


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

Lecture 2

Basic ANOVA and regression

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

T-tests

Reading: LSR, Ch 13

Used for comparing:

- two means that come from different groups with the same variance on a measure (Student *t*-test)
- two means that come from different groups with differing variances (Welch test)
- a group mean and a theoretical value (one-sample *t*-test)
- means recorded by the same people in different conditions (related samples *t*-test)

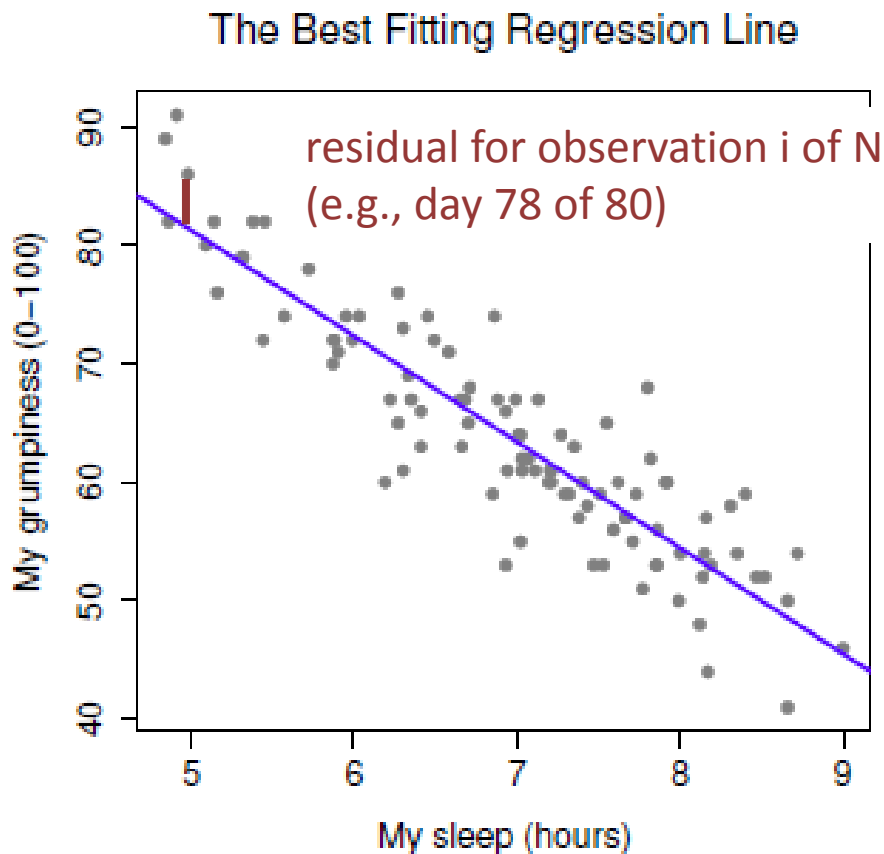
Quick demonstration of the Welch test: Did participants who were asked to think aloud during the soccer game score higher on the measure of supernatural strategising (PostSuploC)?

R has other packages for running *t*-tests but an advantage of the `lsr` package is that it calculates a Cohen's *d*, a measure of effect size – i.e., of the size of differences between two groups.

