Comparing Groups & Covariates: MANOVA, ANOVA, ANCOVA

325-711 Research Methods
2007
Lecturer: Jeromy Anglim

"The complex, thorny, weed-infested and large Andy-eating tarantula inhabited world of statistics suddenly turned into a beautiful meadow filled with tulips and little bleating lambs all jumping for joy at the wonder of life."

Andy Field (2005, p. 304) writing about the illumination that flowed from gaining insight into the general linear model after reading Cohen (1968)

DESCRIPTION:
This seminar will begin by describing what multivariate statistics are and exploring the issues surrounding the use of multivariate statistics. We will the move on to examine how multivariate statistics can be used to test for group differences (e.g managers versus shop floor staff or men versus women). A quick refresher of the basics (e.g t-test) will be given before we move on to spending our time on ANOVA, MANOVA and ANCOVA.
Who knows... You too could have an Andy Field style research methods epiphany. In fact reading Andy Field’s book may be one of the best pathways to experiencing such an epiphany.

Readings
• Tabachnick & Fiddel : Chapter 3 Review of Univariate and Bivariate Statistics
• Web Resources
  • http://www2.chass.ncsu.edu/garson/PA765/manova.htm
  • As always: Google MANOVA or ANCOVA and there’s lots of stuff
  • http://www.bmj.com/statsbk/ - very nice basic introduction to stats

Tabachnick & Fiddel
In a single chapter Tabachnick & Fiddel review univariate and bivariate statistics. This represents the content typically covered in a first and second course in undergraduate statistics in a non-statistics major. It’s a good review and lets you know whether you might need to do further revision in a book like Gravetter & Walneu.
FIELD
Field provides a friendly and accessible introduction to ANCOVA. He shows how to perform ANCOVA in SPSS both using the GLM features and using multiple regression. He does a good job of reminding us that ANCOVA is just another form of multiple regression.

HAIR et al
Typical of Hair et al’s style, the chapter on MANOVA and MANCOVA is long, fairly comprehensive, uses formulas sparingly, and is very readable. Check out http://www.mvstats.com/ if you wish to download the datafiles used in the chapter and see how it was all done in SPSS. I think the best way to learn statistics is combining both the reading of classic text books and then getting down and dirty with some data and your favourite statistics package.

Huberty & Morris (1989)
This article criticises the approach of blindly using a significant MANOVA as a justification and precursor to follow-up univariate ANOVAs. It reminds us of the importance of clarifying the research question before choosing a statistical test and highlights how some research questions are univariate and some are multivariate and that we should choose between multiple ANOVAs and MANOVAs based on the type of research question that we have.

Slide 3

Overview

- Refresher of t-tests, ANOVA & Inferential Statistics
- Reinforcement of core themes of this seminar series
- MANOVA
- ANCOVA
Data analysis can involve a lot of detail to be learnt. I think if all the ideas presented are seen as subservient to this framework, a little bit more structure can be attained. It also orients you towards the right evaluative criteria for assessing your own and other’s analyses.

I.e., is the data good and relevant to answering the research questions? Are the research questions asked interesting? Were the right statistical tools selected and were they interpreted appropriately? Did the data analyses contribute to a better understanding of the research question?

Research question, Research question, Research question, Research question, Research question…. That is the core.

Answering these questions shown below in relation to ANOVA, ANCOVA and MANOVA will be part of what this session is all about.

Access Data - What kinds of Data?
It is often applied to data derived from experiments with grouping data, but may also be applied to correlational designs.

Ask some research questions - What kinds of research questions?

See subsequent Slides, but questions generally relate to whether groups differ in terms of their mean.

Assess your research questions using the right statistical tools - When are they the right ones?; How should they be interpreted?

Answer your research questions - How do they tie back into the underlying research questions?

Need to think about effect sizes in relation to the group differences observed and think about what it means that is being observed. This will be discussed more later.
It all starts with a research question.

So, what kinds of research questions can we answer with t-tests, ANOVA, ANCOVA, MANOVA, and MANCOVA.

The common thread across these techniques is the idea of looking at group differences.

The following topics will pervade this lecture:

**Confidence Intervals:** Often we will be estimating population parameters whether they be group means; the differences between groups; or variance explained. It is generally desirable to have some understanding of the uncertainty of our estimate of the population value.

**Effect Size:** In all the statistical tools presented we want to have some understanding of the practical meaning of the size of the effect. Most of the effects presented today relate to group differences. Thus, we will want to assess the size of the differences between group means observed in our study. In particular we can rely on standardised measures of effect sizes such as r-squared and Cohen's D as well as on the use of scales for the dependent variable that have intrinsic meaning.
**Null Hypothesis Significance Testing:** Any time we are estimating a population parameter (e.g., a mean, a difference between means, a measure of variance explained, etc.), we can potentially conduct a hypothesis test. This involves an understanding of what the null hypothesis is, and how the p-value indicates the probability of obtaining data as divergent or greater than the null hypothesis given the observed data. In essence, we are trying to rule out random sampling as an explanation of the results.

**Power Analysis:** Any time we can run a test of statistical significance, we can also discuss the power of that test. In particular, we will want to know for each of the techniques presented, how to ensure sufficient statistical power. This tends to lead to discussions of what is the necessary sample size.

**Accuracy in Parameter Estimation:** This alternative approach to determining sample size emphasises working out the confidence intervals required to get reasonable estimates of effect sizes.

**Meta Analytic Thinking:** It is important to see research within the context of prior research findings.

**ANOVA is Regression:** One of the aims of the lecture is to show that there are many statistical techniques with different names, which are essentially part of the same underlying model. Thus, while t-tests, ANOVA, ANCOVA and MANOVA, correlation, and multiple regression all have different names, they are all part of the general linear model.

**Remember the research question:**

---

**Slide 7**

---

**Inferential Statistics**

- Theories are based on the population, thus we need to draw inferences about the population from our sample
- Estimate population parameters from sample statistics
  - Means, variances, treatment effects, etc.
  - Greek letters for population parameters
- In any given sample, the statistic will be a little larger or smaller than the population parameter
- If it is an unbiased statistic, the mean of a statistic from many samples will equal the population parameter

For an awesome simulation of this idea, check out:

http://www.ruf.rice.edu/~lane/stat_sim/sampling_dist/index.html

---
Factors
Within the context of ANOVA, a factor is a nonmetric independent variable. The different values that a factor can take are its levels. For example, gender could be a factor and it has two levels (male and female).
Factors differ based on whether they are between subject (different people in each level) or repeated measures (the same people in each level). Repeated measures factors are sometimes called within subjects factors. If we were looking at the effect of working night shift versus day shift on job performance, we could either design the study using a between subjects factor (different works would make up the day shift group and the night shift group) or as a within subjects factor (the same workers are assessed on night shift and day shift).
Designs with both between subjects and within subjects factors are often called mixed designs.
ANOVAs are often defined based on the number of factors. E.g., one-way ANOVA has one factor; two-way ANOVA has two factors.
ANOVAs are also sometimes described by the number of levels of the factors. E.g., a 3 by 2 ANOVA has two factors, one with 3 levels and another with 2 levels.

