International Journal of Development Mathematics Vol 1 Issue 4 Page No. 112 - 123

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

An Assessment of Statistical Power for Analysis of Variance with Missing Data using Multiple Imputation

Alade S. Peter[a*], Okolo Abraham[a], Akinrefon A. Adesupo[a] and Dike I. John[a]

[a]Department of Statistics, Modibbo Adama University, Yola, Adamawa State Nigeria

ARTICLE INFO

Article history:
Received 17 July 2024
Received in revised form 27 September 2024
Accepted 19 October 2024

Keywords:
Analysis of variance (ANOVA), Arbitrary, Incomplete data, Monotone, Multiple imputation, Statistical power

MSC 2020 Subject classification:
62H12, 62P05

ABSTRACT

This study proposes a framework for statistical power assessment in the analysis of variance (ANOVA) models with missing data handled through multiple imputation. While existing literature has examined the power for t-tests with missing data, extending these techniques to ANOVA designs with multiple group means remains an open challenge. The methodology involves incorporating the variance component into an existing power model derived for a pairwise test, enabling explicit power estimation for two-factor, three-factor, and split-plot ANOVA designs with incomplete data. Empirical evaluations across varying effect sizes, missing data rates (8%, 16%, 40%), and numbers of imputations (20, 30, 40, 100) were conducted for both monotone and arbitrary patterns of missing data. The findings reveal that while power remains high at low missing data rates, it declines substantially with increasing missingness, particularly for higher proportions of missing data. Furthermore, arbitrary patterns of missing data exhibit slightly lower power compared to monotone patterns. The number of required imputations to achieve desired power levels depends on the missing data rate, with more imputations needed for higher proportions of missing data. This study provides a framework for power analysis in ANOVA models with missing data, addressing the critical need for reliable guidance in designing studies with incomplete data.

1. Introduction

Missing data poses a formidable obstacle to robust and reliable statistical analysis across a broad spectrum of research disciplines, ranging from the social sciences to biomedical investigations. The manifestation of missingness, emanating from diverse sources such as participant non-response, data entry errors, and study attrition, presents a profound challenge to researchers striving to attain comprehensive and accurate insights from their data. Statistical power, a fundamental construct in hypothesis testing and experimental design, serves as a guiding principle for researchers, illuminating the probability of correctly rejecting a false null hypothesis or, conversely, detecting a true effect when it exists. Achieving adequate statistical power is often beset by challenges, particularly in the presence of missing data, which can lead to underpowered studies, inaccurate conclusions, and inefficient resource allocation (Enders, 2022). Among the array of strategies (complete case analysis, maximum likelihood estimator, listwise deletion, single mean imputation to mention but a few) devised to grapple with missing data, multiple imputation (MI) emerges as a beacon of hope, offering a principled framework for imputing missing values and harnessing the power of augmented datasets. Through the synthesis of these imputed datasets, MI engenders a more comprehensive and robust representation of the underlying data structure (Rubin and Little, 2020). While existing literature has made significant strides in clarifying the impact of missing data on statistical analyses, gaps persist, particularly concerning the estimation of statistical power in the context of analysis of variance (ANOVA). This study endeavours to bridge this gap by extending the framework proposed by Zha and Harel (2019) to encompass ANOVA analyses in designs involving more than two groups with incomplete data.

Statistical power, the probability of correctly rejecting a false null hypothesis and effect size, the estimated strength or magnitude of effects studied, serve as pillars upon which the strength and validity of research conclusions hinge (Lakens *et al*., 2022; Ellis, 2022; Miočević *et al*., 2022). Studies with meaningful effect sizes are essential for drawing accurate inferences and fostering the replicability and generalizability of findings. Missing data presents a

* Corresponding author. Tel.: +2348189183065
E-mail address: segunpeteralade@yahoo.com (Alade Segun.)
https://doi.org/10.62054/ijdm/0104.09

formidable challenge to researchers, with the potential to introduce bias, diminish statistical power, and undermine the integrity of analyses (Graham, 2012; Garson, 2015). The negative effects of missing data on statistical power have been well-documented, with missingness leading to a depletion of sample size and a concomitant reduction in the ability to detect true effects (Kang, 2013; Savla, 2004; Ledolter and Kardon, 2020). Conversely, augmenting sample sizes, either directly or through the restoration of missing observations, can bolster statistical power and improve the precision of parameter estimates (Maxwell *et al*., 2008). It is within this context that multiple imputation (MI) emerges as a potent tool, offering researchers a principled framework for addressing missingness while preserving statistical power. The impact of MI on statistical power has been extensively explored, with studies across domains attesting to its efficacy in enhancing the accuracy, reliability, and validity of research findings (Anderson and Williams, 2017; Smith and Johnson, 2018; Gong *et al.*, 2018). By accounting for missing data and harnessing the latent information embedded within incomplete datasets, MI can improve statistical power, reduce Type II error rates, and increase the precision of estimates, particularly in scenarios with larger amounts of missing data.

Additionally, the optimal number of imputations required to achieve sufficient power and precise parameter estimates has been a subject of ongoing discourse. While early recommendations suggested surprisingly few imputations (Rubin, 1987; Schafer, 1997), more recent studies have advocated for larger numbers, particularly in scenarios with higher missing data rates or more complex models (Graham *et al*., 2007; Bodner, 2008; Von Hippel, 2007; White *et al*., 2011). The determination of the requisite number of imputations is contingent upon factors such as the missing data rate, the number of model parameters, and the requirements for statistical power, underscoring the need for nuanced guidelines tailored to specific research contexts. ANOVA, a versatile and widely employed statistical technique for comparing means across more than two groups or experimental conditions, is predicated upon assumptions of balanced designs with equal sample sizes across groups. However, the presence of missing data can lead to unbalanced designs, violating these assumptions and potentially compromising the validity of statistical inferences (van Ginkel and Kroonenberg, 2014).