Linear regression

Reading: LSR, Ch 15

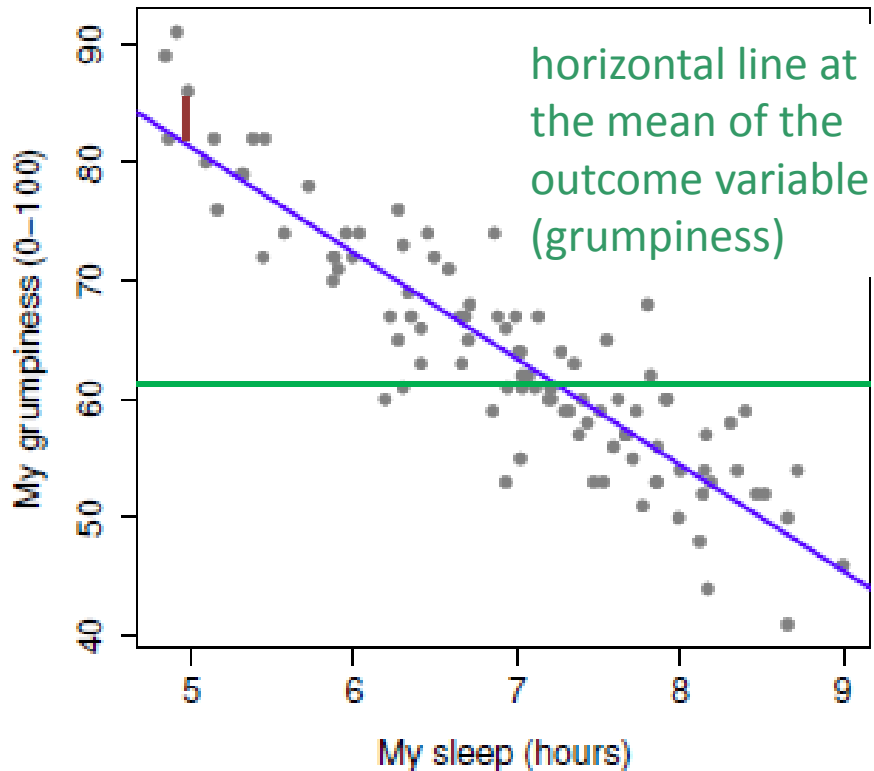
Logic of the analysis – one predictor



- We use a sequence of calculations (maximum likelihood estimation; MLE) to draw a line that minimises the sum of the squared values of the residuals
- MLE makes two key assumptions:
 - Residuals are normally distributed (with mean 0) and have a standard deviation that is the same at every value of the predicted variable/ “outcome” variable (grumpiness)
 - There is a linear relationship between the predictor (sleep) the outcome (grumpiness)

Logic of the analysis – one predictor (continued)

The Best Fitting Regression Line

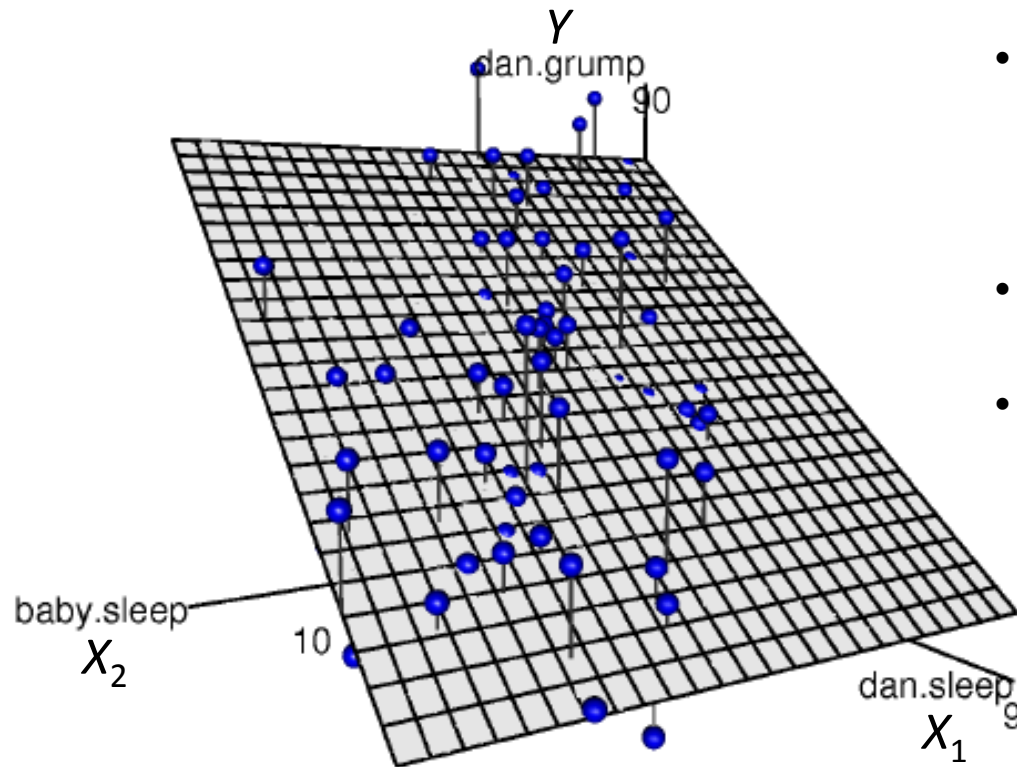


- R^2 tells us the extent to which the sum of squared residuals is smaller than the sum of the following: square of (each value of the outcome variable minus the mean of the outcome variable)
- Two answers to the same question of whether there is a significant relationship between the predictor and the outcome (null hypotheses in green):
 - T-test to determine whether the slope of the regression line (slope coefficient in the model) is significantly different from **zero**
 - F-test (ANOVA) to determine whether the model performs better than an **intercept-only model** (i.e., an equation in which the slope coefficient equals zero and the intercept then equals the outcome variable's mean)

The model: $Y_i = b_1 X_i + b_0 + \epsilon_i$

Slope coefficient \rightarrow b_1 Intercept \rightarrow b_0 Residual \rightarrow ϵ_i