**(M)AN(C)OVA naming conventions**
- **M**: Multivariate
- **ANOVA**: Analysis of Variance
- **C**: Covariance

i.e., the name of your test depends on whether there is a covariate and how many dependent variables you have.
- **ANOVA**: no covariate; one dependent variable
- **ANCOVA**: covariate/s; one dependent variable
- **MANOVA**: no covariate; two or more dependent variables
- **MANCOVA**: covariate; two or more dependent variables
ANOVA is Regression
And they are all the general linear model

- Regression equation
  \[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_n x_{ni} + \varepsilon_i \]

- Best linear composite
- Minimising error in prediction
- Maximising variance explained

See [http://www.statsoft.com/textbook/stglm.html](http://www.statsoft.com/textbook/stglm.html) for an explanation of the idea of the general linear model.
Andy Field’s book does a really nice job of integrating the idea that t-tests and ANOVA and ANCOVA are just another form of regression.

**Slide 10**

**Dummy Coding**

- To represent a nominal variable with k categories you need k-1 binary variables
- A dummy variable is a variable coded 0 and 1 and is used in a set to represent a nominal variable of 3 or more categories
- The category that receives all zeros is called the reference category
- Regression coefficients show how the dummy coded category differs from the reference category
- Note other forms of coding nominal variables exist including effects coding and various forms of orthogonal coding

<table>
<thead>
<tr>
<th>Marital Status</th>
<th>Dummy1</th>
<th>Dummy2</th>
<th>Dummy3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Married</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Divorced</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Widowed</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Dummy coding becomes particularly important when you are attempting to include nominal variables as predictors in multiple regression. In ANOVA, the dummy coding is handled for you by the software.
ANOVA Results

Descriptive Statistics
Dependent Variable: PMA: 4 Choice RT (ms)

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>404.86</td>
<td>96.39</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>397.18</td>
<td>102.59</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>433.24</td>
<td>81.21</td>
<td>6</td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects
Dependent Variable: PMA: 4 Choice RT (ms)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7611.239</td>
<td>1</td>
<td>7611.239</td>
<td>176</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>1097091.092</td>
<td>52</td>
<td>2115.13</td>
<td>56.53</td>
<td>.000</td>
</tr>
<tr>
<td>Corrected Model</td>
<td>1149803.377</td>
<td>51</td>
<td>22670.54</td>
<td>56.53</td>
<td>.000</td>
</tr>
</tbody>
</table>

Parameter Estimates
Dependent Variable: PMA: 4 Choice RT (ms)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>t</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-36.06</td>
<td>28.39</td>
<td>-1.28</td>
<td>.25</td>
<td>-90.59</td>
<td>22.48</td>
</tr>
<tr>
<td>404.15</td>
<td>-36.06</td>
<td>28.39</td>
<td>-1.28</td>
<td>.25</td>
<td>-90.59</td>
<td>22.48</td>
</tr>
<tr>
<td>404.25</td>
<td>-36.06</td>
<td>28.39</td>
<td>-1.28</td>
<td>.25</td>
<td>-90.59</td>
<td>22.48</td>
</tr>
</tbody>
</table>

a. This parameter is set to zero because it is redundant.

Example

- Research Question
  - Is four choice reaction time (milliseconds to respond to flash of light) related to self-reported problems with vision related as measured using the following question:
    - Q4) Do you ever have any recurring problems with your vision?
      - 1 NO, never
      - 2 Yes, but my vision is corrected by visual aids
      - 3 Yes, my vision is substantially impaired

1. Would we expect the two variables to be related?
2. Are the variables categorical or continuous?
3. Are the variables binary, nominal, ordinal, interval or ratio?
4. What statistical techniques would we typically use to explore the relationship?
5. What if we wanted to use multiple regression using dummy coding of vision problems, how many dummy variables would be required?

- It seems plausible that people with problems with vision would take longer to register the flash of light in the task. They also might be older and reaction time has been shown to be related to age.
- Four choice reaction time: continuous vision problems: categorical
- Four choice reaction time: ratio vision problems: probably ordinal, but equally it could be nominal
- ANOVA would be an obvious choice
- K-1 = 3-1 = 2

Slide 12

• Vision corrected was the quickest vision impaired was the slowest
• Roughly 20 / 80 = .25
• No, p = .680
• Dummy 1: no, never
• Dummy 2: vision corrected
• Because we only need k-1 dummy variables
• Because we can work know when someone is this by the fact that they are zero on the
two dummy variables
• The predicted value on the DV, when someone has a value of zero for predictor variables
• An increase of one on the predictor is associated with this much of an increase on the DV
holding all other predictors constant
• 4 Choice RT = 433 - 28[dummy1] – 36[dummy2]
• Vision impaired
• 4 Choice RT = 433 - 28[0] – 36[0] = 433; i.e., the intercept (otherwise known as the
constant)
• 405 - 433 = -28
• The difference between the dummy coded category and the reference category
• 2
• K-1 = 3-1 = 2
• Because qb4 reflects the effect of the two dummy variables pooled together

Slide 13

| SPSS SYNTAX
| recode qb4 (1=1) (2,3=0) into qb4.dummy1.
| recode qb4 (2=1) (1,3=0) into qb4.dummy2.
| exe.
|

| Q4) Do you ever have any recurring problems with your vision? * qb4.dummy1
| Crosstabulation
| Count
| 0 50 50
| 59 0 59
| 6 0 6
| 65 50 115
| Total
| .00 1.00
| qb4.dummy1
| Total
| 1.

| Q4) Do you ever have any recurring problems with your vision? * qb4.dummy2
| Crosstabulation
| Count
| 50 0 50
| 0 59 59
| 6 0 6
| 56 59 115
| Total
| .00 1.00
| qb4.dummy2
| Total
| 1.

| No
| It treats it as a set of ordered categories and assumes the data is interval
| I know I wasn’t
| It keeps things simple; you don’t have to think about how dummy variables are
constructed. Equally things get more complex when you try to dummy code interaction
effects. You would need to know regression, before you can know that ANOVA and t-tests
is just another form of regression.
| You start to see the underlying commonalities in statistics. It starts to become clear that
regression, ANOVA, and t-tests are all basically the same.
| No... inferring causality is not related to the statistical procedure used. It is a function
of study design. In particular, random assignment of participants to groups and a treatment

Manipulation with good experimental control is what is typically required for an inference of causality.

