

## INTRODUCTION

---

Meta-analysis, a term coined by Glass (1976), is intended to provide *the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings*.

Meta-analysis, or research synthesis, or research integration is precisely a scientific method to accomplish this goal by applying sound statistical procedures, and indeed it has a long and old history. The very invention of least squares by Legendre (1805) and Gauss (1809) is an attempt to solve just a unique problem of meta-analysis: use of astronomical observations collected at several observatories to estimate the orbit of comets and to determine meridian arcs in geodesy (Stigler, 1986). In order to determine the relationship between mortality and inoculation with a vaccine for enteric fever, Pearson (1904) used data from five small independent samples and computed a pooled estimate of correlation between mortality and inoculation in order to evaluate the efficacy of the vaccine. As an early application of meta-analysis in the physical sciences, Birge (1932) combined estimates across experiments at different laboratories to establish reference values for some fundamental constants in physics. Early works of Cochran (1937), Yates and Cochran (1938), Tippett (1931), and Fisher (1932) dealt with combining information across experiments in the agricultural sciences in order to derive estimates of treatment effects and test their significance. Likewise, there

are plenty of applications of meta-analysis in the fields of education, medicine, and social sciences, some of which are briefly described below.

In the field of education, meta-analysis is useful in combining studies about coaching effectiveness to improve Scholastic Aptitude Test (SAT) scores in verbal and math (Rubin, 1981; DerSimonian and Laird, 1983), in studying the effect of open education on (i) attitude of students toward school and (ii) student independence and self-reliance, and in combining studies about the relationship between teacher indirectness and student achievement (Hedges and Olkin, 1985). In social science, there is a need to combine several studies of gender differences in separate categories of quantitative ability, verbal ability, and visual-spatial ability (Hedges and Olkin, 1985). For some novel applications of meta-analysis in the field of medicine, we refer to Pauler and Wakefield (2000) for three applications involving dentrifice data, antihypertension data, and preeclampsia data, to Berry (2000) for questions about benefits and risks of mammography of women based on six studies, to Brophy and Joseph (2000) for meta-analysis involving three studies to compare streptokinase and tissue-plasminogen activator to reduce mortality following an acute myocardial infarction, and lastly to Dominici and Parmigiani (2000) for an application of meta-analysis involving studies in which outcomes are reported on continuous variables for some medical outcomes in some studies and on binary variables on similar medical outcomes in some other studies. Of course, there are numerous other diverse applications of meta-analysis in many other fields. We mention several applications below.

**A: Business applications.** In the context of business management and administration, one often encounters several studies with a common effect, and the problem then is of drawing suitable inference about the common effect based on the information from all the studies. Here are some examples. In the context of studying price elasticity, Tellis (1988) reports results from 42 studies! Sethuraman (1995) performed meta-analysis of national brand and store brand cross-promotional price elasticities. Lodish et al. (1995) reported results of 389 real world split cable TV advertising experiments: How TV advertising works? Churchill et al. (1985) reported meta-analysis of the determinants of salesperson performance. Farley and Lehmann (1986, 2001) and Farley, Lehmann, and Sawyer (1995) emphasized the important role of meta-analysis in international research in marketing. A current major thrust in marketing has been an attempt to create global products and brands while retaining local requirements: think global, act local. Deciding which elements of which products can be produced globally and which locally requires meta-analysis of each of the elements.

**B: Environmental applications.** In the context of environmental problems, there are several situations where the meta-analysis methods can be successfully applied. Here is a partial list of such applications.