It is against this backdrop that this study aims to extend the seminal work of Zha and Harel (2019), which developed statistical power analysis techniques for t-tests in the presence of missing data handled via MI. By generalizing their approach to the ANOVA framework, this study seeks to bridge a gap in the literature and provide a power analysis tool tailored to the challenges of non-ignorable missingness and unbalanced designs. The development of a robust power analysis methodology for ANOVA with missing data will empower researchers to design studies more effectively, ensuring that they have adequate statistical power to detect meaningful effects and draw meaningful conclusions, even in the presence of missingness. The study will also contribute to the theoretical understanding of how multiple imputation performs in complex modelling situations, beyond the realm of simple two-group comparisons. By formally quantifying the impact of different missing data assumptions, missingness patterns, and imputation strategies on statistical power, this research will advance the broader field of missing data methodology and its applications in various domains.

In addition to extending the work of Zha and Harel (2019) to the ANOVA context, this study also aims to explore the performance of statistical power under different experimental designs, such as two-factor, three-factor, and split-plot designs, in the presence of missing data. These designs are widely employed in various fields, and their complex interplay with missingness patterns and mechanisms warrants careful investigation. By examining the nuances of power estimation across these diverse experimental setups, this research endeavours to provide a framework that can accommodate the myriad complexities inherent in real-world research scenarios. Furthermore, the study will delve into the interplay between missing data patterns, such as monotone or arbitrary patterns, and their consequences for statistical power. While previous studies have shown the general impact of missing data mechanisms on power, a subtle understanding of their specific effects within the context of ANOVA analyses is imperative for developing tailored and robust methodologies. Lastly, the determination of the optimal number of imputations required to achieve desired levels of power will be a focal point of this investigation. This study aims to provide evidence-based guidelines that can inform decision-making processes, ensuring a balance between computational efficiency and statistical rigour.

2. Methodology

Rubin's (1987) MI framework is reviewed. Power analysis techniques for two-sample t-tests are explained for both complete and incomplete data scenarios based on Zha and Harel (2019). Power calculations framework for two and three factors as well as the split-plot ANOVA models are presented. A step-by-step derivation is then presented to adapt power models for the three different aforementioned ANOVA designs. The mean squared error term substitutes the variance to account for multiple group comparisons in ANOVA, resulting in tailored power estimation equations for such models.

Rubin's Rule for Multiple Imputation

Let θ = estimand of interest

$\hat{\theta} =$ estimator of θ with variance σ^2

$\boldsymbol{Y} = (Y_o, Y_m)$be the complete data

Y_o = observed part of the data

and Y_m represents part of the data that is missing, with $\boldsymbol{X}$ as the covariate.

Thus, the distribution of θ can be represented as:

$$(P(\theta|X, Y_o) = \int P(\theta|X, Y_o, Y_m)\, P(Y_m|Y_o) dY_m \quad (1)$$

The consequences that follow from (1) lead to the combining rules of multiple imputation obtained by Rubin.

Requirement for Reliable Inference Using MI

To test the null hypothesis that a parameter θ is equal to a specific value θ_0, Rubin determined that the statistic t in equation (2) can be employed (van Ginkel and Kroonenberg, 2015; Zha and Harel, 2019):

$$t_v = \frac{\bar{\theta}-\theta}{\sqrt{\bar{\sigma}^2}} \quad (2)$$

which has a t distribution with v degrees of freedom.

For MI to yield valid inferences, the imputation method must be proper and randomization valid. A multiple imputation procedure is randomization valid if the posterior distribution of θ is normally distributed and approximately given as

$$\hat{\theta}_k|\boldsymbol{X}, \boldsymbol{Y_o} \sim N(\bar{\theta}_\infty, \hat{\sigma}^2_{b\,\infty})$$
$$\hat{\sigma}^2_{k\infty}|\boldsymbol{X}, \boldsymbol{Y_o} \sim (\hat{\sigma}^2_{w\,\infty} \ll \hat{\sigma}^2_{b\,\infty}) \quad (3)$$

The quantities $\hat{\sigma}^2_{w\infty}$, $\hat{\sigma}^2_{b\infty}$ and $\bar{\theta}_\infty$ all approaches $\hat{\sigma}^2_w$, $\hat{\sigma}^2_b$ and $\bar{\theta}_k$ respectively as the number of imputations (m) becomes large. This allows one to make probability statements about the hypothesized values of θ. This suffices to say that the expectation of $\hat{\sigma}^2_{w\infty}$, $\hat{\sigma}^2_{b\infty}$ and $\bar{\theta}_\infty$ becomes $\hat{\sigma}^2_{wE}$, $\hat{\sigma}^2_{bE}$ and $\bar{\theta}_{kE}$ respectively. For MI to be valid, the complete data must also be randomization valid (Zha and Harel, 2019; Abowd, 2005).

