Associations between categorical variables

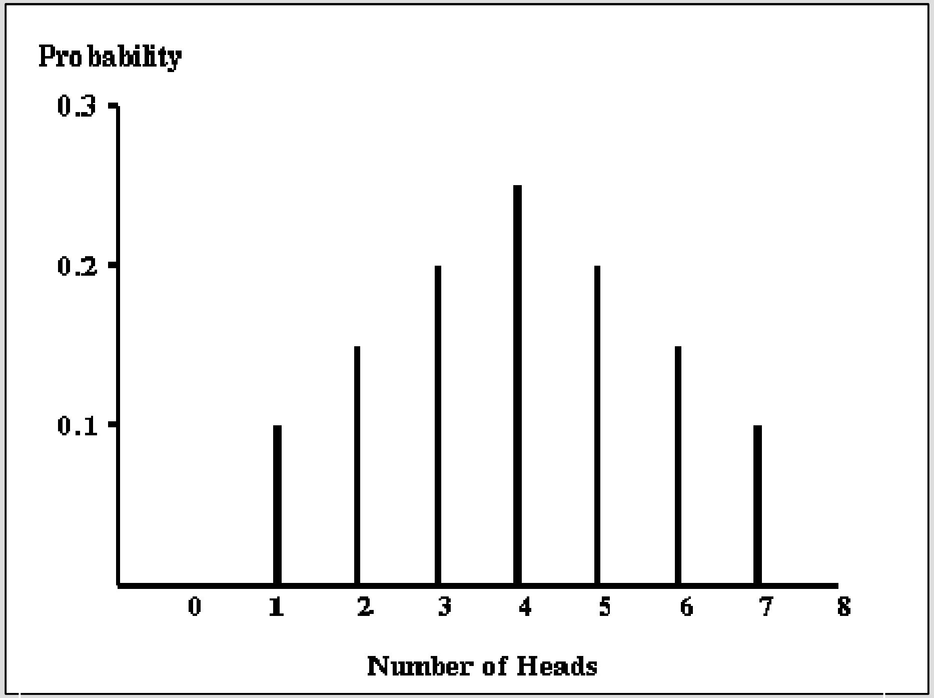
Binomial test

A linguist has collected a sample of sentences including ditransitive verbs from a corpus. Overall, there are 46 sentences in his sample. In 27 sentences the verb occurs with two NP objects, in 19 sentences the verb occurs with an NP and a PP.

(1)	He gives Peter the ball.	V NP NP	(N=27)
(2)	He gives the ball to Peter.	V NP PP	(N=19)

Is the difference in frequency between two categories is significant?

Binomial test



The binomial test is an exact test.

The binomial test is a one sample test.

Binomial test

Construction	Frequency
V NP NP	27
V NP PP	19

Null-hypothesis: The two constructions are equally frequent (suggesting that they are free variants; i.e. there is nothing to explain).

Alternative hypothesis: The two constructions differ in frequency (which must have a reason that needs to be explained).

Binomial test

Test auf Binomialverteilung

		Kategorie	N	Beobachteter Anteil	Testanteil	Asymptotische Signifikanz (2-seitig)
Frequency	Gruppe 1	27,00	27	,59	,50	,302[a]
	Gruppe 2	19,00	19	,41		
	Gesamt		46	1,00		

a. Basiert auf der Z-Approximation.

Example

A researcher examined the effectiveness of two new drugs on chronic pain. The first drug was given and pain assessed (Pain1) then a month later the second drug was given and again pain assessed (Pain2). Following the study the researcher wants to know if the proportions of men and women in the sample used were what would be expected by chance.

Case	Sex	Pain 1	Pain 2
1	male	2	2
2	male	5	3
3	male	2	2
4	male	3	2
5	female	4	2
6	male	5	4
7	female	2	3
8	male	3	3
9	male	4	4
10	female	1	2
11	female	2	4
12	male	3	3
13	male	4	2
14	female	2	1
15	male	3	1

Case	Sex	Pain 1	Pain 2
1	male	2	2
2	male	5	3
3	male	2	2
4	male	3	2
5	female	4	2
6	male	5	4
7	female	2	3
8	male	3	3
9	male	4	4
10	female	1	2
11	female	2	4
12	male	3	3
13	male	4	2
14	female	2	1
15	male	3	1

Case	Sex	Pain 1	Pain 2
1	male	2	2
2	male	5	3
3	male	2	2
4	male	3	2
5	female	4	2
6	male	5	4
7	female	2	3
8	male	3	3
9	male	4	4
10	female	1	2
11	female	2	4
12	male	3	3
13	male	4	2
14	female	2	1
15	male	3	1

X^2 test for goodness-of-fit

X^2 test for goodness-of-fit

The Binomial test is restricted to variables with two levels. If there are more than two levels we use the X^2 test for goodness-of-fit is.

