


R101: A practical guide to making R your everyday statistical tool (PSY532)

# Seminar 2

Basic ANOVA and regression: repeated measures ANOVA and bootstrapping



# Programme

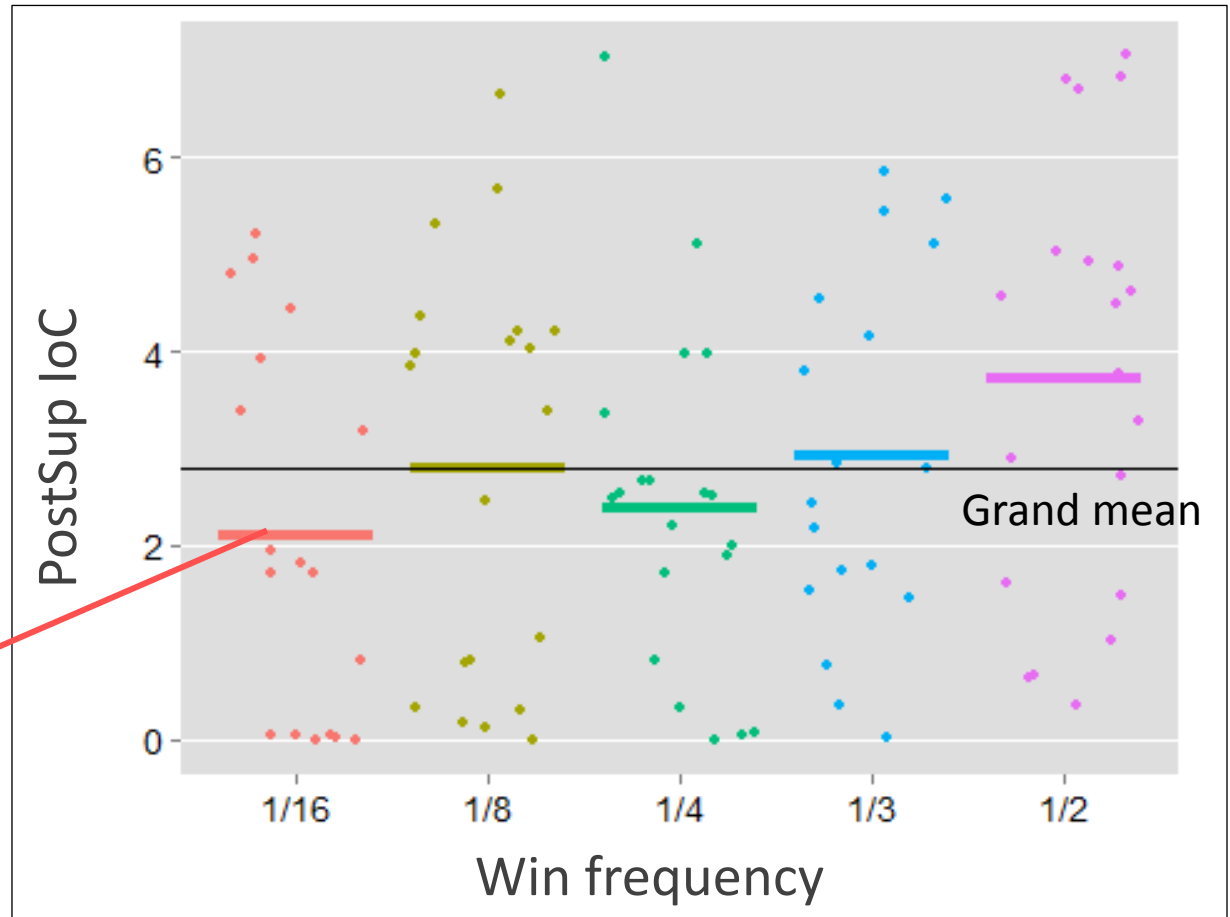
- T-tests
  - Linear regression
  - ANOVA
  - Repeated-measures ANOVA
- 
- Logic of the analysis
  - Hypotheses from our dataset
    - Regression: A hypothesis from a related but slightly different experiment
    - **ANOVA: As for regression + Hypothesis 2 from Lecture 1**
    - **Repeated-measures ANOVA: Hypothesis 1a from Lecture 1**
  - Working together in R:
    - Obtaining descriptive statistics
    - Running the analysis
    - Checking assumptions
  - Reporting the analysis
  - Seminar: Repeated-measures ANOVA; **bootstrapping**

Readings: LSR for everything except Repeated-Measures ANOVA

## Logic of the analysis: The F-statistic as a model comparison

- The F-test, as it is used in both ANOVA and regression, is really a comparison of two statistical models.
- In an ANOVA with one predictor, the F-test is a comparison of an intercept-only model (M0, null hypothesis) to a model involving the intercept *and* the predictor (M1, alternative hypothesis).

