SUMMARY OF COMMONLY USED FORMULAE

1. Prevalence and Incidence

Prevalence Rate

$$= \frac{\text{\# of individuals with a health condition}}{\text{\# individuals in the relevant population}} \text{ at a given point in time}$$

Cumulative Incidence Rate

$$= \frac{\text{\# of new cases of a health outcome during a given period of time}}{\text{\# at risk of developing the outcome at the start of the period}}$$

Incidence Density

$$= \frac{\text{\# new cases of a health outcome}}{\text{Total person time of observation}} \text{ during a given period of time}$$

2. Statistical Estimation

Accuracy = Bias + Reliability

Means, proportions, and rates are all averages:

$$E(\overline{X}) = \mu \quad \text{and} \quad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \quad E(p) = \pi \quad \text{and} \quad p = \frac{\sum_{i=1}^{n} X_i}{n}, \quad E(r) = \lambda \quad \text{and} \quad r = \frac{\sum_{i=1}^{n} X_i}{n}$$

The standard deviation is a measure of the average deviation of *individual observations* around their mean:

$$\text{Standard Deviation of } X = S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_I - \overline{X}\right)^2}{n-1}}$$

The standard error is a measure of the average deviation of *summary statistics* (means, proportions, rates) around their mean assuming infinite repeated sampling:

$$\text{Normal}: \text{s.e.}\left(\overline{X}\right) = \frac{S}{\sqrt{n}} = \sqrt{\frac{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}{n}}$$

The calculation of standard errors for proportions and rates depends on whether they are in their decimal form, (0<=p<=1, 0<=r<=1), or in their integer form (percents, or per 1,000, 10,000, 100,000 etc.):

$$\text{Binomial}: \text{s.e.}(p) = \sqrt{\frac{p(1-p)}{n}}$$

where p takes on values from 0 to 1

or

$$\text{Binomial}: \text{s.e.}(\%) = \sqrt{\frac{\%(100-\%)}{n}}$$

where % takes on values from 1 to 100

$$\text{Poisson}: \text{s.e.}(r) = \sqrt{\frac{r}{n}}$$

where r takes on values from 0 to 1

or

$$\text{Poisson}: \text{s.e.}(\text{rate}) = \sqrt{\frac{\text{rate}}{n} \times \text{multiplier}}$$

where rate takes on values >=1, and
multiplier = 1,000, 10,000, 100,000

Just as summary statistics (means, proportions, and rates) are all averages and analogous to one another, so too, their standard errors are analogous under certain conditions:

$$\sqrt{\frac{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}{n}} \cong \sqrt{\frac{p(1-p)}{n}} \cong \sqrt{\frac{r}{n}}$$

In particular, when an event is very rare (a proportion is very small), the formulas for the binomial and Poisson standard errors are close approximations of one another:

$$\sqrt{\frac{p(1-p)}{n}} \cong \sqrt{\frac{p(1)}{n}} \cong \sqrt{\frac{r}{n}}$$

3. Measures of Association

Difference Measures:

$$\text{Means}: \overline{X}_1 - \overline{X}_2, \quad \text{Proportions}: p_1 - p_2, \quad \text{Rates}: r_1 - r_2$$

		Outcome: Yes	Outcome: No	
Risk Factor	Yes	a	b	n_1
	No	c	d	n_2
		m_1	m_2	N

$$\frac{a}{n_1} - \frac{c}{n_2} = p_1 - p_2 \text{ or } r_1 - r_2$$

$$\text{Attributable Risk} = p_1 - p_2$$

$$\text{Population Attributable Risk} = \text{Attributable Risk} \times \text{Prevalence of Risk}$$

$$p_0 - p_2 = (p_1 - p_2) \times \frac{n_1}{N}$$

$$\text{Preventive Fraction} = \frac{p_2 - p_1}{p_2} = 1 - \text{Relative Risk}$$

