# Quantitative Data Analysis: A Companion for Accounting and Information Systems Research

## Teaching Materials

*Created by Willem Mertens, Amedeo Pugliese & Jan Recker*

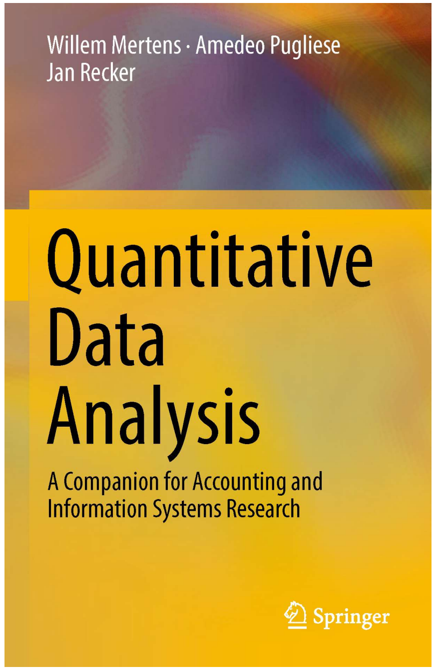
  

# Copyright Notice

# What these materials are about

Offering a guide through the essential steps required in quantitative data analysis

1. Introduction
2. **Comparing Differences Across Groups**
3. **Assessing (Innocuous) Relationships**
4. Models with Latent Concepts and Multiple Relationships: Structural Equation Modeling
5. Nested Data and Multilevel Models: Hierarchical Linear Modeling
6. Analyzing Longitudinal and Panel Data
7. Causality: Endogeneity Biases and Possible Remedies
8. How to Start Analyzing, Test Assumptions and Deal with that Pesky p-Value
9. Keeping Track and Staying Sane



Part 2 and 3:
# Regression and Analysis of Variance Models

# Agenda

1. Regression Models
   - When do we need them?
   - Assumptions
   - Conduct
   - Types
     - Linear
     - Hierarchical
     - Logistic Regression
   - Reporting

2. Analysis of Variance
   - When do we need them?
   - Assumptions
   - Conduct
   - Types
     - ANOVA
     - MANOVA
     - ANCOVA
     - MANCOVA
   - Reporting

# When do we need which test?

| 1 Dependent Variable | 1 Independent Variable | Test |
|---|---|---|
| Binary | Metric | Logistic regression |
| | Non-metric | Chi-square test |
| Non-metric | Metric | Logistic regression |
| | Binary | Mann-Whitney test |
| Metric | Binary | t-test |
| | Metric | Regression analysis |
| | Nominal | Analysis of variance |

# When do we need which test?

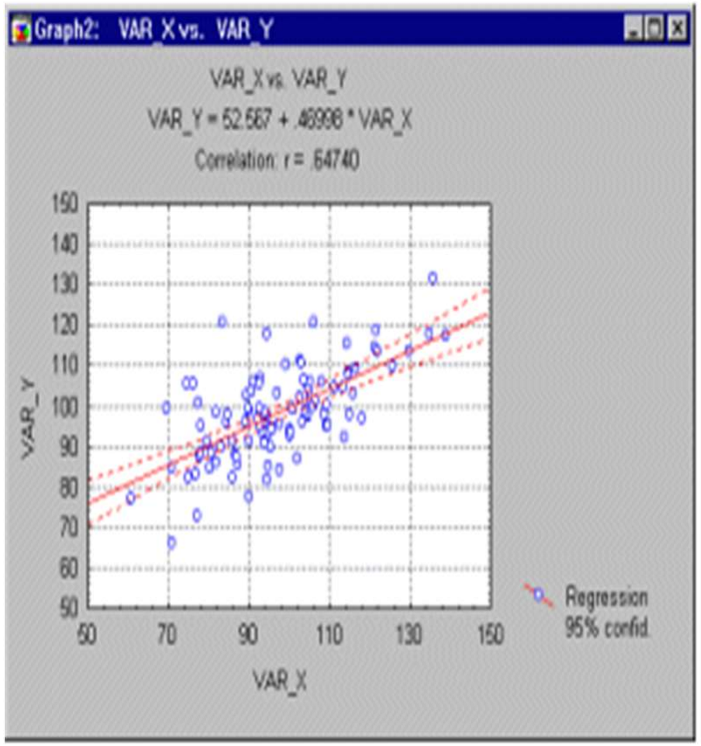
| 1 Dependent Variable | 2 or more Independent Variables | Test |
|---|---|---|
| Non-metric | Metric | Logistic regression |
| | Non-metric | Loglinear analysis |
| Metric | Metric | Multiple regression |
| | Non-metric | Analysis of variance |

# When do we need which test?

| 2 or more Dependent Variables | 2 or more Independent Variables | Test |
|---|---|---|
| Non-metric | Metric | Multivariate multiple regression with dummy variables |
| | Non-metric | Multivariate analysis of variance with dummy variables |
| Metric | Metric | Multivariate multiple regression |
| | Non-metric | Multivariate analysis of variance |

# Regression models

- The purpose of regression models is learn more about the relationship between **several independent** or <u>predictor</u> variables and a **dependent** or <u>criterion</u> variable.
- The computational problem that needs to be solved in regression analysis is to **fit a straight line to a number of points**.
- $Y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + e$



# Types of Regression Models

- Linear regression
  - 1 dependent variable: continuous/scale
  - One or more independent variables: continuous/scale
- Hierarchical regression
  - 1 dependent variable: continuous/scale
  - Multiple blocks of independent variables: continuous/scale
- Logistic regression
  - 1 dependent variable: binary
  - One or more independent variables: continuous/scale

# Linear Regressions: Assumptions

1. **Linearity and additivity** of the relationship between dependent and independent variables:
   1. The expected value of dependent variable is a straight-line function of each independent variable, holding the others fixed.
   2. The slope of that line does not depend on the values of the other variables.
   3. The effects of different independent variables on the expected value of the dependent variable are additive.

