

# USING STATISTICS IN RESEARCH

David A. Stephens

Department of Mathematics, Imperial College

**d.stephens@imperial.ac.uk**

`stats.ma.ic.ac.uk/~das01/StatsShortCourse/`

26th November, 2003

**WEEK 4**

**POWER , SAMPLE SIZE**

**&**

**STUDY DESIGN**

## SECTION 8.

# POWER AND SAMPLE SIZE

General design issues often need to be considered before an experimental study is embarked upon.

- In clinical/animal studies, ethical considerations dictate that the “optimal” number experimental units are considered, and that resources are deployed in an “optimal” fashion.
- Economic forces mitigate against using an expansive study when a smaller one enables the same research hypotheses to be tested.

Data are collected, and hypotheses tested, within a framework of statistical inference and summary; the statistical framework also allows formal assessment of the utility of a study, and allows a statistically optimal study (with respect to a specific hypothesis) to be considered

## 8.1 STATISTICAL HYPOTHESIS TESTING

Recall the basic components of statistical hypothesis testing: in assessing which of two hypotheses,  $H_0$  and  $H_1$

$H_0$  : NULL HYPOTHESIS

$H_1$  : ALTERNATIVE HYPOTHESIS

is preferable in explaining the observed data, we need to specify, and compute the following quantities

- **TEST STATISTIC,  $T$**
- **NULL DISTRIBUTION,  $F_0$**
- **SIGNIFICANCE LEVEL,  $\alpha$**
- **P-VALUE,  $p$**
- **CRITICAL VALUE(S)/CRITICAL REGION  $\mathcal{R}$**

Recall that the **null distribution** is the probability distribution of **test statistic**  $T$  **if the null hypothesis**,  $H_0$ , **is true**; if  $t^*$  is the observed test statistic, lies in the critical region, we **reject**  $H_0$  in favour of  $H_1$ , and **do not reject**  $H_0$  otherwise.

The critical region  $\mathcal{R}$  is defined via the significance level  $\alpha$  by

$$P [T \in \mathcal{R} | H_0 \text{ is TRUE}] \leq \alpha \quad (1)$$

(where  $T \in \mathcal{R}$  means “ $T$  takes a value in the set  $\mathcal{R}$ ”).

Note that (1) considers only the distribution of  $T$  if  $H_0$  is true, and the conditional probability of rejection  $H_0$  in this case.

i.e. it is concerned only with “**false positive**” results.

In a classical test of  $H_0$  (null hypothesis) versus  $H_1$  (alternative hypothesis), there are four possible outcomes, two of which are erroneous:

1. Do not reject  $H_0$  when is  $H_0$  true.
2. Reject  $H_0$  when  $H_0$  is not true.
3. Reject  $H_0$  when  $H_0$  is true (**Type I error**).
4. Do not reject  $H_0$  when  $H_0$  is false (**Type II error**).

	Action	
	Do Not Reject $H_0$	Reject $H_0$
$H_0$ True	✓	<b>Type I Error</b>
$H_0$ not True	<b>Type II Error</b>	✓

**TYPE I : FALSE POSITIVE result**

**TYPE II : FALSE NEGATIVE result**

To construct a test, the distribution of the test statistic under  $H_0$  is used to find a critical region which will ensure that the probability of committing a type I error does not exceed some predetermined significance level  $\alpha$ .

Ideally, we would like to make the probability of making any type of error (false positive and false negative) as small as possible. For a finite sample however, this is not achievable, so a pragmatic approach that bounds the probability of a Type I error is adopted.

NOTE: For an infinite sample, we desire that the probabilities of Type I and Type II errors should both be zero.

## 8.2 POWER CALCULATIONS

The **power**,  $1 - \beta$ , of a statistical test is its ability to **correctly reject the null hypothesis**, or

$$\begin{aligned}1 - \beta &= P[\text{Reject } H_0 | H_0 \text{ is not True}] = P[T \in \mathcal{R} | H_0 \text{ is not True}] \\ &= 1 - P[\text{Do not Reject } H_0 | H_0 \text{ is not True}] \\ &= 1 - P[T \notin \mathcal{R} | H_0 \text{ is not True}]\end{aligned}$$

so that

$$\beta = P[\text{Do not Reject } H_0 | H_0 \text{ is not True}] = P[T \notin \mathcal{R} | H_0 \text{ is not True}]$$

which is based on the distribution of the test statistic under  $H_1$ .



This is the first occasion on which we have had to consider the distribution of the test statistic under the alternative hypothesis; as we shall see, in order to consider a sample size or power calculation, we must **explicitly** consider the alternative hypothesis.

