**Automatisierte Experiment-Auswertung mit generate\_excel\_eval.R**

generate\_excel\_eval ist ein in R geschriebenes Script, das Datenfiles von Unipark standardisiert zusammenfasst und die Auswertung in einer Excel-Datei speichert. R muss auf dem Rechner installiert sein.

Zu jeder Daten-Datei muss ein spezifisches Konfigurationsskript (nachfolgend config file genannt) existieren, welches alle notwendigen Informationen zum Aufbau des Datenfiles beinhaltet.

1. **Ausführung des Skripts**

Ausführung in der Windows Eingabeaufforderung oder im UNIX-Terminal mit config file

Im Terminal (UNIX, MacOS) muss generate\_excel\_eval.R zunächst ausführbar gemacht werden, danach wird das Skript mit angehängtem config file aufgerufen:

*chmod +x generate\_excel\_eval.R*

*./generate\_excel\_eval.R my\_directory/config\_file.R*

In der Windows Eingabeaufforderung lautet der Befehl zum Starten des Skripts

*Rscript generate\_excel\_eval.R my\_directory/config\_file.R*

Beachte: Windows Pfade werden in R mit einem „/“ getrennt, nicht wie üblich mit einem „\“

Ausführung ohne direkte Übergabe des config files

Wird generate\_excel\_eval ohne Parameter gestartet, öffnet sich ein Dialogfenster zur Auswahl des config files. Dies erfordert allerdings die Installation des Pakets *radiant.data*

1. **Anpassung des config files: Globale Variablen**

Im config file müssen zunächst einige globale Einstellungen und die zugehörigen Variablen gesetzt werden:

*name\_data\_file*: Zeichenkette

*name\_excel\_file*: Zeichenkette

*gr\_ID\_name*: Zeichenkette

*gr\_IDs*: numerischer Vektor

*excel\_col\_names*: Liste von Zeichenketten

*name\_data\_file* beinhaltet Pfad und Name der Datendatei. Aktuell werden Datendatein im .sav-Format und im .dta-Format unterstützt. Beispiel:

name\_data\_file <- "directory/my\_data.sav"

*name\_excel\_file* beinhaltet den Namen der zu erzeugenden Excel-Datei. Die Excel-Datei wird in demjenigen Verzeichnis gespeichert, in welchem das Datenfile liegt. Beispiel:

name\_excel\_file <- "my\_excel\_table.xlsx"

*gr\_ID\_name* ist der Name der Spalte, die die IDs der Experiementgruppen enthält. Beispiel:

gr\_ID\_name <- "expgroup\_id"

*gr\_IDs* ist der Vektor aller Gruppen-IDs. Besteht das Experiment aus N Kohorten, hat *gr\_IDs* die Länge N. Der Vektor wird mit dem R-Befel *c(…)* angelegt, für collection. Mit einem Doppelpunkt können Bereiche angegeben werden. Beispiele:

gr\_IDs <- c(1, 2, 3, 4, 11, 12, 13, 14)

gr\_IDs <- c(c(1:4), c(11:14)) #is the same as above

*excel\_col\_names* gibt alle Namen der Gruppen als Liste an. Diese werden in die Titelzeile der Excel-Tabelle geschrieben. Die Liste muss aus N Zeichenketten bestehen, wenn N die Zahl der Kohorten ist. Beispiel:

excel\_col\_names <- list("Treatment 1", "Treatment 2", "Treatment 3")

1. **Anpassung des config files: Zusätzliche Variablen definieren mit *append\_vars***
   1. **Gruppenspezifische Variablen zusammenführen mit *merge\_exp\_columns***

Syntax: *merge\_exp\_colums(varname, col\_name\_list)*

*varname*: Zeichenkette

*col\_name\_list*: geschachtelte Liste von Zeichenketten

*varname* gibt den Spaltennamen der neuen Variable an, in der die einzelnen Gruppenvariablen zusammengeführt werden. Leer- und Sonderzeichen sind nicht erlaubt.

*col\_name\_list* gibt die Struktur an, wie gruppenspezifische Spalten bzw. Variablen zusammengeführt werden. Die Liste muss so viele Einträge haben wie es Gruppen gibt; die Zuordnung wird durch *gr\_IDs* festgelegt. Einzelne Gruppenvariablen können mit dem Symbol *NULL* ausgeschlossen werden. Besteht eine Experiementgruppe aus mehreren Untergruppen, können diese in einer geschachtelten Liste addressiert werden.