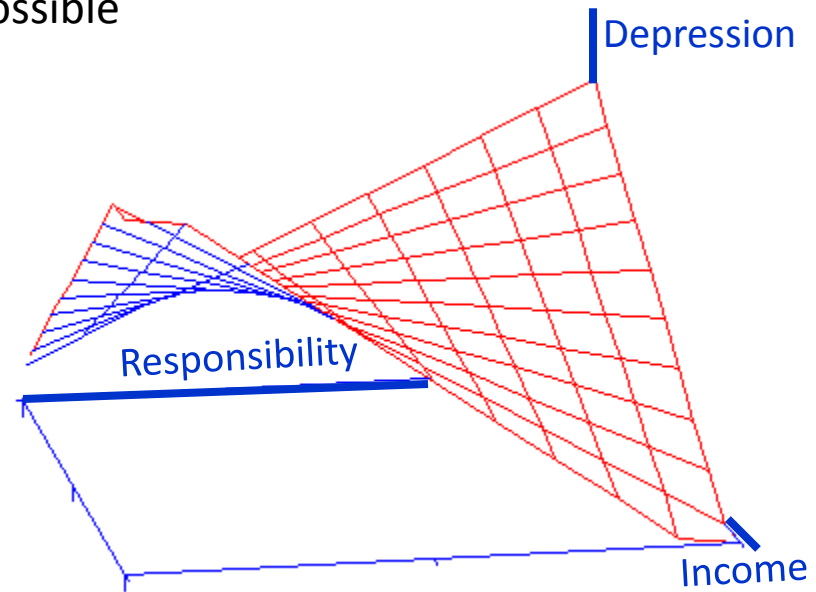
Logic of the analysis – two (or more) predictors



The model:

$$Y_i = b_2 X_{i2} + b_1 X_{i1} + b_0 + \epsilon_i$$

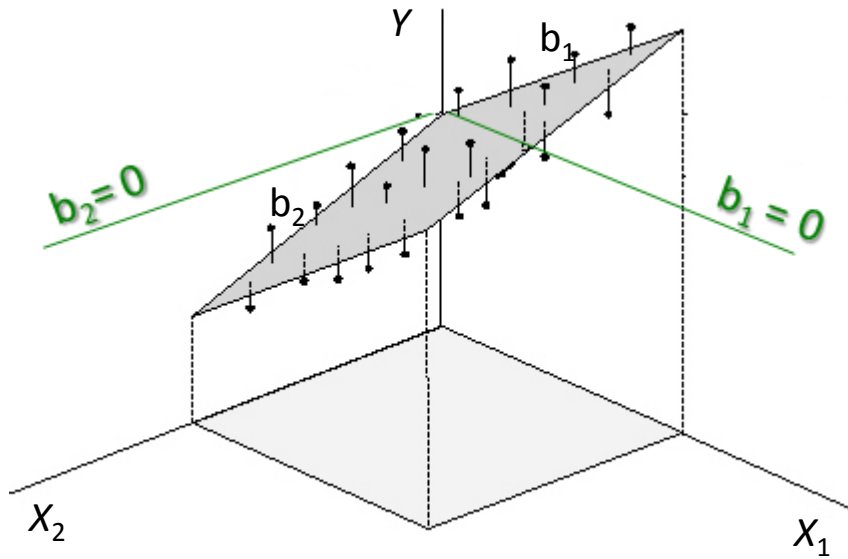
- We use MLE to determine an equation that minimises the sum of the squared values of the residuals (equation of a 3D plane for two predictors)
- MLE makes the same key assumptions as for analyses with a single predictor
- Interactions between the predictors are possible



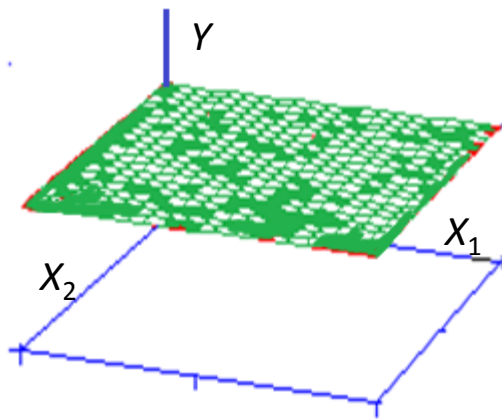
The model:

$$Y_i = b_2 X_{i2} + b_1 X_{i1} + b_3 X_{i1} X_{i2} + b_0 + \epsilon_i$$

Logic of the analysis – two predictors (continued)



Intercept-only model



Horizontal plane from the mean of the outcome variable (Y)

- R^2 has the same meaning, but you can also calculate adjusted R^2 , which is smaller than R^2 if there are many predictors and/or the sample size is small
- Three associated hypothesis tests (null hypotheses in green):
 - T-tests for each coefficient in the model: is it significantly different from **zero**?
 - F-test (ANOVA) to determine whether the model performs better than an **intercept-only model (i.e., an equation in which all slope coefficients equal zero and the intercept is then equal to the outcome variable's mean)**
 - Hierarchical regression: F-test (ANOVA) to determine whether a model featuring one or more additional predictors performs better than the **original model** p. 481

Hypothesis from our dataset (actually from another very similar dataset; SF in your Study Materials/Data folder)

Experiment

- $N = 97$
- 100 trials of the soccer-themed slot-machine task under one of five win-frequency conditions
 - 1 win per 2 trials, per 3 trials, per 4 trials, per 8 trials and per 16 trials
- Pre-game and post-game questionnaires almost identical to those in the Success-Slope experiment from Lecture 1. Final win amount (final credits) also calculated.