# Logic of the analysis – ANOVA as regression (illustration for ANOVA with one predictor)



A group mean selected by researcher (e.g., the lowest win-frequency condition).

When we use the aov function, the chosen group's mean is the intercept (baseline) in a “dummy coded” regression. In this case, the regression has four predictors (see next slide). Using the aov function additionally involves a model comparison (see script, ANOVA Example 1).

Win frequency data (first 6 cases) “dummy coded” with 1/16 as reference group

PNo	SuploC	1/8 ( $X_1$ )	1/4 ( $X_2$ )	1/3 ( $X_3$ )	1/2 ( $X_4$ )
2	0.8333	1	0	0	0
3	0.0000	0	0	0	0
4	2.5000	0	1	0	0
5	4.1667	0	0	1	0
6	0.6667	0	0	0	1
7	4.5000	0	0	0	0

The regression model:

$$Y_p = b_1 X_{1p} + b_2 X_{2p} + b_3 X_{3p} + b_4 X_{4p} + b_0 + \varepsilon_i$$

SuploC of participant p

Participant p's code on  $X_1$

Mean of 1/16 group

Difference between means of 1/2 group and 1/16 group

We determine the values of  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$  and  $b_4$  using the `summary.lm` function.

## Other possible contrasts in the regression component

- The dummy coding in the previous slide contained a treatment contrast.
- Other possible contrasts include Helmert, sum-to-zero (“effect coding”) and manually set orthogonal contrasts.

Win frequency data (first 6 cases) with Helmert contrast and 1/16 as reference group:

PNo0	1/8 ( $X_1$ )	1/4 ( $X_2$ )	1/3 ( $X_3$ )	1/2 ( $X_4$ )
2	1	-1	-1	-1
3	-1	-1	-1	-1
4	0	2	-1	-1
5	0	0	3	-1
6	0	0	0	4
7	-1	-1	-1	-1

This coding enables us to contrast the second level with the reference level, the third with the average of the first two, and so on.

Win frequency data (first 6 cases) with sum-to-zero contrast (“effect coding”) and 1/2 as reference group (as per script):

PNo	1/16 ( $X_1$ )	1/8 ( $X_2$ )	1/4 ( $X_3$ )	1/3 ( $X_4$ )
2	0	1	0	0
3	1	0	0	0
4	0	0	1	0
5	0	0	0	1
6	-1	-1	-1	-1
7	1	0	0	0

This coding enables us to contrast the mean of each group except the reference group with the grand mean. The grand mean is “weighted” (see script) if the groups are not equal in sample size.

The regression model:

$$Y_p = (1/5)b_1 X_{1p} + (1/5)b_2 X_{2p} + (1/5)b_3 X_{3p} + (1/5)b_4 X_{4p} + \frac{b_0}{5} + \varepsilon_i$$

Mean of 1/16 group minus  
weighted grand mean

Mean of 1/8 group minus  
weighted grand mean

(Weighted) grand  
mean

## Rules for manually setting orthogonal contrasts

Rules:

1. The weights within any contrast must sum to zero
2. The weights for any pair of contrasts must sum to zero when the dot product is taken.

Illustration:

- Contrast A = (a, b, c, d, e)
- Contrast B = (f, g, h, i, k)
- Contrast C = (l, m, n, o, p)

Each contrast should compare two sets of means (e.g., mean of a, b and d to the mean of c and e). Chunks with a negative weight (e.g., -1) are compared to chunks with a positive weight. So in this example, we would assign weights like this: (1, 1, -1, 1, -1) or (-1, -1, 1, -1, 1).

