Easy methods for extracting individual regression slopes: Comparing SPSS, R, and Excel

Roland Pfister, Katharina Schwarz, Robyn Carson, Markus Jancyzk

Three different methods for extracting coefficients of linear regression analyses are presented. The focus is on automatic and easy-to-use approaches for common statistical packages: SPSS, R, and MS Excel / Libre Office Calc. Hands-on examples are included for each analysis, followed by a brief description of how a subsequent regression coefficient analysis is performed.

An increasingly popular analysis of within-subjects designs revolves around regression coefficients that are estimated individually for each participant. More precisely, a dependent variable (criterion) is regressed on an independent variable (predictor) individually for each participant. The extracted values for slopes and intercept are then compared between conditions or tested against a population value of 0 via standard significance tests such as paired-samples t-tests or repeated-measures analyses of variance (ANOVA). This procedure is commonly known as regression coefficient analysis (RCA; Lorch & Myers, 1990, Method 3).

RCA circumvents methodological problems of standard regression analysis which assumes different observations to be independent from each other. This assumption is routinely violated by data from within-subjects designs, but it does not apply to the coefficients that were extracted from individual data sets (cf. Lorch & Myers, 1990). In contrast, RCA only assumes a linear relationship between predictor and criterion for each individual participant and can be used for both, continuous and dichotomous predictors (Ahn, Jung, & Kang, 2002; Lorch & Myers, 1990; Myers & Broyles, 2000). RCA thus offers a flexible alternative to more common ANOVA approaches, and it has been applied to numerous different topics, covering as diverse areas as reading, emotion, cognitive control, and numerical cognition (see Table 1 for a more detailed overview).

In light of RCA’s widespread use, the present paper attempts to give a brief overview of how to extract individual regression slopes efficiently with commonly used computer programs. To this end, we compare available methods for extracting regression slopes for three widespread statistical packages: SPSS 19, R 2.15 and MS Excel 2010 / LibreOffice 3.6 Calc. As a hands-on example, we demonstrate how regression slopes can be extracted to probe for spatial-numerical associations in a parity judgment task (Dehaene, Bossini, & Giraux, 1993; cf. Fias, Brysbaert, Geyens, & d’Ydewalle, 1996, Pfister, Schröder, & Kunde, in press).

A major drawback of RCA is that this analysis does not yield any measures of variance-accounted-for (such as R²). Such measures can be obtained from alternative approaches such as linear trend analyses for repeated-measures ANOVA on the one hand (Pinhas, Tzelgov, & Ganor-Stern, 2012) as well as hierarchical linear modeling or mixed modeling on the other hand (Baayen, Davidson, & Bates, 2008; Hoffman, & Rovine, 2007; Quené & van den Bergh, 2008; Van den Noortgate & Onghena, 2006).

Adress correspondence to Dipl.-Psych. Roland Pfister, Julius-Maximilians University of Würzburg, Department of Psychology III, Röntgenring 11, 97070 Würzburg, Germany, Tel +49-931-31-81363, Email: roland.pfister@psychologie.uni-wuerzburg.de.
Extracting regression slopes: Available methods

In a simulation of the mentioned task, we assume four participants to have judged the parity of the numbers 0-9 in a series of trials. Furthermore, participants are assumed to have used two mappings in the experiment: “even” responses were mapped to a left response key during one half of the experiment and to a right response key during the other half. Response times (RTs) can thus be analyzed as a function of target number and response side.

Similar experiments have consistently shown smaller numbers to facilitate left responses and larger numbers to facilitate right responses: the SNARC effect (spatial-numerical association of response codes; Dehaene et al., 1993; for an overview see Wood, Willmes, Nuerk, & Fischer, 2008). This finding is typically quantified as a negative slope when the SNARC effect (RT_{right}-RT_{left}) is regressed on the target number (Fias et al., 1996). The corresponding relations of target number and SNARC effect for the four simulated participants are plotted in Figure 1. These data were generated by adding normally distributed noise (μ = 0; σ = 25) to the mean SNARC effects reported by Dehaene et al. (1993, Exp. 1). Data and scripts for the following slope extractions are available as supplementary material.

