# 14 Power

## Basic concept of power.

Study smoking is bad for your heath
Null hypothesis ⇒ no difference between heath smokers non-smokers

→ Results show you cannot reject the null hypothesis → publish the result → company does **NOT** implement a non-smok policy

5 years later → proves that you made a statistical error + should have rejected the null hypothesis → there is a dift in the health of smokers + non-smokers.

You made a Type II error

Type I → rejecting the null hypothesis when its true
Type II → failing to reject the null hypothesis when its false.

Types of error in hypothesis testing.

| | Ho true | Ho false |
|---|---|---|
| Ho Accepted | Correct | Type II error $\beta$ |
| Ho Rejected | Type I error $\alpha$ | Correct |

Probability of making a type I error = $\alpha$
type II error = $\beta$

Power is the probability of correctly rejecting a false null hypothesis eg: $1 - \beta$

Power of test is related to type II errors.
experiment with more power has a greater chance of rejecting a false null hypothesis than an experiment with less power.

We can often estimate $\beta$ the type II error rate or power.
$1 - \beta$ from other information

Power is determined by 4 factors
→ probability of a Type I error
→ true state of affairs guessed by the alternative hypothesis
→ sample size
→ particular test to be employed.

The non-rejection of the null hypothesis when there is actually a difference between groups has enormous implications in most research groups. → this is why power is so important.
∴ the probability of deciding correctly to reject a false null hypothesis ($H_0$) is known as power, the notation is 1 −

Power diagrams are conceptual / hypothetical aids, + <u>not</u> distributions of observed data.

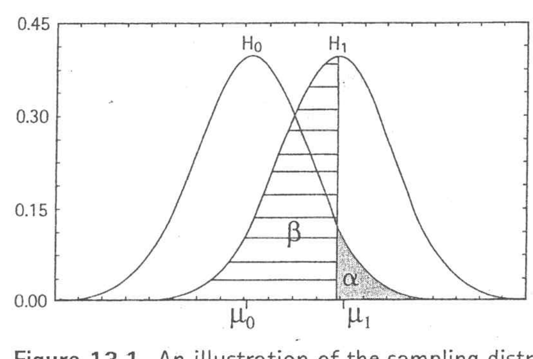


**Figure 13.1** An illustration of the sampling distributions of the mean proposed by the null and alternate hypotheses

- distribution to the left $H_0$ represents the sampling of the mean when the null hypothesis is true $\mu = \mu_0$

→ $\mu$ = true population mean.

- distribution on the right → if the null hypothesis is false an $\mu$ (true population mean) equals $\mu_1$

→ this placement depends on the value of $\mu_1$

Alpha $\alpha$ = probability of a Type 1 error (shaded area of Ho) using a 1-tailed test
(2-tailed shaded area would be $\frac{\alpha}{2}$)

Hi = sampling distribution of the mean Ho if the Ho is false + the true mean is = $\mu_1$

→ even when the null hypothesis is false many of the sample means will fall to the critical left of alpha $\alpha$ causing us to fail to reject the null hypothesis ⇒ making a Type II error.

Possibility of a type II $\beta$ error is the striped area.

→ however if the null hypothesis is false and the sample means fall to the right of the value, we will correctly reject the null hypothesis.

→ probability of correctly rejecting the null hypothesis is what we mean by power and this is the unshaded area of the $H_1$ distribution.

∴ distributions overlap → so the unshaded area includes the area in $H_1$ which the $\alpha$ portion of Ho is shaded (obscured) as $\alpha$ increases (becomes less strict) power increases.

Power is a function of $\alpha$.
→ If we increase $\alpha$ from 0,05 → 0.1 the cutoff point will move to the left, thus decreasing $\beta$ + increasing the probability of a Type I error
But in most real research the researcher is <u>NOT</u> willing to increase Apha $\alpha$ ∴ manipulating $\alpha$ is not a practical strategy for increasing the power of study.

power = 0,6
means that if the null hypotheois is false to the extent that we expect it must be false, there is a probability of 0,6 that the results will cause us to reject the null hypotheois.
→ the greater the power of the experiment, the higher the probability that we will reject the false null hypothesis + the more powerful the experiment is.

Meta - analysis
↳ Statistical method that is used to combine findings of different research projects.
Meta-analysis is used mainly in scientific articles dealing with research in a specific area. (Aron & Aron 1994)

15/ hypothesis tests applied to means : one-way analysis of variance

F - Test - one way variance analysis
↳ ANOVA
ANALYSIS OF VARIANCE

ANOVA
↳ allows us to test the difference between more than two groups of subjects and the influence of more than one independent variable
→ simultaneous testing of differences between more than 2 groups
→ used to test for differences between the means of more than 2 groups + can be used in designs with more than one independent variable.

Rationale for using ANOVA
- typical ANOVA problem - 3 groups of subjects
- null hypothesis for ANOVA is $H_0 = \mu_1 = \mu_2 = \mu_3$
- dealing with a set of possible differences between means
- testing for a set of possible differences instead of testing for a difference between 2 means, we test for an effect
- if we use t-tests for the 3 subjects - it will give an indication where the differences lie, but it is an unsatisfactory approach that leads to an increase in the familywise error rate.