2. **Statistical independence** of the errors

3. **Homoscedasticity** (constant variance) of the errors
   Remember: Variance is equal in different (sub-)samples
   1. versus the predictions
   2. versus any independent variable
   3. versus time (in time series data)

4. **Normality** of the error distribution.

# Linear Regressions: Testing Assumptions

1. **Linearity and additivity** of the relationship between dependent and independent variables:
   Check for systematic patterns in a plot of **How to diagnose**: nonlinearity is usually most evident in a plot of **observed versus predicted values** or a plot of **residuals versus predicted values**.

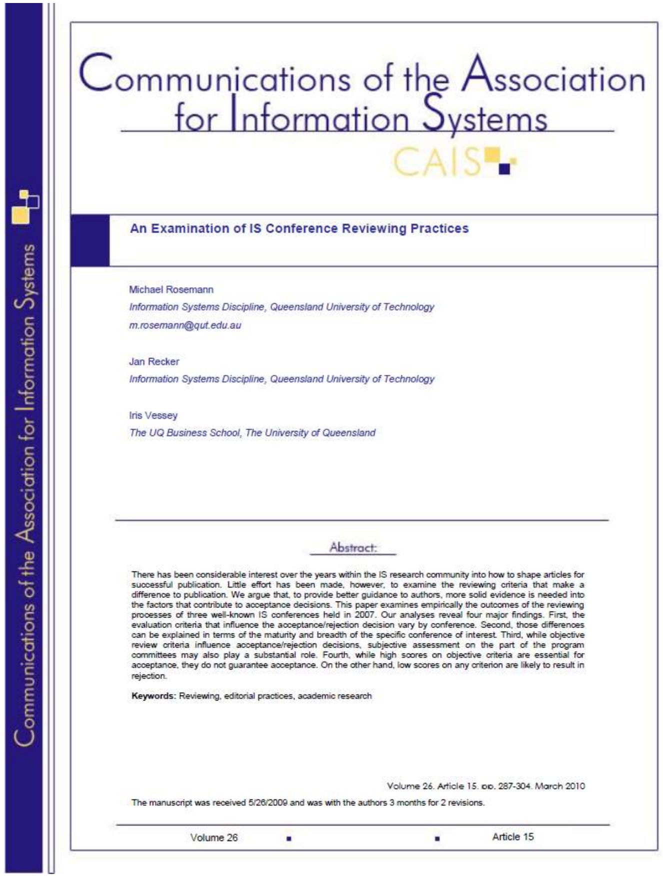
2. **Statistical independence** of the errors:
   Check plots of residuals versus independent variables: residuals should be randomly and symmetrically distributed around zero under all conditions, and in particular **there should be no correlation between consecutive errors no matter how the rows are sorted.**

3. **Homoscedasticity** (constant variance) of the errors:
   Look at a plot of **residuals versus predicted values.** Be alert for evidence of residuals that grow larger either as a function of time or as a function of the predicted value

4. **Normality** of the error distribution:
   Check the plot of residuals for **normal probability**.

# Examples: Regression models

- Analysis of conference reviewing data:
- Which review criteria predict paper acceptance?
- Uses both linear and logistic regression
- Read up at [http://eprints.qut.edu.au/31606/](http://eprints.qut.edu.au/31606/)


Communications of the Association for Information Systems

Communications of the Association for Information Systems

CAIS

An Examination of IS Conference Reviewing Practices

Michael Rosemann
*Information Systems Discipline, Queensland University of Technology*
*m.rosemann@qut.edu.au*

Jan Recker
*Information Systems Discipline, Queensland University of Technology*

Iris Vessey
*The UQ Business School, The University of Queensland*

Abstract:

There has been considerable interest over the years within the IS research community into how to shape articles for successful publication. Little effort has been made, however, to examine the reviewing criteria that make a difference to publication. We argue that, to provide better guidance to authors, more solid evidence is needed into the factors that contribute to acceptance decisions. This paper examines empirically the outcomes of the reviewing processes of three well-known IS conferences held in 2007. Our analyses reveal four major findings. First, the evaluation criteria that influence the acceptance/rejection decision vary by conference. Second, those differences can be explained in terms of the maturity and breadth of the specific conference of interest. Third, while objective review criteria influence acceptance/rejection decisions, subjective assessment on the part of the program committees may also play a substantial role. Fourth, while high scores on objective criteria are essential for acceptance, they do not guarantee acceptance. On the other hand, low scores on any criterion are likely to result in rejection.

**Keywords:** Reviewing, editorial practices, academic research

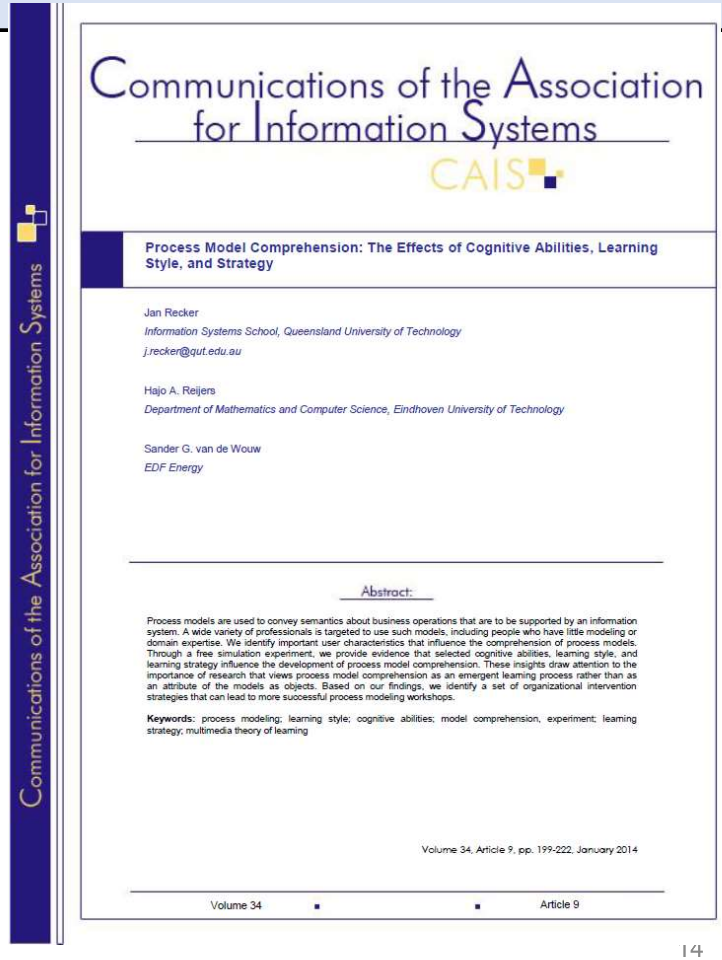
Volume 26, Article 15, pp. 287-304, March 2010

The manuscript was received 5/26/2009 and was with the authors 3 months for 2 revisions.

