
EXPERIMENTAL DESIGN AND STATISTICAL ANALYSES

William Cruz



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright ©2019 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department with the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Experimental Design and Statistical Analyses
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my family

CONTENTS IN BRIEF

PART I 1 EXPERIMENTAL DESIGN

| | |
|----------------------------|----|
| 1 The Experimental Design | 3 |
| 2 Statistical Inference | 13 |
| 3 Single-Factor Experiment | 21 |

CONTENTS

| | |
|-----------------|------|
| List of Figures | ix |
| List of Tables | xi |
| Acronyms | xiii |
| Glossary | xv |
| Introduction | xvii |

PART I 1 EXPERIMENTAL DESIGN

| | | |
|----------|--------------------------------|-----------|
| 1 | The Experimental Design | 3 |
| 1.1 | Basics of Experimental Design | 3 |
| 1.2 | The Experiment | 4 |
| 1.2.1 | Response variables | 4 |
| 1.2.2 | Independent variables | 5 |
| 1.3 | The Design | 6 |
| 1.4 | The Analysis | 7 |
| 1.5 | Examples | 8 |
| 1.5.1 | Salk Vaccine Experiment | 8 |
| 1.5.2 | Cutting Tool Experiment | 9 |
| 1.6 | Summary | 12 |
| 2 | Statistical Inference | 13 |

| | | |
|----------|--|-----------|
| 2.1 | Introduction | 13 |
| 2.1.1 | What is statistics? | 13 |
| 2.1.2 | Statistical Inference | 13 |
| 2.1.3 | The Mean and Variance of a Random Variable | 14 |
| 2.1.4 | Properties of Expected Values | 14 |
| 2.1.5 | Sample statistics | 15 |
| 2.2 | Estimation | 15 |
| 2.2.1 | Point Estimators and Point Estimates | 15 |
| 2.2.2 | Unbiased Statistics | 15 |
| 2.2.3 | Sum of Squares and Mean Squares | 16 |
| 2.2.4 | Interval Estimators and Interval Estimates | 16 |
| 2.3 | Test of hypotheses | 17 |
| 2.3.1 | What is a test of hypothesis? | 17 |
| 2.3.2 | Critical regions | 17 |
| 2.3.3 | Type I and Type II Errors | 17 |
| 2.3.4 | The Classical Approach to Hypothesis Testing | 18 |
| 2.3.5 | The p value of a Test Hypothesis | 19 |
| 2.4 | The Operating Characteristic Curve | 19 |
| 3 | Single-Factor Experiment | 21 |

LIST OF FIGURES

| | | |
|-----|--|----|
| 1.1 | $B \times C$ interaction for power requirement example | 12 |
|-----|--|----|

LIST OF TABLES

| | | |
|-----|---|----|
| 1.1 | Factorial Experiment Arrangement | 5 |
| 1.2 | Nested Experiment Arrangement | 6 |
| 1.3 | Randomization in a Factorial Experiment | 7 |
| 1.4 | Polio Experiment Results | 9 |
| 1.5 | Power Requirement Data | 10 |
| 1.6 | ANOVA for Power Requirement Data | 11 |

ACRONYMS

| | |
|-----|---------------------------|
| LMM | Linear Mixed Model |
| BIB | Balanced Incomplete Block |

GLOSSARY

mean Average among observations

INTRODUCTION

Experimental Design is at the core of research

PART I

THE EXPERIMENTAL DESIGN

CHAPTER 1

THE EXPERIMENTAL DESIGN

1.1 Basics of Experimental Design

As defined by Leedy, research is a systematic quest for undiscovered truth [2]. In this definition, the quest is systematic because it uses scientific procedures to solving a problem and the truth sought, should be something that is unknown at the time, which implies that the researcher has conducted a thoughtful literature search to ascertain that the answer to the experimental question was not addressed in previous studies.

A *true experiment* may be defined as a study in which certain independent variables are manipulated, their effect on one or more dependent variables is determined, and the levels (i.e. values) of these independent variables are assigned at random to the experimental units in the study. Such experiments allow the researcher to develop a statistical model and estimate its validity. An *experimental unit* must be carefully defined; it may be a part, a person, a class, a month, or some other unit depending of the problem. The experimental units make up a *universe* of all such units, present or future, and a *population* is defined as measurements that might be taken on all experimental units in the universe. There may be several populations defined on the same experimental units [1].

Manipulation and randomization are essential for a true experiment from which one may be able to infer cause and effect. Sometimes it is not always feasible to assign some variables to the experimental units at random; although it may be possible to assign some variables to the experimental units at random, it may be possible to run a *quasi-experiment* in which groups of units are assigned to various levels of the independent variable ran-

domly. Randomizing is a technique used to assure a high probability of group equality on many variables that escape of proper control.

There are other types of research that are not experimental. One of the most common is the *ex-post-facto* research. The values of the variables have been determined by circumstances beyond the control of the experimenter, the variables have already acted, and the research measure only what has occurred (e.g. Studies showing a high incidence of cancer in people who smoke heavily). However, such inferences are dangerous, since the researches cannot claim causality unless they are able to rule out other competing hypotheses. In this category are survey research, historical research and developmental research. All these investigations may use statistical methods to analyze data collected but none is really experimental in nature. There are three phases to bear in mind to make a meaningful study: The planning stage, the design stage and the analysis.