Suppose that the hypothesis test concerns a parameter  $\theta$  that can take values in the parameter space  $\Theta$ . Suppose that the null and alternative hypotheses partition  $\Theta$  into two parts,  $\Theta_0$  and  $\Theta_1$ , that is

$$H_0 \quad : \quad \theta \in \Theta_0$$

$$H_1 \quad : \quad \theta \in \Theta_1$$

so that, in the simplest case

$$H_0 \quad : \quad \theta = c$$

$$H_1 \quad : \quad \theta \neq c$$

we have  $\Theta_0 \equiv \{c\}$ ,  $\Theta_1 \equiv \mathbb{R} \setminus \{c\}$

Under  $H_1$ , the probability

$$P[\text{Do not Reject } H_0 | H_0 \text{ is not True}] = P[T \notin \mathcal{R} | \theta \in \Theta_1]$$

which we previously defined as  $\beta$  will vary as the true value of  $\theta$  varies in the set  $\Theta_1$ , hence we should write  $\beta$  as a function of  $\theta$ .

**EXAMPLE:** In a **one-sample test** of a normal mean, we have  $X_1, \dots, X_n$  as a set of random variables relating to the observed data  $x_1, \dots, x_n$ , and an assumption that

$$X_i \sim N(\mu, \sigma^2)$$

for  $i = 1, \dots, n$ . If  $\sigma^2$  is known, to perform a two-sided test of equality the hypotheses would be as follows:

$$\begin{aligned} H_0 & : \mu = \theta_0 \\ H_1 & : \mu \neq \theta_0 \end{aligned}$$

The test statistic is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

and under  $H_0$ ,

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

We reject  $H_0$  at significance level  $\alpha$  if the  $z$  statistic is more extreme than the critical values of the test are

$$\mathcal{R} = \theta_0 \pm C_R \frac{\sigma}{\sqrt{n}} \qquad C_R = \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$$

Now, if  $H_1$  is true, and  $\mu = \theta$  for some value  $\theta$ , then ,  $X \sim N(\theta, \sigma^2)$ , and hence

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N\left(\frac{\theta - \theta_0}{\sigma/\sqrt{n}}, 1\right).$$

so the probability that  $z$  lies in the critical region if  $\mu = \theta$  is

$$\begin{aligned} P[T \in \mathcal{R}|\theta] &= P[Z \leq -C_R|\theta] + P[Z > C_R|\theta] \\ &= \Phi\left(-C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) + \left(1 - \Phi\left(C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right)\right) \end{aligned} \quad (2)$$

where  $\Phi$  is the standard normal distribution function.

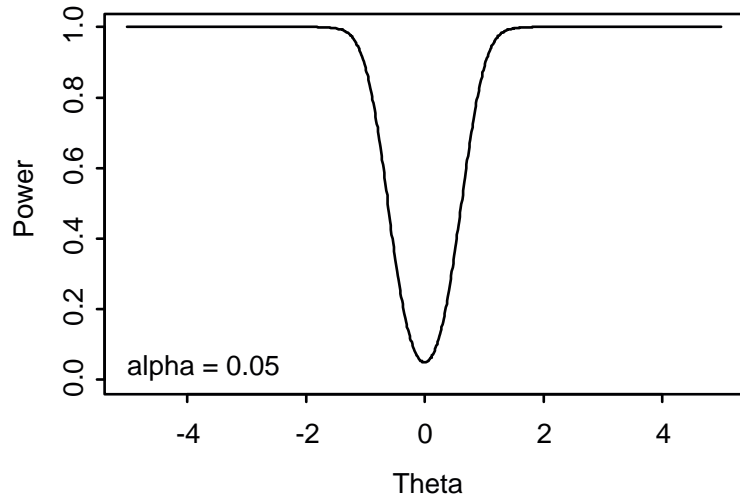
This quantity is the **power function**,  $1 - \beta(\theta)$ , when  $\mu$  is actually equal to  $\theta$ .

Hence the **probability of a Type II error** when the true is  $\beta(\theta)$ , where

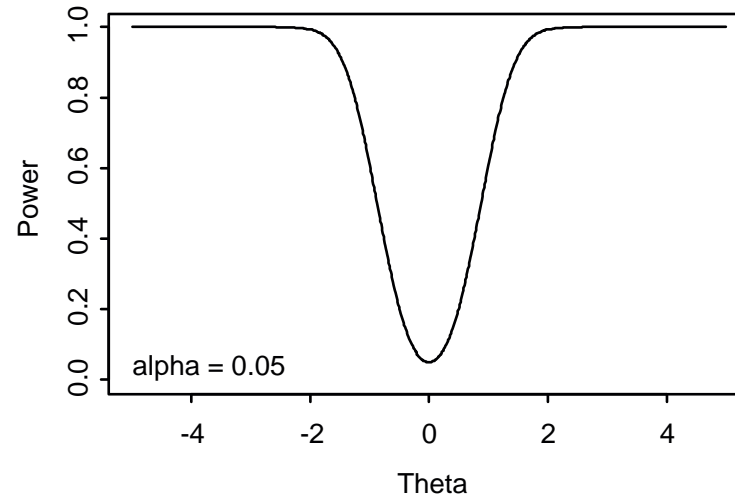
$$\begin{aligned}\beta(\theta) &= 1 - P[T \in \mathcal{R} | \theta] \\ &= \Phi\left(C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) - \Phi\left(-C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) - \left(1 - \Phi\left(C_R + \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right)\right) \\ &= \Phi\left(C_R - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) + \Phi\left(C_R + \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) - 1\end{aligned}$$

The plots below illustrate examples of power functions for different choices of  $\sigma$  and  $n$ , with  $\theta_0 = 0$ .