Hypothesis

In many causal judgement experiments, as the frequency with which two events co-occur increases, conclusions that one event causes the other have been found to increase in strength. Example: treatment with a certain drug and recovery. Here, we expect the same to be the case for wins and choices made during the game. As win frequency increases, choices should come to be considered more causally effective (i.e., more strategic). Win frequency should predict natural or supernatural illusion of control.

Post-game measure of illusion of problem-solving – slight difference from SS data

'Natural' IoC

1. My skill in playing the game.
2. I got better with practice.
3. I developed a logical strategy for playing.
4. Experience in playing computer games.

Natural IoC variable: Average of these items

'Supernatural' IoC

1. I took advantage of moments when my luck was good.
2. I've always been a lucky kind of person.
3. I knew how to make my luck turn good.
4. A certain lucky way of playing just seemed to work for me.
5. The players I chose.
6. I learned how to predict the movements of the goalkeeper.

Supernatural IoC variable: Average of these items

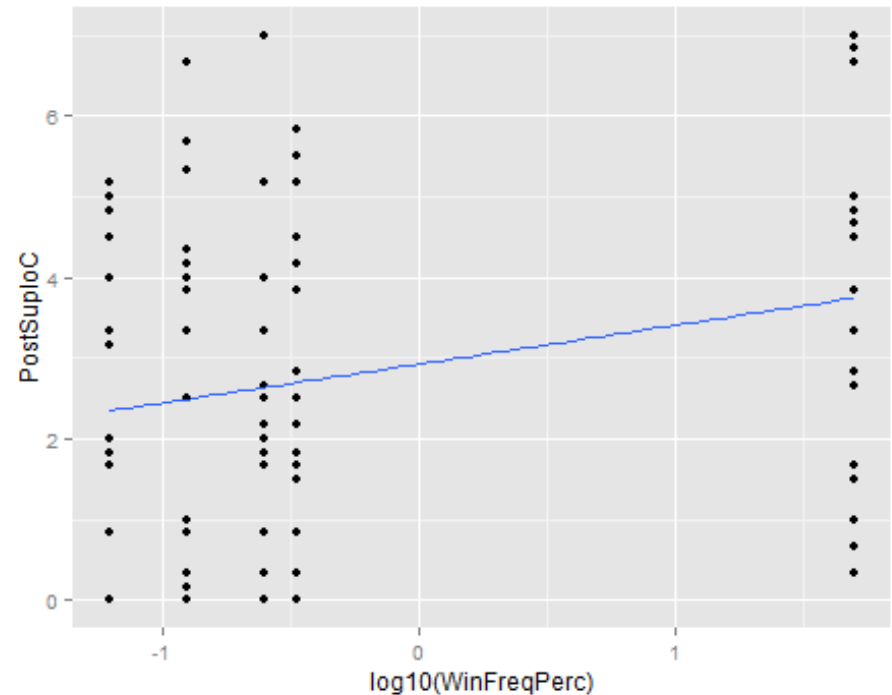
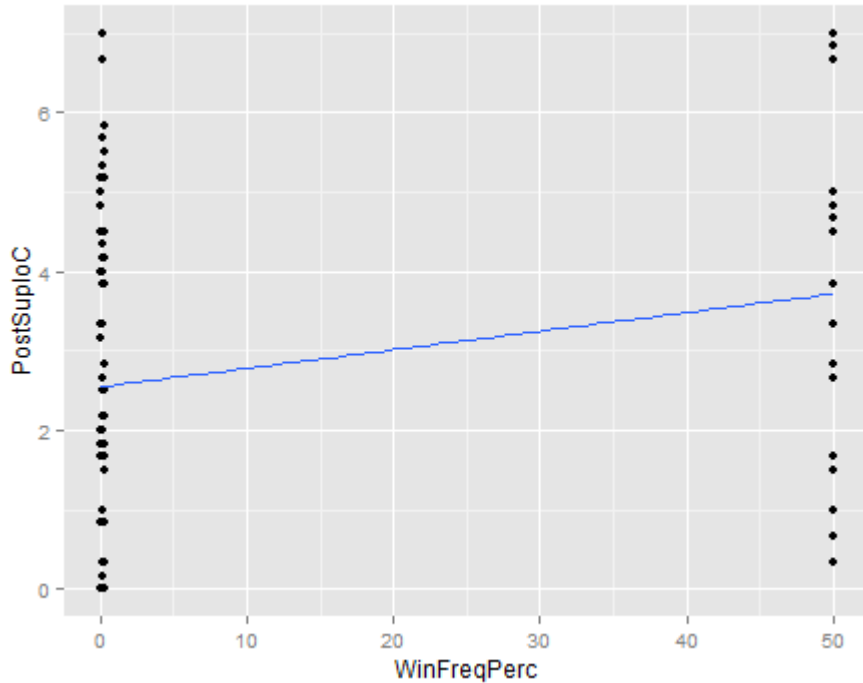
All chance.

It was all chance.

We will use Supernatural IoC (PostSupIoC) as our outcome variable in this demonstration because it has more items and is therefore a potentially more reliable measure of the illusion of problem-solving.