SPSS: Slope extraction via OMS

An efficient way to extract regression slopes with SPSS involves two separate steps (Figure 2). Individual regression analyses are first run for each participant and each condition of interest. The resulting coefficient tables are then automatically read from the output via the Output Management System (OMS).† The two steps are described in detail below.

For the following example of the SNARC effect, we assume the data to consist of three variables (see the supplementary material for the corresponding data set). The variable Subject contains the participant number (1-4), the variable Number codes the target digit (0-9), and the variable SNARC codes the corresponding SNARC effect as the mean difference RT_{right}-RT_{left} in milliseconds. The data thus contains ten cases (rows) per participant, each listing the SNARC effect for one of the target digits.

The two steps described above can then be defined in the syntax editor (see the supplementary material for the entire syntax file). To prepare the individual regression analyses, the data is first split according to the variable Subject using the menu Data > Split File… and the corresponding option Compare groups. Furthermore, a new data set is declared to prepare the call to OMS.

* Analyze each participant separately and initialize
* a new data set for regression output.

SORT CASES BY Subject.
SPLIT FILE BY Subject.

† Alternative ways to obtain regression coefficients are described by Thompson (2008) and by Weaver and Dubois (2012). However, we believe the present solution to offer a more intuitive way because it does not require the user to have advanced background knowledge about SPSS macro facilities and/or training in matrix algebra as previous methods did.
DATASET DECLARE SNARC_Slopes.

Before the actual regression analysis is performed (via Analyze > Regression > Linear...), we need to set up a call to OMS. This can be done either via the menu Utilities > OMS Control Panel... or by writing the syntax directly (which is preferred in this case). Basically, OMS accesses the output of upcoming function calls and can distill any information from the output (such as figures and tables). For the present purposes, we are looking for a specific table that is produced by the analysis command REGRESSION. Furthermore, we want to write the content of this table to the newly defined data set SNARC_Slopes (see above).

* Define the request to the Output Management System (OMS).

OMS
/SELECT TABLES
/IF COMMANDS=['Regression'] SUBTYPES=['Coefficients']
/DESTINATION FORMAT=SAV
OUTFILE = SNARC_Slopes.

Once this call is initialized, OMS starts collecting information from all upcoming regression commands until the call is stopped again by the user. Thus, we can now specify the regression analysis and stop OMS afterward.

* The SNARC effect is regressed on the Number with regression (as in Analyze > Regression > Linear...).

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)

Figure 1. Example data of four participants who are assumed to have completed parity judgment task. SNARC effects (RT_right-RT_left) as a function of the corresponding target number clearly show a negative slope.
Figure 2. Slope extraction with SPSS 19. Individual regression analyses are first run for each participant and each condition of interest. The resulting coefficient tables are then automatically read from the output via the Output Management System (OMS).

/NOORIGIN
/DEPENDENT SNARC
/METHOD=ENTER Number.

* The OMS request is stopped; the SPLIT command * is disabled.

OMSEND.
SPLIT FILE OFF.

Running the code from OMS to OMSEND will now populate the newly defined data set SNARC_Slopes with the data from the coefficient tables of the regression analysis. Among others, the data includes the intercepts and slope coefficients for each participant. The new data set can be cleaned up easily to store only subject numbers and corresponding regression slopes (see the supplementary material). Other efficient ways to constrain the OMS output are explained in the SPSS Command Syntax Reference guide (IBM, 2010).

R: Slope extraction using linear models

Individual regression slopes can be extracted with only a few lines of R code and the most straightforward solution uses the command `lm`. This command allows fitting a linear model to empirical data. The following demonstration is based on a data frame called `snarc_raw` (see the supplementary material). This data frame consists of three variables: The variable Subject codes the participant number (1-4), the variable Number codes the target number (0-9), and the variable SNARC codes the corresponding SNARC effect as the mean difference RT_right - RT_left in milliseconds.