Volume 26 ▪ ▪ Article 15

# Examples: Regression models

- Analysis of
  process model comprehension:
- Which categories of factors are important to being able to understand processes?
- Uses hierarchical regression
- Read up at [http://eprints.qut.edu.au/66531/](http://eprints.qut.edu.au/66531/)




Communications of the Association for Information Systems

CAIS


**Process Model Comprehension: The Effects of Cognitive Abilities, Learning Style, and Strategy**


Jan Recker

*Information Systems School, Queensland University of Technology*

*j.recker@qut.edu.au*

Hajo A. Reijers

*Department of Mathematics and Computer Science, Eindhoven University of Technology*

Sander G. van de Wouw

*EDF Energy*



Abstract:

Process models are used to convey semantics about business operations that are to be supported by an information system. A wide variety of professionals is targeted to use such models, including people who have little modeling or domain expertise. We identify important user characteristics that influence the comprehension of process models. Through a free simulation experiment, we provide evidence that selected cognitive abilities, learning style, and learning strategy influence the development of process model comprehension. These insights draw attention to the importance of research that views process model comprehension as an emergent learning process rather than as an attribute of the models as objects. Based on our findings, we identify a set of organizational intervention strategies that can lead to more successful process modeling workshops.



**Keywords:** process modeling; learning style; cognitive abilities; model comprehension, experiment; learning strategy; multimedia theory of learning

Volume 34, Article 9, pp. 199-222, January 2014

Volume 34 ▪ ▪ Article 9

# Output of SPSS Regression Analyses

Overall Model Fit

**Model Summary**

| Model[b] | R[c] | R Square[d] | Adjusted R Square[e] | Std. Error of the Estimate[f] |
|---|---|---|---|---|
| 1 | .699[a] | .489 | .479 | 7.14817 |

a. Predictors: (Constant), reading score, female, social studies score, math score

b. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

c. **R** - R is the square root of R-Squared and is the correlation between the observed and predicted values of dependent variable.

d. **R-Square** - This is the proportion of variance in the dependent variable (**science**) which can be explained by the independent variables (**math, female**, **socst** and **read**). This is an overall measure of the strength of association and does not reflect the extent to which any particular independent variable is associated with the dependent variable.

e. **Adjusted R-square** - This is an adjustment of the R-squared that penalizes the addition of extraneous predictors to the model. Adjusted R-squared is computed using the formula 1 - ((1 - Rsq)((N - 1) /( N - k - 1)) where k is the number of predictors.

f. **Std. Error of the Estimate** - This is also referred to as the root mean squared error. It is the standard deviation of the error term and the square root of the Mean Square for the Residuals in the ANOVA table (see below).

# Output of SPSS Regression Analyses

Anova Table

**ANOVA[b]**

| Model[c] | | Sum of Squares[e] | df[f] | Mean Square[g] | F[h] | Sig.[h] |
|---|---|---|---|---|---|---|
| 1 | Regression[d] | 9543.721 | 4 | 2385.930 | 46.695 | .000[a] |
| | Residual[d] | 9963.779 | 195 | 51.096 | | |
| | Total[d] | 19507.500 | 199 | | | |

a. Predictors: (Constant), reading score, female, social studies score, math score

b. Dependent Variable: science score

c. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

d. **Regression, Residual, Total** - Looking at the breakdown of variance in the outcome variable, these are the categories we will examine: Regression, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Error).

e. **Sum of Squares** - These are the Sum of Squares associated with the three sources of variance, Total, Model and Residual. The Total variance is partitioned into the variance which can be explained by the independent variables (Regression) and the variance which is not explained by the independent variables (Residual).

f. **df** - These are the degrees of freedom associated with the sources of variance. The total variance has N-1 degrees of freedom. The Regression degrees of freedom corresponds to the number of coefficients estimated minus 1. Including the intercept, there are 5 coefficients, so the model has 5-1=4 degrees of freedom. The Error degrees of freedom is the DF total minus the DF model, 199 - 4 =195.

g. **Mean Square** - These are the Mean Squares, the Sum of Squares divided by their respective DF.

h. **F** and **Sig.** - This is the F-statistic the p-value associated with it. The F-statistic is the Mean Square (Regression) divided by the Mean Square (Residual): 2385.93/51.096 = 46.695. The p-value is compared to some alpha level in testing the null hypothesis that all of the model coefficients are 0.

