a good idea to develop a naming system that will help keep you from mixing things up. Here is an example for these two sets of dummy codes.

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>For the first set of dummy codes:</p> <p>dc_priv0_chn1 & dc_priv0_coop1</p> <p>you know these go together because "priv = 0" for both</p> | <p>For the second set of dummy codes:</p> <p>dc_chn0_priv1 & dc_chn0_coop1</p> <p>you know these go together because "chn = 0" for both</p> |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|