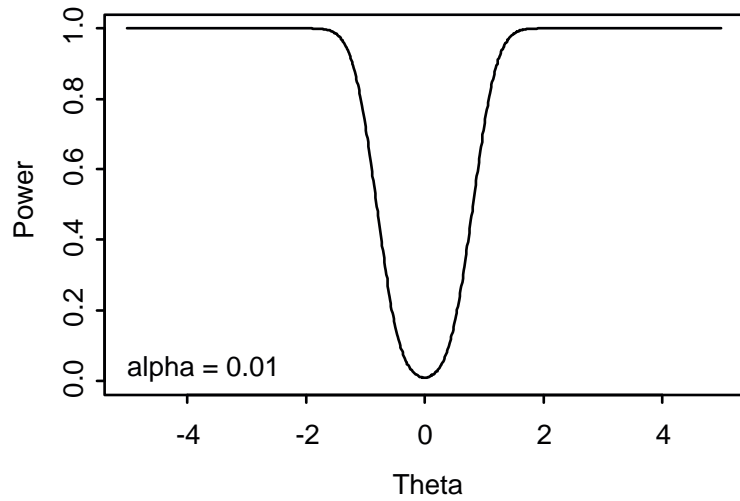
$n=10, \sigma = 1, \theta_0 = 0$



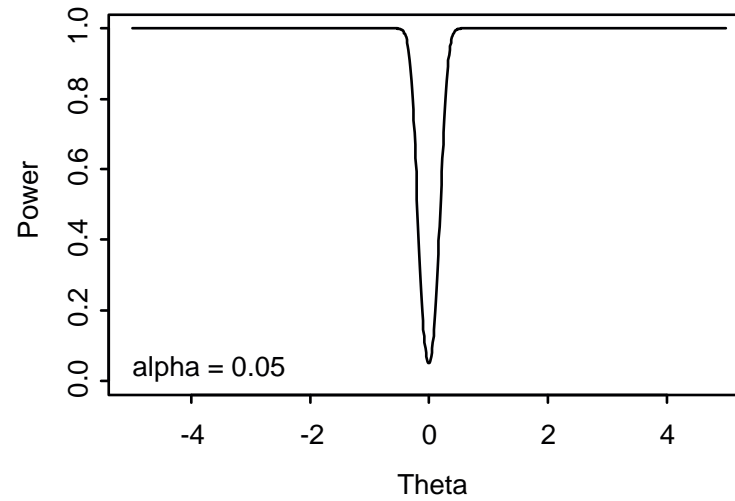
$n=5, \sigma = 1, \theta_0 = 0$



$n=10, \sigma = 1, \theta_0 = 0$



$n=100, \sigma = 1, \theta_0 = 0$



Thus for fixed  $\alpha, \theta_0, \sigma$  and  $n$ , we can compute the power function  $\beta(\theta)$  as  $\theta$  varies.

**NOTE:** The parameters in (2) appear in terms of the ratio

$$\frac{\theta - \theta_0}{\sigma}$$

that is, a **standardized difference** between the hypothesized values of  $\mu$  under the null and alternative hypotheses.

Similar calculations are available for other of the normal distribution-based tests.

## 8.2.1 ONE-SIDED TESTS

To perform a one-sided test of the hypotheses

$$H_0 : \mu = \theta_0$$

$$H_1 : \mu < \theta_0$$

the power function is

$$1 - \beta(\theta) = P[T \in \mathcal{R} | \theta] = P[Z \leq C_R(\alpha) | \theta] = \Phi\left(C_R(\alpha) - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right)$$

where

$$C_R(\alpha) = \Phi^{-1}(\alpha)$$

with a similar calculation if  $H_1 : \mu > \theta_0$

$$1 - \beta(\theta) = P[Z \geq C_R(\alpha) | \theta] = 1 - \Phi\left(C_R(\alpha) - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) \quad C_R(\alpha) = \Phi^{-1}(1 - \alpha)$$



## 8.2.2 UNKNOWN VARIANCE

If  $\sigma^2$  is unknown, to perform a two-sided test of equality the hypotheses would be as follows:

$$H_0 : \mu = \theta_0$$

$$H_1 : \mu \neq \theta_0$$

The test statistic is

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

where  $s$  is the sample standard deviation, and under  $H_0$ ,

$$T = \frac{\bar{X} - \theta_0}{s/\sqrt{n}} \sim Student(n - 1).$$

We reject  $H_0$  at significance level  $\alpha$  if the  $t$  statistic is more extreme than the critical values of the test, with

$$\mathcal{R} = \theta_0 \pm C_R \frac{s}{\sqrt{n}} \quad C_R = F_{t_n}^{-1} \left( 1 - \frac{\alpha}{2} \right)$$

where  $F_{t_k}^{-1}$  is the inverse cdf of the *Student*( $k$ ) distribution

Now, if  $H_1$  is true, and  $\mu = \theta$  for some value  $\theta$ , then

$$\begin{aligned} T &= \frac{\bar{X} - \theta_0}{s/\sqrt{n}} \\ &= \frac{\bar{X} - \theta}{s/\sqrt{n}} + \frac{\theta - \theta_0}{s/\sqrt{n}} = T_0 + \frac{\theta - \theta_0}{s/\sqrt{n}} \end{aligned}$$

where  $T_0 \sim \text{Student}(n - 1)$ .