If the rules are met:

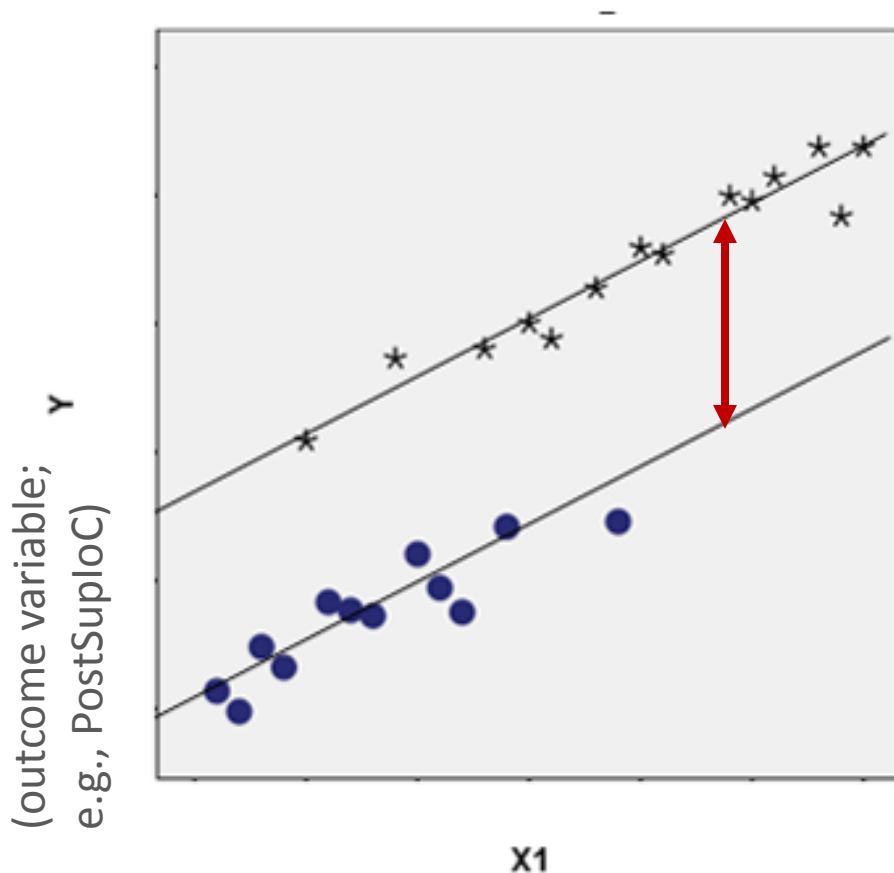
1.  $a + b + c + d + e = 0$ ,  $f + g + h + i + k = 0$ , and  $l + m + n + o + p = 0$
2.  $a*f + b*g + c*h + d*i + e*k = 0$ ,  $l*f + m*g + n*h + o*i + p*k = 0$ , and  $a*l + b*m + c*n + d*o + e*p = 0$

For a worked example, see ANOVA Example 3 in script.

Reading: Field chapter



# Logic of the analysis – ANCOVA (with one predictor and one covariate)



(covariate; e.g., beliefs in value of strategies even before the game – PreDBC\_Sup)

(categorical predictor, here with two levels; e.g., win-freq of 1/2 vs. 1/16)

- If the categorical predictor has more levels (e.g., 5 as in our example), there might be more parallel lines:
  - The vertical distance between lines represents the effect of the categorical predictor
- Two covariates could be visualised as parallel regression planes.
- Parallel slopes (lack of relationship between predictor and covariates) are assumed.
- Covariates can be categorical!

# Logic of the analysis – ANOVA with two or more predictors

From the same data that gives us this table,

		Factor B (2 levels)		total	
Group means – e.g., for group 1,1					
Factor A (3 levels)		$\bar{Y}_{1.1}$	$\bar{Y}_{1.2}$	$\bar{Y}_{1.}$	Row marginal means
		$\bar{Y}_{2.1}$	$\bar{Y}_{2.2}$	$\bar{Y}_{2.}$	
		$\bar{Y}_{3.1}$	$\bar{Y}_{3.2}$	$\bar{Y}_{3.}$	
	total	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{..}$	Grand mean
		Column marginal means			

we can calculate...

Total sum of squares expressing distance between all data points and grand mean

Sum of squares expressing difference between row marginal means and grand mean – variability due to Factor A

Sum of squares expressing difference between column marginal means and grand mean – variability due to Factor B

Sum of squares expressing the extent to which the group means *cannot* be predicted based on the marginal means alone – variability due to interaction between A and B (see next slide)

Four sets of degrees of freedom:

- For Factor A
- For Factor B
- For the interaction between A and B
- For the residuals

$$F_A = \frac{MS_A}{MS_R}$$

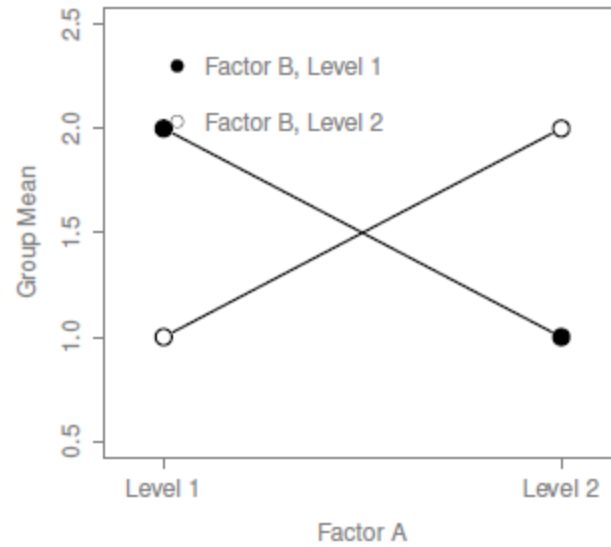
Using the first four quantities, we can calculate the residual sum of squares  
 $SS_R = SS_T - (SS_A + SS_B + SS_{A:B})$

This is all the information we need for computing the **F-value for each predictor and interaction term**. We can also compute an effect size (eta-squared) for each predictor/interaction – e.g., for Factor A:

$$\eta_A^2 = \frac{SS_A}{SS_T}$$

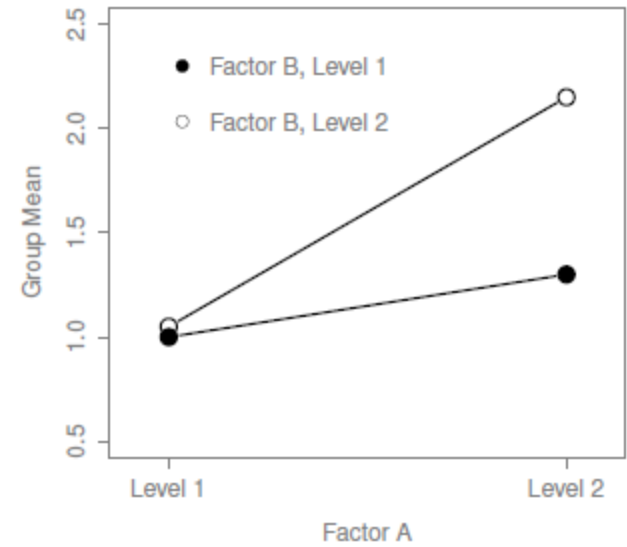
# Interactions

Crossover interaction



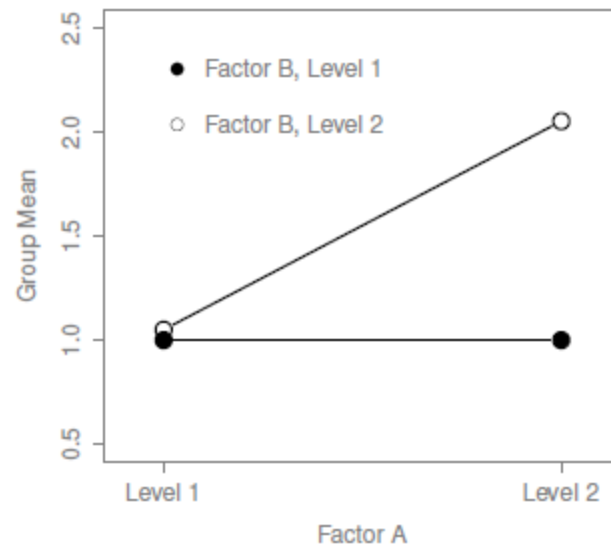
(a)

Effect for one level of Factor A



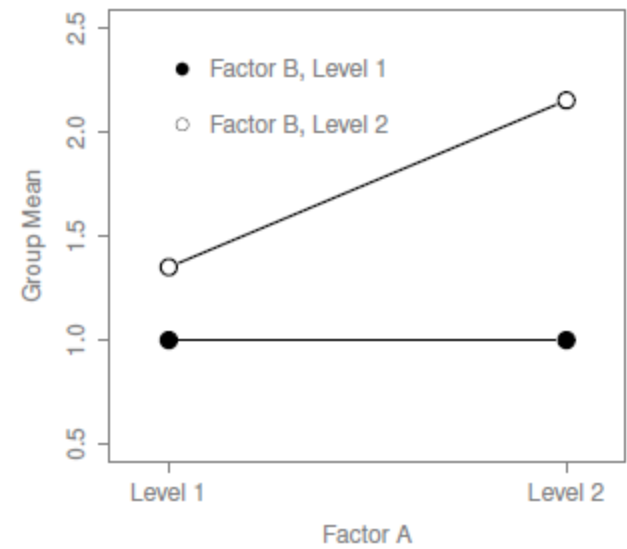
(b)

