Comments for MEDB 5502, Week 04

Topics to be covered

- What you will learn
  - Indicator variables for three or more categories
  - Multiple factor analysis of variance
  - Checking assumptions of analysis of variance
  - Interactions in analysis of variance
  - Interactions in analysis of covariance
  - Interactions in multiple linear regression
  - Unbalanced data

Review oneway analysis of variance

- $H_0 : \ \mu_1 = \mu_2 = \ldots = \mu_k$
- $H_1 : \ \mu_i \neq \mu_j$ for some i, j
  - Reject H_0 if F-ratio is large
- Note: when k=2, use analysis of variance or t-test

Speaker notes

In Biostats-1, we discussed the comparison of three or more means using oneway or single factor analysis of variance. You can actually use analysis of variance when comparing only two means, but an equivalent alternative is the t-test.

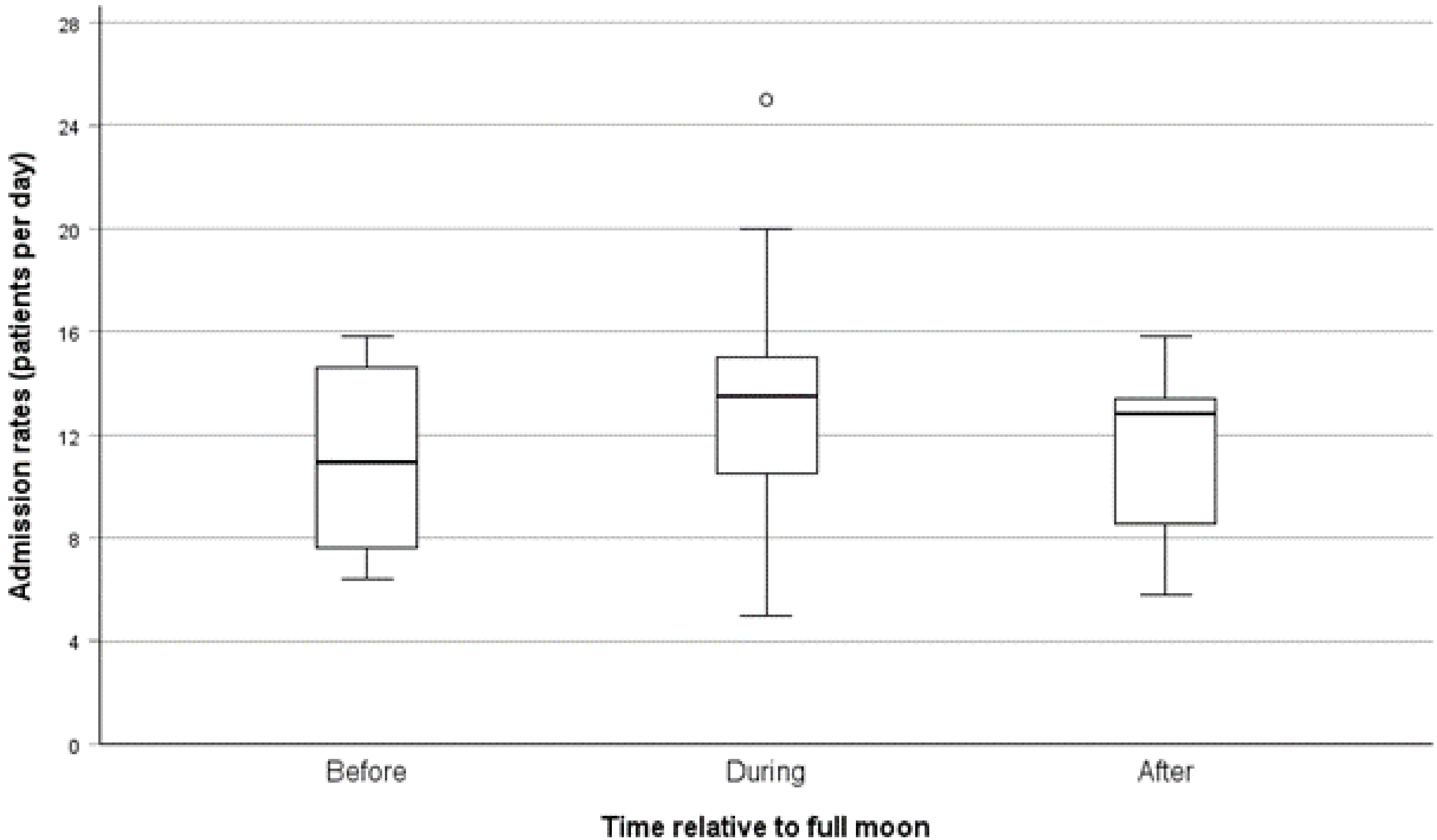
Full moon data

- Admission rates to mental health clinic before, during, and after full moon.
- One year of data

Speaker notes

To illustrate oneway analysis of variance, I found a dataset on mental health clinic admissions.

Boxplot of full moon data



Speaker notes

Here is a boxplot. It looks like admissions during a full moon are a bit higher than before or after a full moon.

Descriptive statistics

Report

Admission

Moon	Mean	N	Std. Deviation
After	11.458	12	3.1094
Before	10.917	12	3.6199
During	13.417	12	5.5014
Total	11.931	36	4.2255

Speaker notes

Here are the means and standard deviations.

Analysis of variance table

ANOVA

Admission

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	41.514	2	20.757	1.174	.322
Within Groups	583.402	33	17.679		
Total	624.916	35			

Speaker notes

The F-ratio is close to 1 and the p-value is large. You would accept the null hypothesis and conclude that the average admission rate during a full moon is the same as at other times.

Tukey post hoc

Multiple Comparisons

Dependent Variable: Admission
Tukey HSD

(I) Moon_code	(J) Moon_code	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
After	Before	.5417	1.7165	.947	-3.670	4.754
	During	-1.9583	1.7165	.496	-6.170	2.254
Before	After	-.5417	1.7165	.947	-4.754	3.670
	During	-2.5000	1.7165	.325	-6.712	1.712
During	After	1.9583	1.7165	.496	-2.254	6.170
	Before	2.5000	1.7165	.325	-1.712	6.712

Admission

Tukey HSD[a]

Moon_code	N	Subset for alpha = 0.05 1
Before	12	10.917
After	12	11.458
During	12	13.417
Sig.		.325

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 12.000.

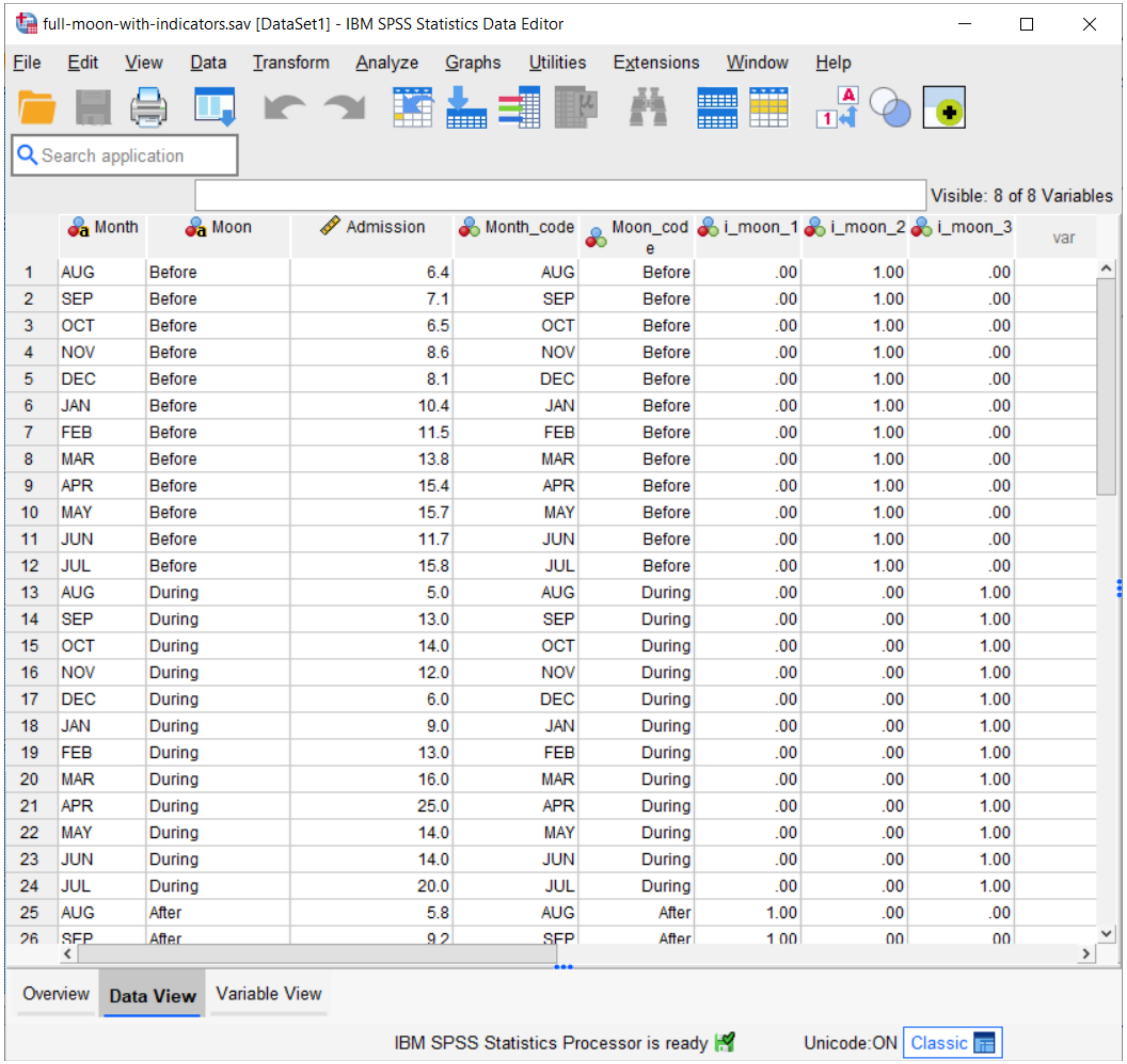
Speaker notes

Normally, you would not run a post hoc test after accepting the null hypothesis. If you did, you would see that every confidence interval includes the value of zero.