# Output of SPSS Regression Analyses

Anova Table

**ANOVA**[b]

| Model[c] | | Sum of Squares[e] | df[f] | Mean Square[g] | F[h] | Sig.[h] |
|---|---|---|---|---|---|---|
| 1 | Regression[d] | 9543.721 | 4 | 2385.930 | 46.695 | .000[a] |
| | Residual[d] | 9963.779 | 195 | 51.096 | | |
| | Total[d] | 19507.500 | 199 | | | |

a. Predictors: (Constant), reading score, female, social studies score, math score

b. Dependent Variable: science score

c. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

d. **Regression, Residual, Total** - Looking at the breakdown of variance in the outcome variable, these are the categories we will examine: Regression, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Error).

e. **Sum of Squares** - These are the Sum of Squares associated with the three sources of variance, Total, Model and Residual. The Total variance is partitioned into the variance which can be explained by the independent variables (Regression) and the variance which is not explained by the independent variables (Residual).

f. **df** - These are the degrees of freedom associated with the sources of variance. The total variance has N-1 degrees of freedom. The Regression degrees of freedom corresponds to the number of coefficients estimated minus 1. Including the intercept, there are 5 coefficients, so the model has 5-1=4 degrees of freedom. The Error degrees of freedom is the DF total minus the DF model, 199 - 4 =195.

g. **Mean Square** - These are the Mean Squares, the Sum of Squares divided by their respective DF.

h. **F** and **Sig.** - This is the F-statistic the p-value associated with it. The F-statistic is the Mean Square (Regression) divided by the Mean Square (Residual): 2385.93/51.096 = 46.695. The p-value is compared to some alpha level in testing the null hypothesis that all of the model coefficients are 0.

# Output of SPSS Regression Analyses

Parameter Estimates

**Coefficients**[a]

| Model[b] | | Unstandardized Coefficients B[d] | Std. Error[e] | Standardized Coefficients Beta[f] | t[g] | Sig.[g] | 95% Confidence Interval for B Lower Bound[h] | Upper Bound[h] |
|---|---|---|---|---|---|---|---|---|
| 1 | (Constant)[c] | 12.325 | 3.194 | | 3.859 | .000 | 6.027 | 18.624 |
| | math score[c] | .389 | .074 | .368 | 5.252 | .000 | .243 | .535 |
| | female[c] | -2.010 | 1.023 | -.101 | -1.965 | .051 | -4.027 | .007 |
| | social studies score[c] | .050 | .062 | .054 | .801 | .424 | -.073 | .173 |
| | reading score[c] | .335 | .073 | .347 | 4.607 | .000 | .192 | .479 |

a. Dependent Variable: science score

b. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

c. This column shows the predictor variables (**constant, math, female**, **socst**, **read**). The first variable (**constant**) represents the constant, also referred to in textbooks as the Y intercept, the height of the regression line when it crosses the Y axis. In other words, this is the predicted value of **science** when all other variables are 0.

d. **B** - These are the values for the regression equation for predicting the dependent variable from the independent variable. The regression equation is presented in many different ways, for example:

**Ypredicted = b0 + b1*x1 + b2*x2 + b3*x3 + b4*x4**

The column of estimates provides the values for b0, b1, b2, b3 and b4 for this equation.

**math** - The coefficient for **math** is .389. So for every unit increase in **math**, a 0.39 unit increase in **science** is predicted, holding all other variables constant.
**female** - For every unit increase in **female**, we expect a -2.010 unit decrease in the **science** score, holding all other variables constant. Because **female** is coded 0/1 (0=male, 1=female), the interpretation is easy: for females, the predicted science score would be 2 points lower than for males.
**socst** - The coefficient for **socst** is .050. So for every unit increase in **socst**, we expect an approximately .05 point increase in the science score, holding all other variables constant.
**read** - The coefficient for **read** is .335. So for every unit increase in **read**, we expect a .34 point increase in the science score.

e. **Std. Error** - These are the standard errors associated with the coefficients.

f. **Beta** - These are the standardized coefficients. These are the coefficients that you would obtain if you standardized all of the variables in the regression, including the dependent and all of the independent variables, and ran the regression. By standardizing the variables before running the regression, you have put all of the variables on the same scale, and you can compare the magnitude of the coefficients to see which one has more of an effect. You will also notice that the larger betas are associated with the larger t-values and lower p-values.

g. **t** and **Sig**. - These are the t-statistics and their associated 2-tailed p-values used in testing whether a given coefficient is significantly different from zero. Using an alpha of 0.05:
The coefficient for **math** (0.389) is significantly different from 0 because its p-value is 0.000, which is smaller than 0.05.
The coefficient for **female** (-2.010) is not significantly different from 0 because its p-value is 0.051, which is larger than 0.05.
The coefficient for **socst** (0.0498443) is not statistically significantly different from 0 because its p-value is definitely larger than 0.05.
The coefficient for **read** (0.3352998) is statistically significant because its p-value of 0.000 is less than .05.
The intercept is significantly different from 0 at the 0.05 alpha level.

h. **95% Confidence Limit for B Lower Bound** and **Upper Bound** - These are the 95% confidence intervals for the coefficients. The confidence intervals are related to the p-values such that the coefficient will not be statistically significant if the confidence interval includes 0. These confidence intervals can help you to put the estimate from the coefficient into perspective by seeing how much the value could vary.

# Reporting Regression Analyses

1. Describe Descriptive Statistics (means, st. dev.) of all variables
2. Report on testing of assumptions – especially if assumptions are violated and what was done about it.
3. Report on model fit statistics (F, df1, df2, $R^2$).
4. Report parameter estimates – for constant and IV
   1. Standardized Beta
   2. T-value and significance
   3. (Confidence intervals)

# Hierarchical Regression Analyses

- Subtype of linear regression models where
  - multiple independent factors exist that
  - can be grouped meaningfully into different categories
- And where an interest exists to compare how predictive the model is given different categories
  - That is, how much better the explanatory power becomes if a particular group of factors is added or deleted.

Example: http://eprints.qut.edu.au/66531/

## Hypothesis Testing

We ran two tests to examine our hypotheses.

First, to examine the data collected on H1–H2, H4, and H5, we estimated two hierarchical regression analyses, one for each process model. These analyses were carried out with SPSS Version 19.0.

One assumption behind the use of regression analysis is that the variables are normally distributed. Our data screening confirmed that these criteria were met for the measures for abstraction ability and selection ability, the dependent variables comp-D1 and comp-D2, as well as for the control variables prior domain knowledge and prior method knowledge. The principal components analysis for the factors' deep learning motive (DM), surface learning motive (SM), deep learning strategy (LS), and surface learning strategy (SS), as well as the control variable self-efficacy (SE), allowed us to extract average total factor scores that also satisfied these assumptions.