Beispiel:

Angenommen es gibt vier Experimentgruppen, in der jeweils eine Partei zu wählen ist. Die Variable *v\_54* beinhaltet die Wahlentscheidung für die erste Gruppe, *v\_78* für die zweite Gruppe. Gruppe 3 soll ignoriert werden und in Gruppe 4 erfolgte die Wahl in zwei disjunkten Variablen *v\_76* und *v\_84*. Dann können alle Abstimmungen in einer Variablen *parteiwahl* zusammengeführt werden:

merge\_exp\_columns("parteiwahl", list("v\_54", "v\_78", NULL,

list("v\_76", "v\_84")))

* 1. **Subgruppen-Variablen erstellen mit *add\_subgroup***

Syntax: *add\_subgroup(varname = logical expression, varname2 = expr2, …)*

*varname*: Zeichenkette

*logical expression*: Term, der Wahrheitswerte TRUE / FALSE ausgibt

*…*: weitere Parameterpaare, um mehrere Subgruppen anzulegen

*varname* gibt den Spaltennamen der neu anzulegenden Variable an, die den Datensatz in zwei disjunkte Subgruppen trennt. Leer- und Sonderzeichen sind nicht erlaubt.

*logical\_expression* bestimmt die Wahrheitswerte der Subgruppenvariable. Es können mehrere Terme durch logische Operatoren verbunden werden. Hierbei bezeichnet “|” das logische Oder, “&” das logische Und. Gleichheit wird mit “==” überprüft.

Beispiel:

Angenommen in einem Experiment werden auf einer Skala Zustimmungswerte für alle großen Parteien abgefragt. Variable *v\_101* speichert die Zustimmung zur SPD, *v\_102* die Zustimmung zu den Grünen. Der Datensatz soll dann auf den Subgruppen untersucht werden, wo (a) die SPD-Identifikation sehr hoch ist, (b) wo die Grünen-Identifikation sehr hoch ist, und (c) wo Teilnehmer:innen eher links eingestellt sind:

add\_subgroup(SPD\_identifikation = (v\_101 >= 7),

Gruene\_identifikation = (v\_102 >= 7),

ist\_halblinks = (v\_101 >= 7) & (v\_102 >= 7))

1. **Anpassung des config files: Ausgewählte Variablen in die Excel-Tabelle schreiben mit *add\_data\_to\_main\_overview***
   1. **Variblenweises Hinzufügen mit *add\_to\_excel***

Syntax: *add\_to\_excel(subgroup\_var, varname, col\_name\_from\_src, FUNC)*

*subgroup\_var*: NULL, Zeichenkette oder Liste von Zeichenketten

*varname*: Zeichenkette

*col\_name\_from\_src*: Zeichenkette

*FUNC*: Funktion

*subgroup\_var* gibt an, ob der Datensatz entlang einer Subgruppe ausgewertet werden soll. Wird *NULL* angegeben, erfolgt keine Subgruppen-Auswertung. Wird nur eine Varibale als Zeichenkette angegeben, erfolgt die Auswertung entlang dieser Variable. Werden mehrere Variablen als Liste angegeben, wird *add\_to\_excel* iterativ für jede einzelne Subgruppe der Liste ausgeführt.

*varname* beinhaltet den lesbaren Namen der auszuwertenden Variable und wird für die linke Spalte in der Excel-Tabelle benutzt

*col\_name\_from\_src* beinhaltet die auszuwertende Variable in Form des Spaltennamens

*FUNC* gibt eine R-Funktion an, mit der *col\_name\_from\_src* gruppenspezfisch ausgewertet werden soll (beispielsweise *min*, *max*, *mean*, *table*). Wird *mean* angegeben, wird automatisch eine ANOVA mit gerechnet, für *table* ein chi2-Test

Beispiel:

Bezogen auf die obigen Beispiele kann die gruppenspezfische Parteiwahl (Variable *parteiwahl*) ohne und mit Subgruppen-Unterteilung ausgewertet werden:

add\_to\_excel(list(NULL, "ist\_halblinks"),

"gewählte Partei",

"parteiwahl",

table)