There are a few times (very, very few) where you accept the null hypothesis and the post hoc test shows differences between two of the groups.

There are also a few times where you reject the null hypothesis, but the post hoc test shows no differences between any of the groups. There is an explanation in both cases, but the cases are so rare that I do not want to delve further into them.

Creating indicator variables



	Month	Moon	Admission	Month_code	Moon_code	i_moon_1	i_moon_2	i_moon_3	var
1	AUG	Before	6.4	AUG	Before	.00	1.00	.00	
2	SEP	Before	7.1	SEP	Before	.00	1.00	.00	
3	OCT	Before	6.5	OCT	Before	.00	1.00	.00	
4	NOV	Before	8.6	NOV	Before	.00	1.00	.00	
5	DEC	Before	8.1	DEC	Before	.00	1.00	.00	
6	JAN	Before	10.4	JAN	Before	.00	1.00	.00	
7	FEB	Before	11.5	FEB	Before	.00	1.00	.00	
8	MAR	Before	13.8	MAR	Before	.00	1.00	.00	
9	APR	Before	15.4	APR	Before	.00	1.00	.00	
10	MAY	Before	15.7	MAY	Before	.00	1.00	.00	
11	JUN	Before	11.7	JUN	Before	.00	1.00	.00	
12	JUL	Before	15.8	JUL	Before	.00	1.00	.00	
13	AUG	During	5.0	AUG	During	.00	.00	1.00	
14	SEP	During	13.0	SEP	During	.00	.00	1.00	
15	OCT	During	14.0	OCT	During	.00	.00	1.00	
16	NOV	During	12.0	NOV	During	.00	.00	1.00	
17	DEC	During	6.0	DEC	During	.00	.00	1.00	
18	JAN	During	9.0	JAN	During	.00	.00	1.00	
19	FEB	During	13.0	FEB	During	.00	.00	1.00	
20	MAR	During	16.0	MAR	During	.00	.00	1.00	
21	APR	During	25.0	APR	During	.00	.00	1.00	
22	MAY	During	14.0	MAY	During	.00	.00	1.00	
23	JUN	During	14.0	JUN	During	.00	.00	1.00	
24	JUL	During	20.0	JUL	During	.00	.00	1.00	
25	AUG	After	5.8	AUG	After	1.00	.00	.00	
26	SEP	After	9.2	SEP	After	1.00	.00	.00	

Speaker notes

This is the data window after asking SPSS to create indicator variables. Note that SPSS created indicator variables alphabetically (after, then before, then during). This is not the order I would have chosen. I would do indicators for before, then during, then after. But I can work with this.

Running general linear model with all indicator variables

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	41.514[a]	2	20.757	1.174	.322
Intercept	.000	0	.	.	.
i_moon_1	.000	0	.	.	.
i_moon_2	.000	0	.	.	.
i_moon_3	.000	0	.	.	.
Error	583.403	33	17.679		
Total	5749.090	36			
Corrected Total	624.916	35			

a. R Squared = .066 (Adjusted R Squared = .010)

Speaker notes

If you try to put all three indicator variables in, you get a disasterous result. There is a perfect collinearity in this data because the sum of all three indicators is constant.

To fix this, you must remove one of the three indicators.

Analysis of variance table with first and second indicators

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	41.514[a]	2	20.757	1.174	.322
Intercept	2160.083	1	2160.083	122.185	<.001
i_moon_1	23.010	1	23.010	1.302	.262
i_moon_2	37.500	1	37.500	2.121	.155
Error	583.403	33	17.679		
Total	5749.090	36			
Corrected Total	624.916	35			

a. R Squared = .066 (Adjusted R Squared = .010)

Speaker notes

This is the analysis of variance table with just the first two indicators. There are extra rows in this table that just cause confusion.

Irrelevant rows removed

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	41.514[a]	2	20.757	1.174	.322
[redacted]					
Error	583.403	33	17.679		
[redacted]					
Corrected Total	624.916	35			

a. R Squared = .066 (Adjusted R Squared = .010)

Speaker notes

Here are the three rows that you should look at: Corrected Model, Error, and Corrected Total.

Parameter estimates, 1 of 3

Parameter Estimates

Dependent Variable: Admission

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	13.417	1.214	11.054	<.001	10.947	15.886
i_moon_1	-1.958	1.717	-1.141	.262	-5.451	1.534
i_moon_2	-2.500	1.717	-1.456	.155	-5.992	.992

- 11.458 - 13.417 = -1.959
- 10.917 - 13.417 = -2.5

Speaker notes

Here are the parameter estimates. The intercept is the estimated average rate when the first and second indicator both equal zero. This is effectively the average rate for the third group (during a full moon).

The first slope is the estimated average difference in rates when you increase the first indicator by one unit and hold the second indicator constant. This is effectively the the average rate for the first group (after a full moon) minus the average rate for the third group (during a full moon). When the days switch from during a full moon to after a full moon, the estimated average admission rate decreases by almost two patients per day.

The second slope is the estimated average difference in rates when you increase the second indicator by one unit and hold the first indicator constant. This is effectively the the average rate for the second group (before a full moon) minus the average rate for the third group (during a full moon). When the days switch from during a full moon to before a full moon, the estimated average admission rate decreases by 2.5 patients per day.

Parameter estimates, 2 of 3

Parameter Estimates

Dependent Variable: Admission

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	10.917	1.214	8.994	<.001	8.447	13.386
i_moon_1	.542	1.717	.316	.754	-2.951	4.034
i_moon_3	2.500	1.717	1.456	.155	-.992	5.992

- 11.458 - 10.917 = 0.541
- 13.417 - 10.917 = 2.5

Speaker notes

You don’t have to remove the third indicator variable. Any will work, but the interpretation changes. If you remove the second indicator variable, the intercept represents the estimated average admission rate of the second group (about 11 patients per day).

The first slope represents the estimated difference between days after a full moon compared to before a full moon. You see about half a patient more per day after a full moon compared to before.The difference is about half a day on average.

The second slope represents the estimated difference between days during a full moon compared to before a full moon. You see about 2.5 more patients on average during a full moon compared to before a full moon.

Parameter estimates, 3 of 3

Speaker notes

You have a third choice: leaving out the first indicator. The intercept now represents the estimated average for the first group (before a full moon) which is about 11.5 patients per day. The first slope tells you that you see about half a patient less on average after a full moon compared to before a full moon. The second slope tells you that you see about 2 patients more on average during a full moon compared to before a full moon.

You may see the term “reference category”. This represents the category associated with the left out indicator variable. The other categories are compared to the reference category.

Which is the best interpretation? It depends a lot on the context of the problem. I would argue that comparing before and after to during makes the most sense, but you could argue for any of these three interpretations.

Or you could synthesize these. Average admission rates are 11.5 patients per day before a full moon, increase by about 2.5 patients per day on average during a full moon and drop by about 2 patients per day after a full moon. Then admission rates jump back up by about half a patient per day when you switch back to days before a full moon.

Using moon as a fixed factor

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	41.514[a]	2	20.757	1.174	.322
Intercept	5124.174	1	5124.174	289.847	<.001
Moon	41.514	2	20.757	1.174	.322
Error	583.403	33	17.679		
Total	5749.090	36			
Corrected Total	624.916	35			

a. R Squared = .066 (Adjusted R Squared = .010)

Speaker notes

You don’t have to create indicator variables yourself. You can let SPSS do it behind the scenes by adding moon to the fixed factor dialog box. SPSS will create indicator variables for the first two categories in alphabetical order and leave out the indicator variable for the category that appears last in alphabetical order.

Removing the unneeded rows

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	41.514[a]	2	20.757	1.174	.322
[redacted]					
Moon	41.514	2	20.757	1.174	.322
Error	583.403	33	17.679		
[redacted]					
Corrected Total	624.916	35			

a. R Squared = .066 (Adjusted R Squared = .010)

Speaker notes

If you run the analysis this way, remove the rows corresponding to Intercept and Total.

Parameter estimates using Moon as a fixed factor

Parameter Estimates

Dependent Variable: Admission

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	13.417	1.214	11.054	<.001	10.947	15.886
[Moon=After]	-1.958	1.717	-1.141	.262	-5.451	1.534
[Moon=Before]	-2.500	1.717	-1.456	.155	-5.992	.992
[Moon=During]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Speaker notes

The parameter estimates are the ones that I showed earlier with the third indicator variable left out. As a general rule, SPSS assigns the first indicator to the first category in alphabetical order (After in this case) and the last indicator to the last category in alphabetical order (During in this case). Then it leaves out the last indicator variable.

Live demo, Multiple factor analysis of variance

Break #1

- What you have learned
  - Indicator variables for three or more categories
- What's coming next
  - Multiple factor analysis of variance

Mathematical model

- $Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$
  - i=1,…,a levels of the first categorical variable
  - j=1,…,b levels of the second categorical variable
  - k=1,…,n replicates with first and second categories

- $H_0 :\ \alpha_i = 0$ for all i

- $H_0 :\ \beta_j = 0$ for all j

Speaker notes

The mathematical model for two factor analysis of variance is a bit more complex than a single factor analysis of variance. You have an overall mean, mu, and deviations from the overall mean associated with the first factor (alpha), deviations from the overall mean associated with the second factor (beta) and an error term (epsilon).