We ran the two three-step hierarchical regression analyses as follows. In step one, we entered prior domain knowledge (PDK-1 and PDK-2), prior method knowledge (PMK), and self-efficacy (SE) as control variables. This was done because they correspond to broad, stable traits whose impacts are well-established in the model understanding literature. In step two, we entered our scores for the two types of cognitive abilities considered, as dynamic traits of relevance to the model-based task at hand. In step three, we added the scores for learning motive and learning strategy as further dynamic traits. This hierarchical analysis allowed us to test whether each of the dynamic traits considered (cognitive abilities, learning process) added significantly to the model. We completed these steps for both domain understanding scores for model 1 and model 2.

Table 4 provides descriptive statistics from the analyses. Tables 5 and 6 provide the details of the two hierarchical regression analyses showing the standardized beta coefficients and significance levels.

**Table 4: Hierarchical Regression Analyses: Descriptive Statistics**

| Variable | Mean | St. Deviation |
|---|---|---|
| Model comprehension model 1 (comp-D1) | 2.92 | 1.15 |
| Model comprehension model 2 (comp-D2) | 2.13 | 1.03 |
| Prior domain knowledge model 1 (PDK-1) | 2.48 | 1.38 |
| Prior domain knowledge model 2 (PDK-2) | 2.61 | 1.37 |
| Prior method knowledge score (PMK) | 1.74 | 2.41 |
| Self-efficacy score (SE) | 3.11 | 0.63 |
| Abstraction Ability score (AA) | 10.94 | 4.21 |
| Selection Ability score (SA) | 5.35 | 3.53 |
| Deep Learning Motive (DM) | 3.31 | 0.76 |
| Surface Learning Motive (SM) | 2.76 | 0.85 |
| Deep Learning Strategy score (LS) | 3.74 | 0.57 |
| Surface Learning Strategy score (SS) | 2.58 | 0.72 |

**Table 5: Hierarchical Regression Analysis (Dependent Variable: Comp-D1)**

| Term | 1: Controls | 2: Cognitive Abilities | 3: Learning Process | Collinearity Statistics | |
|---|---|---|---|---|---|
| | St. Beta | St. Beta | St. Beta | Tolerance | VIF |
| PDK-1 | 0.02 | 0.13 | 0.03 | 0.66 | 1.52 |
| PMK | 0.10 | 0.03 | -0.07 | 0.69 | 1.45 |
| SE | 0.10 | 0.07 | 0.08 | 0.80 | 1.25 |
| AA | | -0.21 | **-0.25*** | 0.55 | 1.83 |
| SA | | **0.46**** | **0.56***** | 0.47 | 2.12 |
| DM | | | 0.01 | 0.66 | 1.52 |
| SM | | | **-0.34**** | 0.80 | 1.25 |
| DS | | | 0.18 | 0.60 | 1.67 |
| SS | | | **0.29*** | 0.74 | 1.34 |
| F | 0.62 | **2.40*** | **3.04**** | | |
| F change | 0.62 | **4.97**** | **3.45**** | | |
| R2 change | 0.02 | **0.11*** | **0.13*** | | |
| R2 | 0.02 | 0.13 | 0.26 | | |
| * p < 0.05; ** p < 0.01; ***p < 0.001. | | | | | |

**Table 6: Hierarchical Regression Analysis (Dependent Variable: Comp-D2)**

| Term | 1: Controls | 2: Cognitive Abilities | 3: Learning Process | Collinearity Statistics | |
|---|---|---|---|---|---|
| | St. Beta | St. Beta | St. Beta | Tolerance | VIF |
| PDK-2 | 0.05 | 0.14 | 0.01 | 0.63 | 1.59 |
| PMK | 0.11 | 0.04 | -0.06 | 0.66 | 1.52 |
| SE | 0.10 | 0.08 | 0.10 | 0.82 | 1.22 |
| AA | | **-0.28*** | **-0.32*** | 0.54 | 1.85 |
| SA | | **0.51***** | **0.61***** | 0.48 | 2.08 |
| DM | | | -0.01 | 0.66 | 1.52 |
| SM | | | **-0.28*** | 0.79 | 1.27 |
| DS | | | 0.16 | 0.59 | 1.68 |
| SS | | | **0.32**** | 0.72 | 1.38 |
| F | 0.70 | **3.04*** | **3.16**** | | |
| F change | 0.70 | **6.41**** | **2.94*** | | |
| R2 change | 0.03 | **0.13**** | **0.11*** | | |
| R2 | 0.03 | 0.16 | 0.27 | | |
| * p < 0.05; ** p < 0.01; ***p < 0.001. | | | | | |

We first examine collinearity statistics. Multi-collinearity is present when tolerance is close to 0 (< 0.01; see Tabachnick and Fidell, 2001) or the VIF is high (> 10), in which case the beta and p coefficients may be unstable. The VIF and tolerance measures shown in Table 5 and Table 6 suggest that multi-collinearity is not an issue in our data.

Perusal of the data in Tables 5 and 6 leads to the following observations.

First, we note that, after controlling for prior domain knowledge (PDK), prior method knowledge (PMK), and self-efficacy (SE) as stable traits, cognitive abilities (AA and SA), and learning approach (DM, SM, DS, and SS) as dynamic traits significantly aid the explanation of domain understanding in both cases considered. Adding these factors step-by-step increased the R2 of the regression models to 0.26 (for comp-D1) and 0.27 (for comp-D2), with the changes in R2 being significant in each step (F change = 4.97 and 3.45, both $p < 0.01$ for model 1; and F change = 6.41, $p < 0.01$ and 2.94, $p < 0.05$ for model 2).

# Logistic Regression Analysis

- Type of regression models where
  - The dependent variable is binary
  - [or ordinal: ordered logistic regression (e.g. 3 categories: low, medium, high)]
- Checks whether we can predict in which category we will land based on the values of the IV.
- Essentially compares a model with predictors (BLOCK 1) against a model without predictors (BLOCK 0):
  - is a prediction with our variables better than random chance?