Example: Experimental Designs

Illustrations of true experiments and ex-post-facto experiments often use techniques such as regression and correlation. For example in a regression problem, levels of the independent variable X are set, a level of X is assigned at random to each experimental unit, observations are made on the dependent variable Y , and an equation for predicting Y from X over the range of specified X 's is sought. Such studies are true experiments. A correlation problem may involve a sample of experimental units for which two variables (X and Y) have acted, the resulting (x and y) pairs are plotted to look for a relationship, and a correlation coefficient is calculated to measure the strength of the relationship. Such studies are ex-post-facto studies.

1.2 The Experiment

The experiment includes a statement of the problem to be solved. A careful statement of the problem goes a long way toward its solution. Much care must be taken to spell out the problem in terms that are well understood and point the way in which the research might be conducted.

1.2.1 Response variables

The statement of the problem must include reference to a least one characteristic of an experimental unit on which information is to be obtained. Such characteristics are called *response* or *dependent variables*. The response variable may be *qualitative* or *quantitative* response. The outcome of a true experiment occurs at random, so a response variables is a random variable. Knowledge of such a variable is essential because the shape of its distribution often dictates what statistical test can be used in subsequent data analysis. In addition to a statement of the problem that includes reference to a response variable, one must ask the following questions about the measurement method.

- What instruments are necessary to measure the variable?
- How accurately can the variable be measured?

1.2.2 Independent variables

Many controllable experimental variables, called independent variables or factors, may contribute to the value of the response variable. *Qualitative factors* have levels that vary by category rather than numerical degree. *Quantitative factors*, have levels that are counts of measurements. The experimenter must decide which factors will be held constant, to be manipulated at certain specified levels, and which will be averaged out by a process of randomization.

Factors whose levels are set at specified values are called *fixed effects*, and those whose levels are randomly selected from all possible levels are called *random effects*. When at least two factors are being considered, one must also ask: How are the factor levels to be combined? Each combination of levels from two or more factors is called a *treatment combination*. If at least one value of the response is observed at each treatment combination, the experiment is called a (complete) *factorial experiment*, as in Table 1.1. An experiment in which the levels of one factor are chosen within the levels of another factor is a *nested experiment*, as in Table 1.2.

Example: Factorial Experiment

Consider an experiment in which the factor A is to be considered at three levels, and a second factor B is to be considered at two levels. Two observations are to be made at each treatment combination. If the $3 \times 2 = 6$ treatment combinations can be set in the following way, a factorial experiment will be possible. Notice that both levels of factor B can be used with the three levels of factor A .

Table 1.1: Factorial Experiment Arrangement

| Factor A | |
|----------|-------|
| Factor B | 1 2 3 |
| 1 | — — — |
| | — — — |
| | — — — |
| 2 | — — — |
| | — — — |
| | — — — |

Example: Nested Experiment

Sometimes one chooses three suppliers and randomly selects two batches of material from each for chemical analysis, letting A denote the supplier and B denote the batch. Such an experiment has the nested arrangement, as illustrated in Table 1.2. This involves six sets of data, like the factorial arrangement in Table 1.1, but now the factors are in a decidedly different arrangement. The two levels of factor B associated with level 1 of factor A cannot be used with levels 2 and 3 of factor A .

Table 1.2: Nested Experiment Arrangement

| Supplier A | | | | | |
|------------|---|---------|---|---------|---|
| 1 | | 2 | | 3 | |
| Batch B | | Batch B | | Batch B | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| — | — | — | — | — | — |
| — | — | — | — | — | — |
| — | — | — | — | — | — |

1.3 The Design

The investigator needs an experimental design for obtaining data that provide objective results with a minimum expenditure of time and resources. In order to correctly assess how many observations are needed, one have to considerate three aspects: how large a difference is to be detected, how much variation is present, and what size risks can be tolerated. If such information is missing, one best advised to take as large a sample as possible. In practice, the sample size is often quite arbitrary; however, the availability of tables and statistical software often make it possible to determine the sample size in an objective fashion [1].

Once a decision has been made to control certain variables at specified levels, there are always a number of other variables that cannot be controlled. To average out the effects of these uncontrolled variables, it is necessary to randomize the order of experimentation. Randomization will allow the experimenter to assume that the measurement errors are independent, a common assumption in most statistical analyses. Independent variables may be purposefully introduced into the study because they are of interest, because they represent restrictions on randomization, or, indeed, because they are somewhat extraneous and not of interest to the experimenter. When deciding how such variables should be handled, one should seek to maximize the systematic variation of factors of interest and to minimize error (i.e. unexplained) variation, including so-called errors of measurement, by controlling the systematic variation of the factors that are not of interest. Kerlinger refers to this as the max-min-con principle of experimental design. Are basically three ways to handle independent variables:

- Rigid control, the factors remain fixed throughout the experiment
- Manipulate or set at levels of interest.
- Randomization

When factors remain fixed, any inferences drawn from the experiment are valid only for the fixed conditions. Variables that are manipulated or set at levels of interest are those whose effects on the response variable are to be studied. They may be either qualitative or quantitative, and their levels may be either fixed or random. Every effort should be made to include extreme levels, to provide the opportunity to maximize any effects that may be present.