There are a total of a and b categories for the two categorical independent variables.

Crosstabulation of categorical predictors

Speaker notes

This table shows how there is one observation for each combination of month and moon phase.

Analysis of variance table for moon data

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	497.097[a]	13	38.238	6.581	<.001
Intercept	5124.174	1	5124.174	881.961	<.001
Moon	41.514	2	20.757	3.573	.045
Month	455.583	11	41.417	7.129	<.001
Error	127.819	22	5.810		
Total	5749.090	36			
Corrected Total	624.916	35			

a. R Squared = .795 (Adjusted R Squared = .675)

Speaker notes

THis is the analysis of variance table. There is one less degree of freedom than the number of categories for each categorical predictor variable. There is a statistically significant difference between the twelve months and a borderline significant difference between the three moon phases.

This differs from the single factor analysis of variance because adding in month as a categorical predictor removed a lot of variation. You are now able to account for almost 80% of the variation in admission rates. Without month in the model, you accounted for less than 7% of the variation.

Removing irrelevant rows

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	497.097[a]	13	38.238	6.581	<.001
[redacted]					
Moon	41.514	2	20.757	3.573	.045
Month	455.583	11	41.417	7.129	<.001
Error	127.819	22	5.810		
[redacted]					
Corrected Total	624.916	35			

a. R Squared = .795 (Adjusted R Squared = .675)

Speaker notes

The rows corresponding to the intercept and the total (uncorrected total) are not needed.

Parameter estimates for the full moon model

Parameter Estimates

Dependent Variable: Admission

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	11.253	1.503	7.486	<.001	8.135	14.370
[Moon=After]	-1.958	.984	-1.990	.059	-3.999	.082
[Moon=Before]	-2.500	.984	-2.541	.019	-4.541	-.459
[Moon=During]	0[a]	.	.	.	.	.
[Month=APR]	8.967	1.968	4.556	<.001	4.885	13.048
[Month=AUG]	-4.033	1.968	-2.049	.053	-8.115	.048
[Month=DEC]	-1.400	1.968	-.711	.484	-5.482	2.682
[Month=FEB]	2.900	1.968	1.474	.155	-1.182	6.982
[Month=JAN]	1.000	1.968	.508	.616	-3.082	5.082
[Month=JUL]	7.000	1.968	3.557	.002	2.918	11.082
[Month=JUN]	3.067	1.968	1.558	.133	-1.015	7.148
[Month=MAR]	4.533	1.968	2.303	.031	.452	8.615
[Month=MAY]	4.567	1.968	2.320	.030	.485	8.648
[Month=NOV]	-.333	1.968	-.169	.867	-4.415	3.748
[Month=OCT]	-.300	1.968	-.152	.880	-4.382	3.782
[Month=SEP]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Speaker notes

The intercept represents the average admission rate during a full moon when the month is September. The two slope terms show how much lower the average admission rates are before and after a full moon, respectively, compared to during a full moon, holding month constant.

Tukey post hoc test

Multiple Comparisons

Dependent Variable: Admission

Tukey HSD

(I) Moon	(J) Moon	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
After	Before	.542	.9840	.847	-1.930	3.014
	During	-1.958	.9840	.138	-4.430	.514
Before	After	-.542	.9840	.847	-3.014	1.930
	During	-2.500*	.9840	.047	-4.972	-.028
During	After	1.958	.9840	.138	-.514	4.430
	Before	2.500*	.9840	.047	.028	4.972

Based on observed means.
The error term is Mean Square(Error) = 5.810.

*. The mean difference is significant at the .05 level.

Admission

Tukey HSD[a,b]

Moon	N	Subset 1	Subset 2
Before	12	10.917	
After	12	11.458	11.458
During	12		13.417
Sig.		.847	.138

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 5.810.

a. Uses Harmonic Mean Sample Size = 12.000.

b. Alpha = .05.

Speaker notes

Use the Tukey posthoc test because the sample sizes are equal across the moon phases. The results are a bit ambiguous because before and after are not statistically different, after and during are not statistically different but before and during are statistically different. This is probably due to a lack of precision and an extra year’s worth of data would help quite a bit.

The analogy I use is travel time. My wife and I live in Leawood. Our son lives in Lee’s Summit. A repair shop we all use is in Olathe. It is not far from Leawood to Olathe. It is not far from Leawood to Lee’s Summit. But it is far from Lee’s Summit to Olathe.

Live demo, Multiple factor analysis of variance

Break #2

- What you have learned
  - Multiple factor analysis of variance
- What's coming next
  - Checking assumptions of analysis of variance

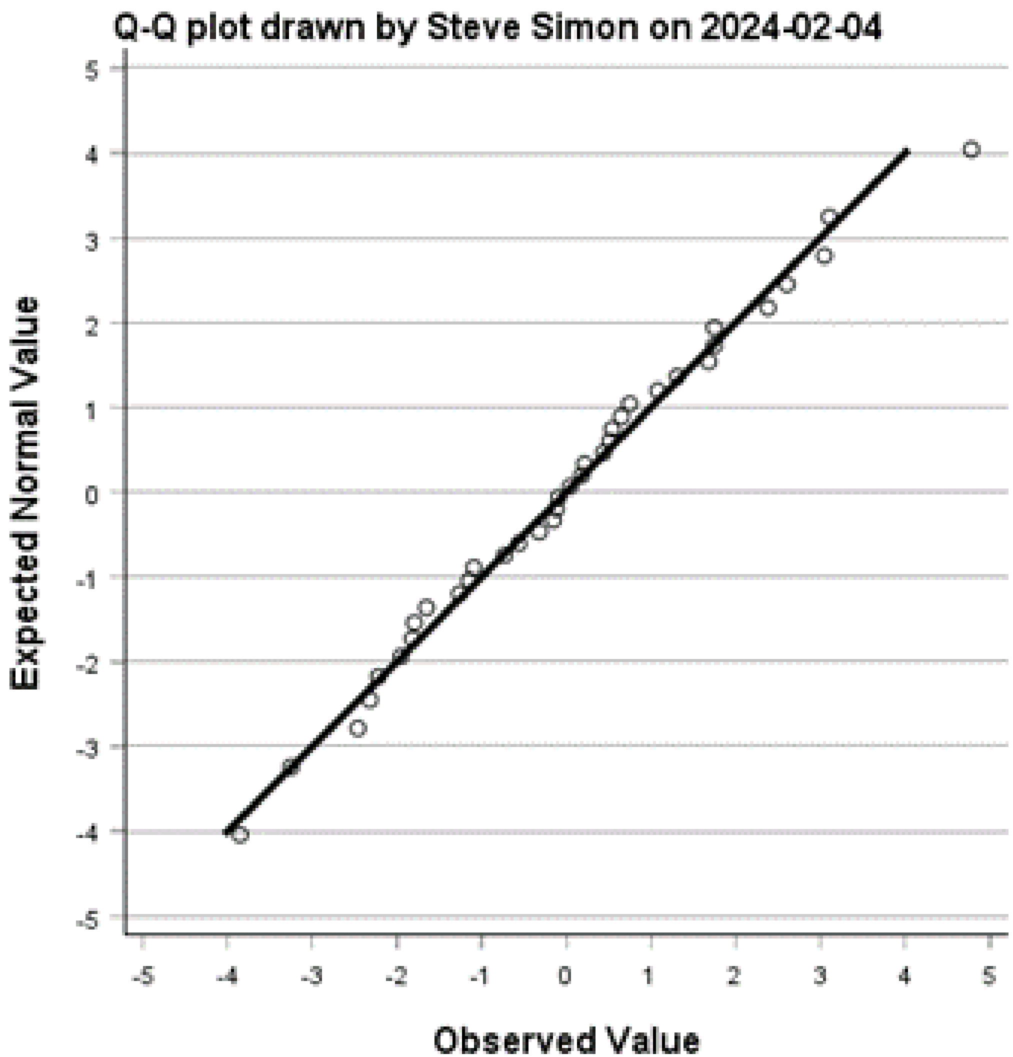
Assumptions

- Normality
- Equal variances
- Independence
- Note: No linearity assumption
    - Only for linear regression and analysis of covariance

Speaker notes

The assumptions for multiple factor analysis of variance are no different than single factor analysis of variance. You must use residuals to check the assumptions of normality and equal variances. The assumption of independence is usually assessed qualitatively.