**Evaluation of superfund cleanup technologies** (Sinha, O'Brien, and Smith, 1991; Sinha and Sinha, 1995). Cleaning up of superfund waste sites (nuclear/chemical/biological) at the National Priorities List (NPL), based on an index comprising four measures, air, groundwater, soil, and surface water, often requires innova-

tive/extremely expensive technology. A critical study of performance of the suggested technologies after a certain amount of time is highly desirable. If found useful, the technologies can be encouraged to continue at the same site. If otherwise, this should be determined as soon as possible so that suitable corrective measures can be taken. Towards this end, a common procedure is to study a prerediation baseline sample and an interim sample taken after a certain period of operation of the technology and test if a desirable percentage of the total contaminant has been removed. Comparison of a few such technologies can be based on several studies, and a data synthesis or pooling of evidence is very natural here in order to determine the final ranking of the technologies.

**Assessment of gasoline quality** (Yu, Sun, and Sinha, 2002). The U.S. Environmental Protection Agency (EPA) evaluates/regulates gasoline quality based on what is known as Reid vapor pressure (RVP). Samples of gasoline are taken from various pumps and RVPs are measured in two ways: on-site at the field level (cheap and quick) and also off-site at the laboratory level (expensive/higher precision). This usually results in two types of data: field data and lab data. Gasoline quality based on RVP can then be determined combining the evidence in both the data sets—a clear application of meta-analysis!

**Water quality in Hillsdale Lake** (Li, Nussbaum, and Sinha, 2000). Hillsdale Lake, a large federal reservoir located about 30 miles from the Kansas City metropolitan area, was authorized by the U.S. Congress in 1954 as part of a comprehensive flood control plan for the Osage and Missouri River Basins. The lake is a major recreational resource—over 500,000 visitors annually—and is also a significant source of drinking water. It is therefore essential that the water quality in this lake, as measured by Secchi depth, be regulated regularly. To achieve this, typically data from a survey of lake users in the various categories of swimming, fishing, boating, skiing, and water sports can be collected and analyzed in order to establish what level of water clarity users perceive as good. Again, it is quite possible that several studies are conducted for this purpose, and there is a need to pool the evidence from such studies to arrive at an overall conclusion about the water clarity level.

**A comparison of CMW and DPW for groundwater monitoring** (Li, 2000). Long-term monitoring of contamination of groundwater at former military land sites is performed by boring wells into the ground at predetermined locations and then assessing trace amounts of certain chemicals. There are two well-known methods for this purpose: an expensive traditional method of conventionally monitored wells (CMWs) and a relatively cheaper new methodology of direct push wells (DPWs). In order to compare these two methods, a joint study was conducted by the United States Air Force with the EPA to evaluate the assessment of pollutants. The former Hanscom Air Force Base (HAFB) located in Middlesex County, Massachusetts, and straddling the towns of Bedford, Concord, Lexington, and Lincoln was selected as the study site, and groundwater samples were collected for an assessment of long-term monitoring with both CMWs and DPWs based on 31 paired well locations. Data were collected on nine volatile organic carbons (VOCs): vinyl chloride, 1,1-DCA, benzene, toluene,

*o*-xylene, *trans*-1,2-DCE, TCE, and 1,4-DCA—labeled as VOC1, VOC2, VOC3, VOC4, VOC5, VOC6, VOC7, VOC8, and VOC9. The site was divided into three regions, and data were collected separately in each region. It is then in the spirit of meta-analysis that we combine the results of the three regions and decide if the two methods perform equally or there is a significant difference.

**Effect of second-hand smoking on women.** This of course is a vital environmental issue with a potential for adverse health effects. Several studies were conducted in many parts of the world to determine if second-hand smoking is harmful for women, and it is absolutely essential that we carry out a meta-analysis, pooling the evidence from all the studies, in order to find out the underlying state of the matter. The relevant data set is reported in Section 18.6. We should mention that based on a suitable meta-analysis of the collected information, an advisory committee of the EPA designated environmental tobacco smoke as a carcinogen.

**C: Health sciences applications.** In the context of medicine or health science problems, there are several situations where the meta-analysis methods can be successfully applied. Here is a partial list of such applications.