On a different account, it is also of prime importance that the order of experimentation¹ is randomized in order to average out the effects of variables that cannot be controlled, such as indicated by the numbers in Table 1.3; such averaging does not remove those effects completely, they only increase the variation in the observed data.

Table 1.3: Randomization in a Factorial Experiment

| Factor A | | | |
|----------|----|----|----|
| Factor B | 1 | 2 | 3 |
| 1 | 10 | 18 | 6 |
| | 11 | 5 | 1 |
| | 17 | 16 | 2 |
| 2 | 9 | 3 | 14 |
| | 4 | 7 | 15 |
| | 12 | 8 | 13 |

Once the experimenters have agreed upon the experimental design and the randomization procedure, they can set up a mathematical model to describe the experiment. This model will show the response variable as a function of all factors to be studied and any restrictions imposed on the experiment as a result of the method of randomization. A typical model description for the experiment layout presented in Table 1.3 is as described in Equation 1.1

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{(ij)} \quad (1.1)$$

with Y_{ij} the measured variable, μ a common effect in all observations, β_i the **Factor B** effect (where $i = 1, 2$), and τ_j the **Factor A** effect (where $j = 1, 2, 3$), and $\varepsilon_{(ij)}$ the random error in the experiment.

Finally and since the objective of the research is to shed light on a stated problem, one next expresses the problem in terms of a testable *hypotheses*. A research hypothesis is a formal statement of what the experimenter expects to find in the data. For example, the i treatments will produce different averages of the response variable. Statisticians usually state their hypotheses in null form, since these are the only easily testable statements. For example, the i treatments average will not have significant differences. A null hypothesis usually can be expressed in terms of the mathematical model set up in the last step of the design phase.

1.4 The Analysis

This step starts with data collection as described in connection with the design. If the experiment is complicated, a pilot run may be used to check details and make sure that no critical details have been overlooked. Once collected, the data should be prepared for manual or computer analysis. The preparation of graphical displays, inspection of those displays, and

¹This means that each observation within an experiment is taken in a random order after tossing a dice or flipping a coin

computation of appropriate numerical summaries follows data collection. The decisions about what analyses to perform should be expressed in terms that are meaningful for both, to the experimenter and to those expected to receive the report. An investigative process has built-in self-correction. It is an iterative process that moves us closer to the truth at each stage. Statistics enters such process at two points: (1) the design of the experiment and (2) the analysis and interpretation of the experimental results. When certain hypotheses appear to be tenable, these results should be used as feedback to design a better experiment.

1.5 Examples

Examples 1.5.1 and 1.5.2 from [1] are but two illustrations of how projects are organized into three phases: experiment, design and analysis.

1.5.1 Salk Vaccine Experiment

The mathematical model of this experiment could be written as

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad (1.2)$$

Where Y_{ij} is equal to 0 or 1: 0 if no polio is diagnosed and 1 if polio is found. The subscript i represents the child, a number from 1 to approximately 200000. The subscript j represents the treatment received by the i^{th} child: 1 if treated with Salk vaccine and 2 if treated with a placebo. μ is a constant or general average of the 0's and 1's for the whole population, τ_j is the effect of the j^{th} treatment, and ε_{ij} is a random error associated with the i^{th} child receiving treatment j .

Letting n_j denote the number of children who receive treatment j , the proportion of children contracting polio who received treatment j is

$$P_j = \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j} \quad (1.3)$$

and the proportion contracting polio in the total sample of $n = n_1 + n_2$ children is

$$P = \frac{\sum_{i=1}^{n_1} Y_{i1} + \sum_{i=1}^{n_2} Y_{i2}}{n_1 + n_2} \quad (1.4)$$

Use of P_j indicates that the proportion contracting polio who receive treatment j is a random variable *before* the experiment is conducted. The observed proportion *after* conducting the experiment is denoted p_j . Letting θ_j denote the true proportion in the j^{th} population who contract polio, the statistical hypothesis to be tested is

$$H_0 = \theta_1 = \theta_2 \quad (1.5)$$

with alternative

$$H_0 = \theta_1 < \theta_2 \quad (1.6)$$

The form of the alternative hypothesis indicates that we are interested in showing that the

true proportion contracting polio will be less for treatment 1 (*vaccine*) than for treatment 2 (*placebo*). Approximately 0.09% of all children contract polio, and $(200,000)(0.0009)=180$ is sufficiently large, so the sample proportions have approximately normal distribution. Thus,

$$Z = \frac{P_1 + P_2}{\sqrt{P(1-P)[(1/n_1) + (1/n_2)]}} \quad (1.7)$$

is an appropriate test statistic. When the null hypothesis ($H_0 = \theta_1 = \theta_2$) is true, Z has an approximate normal distribution with mean 0 and variance 1. The results of the experiment are summarized in Table 1.4. When Equations 1.3, 1.4 and 1.7 are used with those results, an observed value of

$$z \approx \frac{(28 \times 10^{-5}) - (71 \times 10^{-5})}{\sqrt{[49 \times 10^{-5}][1 - (49 \times 10^{-5})][(1/200,745) + (1/201,229)]}} \approx -6.14 \quad (1.8)$$

is obtained. The probability of observing a value of Z as extreme as -6.14 when the null hypothesis is true is $P(Z \leq -6.14) \approx 4.126 \times 10^{-10}$. Thus, the likelihood of observing a difference of this magnitude when the Salk vaccine has no effect is less than one in a billion. This gives a very strong indication that the researchers should reject H_0 and conclude that the vaccine is effective. Conversely the R code for obtaining this result can be found below.