According to Rubin (1987), MI statistics have the following distributions:

$$\hat{\theta}_k|\boldsymbol{X}, \boldsymbol{Y_o} \sim N(\theta, \hat{\sigma}^2_E)$$
$$\hat{\sigma}^2_k|\boldsymbol{X}, \boldsymbol{Y_o} \sim (\hat{\sigma}^2_{w\,E} \ll \hat{\sigma}^2_{b\,E})$$
$$\left((m-1)\frac{\hat{\sigma}^2_b}{\hat{\sigma}^2_{b\,E}}|\boldsymbol{X}, \boldsymbol{Y_o}\right) = q \sim \chi^2_{m-1} \quad (4)$$

Power Calculation with MI

Zha and Harel (2019) derived a general model for calculating statistical power for the hypothesis: H_0: $\hat{\theta} = \theta_0$ versus H_1: $\hat{\theta} = \theta_1$:

T-test for complete data

T-test for complete data is derived as follows from (2)

$$\text{Power} = 1 - \beta$$
$$= P(\text{Rejecting } H_O/H_1)$$
$$= P\left(\frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_w}} > t_{v,\frac{\alpha}{2}}\right) + P\left(\frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_w}} < -t_{v,\frac{\alpha}{2}}\right) \quad (5)$$

As the sample size becomes large the power can be approximated as

$$\approx P\left(\frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_{wE}}} > t_{v,\frac{\alpha}{2}}\right) + P\left(\frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_{wE}}} < -t_{v,\frac{\alpha}{2}}\right)$$

$$= P\left(z > t_{v,\frac{\alpha}{2}} + \frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_{w}}}\right) + P\left(z < -t_{v,\frac{\alpha}{2}} + \frac{\theta_0-\theta_1}{\sqrt{\hat{\sigma}^2_{w}}}\right) \quad (6)$$

where $\hat{\sigma}^2_{wE}$ (The expected value of the within imputation variance) approaches $\hat{\sigma}^2_{w}$ as the sample size becomes large.

T-test for incomplete data

The general formula for calculating statistical power has been derived by Zha and Harel (2019) as:

$$P(\theta \notin C|\boldsymbol{X},\boldsymbol{Y},\ \theta=\theta_0) = P\left(z < \frac{\theta_0-\theta_1}{\left(\hat{\sigma}^2_{wE}(1+r_E)\right)^{\frac{1}{2}}} - z_{\frac{\alpha}{2}}\middle|\boldsymbol{X},\boldsymbol{Y}\right) + P\left(z > \frac{\theta_0-\theta_1}{\left(\hat{\sigma}^2_{wE}(1+r_E)\right)^{\frac{1}{2}}} + z_{\frac{\alpha}{2}}\middle|\boldsymbol{X},\boldsymbol{Y}\right) \quad (7)$$

where $r_E = \frac{\left(1+\frac{1}{m}\right)\hat{\sigma}^2_b}{\hat{\sigma}^2_w}$ (the relative increase in variance due to missingness) (8)

This work extended Zha and Harel's (2019) derivation to a situation of multiple population means, that is analysis of variance. The Cohen's D was used as an estimate of the effect between θ_0 and θ_1.

Generalization to ANOVA designs

Let $\hat{\sigma}^2_{wE}$ denote the population expected value of the within imputation variance. Hansen *et al*. (1953) defined $\hat{\sigma}^2_{wE}$ to be:

$$\hat{\sigma}^2_{wE} = \left(\frac{1}{n_1}-\frac{1}{N}\right)\sigma^2 \quad (9)$$

For Two-factor Design

where n_1 is the number of complete cases and N is the total sample size.

In a two-way ANOVA with factors A and B, the MSE is calculated as:

$$MSE = SSresidual/(N-AB-1) = \frac{\sum\sum(Y_{ijk}-\hat{\mu})^2}{(N-AB-1)} \quad (10)$$

where:

SSresidual = residual sum of squares

A = number of levels of factor A

B = number of levels of factor B

To extend the power analysis model to a two-factor ANOVA, we substituted equation (10) into equation (9) to get:

$$\hat{\sigma}^2_{wE} = \left(\frac{1}{n_1}-\frac{1}{N}\right) * MSE \quad (11)$$

Substituting equation (11), the within-imputation variance $\hat{\sigma}^2_{wE}$ into the power calculation model in equation (7) provides an estimate of power for a two-factor ANOVA with missing data handled via multiple imputation and it is given as:

$$Power = P\left(z < \frac{\theta_0-\theta_1}{\sqrt{\left(\left(\frac{1}{n_1}-\frac{1}{N}\right)*MSE*(1+r_E)\right)}} - z_{\frac{\alpha}{2}}\right) + P\left(z > \frac{\theta_0-\theta_1}{\sqrt{\left(\left(\frac{1}{n_1}-\frac{1}{N}\right)*MSE*(1+r_E)\right)}} + z_{\frac{\alpha}{2}}\right) \quad (12)$$

$$\text{Power} = \text{P}\left(\text{z} < \frac{\theta_0-\theta_1}{\sqrt{\left(\left(\frac{1}{n_1}-\frac{1}{N}\right)*\frac{\sum\sum\left(Y_{ijk}-\hat{\mu}\right)^2}{(N-A-B-1)}*(1+r_E)\right)}} - z_{\frac{\alpha}{2}}\right) + \text{P}\left(\text{z} > \frac{\theta_0-\theta_1}{\sqrt{\left(\left(\frac{1}{n_1}-\frac{1}{N}\right)*\frac{\sum\sum\left(Y_{ijk}-\hat{\mu}\right)^2}{(N-A-B-1)}*(1+r_E)\right)}} + z_{\frac{\alpha}{2}}\right) \quad (13)$$

where θ_0 and θ_1 are the hypothesized means under the null and alternative hypotheses, z is the standard normal distribution and α is the significance level.