Working together in R – descriptive statistics

Graph – ggplot2 commands are in the script



Correlation table

As shown in the script, create a subset data frame of the variables you want to correlate, then use the `correlate` function in the `lsr` package. You should include all possible predictors of the outcome variable for which data is available.

Working together in R – Running the analysis

Revised hypothesis based on the correlation table

Once gambling-related beliefs and soccer interest assessed in the pre-game questionnaire (PreSoccerInterest; PreDBC_Sup) are accounted for, win frequency (LogWinFreqPerc) is a significant predictor of the illusion of supernatural control (PostloCSup).

See script for a demonstration of a hierarchical regression approach to testing this hypothesis. We use the `lm` and `anova` functions for which you do not need to install a package (they are in the 'base' package). We also make use of the `lsr` package.

Working together in R – Checking assumptions

Assumption	Checks available in R	If the assumption is not met...
Normality of residuals	<pre>hist(residuals(model1), breaks = 20) plot(model1, which = 2) shapiro.test(residuals (model1))</pre>	Transform one or more of the predictors
Constant variance of residuals <ul style="list-style-type: none"> • lack of influential points • homogeneity of variance 	<pre>plot(model1, which = 4) plot(model1, which = 5) plot(model1, which = 3) ncvTest(model1)</pre>	Run regression without influential points (see script) Run regression with heteroscedasticity-corrected covariance matrix (see script)
Linearity of the relationship between the outcome and the predictor(s)	Plot of fitted values against observed values <pre>plot(model1, which = 1) residualPlots(model1)</pre>	Transform one or more of the predictors

Reporting the analysis

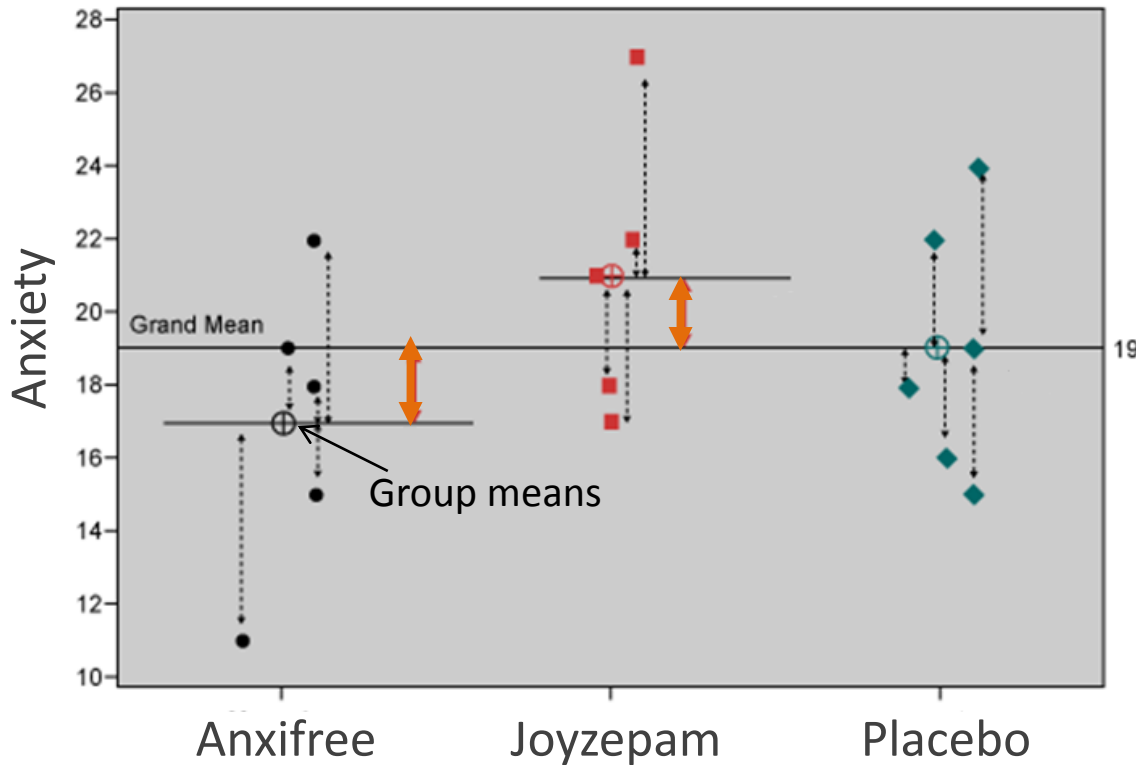
- Table showing coefficients, R^2 , T-test and hierarchical regression results (if any). Standardised coefficients (β) tend to also be reported.
- Summary of results, given the hypothesis: Overall, the analysis indicated that illusion-of-control ratings increased with increases in win-frequency, once the influence of background beliefs and soccer interest was accounted for.