**Antiplatelet drug for patients with ischemic attacks.** In 1988, the question of whether to prescribe an antiplatelet drug for patients with transient ischemic attacks to prevent stroke was controversial. At that time, many randomized trials of antiplatelet drugs to treat patients with cerebrovascular disease have been completed, but the studies were variable in question and their results were contradictory. A meta-analysis of these studies by the Antiplatelet Trialist' Collaboration (1988) found a highly significant 22% reduction in the estimated relative risk of stroke, myocardial infarction, and vascular death in patients with cerebrovascular disease who were treated with an antiplatelet drug.

**Functional dyspepsia** (Allescher et al., 2001). Nonulcer dyspepsia is characterized by a variety of upper abdominal symptoms in the absence of organic disease. Within the general population, dyspepsia is very common, and as a result, empirical therapy without prior diagnostic procedures has been recommended for the management of these patients. Both acid-suppressive substances such as histamine  $H_2$ -receptor antagonists ( $H_2$ -RAs) and gastroprokinetics have been suggested as first-line, empirical therapy. Clinical trials of  $H_2$ -RAs have yielded somewhat contradictory results and benefit seems largely confined to refluxlike or ulcerlike dyspepsia subgroups. All these findings came out of an appropriate meta-analysis study.

**Dentifrice** (Johnson, 1993). In a series of nine randomized controlled clinical trials, sodium monofluorophosphate (SMFP) was compared to sodium fluoride (NaF) dentifrice in the prevention of caries development. The data consist of treatment differences,  $NaF_i - SMFP_i$ , where  $NaF_i$  is the change from baseline in the decayed/missing (due to caries)/filled-surface dental index at three years follow-up for regular use of NaF and  $SMFP_i$  is defined similarly for  $i = 1, \dots, 9$ . A statistical meta-analysis is in order here.

**Recovery time after anaesthesia** (Whitehead, 2002). A multicenter study with nine centers was undertaken to compare two anaesthetic agents undergoing short surgical procedures, where rapid recovery is important. The response of interest is the recovery time [time from when the anaesthetic gases are turned off until the patients open their eyes (in minutes)]. A meta-analysis would be quite appropriate in this context. Incidentally, a logarithmic transformation of the underlying data produces almost normal data.

As the scope of meta-analysis grew over the years, several terminologies also came into existence, such as quantitative research synthesis, pooling of evidence, or creating an overview. While most of the early works, including Mosteller and Bush (1954), provided a logical foundation for meta-analysis, the appearance of several books, notably Glass, McGaw, and Smith (1981), Hunter, Schmidt, and Jackson (1982), Rosenthal (1984), Hedges and Olkin (1985), and the edited volume by Cooper and Hedges (1994), and literally thousands of meta-analytic papers during the last 20 years or so, primarily covering applications in health sciences and education, has made the subject to have a very special role in diverse fields of applications.

The essential character of meta-analysis is that it is the statistical analysis of the summary findings of many empirical studies, which are called primary analyses, all targeted towards a common goal. However, differences among the constituent studies due to sampling designs, presence of different covariates, and so on, can and do exist while sharing a common objective. A fundamental assumption behind conducting a meta-analysis or pooling of evidence or information or data across studies in order to obtain an average effect across all studies is that the size of the effect (basic parameter of interest) reported in each study is an estimate of a common effect size of the whole population of studies. It is therefore essential to test for homogeneity of population effect sizes across studies before conducting a meta-analysis if obtaining an estimate of average effect or its test is the primary goal of the meta-analysis.

The notion of effect size is central to many meta-analysis studies which often deal with comparing two treatments, control and experimental, in an effort to find out if there is a significant difference between the two. In the case of continuous measurements, a standardized mean difference plays an important role to measure such a difference. In the case of qualitative attributes, the difference or ratio of two proportions, odds ratio, and  $\varphi$  coefficient are used to capture such differences. Again, when the objective is to study the relationship between two variables, an obvious choice is the usual correlation coefficient.