- the familywise error rate is the probability of rejecting at least one null hypothesis when it is true., in a set (family) of comparisons. The overall Type I error rate (family wise error rate) for a set is controlled by ANOVA

$$1-(1-\alpha)^k$$

k = number of significance tests

ANOVA provides an omnibus test for an effect.

Anova → analysis of variance.

→ helps to deal with the problem of familywise error by countering the increase in alpha that occurs when we compare more than 2 groups mean

→ employs a single calculation to test all possible comparisons at once. If the effect is significant, then we know at least one of the 3 comparisons is significant.

- Based on 3 assumptions
  * Scores are distributed normally around the population mean
  * homogeneity of variance → variances of the populations' score are the same.
  * observations are independent.

- dependent variable is always some kind of measurement
  ↳ interval scale for parametric tests
  ↳ ordinal scale for non-parametric tests

Anova is a parametric test

→ based on a normal distribution
→ researcher has no control over the dependent variable :
↳ it is the outcome of manipulation of independent variable.

→ independent variable → manipulated by researcher.
↳ researcher decides on number of levels of this variable + makes sure that members of each group are participated to only one of these levels
→ 2 types of independent variable.
* Quantitive (interval measurement) eg. hours of study per wee
* Qualitive (nominal measurement) eg. teaching method like lecture, video, book

Logic of analysis of variance.
* total variability of a set of scores has 2 components
↳ Variability within groups - measures the normal variability of random errors in the data
* error effect that is, variability resulting from random sampling and Chance events.

↳ Variability between groups - measures the difference between group means.
treatment effect → measure of the effect that the experimental treatment has on participants

F-Ratio is the ratio of variability between groups to variability within groups.

$$F = \frac{\text{variability between groups}}{\text{variability within groups}}$$

$$= \frac{\text{treatment effect}}{\text{error effect}}$$

when the treatment effect exceeds the error effect by a certain critical value, the F test is significant. we conclude that there is an significant difference between group means and we reject the null hypotheosis

NB Always Report ANOVA in a summary table.

<u>example</u>:

| Source | df | SS | MS | F |
|---|---|---|---|---|
| Group | | | | |
| Error | | | | |
| Total. | | | | |

→ SS, MS, F can <u>never</u> have a negative value (check calculations if you get a negative.

→ all SS + MS values are squares

→ Quick Check of interim computation do the following 5 computions

1.* df group + df error = df total

2. SS group + SS error = SS total

3 - SS grop ÷ df grop = MS total

4 - SS error ÷ df error = MS error

5 - MS group ÷ MS error = F

$$SS_{total} = \Sigma x^2 - \frac{(\Sigma X)^2}{N}$$

$$SS_{group} = n \, \Sigma(\bar{x}_j - \bar{X})^2 -$$ $\bar{X} = \Sigma X$ ~~group~~ total $/ N$ (total no of participant)

$$SS_{error} = SS_{total} + SS_{group}$$

table form calculate $X_1$ $X_2$ $X_3$ $X_1^2$ $X_2^2$ $X_3^2$

$\Sigma X_1$ $\Sigma X_2$ $\Sigma X_3$ $\Sigma X_1^2$ $\Sigma X_2^2$ $\Sigma X_3^2$

n = no of participants in each group

N = total no of participants

$df_{total} = N - 1$

$df_{group} = k - 1$ (k = no of groups)

$df_{error} = k(n-1)$  k = no of groups
n = no of participants in each group.

MS error

F / df error

Do the <u>table</u>

| Source | df | SS | MS | F |
|---|---|---|---|---|
| Groups | | | | |
| error | | | | |
| Total | | | | |

15 Non - parametric equivalent of ANOVA.

-> Kruskal wallis one way ANOVA is the nonparametric or distribution free equivalent of the standard one way ANOVA or F-Test

-> in this technique it is assumed that the dependent variable has an underlying continuous distribution of at least ordinal measurement.

16 Chi - square test

-> qualitive (categorical) data

-> helps to analyse frequency or categorical data.

-> categorical data represent no of observations (scores) in each category, whereas measurement data are obtained by measuring objects.

-> Can only work with 2 categorical variables

↳ one classification variable

↳ 2 classification variable.

exhaustive classification - sufficient categories are provided to "exhaust" or encompass all members of the population.

mutualy exclusive -> each member of the population can be assigned to one and only one classification

-> when data can not be classified ~~into~~ as a single variable -> sometimes data fall into 2 categories and the variable (data) is classified as 2 variables.

↳ contingency table.

-> contingency table is also commonly known as a crosstabulation -> crosstab for short.

-> contingency table is typically referred to as an r x c table eg 3x2 table where r = number of rows; c = the number of columns and r x c = the total number of cells in the table.