Hierarchical regression step	Predictors	b	SE b	β	t	p	Adj R^2
1	Intercept						
	DBC total						
2	Intercept						
	DBC total						
	Win-frequency						

ANOVA: Independent measures

Reading: LSR, Ch 14 and 16

Logic of the analysis – one predictor (here, drug type, with 3 levels)



- We calculate two quantities:
 - Sum of squares expressing difference between each individual score and the group mean:
 - Sum of squares expressing **difference** between group means and grand mean – variability due to factor (drug type)
- These enable us to compute an F-value, which can then be tested for significance.

*Note: Data in illustration does not correspond to textbook

	df	sum of squares	mean squares	F-statistic	p-value
between groups	$df_b = G - 1$	$SS_b = \sum_{k=1}^G N_k (\bar{Y}_k - \bar{Y})^2$	$MS_b = \frac{SS_b}{df_b}$	$F = \frac{MS_b}{MS_w}$	[complicated]
within groups	$df_w = N - G$	$SS_w = \sum_{k=1}^G \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y}_k)^2$	$MS_w = \frac{SS_w}{df_w}$	-	-

- N is number of participants
- G is number of groups
- i is a participant number
- k is an integer representing the group number/factor level
- Effect size – eta-squared:

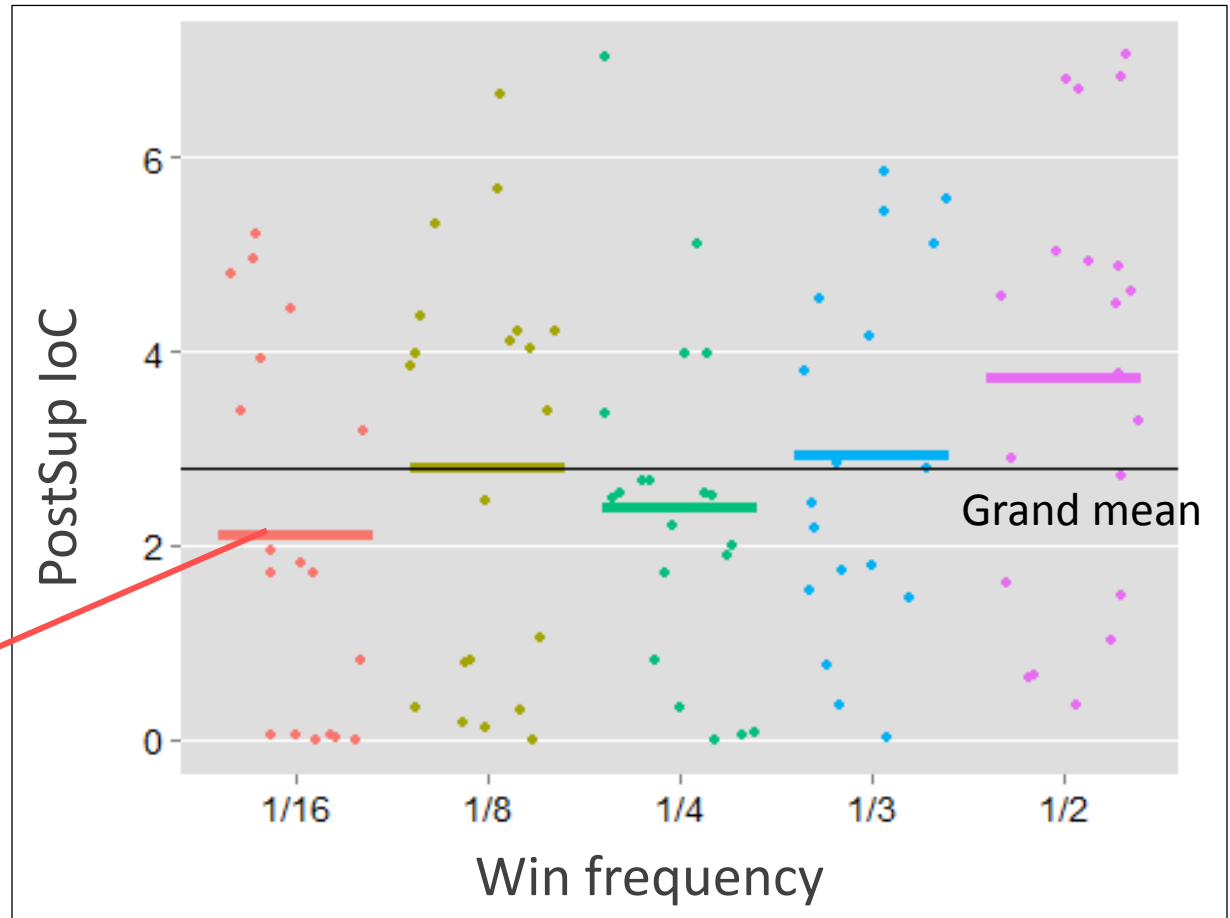
$$\eta^2 = \frac{SS_b}{SS_{tot}}$$

SS expressing difference between all scores (regardless of group) and grand mean

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).