The X^2 test is not an exact test and has certain preconditions:

> You must not have more than 25% of cells with an expected frequency of less than 5.

X^2 test for goodness-of-fit

A linguist has collected relative clauses from a corpus, which he divided into four types: (1) subjects relatives, (2) object relatives, and (3) oblique relatives, (4) genitive relatives. Is the sample difference sufficient to assume that the four types of relative clause differ in frequency in the true population?

	Subject	Object	Oblique	Genitive	Total
Freq	55	53	39	4	151
Exp.					

X^2 test for goodness-of-fit

A linguist has collected relative clauses from a corpus, which he divided into four types: (1) subjects relatives, (2) object relatives, and (3) oblique relatives, (4) genitive relatives. Is the sample difference sufficient to assume that the four types of relative clause differ in frequency in the true population?

	Subject	Object	Oblique	Genitive	Total
Freq	55	53	39	4	151
Exp.	37.75	37.75	37.75	37.75	

X^2 test for goodness-of-fit

Null-hypothesis:

The four types of relative clauses are equally frequent in the true population.

Alternative hypothesis:

The four types of relative clauses are not equally frequent in the true population.

X^2 test for goodness-of-fit

$$\Sigma \; \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4					

X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4	37.75 37.75 37.75 37.75				

X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4	37.75 37.75 37.75 37.75	17.25 15.25 1.25 -33.75			

X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4	37.75 37.75 37.75 37.75	17.25 15.25 1.25 -33.75	297.56 232.56 1.56 1139.06		

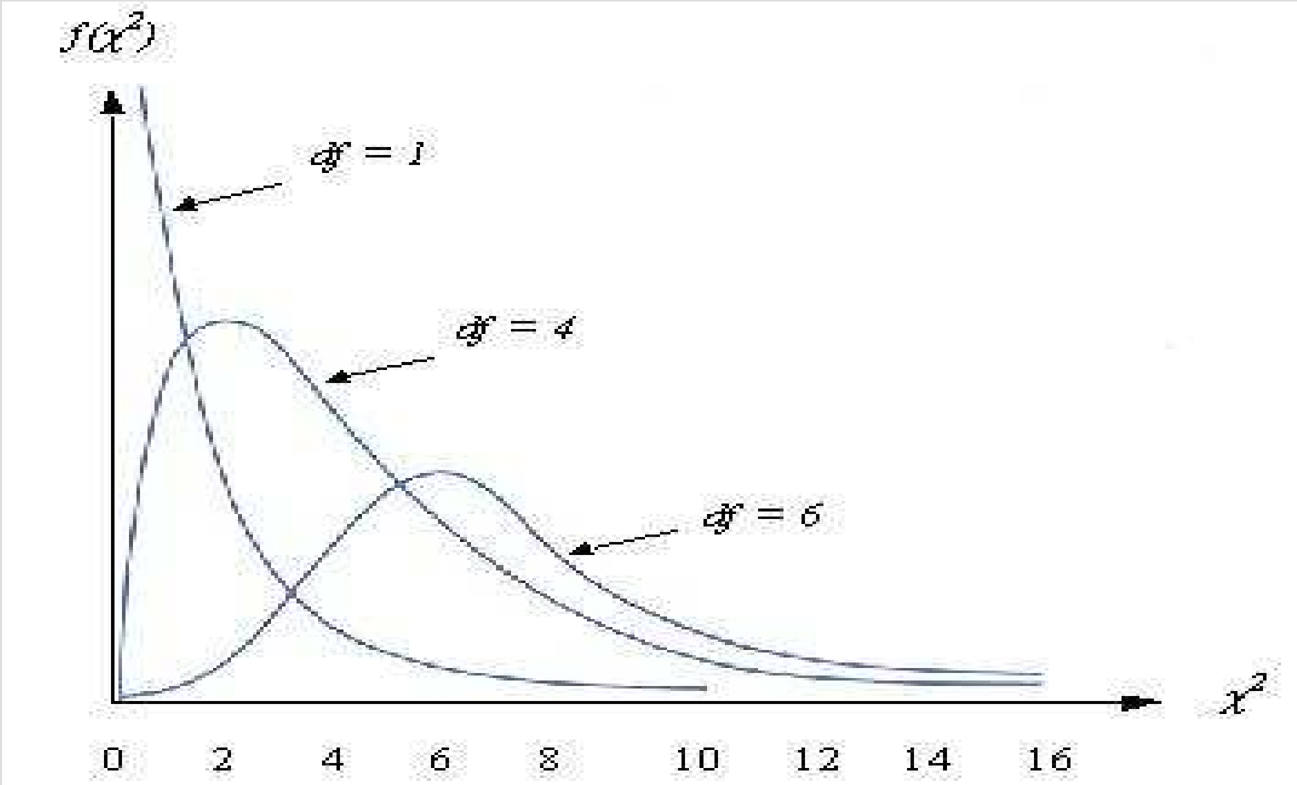
X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4	37.75 37.75 37.75 37.75	17.25 15.25 1.25 -33.75	297.56 232.56 1.56 1139.06	1670	