Then the probability that  $T$  lies in the critical region is

$$\begin{aligned}1 - \beta(\theta) &= P[T \in \mathcal{R}|\theta] && (3) \\&= P\left[\frac{\bar{X} - \theta}{s/\sqrt{n}} + \frac{\theta - \theta_0}{s/\sqrt{n}} \leq -C_R|\theta\right] + P\left[\frac{\bar{X} - \theta}{s/\sqrt{n}} + \frac{\theta - \theta_0}{s/\sqrt{n}} > C_R|\theta\right] \\&= P\left[\frac{\bar{X} - \theta}{s/\sqrt{n}} \leq -C_R - \frac{\theta - \theta_0}{s/\sqrt{n}}|\theta\right] + P\left[\frac{\bar{X} - \theta}{s/\sqrt{n}} > C - \frac{\theta - \theta_0}{s/\sqrt{n}}|\theta\right] \\&= F_{t_n}^{-1}\left(-C_R - \frac{\theta - \theta_0}{s/\sqrt{n}}\right) + \left(1 - F_{t_n}^{-1}\left(C_R - \frac{\theta - \theta_0}{s/\sqrt{n}}\right)\right)\end{aligned}$$

### 8.2.3 TWO SAMPLE TESTS

In a two sample problem, if  $\sigma^2$  is unknown but common for both samples, to perform a test of the hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 = \delta$$

The test statistic is

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where  $s_P$  is the pooled sample standard deviation, and under  $H_0$ ,

$$T \sim Student(n_1 + n_2 - 2).$$

We reject  $H_0$  at significance level  $\alpha$  if the  $t$  statistic is more extreme than the critical values of the test are

$$\mathcal{R} = \pm C_R \frac{s}{\sqrt{n}} \quad C_R = F_{t_{n_1+n_2-2}}^{-1} \left( 1 - \frac{\alpha}{2} \right)$$

Now, if  $H_1$  is true, for the particular value of  $\delta$  specified

$$\begin{aligned} T &= \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{(\bar{X}_1 - \bar{X}_2) - \delta}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} + \frac{\delta}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = T_0 + \delta_0 \end{aligned}$$

say, where  $T_0 \sim Student(n_1 + n_2 - 2)$ .

Then the probability that  $T$  lies in the critical region is

$$\begin{aligned} 1 - \beta(\theta) &= P[T \in \mathcal{R}|\theta] && (4) \\ &= P[T_0 + \delta_0 \leq -C_R|\delta] + P[T_0 + \delta_0 > C_R|\delta] \\ &= P[T_0 + \delta_0 \leq -C_R - \delta_0|\delta] + P[T_0 > C_R - \delta_0|\delta] \\ &= F_{t_{n_1+n_2-2}}^{-1}(-C_R - \delta_0) + \left(1 - F_{t_{n_1+n_2-2}}^{-1}(C_R - \delta_0)\right) \end{aligned}$$

and thus the power function is calculable for any combination of  $\alpha, n_1, n_2$  and  $\delta$ .

**SUMMARY:** The adequacy of a test to distinguish between two hypotheses is a function of

- The null and alternative hypotheses;
- The target significance level  $\alpha$ ;
- The desired power to detect  $H_1$  for a specific  $\theta$ ,  $\beta(\theta)$ ;
- The variability within the population(s) under study as measured by  $\sigma$
- The sample size  $n$  (or  $n_1$  and  $n_2$ ).

Our objective is to find a relationship between the above factors and the sample size that enables us to select a sample size consistent with the desired  $\alpha$  and  $\beta(\theta)$ , typically, we will hypothesize a specific value of  $\theta$  and compute the corresponding  $\beta$ .

## 8.2.4 GENERAL POWER CONSIDERATIONS

The principles outlined above can be applied in more complicated situations

- NON-PARAMETRIC TESTS
- NON-NORMAL DATA TESTS
  - Approximate Binomial
  - Exact Binomial
- ONE-WAY/TWO-WAY ANOVA
  - number of groups/cross-categories,  $K$
  - number of observations per category,  $n_K$
  - category levels  $\theta_1, \dots, \theta_K$
- REPEATED MEASURES



The details of the power calculation are more complicated as the complexity of the experimental procedure increases, but the principles remain the same; we compute

the probability of rejecting a specified null hypothesis  
when  
a specific alternative hypothesis corresponds the actual truth

that is, we are obliged to consider both null **and** alternative hypotheses, and their impact on the distribution of the test statistic.

This is fundamentally different from the simple hypothesis testing situation, where we only consider the **null** distribution.