For three-factor Design

In a three-factor ANOVA model with factors A, B, and C, the MSE is calculated as:

$$MSE = SS_{residual}/(N-ABC-1) = \frac{\sum\sum\sum\left(Y_{ijkl}-\hat{\mu}\right)^2}{(N-ABC-1)} \quad (14)$$

where:

$SSresidual = \sum\sum\sum\left(Y_{ijkl}-\hat{\mu}\right)^2$ (the residual sum of squares)

Y_{ijkl}= The observation in the $(ijkl)$th cell

$\hat{\mu}$ = The overall estimated mean

N = Total number of observations

A = Number of levels of factor A

B = Number of levels of factor B

C = Number of levels of factor C

Likewise, to extend the power analysis formula to a three-factor ANOVA, we substitute equation (14) into equation (11) and further on into the power model in (12), the power estimate for a three-factor ANOVA is given thus:

$$Power = P\left(z < \frac{\theta_0-\theta_1}{\sqrt{\left(\frac{1}{n_1}-\frac{1}{N}\right)*\left[\frac{\sum\sum\sum\left(Y_{ijkl}-\hat{\mu}\right)^2}{N-ABC-1}\right]*(1+r_E)}} - z_{\frac{\alpha}{2}}\right) + P\left(z > \frac{\theta_0-\theta_1}{\sqrt{\left(\frac{1}{n_1}-\frac{1}{N}\right)*\left[\frac{\sum\sum\sum\left(Y_{ijkl}-\hat{\mu}\right)^2}{N-ABC-1}\right]*(1+r_E)}} + z_{\frac{\alpha}{2}}\right) \quad (15)$$

Equation (15) provides the power calculation for a three-factor ANOVA with multiple imputation.

For Split-Plot Design

In a three-factor ANOVA model with factors A, B, and C, the MSE is calculated as:

$$MSE_w = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{i..}\right)^2}{ab(n-1)} \quad (16)$$

where:

$(Y_{ijk})$is the observed value of the k-th replicate in the j-th subplot of the i-th whole plot.

$(\bar{Y}_{i..})$ is the mean of all observations within the i-th whole plot.

(a) is the number of whole plot treatments.

(b) is the number of subplot treatments.

(n) is the number of replicates.

To extend the power analysis formula to a split-plot design, we substitute equation (16) into equation (11) and further on into the power model in (12), the power estimate is thus given as:

$$\text{Power} = P\left(z < \frac{\theta_0 - \theta_1}{\sqrt{\left(\frac{1}{n_1} - \frac{1}{N}\right) \times \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{i..}\right)^2}{ab(n-1)} \times (1 + r_E)}} - z_{\frac{\alpha}{2}} \right) + P\left(z > \frac{\theta_0 - \theta_1}{\sqrt{\left(\frac{1}{n_1} - \frac{1}{N}\right) \times \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{i..}\right)^2}{ab(n-1)} \times (1 + r_E)}} + z_{\frac{\alpha}{2}} \right) \quad (17)$$

Equation (17) provides the power calculation for a split-plot ANOVA design with multiple imputation.

To address missing data, multiple imputation was conducted using SPSS version 27. Missingness was treated at 8%, 16%, and 40% for each design. Additionally, 20, 30, 40, and 100 imputations were performed for each design at each level of missingness. The results of the imputations were combined or pooled using an Office 2019 spreadsheet to obtain the Rubin rules. The primary focus of the analysis is the mean square error (MSE). Alongside the parameter of interest θ (MSE), estimates were obtained for the between imputation variance, $\hat{\sigma}_b^2$, the fraction of missing Information (FMI), the relative increase in variance (RIV), and relative efficiency (RE). Cohen's D measure of effect size (Smith, 2021; Jones, 2022), with values of 0.2, 0.5, and 0.8 representing small, medium, and large effects, respectively, were employed. Tables 2, 4 and 6 show the results of the different configurations based on equations 13, 15 and 17 for the two-factor, three-factor and split-plot designs respectively.

3. Results and Discussions

The study used data from two experiments: one assessing the antifungal activity of B. prostrata extracts against fungal agents deteriorating postharvest tomatoes in Plateau State, and another evaluating different oat varieties and manure levels in a split-plot design. The first experiment by the Plateau Agricultural Development Programme (PADP) produced two data sets: a two-factor table with 60 observations analyzing aqueous and methanolic extracts, and a three-factor table with 320 observations examining various leaf extract fractions. The second experiment, reported by F. Yates in 1935, involved a split-plot design with three oat varieties and four manure levels, conducted in six blocks each subdivided into whole and split plots, ensuring a randomized complete block design for both factors (www.stat.purdue.edu/~zhanghao).

Table 1: Power estimates in two-factor, three-factor and Split-plot Designs for Complete data

Effect Size	Two-Factor	Three-factor	Split Plot
Small=0.2	0.289462	0.632742	0.099417
Medium= 0.5	0.811255	0.998408	0.495403
Large = 0.8	0.990460	0.999996	0.997179

Table 1 provides a reference benchmark, offering power estimates for complete cases in the three different ANOVA designs. This benchmark serves as a standard, illustrating optimal power conditions against which the analyses involving missing data can be compared.