Table 1.4: Polio Experiment Results

| Treatment | Sample Size | Number Contracting Polio | Proportion Contracting Polio |
|---------------------|-----------------|--------------------------|---------------------------------|
| <i>Salk Vaccine</i> | $n_1 = 200,745$ | 56 | $p_1 \approx 28 \times 10^{-5}$ |
| <i>Placebo</i> | $n_2 = 201,229$ | 142 | $p_2 \approx 71 \times 10^{-5}$ |
| <i>Totals</i> | $n = 401,974$ | 198 | $p \approx 49 \times 10^{-5}$ |

```
n1 <- 200745 # sample size of those who received Salk vaccine as treatment
n2 <- 201229 # sample size of those who received placebo as treatment
n <- n1 + n2 # Total sample size
num_pol1 <- 56 # Number of children who contracted polio and received Salk vaccine
num_pol2 <- 142 # Number of children who contracted polio and received Placebo
num_pol <- num_pol1+num_pol2 # Total number of children who contracted polio
P1 <- num_pol1/n1 # Proportion of children who were diagnosed with polio and received
Salk vaccine 1-((n1-num_pol1)/n1)
P2 <- num_pol2/n2 # Proportion of children who were diagnosed with polio and received
Placebo 1-((n2-num_pol2)/n2)
P <- num_pol/n
z <- (P1 - P2)/sqrt((P*(1-P))*((1/n1)+(1/n2)))
```

1.5.2 Cutting Tool Experiment

An experiment was to be designed to study the effect of several factors on the power requirements for cutting metal with ceramic tools. The metal was cut on a lathe, and the vertical component of a dynamometer reading. Since this component is proportional to the horsepower requirements in making the cut, it was taken as the measured variable, Y .

The vertical component is measured in millimeters of deflection on a recording instrument. Some of the factors that might affect this deflection are tool types, angle of tool edge bevel, type of cut, depth of cut, feed rate, and spindle speed. It was agreed to hold some factors constant because they represent typical operating conditions. The main objective of the study was to determine the effect of three factors (tool type, angle of edge bevel, and type of cut) on the power requirements. Since only two ceramic tool types were available, this factor was considered at two levels. The angle of tool edge bevel was also set at two levels. 15 and 30, representing the extremes for normal operation. The type of cut was either continuous or interrupted, two levels.

There are therefore two fixed levels for each of three factors or 2^3 conditions, this is called a 2^3 factorial experiment. The levels of two factors (type of tool and type of cut) are qualitative, whereas the angle of edge bevel is a quantitative factor. It was decided to take four observations under each of the eight conditions, making a total of 32 runs in a completely randomized order. Complete randomization ensures the averaging out of any effects that might be correlated with the time of the experiment. The mathematical model for this experiment and design would be

$$Y_{ijkm} = \mu + T_i + B_j + TB_{ij} + C_k + TC_{ik} + BC_{jk} + TBC_{ijk} + E_{m(ijk)} \quad (1.9)$$

with Y_{ijkm} the measured variable, μ a common effect in all observations (the true mean of the population from which all the data came), T_i the tool type effect (where $i = 1, 2$), B_j the angle of bevel (where $j = 1, 2$) with level 1 the 15 angle, C_k the type of cut (where $k = 1, 2$) with level 1 the continuous cut, and $\varepsilon_{m(ijk)}$ the random error in the experiment (where $m = 1, 2, 3, 4$). The other terms stand for interactions between the main factors T , B , and C . The results in millimeter deflection are given in Table 1.5, below it can also be found the R code for generating this data frame and using the package *EMSaov* with the given model and fixed effects produces the ANOVA summary in Table 1.6.

Table 1.5: Power Requirement Data

| Type of Cut, C | Tool Type, T | | | |
|--------------------|------------------|------|------------------|------|
| | 1 | | 2 | |
| | Bevel Angle, B | | Bevel Angle, B | |
| | 15° | 30° | 15° | 30° |
| <i>Continuous</i> | 29.0 | 28.5 | 28.0 | 29.5 |
| | 26.5 | 28.5 | 28.5 | 32.0 |
| | 30.5 | 30.0 | 28.0 | 29.0 |
| | 27.0 | 32.5 | 25.0 | 28.0 |
| <i>Interrupted</i> | 28.0 | 27.0 | 24.5 | 27.5 |
| | 25.0 | 29.0 | 25.0 | 28.0 |
| | 26.5 | 27.5 | 28.0 | 27.0 |
| | 26.5 | 27.5 | 26.0 | 26.0 |