Therefore, a power calculation **necessarily** involves consideration of a specific alternative hypothesis, that is, equivalently, the magnitude of

- $\frac{\theta - \theta_0}{\sigma}$  in the Normal sample case with known variance  $\sigma^2$
- $\delta$  if  $\sigma^2$  is unknown
- $\delta_\pi = \pi_1 - \pi_2$  in a two-sample Binomial problem, and test of

$$H_0 \quad : \quad \pi_1 - \pi_2 = 0$$

$$H_1 \quad : \quad \pi_1 - \pi_2 = \delta_\pi$$

and so on.

How do we choose these quantities ?

- usually by consideration of a “clinically” or ”experimentally” significant difference, or an “anticipated” effect size..

## 8.3 EXAMPLES

(see Machin et al, 1997, *Sample Size Tables for Clinical Studies*)

- power/sample size for independent groups of binary, ordered, categorical and continuous data
- paired/repeated measures data
- for equivalence studies
- survival
- observer (inter-rater) agreement

## 8.4 SIMULATION-BASED CALCULATION

When analytic expressions for the power/Type II error probability are not easily available, we can do approximate power calculations by simulation means

- we formulate the test (null and alternative hypotheses, test statistic) in the usual way
- we repeatedly simulate data under the alternative hypothesis model (for fixed sample size, null model)
- we compute the power/Type II error probability empirically by evaluating the frequency with which the null hypothesis is correctly rejected.

For complicated designs (correlated data, clustered/grouped data), this is often the simplest solution.

## 8.5 SAMPLE SIZE CALCULATIONS

In all of the above, we have concentrated on computing the **achieved power** for detecting a particular effect (relative effect) in a **fixed** study (perhaps that has already been carried out).

Often it is desirable to reverse the logic and to ask if a certain power  $\beta$  to detect an effect (if it is there) is required for a specified significance level  $\alpha$ , how large would sample size  $n$  need to be ?

Such a consideration is of strategic importance in study design, and can give insight into the practicability of the proposed study.

Recall the simple concept of standard error in a mean;

$$s.e.(\bar{X}) = \frac{s}{\sqrt{n}}$$

Clearly as  $n$  increases, the standard error decreases. Thus if we wanted a standard error that was no larger than some quantity  $\epsilon$ , we would have to choose  $n$  large enough to ensure this, that is,

$$\frac{s}{\sqrt{n}} \leq \epsilon \Leftrightarrow n \geq \left(\frac{s}{\epsilon}\right)^2$$

This simple idea extends naturally to confidence intervals, and to hypothesis tests, and hence to power assessments.

In the simple case of a single normal sample with known variance, the power equation in (2) can be rearranged to be explicit in one of the other parameters if  $\beta$  is regarded as fixed.

For example, if  $\alpha, \beta, \theta_0$  and  $\theta_1$  are fixed, we can rearrange to get a sample size calculation to test for fixed difference  $\delta = \theta_1 - \theta_0$

$$n = \frac{\sigma^2 (C_R + \Phi^{-1}(1 - \beta))^2}{(\theta_1 - \theta_0)^2}$$

or standardized difference  $\Delta = \frac{|\theta_1 - \theta_0|}{\sigma}$

$$n = \frac{(C_R + \Phi^{-1}(1 - \beta))^2}{\Delta^2}$$

This idea of rearranging the power calculation to obtain a sample size extends to the general cases described above.

Other issues do need to be considered

- one-sided vs two-sided tests
- in two sample problems, the deployment of the samples to be used
  - equal proportions in the two groups
  - fixed unequal allocation ratio between subjects assigned to the two groups (in observational studies this may be necessary)
- allocation by randomization: exchangeable subjects



## 8.6 STUDY DESIGN ISSUES

The method of data collection can sometimes influence how the data are subsequently analyzed. Typically, we wish to examine the variability in a **incidence** of the response event with some **exposure** factor, possibly with the presence of **confounding** factors.

In clinical, medical or epidemiological studies, there are two types of study;

- **OBSERVATIONAL** : where the exposure **arises naturally**, and the experimenter attempts to detect differences in response
- **EXPERIMENTAL** : where the exposure is **determined by the experimenter**

The type of study used influences how the data are analyzed.

## 8.6.1 OBSERVATIONAL STUDIES

Consider the following representation of an observational study; let

- $S$  denote the **inclusion of a subject in the study**,
- $E$  denote **exposure**
- $F$  denote **incidence**; if  $F$  occurs, then we observe a **case**.

We will try to examine variation in **incidence** rate across different levels of the **exposure** factor.