- We can manually dummy code them
- Recode the variable qb4 into a new variable called qb4.dummy1 making values of 1 equal 1 and values of 2 and 3 equal 0.
- Recode the variable qb4 into a new variable called qb4.dummy2 making values of 2 equal 1 and values of 1 and 3 equal 0.
- Execute this transformation.
- It’s a way of checking that the recoding worked the way that it was meant to
- Substantially impaired; category 3

**Slide 14**

**Regression**

1. What is r-squared? How does it compare to the r-square for the ANOVA?
2. What is the p-value for the overall model? How does it compare to the r-square for qb4?
3. How does the regression equation compare to the parameter estimates in the ANOVA?
4. If it’s all the same, why would you choose to do it using ANOVA?
5. If it’s all the same, why would you choose to do it using regression?

- .007; it’s the same
- P=.680; it’s the same
- It’s exactly the same
- You don’t have to construct the dummy variables yourself;
  You want other bits of output like levene’s test, observed power and estimates of marginal means which are easier to obtain in ANOVA
  You want a pooled measure of the variance explained by the collection of dummy variables
- You don’t mind constructing dummy variables and in fact prefer being explicit about what the model is doing.
  You could explore other ways of splitting the nominal variable into k-1 binary variables
  (there are other coding methods beside dummy coding, including something called effects coding)
  There’s some output that’s easier to obtain in regression such as certain residual plots, semi-partial correlations, etc.
Explaining Variability
• The basis of theory building
• Prediction
• Attributing variability to sources; Partitioning the explained variation

A word about causality
• Association versus Causation
• Designs
  – Experimental
  – Quasi-experimental
  – Correlational
• Causal inferences are based on designs, not particular statistical techniques

Understanding naturally occurring patterns of association and understanding causation are two important elements of theory building.
You may have heard the phrase: “correlation does not mean causation”. But what does this mean?
A more accurate statement would be: Naturally occurring association between two variables does not guaranty that one of the variables causes the other.

Designs
Experimental
This typically involves at two or more groups where an independent variable is manipulated across the groups and participants are randomly allocated into groups.
This is typified by randomised control trials (RCTs) particularly popular in the medical literature. Once random sampling has been ruled out as an explanation (i.e., through a test of statistical significance), some of the differences between the groups following administration of the intervention are assumed to have been caused by some aspect of the intervention.

Quasi-experimental designs
Groups are based on pre-existing groups. Examples include when interventions are applied to particular existing classrooms or worksites. Issues arise as to whether the groups were different to begin with.
Repeated measures designs could also be considered quasi-experimental as a range of issues relating to carry-over effects (e.g., learning, fatigue, etc.) between conditions make inferences of causality more difficult.

**Correlational Designs**

In correlational designs, no variable is manipulated. Thus, both the independent and dependent variable are observed in their naturally occurring patterns. This typically results in both variables being continuous, but this is not necessarily the case. For example, gender is a binary variable and is not experimentally manipulated. In correlational designs inferring causality becomes even more difficult as many alternative explanations exist for the observed correlations.

**Causal inferences are based on designs, not particular statistical techniques**

T-tests and ANOVA are not exclusively used to analyse experimental and quasi-experimental designs. Regression and correlation are not exclusively used to analyse correlational designs. If you wish to infer causality, first ask what kind of design was used to collect the data. Then, based on the design, think about the associated limitations for inferring causality. These limitations are progressively greater as we move from experimental to quasi-experimental to correlational designs.

The core message is that it is the design and not the statistical technique which is relevant to the question of inferring causality.

**Slide 17**

---

**One Sample Z-test**

- **Context**
  - One metric dependent variable
  - One sample compared to known population mean and population standard deviation
- **Null Hypothesis**
  - Sample mean equals population mean
- **Formula**
  \[ z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \]
- **Look up obtained**
  \[ z \] against Table of normal distribution

**Formula:**

The difference between the sample mean and the population mean in terms of the standard error of the mean.
One-Sample t-test

- Context
  - One metric dependent variable
  - One sample compared to known population mean but unknown population standard deviation
- Formula
  \[ t = \frac{\bar{X} - \mu}{s / \sqrt{n}} \]
  \[ \bar{X} = \text{sample mean of } X \]
  \[ \mu = \text{population mean of } X \]
  \[ s = \text{sample standard deviation of } X \]
  \[ n = \text{sample size} \]

**Theory**
When we know the population mean, but not the population standard deviation, we may want to know whether the mean obtained in our sample is significantly different from the population. In this situation we use the one sample t-test.

**Assumptions**
- **Normality**
  It is assumed that the dependent variable is normally distributed in the population. The procedure is relatively robust to modest violations of this assumption
- **Homogeneity of variance**
  The variance in the comparison group and in your sample is assumed to be the same. This can be tested statistically with Levene’s test of Homogeneity of variance.

**Slide 19**

**One Sample T-test Output**

<table>
<thead>
<tr>
<th>Hypothetical Example: Testing whether job satisfaction is different from target</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Sample Statistics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>jobsatisfaction</td>
</tr>
<tr>
<td>One-Sample Test</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>jobsatisfaction</td>
</tr>
</tbody>
</table>

**SPSS**

*Overview of Study*
Imagine you had collected job satisfaction data from the employees in your organisation on a five-point scale. You have data from a benchmarking agency to show that the mean job satisfaction level for your industry is 3.4. You want to know whether your organisation’s level of job satisfaction is significantly different from the benchmark.

**Running**

Analyze >> One Sample T Test

Place the dependent variable into “Test Variables”
Enter the value of the population mean into “Test Value”

**Output**

The first table provides descriptive statistics about the mean of your sample. We can see that in the organisation job satisfaction is lower (2.99) in comparison to the population benchmark (3.4). The next table tells us whether this is statistically significant. The sig (2-tailed) section provides a p value less than .05. Thus, we can conclude that the difference between our sample and the population mean is statistically significant.

**Write-up**

A comparison was made between the job satisfaction results for the organisation (mean=2.99, sd=1) and the benchmark for the industry (mean =3.4). A one sample t-test showed that this was a statistically significant difference, t (99) = -3.98, p<.001.

---

### Independent groups t-test

- **Application**
  - Binary Independent Variable
  - Metric Dependent Variable
- **Assumptions**
  - Independence of observations
  - Homogeneity of variance
  - Normality of the dependent variable

---

**Theory**

**Overview**

The independent-groups t-test is used to test whether two groups differ in terms of their means on a continuous dependent variable. Examples of the kinds of questions that could be answered include: Are there gender differences in intelligence? Does a group that receives training perform better than a group that does not? Is there a difference in job satisfaction between older and younger workers?