In the two-factor design, the complete case power is exceptionally high for medium and large effect sizes, peaking at 0.990460 for a substantial effect size of 0.8. Similarly, the three-factor design shows an even higher maximum power of 0.999996 at the same effect size of 0.8. These results indicate that, with complete data, both designs demonstrate strong statistical power across various effect sizes when a large number of observations are included. For the split-plot design, power estimates are significantly lower for small effect sizes, with a power of 0.099417 at an effect size of 0.2. However, the power increases remarkably with larger effect sizes, reaching 0.495403 for a medium effect size of 0.5 and 0.997179 for a large effect size of 0.8. This suggests that while the split-plot design is less effective at detecting small effect sizes, it performs similarly to the other designs for medium and large effect sizes.

Table 2: Estimations from a two-factor Design with various levels of Missing Data

		Monotone				**Arbitrary**			
		Number of Imputations				***Number of Imputations***			
Number of Missing Observations	Estimate	***20***	***30***	***40***	***100***	***20***	***30***	***40***	***100***
5 (8%)	$\boldsymbol{\theta}$	160.690	160.214	158.538	159.631	185.880	163.730	162.646	161.488
	$\boldsymbol{\sigma_b^2}$	333.10	83.074	86.624	132.102	1160.56	125.745	185.446	97.337
	FMI	0.129	0.133	0.120	0.110	0.066	0.077	0.103	0.076
	RIV	0.146	0.152	0.135	0.123	0.070	0.083	0.114	0.082
	RE	0.994	0.996	0.997	0.999	0.997	0.997	0.997	0.999
10 (16%)	$\boldsymbol{\theta}$	160.187	160.907	164.668	161.155	158.942	157.517	157.583	154.997
	$\boldsymbol{\sigma_b^2}$	293.098	248.273	276.047	252.934	234.134	222.381	149.505	156.014
	FMI	0.207	0.318	0.215	0.312	0.145	0.139	0.127	0.142
	RIV	0.256	0.456	0.272	0.451	0.167	0.160	0.144	0.0165
	RE	0.990	0.990	0.995	0.997	0.993	0.995	0.997	0.999
24 (40%)	$\boldsymbol{\theta}$	182.405	162.775	180.601	171.58	172.790	169.804	182.231	179.319
	$\boldsymbol{\sigma_b^2}$	1128.801	467.668	1481.198	885.193	736.453	1008.78	1378.65	1129.122
	FMI	0.808	0.810	0.816	0.831	0.456	0.438	0.379	0.376
	RIV	0.894	4.041	4.264	4.848	0.803	0.756	0.599	0.597
	RE	0.961	0.974	0.980	0.992	0.978	0.986	0.991	0.996

Table 2 shows estimates from a two-factor design with varying levels of missing data rates, imputation numbers, and missing data patterns (monotone and arbitrary). As the number of imputations increases from 30 to 100, the fraction of missing information (FMI) and the relative increase in variance (RIV) decreases in some of the scenarios, while relative efficiency (RE) increases. The power models show that more imputations improve the results.

The RE levels are slightly higher in the arbitrary missingness than in the monotone. This indicates arbitrary patterns better maintain statistical efficiency relative to monotone patterns for comparable imputation numbers. However, at higher missingness rates like 40%, there is a slight drop in RE, but FMI and RIV remain very high, even with 100 imputations. This gauges the efficiency loss even at high imputation numbers when missing data rates are extreme (40%).

Table 3: Statistical Power Estimates for Two-factor Design

		Monotone				**Arbitrary**			
		Number of Imputations				***Number of Imputations***			
Number of Missing Observations	***Effect Size***	***20***	***30***	***40***	***100***	***20***	***30***	***40***	***100***
5 (8%)	***0.2***	0.06657	0.066536	0.066965	0.06703	0.065335	0.067219	0.066848	0.067476
	0.5	0.157256	0.157007	0.159835	0.16026	0.149095	0.161504	0.159059	0.163202
	0.8	0.328243	0.327628	0.334615	0.335662	0.307909	0.338722	0.332699	0.34289
10 (16%)	***0.2***	0.056856	0.055884	0.056584	0.055895	0.057439	0.057552	0.057655	0.057643
	0.5	0.093673	0.087400	0.091918	0.087473	0.097449	0.098182	0.098849	0.098768
	0.8	0.164432	0.147860	0.159796	0.148051	0.174399	0.176332	0.178092	0.177876
24 (40%)	***0.2***	0.050458	0.050498	0.050430	0.050407	0.051320	0.051380	0.051412	0.051437
	0.5	0.052888	0.053143	0.052711	0.052568	0.058309	0.058684	0.058890	0.059044
	0.8	0.057419	0.058075	0.056965	0.056597	0.071431	0.072403	0.072931	0.073338

Table 3 shows the estimated statistical power for the two-factor design under the different missing data configurations using equation (13). The table provides a power analysis, quantifying the impacts of varying effect sizes, missing data rates, patterns, and imputation numbers. It can be observed that power remains high at a low (8%) missingness but drops substantially as missing rates increase, especially at 40% missingness where power is very low even for a 0.8 (large) effect size. This signifies the dominance of missing data rate on power over factors like effect size.

Statistical power patterns between monotone and arbitrary patterns are generally comparable. At 40% missingness, the level of statistical power was very poor in both patterns. This shows that the higher the missingness, the poorer the level of power.