Recent meta-analytic work, however, concentrates on discovering and explaining variations in effect sizes rather than assuming that they remain the same across studies, which is perhaps rarely the case owing to uncontrollable differences in study contexts, designs, treatments, and subjects. When results of several scientific studies of the same phenomena exist and more or less agree, by conducting an appropriate test of homogeneity and accepting the hypothesis of homogeneity, the case for summarizing results of all studies with a single average effect size can be strengthened and defended. If, however, this hypothesis is rejected, no single number can adequately account for the variety of reported results. Thus, if the results from various studies differ either

significantly or even marginally, we should make an attempt to investigate methods to account for the variability by further work. This is precisely the spirit of some recent research in meta-analysis using random and mixed effects models, allowing inclusion of trial-specific covariates which may explain a part of the observed heterogeneity. In other words, a set of conflicting findings from different studies is looked upon as an opportunity for learning and discovering the sources of variation among the reported outcomes rather than a cause for dismay.

While most common meta-analysis applications involve comparison of just one variable (experimental) with another (control), multivariate data can also arise in meta-analysis due to several reasons. First, the primary studies themselves can be multivariate in nature because these studies may measure multiple outcomes for each subject and are typically known as multiple-endpoint studies. It should, however, be noted that not all studies in a review would have the same set of outcomes. For example, studies of SAT do not all report math and verbal scores. In fact, only about half of the studies dealt with in Becker (1990) provided coaching results for both math and verbal! Secondly, multivariate data may arise when primary studies involve several comparisons among groups based on a single outcome. As an example, Ryan, Blakeslee, and Furst (1986) studied the effects of practice on motor skill levels on the basis of a five-group design, four different kinds of practice groups and one no-practice group, thus leading to comparisons of multivariate data. These kinds of studies are usually known as multiple-treatment studies.

As mentioned earlier, although most statistical methods of meta-analysis focus on deriving and studying properties of a common estimated effect which is supposed to exist across all studies, when heterogeneity across studies is believed to exist, a meta-analyst must estimate the extent and sources of heterogeneity among studies if the hypothesis of homogeneity is not found to be tenable. While fixed effects models discussed in this book under the assumption of homogeneous effects sizes continue to be the most common method of meta-analysis, the assumption of homogeneity given variability among studies due to varying research and evaluation protocols may be unrealistic. In such cases, a random effects model which avoids the homogeneity assumption and models effects as random and coming from a distribution is recommended. The various study effects are believed to arise from a population, and random effects models borrow strength across studies in providing estimates of both study-specific effects and underlying population effect.

Whether a fixed effects model or a random effects model, a Bayesian approach considers all parameters (population effect sizes for fixed effects models, in particular) as random and coming from a superpopulation with its own parameters. There are several advantages for a Bayesian approach to meta-analysis. The Bayesian paradigm provides in a very natural way a method for data synthesis from all studies by incorporating model and parameter uncertainty. Moreover, a predictive distribution for future observations coming from any study, which may be a quantity of central interest to some decision makers, can be easily developed based on what have been already observed. The use of Bayesian hierarchical models often leads to more appropriate estimates of parameters compared to the asymptotic ones arising from maximum

likelihood, especially in the case of small sample sizes of component studies, which is typical in meta-analysis.

There are at least two other vital issues with meta-analysis procedures. Although it is true that most of the primary studies to be included in a meta-analysis provide a complete background of the problem being considered along with relevant entire or summary data, it also happens sometimes that some studies report only the ultimate finding in terms of the sign of the estimated underlying effect size being positive or negative or in terms of the significance or nonsignificance of the test for the absence of an effect size. It then poses a challenge for the statisticians to develop suitable statistical procedures to take into account this kind of incomplete or scanty information to carry out meta-analysis. Fortunately, there are techniques under the category of vote counting procedures to effectively deal with such situations.