As a basic procedure, we use the for command to loop through the data frame and to compute separate analyses for each participant. For the example data set, the loop index \( i \) can thus be defined as a counter from 1 to 4 (a more flexible definition of the loop index is described in the supplementary material). Furthermore, we define the empty variable `snarc_coefs` to store the to-be-extracted regression slopes:

```r
snarc_coefs = c(NA, NA, NA, NA)
for (i in c(1:4)) {
  # [...] run participant-wise analyses
}
```

Inside the loop, we perform four separate steps. First, the
relevant observations of the data frame `snarc_raw` is copied to a temporary data frame `snarc_tmp`. Then, the command `lm` is used to perform the linear regression of the SNARC effect on the corresponding number; the results of this analysis are saved as the linear model `reg_result`. The function `coef` is used subsequently to access the linear model. For a simple linear regression, the `coef` function returns a vector of two elements: the first element corresponds to the intercept, the second element is the slope coefficient. Finally, this coefficient is saved inside the summary vector `snarc_coefs`:

```r
# Create temporary data frame:
snarc_tmp <- snarc_raw[snarc_raw$Subject==i,]
# Perform regression:
reg_result <- lm(snarc_tmp$SNARC ~ snarc_tmp$Number)
# Get coefficient:
tmp_coef   <- coef(reg_result)
# Store coefficient:
snarc_coefs[i] <- tmp_coef[2]
```

If the above series of steps is implemented inside a suitable loop, extracted coefficients will be readily available via the vector `snarc_coefs`.

**Excel / Calc: The SLOPE function**

Extracting regression slopes is remarkably convenient with Excel 2010 and its open source counterpart Calc. For the simulated data, we can simply use the function `=SLOPE` to request the slope coefficient (similarly, `=INTERCEPT` returns the intercept). This function takes the criterion data (SNARC effect) as first argument and the predictor data (numbers) as second argument (Figure 3).

This function can then be applied to each individual participant by dragging down the formula using the Fill Handle tool. To fix the predictor data for this operation, we need to define an absolute range of x values using the `$` operator. In the example of Figure 3, the correct specification of the x range would be `C$3:L$3`.

**Concluding remarks**

The preceding description of different methods for slope extraction can be summarized as follows. In SPSS, extracting regression coefficients involves two separate steps during which separate regression analyses are carried out and the corresponding output is fed back to a new data set using the OMS facilities. In R, we follow a similar procedure and loop through the data set while storing the output of a linear regression in each iteration. In Excel / Calc, regression slopes and intercepts for simple linear regressions are readily available via built-in functions.

Thus, Excel and Calc do seem to offer a very efficient way to perform participant-wise regression analyses (Lorch & Myers, 1990, Method 3) and we do indeed suggest that these programs offer an interesting alternative to standard statistical packages such as R or SPSS. This conclusion, however, only holds true for simple linear regressions using one predictor variable. More advanced setups (starting with multiple linear regression) are more difficult to handle with Excel or Calc, whereas the described methods for SPSS and R generalize easily and intuitively to these settings. In fact, the demonstrated algorithms for SPSS and R will work just as well for multiple regression and will simply enhance the output by data relating to the additional predictors.

Independent of the extraction method used, the obtained coefficients are then ready for subsequent RCA (Lorch &
Myers, 1990; for limitations of this approach and possible alternatives, see Van den Noortgate & Onghena, 2006). Such analyses may include rather simple comparisons such as testing whether the obtained coefficients differ significantly from zero via a one-sample t-test. This simple comparison would indeed be appropriate for the data presented in the above example of the SNARC effect; however, RCA can of course also involve any statistical test depending on the research question and the data obtained.

References


Thompson, G. L. (2008). Eliminating aggregation bias in experimental research: Random Coefficient Analysis as an alternative to performing a ‘by-subjects’ and/or ‘by-items’ ANOVA. *Tutorials in Quantitative Methods for Psychology, 4*(1), 21-34.


Manuscript received 19 December 2012

Manuscript accepted 26 February 2013