Using probability theory

$$P(E \cap F \cap S) = P(E)P(F|E)P(S|E \cap F).$$

We will use this factorization to deduce estimable quantities from different observational studies that comprise the  $S$  “margin” of a  $2 \times 2 \times 2$  events table with the recorded number of observations as follows; for the events

	$E \cap S$	$E' \cap S$
$F \cap S$	$E \cap F \cap S$	$E' \cap F \cap S$
$F' \cap S$	$E \cap F' \cap S$	$E' \cap F' \cap S$

and the counts data

	$E \cap S$	$E' \cap S$	TOTAL
$F \cap S$	$n_{11}$	$n_{12}$	$n_{1.}$
$F' \cap S$	$n_{21}$	$n_{22}$	$n_{2.}$
TOTAL	$n_{.1}$	$n_{.2}$	$n_{..}$

## 8.6.2 COHORT STUDY

In a **cohort** study, the defining feature is that  $E$  and  $F$  are **independent** of  $S$  so that

$$P(E \cap F \cap S) = P(E)P(F|E)P(S) \quad \implies$$

	$E$	$E'$
$F$	$E \cap F$	$E' \cap F$
$F'$	$E \cap F'$	$E' \cap F'$

as the  $S$  and  $S'$  margins are **identical**.

It follows that **all of the following quantities are estimable**:

- **RATES OF EXPOSURE AND INCIDENCE**

$$\theta = P(E) = P(E \cap F) + P(E \cap F')$$

and

$$\phi = P(F) = P(E \cap F) + P(E' \cap F)$$

with estimates

$$\hat{\theta} = \frac{n_{.1}}{n_{..}} \qquad \hat{\phi} = \frac{n_{1.}}{n_{..}}$$

- **INCIDENCE RATES IN THE EXPOSED/UNEXPOSED GROUPS**

$$\pi_1 = P(F|E) = \frac{P(E \cap F)}{P(E)}$$

$$\pi_0 = P(F|E') = \frac{P(E' \cap F)}{P(E')}$$

with estimates

$$\hat{\pi}_1 = \frac{n_{11}}{n_{.1}} \qquad \hat{\pi}_0 = \frac{n_{12}}{n_{.2}}$$

- **THE RELATIVE RISK**

$$\rho = \frac{\pi_1}{\pi_0} = \frac{P(E \cap F)/P(E)}{P(E' \cap F)/P(E')}$$

with estimate

$$\hat{\rho} = \frac{\hat{\pi}_1}{\hat{\pi}_0} = \frac{n_{11}/n_{.1}}{n_{12}/n_{.2}}$$

- **EXPOSURE RATES IN THE CASE AND CONTROL GROUPS**

$$\gamma_1 = P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$\gamma_0 = P(E|F') = \frac{P(E \cap F')}{P(F')}$$

with estimates

$$\hat{\gamma}_1 = \frac{n_{11}}{n_1} \qquad \hat{\gamma}_0 = \frac{n_{21}}{n_2}.$$



- **ODDS ON INCIDENCE IN THE EXPOSED AND UNEXPOSED GROUPS**

$$\omega_1 = \frac{\pi_1}{1 - \pi_1} = \frac{P(E \cap F)}{P(E \cap F')}$$

$$\omega_0 = \frac{\pi_0}{1 - \pi_0} = \frac{P(E' \cap F)}{P(E' \cap F')}$$

with estimates

$$\hat{\omega}_1 = \frac{\hat{\pi}_1}{1 - \hat{\pi}_1} = \frac{n_{11}}{n_{21}} \qquad \hat{\omega}_0 = \frac{\hat{\pi}_0}{1 - \hat{\pi}_0} = \frac{n_{12}}{n_{22}}$$

- **ODDS ON EXPOSURE IN THE CASE AND CONTROL GROUPS**

$$\Omega_1 = \frac{\gamma_1}{1 - \gamma_1} = \frac{P(E \cap F)}{P(E' \cap F)}$$

$$\Omega_0 = \frac{\gamma_0}{1 - \gamma_0} = \frac{P(E \cap F')}{P(E' \cap F')}$$

with estimates

$$\hat{\Omega}_1 = \frac{\hat{\gamma}_1}{1 - \hat{\gamma}_1} = \frac{n_{11}}{n_{12}} \qquad \hat{\Omega}_0 = \frac{\hat{\gamma}_0}{1 - \hat{\gamma}_0} = \frac{n_{21}}{n_{22}}$$

- **ODDS RATIO**

$$\psi = \left\{ \begin{array}{l} \frac{P(F|E)/P(F'|E)}{P(F|E')/P(F'|E')} = \frac{P(E \cap F)/P(E \cap F')}{P(E' \cap F)/P(E' \cap F')} = \frac{\omega_1}{\omega_0} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \\ \frac{P(E|F)/P(E'|F)}{P(E|F')/P(E'|F')} = \frac{P(E \cap F)/P(E' \cap F)}{P(E \cap F')/P(E' \cap F')} = \frac{\Omega_1}{\Omega_0} = \frac{\gamma_1/(1 - \gamma_1)}{\gamma_0/(1 - \gamma_0)} \end{array} \right.$$

with estimate

$$\hat{\psi} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

### 8.6.3 CASE-CONTROL STUDY

In a **case-control** study, the defining feature is that  $E$  is **independent** of  $S$  **given**  $F$  and **given**  $F'$ , but

$$\begin{aligned} P(S|E \cap F) &= P(S|E' \cap F) & P(S|E \cap F') &= P(S|E' \cap F') \\ P(E|S \cap F) &= P(E|S' \cap F) & P(E|S \cap F') &= P(E|S' \cap F') \end{aligned}$$

In practice the design proceeds as follows; we look for incidences or **cases** and automatically include them in the study, and then we find a set of controls who do not have the “case response” and include them also. Our assumption of conditional independence of  $E$  and  $S$  given  $F$  means corresponds to an assumption of no probabilistic dependence between exposure and inclusion in the study.