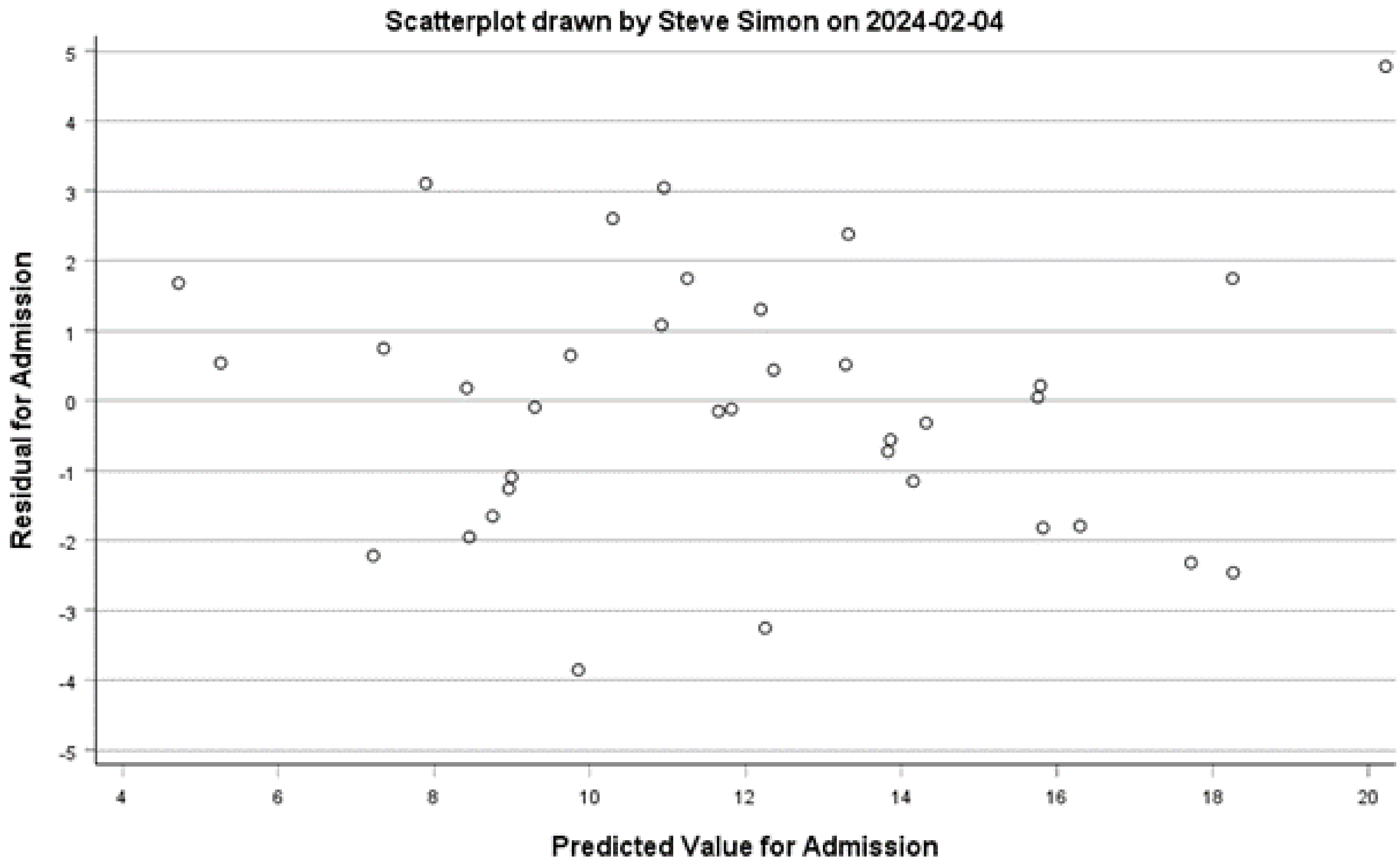
Q-Q plot of residuals



Speaker notes

The residuals from the full moon analysis of variance model appear to be normally distributed.

Residual versus predicted value plot



Speaker notes

There is no evidence of unequal variances.

Live demo, Checking assumptions of analysis of variance

Break #3

- What you have learned
  - Checking assumptions of analysis of variance
- What's coming next
  - Interactions in analysis of variance

What is an interaction

- Impact of one variable is influenced by a second variable
- Example, influence of alcohol on sleeping pills
- Three types of interactions
  - Between two categorical predictors
  - Between a categorical and a continuous predictor
  - Between two continuous predictors
- Interactions greatly complicate interpretation

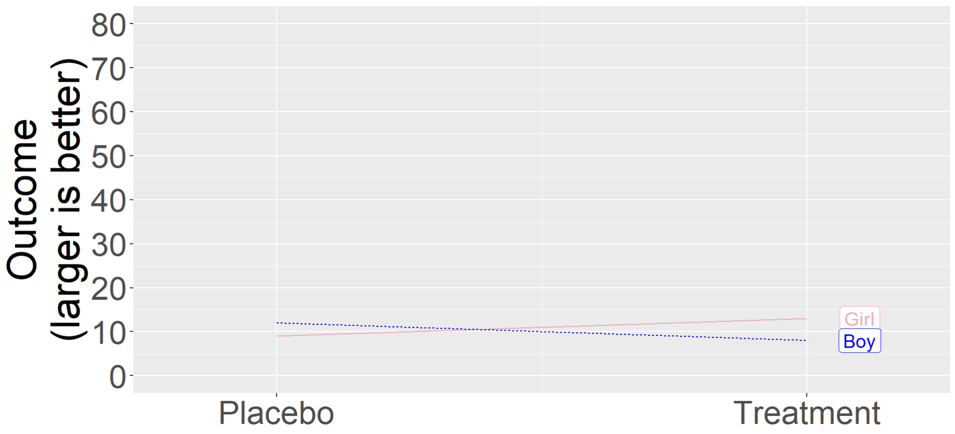
Speaker notes

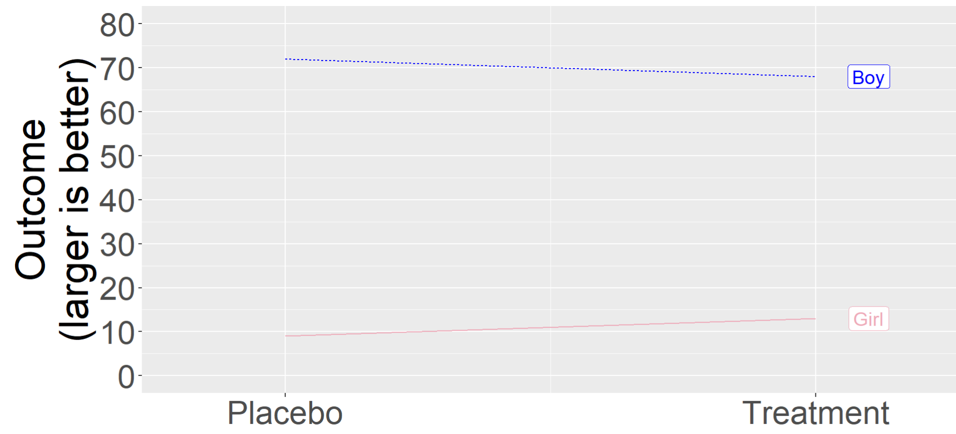
Interactions are important to look for, but if you find one, don’t rejoice. Interactions are a headache. They tell you that a simple interpretation of your research won’t work. That’s important to know, of course, but it also means that you will have to spend more time explaining your results in a paper or presentation.

Interaction plot

- X axis, first categorical variable
- Separate lines for second categorical variable
- Y axis, average outcome

Hypothetical interaction plots, 1 of 4





- No interaction
- Ineffective treatment
- Boys/girls similar

- No interaction
- Ineffective treatment
- Boys fare better than girls

Speaker notes

An interaction plot shows the mean values for each of the two categories. In this example, there is a placebo and a treatment. The outcome is unspecified, but a larger value is presumed to represent a better outcome. This is a pediatric example and the data is subdivided into two populations, boys and girls.

The flatness or steepness of the lines indicates whether patients given the treatment fare better than patients given the placebo.

The separation (if there is any) between the lines measures whether boys fare better or worse than girls.

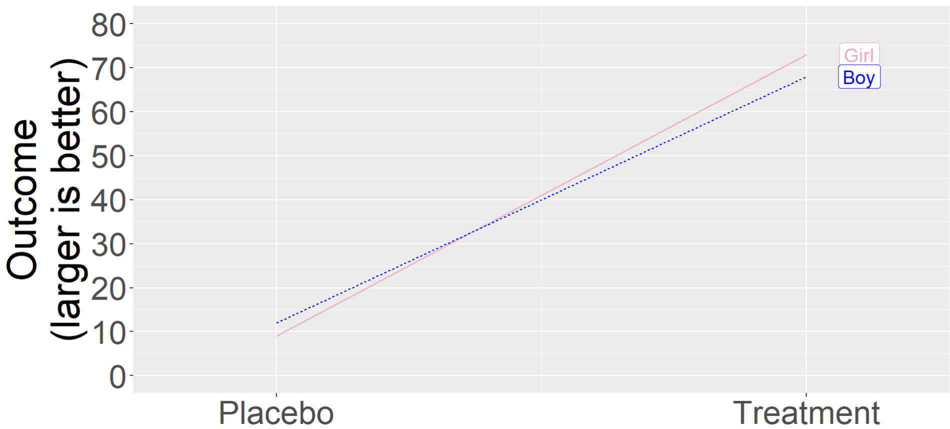
If the lines have roughly the same slope (both are flat or both are steep), then there is no interaction.

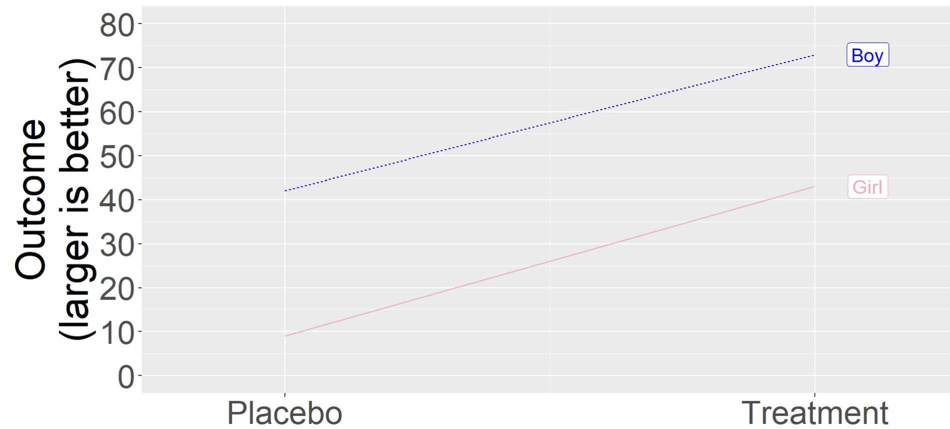
In the plot on the left, the two lines are flat, indicating that the treatment is ineffective. The outcome is not changed from the placebo.