The problem of selection or publication bias is rather crucial in the context of meta-analysis since the reported studies on which meta-analysis is typically based tend to be mostly significant and there could be many potential nonsignificant studies which are not reported at all simply because of their nonsignificant findings and hence these studies are not amenable to meta-analysis considerations. Such a situation is bound to happen in almost any meta-analysis scenario in spite of one's best attempt to get hold of all relevant studies, and statistically valid corrective measures should be developed and followed to deal with such a serious publication bias issue. Again, fortunately, there are some valid statistical procedures to tackle this vital problem.

We now point out that in the context of statistical meta-analysis, there are four important stages of research synthesis:

- (i) **problem formulation stage**
- (ii) **data collection stage**
- (iii) **data evaluation stage**
- (iv) **data analysis and interpretation stage.**

We describe these four stages below.

At the formulation stage of the research synthesis problem, we clearly spell out the universe to which generalizations are made (fixed effects model and random effects model) and the nature of the effect size parameters to be inferred upon (Hedges, 1994). Since research synthesis extends our knowledge through the combination and comparison of primary studies, it is important for us to indicate the perspective of the fixed effects model where the universe to which generalizations are made consists of ensembles of studies identical to those in the study sample. On the other hand, the random effects model perspective is relevant when the universe to which generalizations are made consists of a population of studies from which the study sample is drawn. Objectively and clearly defining the nature of the effect size parameter to be estimated or tested in a meta-analysis problem is also fundamental. One instance about the inference of an effect size is to ascertain the relationship between two variables  $X$  and  $Y$  in terms of either (a) estimation of the magnitude of the relationship (effect size) along with an indication of the accuracy or the reliability of the estimated

effect size (standard error or confidence interval) or (b) a test of significance of the difference between the realized effect size and the effect size expected under the null hypothesis of no relation between  $X$  and  $Y$ . Some other common effect size measures are given by the standardized difference of two means, standardized difference of two proportions, difference of two correlations, ratio of proportions, odds ratio, risk ratio, and so on. We have elaborated on all these measures in Chapter 2.

The data collection or literature search stage in research synthesis is indeed very challenging. This is of course different from primary analysis of studies. There are usually five major modes of searching for sources of primary research, namely, manual and computer search of subject indexes from abstract databases, footnote chasing (references in review/nonreview papers and books), consultation (formal/informal requests, conferences), browsing through library shelves, and manual and computer citation searches (White, 1994). While it is hoped that these search procedures, in addition to reviewing books/book chapters, research/technical reports, conference papers, and other possible sources, would lead to an exhaustive collection of relevant literature for the problem under study, sometimes we also need to use special ways and means to retrieve what are known as fugitive literature and information appearing in unpublished papers/technical reports, unpublished dissertations/master's theses, and the like. In this context, publication bias is quite relevant while doing the research synthesis, bearing in mind the fact that often research leading to nonsignificant conclusions are not reported at all or rarely so (the well known file-drawer problem). We have addressed this important issue in Chapter 13.

Not all studies available for meta-analysis may qualify for inclusion due to various reasons. The data evaluation stage consists of carefully checking the nature and sources of primary research data, missing observations in primary data, and sources of potential bias in the primary data, all in an attempt to assign suitable weights to the various primary data sources at the time of carrying out meta-analysis or data synthesis.

Finally, the data analysis stage, which is the main purpose of this book, deals with statistically describing and combining various primary studies. Naturally what we need here is a wide collection of sound statistical methods depending on the nature of the underlying problem. We describe ways to combine various measures of effect sizes either for estimation or test or confidence interval and also ways to deal with missing values in primary studies as well as publication bias. In the sequel, we need statistical procedures for univariate and multivariate cases, discrete and continuous cases, and also frequentist and Bayesian methods.

