Categorical Data Examples

Testing a single proportion

Researchers tested whether Underdogs in American football tend to "cover the spread" in a majority of games. For bet to "pay-off" in long-run, the Underdog would have to win the bet with Probability > .5238. Based on Arena Football League:

$H_0: \pi \leq 0.5238 \quad H_A: \pi > .5238$ Sample: $n = 992$ games $\quad y = 534$ underdog wins bet

Comparing 2 Proportions – Independent / Large Samples

Researchers compared use of tanning beds among students who do and do not watch reality beauty shows

$H_0: \pi_1 = \pi_2 \quad H_A: \pi_1 \neq \pi_2$

Group 1: Watch Reality Beauty Shows: $n_1 = 224 \quad y_1 = 29$ use tanning beds

Group 1: Do not Watch Reality Beauty Shows: $n_2 = 352 \quad y_2 = 13$ use tanning beds

Comparing 2 Proportions – Independent / Small Samples

John Lister Data on Deaths from upper limb amputations: $H_0: \pi_{\text{Post}} \geq \pi_{\text{Pre}} \quad H_A: \pi_{\text{Post}} < \pi_{\text{Pre}}$

	Death	NoDeath	Total
Pre-Antiseptic	6	6	12
Post-Antiseptic	1	11	12
Total	7	17	24

Comparing 2 Proportions – McNemar's Test

A sample of 150 people asked to choose between a Sure Thing and a Risky alternative in terms of a Gain, and in terms of a loss. Let π be the probability a person chooses the Risky alternative. $H_0: \pi_G = \pi_L \quad H_A: \pi_G \neq \pi_L$

Gain\Loss	Sure	Risky
Sure	16	110
Risky	4	20

Goodness-of-Fit for Multinomial Distribution

120 Subjects were given warm beverage in cups of 4 colors (Blue, Green, Yellow, Red) and asked to rate which was warmest (all were of same temperature). Researchers were testing whether there are inherent differences of choices from random selection.

$H_0: \pi_B = \pi_G = \pi_Y = \pi_R = 0.25 \qquad$ Data: $n_B = 16, n_G = 24, n_Y = 34, n_R = 46$

Chi-Square Test

People were assigned to either Simultaneous choosing items for 3 future product uses, or Sequential choice of items for a single usage on 3 occasions. Subjects were classified as having selected either High, Medium, or Low variety based on their purchase choices. H_0: No association between choice condition and Variety Seeking. H_A: Association exists.

Observed	H	M	L	Total
Sim	141	67	22	230
Seq	91	85	62	238
Total	232	152	84	468

Ordinal Association

People were told a beer was either: low, medium, or high price; and then after tasting it, rating it as: undrinkable (0), poor (1), fair (2), good (3), or very favorable 4. Compute the sample value of γ, the measure of ordinal association.

		Price		
		1	2	3
Quality	0	4	1	4
	1	20	21	8
	2	23	22	26
	3	9	12	15
	4	4	4	7

Large-Sample Z-test for Proportions and Chi-Square Test in 2x2 Table

Prospect Theory: (Independent Samples)

Problem 11: In addition to what you own, you have been given 1000. (Gain Condition)

Choice A: 50% Chance of Win 1000, 50% Chance of Win 0 Choice B: 100% Chance of Win 500 (n_1 = 70, 58 take B)

Problem 12: In addition to what you own, you have been given 2000. (Loss Condition)

Choice C: 50% Chance of Lose 1000, 50% Chance of Lose 0 Choice D: 100% Chance of Lose 500 (n_2 = 68, 21 take D)

Goal: Test whether true proportions choosing sure bet are equal for 2 Conditions. $H_0 : \pi_G = \pi_L$

Sample proportions in Each Condition (and overall) choosing sure Bet and Z-test:

$$\hat{\pi}_G = \frac{59}{70} = 0.842857 \quad \hat{\pi}_L = \frac{21}{68} = 0.308824 \quad \hat{\bar{\pi}} = \frac{59+21}{70+68} = \frac{80}{138} = 0.579710 \quad \sqrt{0.579710\left(0.420290\right)\left(\frac{1}{70}+\frac{1}{68}\right)} = 0.084046$$

$$TS : Z_{obs} = \frac{0.842857 - 0.308824}{0.084046} = 6.354056 \quad \text{RR:} \quad \left|Z_{obs}\right| \geq z_{.025} = 1.96$$

Chi-Square Test: H_0: No association between Gain/Loss perspective and Choice

obs	Sure	Risk	
Gain	59	11	70
Loss	21	47	68
	80	58	138

exp	Sure	Risk	
Gain	40.57971	29.42029	70
Loss	39.42029	28.57971	68
	80	58	138

Chi-squar	Sure	Risk	
Gain	8.361496	11.5331	
Loss	8.607422	11.87231	
			40.37432

Observed Counts: $n_{ij} \quad i=1,2;\ j=1,2$ Expected Counts: $E_{ij} = \frac{n_{i\bullet}n_{\bullet j}}{n_{\bullet\bullet}} \quad i=1,2;\ j=1,2$

$$TS: X^2_{obs} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij}-E_{ij}\right)^2}{E_{ij}} = 40.37432 \quad RR: X^2_{obs} \geq \chi^2_{.05} = 3.841$$

Note: $Z^2_{obs} = X^2_{obs}$ and $Z^2_{.025} = \chi^2_{.05}$

Odds Ratio

Study compared Olestra and triglyceride potato chips for gastrointestinal symptoms.

Olestra: $n_O = 563$ $y_O = 89$ Triglyceride: $n_T = 529$ $y_T = 93$

Mantel –Haenszel Test – Combining Results from 2x2 Contingency Tables

Sleep apnea study. Combining results from *h* = 3 tables. Exposed = Modanafil Unexposed = Placebo

Outcome = Presence / Absence of Good Event

				N1	N2	N1-N2	D
7	2	9		7	6.3	0.7	1.094211
7	4	11					
14	6	20					
36	16	52		36	25.08772	10.91228	7.123962
19	43	62					
55	59	114					
38	37	75		38	28.73377	9.266234	9.152318
21	58	79					
59	95	154					
				MHChiSq	25.09499		
				MHCC	23.90743		

$$\text{Test Statisic: } \chi^2_{MH} = \frac{\left[\sum_{h=1}^{q}\left(n_{h11} - \frac{n_{h1\bullet}n_{h\bullet1}}{n_{h\bullet\bullet}}\right)\right]^2}{\sum_{h=1}^{q}\frac{n_{h1\bullet}n_{h2\bullet}n_{h\bullet1}n_{h\bullet2}}{n^2_{h\bullet\bullet}\left(n_{h\bullet\bullet}-1\right)}}$$

$$\text{Rejection Region: } \chi^2_{MH} \geq \chi^2_{\alpha,1} \qquad \text{P-value: } P\left(\chi^2_1 \geq \chi^2_{MH}\right)$$

$$\overset{\wedge}{OR}_{MH} = \frac{R}{S} \qquad R = \sum_{h=1}^{q}\frac{n_{h11}n_{h22}}{n_{h\bullet\bullet}} \qquad S = \sum_{h=1}^{q}\frac{n_{h12}n_{h21}}{n_{h\bullet\bullet}}$$

$$v = \hat{V}\left(\overset{\wedge}{OR}_{MH}\right) = \frac{1}{S^2}\sum_{h=1}^{q}S_h^2\left[\frac{1}{n_{h11}}+\frac{1}{n_{h12}}+\frac{1}{n_{h21}}+\frac{1}{n_{h22}}\right]$$

$$\text{95\% CI for Overall Odds Ratio: } \left(\overset{\wedge}{OR}_{MH}\, e^{-1.96\sqrt{v}}, \overset{\wedge}{OR}_{MH}\, e^{1.96\sqrt{v}}\right)$$

Inter-Rater Agreement – Cohen's Kappa and Weighted Kappa

Movie ratings by Gene Siskel and Roger Ebert (n = 160 movies).

Counts						Proportions				
S\E	1	2	3			S\E	1	2	3	
1	24	8	13	45		1	0.15	0.05	0.08125	0.28125
2	8	13	11	32		2	0.05	0.08125	0.06875	0.2
3	10	9	64	83		3	0.0625	0.05625	0.4	0.51875
	42	30	88	160			0.2625	0.1875	0.55	1
Weights						Expected				
S\E	1	2	3			S\E	1	2	3	
1	1	0.75	0			1	0.073828	0.052734	0.154688	
2	0.75	1	0.75			2	0.0525	0.0375	0.11	
3	0	0.75	1			3	0.136172	0.097266	0.285313	
p_obs	0.63125		p_ow	0.8						
p_chance	0.396641		p_ew	0.631016						
kappa	0.388839		kappa_w	0.457972						

$$p_{obs} = \sum_{i=1}^{k} p_{ii} \qquad p_{\exp} = \sum_{i=1}^{k} p_{i\bullet} p_{\bullet i} \qquad \hat{\kappa} = \frac{p_{obs} - p_{\exp}}{1 - p_{\exp}}$$

$$w_{ij} = 1 - \frac{\left(i-j\right)^2}{\left(k-1\right)^2} \qquad p_{obs}^{w} = \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} p_{ij} \qquad p_{\exp}^{w} = \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} p_{i\bullet} p_{\bullet j} \qquad \hat{\kappa}_w = \frac{p_{obs}^{w} - p_{\exp}^{w}}{1 - p_{\exp}^{w}}$$