The two lines lie more or less on top of one another. This indicates that there is no difference in average outcome between boys are girls.

In the plot on the right, the two lines are flat. The treatment is ineffective. There is, however, a difference. The average outcome for boys is a lot better both in the placebo group and the treatment group. The lines are roughly parallel, indicating no interaction.

Hypothetical interaction plots, 2 of 4





- No interaction
- Effective treatment
- Boys/girls similar

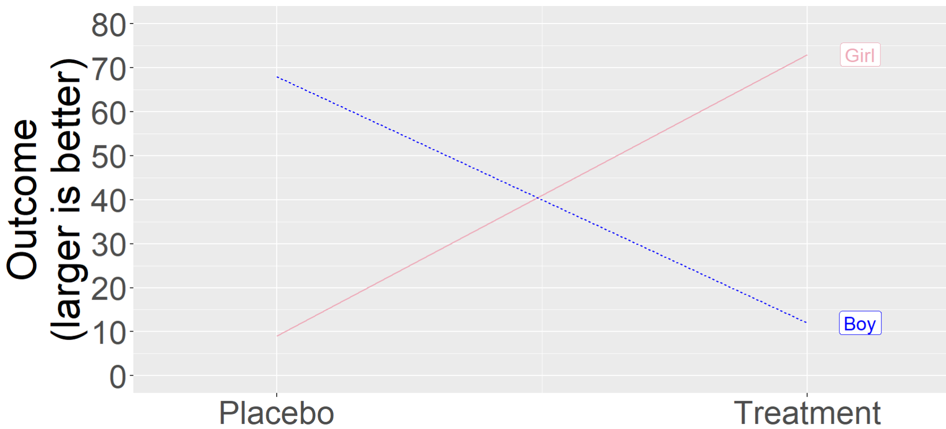
- No interaction
- Effective treatment
- Boys fare better than girls

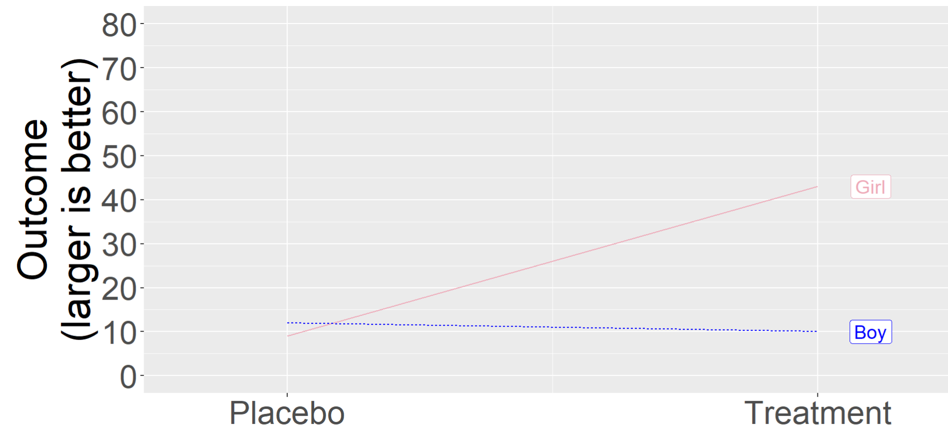
Speaker notes

In the plot on the left, there is a steep slope for both boys and girls. The treatment is effective. There is no separation in the lines. Boys do not fare any better or worse on average than girls.

In the plot on the left, there is a steep slope and a separation between the lines. Boys fare better than girls on average. Both lines have a steep slope. The treatment. The lines are parallel, so there is no interaction.

Hypothetical interaction plots, 3 of 4





- Significant interaction
- Harmful treatment in boys
- Effective treatment in girls

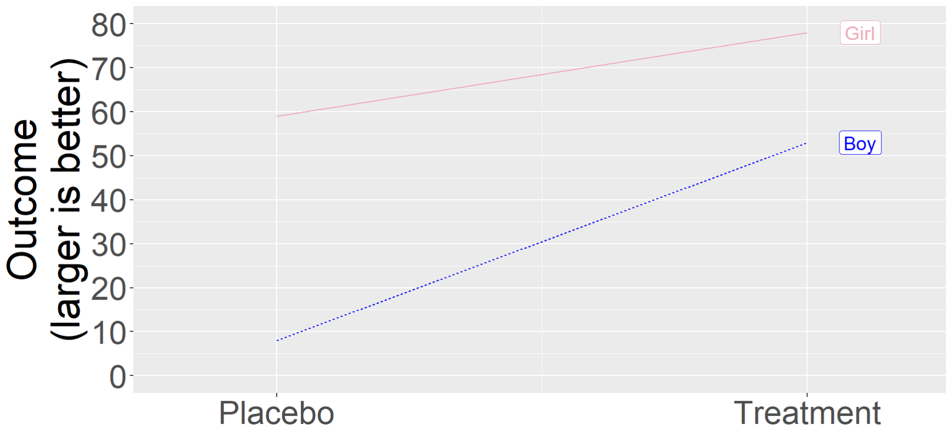
- Significant interaction
- Ineffective treatment in boys
- Effective treatment in girls

Speaker notes

In the plot on the left, the lines are not parallel, so this is evidence of an interaction. In fact, the two lines cross. This is an extreme interaction. Boys fare better on the treatment and girls fare better on the placebo.

In the plot on the right, the lines are not parallel, so this is also evidence of an interaction, but a different sort of interaction. The line for boys is flat and the line for girls is steep. The treatment is worthless for boys, but quite helpful for girls.

Hypothetical interaction plots, 4 of 4



- Significant interaction
- Girls fare better overall
- Effective treatment
- Much more effective in boys

Speaker notes

In this final plot, the lines are not parallel, indicating a third type of interaction. The slope is much steeper for boys. Girls see a moderate improvement on average, but boys see a really large improvement.

Data dictionary for exercise data, 1 of 3

```
data_dictionary: exercise.sas7bdat

description: |
  This dataset is used in a tutorial about interactions. A description from
the original source: The dataset consists of data describing the amount of
weight loss achieved by 900 participants in a year-long study of 3 different
exercise programs, a jogging program, a swimming program, and a reading
program which serves as a control activity. Researchers were interested in how
the weekly number of hours subjects chose to exercise predicted weight loss.
```

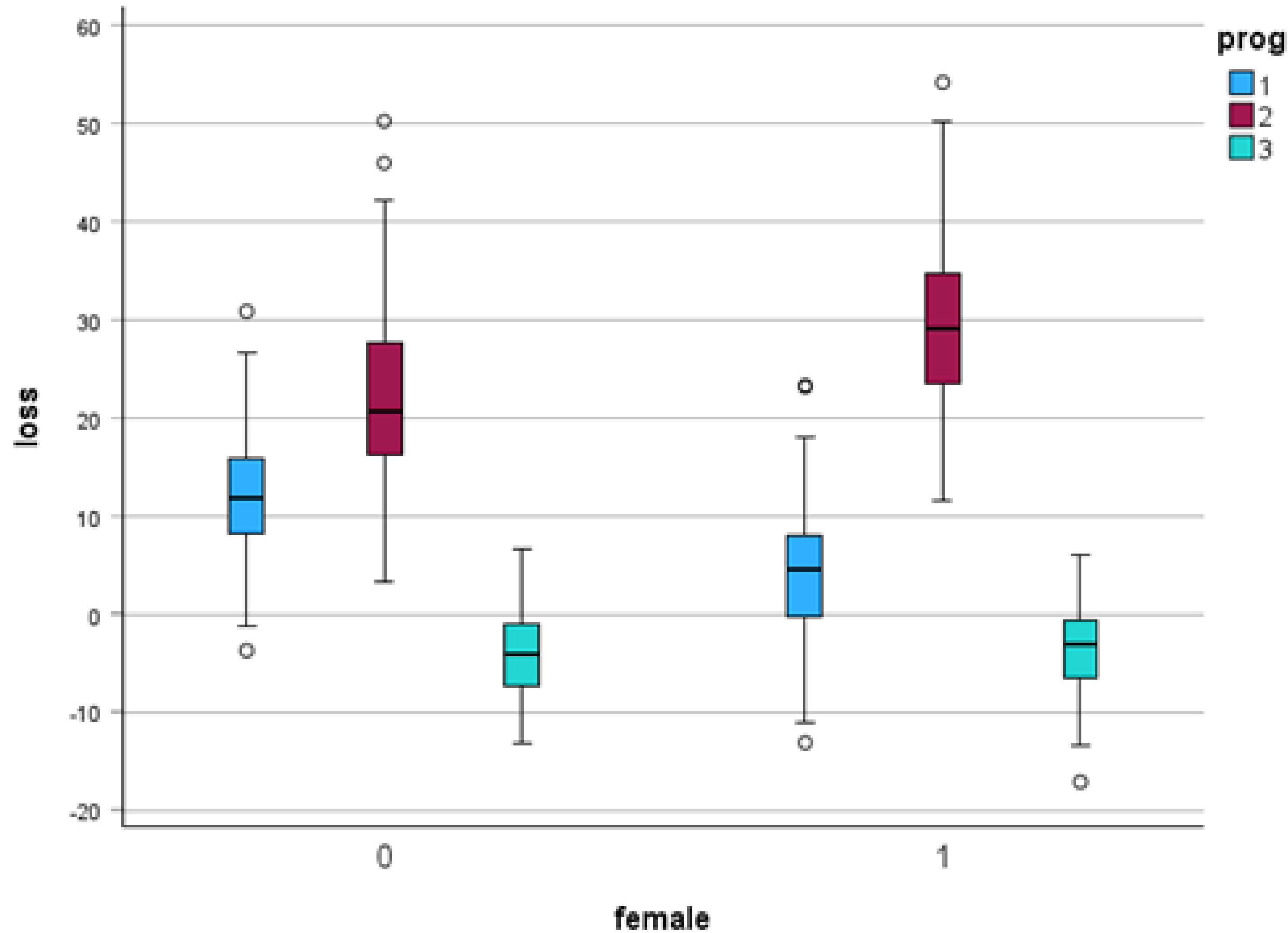
Data dictionary for exercise data, 2 of 3

```
loss:
  label: Average weekly weight loss
  note: negative scores denote weight gain
  scale: real
  unit: unknown, presumably pounds
hours:
  label: weekly average amount of exercise
  unit: hours
effort:
  label: weekly effort scores
  note: self report
  scale: non-negative integer
  range: 0 through 50
  direction: 50 denoting maximum physical effort
```