Given the above broad spectrum of topics that can be covered under the umbrella of a book on meta-analysis, our goal in writing this book is primarily concerned with some statistical aspects of meta-analysis. As already mentioned, the heart of the enterprise of carrying out meta-analysis or synthesizing research consists of comparing and combining the results of individual primary studies of a particular, focused research question, and the emphasis is essentially on two types of statistical analysis: combining results of tests of significance of effect size and combining estimates of effect size. The effect size, as explained earlier, is a generic term referring to the magnitude of an effect or more generally the size of the relation between two variables.



Moreover, in case of diverse research findings from comparable studies, an attempt must be made to understand and point out reasons for such differences.

Keeping the above general points in mind, the outline of the book is as follows.

Chapter 2 describes various standard measures of effect size based on means, proportions,  $\varphi$  coefficient, odds ratio, and correlations. Some illustrative examples to explain the related computations and concepts are included.

Chapter 3 deals with methods of combining individual tests based on primary research with plenty of applications. This chapter is exclusively based on combination of  $P$ -values mainly because the studies which are meant for meta-analysis more often report their  $P$ -values than other details of the study. The methods described here are exact and also appear in standard textbooks on meta-analysis. It should be mentioned that there are other methods based on suitably combining often independent component test statistics. However, the sampling distributions of the combination of such test statistics may not be readily available. We discuss these aspects in detail in Chapter 5 in the context of inference about a common mean of several univariate normal populations.

Chapter 4 describes methods of combining individual estimates of effect sizes based on primary research to efficiently estimate the common effect size parameter as well as to construct its confidence interval. The methods suggested in this chapter are mainly asymptotic in nature and again are quite routine.

Chapter 5 is devoted to a detailed analysis of a special kind of meta-analysis problem, namely, inference about the common mean of several univariate normal populations with unknown and unequal variances. This problem has a long and rich history and is very significant in applications. Most of the results presented here are new and have not appeared in any textbook before. Two classic data sets are used throughout to explain the concepts.

Chapter 6 describes various tests of the important hypothesis of the homogeneity of population effect sizes in some particular models. In the context of statistical meta-analysis, one should carry out these tests of homogeneity of effect sizes before applying tools of combining the effect sizes. The results presented here are based on the review papers by Hartung and his students.

One-way random effects models, useful when the basic hypothesis of homogeneity of effect sizes does not hold, is taken up in Chapter 7. There is a huge literature on this topic and we have made an attempt to present all the important results in this connection. Typically, there are two scenarios: error variances are all equal (homogeneous case) and error variances are not equal (heterogeneous case). We have dealt with both cases. The reader will find a variety of new and novel solutions in this chapter.

Chapter 8 extends the results of the previous three chapters to the meta-analysis of comparative trials with normal outcome. Results in the fixed effects as well as in the random effects model are provided for the effect sizes difference of means, standardized difference of means, and ratio of means.

Meta-analysis procedures to analyze categorical data of both binary and ordinal nature are presented in Chapter 9. We have provided fixed effects as well as random effects results with motivating examples. This is another nice feature of the book.

Meta-regression, multivariate meta-analysis, and Bayesian meta-analysis are, respectively, presented in three subsequent chapters—10, 11, and 12. In Chapter 10, we describe meta-analysis regression procedures with one and more than one covariate with illustrations. In Chapter 11, we describe both aspects of multiple-endpoint and multiple-treatment studies. We have provided a unified Bayesian approach to meta-analysis with some examples in Chapter 12. All these three chapters provide unique features to our book.

The important concepts of publication bias and vote counting procedures, which also appear in standard textbooks, are taken up in Chapters 13 and 16. As mentioned earlier, these problems arise when we do not have access to all the literature on the subject under study and also when there is not enough evidence in the studies which are indeed available.