Example: http://eprints.qut.edu.au/31606/

# Logistic Regression Analysis: Output

**Block 0: Beginning Block**

**Classification Table[a,b]**

| | | | Predicted[e] | | |
|---|---|---|---|---|---|
| | | | honcomp | | Percentage Correct |
| | Observed[d] | | .00 | 1.00 | |
| Step 0[c] | honcomp | .00 | 147 | 0 | 100.0 |
| | | 1.00 | 53 | 0 | .0 |
| | Overall Percentage[f] | | | | 73.5 |

a. Constant is included in the model.

b. The cut value is .500

**Variables in the Equation**

| | | B[g] | S.E.[h] | Wald[i] | df[j] | Sig.[i] | Exp(B)[k] |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | -1.020 | .160 | 40.540 | 1 | .000 | .361 |

**Variables not in the Equation**

| | | | Score[l] | df[m] | Sig.[l] |
|---|---|---|---|---|---|
| Step 0 | Variables | read | 47.906 | 1 | .000 |
| | | science | 34.862 | 1 | .000 |
| | | ses | 14.783 | 2 | .001 |
| | | ses(1) | .302 | 1 | .582 |
| | | ses(2) | 8.666 | 1 | .003 |
| | Overall Statistics[n] | | 58.644 | 4 | .000 |

c. **Step 0** - SPSS allows you to have different steps in your logistic regression model. The difference between the steps is the predictors that are included. This is similar to blocking variables into groups and then entering them into the equation one group at a time. By default, SPSS logistic regression is run in two steps. The first step, called Step 0, includes no predictors and just the intercept. Often, this model is not interesting to researchers.

d. **Observed** - This indicates the number of 0's and 1's that are observed in the dependent variable.

e. **Predicted** - In this null model, SPSS has predicted that all cases are 0 on the dependent variable.

f. **Overall Percentage** - This gives the percent of cases for which the dependent variables was correctly predicted given the model. In this part of the output, this is the null model. 73.5 = 147/200.

g. **B** - This is the coefficient for the constant (also called the "intercept") in the null model.

h. **S.E.** - This is the standard error around the coefficient for the constant.

i. **Wald and Sig**. - This is the Wald chi-square test that tests the null hypothesis that the constant equals 0. This hypothesis is rejected because the p-value (listed in the column called "Sig.") is smaller than the critical p-value of .05 (or .01). Hence, we conclude that the constant is not 0. Usually, this finding is not of interest to researchers.

j. **df** - This is the degrees of freedom for the Wald chi-square test. There is only one degree of freedom because there is only one predictor in the model, namely the constant.

k. **Exp(B)** - This is the exponentiation of the B coefficient, which is an odds ratio. This value is given by default because odds ratios can be easier to interpret than the coefficient, which is in log-odds units. This is the odds: 53/147 = .361.

# Logistic Regression Analysis: Output

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square c | df d | Sig. c |
|---|---|---|---|---|
| Step 1 b | Step | 65.588 | 4 | .000 |
| | Block | 65.588 | 4 | .000 |
| | Model | 65.588 | 4 | .000 |

**Model Summary**

| Step | -2 Log likelihood e | Cox & Snell R Square f | Nagelkerke R Square f |
|---|---|---|---|
| 1 | 165.701 a | .280 | .408 |

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

**Classification Table a**

| Observed g | | | Predicted h: honcomp .00 | honcomp 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | honcomp | .00 | 132 | 15 | 89.8 |
| | | 1.00 | 26 | 27 | 50.9 |
| | Overall Percentage i | | | | 79.5 |

a. The cut value is .500

**Variables in the Equation**

| | | B j | S.E. k | Wald l | df m | Sig. l | Exp(B) n |
|---|---|---|---|---|---|---|---|
| Step 1 a | read | .098 | .025 | 15.199 | 1 | .000 | 1.103 |
| | science | .066 | .027 | 5.867 | 1 | .015 | 1.068 |
| | ses | | | 6.690 | 2 | .035 | |
| | ses(1) | .058 | .532 | .012 | 1 | .913 | 1.060 |
| | ses(2) | -1.013 | .444 | 5.212 | 1 | .022 | .363 |
| | Constant | -9.561 | 1.662 | 33.112 | 1 | .000 | .000 |

a. Variable(s) entered on step 1: read, science, ses.

b. **Step 1** - This is the first step (or model) with predictors in it. In this case, it is the full model that we specified in the **logistic regression** command. You can have more steps if you do stepwise or use blocking of variables.

c. **Chi-square** and **Sig.** - This is the chi-square statistic and its significance level. In this example, the statistics for the Step, Model and Block are the same because we have not used stepwise logistic regression or blocking. The value given in the Sig. column is the probability of obtaining the chi-square statistic given that the null hypothesis is true. In other words, this is the probability of obtaining this chi-square statistic (65.588) if there is in fact no effect of the independent variables, taken together, on the dependent variable. This is, of course, the p-value, which is compared to a critical value, perhaps .05 or .01 to determine if the overall model is statistically significant. In this case, the model is statistically significant because the p-value is less than .000.

d. **df** - This is the number of degrees of freedom for the model. There is one degree of freedom for each predictor in the model. In this example, we have four predictors: **read**, **write** and two dummies for **ses** (because there are three levels of **ses**).

e. **-2 Log likelihood** - This is the -2 log likelihood for the final model. By itself, this number is not very informative. However, it can be used to compare nested (reduced) models.

f. **Cox & Snell R Square** and **Nagelkerke R Square** - These are pseudo R-squares. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics (these are only two of them). Because this statistic does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic with great caution.

g. **Observed** - This indicates the number of 0's and 1's that are observed in the dependent variable.

h. **Predicted** - These are the predicted values of the dependent variable based on the full logistic regression model. This table shows how many cases are correctly predicted (132 cases are observed to be 0 and are correctly predicted to be 0; 27 cases are observed to be 1 and are correctly predicted to be 1), and how many cases are not correctly predicted (15 cases are observed to be 0 but are predicted to be 1; 26 cases are observed to be 1 but are predicted to be 0).

i. **Overall Percentage** - This gives the overall percent of cases that are correctly predicted by the model (in this case, the full model that we specified). As you can see, this percentage has increased from 73.5 for the null model to 79.5 for the full model.

j. **B** - These are the values for the logistic regression equation for predicting the dependent variable from the independent variable. They are in log-odds units. Similar to OLS regression, the prediction equation is

# Reporting Logistic Regression Analyses

In stepwise logistic regression, several measures of model significance may be used [Hosmer and Lemeshow, 2000]. Table 6 shows such measures for the models of each conference. Specifically, the Hosmer-Lemeshow goodness-of-fit test shows that each of the final regression models is significantly better at determining acceptance/rejection decisions than random chance. The results of the other tests support this finding.