Data dictionary for exercise data, 3 of 3

```
prog:
  label: exercise program
  note: reading is a control
  values:
    Jogging: 1
    Swimming: 2
    Reading: 3
female:
  scale: binary categorical
  values:
    male: 0
    female: 1
satisfied:
  label: satisfied with weight lost
  scale: binary categorical
  values:
    Dissatisfied: 0
    Satisfied: 1
```

Box plots of exercise data



Speaker notes

Here is a clustered boxplot of weight loss in a study of two exercise interventions: jogging and swimming, and a control intervention, reading. This study shows a clear interaction between gender and exercise. Both jogging and swimming provide greater weight loss than reading (no big surprise). Swimming appears to be better than jogging. But it is a lot better for females and only a little bit better for males.

Mean values for the interaction

Report

loss

female	prog	Mean	N	Std. Deviation
0	1	11.7720	150	5.98881
	2	22.5224	150	8.59176
	3	-3.9556	150	4.14022
	Total	10.1129	450	12.67177
1	1	4.2887	150	6.88500
	2	29.1177	150	7.99691
	3	-3.6201	150	4.09571
	Total	9.9287	450	15.41056
Total	1	8.0304	300	7.45267
	2	25.8200	300	8.91992
	3	-3.7879	300	4.11456
	Total	10.0208	900	14.10024

Speaker notes

In this table of means, notice that men lose about 11 pounds on the jogging program, and 22 pounds on the swimming program. So swimming is better. For women, the losses are about 4 pounds on average with jogging and 30 pounds on swimming. The extra benefits of swimming are so much larger in females than in males.

Analysis of variance table for interaction model

Tests of Between-Subjects Effects

Dependent Variable: loss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	140748.038[a]	5	28149.608	662.463	<.001
Intercept	90375.473	1	90375.473	2126.864	<.001
female	7.634	1	7.634	.180	.672
prog	133277.206	2	66638.603	1568.249	<.001
female * prog	7463.198	2	3731.599	87.818	<.001
Error	37988.163	894	42.492		
Total	269111.674	900			
Corrected Total	178736.201	899			

a. R Squared = .787 (Adjusted R Squared = .786)

Speaker notes

The analysis of variance table shows a large F ratio for the interaction between exercise program and sex.

Parameter estimates for the interaction model

Parameter Estimates

Dependent Variable: loss

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	-3.620	.532	-6.802	<.001	-4.665	-2.576
[female=0]	-.335	.753	-.446	.656	-1.813	1.142
[female=1]	0[a]	.	.	.	.	.
[prog=1]	7.909	.753	10.507	<.001	6.432	9.386
[prog=2]	32.738	.753	43.494	<.001	31.261	34.215
[prog=3]	0[a]	.	.	.	.	.
[female=0] * [prog=1]	7.819	1.064	7.345	<.001	5.730	9.908
[female=0] * [prog=2]	-6.260	1.064	-5.881	<.001	-8.349	-4.171
[female=0] * [prog=3]	0[a]	.	.	.	.	.
[female=1] * [prog=1]	0[a]	.	.	.	.	.
[female=1] * [prog=2]	0[a]	.	.	.	.	.
[female=1] * [prog=3]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Speaker notes

The intercept is the estimated average weight loss when all the indicator variables are equal to zero. That means the two reference categories: females for gender and reading for the exercise group. If you join a book club expect to gain more than three pounds.

The estimate for female=0 is the how much different the weight loss is in the reading group when you change gender from female to male. It's a small change and not statistically significant.

The estimate for group=1 measures how much difference you see in the average weight loss in the jogging program compared to the reading program in the reference category (female). There is almost an 8 pound difference and this is statistically significant.

The estimate for group=2 measures how much difference you see in the average weight loss in the swimming program compared to the reading program in the reference category (female). There is an astounding 32 plus pound difference between swimming and reading, but again for the reference category of females.

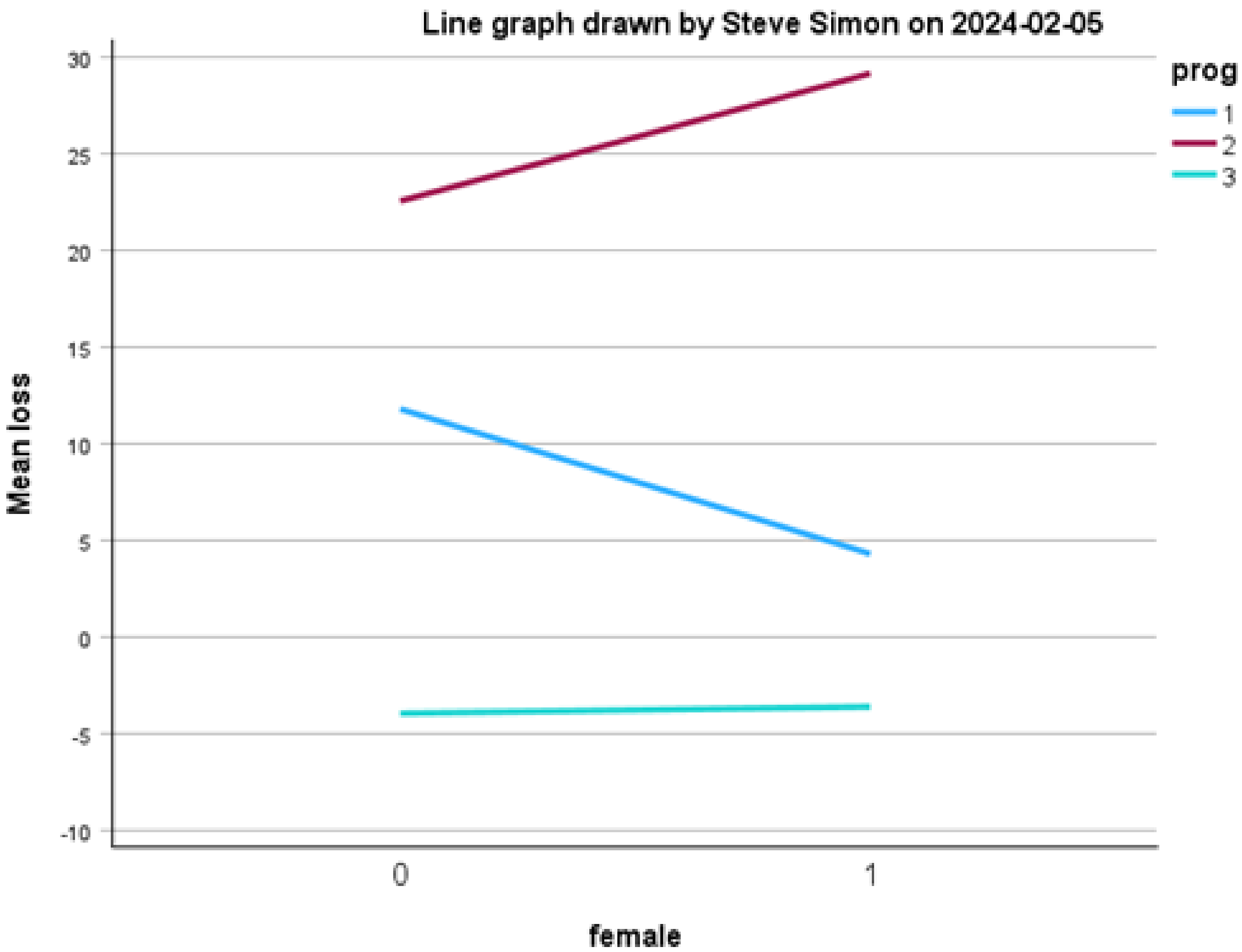
The estimate for female=0*prog=1 shows how much different the benefit of jogging over reading is for men compared to women. Men did pretty well on the jogging program, losing 11 pounds instead of gaining more than 3 pounds. This is more impressive than for women who lost 4 pounds instead of gaining 3.

The estimate for female=0*prog=2 shows how much different the benefit of swimming is for men than for women. Both groups are better off swimming than reading, men losing 22 pounds versus a 3 pound gain. But the falls short of the benefit of swimming over reading for women a 29 pound loss versus a 3 pound gain.

Both of the parameters associated with the interaction are statistically significant.

Bottom line is that the benefits of jogging over reading and the benefits of swimming over reading are not the same for men and women.

