General Biostatistics

Part 6

Common Statistical Tests

Comparing Proportions of Two Groups

Outline

- Comparing two groups with dichotomous outcomes
- 2 x 2 tables
- Summarizing information in 2 x 2 tables
- Chi-squared (χ^2) statistic
- Short-cut formula for χ^2 statistic for 2 x 2 tables

Comparing Two Groups

- We are interested in whether 2 groups are the same with respect to some dichotomous characteristic or outcome.
- Data may be displayed in a 2 x 2 table

2 x 2 Tables

Treatment	Response to treatment		Total
	Yes	No	
A	37	13	50
B	17	53	70
Total	54	66	120

Summarizing Information

Treatment	Response to treatment		Total	Proportion Responding
	Yes	No		
A	37	13	50	37/50= 0.74
B	17	53	70	17/70= 0.24
Total	54	66	120	54/120=0.45

Summarizing Information

- 74% respond to Treatment A
- 24% respond to Treatment B
- There is a 50% absolute difference in response rate between Treatments A and B
- The relative difference in response rate between Treatments A and B is a 208% relative increase
- (0.74-0.24)/0.24 = 2.08 x100 = 208%

Summarizing Information

- The odds of responding to Treatment A:
  P(responding)/P(not responding) =
  (37/50) / (13/50) = 2.85
- The odds of responding to Treatment B:
  P(responding)/P(not responding) =
  (17/70) / (53/70) = 0.32
- The odds ratio of response (comparing Treatment A to Treatment B):
  (odds in A) / (odds in B) = 2.85 / 0.32 =8.9

Comparing Two Groups

- Is the difference in response rates between the two treatments bigger than what we would expect due to chance alone?

Comparing Two Groups

- What frequencies would we expect if response was the same by treatment?

	Response to treatment		
Treatment	Yes	No	Total
A			50
B			70
Total	54	66	120

Comparing Two Groups

- What frequencies would we expect if response was the same by treatment?

	Response to treatment		
Treatment	Yes	No	Total
A	0.45(50)=22.5		50
B	0.45(70)=31.5		70
Total	54	66	120

Comparing Two Groups

- What frequencies would we expect if response was the same by treatment?

	Response to treatment		
Treatment	Yes	No	Total
A	22.5	27.5	50
B	31.5	38.5	70
Total	54	66	120

Observed vs Expected Counts

- Observed counts
- Expected counts

Treatment	Response to treatment		Total
	Yes	No	
A	37	13	50
B	17	53	70
Total	54	66	120

Treatment	Response to treatment		Total
	Yes	No	
A	22.5	27.5	50
B	31.5	38.5	70
Total	54	66	120

Chi-Squared Statistic

- χ^2 test statistic
  - Compares observed to expected counts across all k cells of a contingency table
  - Calculations use frequencies (counts) not proportions

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Chi-Squared Statistic

- χ^2 test statistic
  - Useful as a test of independence (no association; homogenous)
  - Test Ho that the samples are drawn from populations that are homogenous
- χ^2 test statistic follows a χ^2 distribution
  - family of distributions
  - its shape depends on its degrees of freedom

Chi-Squared Statistic

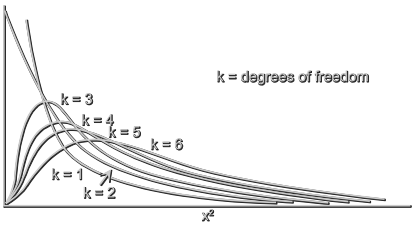
- χ^2 test statistic

$$\chi^2 = \sum_{i=1}^{k}\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\chi^2 = \frac{(37-22.5)^2}{22.5} + \frac{(13-27.5)^2}{27.5} + \frac{(17-31.5)^2}{31.5} + \frac{(53-38.5)^2}{38.5}$$

$$\chi^2 = 29.1 \text{ with } 1 \text{ df}$$

χ^2 Distributions



Chi-Squared Statistic

- Short-cut formula for 2 x 2 table

	Response to treatment		
Treatment	Yes	No	Total
A	37=a	13=b	50=a+b
B	17=c	53=d	70=c+d
Total	54=a+c	66=b+d	120=n

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Chi-Squared Statistic

- Short-cut formula for 2 x 2 table

	Response to treatment		
Treatment	Yes	No	Total
A	37=a	13=b	50=a+b
B	17=c	53=d	70=c+d
Total	54=a+c	66=b+d	120=n

Chi-Squared Statistic

- Short-cut formula for 2 x 2 table

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$\chi^2 = \frac{120(37(53)-13(17))^2}{(54)(66)(50)(70)}$$

$$\chi^2 = 29.1 \text{ with } 1 \text{ df}$$

Other Methods for 2 x 2 Tables

- Z test of difference in proportions between 2 groups
- Fisher's exact test
- Confidence interval for the true difference in proportions between 2 groups
- Odds ratio and confidence interval
- Mantel-Haenszel statistic for combining results from multiple 2 x 2 tables

Summary

- Summarizing information in 2 x 2 tables
- Testing for differences between 2 groups
  - Are the groups “homogeneous” with respect to some characteristic?
  - Is there an association between group and characteristic?
  - Are group and the characteristic independent?