One cell is different



(c)

Effect for one level of Factor B



(d)

## Logic of the analysis: Different types of hypothesis tests (model comparisons) in unbalanced designs

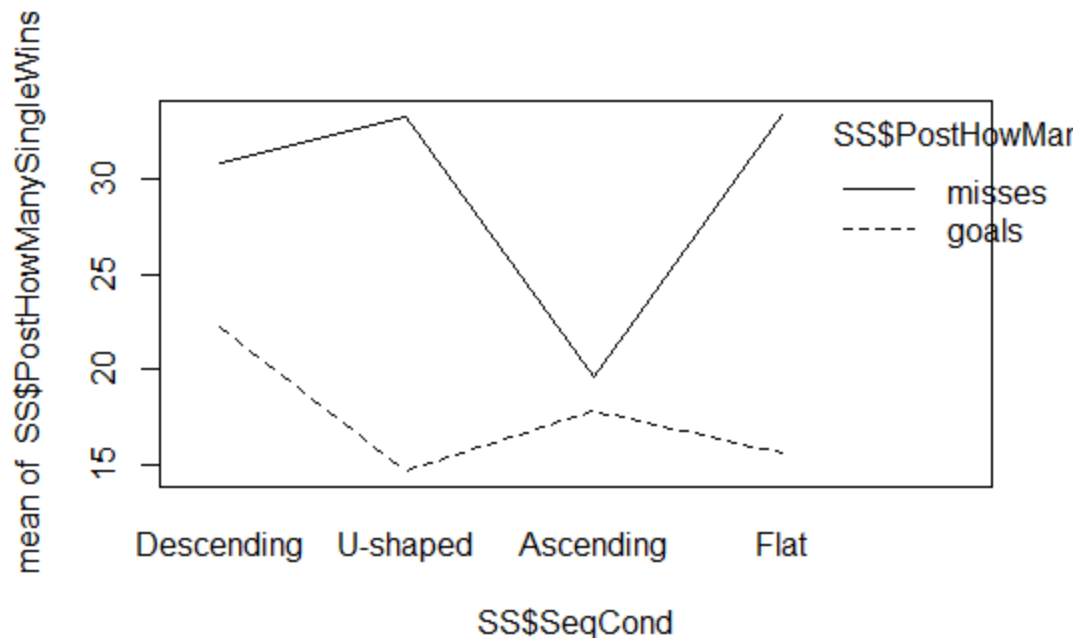
- Different types of hypothesis tests are an issue to consider in any factorial ANOVA (i.e., ANOVA with two or more predictors) *where group sample sizes are not equal* (e.g., where group 1,1 has  $N = 25$  and group 3,1 has  $N = 17$ )
- To do with the F statistic as a model comparison (see earlier slide)

Name	Model comparison method	Recommended for	Not recommended for
Type I Sums of Squares (R default)	Sequential; The first term that is entered “grabs” all the variance in $Y$ that it can. The second term grabs as much as possible of the remaining variance, and so on.	Situations where cell sizes (1,1; 1,2 etc) reflect differences in proportions in the population.  Situations where it is crucial to know the effect size (eta squared).	Situations where you do not have a theoretical justification for the ordering of predictors.
Type II Sums of Squares	Non-sequential, hierarchical; The null model always contains less terms, so that the term whose significance we are trying to test is not part of a higher-order term in the model (i.e., an interaction).	Most situations	
Type III Sums of Squares (SPSS default)	Non-sequential, unique; The null model always contains one less term, corresponding to the term whose significance we are trying to test.		Situations where you expect a significant main effect <i>and</i> an interaction. The main effects are meaningless when there is a significant interaction.

# Working together in R – descriptive statistics

## Interaction plot

Plot suggests that there might be an interaction.



## Descriptive statistics

As shown in the script, check for a correlation between the outcome variable and any proposed covariates. Also use the `psych` package to generate relevant descriptive statistics, as we did in the last lecture.

## Working together in R – Running the analysis

### **A different hypothesis – this time from our SS data (Hypothesis #2)**

Once gambling-related beliefs (PreDBC\_Total) are accounted for, a higher percentage of wins (PostHowManySingleWins) should be remembered in the *descending condition* relative to the others (SeqCond). Sequence condition could interact with question wording (PostHowManySingleCaptionType).