Interaction plot, 1 of 2

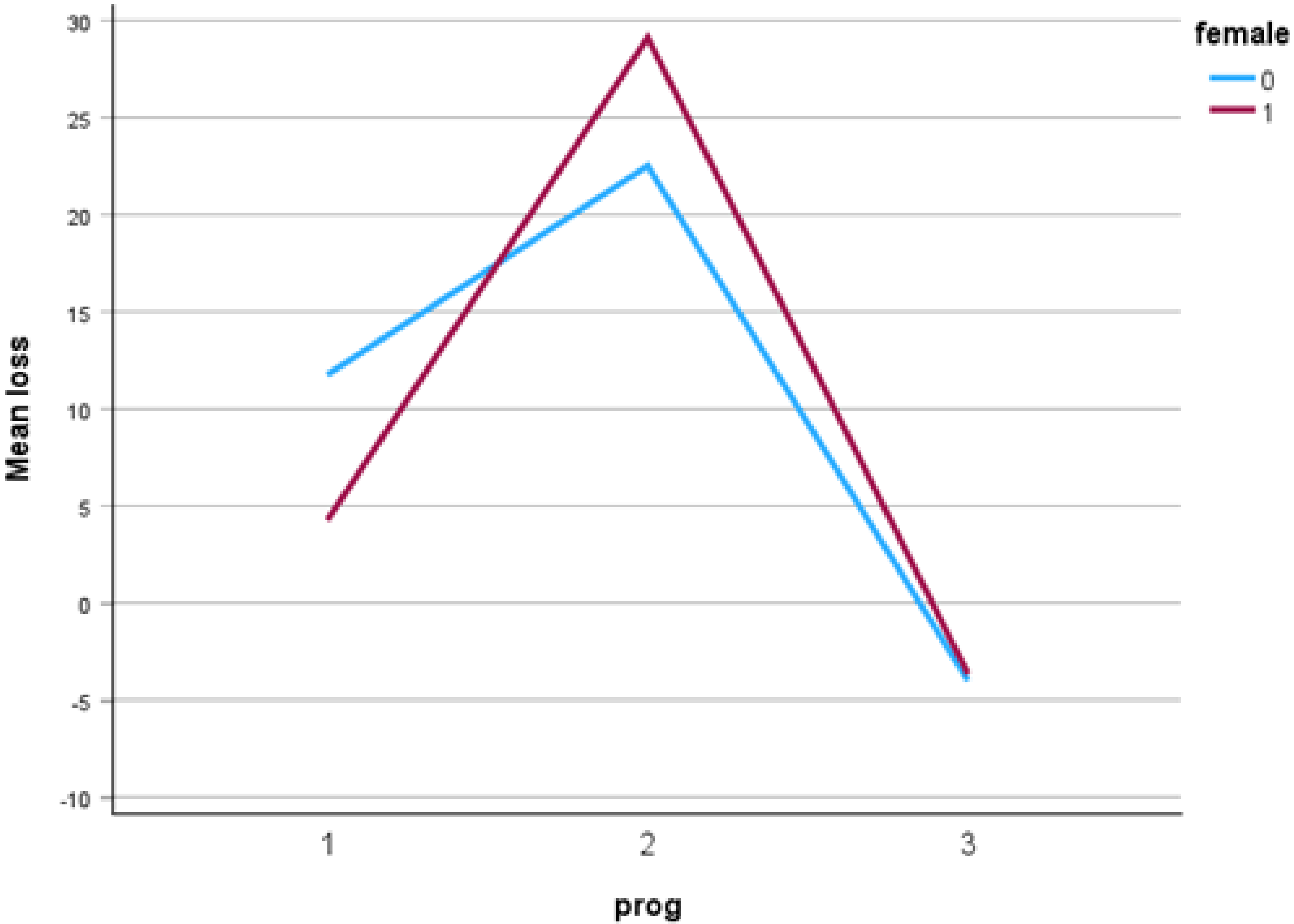


Speaker notes

An interaction plot draws a multiple lines connecting means. In this graph the top line shows the mean weight loss in the swimming program with the men’s mean on the left and the women’s mean on the right. The middle line shows the mean weight loss in the jogging program, again with the men’s mean on the left and the women’s on the right. The bottom line shows the mean weight loss (actually a weight gain!) in the reading program. Clearly the swimming program is the best, a bit better for women than men because the line slopes upward. The jogging program is second best, a bit worse for women than for men. The reading program is worst and the benefit, if you could call it that being about the same for men and women (the lines are flat).

It is a lack of parallelism that is the hallmark of an interaction.

Interaction plot, 2 of 2



Speaker notes

You could draw the interaction plot differently with a line for women (red) and a line for men (blue). The interpretation is about the same, perhaps, but the emphasis is different. The superior weight loss for men in the jogging program and the superior weight loss for women in the swimming program is emphasized by the crossing lines. Both the men's and womens lines almost touching for the reading program emphasizes the equivalent results for the two genders.

Which plot is better? I like the first one, but would not complain if you liked the second one better. It depends on what story you want to emphasize. The first graph emphasizes the difference between the diets a bit more strongly and the second emphasizes the differences between the genders a bit more strongly.

When you can’t estimate an interaction

- Special case, n=1
  - Only one observation for categorical combination

Speaker notes

There is a special case where you have two categorical independent variables and you cannot estimate an interaction. If you have n=1, exactly one observation for each combination of your two categorical variables, then you don’t have enough degrees of freedom to estimate an interaction and still be able to test whether that interaction is statistically significant.

It’s sort of like that old joke I told about married life (it’s okay but you lose a degree of freedom). Interactions cause an even bigger loss of degrees of freedom and in the case with only one observation per combination of categories, you lose enough degrees of freedom that it is not marriage, it being in prison.

Example, full moon study, 1 of 2

Month * Moon Crosstabulation

Count

		Moon			
		After	Before	During	Total
Month	APR	1	1	1	3
	AUG	1	1	1	3
	DEC	1	1	1	3
	FEB	1	1	1	3
	JAN	1	1	1	3
	JUL	1	1	1	3
	JUN	1	1	1	3
	MAR	1	1	1	3
	MAY	1	1	1	3
	NOV	1	1	1	3
	OCT	1	1	1	3
	SEP	1	1	1	3
Total		12	12	12	36

Speaker notes

Here is an example where you only have one observation for each combination of categories.

Example, full moon study, 2 of 2

Tests of Between-Subjects Effects

Dependent Variable: Admission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Moon	41.514	2	20.757	.	.
Month	455.583	11	41.417	.	.
Moon * Month	127.819	22	5.810	.	.
Error	.000	0	.		
Corrected Total	624.916	35			

Speaker notes

You lose two degrees of freedom for moon (3 phases: before, during, and after). You lose 11 degrees of freedom for month (12 months -1). The interaction has 2 times 11 or 22 degrees of freedom. You only started with 35 degrees of freedom. Subtract 2, 11, and 22, and you are left with zero degrees of freedom for error.

If you find yourself in this situation, just state that no test for interaction was possible in your methods section and highlight this as a weakness of your study in the discussion section.

Live demo, Interactions in analysis of variance

Break #4

- What you have learned
  - Interactions in analysis of variance
- What's coming next
  - Interactions in analysis of covariance

A second type of interaction

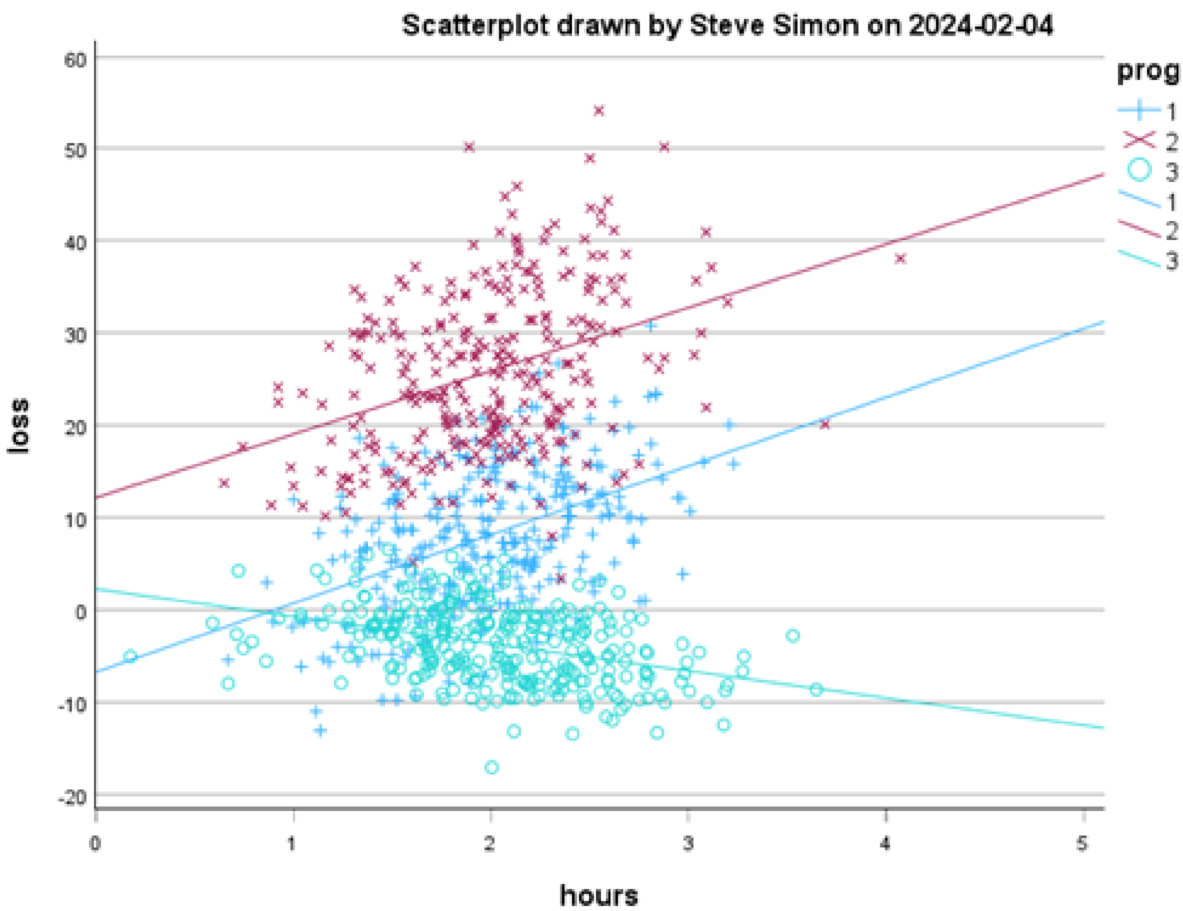
- Interactions in analysis of covariance
  - Between categorical predictor and continuous predictor
  - Different slopes within each category

Speaker notes

Last week, I touched briefly on the interaction in an analysis of covariance model. That model makes an assumption that the effect of a covariate on the outcome is the same in both the treatment group and the control group. That implies that the trend lines are parallel.