Now, consider the estimation of

$$\Omega_1 = \frac{P(E|F)}{P(E'|F)} = \frac{\gamma_1}{1 - \gamma_1}$$

We have that

$$\begin{aligned}\frac{P(E \cap F \cap S)}{P(E' \cap F \cap S)} &= \frac{P(E)P(F|E)P(S|E \cap F)}{P(E')P(F|E')P(S|E' \cap F)} \\ &= \frac{P(F|E)P(E)}{P(F|E')P(E')} = \frac{P(E|F)}{P(E'|F)}\end{aligned}$$

as, by assumption,  $P(S|E \cap F) = P(S|E' \cap F)$ . A similar factorization is possible for

$$\Omega_0 = \frac{P(E|F')}{P(E'|F')} = \frac{\gamma_0}{1 - \gamma_0}$$

However, if we try to proceed in the same way for the **incidence ratio** in cases and controls, then

$$\frac{\pi_1}{\pi_0} = \frac{P(F|E)}{P(F|E')} = \frac{P(E|F) P(E')}{P(E'|F) P(E)} = \Omega_1 \left( \frac{1 - \theta}{\theta} \right)$$

and the simplification cannot proceed further; we have no way of estimating  $\theta$ , the exposure rate in the population. Furthermore, for the **odds on incidence** in the exposed group, we might try a similar approach to above and examine

$$\begin{aligned} \frac{P(E \cap F \cap S)}{P(E \cap F' \cap S)} &= \frac{P(E)P(F|E)P(S|E \cap F)}{P(E)P(F'|E)P(S|E \cap F')} \\ &= \frac{P(F|E)P(S|E \cap F)}{P(F'|E)P(S|E \cap F')} = \omega_1 \frac{P(S|E \cap F)}{P(S|E \cap F')} \end{aligned}$$

Again the simplification cannot proceed further as

$$\frac{P(S|E \cap F)}{P(S|E \cap F')}$$

is indeterminate.

The case-control study design, therefore, is perhaps more efficient, but does not allow the full range of inferences to be made.

However

$$\begin{aligned} \frac{P(E \cap F \cap S) / P(E \cap F' \cap S)}{P(E' \cap F \cap S) / P(E' \cap F' \cap S)} &= \frac{P(E \cap F \cap S) / P(E' \cap F \cap S)}{P(E \cap F' \cap S) / P(E' \cap F' \cap S)} \\ &= \frac{P(E|F) / P(E'|F)}{P(E|F') / P(E'|F')} = \frac{\gamma_1 / (1 - \gamma_1)}{\gamma_0 / (1 - \gamma_0)} = \frac{\Omega_1}{\Omega_0} = \varphi \end{aligned}$$

and also

$$\frac{P(E|F) / P(E'|F)}{P(E|F') / P(E'|F')} = \frac{P(F|E) / P(F|E')}{P(F'|E) / P(F'|E')} = \frac{\pi_1 / \pi_0}{(1 - \pi_1) / (1 - \pi_0)} = \frac{\pi_1 / (1 - \pi_1)}{\pi_0 / (1 - \pi_0)} = \psi$$

and finally

$$\frac{P(E|F) / P(E'|F)}{P(E|F') / P(E'|F')} = \frac{P(F|E) / P(F'|E)}{P(F|E') / P(F'|E')} = \frac{\pi_1 / (1 - \pi_1)}{\pi_0 / (1 - \pi_0)} = \frac{\omega_1}{\omega_0} = \psi$$



It follows that **only the following quantities are estimable in the absence of other knowledge**

- **EXPOSURE RATES IN THE CASE AND CONTROL GROUPS**

$$\gamma_1 = P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$\gamma_0 = P(E|F') = \frac{P(E \cap F)}{P(F')}$$

with estimates

$$\hat{\gamma}_1 = \frac{n_{11}}{n_1.} \qquad \hat{\gamma}_0 = \frac{n_{21}}{n_2.}$$

- **ODDS ON EXPOSURE IN THE CASE AND CONTROL GROUPS**

$$\Omega_1 = \frac{\gamma_1}{1 - \gamma_1} = \frac{P(E \cap F)}{P(E' \cap F)}$$

$$\Omega_0 = \frac{\gamma_0}{1 - \gamma_0} = \frac{P(E \cap F')}{P(E' \cap F')}$$

with estimates

$$\hat{\Omega}_1 = \frac{\hat{\gamma}_1}{1 - \hat{\gamma}_1} = \frac{n_{11}}{n_{12}} \qquad \hat{\Omega}_0 = \frac{\hat{\gamma}_0}{1 - \hat{\gamma}_0} = \frac{n_{21}}{n_{22}}$$