**Assumptions**

- **Independence of Observations**
  Independence of observations is the assumption that there is no relationship between one observation and the next. This assumption is usually satisfied.
- **Homogeneity of Variance**

Homogeneity of variance is the assumption that the within group variance is equal across groups in the population. SPSS provides Levene’s test, which test the assumption. If Levene’s test has a significance level less than .05, then the assumption is typically held to be violated. T-tests are relatively robust to violations of the assumption especially if group sizes are relatively equal.

**Normality of the Dependent Variable**
Normality is the assumption that the scores on the dependent variable are normally distributed in each group.

**Slide 21**

The following graph was created using G-power 3. Assuming a two tailed independent groups t-test with alpha set at the typical .05 level and equal group sizes, it shows how statistical power varies as a function of the effect size (Cohen’s D) and total sample size. I would thoroughly recommend you download this free software and have a play around with it.

Download it at: wwwpsychouni-duesseldorfde/abteilungen/aap/gpower3
Or just type “g power 3” into Google

Frequently as part of your ethics application, your thesis confirmation, or as part of a journal article you will be required to do a power analysis. Equally you may be wishing to assess the adequacy of the sample size of an article you are reading.

A graph like that shown above can be very useful. Often we don’t know exactly what the population effect size is in a particular context. We may read various articles and meta-analyses and form an estimate. However, it can be useful to state something like this in your ethics application or thesis showing a graph like that above tailored to your particular context:

“Based on a review of the literature the standardised difference between group means is assumed to be somewhere between a lower medium (cohen’s d = .4) and large effect (cohen’s d=.8). Assuming alpha = .05 and a two tailed test is adopted, a total sample size of 150 will give us reasonable power if the effect size is .4 (about 68% power) and excellent power if it is a large effect (almost 100% power). Thus, the selected sample size of 150 seems to offer a reasonable chance of obtaining a statistically significant result, especially if the...
effect size is towards the large end of the effect size spectrum (See figure... for detailed power curves)"

I think in general it is good to start getting an intuitive sense of how power, effect size and sample size are related and exploring a few of these graphs is an excellent way to do it.

Slide 22

This Slide makes more explicit what statistical power means in relation to null hypothesis significance testing.

This is a power analysis looking at achieved power of a study with 50 participants in each group and assuming a Cohen’s D of .5 (i.e., half a standard deviation between the two groups), two tail t-test and alpha = .05.

Our obtained power was 69.7% (i.e., reasonable, but not great)

The red distribution shows the spread of t-test values we would get in samples if the null hypothesis were true (i.e., no group differences).

The blue line shows the distribution of t-test values we would get if the population effect size was .5.

The green line reflects the critical t-value. When t is larger than critical t we conclude that such a result would be unlikely to occur if the null hypothesis were true and as a consequence, we reject the null hypothesis.

We can see that only some of the time is blue distribution (the non central distribution) above the green line. Thus, some of the time we will do a study and even though the null hypothesis is false, our obtained t-value just by random sampling will not be large enough to rule out the null hypothesis.

Thus, statistical power is the probability of correctly rejecting the null hypothesis, when the null hypothesis is false.
Accuracy in Parameter Estimation Approach
Size of 95% Confidence interval around Cohen’s D
Assumes medium (d=.5) effect size and equal sample sizes for the two groups

E.g., Sample size of 300 and effect size of .5 would have confidence interval Plus or minus .22
e.g., between .28 and .72

The above graph was created in the free open source software package R. The code used to create the graph is available on the digital version of the Slide, if you are interested. The code may appear complex at first but the more you use R and the more online tutorials you work through, the more familiar it all appears. In particular obtaining the basic information about confidence intervals for effect sizes is not too difficult once you know the basics. The help files are quite extensive.

In terms of interpreting the graph, the key point is that if we are trying to assess the population effect size for our study, we want to know what level of confidence we want to have. Thus, instead of choosing a sample size based on power analysis, we choose a sample based on what will give us a desired confidence of the population effect size. We might decide that plus or minus .3 in terms of Cohen’s D would be adequate. Reading off the graph

Hypothetical Example: Gender & Liking Advertising Campaign

SPSS: Independent Group T-test

<table>
<thead>
<tr>
<th>gender</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>liking</td>
<td>male</td>
<td>2.0000</td>
<td>.81650</td>
<td>.25820</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>3.3000</td>
<td>1.15950</td>
<td>.36667</td>
</tr>
</tbody>
</table>

Independent Samples Test

<table>
<thead>
<tr>
<th>liking</th>
<th>F</th>
<th>Sig.</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variances assumed</td>
<td>3.875</td>
<td>.055</td>
<td>2.899</td>
<td>18</td>
<td>.010</td>
<td>-1.33000</td>
<td>.44845</td>
<td>.44845</td>
<td>-2.43717</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.689</td>
<td>.101</td>
<td>2.899</td>
<td>18</td>
<td>.010</td>
<td>-1.33000</td>
<td>.44845</td>
<td>.44845</td>
<td>-2.43717</td>
</tr>
</tbody>
</table>
Overview of Study

Imagine you were in Market research and you wanted to know whether there was any difference between males and females in liking for your new advertising campaign. You showed the advertisement to 10 females and 10 males. Each participant rated their liking for the advertising on a 5-point scale where higher scores indicated greater liking of the advertising.

You then go to: Analyze >> Compare Means >> Independent-Samples T Test
Place gender into the Grouping Variable and press Define Groups to specify what numbers represent the groups in the data file.
Place ‘liking’ in the Test Variable (Dependent Variable)

Output

This first table shows the descriptive statistics (mean, standard deviation and sample size) associated with the two groups. From an initial look at the means it would appear that females liked the advertisement more than males. But is this difference statistically significant?

The table below shows two rows of data. One assumes homogeneity of variance and one does not. The process is to first look at Levene’s test to see whether homogeneity of variance is a reasonable assumption. As the p-value is not less than .05, we can assume homogeneity of variance. We then proceed to analyse “Equal variances assumed” row. The p-value associated with the test is statistically significant at the .05 level.

Write-up

An independent-groups t-test was performed examining difference between males and females on liking for a new advertising campaign. A significant difference was found between males and females for liking of the program, t (18) = 2.90, p = 0.1. Examination of the means showed that females (mean = 3.3, sd = 1.2) like the advertising more than males (mean = 2.0, sd = 0.8).