See script for a demonstration of a Type II Sums of Squares ANCOVA test of this hypothesis. We use the `lm` function for which you do not need to install a package. The `Anova` function we use is in the `car` package. We also make use of the `psych` package (`describeBy`), and the `effects` package (function: `effect`).



# Working together in R – Checking assumptions

Assumption	Checks available in R	If the assumption is not met...
Normality of residuals	<pre>hist(residuals(anova_ SSHyp2)) shapiro.test(residuals (anova_SSHyp2))</pre>	Try a generalised linear model – discussed in a few lectures' time
Constant variance of residuals across predicted group means – homogeneity of variance	<pre>leveneTest(formula) – car package. Formula must specify a saturated model (i.e., a model with all possible main effects and interactions) with no covariates.</pre>	<pre>oneway.test() kruskal.test()</pre>
Homogeneity of regression slopes (ANCOVA)	<pre>HRS &lt;- aov(outcome variable ~ predictor*covariate) or with multiple predictors: HRS &lt;- aov(outcome variable ~predictor1*predictor2 *covariate) Anova(HRS, type = 2)</pre>	Try a more complex model where the covariate is a predictor
Independence between the covariate and predictor(s) (ANCOVA)	<pre>aov(predictor1*predictor2~covariate)</pre>	Try a more complex model where the covariate is a predictor.

## Reporting the analysis – as in Results section

- Table (or very clear graph) showing means and SDs across factor levels. As in the interaction plot.
- In text: An ANCOVA (with Type II Sums of Squares) was conducted with percentage of remembered wins as the outcome variable, success-slope and question wording as predictors, and background beliefs (*Drake Beliefs About Chance* total score) as a covariate. After the significant influence of background beliefs was accounted for ( $F(1,325) = 11.32, p < .001, \eta^2 = .03$ ), the analysis revealed a significant main effect of success-slope ( $F(3,325) = 3.10, p = .03, \eta^2 = .02$ ), a significant main effect of question wording ( $F(1,325) = 38.08, p < .001, \eta^2 = .09$ ), and a significant interaction effect ( $F(3,325) = 3.83, p = .01, \eta^2 = .03$ ). Planned comparisons of the Descending condition's mean to those of other groups under a treatment contrast revealed a significant difference between the Ascending and Descending groups ( $p = .05$ ). As regards the interaction, the effect of question wording was found to be marginally significantly different in the Ascending, as compared to the Descending, condition ( $p = .07$ ). As the descriptive statistics suggest, question wording was irrelevant to the win-frequency estimates of participants in the Ascending condition. **Notably, the homogeneity of variance assumption was violated in the analysis.**
- $F$  values (with degrees of freedom),  $p$  values and effect sizes can also be reported in a table.
- A table showing estimated marginal means could also be included.

## Discussing the analysis – as in Discussion section

The results suggest that more wins were remembered when most wins were concentrated early in the experienced sequence, rather than late in the sequence. This is partly consistent with our expectation that memory for wins would resemble memory for word lists, where the words at the top of the list are remembered more clearly. Interestingly, the early-wins condition did not differ from the evenly-spaced and U-shaped conditions in terms of remembered wins. For the U-shaped condition, a likely explanation is that the early wins there were clearly remembered. For the evenly-spaced condition, it is possible that memory was boosted by the “spacing” of the wins. The effects of spacing are well-known in the memory literature. Words tend to be remembered better the wider their spacing across time. The spacing effect is also likely to have been responsible for the effects of question wording. People seemed to have been underestimating the frequency of losses, possibly because these were not as widely spaced as wins. Why this effect of question wording was not observed in the late-wins (Ascending) condition is unclear.

# Repeated measures ANOVA

Reading: Baguley Ch 16

## Logic of the analysis

A number of analyses can be labelled repeated-measures ANOVA:

- One-way ANOVA with repeated measures: each participant provides measures on all levels of the predictor variable (Factor A)
- Factorial ANOVA with repeated measures on all factors: there are two or more predictor variables (e.g., Factor A and Factor B), and each participant provides measures on all factors of all measures
- Mixed measures factorial ANOVA: there is at least one predictor factor with independent measures and at least one predictor factor with repeated measures (e.g., Factor A is a repeated measures factor while Factor B is an independent measures factor).
- Any of the above with one or more covariates.