-> we work with an classification in which the observation is classified in respect of 2 variables simultaneously

→ one classification variable
(chi-square goodness of fit test)
goodness of fit test → name of this type of chi square
↳ used for only 1 variable
↳ compare 1 set of observed data with a theoretical prediction of frequencies
↳ we work with observed O + expected E scores

$x^2$ significance test
→ typically used to test for association between 2 categorical variables.
→ null hypothesis is that NO association exists between sets of categories (left handedness is NOT associated with unusual prowess at tennis)
→ values in the cells of a contingency table must be absolute frequencies and not proportion or persentages
Always retain the original frequencies even if you calculate persentages from your data.
→ expected frequency is a hypothetical frequency calculated either according to
① a theoretical model of a categorical association
OR ② chance expectation (the classic 2 dimensional $x^2$ test of association

$$\text{expected frequency} = \frac{\text{total cell rows} \times \text{total cell colums}}{\text{grand total of all subjects}}$$

example.

| | Heavy tea | Moderate | Abstainer | Total |
|---|---|---|---|---|
| Short | 28 | 32 | 40 | 100 |
| long | 27 | 52 | 21 | 100 |
| Total | 55 | 84 | 61 | 200 |

$100 \times 84 / 200 = 42$

→ since the formula for $x^2$ is the sum of squared numbers $x^2$ will always be a possitive number.

$$x^2 = \sum \frac{(O - E)^2}{E}$$

O = obseved frequency for a cell
E = expected frequency for a cell

form of the $x^2$ significance test is conceptually similar to other significance tests you have encountered → $x^2$ = observed difference ÷ sampling error of expectation

$$df = (r - 1) \times (c - 1)$$

r = no of rows in the table
c = no of columns in the table

$$\text{expected } x^2 = \frac{\text{total no of subjects}}{\text{no of categories}}$$

$x^2$ crit = look up in $x^2$ table.

17 Multivariate statistical techniques

Multivariate statistics are all statistical techniques that ~~analy~~ analyse multiple (more than 2) measurements of each individual or object simultaneously.
extentions of the following

* univariate distributions
  - mean
  - variance + standard deviation
  - t-tests (related + independent)
  - one way ANOVA (one dependent variable)

* bivariate analysis
  - correlation
  - regression (simple regression with one independent variable)

* Purpose of multivariate statistical analysis is to measure, explain and predict relations between variates (weighted combinations of variables.)
  → each type of multivariate statistical technique requires a special measurement scale
    * nominal } nonmetric measurement scales
    * ordinal }
    * interval } metric measurement scales
    * ratio }

Definition of multivariate statistical techniques

Multiple regression → general linear modeling
→ analyses the relation between a single metric dependent variable (criterion) + various metric / non metric independent variables / predictors
→ purpose is to predict changes in the dependent variable on the basis of changes in the independent variables

Multiple discriminant Analysis →
→ used to determine relationship between a single non-metric variable + various metric independent variables.
→ purpose to predict group membership of an entity (person/object) by means of selected independent variables

Multivariate ANOVA
→ extention of one-way ANOVA to accomodate more than 1 independent variable
→ used to study relationship between several nonmetric independent variables (treatments) and 2 or more dependent variables

Canonical correlation
→ purpose to correlate several (set of) metric or nonmetric dependent variables (criteria) with several (set of) metric / non metric independent variables (predicters) simultaneously
→ used to predict multiple dependent variables on the basis of multiple independent variables, as opposed to

multiple regression equations which predict one dependent variable on the basis of multiple independent variables

Structural equation modelling → LISREL

→ determines seperate relationships for each set of dependent correlations.

→ provides most efficient estimation technique for a series of seperate multiple regression equations estimated simultaneously

→ often a dependent variable is used as an independent variable in subsequent dependent correlations.

→ dependent variable is metric

→ independent variable is metric or non metric.

Factor analysis

→ analysis the correlation between a large number of variables to determine common underlying factors / dimensions

→ purpose is to condense data contained in a large number of original variables to smaller more managable number of factors / components

→ metric measurement scale usually but may be nonmetric provided that the latter has been converted into metric scale

Cluster analysis

→ identify significant subgroups of individuals, respondents or objects so that each cluster contains its own distinctive features that set it apart from other clusters

→ 2 steps

-> measuring the correspondence between entities to determine how many groups a sample contains

-> profile anlaysis of the persons or variables to identify their structures

-> measurement scale normally metric but can be non-metric ~~prof~~ provided it's been converted to metric

Multidivisional scaling -> perceptual mapping

-> consists of a series of techniniques to identify certain key dimentions underlying respondent's evaluations of objects

-> often used in marketing to analyse consumer ratings of products, services and organisations

-> measurement scale can be metric or non metric

18 Choice of appropriate analysis

4 Questions to decide what technique to use

* Do I have to describe the data / descriptive statistics or is it a study whose results have to be generalized from a sample to a population (inferential statistics)

* Is the dependent variable measurement data or categorical data?

* How many samples 2,3 or more?

* Are samples related or independent?