Repeated Measures T-test

- Design
  - Same as a one-sample t-test of difference scores
  - Difference score: Time 2 – Time 1

- Assumptions
  - Normality of difference scores
  - Independence of observations
    \[
    t = \frac{\bar{D} - 0}{s_D / \sqrt{n}}
    \]
    \[
    \bar{D} = \text{sample mean of difference score (D)}
    \]
    \[
    s_D = \text{sample standard of difference score}
    \]
    \[
    n = \text{sample size}
    \]

Theory

Overview

The repeated measures t-test is used to test whether means differ on two levels of a variable obtained from the same set of individuals. Examples could include looking at knowledge
before and after training, health levels before and after receiving a drug, or satisfaction with two different computer interfaces. Repeated measures designs are generally more powerful than between-subjects designs, because we are able to remove the effect due to individual differences.

**Assumptions**

**Independence of observations**
Assumption is that observations within a treatment are independent. It is not assumed that observations on the same person are independent. By definition people will be related to themselves from time 1 to time 2.

**Normality of difference scores**
The assumption is that in the population the difference scores are normally distributed. A difference score is the difference between a participants score in the first condition minus the second condition. This can be computed using SPSS: Transform >> Compute; then write an expression that reflects one variable minus the other. This can then be plotted as a histogram.

**Overview of Study**
Imagine a scenario where you are a Human Resource manager and have just implemented a program to attempt to increase job satisfaction. You have measured job satisfaction at time 1 prior to your program. Six months later (time2) after implementing your program, you have then measured job satisfaction on the same employees. Job satisfaction was measured on a 7 point scale where higher scores indicate greater job satisfaction.

**Analyze >> Compare Means >> Paired Samples T Test**
Click on time1 then on time2 and copy across to paired variables.

**Output**
The following table shows the means, sample size and standard deviation of job satisfaction at the two time points. Which time point looks like it had higher job satisfaction? It looks like job satisfaction went up between time1 and time2.

This table shows the correlation between scores at time1 and time2. This shows whether the people who were more satisfied at time1 were also more satisfied at time2. This is the stable individual difference factor that is removed in repeated measures design. Thus, the larger
the correlation, the more stable individual differences are, and the more powerful the repeated measures design is.

This table shows whether the difference between the two time points is statistically significant. Based on an alpha of .05, if the significance column is less than .05, we would conclude there was a significant difference.

**Write-up**

A repeated measures t-test was performed to assess whether job satisfaction changed over time following an intervention that aimed to increase job satisfaction. Prior to the intervention, mean job satisfaction was 3.4 (sd = .99) and after intervention mean job satisfaction was 4.4 (sd = 1.59). Based on a paired samples t-test this represented a significant increase in job satisfaction over time, t(14) = -2.82, p = .014.


Data looks at the effect of self reported exercise status on
After performing an omnibus ANOVA assessing the null hypothesis that all groups have equal means, it is typical to want to know which groups differ from which other groups. There are several different ways of doing this. Broadly they can be classified in terms of being planned comparisons or pairwise comparisons and whether they are a priori or post hoc.

**A priori:**
A priori comparisons are planned prior to inspecting the sample means. They are grounded in theory and typically reflect a subset of all possible comparisons.

**Post-hoc:**
Post hoc comparisons are performed after seeing the data and involve either explicitly or implicitly performing all possible comparisons.

**Planned comparisons:**
Trend analysis is used when the independent variable involves a series of ordered levels. It can be used to test for linear, quadratic or higher order relationship between the independent grouping variable and the dependent variable. Simple comparisons involve comparing just two groups. Compound comparisons involve comparing two sets of groups where at least one set is made up of two or more groups. An example would be to compare the average mean of groups 1 and 2 with the average of groups 3, 4 and 5.

Comparisons are orthogonal when they are asking non-correlated research questions. Orthogonality is assessed by checking that the sum of multiplied pairs of weights sum to zero. For more information see Howell. Orthogonality is often considered a desirable property and also links up with topics related to including. There can only ever by k – 1 orthogonal contrasts for an independent variable with K groups (i.e., the same as the group df).

**Post hoc tests:**
Post hoc tests typically involve performing all possible pairs of comparisons. A quick look at SPSS will give you a sense of the many different types. There are different options depending on such things as whether homogeneity of variance is satisfied, how you wish to trade-off...
Factorial ANOVA

• Design
  – Two or more categorical independent variables
    and one continuous dependent variable

• ANOVA Table & F tests
  – Understand each of the elements conceptually
    and procedurally

• Assumptions
  – Independence of observations
  – Normality
  – Homogeneity of variance

Factorial ANOVA

• Main & Interaction Effects
  – Main effect
    • Difference in marginal means of one factor
  – Interaction effect
    • IN WORDS: The effect of one factor depends on the level of the other factor
    • GRAPHICALLY: Graph of means shows non-parallel lines

“Interaction” effects are the same as moderator effects in moderator regression. When we say that there is an interaction effect between Factor A and Factor B, this is the same thing as saying factor A moderates the relationship between Factor B and the dependent variable.
Factorial ANOVA

- Following up overall factorial ANOVA
  - If interaction is significant:
    - Analysis of Simple Effects
      - ANOVA and potentially post-hoc tests for one factor at each level of another factor
      - Decide whether to use overall error term
    - OR Interaction contrasts
      - Determine weights of contrasts for both factors and multiply them for each corresponding cell
  - If interaction is not significant
    - Main effect contrasts
      - \( n \) = sample size on which marginal mean was based
      - DF error and ms error is the same as for the corresponding factor in the ANOVA
    - Main effect Post hoc tests

Error Term in different models

- Fixed Variable
  - All levels of theoretical interest are represented in the treatment variable (the typical scenario)
  - Want to make inferences only about selected levels
  - E.g., treatment vs control
- Random Variable
  - A random selection of possible levels has been selected
  - Want to make inferences only broader population of levels
  - E.g., a subset of schools; train routes; departments, organisations
- Three types of models
  - Fixed effects Model
  - Random effects Model
  - Mixed effects Model
- The rule:
  - If the factor is crossed with a random factor, it takes the interaction MS as its error term

One way of thinking about the rule (e.g., effect of math program [fixed] on subset of Victorian schools [random]) is that if the effect of the fixed factor (e.g., receiving not receiving math intervention) varies across levels of the random factor (e.g., maths program is more effective in some schools than others), the interaction MS would be bigger. Bigger Interaction MS as an error term leads to reduced power which just reflect the uncertainty in knowing the effect of the intervention, because it is varying across schools.
Repeated Measures ANOVA

- Repeated Measures Variables
- ANOVA Table
  - Sums of Squares Partitioning – removes systematic effect of subject
  - Tends to increase power
- Problem of Carry over effects
  - Strategies for dealing with carry over effects
    - Randomisation
    - Counterbalancing
    - Latin Squares Designs
    - Randomised Blocks Design
- Planned comparisons for repeated measures design
- Also see more advanced topics:
  - Multilevel modelling
  - Latent growth curve analysis

For more information on repeated measures ANOVA
See: Howell’s text book
SPSS annotated output example:
http://www.ats.ucla.edu/stat/spss/seminars/Repeated_Measures/default.htm

ANOVA Table
- Sums of Squares Partitioning – removes systematic effect of subject
  - Tends to increase power, because we are removing a systematic source of variance that tends to be quite large.

Problem of Carry over effects
Strategies for dealing with carry over effects
- Randomisation: Order of conditions is random
- Counterbalancing: All possible orders of are represented equally
- Latin Squares Designs: Each condition appears equally in each rank order position equally
- Randomised Blocks Design: Participants are paired based on matched scores on a powerful predictor of the DV

Planned comparisons for repeated measures design
Repeated Measures ANOVA

- Assumptions
  - Compound symmetry
    - Covariance between levels of repeated measures variable is equal in the population
  - Sphericity
    - Actually, the assumption is sphericity of which compound symmetry is a sufficient but not necessary requirement
  - Assessment - Mauchly’s test
  - Consequence of violation
    - Significance testing is biased without a correction

- Different F tests
  - Sphericity assumed
    - Univariate (sphericity assumed)
  - Sphericity violated
    - Corrected Univariate - crit. F_{\epsilon(k-1), \epsilon(n-1)(k-1)}
  - Multivariate

When sphericity is violated, the choice between corrected univariate and multivariate is like that between Pepsi and Coke.
Neither is definitively superior.

Mixed ANOVA

- Design
  - Mix of between and within subjects factors
- ANOVA Table
  - Sums of Squares Partitioning
- Assumptions
  - Between subjects factors:
    - Homogeneity of Covariance
      - Tested using Box’s M test
    - Normality, independence of observations
  - Repeated measures factors and interaction: sphericity

This book tries to show the common themes across multivariate statistics.

So what does multivariate mean?

Literally it means multiple variables. It is sometimes used in contrast to univariate (describing one variable) and bivariate (describing two variables).

Other times it is reserved for situations were we have multiple dependent variables (e.g., as in MANOVA) or where there is a set of variables which are neither independent or dependent (e.g., Principal Components Analysis, Factor Analysis, Cluster Analysis, Multidimensional Scaling). However, techniques where there is only one dependent variable, such as multiple regression or discriminant function analysis are often called multivariate procedures. In the end, perhaps the meaning of the word is not so important.

More important than working out the exact meaning of multivariate is getting some general ideas about what are the common themes of multivariate analysis.

Harlow tried to synthesise this idea in terms of the many multiplicities of multivariate thinking. She tries to explain multivariate statistics in terms of a way of thinking.

Some of these ideas include:

- **Multiple Measures**: Including multiple measures allows for more sophisticated analyses of the construct.
- **Multiple Statistical methods**: There is often not a single statistical tool for answering a particular research problem
MANOVA

• Elements
  – One or more between subjects factors
  – Two or more metric dependent variables

• Questions
  – Do group means differ on some form of composite of the dependent variables?
  – What is the composite that groups differ on? How do they differ?

• Annotated SPSS Output
  – http://www2.chass.ncsu.edu/garson/PA765/manospss.htm

Warning about MANOVAs

• MANOVAs can be confusing to interpret
  – Without knowing the nature of the composite dependent variable and where the groups sit on this composite, MANOVA can be misleading
  – SPSS does not make it easy to find out this information

• Alternatives
  – Construct your own composite of the dependent variables
  – Stick to univariate ANOVAs

If you are only using the MANOVA as a precursor to doing univariate ANOVAs and are not interested in the composite of the dependent variables, then there is little point in doing the MANOVA.

Huberty & Morris (see reading list) provide a good overview of these issues.

Construct your own composite of the dependent variables
An alternative approach is to create your own composite of the relevant dependent variables. Such a composite might be influenced by both theoretical considerations and correlations in the data. If you have a series of variables all looking at satisfaction with a product, such as loyalty, attitude, and overall satisfaction, you could argue that theoretically they are related and should be combined to form an overall measure. If the data shows that the three measures are all strongly correlated this would strengthen the argument for creating a composite. You could then perform a univariate ANOVA on this variable.

Stick to Univariate ANOVAs
If the univariate scales have intrinsic meaning to you and your readership, it may be preferable to stick to univariate ANOVAs. In this case the interpretation of the results are clear and probably more closely correspond to the kinds of questions people are asking. For example, if you had variables called income, years of education and occupational status, you could run a MANOVA on all three variables and this might be interesting. Equally, it might be interesting just to know whether males and females differ in income in and of itself.

**Slide 39**

**MANOVA Procedure**

- Find a set of weights for the dependent variables and the predictor variables that maximises the correlation between the two.
- MANOVA F tests
  - Wilk’s Lambda
  - Pillai’s Trace
  - Hotellings
  - Roy’s Largest Root

**Slide 40**

**MANOVA: working example**

- **Research Question**
  - Are there differences between American, European & Japanese cars in terms Miles per Gallon, Engine Displacement, Horsepower, Vehicle Weight, and Time to Accelerate to 60mph?

Cars.sav is a standard SPSS datafile that comes with any SPSS installation. So you can try to replicate these results if you wish.

Syntax I used in SPSS – I created this from the menus

```plaintext
GLM mpg engine horse weight accel  BY origin 
/METHOD = SSTYPE(3) 
/INTERCEPT = INCLUDE
```
What do we get from the descriptive statistics?
We learn about the basic scale of measurement of the variables.
We learn about the group sizes which indicates something about the confidence we should put in the findings.
We can also start to get a descriptive sense of whether the groups appear to differ in any substantial way.
In particular it is worth calculating Cohen's d in one's head. E.g., if the standard deviation for miles per gallon is around 6.5. The difference between American and Japanese cars in relation to Miles per gallon is about 1.5 standard deviations. This would be considered a very large effect. In this case because the metric is inherently meaningful (i.e., miles per gallon), you can think about what a difference of ten miles per gallon would mean to you or a hypothetical consumer. The main point is that you can start thinking about the practical size of the differences between groups.
Correlations between the variables

<table>
<thead>
<tr>
<th></th>
<th>Miles per Gallon</th>
<th>Engine Displacement (cu. inches)</th>
<th>Horsepower</th>
<th>Vehicle Weight (lbs.)</th>
<th>Time to Accelerate from 0 to 60 mph (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miles per Gallon</td>
<td></td>
<td>- .789**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>398</td>
<td>398</td>
<td>398</td>
<td>398</td>
<td></td>
</tr>
<tr>
<td>Engine Displacement (cu. inches)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>398</td>
<td>406</td>
<td>400</td>
<td>406</td>
<td></td>
</tr>
<tr>
<td>Horsepower</td>
<td></td>
<td>- .771**</td>
<td>1</td>
<td></td>
<td>-.701**</td>
</tr>
<tr>
<td>N</td>
<td>392</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>Vehicle Weight (lbs.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>398</td>
<td>406</td>
<td>400</td>
<td>406</td>
<td></td>
</tr>
<tr>
<td>Time to Accelerate from 0 to 60 mph (sec)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>398</td>
<td>406</td>
<td>400</td>
<td>406</td>
<td></td>
</tr>
</tbody>
</table>

Correlation is significant at the 0.01 level (2-tailed).

It is useful to look at the correlations between the dependent variables in a MANOVA. In this case all the variables are relatively strongly correlated in ways fairly consistent with a basic understanding of car dynamics. More powerful cars with bigger engines and greater weight use more petrol.

Slide 43

MANOVA: Assumptions

- **Univariate Assumptions**
  - Homogeneity of variance
  - Independence of observations across cases
  - Normality

- **Multivariate Assumptions**
  - Homogeneity of Covariance Matrices
  - Multivariate Normality
Assumption tests

Box's Test of Equality of Covariance Matrices:

<table>
<thead>
<tr>
<th>Box's M</th>
<th>584.373</th>
</tr>
</thead>
<tbody>
<tr>
<td>df1</td>
<td>18.967</td>
</tr>
<tr>
<td>df2</td>
<td>30</td>
</tr>
<tr>
<td>Sig.</td>
<td>128122.1</td>
</tr>
<tr>
<td></td>
<td>.000</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Design: Intercept+origin

Levene's Test of Equality of Error Variances:

<table>
<thead>
<tr>
<th>Miles per Gallon</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.666</td>
<td>2</td>
<td>388</td>
<td>.869</td>
<td></td>
</tr>
<tr>
<td>105.928</td>
<td>2</td>
<td>388</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>48.330</td>
<td>2</td>
<td>388</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>43.588</td>
<td>2</td>
<td>388</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>6.661</td>
<td>2</td>
<td>388</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Design: Intercept+origin

Here we see the basic statistical test first of homogeneity of covariance matrices and then the test of homogeneity of variance. When p is less than .05, we often say that the assumption has been violated.

Box’s M tests for homogeneity of covariance matrices.
Levene’s Test tests for homogeneity of variance for each of the dependent variables. When sample sizes are large these assumption tests can be a little sensitive.

Slide 45

Normality
### Slide 46

#### Multivariate test

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Asymptotic Sig.</th>
<th>Cost of</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai’s Trace</td>
<td>0.30</td>
<td>1437.59</td>
<td>0.000</td>
<td>364.00</td>
<td>0.000</td>
<td>.995</td>
<td>71892.980</td>
<td>1.000</td>
</tr>
<tr>
<td>Wilks’ Lambda</td>
<td>0.995</td>
<td>1437.59</td>
<td>0.000</td>
<td>364.00</td>
<td>0.000</td>
<td>.995</td>
<td>71892.980</td>
<td>1.000</td>
</tr>
<tr>
<td>Hotelling’s Trace</td>
<td>0.005</td>
<td>1437.59</td>
<td>0.000</td>
<td>364.00</td>
<td>0.000</td>
<td>.995</td>
<td>71892.980</td>
<td>1.000</td>
</tr>
<tr>
<td>Roy’s Largest Root</td>
<td>0.005</td>
<td>1437.59</td>
<td>0.000</td>
<td>364.00</td>
<td>0.000</td>
<td>.995</td>
<td>71892.980</td>
<td>1.000</td>
</tr>
</tbody>
</table>

- a. Computed using alpha = .05
- b. Exact statistic
- c. The statistic is an upper bound on F that yields a lower bound on the significance level.
- d. Design: Intercept+origin

### Slide 47

#### MANOVA: Follow up

- **Follow-up Strategies**
  - Series of univariate ANOVAs
  - Discriminant Function Analysis
  - Canonical Variate Analysis
  - Roy Bargman’s Step Down Analysis

There are several follow-up strategies once you have assessed the overall MANOVA. The most common is to do a series of univariate ANOVAs. This is the default output displayed in SPSS when you perform a MANOVA. Remember at the point at which we are looking at univariate ANOVAs, we are no longer doing a MANOVA. It should be noted that this is not answering the question of how the groups differ on some composite of the dependent variables. Thus, it is not directly explaining the results of the MANOVA. To explain the results of the MANOVA, discriminant function analysis is often more useful. Roy Bargman’s step down analysis is explained in Tabachnick & Fiddel, but it is relatively difficult to implement and in my opinion not especially enlightening in terms of providing an understanding the composite of the dependent variables.
Here we see that each of the univariate ANOVAs is statistically significant at the .05 level. I followed these overall ANOVAs up with tukey post hoc tests to discern which groups were significantly different from which other groups.

Looking at it holistically, a clear impression emerges consistent perhaps with popular stereotypes that the American made cars are large, powerful, and petrol guzzling cars in comparison to European and Japanese cars.
Slide 50

### Series of Univariate ANOVAs

- **Pros**
  - Shows which of the dependent variables the groups differ on

- **Cons**
  - Does not show what the composite dependent variable is
  - If you were only interested in univariate effects, why do the MANOVA at all

---

**Approach:**

This follow-up strategy is one of the most commonly applied. It is the default approach in SPSS GLM Multivariate. It involves running a series of univariate ANOVAs on all the dependent variables.

**Warning:**

Many people think that because they are using the SPSS GLM Multivariate procedure that all the output is a MANOVA. Actually, only the multivariate tests are the MANOVA. The series of univariate ANOVAs is not MANOVA. The individual ANOVA output saves you the time or running all these separately.

Slide 51

### Discriminant Function Analysis

- **Pro**
  - Shows what the composite of the dependent variables looks like and where the groups stand on this composite

- **Con**
  - Does not work if you have covariates or if you have multiple independent variables

- **Also**
  - Analysis technique in its own right
  - Multiple IVS (usually metric)
  - One nonmetric DV

Discriminant function is often used in market research to develop predictors of market segments in terms of multiple predictor variables. Mathematically, it is basically MANOVA in reverse. Hair et al provides a discussion of the different emphases.
In some ways it is a more appropriate follow-up procedure to MANOVA than a series of univariate tests, because it aims to explain what the composite dependent variate is in the MANOVA.