As on slide 11, the following still need to be calculated for every factor in the model (repeated or independent). For each **repeated-measures factor**, additional sets of sums of squares need to be calculated. In the description below, Factor B is an independent-measures factor, while Factor A is a repeated-measures factor.

Total sum of squares expressing distance between all data points and grand mean

Sum of squares expressing difference between row marginal means and grand mean – variability due to Factor A

Sum of squares expressing difference between column marginal means and grand mean – variability due to Factor B

Sum of squares expressing the extent to which the group means *cannot* be predicted based on the marginal means alone – variability due to interaction between A and B (see next slide)

Sum of squares expressing distance between each participant's mean score across all levels of the repeated-measures factor and the grand mean. This is the variability due to random individual differences on Factor A. It is also called *within-subjects variance on A*.

Sum of squares expressing variability due to interaction between Factor B and random individual differences on Factor A

Degrees of freedom for all six quantities – see Baguley p. 639

Again, this is all the information we need for computing the F-value for each predictor and interaction term.

## Logic of the analysis: The sphericity assumption

- Apart from the standard assumption of the normality of residuals, repeated-measures ANOVA has a sphericity assumption. Here is what this assumption means in relation to a repeated-measures factor with three levels. If the difference between scores on Levels 1 and 2 is calculated for all participants, the variance of the difference quantities must equal the variance of the difference quantities obtained by comparing Levels 1 and 3, as well as Levels 2 and 3.
- Clearly, the assumption is met automatically if the repeated-measures factors in the model contain only two levels (e.g., Time 1 and Time 2).
- The sphericity assumption is violated so often that reporting the results means reporting one of two *corrections* to the  $p$ -value given the degree of sphericity violation. The possible corrections are the Greenhouse-Geisser correction and the Huynh-Feldt correction. If the average of the epsilon values generated by these procedures (see script) is .75 or greater, the Greenhouse-Geisser  $p$ -values should be reported (p. 633).

## Working together in R

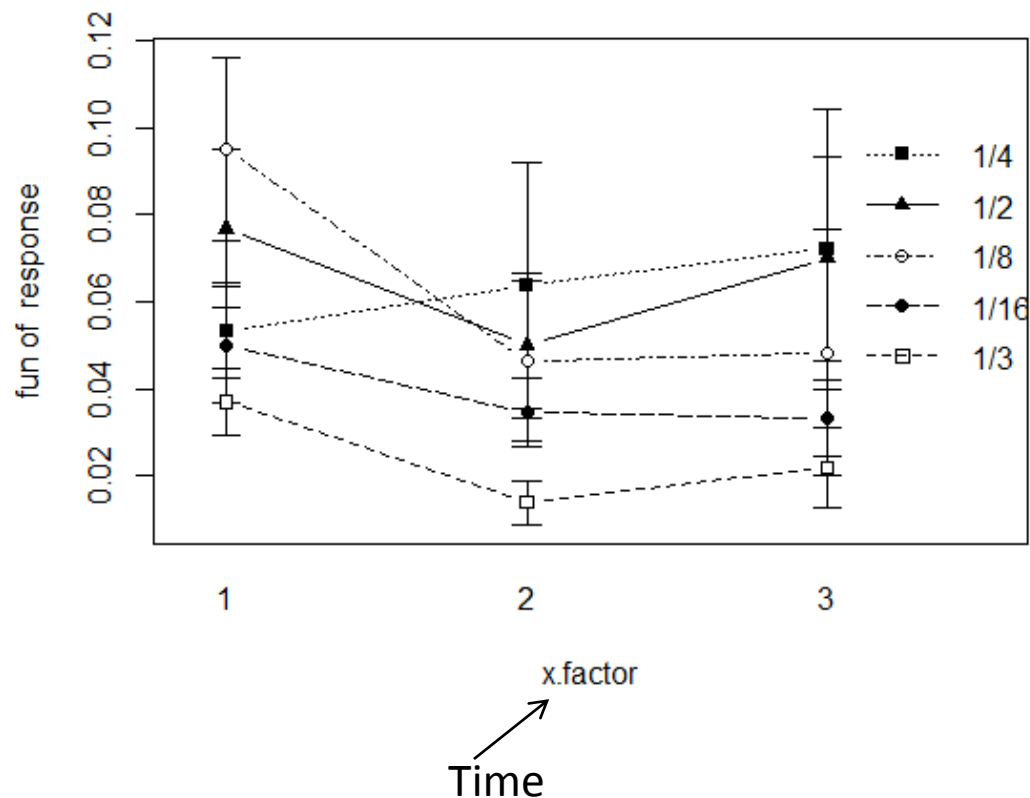
Two examples:

1. SS data Hypothesis 1a: If people perceive themselves to be problem-solving (learning a strategy) in games of chance, the number of player profile changes and the degree of kick-direction entropy should decrease over time in the Ascending slope condition.
2. SF data: If people's perceptions of correct problem-solving increase with win-frequency, participants experiencing higher win-frequency should exhibit fewer player profile changes in the last 30 rounds compared to the middle 40 and first 30 rounds. This should not be observed in the lower win-frequency conditions.