Table 4: Estimates Obtained for Three-factor Design Imputation of Missing Data

Number of Missing Observations		*Number of Imputations*				*Number of Imputations*			
	Estimate	***20***	***30***	***40***	***100***	***20***	***30***	***40***	***100***
25 (8%)	$\boldsymbol{\theta}$	1.785	1.854	2.020	1.936	4.651	4.643	4.950	4.796
	$\boldsymbol{\sigma_b^2}$	0.784	0.366	0.526	0.452	1.486	1.785	2.149	1.305
	FMI	0.718	0.614	0.587	0.663	0.281	0.329	0.345	0.373
	RIV	2.371	1.530	1.378	1.941	0.380	0.479	0.517	0.590
	RE	0.965	0.980	0.986	0.993	0.986	0.989	0.991	0.996
50 (16%)	$\boldsymbol{\theta}$	3.968	3.816	4.235	3.793	9.045	8.919	8.507	8.734
	$\boldsymbol{\sigma_b^2}$	1.971	1.587	1.707	1.171	1.742	2.725	1.549	2.090
	FMI	0.803	0.747	0.745	0.681	0.319	0.362	0.319	0.416
	RIV	3.769	2.805	2.817	2.109	0.454	0.553	0.461	0.707
	RE	0.961	0.976	0.982	0.993	0.984	0.988	0.992	0.996
128 (40%)	$\boldsymbol{\theta}$	12.637	11.491	11.961	11.275	19.565	17.780	18.608	18.113
	$\boldsymbol{\sigma_b^2}$	5.014	7.458	10.595	5.256	9.973	7.412	5.931	5.828
	FMI	0.883	0.926	0.904	0.906	0.703	0.525	0.487	0.552
	RIV	6.924	11.723	9.043	9.446	2.207	1.066	0.926	1.220
	RE	0.958	0.970	0.978	0.991	0.966	0.983	0.988	0.995

It can be observed from Table 4 that the between-imputation variance, the FMI and RIV increase with more missingness, this establishes the fact that with more missingness there is a higher loss of information. On the other hand, with more imputations the relative efficiency increases, indicating a more robust estimation with higher imputation numbers for the three-factor case as well. But even with a three-factor design, as can be seen, a high 40% missingness still leads to substantially lower relative efficiencies and much higher FMI and RIV values compared with that of a two-factor design.

Table 5: Statistical Power Estimates for Three-factor Design

Number of Missing Observations		**Monotone**				**Arbitrary**			
		Number of Imputations				***Number of Imputations***			
	Effect Size	**20**	**30**	**40**	**100**	**20**	**30**	**40**	**100**
25 (8%)	**0.2**	0.998856	0.999898	0.999868	0.99929	0.99808	0.996832	0.994213	0.993622
	0.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
	0.8	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
50 (16%)	**0.2**	0.480587	0.587793	0.542961	0.677617	0.630267	0.607922	0.654782	0.576607
	0.5	0.997583	0.999763	0.999335	0.999979	0.999919	0.999855	0.999959	0.99969
	0.8	1.00000	1.00000	1.00000	1.000000	1.00000	1.00000	1.00000	1.00000
128 (40%)	**0.2**	0.072235	0.065171	0.068498	0.06887	0.085722	0.111686	0.113263	0.106233
	0.5	0.194564	0.148015	0.169938	0.172392	0.282362	0.439448	0.448294	0.408189
	0.8	0.417532	0.305198	0.359300	0.365229	0.599762	0.824300	0.833414	0.789304

Table 5 shows the estimated statistical power for the three-factor design across the same missing data scenarios, these values were computed using equation (15). At a low 8% missingness, power remains very high even for a modest 0.2 effect size, with higher effect sizes maintaining perfect power uniformly. So low missingness does not appear to influence power drastically. But at 40% missingness, a stark power reduction can be noted as before. For example, power falls to approximately 0.4 even for a large effect size of 0.8 in the monotone case.

This phenomenon mirrors what is seen for the two-factor design, highlighting again the extreme statistical impacts of very high missing data rates regardless of aspects like the effect size or design complexity (additional factors).

Table 6: Estimations from a Split Plot Design with various levels of Missing Data

		Monotone				**Arbitrary**			
		Number of Imputations				***Number of Imputations***			
Number of Missing Observations	Estimate	***20***	***30***	***40***	***100***	***20***	***30***	***40***	***100***
	$\boldsymbol{\theta}$	176.304	183.516	174.975	182.822	187.136	185.600	185.772	183.904
	$\boldsymbol{\sigma_b^2}$	257.215	871.906	257.947	473.882	340.21	333.934	339.196	233.174
6 (8%)	***FMI***	0.106	0.150	0.120	0.155	0.294	0.142	0.239	0.184
	RIV	0.117	0.174	0.136	0.184	0.405	0.164	0.311	0.224
	RE	0.995	0.995	0.997	0.998	0.985	0.995	0.994	0.998
	$\boldsymbol{\theta}$	177.810	183.933	180.670	188.414	199.862	202.698	201.438	195.741
	$\boldsymbol{\sigma_b^2}$	733.879	641.066	439.812	994.767	1604.052	1328.813	922.464	723.960
12 (16%)	***FMI***	0.283	0.415	0.457	0.448	0.333	0.342	0.258	0.298
	RIV	0.384	0.689	0.822	0.803	0.483	0.508	0.343	0.422
	RE	0.986	0.986	0.989	0.996	0.984	0.989	0.994	0.997
	$\boldsymbol{\theta}$	233.804	262.623	251.184	246.089	216.785	206.771	216.340	209.150
	$\boldsymbol{\sigma_b^2}$	2357.313	5196.930	3530.881	2758.647	455.211	238.142	423.172	505.442
29 (40%)	***FMI***	0.487	0.354	0.441	0.477	0.182	0.196	0.187	0.176
	RIV	0.905	0.536	0.772	0.903	0.218	0.241	0.228	0.212
	RE	0.976	0.988	0.989	0.995	0.991	0.994	0.995	0.998