- **ODDS RATIO**

$$\psi = \left\{ \begin{array}{l} \frac{P(F|E)/P(F'|E)}{P(F|E')/P(F'|E')} = \frac{P(E \cap F)/P(E \cap F')}{P(E' \cap F)/P(E' \cap F')} = \frac{\omega_1}{\omega_0} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \\ \frac{P(E|F)/P(E'|F)}{P(E|F')/P(E'|F')} = \frac{P(E \cap F)/P(E' \cap F)}{P(E \cap F')/P(E' \cap F')} = \frac{\Omega_1}{\Omega_0} = \frac{\gamma_1/(1 - \gamma_1)}{\gamma_0/(1 - \gamma_0)} \end{array} \right.$$

with estimate

$$\hat{\psi} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

**EXAMPLE: LIMITATION OF CASE CONTROL STUDIES**

An illustration of why case-control studies are limited in their usefulness is presented below; fixing  $\gamma_1 = 0.2$  and  $\gamma_0 = 0.1$  and changing the size of the CONTROLS group. In Table 1

	$E \cap S$	$E' \cap S$	TOTAL
CASES	20	80	100
CONTROLS	100	900	1000
TOTAL	120	980	1100

and in Table 2

	$E \cap S$	$E' \cap S$	TOTAL
CASES	20	80	100
CONTROLS	500	4500	5000
TOTAL	520	4580	5100

Then clearly if we estimate  $\gamma_1$  and  $\gamma_0$ , we recover the true values 0.2 and 0.1, and in each case

$$\text{TABLE 1: } \hat{\psi} = \frac{20 \times 900}{80 \times 100} = \frac{9}{4}$$

$$\text{TABLE 2: } \hat{\psi} = \frac{20 \times 4500}{80 \times 500} = \frac{9}{4}$$

but if we try to estimate, for example  $\pi_1$  and  $\pi_0$  in the same way that we would for a cohort study, we get different results from the two tables

$$\text{TABLE 1: } \hat{\pi}_1 = \frac{20}{120} = \frac{1}{6} \qquad \hat{\pi}_0 = \frac{80}{980} = \frac{4}{49}$$

$$\text{TABLE 2: } \hat{\pi}_1 = \frac{20}{520} = \frac{1}{26} \qquad \hat{\pi}_0 = \frac{80}{4580} = \frac{4}{229}$$

The row totals, corresponding to the total numbers of **cases** and **controls**,  $n_1$  and  $n_2$ , are fixed by the experimenter, and we do **not have a random sample of exposed and unexposed individuals** from the population. In a cohort study, only the total cohort size,  $n_{..}$ , is fixed.

## 8.6.4 STANDARD ERRORS FOR EFFECT SIZES

In a  $2 \times 2$  table analysis, our estimates of key parameters are functions of the counts in the table; these estimates have associated (estimated) standard errors that allow construction of confidence intervals for the parameters, and hence permit hypothesis testing.

Recall the counts data for individuals in the study

	$E$	$E'$	TOTAL
$F$	$n_{11}$	$n_{12}$	$n_{1.}$
$F'$	$n_{21}$	$n_{22}$	$n_{2.}$
TOTAL	$n_{.1}$	$n_{.2}$	$n_{..}$

Then we have the following estimates and estimated standard errors for effect sizes; we typically examine such quantities on the (natural) log scale:

- The log **relative-risk**

$$\log \hat{\rho} = \log \frac{\hat{\pi}_1}{\hat{\pi}_0} = \log \left( \frac{n_{11}/n_{.1}}{n_{12}/n_{.2}} \right)$$

with **estimated standard error**

$$\sqrt{\left( \frac{1}{n_{11}} - \frac{1}{n_{11} + n_{21}} \right) + \left( \frac{1}{n_{12}} - \frac{1}{n_{12} + n_{22}} \right)}$$

- The log **odds ratio**

$$\log \hat{\psi} = \log \left( \frac{n_{11}n_{22}}{n_{12}n_{21}} \right)$$

with **estimated standard error**

$$\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{21}} + \frac{1}{n_{12}} + \frac{1}{n_{22}}}$$

## 8.6.5 EXPERIMENTAL STUDIES

Experimental studies are studies where the exposure factor is determined by the experimenter during the study

- treatment/control
- drug/placebo
- dose level 1, 2, 3, ...,  $K$

An experimental study is a special type of **cohort** study



The most common type of experimental study is a **randomized controlled trial**.

This is a study design where treatments, interventions, or enrollment into different study groups are assigned by **random** allocation rather than by conscious decisions of clinicians or patients.

If the sample size is large enough, this study design avoids problems of **bias** and **confounding** variables by assuring that both known and unknown determinants of outcome are **evenly distributed between treatment and control groups**.

*“The importance of randomization cannot be over stressed. Randomization is necessary for conclusions drawn from the experiment to be correct, unambiguous and defensible.”*

<http://www.itl.nist.gov/div898/handbook/pri/section7/pri7.htm>

Strategy for randomization:

- identify possible confounding factors
- partition (prospectively/retrospectively) experimental units into homogeneous subgroups according to confounding factors
- within in subgroup, allocate units to treatment/exposure groups at random.