# Working together in R – Descriptive statistics

## Descriptive statistics and interaction plot following data restructuring from wide form to long form

See the script for a demonstration of using the `reshape` package to restructure the data before using the package `sciplot` to draw a quick graph. The `psych` package (`describeBy` function) and the `cor.test` function in the base package are used again, as for ANOVA. `cor.test` is used to assess whether prior beliefs (`PreDBC_Total`) should be included as a covariate. In Example 2, the inclusion of prior beliefs as a covariate proves unnecessary.





## Working together in R – Running the analysis

### **Two packages**

The script shows two packages for repeated-measures ANOVA in R. The `nlme` package (function: `lme`) is best to use when there are covariates in the model (as in Example 1). The alternative `ez` package (function: `ezANOVA`) is yet to be fully developed for analysis of covariance, but it has the advantage of including sphericity tests and corrections as part of the output.

## Working together in R – Checking assumptions

Assumption	Checks available in R	If the assumption is not met...
Normality of residuals	<pre>hist(as.numeric(residuals(lmeModelName)),breaks = 20)</pre> <pre>shapiro.test(residuals(lmeModelName))</pre>	
Sphercity	Mauchly's test of sphercity is available as part of the output for ezANOVA, but Baguley recommends reporting corrected $p$ -values instead (see p. 633, slide 22 and script Example 2).	Try moulding your repeated-measures factor into a factor containing just two levels.

## Reporting the analysis (Example 2) – as in Results section

- Table (or very clear graph) showing means and SDs across factor levels. As in the interaction plot (sciplot).
- In text: A repeated-measures ANOVA (with Type II Sums of Squares) was conducted. The outcome variable was the percentage of rounds in which the player profile was changed. The predictors were win frequency (1/16 vs. 1/8 vs. 1/4 vs. 1/3 vs. 1/2) and time period (first 30 rounds vs. middle 40 rounds vs. last 30 rounds). The analysis revealed a significant main effect of time ( $F(2,182) = 4.04$ ,  $p_{G-G} = .02$ ,  $ges = .01$ ). The time periods did not differ in the mean number of player changes, but the descriptive statistics suggest that the significant effect of time involved a decrease in the number of changes over time. Neither the effect of win-frequency ( $F(4,91) = 1.20$ ,  $p_{G-G} = .32$ ) nor the interaction effect ( $F(8,182) = 1.36$ ,  $p_{G-G} = .22$ ) were significant.
- $F$  values (with degrees of freedom), corrected  $p$  values and effect sizes ( $ges$ ) could alternatively be reported in a table.

Discussing the analysis – as in Discussion section

The results suggest that participants did not come to prefer a particular player over time in the higher win-frequency conditions. Instead, the results point to a decrease in player profile changes over time across all win-frequency conditions. This could be the result of boredom or strategising.

# Bootstrapping: worked examples

- Basic logic:
  - randomly draw out values from your sample (with replacement), until you have exactly the same number of values as there are in your sample (e.g., 96)
  - do this 500 to 100 000 times, depending on how long you wish to look at the computer while it “plays the lottery”
  - calculate the relevant test statistic (e.g., difference between means or regression coefficient) in each sample
  - determine confidence intervals around your test statistic based on a pooling of the bootstrapped estimates
- R packages generally calculate bias-corrected and accelerated bootstrapped confidence intervals ( $BC_a$ ). These are a good default.

# Reading

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. Available online: <http://health.adelaide.edu.au/psychology/ccs/teaching/lsr/>. Chapters 13-16.

Baguley, T. *Serious Stats: A Guide to Advanced Statistics for the Behavioural Sciences*. Palgrave Macmillan: UK. Chapter 16 “Repeated Measures ANOVA” (pdf in Study Materials/Readings).

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UK. Chapter 10. Comparing several means: ANOVA (pdf in Study Materials/Readings).