If there is an interaction, then the slope differs within each category. It might be steeper in the treatment group for example, and shallower in the control group.

Interaction between exercise program and hours spent exercising



Speaker notes

This plot shows a marked interaction. If you looked just at prog 1 (jogging program) and prog 2 (swimming program), there is no interaction. The more hours you spent on the program the more weight you lost.

The reading program, however, has a different slope. Instead of increasing steadily, the more time you spent reading the less weight that you lost. You might even gain some weight if you followed the reading program religiously.

Testing for interaction in analysis of covariance

Tests of Between-Subjects Effects

Dependent Variable: loss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	140940.743[a]	5	28188.149	666.752	<.001
Intercept	326.802	1	326.802	7.730	.006
prog	2901.898	2	1450.949	34.320	<.001
hours	3116.205	1	3116.205	73.710	<.001
prog * hours	5319.048	2	2659.524	62.907	<.001
Error	37795.457	894	42.277		
Total	269111.674	900			
Corrected Total	178736.201	899			

a. R Squared = .789 (Adjusted R Squared = .787)

Speaker notes

This table shows a statistically significant interaction. The F ratio is large and the p-value is small.

Table with irrelevant rows removed

Tests of Between-Subjects Effects

Dependent Variable: loss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	140940.743[a]	5	28188.149	666.752	<.001
[redacted]					
prog	2901.898	2	1450.949	34.320	<.001
hours	3116.205	1	3116.205	73.710	<.001
prog * hours	5319.048	2	2659.524	62.907	<.001
Error	37795.457	894	42.277		
[redacted]					
Corrected Total	178736.201	899			

a. R Squared = .789 (Adjusted R Squared = .787)

Speaker notes

Remove the rows associated with the intercept and the (uncorrected) total.

Parameter estimates

Parameter Estimates

Dependent Variable: loss

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	2.216	1.486	1.491	.136	-.700	5.133
[prog=1]	-8.997	2.216	-4.060	<.001	-13.346	-4.648
[prog=2]	9.933	2.177	4.562	<.001	5.660	14.205
[prog=3]	0[a]	.	.	.	.	.
hours	-2.956	.708	-4.176	<.001	-4.346	-1.567
[prog=1] * hours	10.409	1.072	9.708	<.001	8.304	12.513
[prog=2] * hours	9.830	1.051	9.349	<.001	7.767	11.894
[prog=3] * hours	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

- Intercept for prog=1, -8.997 + 2.216 = -6.781
- Intercept for prog=2, 9.993 + 2.216 = 12.209
- Intercept for prog=3, 2.216
- Slope for prog=1, 10.409 + -2.956 = 7.453
- Slope for prog=2, 9.83 + -2.956 = 6.874
- Slope for prog=3, -2.956

Speaker notes

The first three coefficients are a bit of an extrapolation because they represent patients who spend zero hours on the proposed intervention.

The intercept is the estimated average weight loss in the reference category (group=3 or reading program) when the number of hours devoted to the program is zero. In this case it is a weight gain of 2 pounds.

The coefficient for group=1 is how much better the weight loss is when you switch from the reading program to the jogging program, but still put in zero hours.

The coefficient for group=2 is how much better the weight loss is on average when you switch from the reading program to the jogging program, but still put in zero hours.

The coefficient for hours is how much better the weight loss is when you add an extra hour of effort and you are in the reading program. Actually, it is how much worse. Each extra hour of reading and you gain an extra three pounds on average.

The coefficient for hours*group=1 shows the change in slope when you switch from the reading program to the jogging program. Each hour invested is 10 pounds better (a loss of 7 pounds per hour invested instead of a gain of 3 pounds per hour invested) when you switch from the reading program to the jogging program.

The interpretation for hours*group=2 is similar It shows the change in slope when you switch from the reading program to the swimming program. Each hour invested is 10 pounds better (a loss of 7 pounds per hour invested instead of a gain of 3 pounds per hour invested) when you switch from the reading program to the swimming program.

Live demo, Interactions in analysis of covariance

Break #5

- What you have learned
  - Interactions in analysis of covariance
- What's coming next
  - Interactions in multiple linear regression

Analysis of variance table

Parameter Estimates

Dependent Variable: loss

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	7.799	11.604	.672	.502	-14.975	30.572
hours	-9.376	5.664	-1.655	.098	-20.492	1.740
effort	-.080	.385	-.209	.835	-.835	.675
hours * effort	.393	.188	2.098	.036	.025	.761

Speaker notes

Table of means

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
hours	900	.18	4.07	2.0024	.49454
effort	900	12.95	44.08	29.6592	5.14276
Valid N (listwise)	900				

Centered analysis

Parameter Estimates

Dependent Variable: loss

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	10.005	.452	22.130	<.001	9.117	10.892
hours_centered	2.291	.915	2.503	.012	.495	4.087
effort_centered	.707	.088	8.042	<.001	.535	.880
hours_centered * effort_centered	.393	.188	2.098	.036	.025	.761

Speaker notes

Weight loss at various conditions

- hours = 2 (mean), effort = 30 (mean),
  - $\hat{Y}$ = 10.005
- hours = 4 (mean+2), effort = 30 (mean),
  - $\hat{y}$ = `10.005 + 2.291*2` = 14.587
- hours = 2 (mean), effort = 40 (mean+20)
  - $\hat{Y}$ = `10.005 + 0.707*20` = 24.145
- hours = 4 (mean+2), effort = 40 (mean+20)
  - $\hat{Y}$ = `10.005 + 2.291*2 + 0.707*20 + 0.393*2*20` = 44.447

Live demo, Interactions in multiple linear regression

Break #6

- What you have learned
  - Interactions in multiple linear regression
- What's coming next
  - Unbalanced data

FEV data

Descriptive Abstract: Sample of 654 youths, aged 3 to 19, in the area of East Boston during middle to late 1970's. Interest concerns the relationship between smoking and FEV. Since the study is necessarily observational, statistical adjustment via regression models clarifies the relationship.

- fev continuous measure (liters)
- sex discrete/nominal (Female coded 0, Male coded 1)
- smoke discrete/nominal (Nonsmoker coded 0, Smoker coded 1)
  - Source: https://jse.amstat.org/datasets/fev.txt

Line plots of means for unbalanced data

Speaker notes

Table of means

Report

fev

sex	smoke	Mean	N	Std. Deviation
Female	0	2.38	279	.64
	1	2.97	39	.42
	Total	2.45	318	.65
Male	0	2.73	310	.97
	1	3.74	26	.89
	Total	2.81	336	1.00
Total	0	2.57	589	.85
	1	3.28	65	.75
	Total	2.64	654	.87

Speaker notes

Table of frequencies and column percentages

sex * smoke Crosstabulation

			smoke 0	smoke 1	Total
sex	Female	Count	279	39	318
		% within smoke	47.4%	60.0%	48.6%
	Male	Count	310	26	336
		% within smoke	52.6%	40.0%	51.4%
Total		Count	589	65	654
		% within smoke	100.0%	100.0%	100.0%

Speaker notes

Live demo, Unbalanced data

Summary

- What you have learned
  - Indicator variables for three or more categories
  - Multiple factor analysis of variance
  - Checking assumptions of analysis of variance
  - Interactions in analysis of variance
  - Interactions in analysis of covariance
  - Interactions in multiple linear regression
  - Unbalanced data

Additional topics??

Speaker notes

- Learning objectives for multifactor ANOVA
  - Define Factor
  - Define Level
  - Assess types of ANOVA based on factors and levels
  - Calculate the maximum number of groups in a multi-factorial ANOVA
  - Define interaction
  - Explain how to determine if an interaction between the factors exists
  - Explain how to “name” a multi-factorial ANOVA
  - Write the Null and Alternative Hypothesis for a multi-factorial ANOVA
  - Write the Research Questions for a multi-factorial ANOVA
  - Write the Research Interaction Questions for a multi-factorial ANOVA
  - Define and explain the order of interpretation for a multi-factorial ANOVA
  - Define Main Effects
  - Explain the test of the Main Effects
  - Define Simple Main Effects
  - Explain the test of the Simple Main Effects
  - Define Interaction Effect
  - Explain the test of the Interaction Effect
  - Explain what information the Interaction Effect test provides
  - Explain the difference in information obtained between a 2-way ANOVA and 2 one way ANOVA’s
  - Define the “Differences Between the Differences”
  - Calculate the “Differences Between the Differences”
  - Draw conclusions about the “Differences Between the Differences”
  - Interpret the Interaction Effect test
  - Determine the next steps of interpretation with a significant interaction effect
  - Determine the next steps of interpretation with a non-significant interaction effect
  - Interpret the Main Effects
  - Interpret the Simple Main Effects
  - List and define the Assumptions and tests for a multi-factorial ANOVA
  - Define Marginal Means
  - Explain the difference between Marginal Means and Means

- Explain the difference between Marginal Means and Means
- Complete a statistical write-up of a multi-factorial ANOVA with all elements included