Table 1.6: ANOVA for Power Requirement Data

| Source | Df | SS | MS | F value | P value | Sig |
|----------------|----|------------|------------|---------|---------|-----|
| tool | 1 | 2.8203125 | 2.8203125 | 1.2667 | 0.2715 | |
| bevel | 1 | 20.3203125 | 20.3203125 | 9.1263 | 0.0059 | ** |
| tool:bevel | 1 | 0.1953125 | 0.1953125 | 0.0877 | 0.7696 | |
| cut | 1 | 31.0078125 | 31.0078125 | 13.9263 | 0.001 | ** |
| tool:cut | 1 | 0.0078125 | 0.0078125 | 0.0035 | 0.9533 | |
| bevel:cut | 1 | 0.9453125 | 0.9453125 | 0.4246 | 0.5209 | |
| tool:bevel:cut | 1 | 0.1953125 | 0.1953125 | 0.0877 | 0.7696 | |
| Residuals | 24 | 53.4375000 | 2.2265625 | | | |
| Totals | 31 | 108.9297 | | | | |

```
# Power Requirement Data Frame
library(EMSAov) # Required library
power_cut <- structure(list(cut = structure(c(rep(2L,16),c(rep(1L,16))),
                                           .Label = c("Continuous","Interrupted"),
                                           class = "factor"),
  tool = structure(c(rep(c(rep(2L,8),rep(1L,8)),2)),
                  .Label = c("Tool1","Tool2"),
                  class = "factor"),
  bevel = structure(c(rep(c(rep(2L,4),rep(1L,4)),4)),
                   .Label = c("15","30"),
                   class = "factor"),
  mil_def = as.numeric(c(27.5, 28, 27, 26, 24.5, 25, 28,
                        26, 27, 29, 27.5, 27.5, 28, 25,
                        26.5, 26.5, 29.5, 32, 29, 28, 28,
                        28.5, 28, 25, 28.5, 28.5, 30,
                        32.5, 29, 26.5, 30.5, 27))),
                      .Names = c("cut","tool","bevel","mil_def"),
                      class = "data.frame",
                      row.names = c(NA, -32L))
# Analysis of Variance (ANOVA)
Anova_T <- EMSanova(mil_def~tool*bevel*cut,data=power_cut,
                    type=c("F","F","F"),nested=c(NA,NA,NA))
```

```
# Interaction plot between factor B and C
library(ggplot2,plyr) # Required libraries
desc_power <- ddply(power_cut,.(cut,bevel),summarise,pw_mean = mean(mil_def),
                    pw_ci = 1.96*sd(mil_def)/sqrt(length(mil_def)))
pd <- position_dodge(width = 0.2)
(AnsPlot <- ggplot(data = desc_power,aes(x=bevel,y=pw_mean,group = cut))+
  geom_line(aes(linetype = cut), position = pd) +
  geom_errorbar(aes(ymin= pw_mean - pw_ci, ymax= pw_mean + pw_ci),
               width = .1,position = pd, linetype = 1) +
  geom_point(size = 4, position = pd) +
  geom_point(size = 3, position = pd, color = "white") +
  guides(linetype = guide_legend("Type of Cut")) +
  labs(title = c(),x = c("Bevel (degrees)",
                        y = c("Sample mean")) +
  theme(
    panel.background = element_rect(fill = "white"),
    legend.key = element_rect(fill = "white"),
    axis.line.x = element_line(colour = "black", size = 1),
    axis.line.y = element_line(colour = "black", size = 1)
  )
)
```

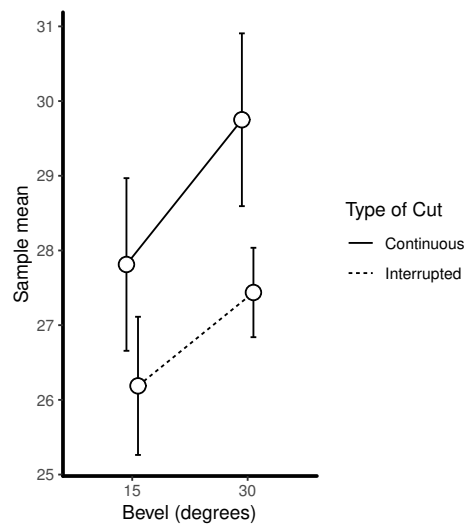


Figure 1.1: $B \times C$ interaction for power requirement example

1.6 Summary

1. Experiment

- (a) Statement of problem
- (b) Choice of response (dependent) variable
- (c) Selection of factors (independent variables) to be varied
- (d) Choice of levels of these factors
 - i. Quantitative or Qualitative
 - ii. Fixed or random

2. Design

- (a) Number of observations to be taken
- (b) Order of experimentation
- (c) Method of randomization to be used
- (d) Mathematical model to describe the experiment
- (e) Hypotheses to be tested

3. Analysis

- (a) Data collection and processing
- (b) Computation of test statistics and preparation of graphics
- (c) Interpretation of results for the experimenter

CHAPTER 2

STATISTICAL INFERENCE

2.1 Introduction

In any experiment, the experimenter is attempting to draw inferences or make a decision about some hypothesis concerning a situation being studied.