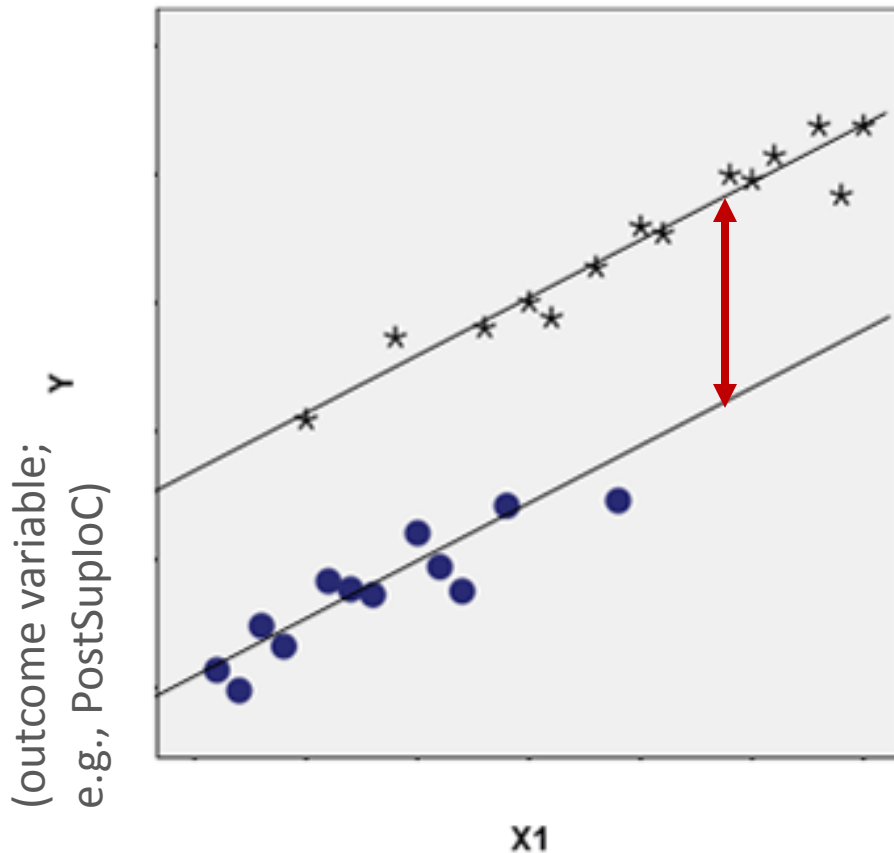
If the rules are met:

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

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

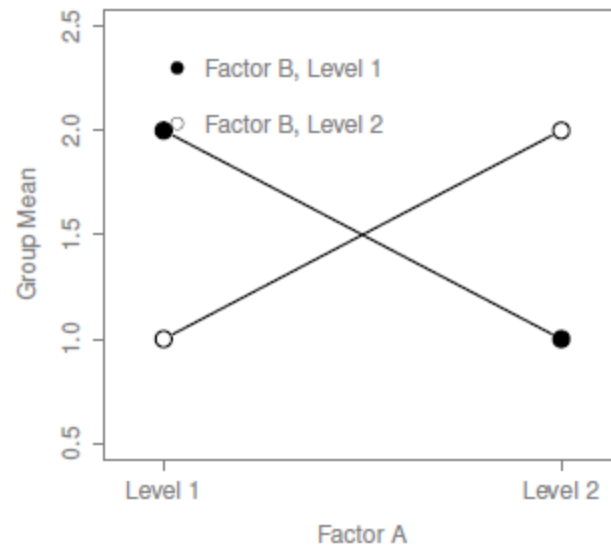
Using the first four quantities, we can calculate the residual sum of squares
 $SS_R = SS_T - (SS_A + SS_B + SS_{A:B})$

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:

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

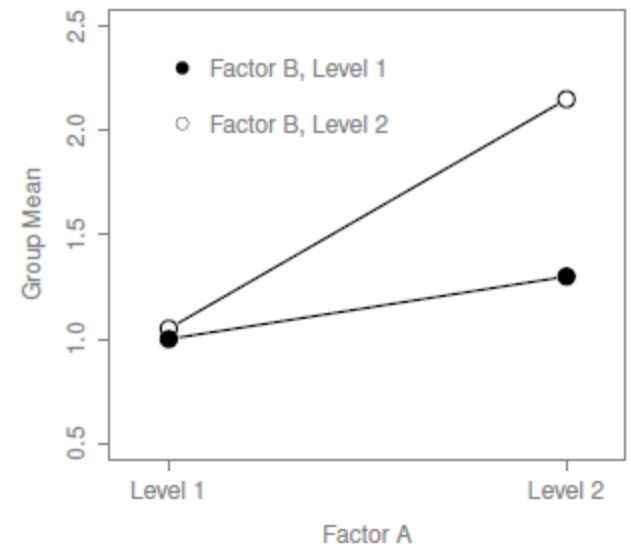
Interactions

Crossover interaction



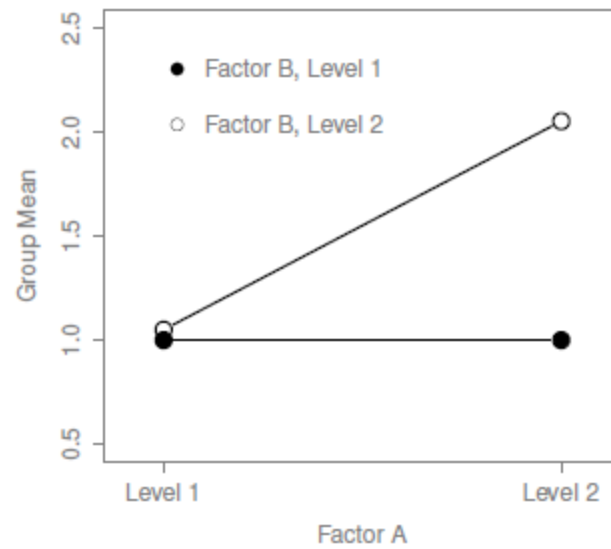
(a)

Effect for one level of Factor A



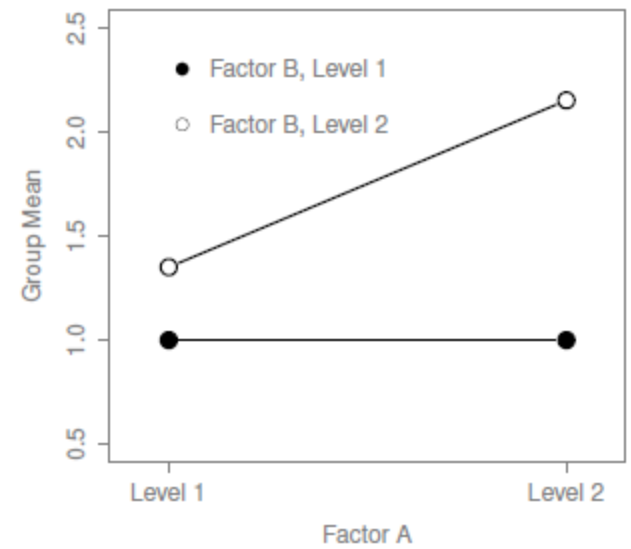
(b)

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

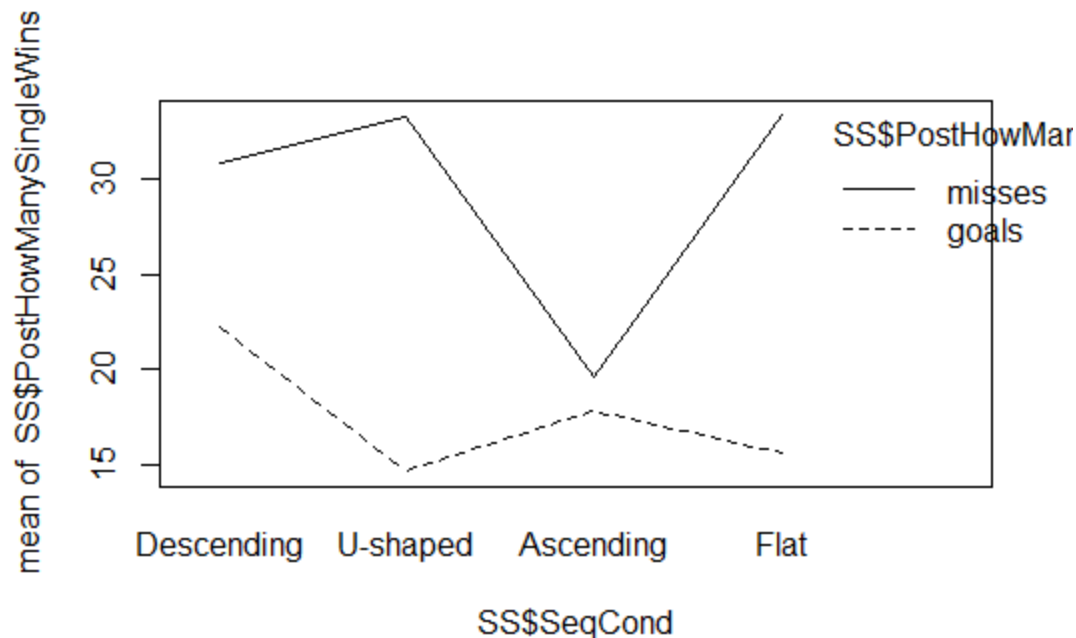
- An issue to consider in any factorial ANOVA (i.e., ANOVA with two or more predictors) *where group sample sizes are not equal* (e.g., where group 1,1 has $N = 25$ and group 3,1 has $N = 17$)
- To do with the F statistic as a model comparison (see earlier slide)

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.	<p>Situations where cell sizes (1,1; 1,2 etc) reflect differences in proportions in the population.</p> <p>Situations where it is crucial to know the effect size (eta squared).</p>	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 <- aov(outcome variable ~ predictor*covariate) or with multiple predictors: HRS <- 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.

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).