X^2 test for goodness-of-fit

Observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55 53 39 4	37.75 37.75 37.75 37.75	17.25 15.25 1.25 -33.75	297.56 232.56 1.56 1139.06	1670	$\chi 2$ = 44.25

X^2 test for goodness-of-fit



df = [number of levels] – [1]

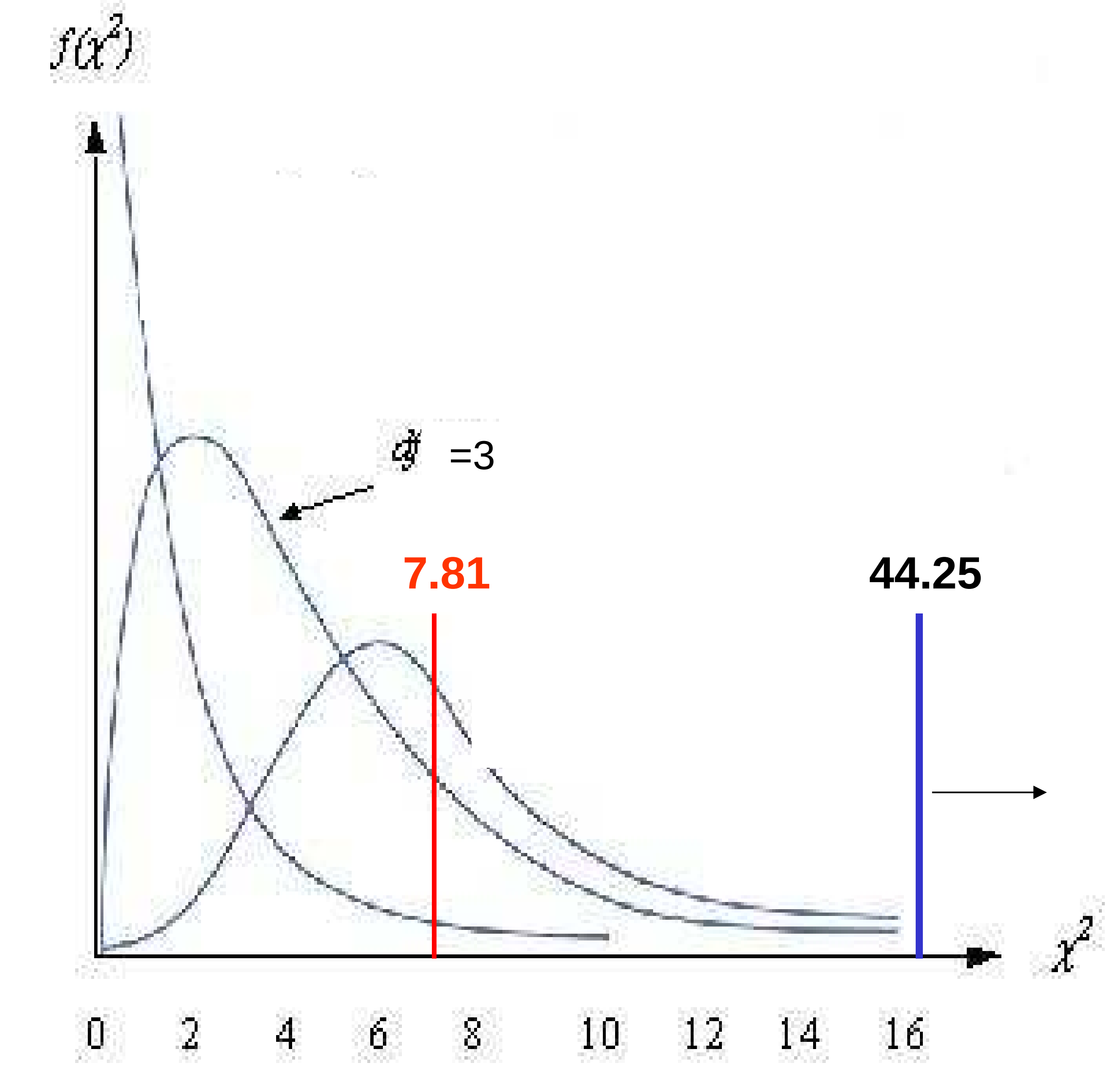
The Chi-Square (χ^2) Distribution

Degrees of freedom	Area to the Right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Donald B. Owen, *Handbook of Statistical Tables,* U.S. Department of Energy (Reading, Mass.: Addison-Wesley, 1962). Reprinted with permission of the publisher.

X^2 table

	.995	.99	.975	.95	.90	.10	**.05**	.025	.01	.005
1 df										
2 df										
3 df	0.072	0.115	0.216	0.352	0.584	6.25	**7.81**	9.35	11.34	12.84
4 df										



$f(\chi^2)$
df =3
7.81
44.25
χ^2
0
2
4
6
8
10
12
14
16

X^2 test for independence

X^2 test for independence

A linguist wants to find out if subject and object are expressed by the same type of nouns. Specifically, he wants to know if lexical and pronominal NPs are equally distributed in subject and object NPs. In order to test this hypothesis, he collected the following frequency data from a small corpus.

	Subject	**Object**	**Total**
Pronominal	47	17	64
Lexical	41	52	93
Total	88	69	157

X^2 test for independence

$$\text{Exp} = \frac{(\text{marginal} \times \text{marginal})}{\text{grand total}}$$

X^2 test for independence

	Subject	Object	Total
Pro	**47** 64×88/157 = **…**	**17** 64×69/157 = **…**	64
Lex	**41** 93×88/157 = **…**	**52** 93×69/157 = **…**	93
Total	88	69	157

X^2 test for independence

	Subject	Object	Total
Pro	**47** (35.9)	**17** (28.1)	64
Lex	**41** (52.1)	**52** 9(40.9)	93
Total	88	69	157

X^2 test for independence

$$\Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

X^2 test for independence

$$\chi^2 = \frac{(47-39.9)^2}{35.9} + \frac{(17-28.1)^2}{28.1} + \frac{(41-52.1)^2}{52.1} + \frac{(52-40.9)^2}{40.9} = 13.18$$

df = [row - 1] $\times$ [column - 1]

Review

χ^2 test

- Explain the difference between the χ^2 test for goodness of fit and the χ^2 test for independence.
- A reseascher collected a sample of adverbial clauses from a spoken corpus of English. The sample includes 52 conditional clauses, 68 causal clauses, and 82 temporal clauses. Are these data sufficient to claim that temporal clauses are more frequent than conditional and causal clauses?

χ^2 test

	Conditional	Causal	Temporal
Frequency	52	68	82

χ^2 test

	Conditional	Causal	Temporal
Frequency	52	68	82
Exp. Freq.	67.33	67.33	67.33

Statistik für Test

	Freq
Chi-Quadrat	6,693[a]
df	2
Asymptotische Signifikanz	,035

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 67,3.

χ^2 test

- Explain the difference between the χ^2 test for goodness of fit and the χ^2 test for independence.
- A reseascher collected a sample of adverbial clauses from a spoken corpus of English. The sample includes 52 conditional clauses, 68 causal clauses, and 82 temporal clauses. Are these data sufficient to claim that temporal clauses are more frequent than conditional and causal clauses?
- In a second step the researcher collects an additional sample of causal and temporal clauses from a written corpus. This time causal clauses (N = 103) are more frequent than temporal clauses (N = 79). Does the frequency of causal and temporal clauses vary with register (i.e. spoken vs. written). Determine the expected frequencies and submit the data to statistical analysis.

χ^2 test

	Causal	Temporal	Total
Spoken	68	62	130
Written	103	79	182
Total	171	141	212

AC * Register Kreuztabelle

			Register		
			spoken	written	Gesamt
AC	causal	Anzahl	68	103	171
		Erwartete Anzahl	65,1	105,9	171,0
	temporal	Anzahl	82	141	223
		Erwartete Anzahl	84,9	138,1	223,0
Gesamt		Anzahl	150	244	394
		Erwartete Anzahl	150,0	244,0	394,0

χ^2 test

	Causal	Temporal	Total
Spoken	68	62	130
Written	103	79	182
Total	171	141	212

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,368[a]	1	,544		
Kontinuitätskorrektur[b]	,252	1	,616		
Likelihood-Quotient	,368	1	,544		
Exakter Test nach Fisher				,601	,308
Anzahl der gültigen Fälle	394				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 65,10.

b. Wird nur für eine 2x2-Tabelle berechnet

χ^2 test

	Causal	Temporal	Total
Spoken	68	62	130
Written	103	79	182
Total	171	141	212

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,031	,544
	Cramer-V	,031	,544
Anzahl der gültigen Fälle		394	

X^2 test for independence

1. Each subject provides a score for only one cell
2. None of the cells is empty
3. Not more than 25% of the cells has an expected frequency of less than 5 (which is one cell in a 2×2 table)

Alternative: Fisher Exact

X^2 test for independence: r×c

Is there a correlation between smoking and drinking? To examine this question 337 subjects were divided into the following groups.

	Smoker			
Drinker	**Heavy**	**Light**	**Non**	**Total**
Heavy	33	32	35	100
Light	56	23	34	113
Non	42	28	54	124
Total	131	83	123	337

McNemar

McNemar

Does the ambient language influence the acquisition of grammatical constructions? In order to examine this question a child language researcher asked 100 German-speaking children to repeat a ditransitive sentence including10 words (e.g. *Der Mann gibt dem kleinen Jungen einen sehr großen Ballon*). The children have to repeat the sentences at two different times: First at the beginning of the study and second after a training phase during which they encounter ten ditransitive constructions embedded in a one hour conversation.

McNemar

Results:

- 31 children repeated the ditransitive sentence correctly both before and after ‘training’.
- 17children repeated the ditransitive sentence incorrectly both before and after ‘training’.
- 39 children repeated the ditransitive sentence incorretly before training and correctly after ‘training’.
- 13 children repeated the ditransitive sentence correctly before training and incorrectly after ‘training’.

Does training influence the children‘s performance?

McNemar

		prior		
		Correct	**False**	**Total**
posterior	**Correct**	31	39	70
	False	13	17	30
	Total	44	56	100

McNemar

Extensions of McNemar:

Bowker: The DV has more than two levels (correct – partly correct - false).

Cochran Q: Subjects are tested multiple times (i.e. at least three times).

Overview of statistical tests

Correlation + Regression

	Correlation	Regression
Nominal		
Ordinal		
Interval		

Correlation + Regression

	Correlation	**Regression**
Nominal		
Ordinal		
Interval	1. Pearson’s r	

Correlation + Regression

	Correlation	Regression
Nominal		
Ordinal	1. Spearman's Rho 2. Kendall's Tau	
Interval	1. Pearson's r	

Correlation + Regression

	Correlation	**Regression**
Nominal	1. Pearson $\chi 2$ test 2. Phi coefficient 3. Cramer's V	
Ordinal	1. Spearman's Rho 2. Kendall's Tau	
Interval	1. Pearson's r	

Correlation + Regression

	Correlation	**Regression**
Nominal	1. Pearson $\chi 2$ test 2. Phi coefficient 3. Cramer's V	
Ordinal	1. Spearman's Rho 2. Kendall's Tau	
Interval	1. Pearson's r	1. Simple bivariate regression 2. Multiple regression

Correlation + Regression

	Correlation	**Regression**
Nominal	1. Pearson $\chi 2$ test 2. Phi coefficient 3. Cramer's V	1. Logistic regression 2. Discriminant analysis
Ordinal	1. Spearman's Rho 2. Kendall's Tau	
Interval	1. Pearson's r	1. Simple bivariate regression 2. Multiple regression

Frequency data

	1×k Table	2×2 Table	
	1 sample test	Within (vorher-nachher)	Between
Nominal			

Frequency data

	1×k Table	2×2 Table	
	1 sample test	**Within (vorher-nachher)**	**Between**
Nominal	1. Binomial 2. χ2 goodness of fit		

Frequency data

	1×k Table	2×2 Table	
	1 sample test	**Within (vorher-nachher)**	**Between**
Nominal	1. Binomial 2. $\chi 2$ goodness of fit 3. Runs test 4. Kolmogorov-Smirnov		

Frequency data

	1×k Table	**2×2 Table**	
	1 sample test	**Within (vorher-nachher)**	**Between**
Nominal	1. Binomial 2. $\chi 2$ goodness of fit 3. Runs test 4. Kolmogorov-Smirnov	McNemar	

Frequency data

	1×k Table	2×2 Table	
	1 sample test	**Within (vorher-nachher)**	**Between**
Nominal	1. Binomial 2. $\chi2$ goodness of fit 3. Runs test 4. Kolmogorov-Smirnov	McNemar	1. $\chi2$ of independ. 2. Fischer exact

Frequency data

k×k Table		k×k×k Table	
Within (vorher-nachher)	**Between**	**Within**	**Between**

Frequency data

k×k Table		k×k×k Table	
Within (vorher-nachher)	**Between**	**Within**	**Between**
1. Bowker 2. Cochran Q			

Frequency data

k×k Table		**k×k×k Table**	
Within (vorher-nachher)	**Between**	**Within**	**Between**
1. Bowker 2. Cochran Q	r*c $\chi 2$		

Frequency data

k×k Table		k×k×k Table	
Within (vorher-nachher)	**Between**	**Within**	**Between**
1. Bowker 2. Cochran Q	r*c $\chi 2$	1. Loglinear analysis 2. CFA	1. Loglinear analysis 2. CFA

Ordinal and interval data

	1 IV	1 IV	
	1 sample test	2 sample tests	
		Within	Between
Ordinal			
Interval			

Ordinal and interval data

	1 IV	**1 IV**	
	1 sample test	**2 sample tests**	
		Within	**Between**
Ordinal			
Interval	1. One-sample t-test 2. Confidence intervals		

Ordinal and interval data

	1 IV	1 IV	
	1 sample test	2 sample tests	
		Within	Between
Ordinal			
Interval	1. One-sample t-test 2. Confidence intervals	Paired t-test	

Ordinal and interval data

	1 IV	1 IV	
	1 sample test	**2 sample tests**	
		Within	**Between**
Ordinal			
Interval	1. One-sample t-test 2. Confidence intervals	Paired t-test	Dependent t-test

Ordinal and interval data

	1 IV	**1 IV**	
	1 sample test	**2 sample tests**	
		Within	**Between**
Ordinal		1. Wilcoxon 2. Sign test	
Interval	1. One-sample t-test 2. Confidence intervals	Paired t-test	Dependent t-test

Ordinal and interval data

	1 IV	**1 IV**	
	1 sample test	**2 sample tests**	
		Within	**Between**
Ordinal		1. Wilcoxon 2. Sign test	1. Mann-Whitney U 2. Kolmogorov-Smirnov
Interval	1. One-sample t-test 2. Confidence intervals	Paired t-test	Dependent t-test

Ordinal and interval data

	1 IV		2 IVs	
	2+ sample tests		2+ sample tests	
	Within	**Between**	**Within**	**Between**
Ordinal				
Interval				

Ordinal and interval data

	1 IV		2 IVs	
	2+ sample tests		2+ sample tests	
	Within	Between	Within	Between
Ordinal				
Interval	One-way ANOVA	One-way ANOVA		

Ordinal and interval data

	1 IV		2 IVs	
	2+ sample tests		**2+ sample tests**	
	Within	**Between**	**Within**	**Between**
Ordinal	Friedman	Kruskal-Wallis		
Interval	One-way ANOVA	One-way ANOVA		

Ordinal and interval data

	1 IV		2 IVs	
	2+ sample tests		2+ sample tests	
	Within	**Between**	**Within**	**Between**
Ordinal	Friedman	Kruskal-Wallis		
Interval	One-way ANOVA	One-way ANOVA	Factorial ANOVA	Factorial ANOVA

Configurational Frequency Analysis

Configurational frequency analysis

	Declarative clauses		Questions		
	Transitive	Intransitive	Transitive	Transitive	Total
Written	23	45	56	12	136
Spoken	34	56	32	22	144
Total	57	101	88	34	280

Variables:
1. Genre (spoken – written)
2. Sentence type (declarative – interrogative)
3. Transitivity (transitive - intransitive

Configurational frequency analysis

What determines the choice between the s-gentive and the of-attributive construction:

(1) The book‘s cover
(2) The cover of the book

Variables:

1. Construction type: -s vs. of
2. Meaning of possessor: abstract, concrete, human
3. Meaning of possessed: abstract, concrete, human

Configurational frequency analysis

Possessed	**Abstract**		**Concrete**		**Human**		**Total**		
Possessor	*of*	*s*	*of*	*s*	*of*	*s*	*of*	*s*	Total
Abstract	80	37	9	8	3	2	92	47	139
Concrete	22	0	20	1	0	0	42	1	43
Human	9	58	1	35	6	9	16	102	118
Total	111	95	30	44	9	11	150	150	300
	206		74		20				

Expected frequencies: Multiplication of marginal frequencies

Configurational frequency analysis

Possessed	**Abstract**		**Concrete**		**Human**		**Total**		
Possessor	*of*	*s*	*of*	*s*	*of*	*s*	*of*	*s*	Total
Abstract	80	37	9	8	3	2	92	47	139
Concrete	22	0	20	1	0	0	42	1	43
Human	9	58	1	35	6	9	16	102	118
Total	111	95	30	44	9	11	150	150	300
	206		74		20				

Configurational frequency analysis

Possessed	**Abstract**		**Concrete**		**Human**		**Total**		
Possessor	*of*	*s*	*of*	*s*	*of*	*s*	*of*	*s*	Total
Abstract	80	37	9	8	3	2	92	47	139
Concrete	22	0	20	1	0	0	42	1	43
Human	9	58	1	35	6	9	16	102	118
Total	111	95	30	44	9	11	150	150	300
	206		74		20				

Possessor	Possessed	Type	Observed	Expected	Residuals
Abstract	**Abstract**	**of**	80	$\frac{139 \times 206 \times 150}{300^2} = 47.72$	$\frac{(80 - 47.72)^2}{47.72} = 21.83$
Abstract	**Abstract**	**s**	37	37.72	2,41
Abstract	**Concrete**	**of**	9	17.14	3.87
Abstract	**Concrete**	**s**	8	17.14	4.88
Abstract	**Human**	**of**	3	4.63	0.58
Abstract	**Human**	**s**	2	4.63	1.5
Concrete	**Concrete**	**s**	22	14.76	3.55
Concrete	**Concrete**	**of**	0	14.76	14.76
Concrete	**Abstract**	**s**	20	5.3	40.73
Concrete	**Abstract**	**of**	1	5.3	3.49
Concrete	**Human**	**s**	0	1.43	1.43
Concrete	**Human**	**s**	0	1.43	1.43
Human	**Abstract**	**s**	9	40.51	24.51
Human	**Abstract**	**of**	58	40.51	7.55
Human	**Concrete**	**s**	1	14.55	12.62
Human	**Concrete**	**of**	35	14.55	28.73
Human	**Human**	**s**	5	3.93	1.09
Human	**Human**	**s**	9	3.93	6.53
Summen			300	300	181,54 ($\chi 2$)

Configurational frequency analysis

- Determine the expected frequencies for all variable combinations.
- Divide the p-value (i.e. the asociated X^2-value) by the total number of tests (here: 18 tests= $X^2 = 8.95$).
- Compare the X^2-values of each variable combination to the X^2-value of the adjusted p-value (i.e. the one divided by the total number of tests).
- Variable combinations with a X^2-value higher than the adjusted X^2-value are significant ‘types’.

Possessor	Possessed	Type	Observed	Expected	Residuals
Abstract	**Abstract**	**of**	80	$\frac{139\times206\times150}{300^2}$ = 47.72	$\frac{(80-47.72)^2}{47.72}$ = 21.83
Abstract	**Abstract**	**s**	37	37.72	2,41
Abstract	**Concrete**	**of**	9	17.14	3.87
Abstract	**Concrete**	**s**	8	17.14	4.88
Abstract	**Human**	**of**	3	4.63	0.58
Abstract	**Human**	**s**	2	4.63	1.5
Concrete	**Concrete**	**s**	22	14.76	3.55
Concrete	**Concrete**	**of**	0	14.76	**14.76**
Concrete	**Abstract**	**s**	20	5.3	**40.73**
Concrete	**Abstract**	**of**	1	5.3	3.49
Concrete	**Human**	**s**	0	1.43	1.43
Concrete	**Human**	**s**	0	1.43	1.43
Human	**Abstract**	**s**	9	40.51	**24.51**
Human	**Abstract**	**of**	58	40.51	7.55
Human	**Concrete**	**s**	1	14.55	**12.62**
Human	**Concrete**	**of**	35	14.55	**28.73**
Human	**Human**	**s**	5	3.93	1.09
Human	**Human**	**s**	9	3.93	6.53
Summen			300	300	181,54 ($\chi2$)

Logistic regression

Logistic regression

- Multiple predictor variables (continuous + categorical)
- A categorical dependent variable (with two or more levels)

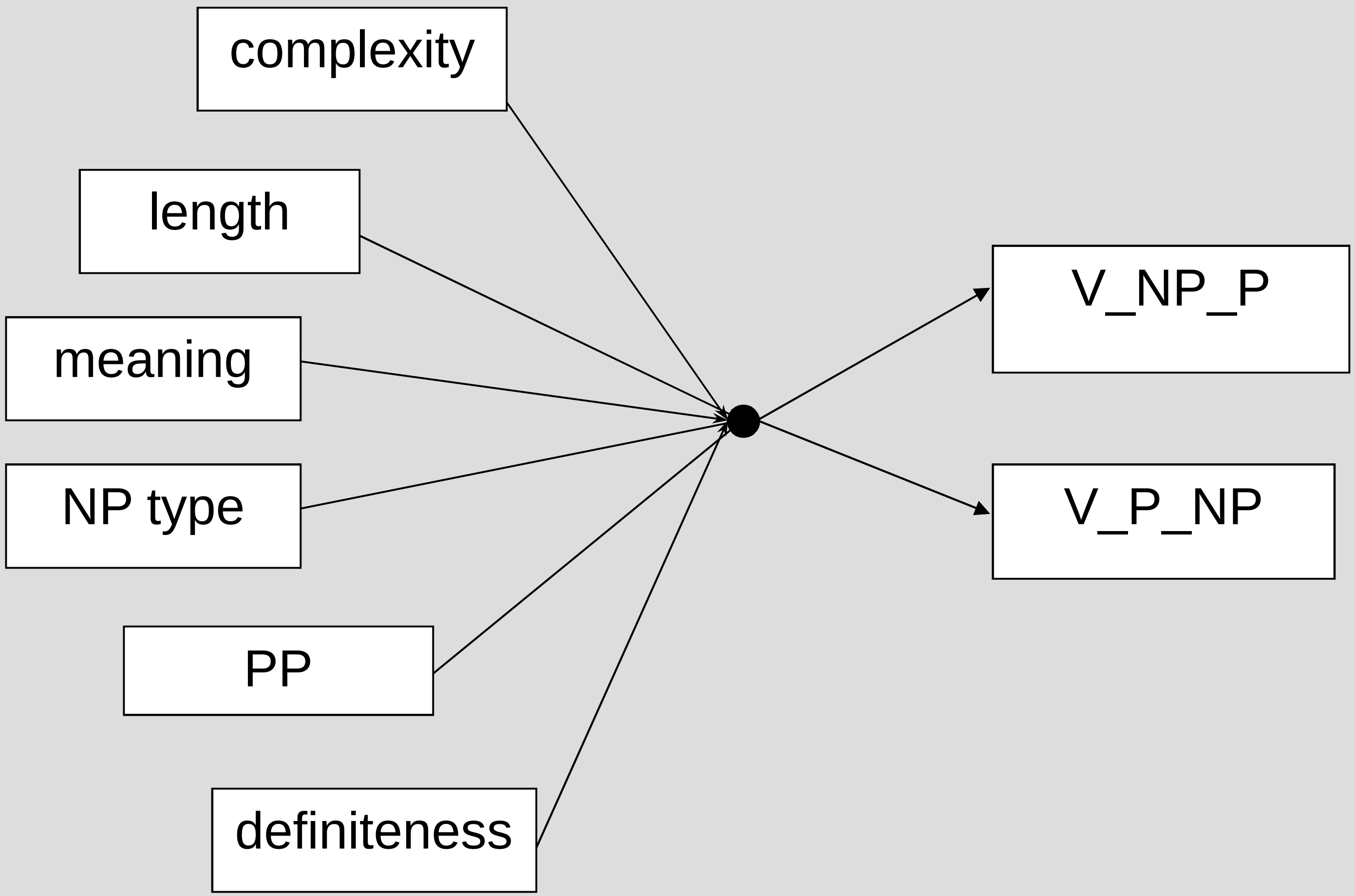
Logistic regression

What determines the order of object and particle in the English verb particle construction?

(1) He looked the number up.
(2) He looked up the number.

Previous research suggests that the following factors may be relevant: the length and complexity of the direct object, the meaning and definiteness of the object, the NP type of the object (pronoun vs. lexical NP), and the occurrence of a locational PP at the end of the sentence.

Logistic regression



Exercises

Exercise 1. A researcher has collected a sample of 575 complex sentences including temporal adverbial clauses. 201 temporal clauses refer to an event that occurred prior to the main clause, 161 temporal clauses occur simultaneously to the main clause, and 213 temporal clauses occur posterior to the main clause:

(1) After we left Jena, it began to rain.
(2) When we arrived in Jena, it began to rain.
(3) Before we arrived in Jena, it began to rain.

Is the variation between prior, posterior, and simultaneous temporal clauses still the range of what one would expect by chance, or is the distribution skewed?

Determine the expected frequencies and calculate the $\chi 2$ value (by hand and with SPSS).

How many degrees of freedom do we have?

	N
prior	201
simultaneous	161
posterior	213

Statistik für Test

	frequency
Chi-Quadrat[a]	7,736
df	2
Asymptotische Signifikanz	,021

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 191,7.

Exercise 2. A researchers wants to find out if and to what extend age and time spend in a pre-school affect the acquisition of complex sentences. The command of complex sentences was tested in a comprehension experiment with 15 children.

Age (month)	Preschool (weeks)	Test score
33,00	45,00	61,00
64,00	68,00	72,00
33,00	100,00	84,00
22,00	44,00	39,00
70,00	62,00	50,00
66,00	61,00	55,00
59,00	52,00	71,00
84,00	66,00	71,00
56,00	79,00	66,00
44,00	44,00	51,00
22,00	16,00	29,00
44,00	61,00	45,00
80,00	60,00	70,00
66,00	61,00	58,00
79,00	60,00	65,00

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,810[a]	,656	,599	9,19051

a. Einflußvariablen : (Konstante), anxiety, studytime

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	17,891	9,088		1,969	,073
	studytime	,543	,142	,691	3,827	,002
	anxiety	,172	,127	,244	1,353	,201

a. Abhängige Variable: score

Exercise 3. Adverbial clauses can precede or follow the main clause. Is clause order dependent on the semantic link between main and subordinate clause? The answer this question examine the following data from a corpus.

	Initial	Final	Central
Causal	5	45	5
Conditional	37	16	2
Temporal	45	36	7
Concessive	17	15	3

1. Determine the expected frequencies (manually and by using SPSS).
2. Analyse the association between order and meaning.

AC * Position Kreuztabelle

			Position			Gesamt
			central	final	initial	
AC	causal	Anzahl	5	45	5	55
		Erwartete Anzahl	4,0	26,4	24,5	55,0
	concessive	Anzahl	3	15	17	35
		Erwartete Anzahl	2,6	16,8	15,6	35,0
	conditional	Anzahl	2	16	37	55
		Erwartete Anzahl	4,0	26,4	24,5	55,0
	temporal	Anzahl	7	36	45	88
		Erwartete Anzahl	6,4	42,3	39,3	88,0
Gesamt		Anzahl	17	112	104	233
		Erwartete Anzahl	17,0	112,0	104,0	233,0

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	42,510[a]	6	,000
Likelihood-Quotient	47,889	6	,000
Anzahl der gültigen Fälle	233		

a. 3 Zellen (25,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 2,55.

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,427	,000
	Cramer-V	,302	,000
Anzahl der gültigen Fälle		233	