2.1.1 What is statistics?

Statistics may be defined as a tool for decision making in the light of uncertainty¹. When the observed results of an experiment could have occurred only 5 times in 100 by chance alone, most experimenters consider that this is a rare event and will state that the results are statistically significant at the 5% significance level. In such cases the hypothesis being tested is usually rejected as untenable. When statistical methods are used in experimentation, one can assess the magnitude of the risks taken in making a particular decision [1].

2.1.2 Statistical Inference

Inferring something about a population from a sample drawn from that population is called *statistical inference*. A *population* is defined as all possible values of a response variable Y obtained from all experimental units in the universe under consideration. Characteristics

¹Cases in which the exact outcome is not completely predictable

of the population associated with this random variable are called *parameters*. Statistical inference is divided in two parts: (1) estimation of a parameter and (2) testing a hypothesis about a parameter.

2.1.3 The Mean and Variance of a Random Variable

The *mean* or *expected value* of the random variable Y is designated as $E[Y] = \mu$, if the probability function defining the random variable is known,

$$E[Y] = \sum_i y_i p(y_i) \quad (2.1)$$

when Y is a discrete random variable with discrete probability function $p(y_i) = P(Y = y_i)$, and

$$E[Y] = \int_{-\infty}^{\infty} y f(y) dy \quad (2.2)$$

when Y is a continuous random variable with probability density function $f(y)$. The long-range average of squared deviations from the mean of a random variable is called the *variance* of that variable. Designating the variance of Y by $Var(Y) = \sigma_Y^2$,

$$Var(Y) = E[(Y - \mu_Y)^2] \quad (2.3)$$

In the discrete case,

$$Var(Y) = \sum_i (y_i - \mu_Y)^2 p(y_i) \quad (2.4)$$

and

$$Var(Y) = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f(y) dy \quad (2.5)$$

In the continuous case. The positive square root of the variance of Y is called the *standard deviation* of Y , denoted $SD(Y) = \sigma_Y$.

2.1.4 Properties of Expected Values

Basic theorems

1. $E[k] = k$ for k any constant
2. $E[kY] = kE[Y]$
3. $E[Y_1 + Y_2] = E[Y_1] + E[Y_2]$
4. $Var(Y) = E[(Y - \mu_Y)^2] = E[Y^2] - \mu_Y^2$

2.1.5 Sample statistics

Most statistical theory is based on the assumption that samples used are random samples, this is a randomly selected collection of units is called a random sample of n experimental units. The collection of n values of the response variable associated with the random sample of experimental units is called a *random sample of size n* from the corresponding population. Quantities computed from those sample values are called *sample statistics*. This include the *sample mean*

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad (2.6)$$

and the sample variance

$$S^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1} \quad (2.7)$$

Before sampling occurs Y_1, \dots, Y_n are random variables, so \bar{Y} and S^2 are random variables. Once sampling has occurred and the values y_1, \dots, y_n are obtained,

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (2.8)$$

And

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} \quad (2.9)$$

Are the observed values of \bar{Y} and S^2 , respectively. The positive square root of S^2 , denoted S , is called the *sample standard deviation* and its observed value is denoted by s .

2.2 Estimation

The objective of statistical estimation is to use a statistic for a sample drawn from a population to estimate a parameter of that population. Estimates of two types are usually needed, *point estimates* and *interval estimates*.

2.2.1 Point Estimators and Point Estimates

A random variable such as \bar{Y} or S^2 , when used to obtain estimates of a population parameter is called a *point estimator* of that parameter, and an observed value of such statistic is called a *point estimate* of the parameter. Since they are random variables, the commonly used point estimators have probability distributions with means and variances. Statisticians usually use point estimators that are unbiased statistics.

2.2.2 Unbiased Statistics

A point estimator W of a population parameter θ is an *unbiased statistic* if $E[W] = \theta$. That is, the expected or average value of W taken over an infinite number of similar samples equals the population parameter being estimated. Since $E[kY] = kE[Y]$ and $E[Y_1 + Y_2] = E[Y_1] + E[Y_2]$ can be used to prove that

$$E[\bar{Y}] = E[Y] = \mu_Y (\text{or simply, } \mu) \quad (2.10)$$

\bar{Y} is an unbiased estimator of μ and \bar{y} is an unbiased estimate. Likewise,

$$E[S^2] = \text{Var}(Y) = \sigma_Y^2 (\text{or simply, } \sigma^2) \quad (2.11)$$

so S^2 is an unbiased estimator of σ^2 and s^2 is an unbiased estimate. On the other hand, the sample standard deviation S is not an unbiased estimator of σ , since $E[S] \neq \sigma$. For a normal population, $E[S] = c_4\sigma$ for c_4 a positive constant based on the sample size.