**Table 6: Model Fit for the Effect of Review Criteria Scores on the Acceptance/Rejection Decision**

| Conference | -2 Log Likelihood | Cox & Snell $R^2$ | Nagelkerke $R^2$ | Hosmer-Lemeshow Goodness-of-fit (chi-square, p) |
|---|---|---|---|---|
| ECIS 2007 | 364.67 | 0.48 | 0.67 | (5.30, p = 0.73) |
| BPM 2007 | 62.35 | 0.33 | 0.59 | (2.81, p = 0.95) |
| ER 2007 | 38.96 | 0.57 | 0.86 | (2.53, p = 0.96) |

Table 7 presents several measures describing the importance of the criteria in each of the final regression models. The significance of each criterion was assessed based on the significance of the Wald statistic [Tabachnick and Fidell, 2001].

**Table 7: Effect of Review Criteria Scores on the Acceptance/Rejection Decision**

| Conference | Review Criterion | Beta | SE | Wald | Sig. | Exp (B) |
|---|---|---|---|---|---|---|
| ECIS | **Significance/ Contribution** | 1.12 | 0.25 | 20.50 | **0.00** | 3.05 |
| | **Theoretical Strength** | 0.75 | 0.1 | 18.02 | **0.00** | 2.12 |
| | **Presentation** | 0.73 | 0.18 | 16.48 | **0.00** | 2.07 |
| | **Appeal to Audience** | 0.63 | 0.22 | 8.41 | **0.00** | 1.88 |
| | Methodology used | | | | 0.11 | |
| | Relevance to ECIS | | | | 0.19 | |
| BPM 2007 | **Originality** | 2.07 | 0.53 | 15.15 | **0.00** | 7.96 |
| | **Technical Soundness** | 1.15 | 0.48 | 5.72 | **0.02** | 3.17 |
| | Practical Impact | | | | 0.09 | |
| | Presentation | | | | 0.22 | |
| | Relevance to BPM | | | | 0.26 | |
| ER 2007 | **Technical Quality** | 4.52 | 1.22 | 13.68 | **0.00** | 91.31 |
| | **Significance** | 2.76 | 1.13 | 5.93 | **0.02** | 15.74 |
| | **Relevance to ER** | 2.34 | 0.81 | 8.34 | **0.00** | 10.42 |
| | **Originality** | 1.69 | 0.78 | 4.75 | **0.03** | 5.42 |
| | Presentation | | | | 0.24 | |

Perusal of Table 7 leads to the following observations. First, while all review criteria are significant predictors of the overall evaluation of a paper (as shown in Table 5), they are not necessarily significant predictors of the acceptance/rejection decision. The stepwise regression identified a number of review criteria scores that do not significantly influence the acceptance/rejection decision. For ECIS 2007, four of six review criteria significantly influenced the acceptance/rejection decision: "Significance/contribution," "Theoretical strength," "Presentation," and Appeal to audience." For BPM 2007, just two of five review criteria, "Originality" and "Technical soundness," were significantly associated with the acceptance/rejection decision. Finally, for ER 2007, we found that all review criteria with the exception of "Presentation" significantly influenced the acceptance/rejection decision.

Second, the review criteria that influence the acceptance/rejection decision differ across conferences. We see this, for example, in the common review criteria, "Presentation" and "Relevance to conference." "Presentation" is a significant predictor in the acceptance/rejection decision for ECIS ($\beta$ = 0.73, p = 0.00), but not for BPM (p = 0.22), or ER (p = 0.24). Relevance to the conference, on the other hand, is a significant predictor in the acceptance/rejection decision for ER ($\beta$ = 2.34, p = 0.00), but not for ECIS (p = 0.19) or BPM (p = 0.26). We further note that the originality criterion is a significant predictor for both BPM and ER, while the significance/contribution criterion is a significant predictor for both ECIS ($\beta$ = 1.12, p = 0.00) and ER ($\beta$ = 2.76, p = 0.00).

# Agenda

1. Regression Models
   - When do we need them?
   - Assumptions
   - Conduct
   - Types
     - Linear
     - Hierarchical
     - Logistic Regression
   - Reporting

2. Analysis of Variance
   - When do we need them?
   - Assumptions
   - Conduct
   - Types
     - ANOVA
     - MANOVA
     - ANCOVA
     - MANCOVA
   - Reporting

# Analysis of Variance Models

- a statistical method used to test differences between two or more means.
- Inferences about means are made by analyzing variance.
- Think of it as an extension of t-tests
  - To two or more groups
  - To means+variance rather than only means.
- In a typical ANOVA, the null hypothesis is that all groups are random samples of the same population.
  - For example, when studying the effect of different treatments on similar samples of patients, the null hypothesis would be that all treatments have the same effect (perhaps none).
- Rejecting the null hypothesis would imply that different treatments result in altered effects.
- Often used in experimental research, to study effects of treatments.

