Summarizing Categorical Data:  
Risk, Relative Risk, Odds, Odds Ratios

A microwave popcorn production plant had a suspicious incidence of chronic airway obstruction among its workers. Air and dust samples at different locations within the plant were collected to determine worker exposure to a particular chemical used in the production process. Employees were classified into two groups: a “low exposure group” and a “high exposure group.”

<table>
<thead>
<tr>
<th></th>
<th>low exposure</th>
<th>high exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>airway obstructed</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>airway not obstructed</td>
<td>52</td>
<td>43</td>
</tr>
</tbody>
</table>

Numerical Summaries

**Conditional proportions:** calculate proportion of “successes” separately for each group, “conditioning” on the explanatory variable group

let's morbidly define a “success” to be “having airway obstruction”

conditional proportion of employees with airway obstruction -

- in the high exposure group: \( \frac{15}{58} = .259 \)  (risk)
- in the low exposure group: \( \frac{6}{58} = .103 \)  (risk)

difference in the 2 proportions: \( .259 - .103 = .156 \)  (does it seem large?)

what if the proportions had been .494 and .650? **same difference of .156**  (large?)

**Relative risk:** the ratio of the risks for each group.

\[
relative \ risk = \frac{\text{risk of group 1}}{\text{risk of group 2}} \quad \text{or} \quad \frac{\text{proportion of successes in group 1}}{\text{proportion of successes in group 2}}
\]

Note: it is often standard to put the group with the lower risk in the denominator.

\[
relative \ risk = \frac{\text{risk group 1}}{\text{risk group 2}} = \frac{.259}{.103} = 2.51
\]
**INTERPRETATION:**
the risk of airway obstruction is 2.5 greater for employees in the high exposure group than for employees in the low exposure group

**ODDS:**

*Odds of success:* the ratio of the proportion of “successes” to the proportion of “failures”, which simplifies to the ratio of the number of successes to the number of failures.

\[
\text{odds} = \frac{\text{proportion of success in the group}}{\text{proportion of failures in the group}} = \frac{\text{number of successes in the group}}{\text{number of failures in the group}}
\]

**EXAMPLES**

i) If the odds are 2 (commonly expressed 2-to-1) in favor of an outcome, we expect a success twice as often as a failure in the long run.

ii) If the odds are 9-to-5 in favor of an outcome, we expect 9 successes for every 5 failures in the long run.

iii) It is important to know how the “outcome” is defined. For example, in horse racing, odds are typically presented in terms of “losing the race,” so if a horse is given 2-to-1 odds against winning a race, we expect the horse to lose two-thirds of the races in the long run.

**EXAMPLE:**

Success: “no airway obstruction”. Calculate the overall odds of no airway obstruction for these 116 workers.

\[
\text{odds} = \frac{95}{116} = \frac{95}{21} = 4.52 \text{ (to 1)}
\]

Interpretation: for every 4.5 workers with airway obstruction, we expect 1 worker without airway obstruction.

We often need to compare the odds of successes for two different groups

**ODDS RATIO**

To compare the odds of successes for two different groups, compute the *odds ratio* between group 1 and group 2 (using group 2 as the reference group):

\[
\text{odds ratio} = \frac{\text{odds of success in group 1}}{\text{odds of success in group 2}}
\]
NOTE: it is often standard to put the group with the lower odds in the denominator so that the odds ratio is larger than 1.

EXAMPLE
Compute the odds ratio of airway obstruction between the high exposure group and low exposure group.

odds of airway obstruction in high exposure group: \( \frac{15}{43} = .35 \) (group 1)

odds of airway obstruction in low exposure group: \( \frac{6}{52} = .12 \) (group 2)

\[ \text{odds ratio} = \frac{.35}{.12} = 2.92 \]

Interpretation: odds of airway obstruction in high exposure group is 2.92 times larger than the odds of airway obstruction in the low exposure group.

CONCLUSION FROM THIS DATA SET
Employees with high exposure (sample size 58) are 2.51 times more likely to experience airway obstruction than those in the low exposure group (sample size 58). Equivalently, the odds of airway obstruction for high exposure workers are 2.92 times higher than the odds for the low exposure group.

EXAMPLE
Near-sightedness typically develops during the childhood years. Recent studies have explored whether there is an association between development of near-sightedness and the use of night-lights with infants. In one study the parents of children who were seen as outpatients in a university pediatric ophthalmology clinic completed a questionnaire concerning their child's bedroom lighting at night before the age of 2 years. The results are shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th>near-sighted</th>
<th>not near-sighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>used night light</td>
<td>78</td>
<td>115</td>
</tr>
<tr>
<td>no night light</td>
<td>18</td>
<td>114</td>
</tr>
</tbody>
</table>

“Success”: near-sighted

Part a:
conditional proportion of children who are near-sighted -

in the night-light group: \( \frac{78}{78+115} = .404 \) (risk)

in the no night-light group: \( \frac{18}{18+114} = .136 \) (risk)
Part b:
relative risk = \( \frac{\text{risk group } 1}{\text{risk group } 2} = \frac{.404}{.136} = 2.97 \)

Part c:
Calculate the overall odds of “not near-sighted” for these 325 children.

\[
\text{odds} = \frac{229}{325} = \frac{229}{96} = 2.39 \text{ (to 1)}
\]

**Interpretation:** for every 2.39 children that are not near-sighted we expect 1 child with near-sightedness.

Part d:
Compute the odds ratio of near-sightedness between the night-light group and no night-light group.

odds of near-sightedness in night-light group: \( \frac{78}{115} = \frac{78}{115} = 0.68 \) (group 1)

odds of near-sightedness in no night-light group: \( \frac{18}{114} = \frac{18}{114} = 0.16 \) (group 2)

odds ratio = \( \frac{.68}{.16} = 4.25 \)

**Interpretation:** the odds of near-sightedness for night-light users are 4.25 times larger than the odds of near-sightedness for non-night-light users.