2.2.3 Sum of Squares and Mean Squares

For Y_1, \dots, Y_n a random sample from the distribution of a random variable Y with mean μ and standard deviation σ , the *sum of squares* is the statistic.

$$SS_Y = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2.12)$$

It is the sum of squares of the deviations of the sample values from the mean of the sample. For y_1, \dots, y_n the observed sample values, the calculated constant

$$\sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.13)$$

is also denoted SS_Y . Since $SS_Y = (n-1)S_Y^2$, $E[(n-1)S_Y^2] = (n-1)E[S_Y^2]$, and $E[S_Y^2] = \sigma_Y^2$, we find that

$$E[SS_Y] = (n-1)\sigma_Y^2 \quad (2.14)$$

where $n-1$ represents the degrees of freedom (*df*) associated with SS_Y . Suppose the sum of squares SS_W is associated with a random variable W . Division of SS_W by its degrees of freedom indicates an averaging of the squared deviations from the mean. The resulting statistic, denoted MS_W is called a *mean square*. In simple cases like SS_Y considered in this section, $df = n-1$, $MS_Y = SS_Y/(n-1) = S_Y^2$, and $E[MS_Y] = \text{Var}(Y)$.

2.2.4 Interval Estimators and Interval Estimates

A $100(1-\alpha)\%$ *confidence interval estimator* of a population parameter is a random interval that is asserted to include the parameter in question with probability $1-\alpha$ for some $0 < \alpha < 1$. The interval with end points determined from observed sample values is called a $100(1-\alpha)\%$ *confidence interval (estimate)* for that parameter. For example, a 95% confidence interval for the mean μ of a normal distribution is given by

$$\bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}} \quad (2.15)$$

Where $P(Z \leq -1.96) = 0.025$, $P(Z \geq 1.96) = 0.025$, n is the sample size, and σ is

the population standard deviation. In general, if Y has a normal distribution with mean μ and standard deviation σ , with \bar{y} the mean of the observed values y_1, \dots, y_n for a random sample from the distribution of Y and $P(Z \leq Z_{1-\alpha/2}) = 1 - \alpha/2$ with $Z_{1-\alpha/2}$ obtained from Statistical Table A, the interval with end points

$$\bar{y} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (2.16)$$

is a $100(1 - \alpha)\%$ confidence interval for μ . If σ unknown and the population is normal, Students t distribution (Statistical Table B) is used, and a $100(1 - \alpha)\%$ confidence interval is given by

$$\bar{y} \pm t_{1-\alpha/2} \frac{s}{\sqrt{n}} \quad (2.17)$$

where t has $n - 1$ degrees of freedom, $P(t \leq t_{1-\alpha/2}) = 1 - \alpha/2$, and s is the observed sample standard deviation. When Y is continuous, non-normal and n is at least 30, either Equation 2.16 or 2.17 can be used to obtain an approximate $100(1 - \alpha)\%$ confidence interval for μ .

2.3 Test of hypotheses

A *statistical hypothesis* is an assertion about the sampled population that is to be tested. It usually consist of assigning a value to one or more parameters of the population as in $H_0 : \mu = 19.5$. The basis for the assignment of this value to μ usually rest on past experience. One also might hypothesize about the variance such that $H_0 : \sigma^2 = 40$. When such hypotheses are to be tested, the other parameters of the population are either assumed or estimated from data taken on a random sample of experimental units from the universe. Assumptions about the population and its parameters are taken as true an can be checked using graphical techniques such as histograms, box plots, and normal quantile plots.

2.3.1 What is a test of hypothesis?

A *test of hypothesis* is a rule by which one determines whether a hypothesis is rejected or not rejected. Such a rule is based on sample statistics, called *test statistics* when they are used to test hypotheses.

2.3.2 Critical regions

The *critical region* of a test statistic consist of all values of the test statistic for which the decision is made to reject the null hypothesis H_0

2.3.3 Type I and Type II Errors

Since hypothesis testing is based on observed sample statistics computed on n observations, the decision is always subject to error. If a null hypothesis that is really true is rejected, a *type I error* has been committed, and its magnitude is denoted as α . If the null hypothesis is not true but is not rejected because the sample provides insufficient evidence

for such rejection, a *type II error* has been made. The magnitude of that error is denoted $\beta(\mu)$

These α and β error probabilities are often referred to as the risks of making incorrect decisions. One of the objectives in hypothesis testing is to design a test for which α is small and β is known to be small for at least one value of the parameter at which the experimenter would like rejection to occur. In most such test procedures α is set at some predetermined level.

2.3.4 The Classical Approach to Hypothesis Testing

In many cases, the critical region of the test statistic, based on the accepted magnitude of α is determined before sampling occurs. Then the value of the test statistic is calculated for the values of the obtained sample, and the null hypothesis is rejected when that value is in the critical region. A series of steps applicable to hypotheses and test statistics of most types can be taken when this classical approach is used. The following example illustrate the procedure.

Example: Classical Approach to Hypothesis Testing

Suppose the average number of miles per gallon obtained by a certain carburetor is to be at least 19.5 mpg. Further suppose that a customer of the supplier of those carburetors has reason to believe that the average is less than 19.5 mpg. The following parallel steps illustrate one way for the customer to test the hypothesis.

Steps in Hypothesis Testing

1. Set up the hypothesis and its alternative - $H_0 : \mu = 19.5 \text{ mpg}$ and $H_1 : \mu < 19.5 \text{ mpg}$
2. Set the significance level, α , of the test and the sample size n . - $\alpha = 0.05, n = 25$
3. Choose a test statistic to test H_0 , noting any assumptions necessary when applying this statistic. - Test statistic: Standardized \bar{Y} , $Z = (\bar{Y} - \mu)/(\sigma/\sqrt{n})$ (Assume σ is known to be 2mpg).
4. Determine the sampling distribution of this test statistic when H_0 is true. - $Z = (\bar{Y} - 19.5)/(2/\sqrt{25}) = 5(\bar{Y} - 19.5)/2$ has a standard normal distribution
5. Set up a critical region on this test statistic where the probability of rejecting H_0 is α when H_0 is true. - Since $P(Z \leq -1.645) = 0.05$, the critical region is $z|z \leq -1.645$. That is, reject H_0 is $z \leq -1.645$.
6. Choose a random sample of n observation, compute the test statistic, and make a decision on H_0 . - If $n = 25$ and $\bar{y} = 18.9 \text{ mpg}$, $z = \frac{5(18.9 - 19.5)}{2} = -1.5$. As $-1.5 > -1.645$, do not reject H_0 .

Without consideration of the magnitude of a type II error, H_0 should not be accepted as true. For that reason, the decision in step 6 is *do not reject H_0*

In the example above a one-sided or one-tailed test was used. This is dictated by the alternative hypothesis, since we wish to reject H_0 only when low values of \bar{Y} are observed.

When Y is normally distributed, the null hypothesis is true, and $\sigma = 2$; \bar{Y} is normally distributed with mean 19.5 and standard deviation $\sigma/\sqrt{n} = 2/\sqrt{25} = 0.4$. The decision rule is expressing the experimenter's beliefs that a sample mean that is at least 1.645 standard deviations below 19.5 [i.e., \bar{y} than $19.5 - 1.645(0.4) = 18.842\text{mpg}$] deviates far enough from the hypothesized mean to be due to something other than chance. Since $P(\bar{Y} \leq 18.842) \approx 0.05$, the significance level is $\alpha = 0.05$. The size of the significance level is a choice of the experimenter and is often set, by tradition, to values such as 0.05 or 0.01.

In general, if Y has a normal distribution with mean μ and standard deviation σ : \bar{y} is the mean of the observed value y_1, \dots, y_n for a random sample from the distribution of Y ; and $P(Z \leq Z_{1-\alpha/2}) = 1 - \alpha/2$ with $Z_{1-\alpha/2}$ obtained from statistical table A, a test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ [or $\mu > \mu_0$ or $\mu < \mu_0$] can be conducted at the $100\alpha\%$ significance level by using

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \quad (2.18)$$

as the test statistic and rejecting H_0 when $z = (\bar{y} - \mu_0)/(\sigma/\sqrt{n})$ and $|z| \geq Z_{1-\alpha/2}$ [or $z \geq Z_{1-\alpha}$ or $z \leq Z_\alpha$]. If σ is unknown and the population is normal, Student's t distribution (Statistical Table B) is used with the test statistic

$$t = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \quad (2.19)$$

where t has $n - 1$ degrees of freedom, $P(t \leq t_{1-\alpha/2}) = 1 - \alpha/2$, S is the sample standard deviation, and H_0 is rejected when $t = (\bar{y} - \mu_0)/(s/\sqrt{n})$ and $|t| \geq t_{1-\alpha/2}$ [or $t \geq t_{1-\alpha}$ or $t \leq t_\alpha$]. When Y is nonnormal and n is at least 30, either Equation 2.18 or 2.19 can be used to test at an approximate $100\alpha\%$ significance level.

2.3.5 The p value of a Test Hypothesis

In the example of the previous section, $Z = (\bar{Y} - 19.5)(2/\sqrt{25}) = (\bar{Y} - 19.5)/0.4$ has a standard normal distribution when Y is normally distributed with $\sigma = 2$ and the null hypothesis is true. After sampling, $\bar{y} = 18.9\text{mpg}$ and $z = -1.5$ were observed. The probability of observing a value of Z at least as extreme as -1.5 when $H_0 : \mu = 19.5\text{mpg}$ is true is $P(Z \leq -1.5) = 0.0668$ from statistical Table A. This probability, called the p value of the test, indicates that the sample evidence allows one to reject the null hypothesis for any $\alpha \geq 0.0668$.

In general, the p value of a test is the probability of obtaining a value of the test statistic that is at least as extreme as the calculated value when the null hypothesis is true. It is the smallest significance level at which the null hypothesis can be rejected. Thus, the agreed-upon value of α is less than the observed p value the null hypothesis is not rejected.

2.4 The Operating Characteristic Curve

$$\alpha \quad (2.20)$$

CHAPTER 3

```
saludo <- 'hola'
```

Bibliography

- [1] Charles Robert Hicks and Kenneth V Turner. *Fundamental concepts in the design of experiments*. Oxford University Press New York, 4 edition, 1999.
- [2] Paul D Leedy and Jeanne Ellis Ormrod. *Practical research*. Pearson Custom, 2005.