Ratio Measures:

$$\text{RR and RP} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{\frac{a}{n_1}}{\frac{c}{n_2}} = \frac{r_1}{r_2} \text{ or } \frac{p_1}{p_2}$$

Crude:

$$\text{OR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

Rothman - Boice Summary Relative Risk :

$$= \frac{\sum_{i=1}^{\#\,strata} \frac{a_i n_{2i}}{N_i}}{\sum_{i=1}^{\#\,strata} \frac{c_i n_{1i}}{N_i}}$$

Adjusted:

Mantel - Haenszel Summary Odds Ratio :

$$= \frac{\sum_{i=1}^{\#\,strata} \frac{a_i d_i}{N_i}}{\sum_{i=1}^{\#\,strata} \frac{b_i c_i}{N_i}}$$

4. Test Statistics

General form of test statistics:

$$\text{Test Statistic} = \frac{\text{Observed Association - Expected Association}}{\text{Standard Error of the Association}}$$

Test for the Difference Between Two Independent Means:

$$t = \frac{\overline{X}_1 - \overline{X}_2 - 0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where S is the standard deviation of the observed values

Test for the Difference Between Two Independent Proportions or Rates

$$\chi^2 = \sum_{i=a}^{d} \frac{(O_i - E_i)^2}{E_i}$$

where O_i = the observed value in each cell

and $E_i = \frac{\text{row total} \times \text{column total}}{N}$

or

$$z = \frac{(p_1 - p_2) - 0}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{and} \quad z = \frac{r_1 - r_2 - 0}{\sqrt{r_0\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

For the z tests, p_1 and $r_1 = \frac{a}{n_1}$, p_2 and $r_2 = \frac{c}{n_2}$, and p_0 and $r_0 = \frac{m_1}{N}$

Test for Difference Between a Proportion or Rate and a Standard

$$z = \frac{p_1 - \text{Standard}}{\sqrt{\frac{\text{Standard}(1 - \text{Standard})}{n_1}}} \quad \text{or} \quad z = \frac{r_1 - \text{Standard}}{\sqrt{\frac{\text{Standard}}{n_1}}}$$

Test for the Relative Risk and Relative Prevalence:

$$z = \frac{\ln\left(\frac{r_1}{r_2}\right) - 0}{\sqrt{\left(\frac{1}{a} \times \frac{b}{n_1}\right) + \left(\frac{1}{c} \times \frac{d}{n_2}\right)}}$$

Test for the Odds Ratio

$$z = \frac{\ln\left(\frac{a \times d}{b \times c}\right) - 0}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

5. Confidence Intervals

General form of confidence intervals:

$$\text{CI} = \text{Estimate} \pm \text{Critical Value} \times \text{Standard Error of the Estimate}$$

$$\text{CI} = \text{Association} \pm \text{Critical Value} \times \text{Standard Error of the Association}$$

Confidence intervals around single estimates:

$$CI(\overline{X}) = \overline{X} \pm 1.96 \frac{S}{\sqrt{n}}$$

$$CI(p) = p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

$$CI(r) = r \pm 1.96 \sqrt{\frac{r}{n}}$$

$$CI(d) = d \pm 1.96 \sqrt{d}$$

where d is a count
of rare health events

Confidence intervals around measures of association:

Difference Measures

$$\text{Normal}: CI = \overline{X}_1 - \overline{X}_2 \pm 1.96 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$\text{Binomial}: CI = p_1 - p_2 \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\text{Poisson}: CI = r_1 - r_2 \pm 1.96 \sqrt{\frac{r_1}{n_1} + \frac{r_2}{n_2}}$$

Ratio Measures

$$CI_{\text{RR and RP}} = e^{\left(\ln\left(\frac{r_1}{r_2}\right) \pm 1.96 \sqrt{\left(\frac{1}{a} \times \frac{b}{n_1}\right) + \left(\frac{1}{c} \times \frac{d}{n_2}\right)}\right)}$$

$$CI_{OR} = e^{\left(\ln\left(\frac{a \times d}{b \times c}\right) \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right)}$$