For more information about the technique:
Read Hair et al or Tabachnick & Fiddel
Or on the web:
http://www2.chass.ncsu.edu/garson/pa765/discrim.htm
Or just Google “Discriminant Function Analysis”

Slide 52

The discriminant function analysis output shows the relationship between the composite of variate based on the DVS and it shows where each of the counties of origin fall on the composite variate.

Looking at the eigenvalues we see that the first function has the storn

SPSS syntax created from the menus
DISCRIMINANT
/GROUPS=origin(1 3)
/VARIABLES=mpg engine horse weight accel
/ANALYSIS ALL
/PRIORS EQUAL
/CLASSIFY=NONMISSING POOLED .

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Function</th>
<th>Max. Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>.992</td>
<td>95.8</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.706</td>
<td>4.2</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functions at Group Centroids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country of Origin</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>1. American</td>
</tr>
<tr>
<td>2. European</td>
</tr>
<tr>
<td>3. Japanese</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standardized Canonical Discriminant Function Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
</tr>
<tr>
<td>Miles per Gallon</td>
</tr>
<tr>
<td>Engine Displacement (cu. inches)</td>
</tr>
<tr>
<td>Horsepower</td>
</tr>
<tr>
<td>Vehicle Weight (lbs.)</td>
</tr>
<tr>
<td>Time to Accelerate from 0 to 60 mph (sec)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engine Displacement (cu. inches) *</td>
</tr>
<tr>
<td>Vehicle Weight (lbs.) *</td>
</tr>
<tr>
<td>Miles per Gallon</td>
</tr>
<tr>
<td>Horsepower</td>
</tr>
<tr>
<td>Time to Accelerate from 0 to 60 mph (sec)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables ordered by absolute size of correlation within function. Largest absolute correlation between each variable and any discriminant function</td>
</tr>
<tr>
<td>Miles per Gallon</td>
</tr>
<tr>
<td>Vehicle Weight (lbs.)</td>
</tr>
<tr>
<td>Horsepower</td>
</tr>
<tr>
<td>Vehicle Weight (lbs.)</td>
</tr>
<tr>
<td>Time to Accelerate from 0 to 60 mph (sec)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functions at Group Centroids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country of Origin</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>1. American</td>
</tr>
<tr>
<td>2. European</td>
</tr>
<tr>
<td>3. Japanese</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unstandardized canonical discriminant functions evaluated at group means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>1. American</td>
</tr>
<tr>
<td>2. European</td>
</tr>
<tr>
<td>3. Japanese</td>
</tr>
</tbody>
</table>

Abbreviated Output from DFA

Slide 52
Canonical Correlation

- **Pro**
  - Shows what the composite of the dependent variables looks like and where the groups stand on this composite
  - Can handle multiple IVs and covariates

- **Con**
  - Output is often a little mysterious at first
  - Have to set up the dummy variables for nominal variables yourself

- **Also**
  - Technique in its own right
    - Multiple IVs related to multiple DVs
  - Most general procedure of the general linear model

See `cancor` in SPSS

MANOVA the old school SPSS syntax will also approximate it

Canonical correlation is the most general procedures of the general linear model. All other tests can be seen as specific instances of canonical correlation.

ANOVA

- **Elements**
  - One or more nominal
  - One or more metric independent variables
  - One metric dependent variable
  - For MANCOVA

- **Questions**
  - Prediction & Model building
    - How well do the nominal and metric variables predict the outcome?
  - Interaction / moderation
    - Are the regression slopes homogeneous?
  - Controlling for extraneous factors
    - What is the difference between the groups in terms of the dependent variable controlling for the covariate?
ANCOVA Working Example

- Do salary's of professors differ between males and females after controlling for number of publications, times since getting PhD and number of citations?

- What types of variables do we have?


Without any covariates we see that the males are on average earning more, but that this is not statistically significant. Nonetheless, Cohen’s d is about .4. We might still be concerned. The p value is not less than .05, but perhaps we’d be comforted if we found out that the differences were smaller after controlling for other job relevant criteria such as tenure and job performance indicators such as citations and number of publications.
Slide 57

**ANCOVA Assumptions**

- Standard One-way ANOVA assumptions
  - Normality
  - Homogeneity of variance
  - Independence of observations
- Plus one new one
  - Homogeneity of regression slopes assumption

Slide 58

**Homogeneity of regression slopes assumption**

- Definition
  - Homogeneity = the same (i.e., in different groups)
  - Regression slopes = relationship between covariate and dependent variable
- Assessment
  - Graphical (very intuitive)
    - Perform scatterplot between covariate and DV with different markers for each group and plot regression line for each group
  - Significance test
    - Test covariate by grouping IV interaction effect
    - Typically specified using a custom model

When we wish to control for the influence of a covariate to get a more powerful test of group differences, we require the homogeneity of regression slopes assumption to be satisfied.

Interestingly, when our purpose is prediction or when we are explicitly interested in the interaction effect, it ceases to be an assumption. Rather, analysis of homogeneity of regression slopes becomes a substantive question of interest in its own right. What we are testing is just another form of interaction effect. This is also what we refer to moderator regression. Does the covariate (e.g., gender) moderate the relationship between a covariate (e.g., publications) and the DV (e.g., salary).
None of the interactions between the sex and the covariates are statistically significant. A more sophisticated test would pool all these interaction tests into one. It may also be the case that given the small sample, we do not have adequate statistical power to test the assumption. A visual examination of the relationship between the covariates and the DV with markers indicating the levels of the grouping variable and separate regression lines for each group can be very helpful in assessing the assumption. Examination of the scatterplot suggests some tendency towards a stronger relationship between publications and salary for males, but given that it is not statistically significant, we can not rule out random sampling as an explanation.

The above tables suggests that it was mainly time in the job and citation that predicted salary in this sample of professors. The parameter estimates show what this model would look like, if formulated as a regression. In particular the parameter estimates are useful for determining the direction of the relationship of a covariate. In this case, every year on the job predicts an addition $857
and every citation predicts an additional 200 dollars. We can see how the procedure has dummy coded gender to indicate that males are predicted to earn $917 more than females after controlling for all other variables, although this was not statistically significant. The bottom table shows the difference between males and females after controlling for covariates. In this case the difference appears smaller than it was before controlling for the covariates. This might suggest that some of the observed differences in salary may be related to job relevant factors. Once again, the small sample size precludes making definitive findings on these matters.

Slide 61

Assumptions vs Diagnostic Threats

• Assumptions
  – Mathematical expectations about distributions and sampling that underlie significance testing
  – Formal and mathematical
• Diagnostic Threats
  – Concerned with making valid inferences to the population
  – More Pragmatic
  – Closely related to assumption testing
  – Incorporate prior expectations