In Table 6, with the monotone pattern of missingness, it can be observed that as the number of imputations increases from 20 to 100, the fraction of missing information (FMI) and the relative increase in variance (RIV) tend to decrease, while the relative efficiency (RE) increases. This trend is consistent across all levels of missing data, showing the beneficial effect of higher imputation numbers on the overall quality and reliability of the imputed datasets. Conversely, the relative efficiency steadily improves, rising from 0.995 with 20 imputations to 0.998 with 100 imputations, signalling a higher degree of statistical efficiency and reliability in the imputed datasets. However, as the rate of missing data escalates to 16% and 40%, the effects on these parameters become more pronounced. At a 40% missing data rate, the FMI and RIV reach substantially higher levels, even with 100 imputations, compared to the 8% and 16% missing data rates. Specifically, the FMI and RIV attain values of 0.477 and 0.903, respectively, with 100 imputations, indicating a substantial loss of information and an increase in variance due to the high proportion of missing data.

In the arbitrary pattern of missingness, similar trends were observed, albeit with some notable differences. At an 8% missing data rate, the FMI, RIV, and RE exhibit comparable patterns to those observed in the monotone case, with the FMI decreasing from 0.294 to 0.184, the RIV declining from 0.405 to 0.224, and the RE improving from 0.985 to 0.998 as the number of imputations increases from 20 to 100. However, at higher missing data rates, the arbitrary pattern appears to have a slight advantage over the monotone pattern in terms of relative efficiency. For instance, at a 40% missing data rate, the RE with 100 imputations is 0.998 for the arbitrary pattern, compared to 0.995 for the monotone pattern, suggesting that the arbitrary pattern may be better equipped to maintain statistical efficiency in the face of extreme missing data scenarios.

Table 7: Statistical Power Estimates for the Split Plot Design

		Monotone				**Arbitrary**			
		Number of Imputations				***Number of Imputations***			
Number of Missing Observations	***Effect Size***	***20***	***30***	***40***	***100***	***20***	***30***	***40***	***100***
6 (8%)	***0.2***	0.068619	0.067004	0.068445	0.066923	0.063909	0.066957	0.065025	0.066268
	0.5	0.170735	0.160087	0.169587	0.159558	0.139714	0.159778	0.147057	0.15524
	0.8	0.361228	0.335236	0.35845	0.33393	0.284226	0.334473	0.30279	0.323245
12 (16%)	***0.2***	0.056725	0.055322	0.055022	0.054865	0.055579	0.055409	0.056115	0.055943
	0.5	0.092831	0.083789	0.081858	0.080856	0.085443	0.084350	0.088873	0.087782
	0.8	0.162209	0.138317	0.133218	0.130571	0.142687	0.139799	0.151804	0.148869
29 (40%)	***0.2***	0.051095	0.051209	0.051095	0.051041	0.05185	0.051904	0.051839	0.051928
	0.5	0.056888	0.057609	0.056893	0.056549	0.061653	0.061994	0.061582	0.062142
	0.8	0.067747	0.069614	0.067759	0.066870	0.080127	0.081014	0.079941	0.081401

Table 7 presents the estimated statistical power for the split-plot design under various missing data configurations using equation (17). Examining the results for an 8% missing data rate, it can be observed that statistical power remains relatively high across all effect sizes, even with a small effect size of 0.2. For instance, under the monotone pattern with 20 imputations, the power is 0.068619 for a small effect, 0.170735 for a medium effect, and 0.361228 for a large effect. These power levels, while lower than the complete case scenario presented in Table 1, still suggest a reasonable probability of detecting true effects, particularly for medium and large effect sizes.

However, as the rate of missing data increases to 16%, a decline is witnessed in statistical power across all effect sizes and missing data patterns. For example, under the monotone pattern with 20 imputations and a large effect size of 0.8, the power drops from 0.361228 at 8% missingness to 0.162209 at 16% missingness. This trend shows the sensitivity of statistical power to higher missing data rates, even with multiple imputation techniques. The impact of missing data on power becomes even more pronounced at the 40% missing data rate, where power levels plummet across all effect sizes and imputation scenarios. For instance, under the monotone pattern with 20 imputations and a large effect size of 0.8, the power further decreases to a mere 0.067747, indicating a substantial reduction in the ability to detect true effects in the presence of such extreme missingness.

In comparing the monotone and arbitrary patterns of missingness, it was observed that the arbitrary pattern generally yields slightly higher power estimates, particularly at higher missing data rates. However, it is important to note that the differences in power between the two patterns are relatively modest, suggesting that the impact of the missing data rate may supersede the influence of the missingness pattern itself.

Contrasting these results with the complete case scenario presented in Table 1, it becomes evident that missing data, even at moderate levels, can significantly erode statistical power in split-plot designs. While the complete case power for a large effect size of 0.8 is 0.997179, the corresponding power under an 8% missing data rate with 20 imputations and the monotone pattern is only 0.361228 – a substantial reduction. This discrepancy highlights the critical importance of accounting for missing data and employing appropriate imputation strategies to mitigate the adverse effects on statistical power. Furthermore, the number of imputations appears to have a limited impact on power beyond a certain threshold, particularly for lower effect sizes and higher missing data rates. For instance, at a 40% missing data rate and a small effect size of 0.2, the power remains relatively unchanged across different imputation numbers, ranging from 0.051095 with 20 imputations to 0.051041 with 100 imputations under the monotone pattern. This suggests that while multiple imputation can partially alleviate the power loss associated with missing data, there may be diminishing returns in terms of power gains beyond a certain number of imputations, especially in scenarios with extreme missingness and small effect sizes.

4. Findings and Conclusion

This research aimed to generalize statistical power estimation from two-sample t-tests to ANOVA with multiple sample means, assess power efficiency under monotone and arbitrary missing data patterns, compare power across

designs with missing data, and determine the optimal number of imputations for desired power levels. The methodology involved incorporating the variance component from ANOVA into Zha and Harel's (2019) power calculation model, enabling explicit power estimation for two-factor, three-factor, and split-plot ANOVA designs with incomplete data handled via multiple imputation. Comparison to complete case analysis showed higher power estimates for complete data scenarios. The analysis revealed that power increased with larger effect sizes, decreased sharply with higher missing rates (especially at 40%), and was slightly lower for arbitrary missingness patterns compared to monotone. The derived power models effectively captured the impact of key factors such as effect size, missing data rate, imputation number, and missingness pattern, validating their utility for assessing power in ANOVA models with missing data. Findings showed that at low 8% missing data rates, power remained high, but dropped significantly at 40% missingness, even for large effect sizes. Arbitrary missing patterns showed slightly higher relative efficiency than monotone patterns, but both declined sharply at 40% missingness. Across all designs, the number of imputations displayed diminishing returns on power beyond certain thresholds, particularly for lower effect sizes and higher missing rates. The study also compared 8% and 16% missing data rates, revealing that while power remained robust at 8%, it declined noticeably at 16%. The split-plot design was especially sensitive to higher missing data rates. The findings underscore the critical need for robust missing data handling and careful power assessments to ensure the validity and reproducibility of ANOVA-based research.

References

Abowd, J. M. (2005). *Multiple Imputation. In the Encyclopedia of Social Measurement*, (2) 753-758. Academic Press

Anderson, E., and Williams, R. (2017). Examining the influence of multiple imputation on statistical power in educational research. *Journal of Educational Measurement*, 40(4) 423-438.

Enders, C. K. (2022). *Applied Missing Data Analysis* (2nd ed.). Guilford Press.

Garson, G.D. (2015). Missing Values Analysis and Data Imputation. Asheboro, NC: Statistical Associates Publishers.

Gong, W., Wan, L., Lu, W., Ma, L., Cheng, F., Cheng, W., Grünewald, S., and Feng, J. (2018). Statistical testing and power analysis for brain-wide association study. Medical Image Analysis, 47, 15-30. https://doi.org/10.1016/j.media.2018.03.014

Graham, J.W., Olchowski, A.E., and Gilreath, T.D. (2007) How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Society for Prevention Research*, 8:206–213

Graham, J. W. (2012), Missing Data: Analysis and Design, Statistics for Social and Behavioural Sciences. Springer Science and Business Media, New York.

Hansen, M. H., Hurwitz, W. N., and Madow, W.G. (1953) Sample survey methods and Survey theory, 1st ed. Wiley, New York

Jones, B.C. (2022). Practical Significance Levels for Cohen's d. *Psychological Reports*, 110(1), 28–34.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*. 64. 402-6. 10.4097/kjae.2013.64.5.402.

Ledolter, J. and Kardon, R.H. (2020) Focus on Data: Statistical Design of Experiments and Sample Size Selection Using Power Analysis, Invest Ophthalmol Vis Sci. 2020;6(8):11 https://doi.org/10.1167/iovs.61.8.11

Maxwell, S.E., Kelley, K., and Rausch, J.R. (2008) Sample size planning for statistical power and accuracy in parameter estimation. *The Annual Review of Psychology*. (59) 537-563

Muthén, L. K., and Muthén, B. O. (2012). Mplus user's guide (7th ed.). Muthén and Muthén

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons Inc.

Rubin, D. B. and Little, R. J. A. (2020). *Statistical Analysis with Missing Data* (3rd ed.). Wiley.

Savla, J. S. (2004) Effects of missing data on statistical power to detect change in family-based preventive intervention research. A dissertation submitted to the graduate faculty of the University of Georgia.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Smith, A. (2021). Using Cohen's d to Interpret Effect Size. *Journal of Statistics*, 57(3), 105-117.

Smith, J., and Johnson, A. (2018). The impact of multiple imputation on statistical power in social science research. *Journal of Applied Social Psychology*, 42(3) 567-582

van Ginkel, J.R. and Kroonenberg, P.M (2015). Analysis of Variance of Multiply Imputed Data Multivariate Behav Res. *National Institute of Health*; 49(1) 78–91. doi:10.1080/00273171.2013.855890.

Von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology,* 37(1), 83-117.

White, I. R., Carlin, J. B., and Hobbs, B. P. (2011). Imputation of missing covariate values for the Cox model. Statistics in Medicine, 30(18), 2122-2137

Zha, R. and Harel, O. (2019) Power calculations in multiply imputed data. Statistical Papers. https://doi.org/10.1007/s00362-019-01098-8