We describe in Chapter 14 the statistical methods for recovery of interblock information. One of the earliest applications of statistical meta-analysis (Yates, 1939, 1940; Rao, 1947) consists of combining what are known as intrablock and interblock estimates of treatment effects in the presence of random block effects in two-way mixed effects models. Early tests of significance of treatment effects in this context were based on combining the  $P$ -values using Fisher's method. Since this method does not take into account the underlying statistical structure, methods to improve it were suggested in several papers (Fedingold, 1985, 1988; Cohen and Sackrowitz, 1989; Mathew, Sinha, and Zhou, 1993; Zhou and Mathew, 1993). We describe all these procedures in detail in this chapter, which is a new and novel contribution to the vast literature on statistical meta-analysis.

A different kind of meta-analysis dealing with a combination of polls is presented in Chapter 15. This particular topic has applications in market research and is completely new. This is based on a technical report by Dasgupta and Sinha (2006).

There are many computational aspects of statistical meta-analysis which are taken up in Chapter 17 using the general statistical software packages SAS and R. Sample programs for both softwares are explained with examples.

Finally, sample data sets which are analyzed throughout the book are included in the final section of the book, Chapter 18. The References section at the end contains a long list of papers referred to in this book.

We conclude this introductory chapter with two observations. First, although we have described a variety of diverse scenarios where meta-analysis methods can be successfully applied, we have not made an attempt to do so. Our illustrations of the methods are naturally limited to our judgment and own experiences. Second, except in some chapters, virtually all of the statistical methods described in this book are based on standard large sample results for the (asymptotic) distributions of sample means, sample proportions, sample correlations, and so on, and hence due caution should be exercised when using these methods. Some of the frequently quoted results are listed below for ready reference (see Rao, 1973; Rohatgi, 1976):

1.  $X_1, \dots, X_n$  are independently and identically distributed (iid) with mean  $\mu$  and variance  $\sigma^2$ . Then, for large  $n$ ,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

that is,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).$$

This is a standard version of the celebrated central limit theorem (CLT).

2.  $X_1, \dots, X_n$  are iid with mean  $\mu$  and variance  $\sigma^2$ . Then, for large  $n$ ,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim N(0, 1),$$

where  $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ . This is an application of CLT coupled with Cramer's theorem (Slutsky's theorem).

3.  $X \sim B(n, P)$ . Then, for large  $n$ ,

$$\frac{X - nP}{\sqrt{nPQ}} \sim N(0, 1),$$

where  $Q = 1 - P$ , that is,

$$\frac{\sqrt{n}(p - P)}{\sqrt{PQ}} \sim N(0, 1),$$

where  $p = X/n$ . This is a standard application of the CLT.

4.  $X \sim B(n, P)$ . Then, for large  $n$ , writing  $p = X/n$ ,

$$\sin^{-1} \sqrt{p} \sim N\left(\sin^{-1} \sqrt{P}, \frac{1}{4n}\right).$$

This is a well-known version of Fisher's variance-stabilizing transformation applied to the binomial proportion.

5.  $(X_1, Y_1), \dots, (X_n, Y_n)$  are iid from a bivariate distribution with means  $(\mu_1, \mu_2)$ , variances  $(\sigma_1^2, \sigma_2^2)$ , and correlation  $\rho$ . Then, for large  $n$ ,

$$r \sim N\left(\rho, \frac{(1 - \rho^2)^2}{n - 1}\right),$$

where  $r$  is the usual sample correlation defined as

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2\right]^{1/2}}.$$

This is also an application of CLT coupled with Cramer's theorem (Slutsky's theorem; see Rao, 1973).

6.  $(X_1, Y_1), \dots, (X_n, Y_n)$  are iid from a bivariate distribution with means  $(\mu_1, \mu_2)$ , variances  $(\sigma_1^2, \sigma_2^2)$ , and correlation  $\rho$ . Then, for large  $n$ ,

$$z \sim N\left(\zeta, \frac{1}{n-3}\right),$$

where

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \quad \text{and} \quad \zeta = \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right).$$

This is a well-known version of Fisher's variance-stabilizing transformation applied to the sample correlation coefficient.