# Types of Analysis of Variance Models

- **One-way ANOVA**
  - used to test for differences among two or more independent groups (means).
  - Typically, however, the one-way ANOVA is used to test for differences among at least three groups, since the two-group case can be covered by a t-test (when there are only two means to compare, the t-test and the ANOVA F-test are equivalent).
- **Factorial ANOVA**
  - used when the experimenter wants to study the interaction effects among the treatments.
- **Repeated measures ANOVA**
  - used when the same subjects are used for each treatment (e.g., in a longitudinal study).
- **Multivariate analysis of variance (MANOVA)**
  - used when there is more than one dependent variable.
- **Analysis of covariance (ANCOVA)**
  - blends ANOVA and regression: evaluates whether population means of a DV are equal across levels of a categorical IV [treatment], while statistically controlling for the effects of other continuous variables that are not of primary interest [covariates].

# ANOVA and Research Designs

- The type of ANOVA model is highly dependent on your research design and theory; in particular:

- What are between-subject factors? How many?
- What are within-subject factors? How many?
- What are treatments? How many?
- Which factors are theoretically relevant, which are mere controls?

# ANOVA Assumptions

- Independence, normality and homogeneity of the variances of the residuals
  - Like we discussed last week.
- Note there are *no* necessary assumptions for ANOVA in its full generality, but the F-test used for ANOVA hypothesis testing has assumptions and practical limitations.

# One-way and two-way ANOVA

- **One-way**
  = one-way between groups model
  - E.g., school performance between boys versus girls

- **Two-way**
  = two one-ways for each factor
  PLUS
  interaction between two factors
  - E.g., school performance between boys versus girls and locals versus internationals

- Three-way
  - You get the idea…

# Illustration: Analysis of Variance

- Injuries sustained by kids wearing superhero costumes
- Does it depend on which costume they wear?
  - Superman, Spiderman, Hulk, Ninja Turtle?

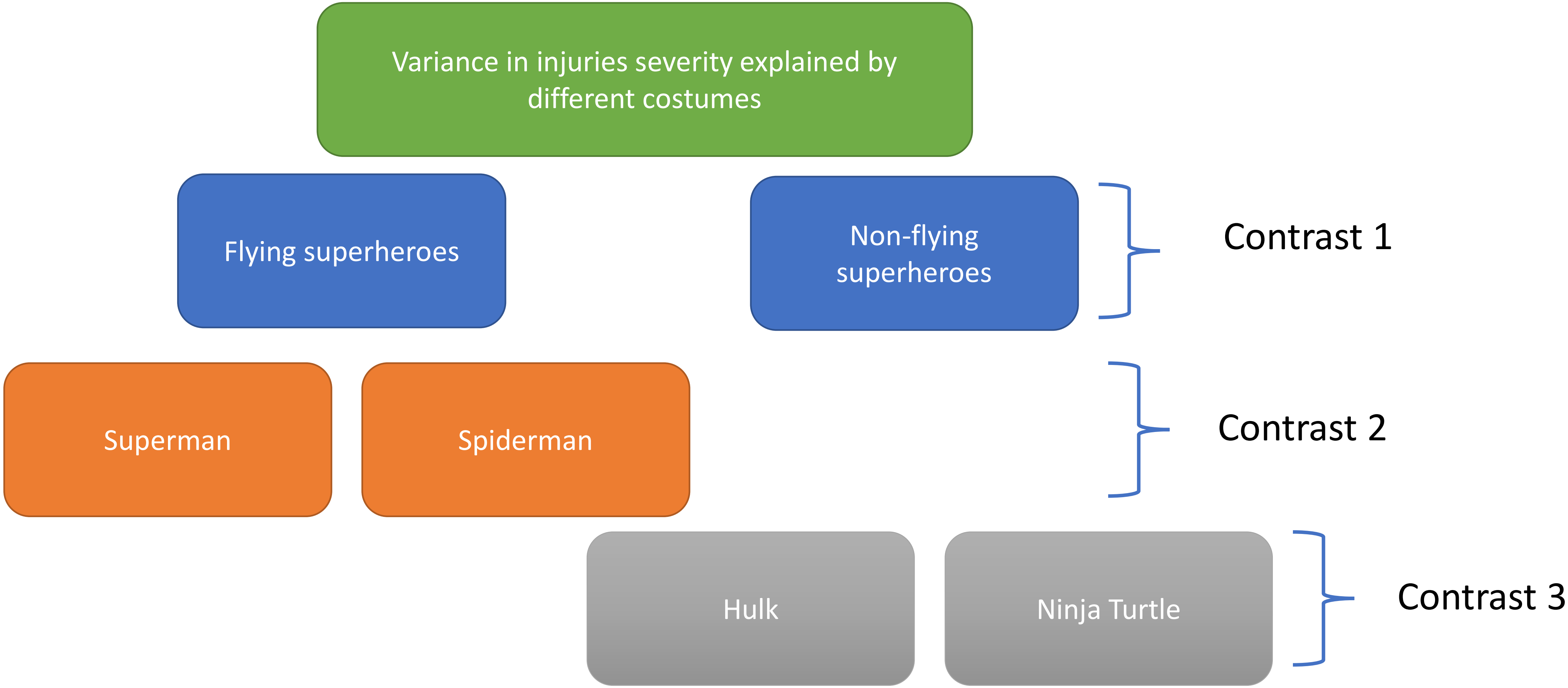
- Adopted from http://www.statisticshell.com/docs/onewayanova.pdf

# What ANOVA could tell us

- Are injuries sustained random or significantly dependent on wearing superhero costumes?

- Is there any order of injuries sustained by type of costume?

# What ANOVA could tell us



# Translated into contrasts

| Contrast | Costume | | | |
|---|---|---|---|---|
| | superman | spiderman | hulk | ninja turtle |
| 1 | 2 | 2 | -2 | -2 |
| 2 | 1 | -1 | 0 | 0 |
| 3 | 0 | 0 | 1 | -1 |

# Important elements of any ANOVA test

- Descriptive Statistics: means, errors, 95% CI
- Levene's test of homogeneity → should be insignificant
- ANOVA results
  - Between groups (model)
  - Within groups (residual variance)
  - Contrast tests (depending on Levene's test results)
  - Post hoc tests (if conducted)

# Illustration: MANOVA

- Usefulness, difficulty and importance of dietary information from three sources
  - