Biost 517 / Biost 514 Applied Biostatistics I / Biostatistics I

Timothy A. Thornton, Ph.D.
Associate Professor of Biostatistics
University of Washington

Lecture 11:

Ratios of Two Proportions; Relative Risk and Odds Ratio; Fisher's Exact Test

Comparing two proportions: risk difference

- In the previous lecture we discussed comparing two proportions for two populations (or two subpopulations from the overall population) using the **risk difference.**
- For sufficiently large sample sizes n_1 and n_2, the normal distribution can be used for inference on the risk difference:

Consider two subpopulations A and B with unknown proportions p_1 and p_2 respectively. A simple random sample (SRS) of size n_1 from subpopulation A yields $\hat{p}_1$, and an independent SRS of size n_2 from subpopulation B yields $\hat{p}_2$. Then the distribution for risk difference is as follows:

$$(\hat{p}_1 - \hat{p}_2) \dot{\sim} N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

when n_1 and n_2 are sufficiently large.

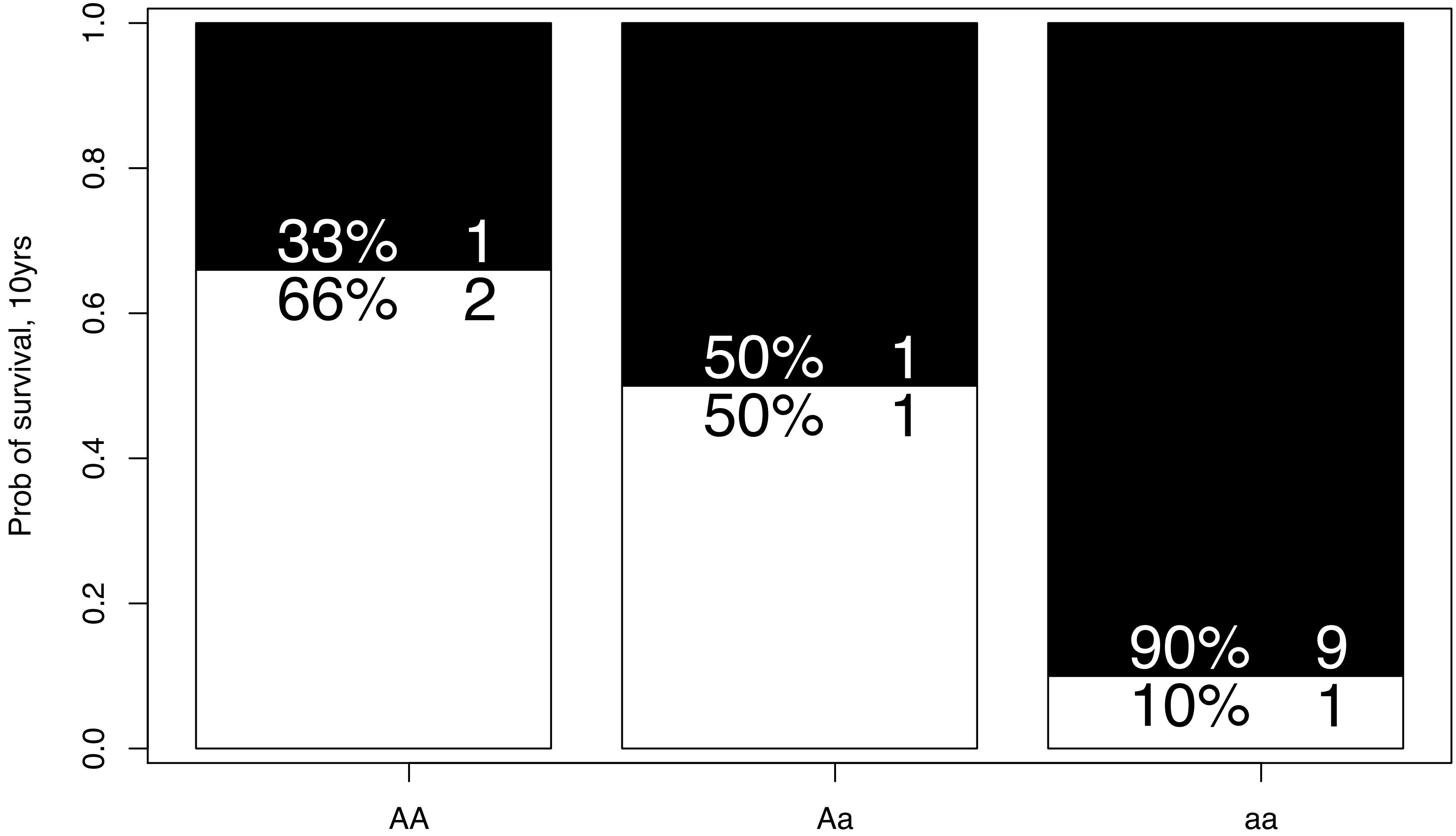
Summary Measures for Comparisons of Two Proportions

- Summary measures of interest for a Bernoulli random variable are pretty much limited to either
  - The proportion *p* (a mean), or
  - The odds $o = p / (1-p)$
    - Note that a proportion *p* must have $0 \le p \le 1$, while odds can take values between 0 and infinity, i.e., $0 \le o \le \infty$
- Contrasts used to compare the distribution of a Bernoulli random variable between two subpopulations thus include:
  - Difference in proportions: $p_1 - p_0$ (risk difference (RD))
  - Ratio of proportions : p_1 / p_0 (risk ratio (RR))
  - Odds ratio : $o_1 / o_0 = \dfrac{p_1 / (1 - p_1)}{p_0 / (1 - p_0)}$ (odds ratio (OR))

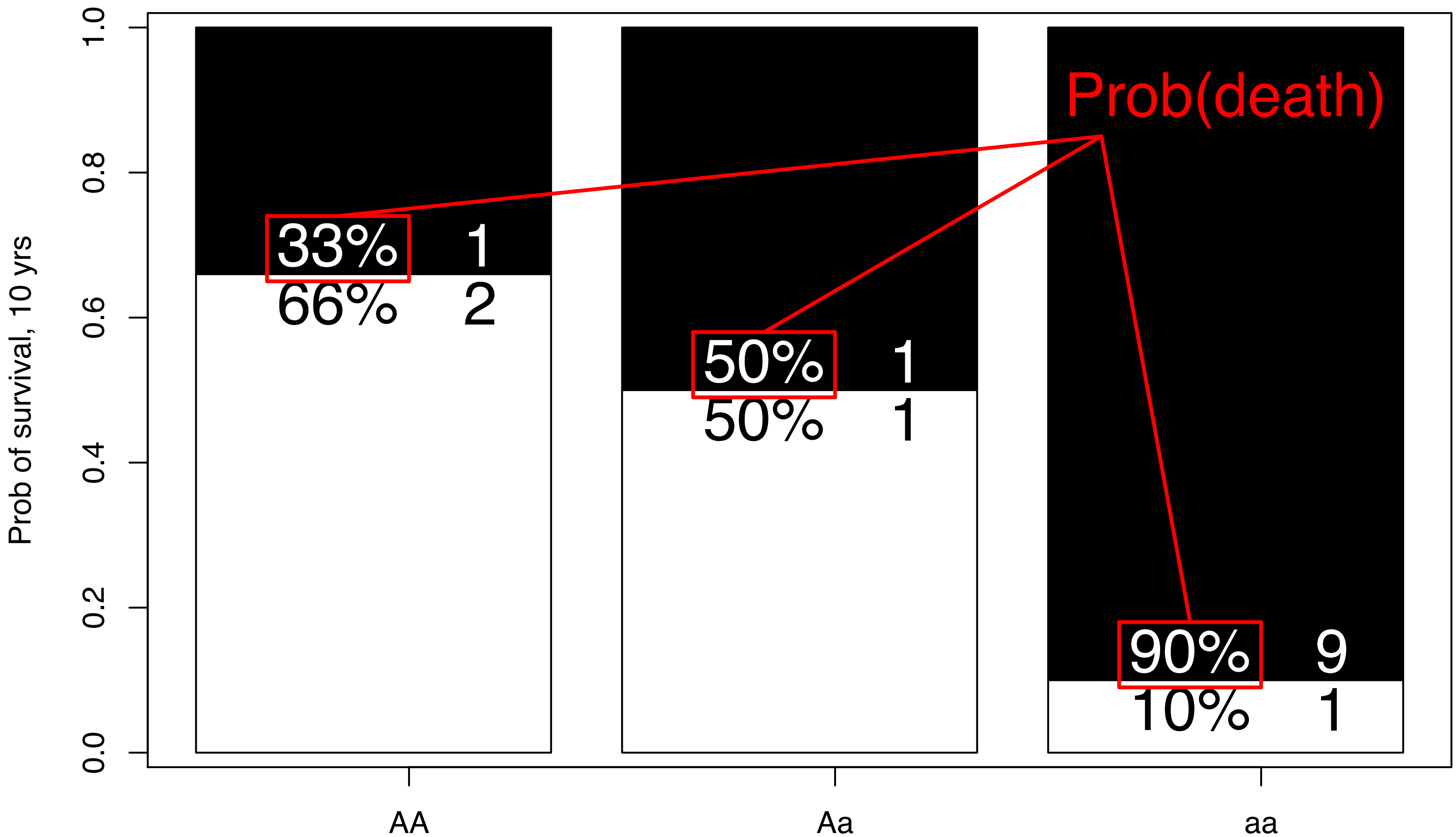
Example: Odds and Ratios

- What are odds? It really is just probability
- It is a [gambling friendly] measure of chance
- Consider a sample of individuals who are in remission for prostate cancer. Consider a genetic marker that is a biomarker for how aggressive the prostate cancer. Groupings are based on the genotype at the marker and three groups are: **AA**, **Aa**, and **aa**
- We are interested in comparing the proportion of individuals in each group who survives at least 10 years after being in remission

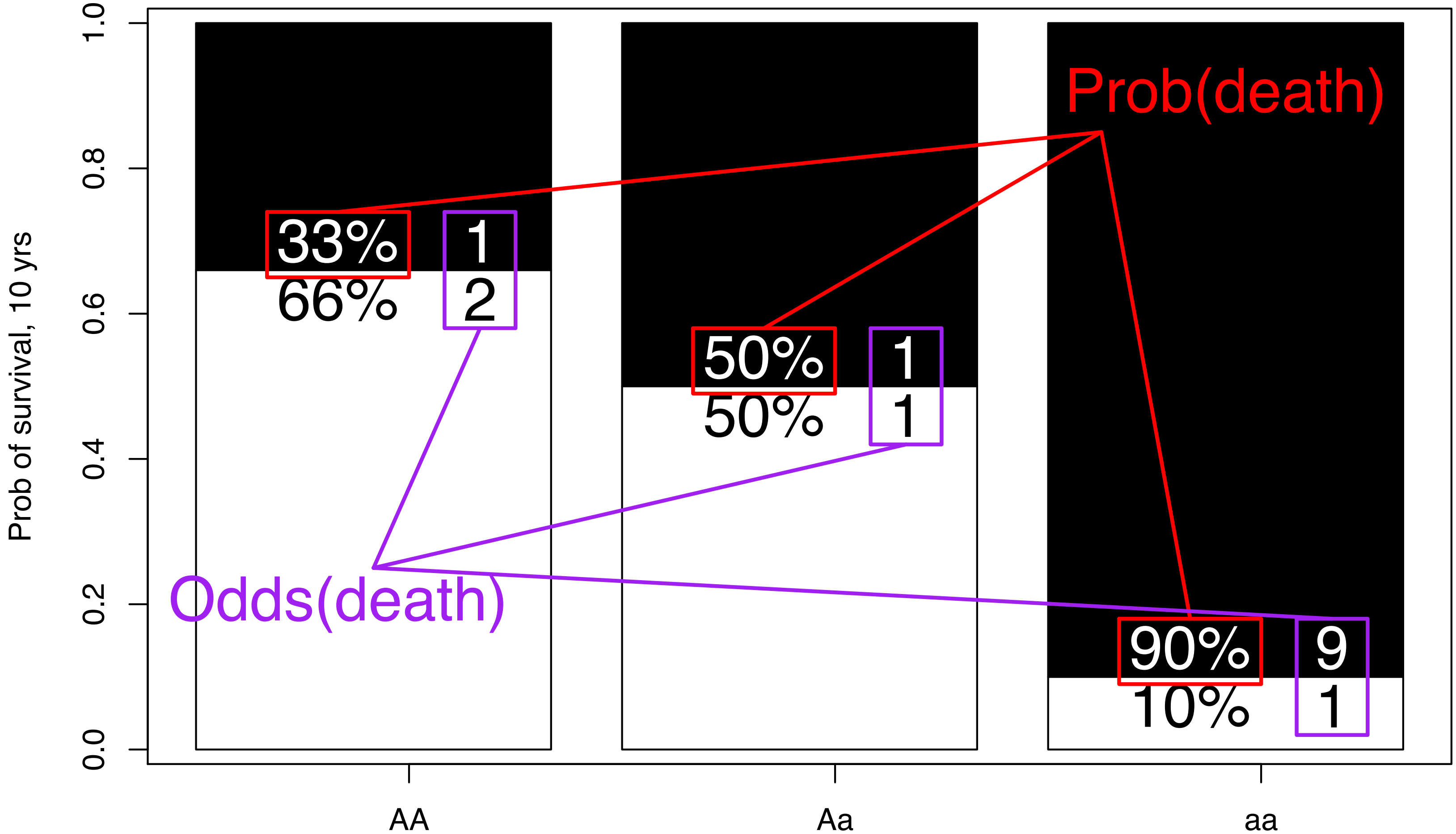
Example: Odds and Ratios



Example: Odds and Ratios



Example: Odds and Odds Ratios



Relative Risk versus Odds Ratios

- What about “Relative Risk” (RR) or “Risk Ratio” of death within 10 years?
- The RR for group Aa versus group AA is (0.5/0.33)=1.515
  - In comparison, odds ratio (OR) of Aa versus AA is (1/1)/(1/2))=2
- The RR for group aa versus group AA is (0.9/0.33)=2.727
  - In comparison, odds ratio (OR) of aa versus AA is (9/1)/(1/2))=18

Comparing two proportions

- Suppose we are interested in comparing the proportions of events that occurred in two treatment groups: an exposed group and an unexposed group.   and placebo.
- Consider the 2x2 table below with the counts of events that occurred and that did not occur for each group.
- Let ***a*** and ***b*** be the counts for the number events that occurred and the did not occur, respectively in the exposed group. Let ***c*** and ***d*** be the counts for the number events that occurred and that did not occur, respectively, in the unexposed group.

Group	**Event Occurred**	
	Yes	No
Exposure	a	b
Control	c	d

Comparing two proportions: RR and OR

Group	Event Occurred	
	Yes	No
Exposure	a	b
Control	c	d

$$\widehat{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{a/(a+b)}{c/(c+d)}$$

$$\widehat{OR} = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{a/b}{c/d} = \frac{ad}{cb}$$

Inference for Relative Risk

- RR=1 implies no association between exposure group and outcome
- If RR>1 or RR<1 then there is an association between exposure group and outcome
- The distribution of the $\log(\widehat{RR})$ is normally distribution with mean equal to $\log(RR)$
- The approximation typically works well if the counts in each of the cells in the contingency table is least 5

Inference for Relative Risk

Group	Event Occurred	
	Yes	No
Exposure	a	b
Control	c	d

- $log(\hat{RR}) = log\left(\frac{a/(a+b)}{c/(c+d)}\right)$ is approximately normally distributed with

$$s.e.\left(log(\hat{RR})\right) = \sqrt{\frac{1}{a} + \frac{1}{c} - \frac{1}{a+b} - \frac{1}{c+d}}$$

- Exponentiate to obtain lower limit of 95% CI for RR:

$$= exp(log(\hat{RR}) - 1.96 \times s.e.(log(\hat{RR})))$$

- Similarly exponentiate to obtain upper limit of 95% CI for RR:

$$= exp(log(\hat{RR}) + 1.96 \times s.e.(log(\hat{RR})))$$

Inference for Odds Ratio

- OR=1 implies no association between exposure group and outcome
- If OR>1 or OR<1 then there is an association between exposure group and outcome
- For large enough sample sizes the distribution of the $\log(\widehat{OR})$ is approximately normal with mean equal to $\log(OR)$
- The approximation typically works well if the counts in each of the cells in the contingency table is least 5

Inference for Odds Ratio

Group	Event Occurred	
	Yes	No
Exposure	a	b
Control	c	d

- $log(\hat{OR}) = log\left(\frac{ad}{cb}\right)$ is approximately normally distributed with

$$s.e.(log(\hat{OR})) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- Exponentiate to obtain lower limit of 95% CI for OR:

$$= exp(log(\hat{OR}) - 1.96 \times s.e.(log(\hat{OR})))$$

- Similarly exponentiate to obtain upper limit of 95% CI for OR:

$$= exp(log(\hat{OR}) + 1.96 \times s.e.(log(\hat{OR})))$$

Choice of Summary Measure / Contrast

- In choosing among these different comparison measures for proportions, one should consider important scientific issues including
  - Any desire to accentuate public health impact
  - Any desire to accentuate differences with rare events
  - Interplay of the following two questions:
    - How we want to eventually use the results of the analysis
      - Which variable do we ideally want to condition on and talk about the distribution of the other?
    - And then we also have to consider, based on the first question, what is the impact on how we sampled our data, i.e., the study design
      - Were sample sizes in any subpopulations fixed by design?

Choice of Summary Measure / Contrast

- Public health impact is typically best measured by the difference in proportions ➔ whenever considering public health impact, perhaps analysis using RD should be seriously considered
  - RD, as a difference in proportions.
    - From RD can estimate the number of affected people in a larger population
- Example: We believe that radon is harmful in terms of causing lung cancer. How many people really die because of their exposure to radon? To answer this question, we would want to use the risk difference.
  - Goal is to estimate of the total number at risk. Can do this by first finding the difference in proportions who get cancer in each of the following two groups: radon exposed and non-radon exposed
  - Apply this risk probability to the entire population.
- So to accentuate public health impact, risk difference might be more useful

Choice of Summary Measure / Contrast

- On the other hand, if we are trying to discover causes of rare disease, for example, then the existence of an association is best demonstrated and described using ratios ➔ perhaps we prefer RR or OR
  - Classic example: though it is unlikely in either case, I am many more times likely to win the lottery by buying a ticket than by finding a winning ticket
  - If we just talked about risk difference, the two probabilities are so negligibly small that it would be difficult to actually grasp how much less likely I am to win because I refuse to buy a ticket. The ratio, however, better accentuates those differences.

Comparing Independent Proportions with Small Samples

Small Sample CI

- For making inference about a single proportion, we could use exact methods
  - Consider all values of p and ask: are our data unusual for that value?
- Unfortunately, we have no way of obtaining exact CI for measures comparing the proportions
- We could consider all possible values of the two proportions, and see whether a test would reject each combination
  - But the result would not always be the same for two proportions that differ by the same amount
    - E.g, it might reject .10 and .20, but not .40 and .50
  - In other words, we need to know the value of the two proportions being compared in order to find the exact distribution of the difference

Small Sample Tests

- We can, however, describe the exact distribution of the data under the null hypothesis conditional on all the “margins” of a contingency table

- We imagine randomly assigning observations while keeping the marginal totals constant

Hypogeometric Distrbution

- The **hypergeometric distribution** is a discrete probability distribution that describes the probability of having *k* successes in *n* draws, without replacement, from a finite population of size *N* that contains exactly *K* objects with that feature, wherein each draw is either a success or a failure.

$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Fisher's Exact Test

- Probability of more extreme contingency tables with the same marginal totals
- Probabilities by **hypergeometric distribution**
  - (Use a computer)

$$
\begin{array}{ll}
 & \text{Response} \\
 & \begin{array}{ccc|c}
 & + & - & \\
0 & a-k & b+k & n_0 \\
1 & c+k & d-k & n_1 \\
\hline
 & m_0 & m_1 & N
\end{array}
\end{array}
$$

Group

$$\text{Consider all possible values of } k$$

Fisher's Exact Test: An example

- In a retrospective study among men ages 50-54 who died the same month, investigators included approximately an equal number of men who died from CVD (cases) and who died from other causes (controls). Of 35 men who died from CVD, 5 had a high-salt diet before they died. Among the 25 controls, 2 were on a high salt diet.

Rosner text, 10.3 (p. 367 in 7th edition)

Fisher's exact test: An Example

	High Salt	Low Salt	Total
non-CVD	2	23	25
CVD	5	30	35
Total	7	53	60

Were men who died with and without CVD equally likely to have had a high salt diet?

Fisher's exact test: An Example

- All possible tables with fixed margins

0		25
		35
7	53	60

1		25
		35
7	53	60

Observed table

2		25
		35
7	53	60

3		25
		35
7	53	60

4		25
		35
7	53	60

5		25
		35
7	53	60

6		25
		35
7	53	60

7		25
		35
7	53	60

Fisher's exact test: An Example

- The 8 tables on the previous slide are the only tables consistent with the margin counts.
- Probabilities for each possible table can be calculated under the assumption that CVD is independent of low-salt/high-salt diet.
  - Come from the hypergeometric distribution
  - (You won't need to do this by hand.)
- Each table gets a probability
- P-value = sum of these probs for observed table and more extreme tables.

Fisher’s exact test: An Example

- Probability of each table assuming no association.

prob=0.017

0		25
		35
7	53	60

prob=0.105

1		25
		35
7	53	60

Observed table prob=0.252

2		25
		35
7	53	60

prob=0.312

3		25
		35
7	53	60

prob=0.214

4		25
		35
7	53	60

5		25
		35
7	53	60

prob=0.082

prob=0.016

6		25
		35
7	53	60

prob=0.001

7		25
		35
7	53	60

Fisher's exact test in R

- The ***fisher.test()*** function in R can be used to perform a fisher exact text.
- It takes as input a 2x2 matrix (or table)
- Performs Fisher's exact test using the odds ratio as the measure for comparing proportions between the row variable and the column variable
  - Provides a p-value for hypothesis testing of the odds ratio
  - Provides a 95% confidence interval for the odds ratio

Fisher's exact test in R: Example

```
> DataTable<-matrix(c(2, 23, 5, 30),nrow=2)
> DataTable
     [,1] [,2]
[1,]    2    5
[2,]   23   30
> fisher.test(DataTable)

        Fisher's Exact Test for Count Data

data:  DataTable
p-value = 0.6882
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.04625243 3.58478157
sample estimates:
odds ratio
  0.527113
```

Fisher's exact test: An Example

- 1-sided p-value sums these three probabilities

prob=0.017

0		25
		35
7	53	60

prob=0.105

1		25
		35
7	53	60

prob=0.252

Observed table

2		25
		35
7	53	60

prob=0.312

3		25
		35
7	53	60

prob=0.214

4		25
		35
7	53	60

5		25
		35
7	53	60

prob=0.082

prob=0.016

6		25
		35
7	53	60

prob=0.001

7		25
		35
7	53	60

Fisher's exact test: An Example

- 2-sided p-value sums 7 probabilities (R method)

prob=0.017

0		25
		35
7	53	60

prob=0.105

1		25
		35
7	53	60

prob=0.252

Observed table

2		25
		35
7	53	60

prob=0.312

3		25
		35
7	53	60

prob=0.214

4		25
		35
7	53	60

prob=0.082

5		25
		35
7	53	60

prob=0.016

6		25
		35
7	53	60

prob=0.001

7		25
		35
7	53	60

Fisher's exact: 2-sided p-value

- Different software may give different two-sided p-values
  - sum the probabilities of all tables with smaller probabilities than the observed table (R)
  - double the p-value for the 1-sided test
  - Sum the probabilities of all tables with the odds ratio at least as far from 1 as the observed table
  - … or the relative risk at least as far from 1 as the observed table
  - …or risk difference at least as far from zero as the observed table

Normal Approximation vs. Fisher's exact

- Fisher's exact test is valid for small samples, but not a great test
  - Can be excessively conservative – the set of possibly p-values is very "chunky"
    - Suppose $\alpha=0.05$. In the example, how could one get a p-value < 0.05?
    - With small sample sizes power is already an issue, exacerbated by using a test with low power.
  - The name "exact" sounds very scientific/precise, but that's misleading in some sense…
    - recommend using the normal approximation (e.g, a Z test) when appropriate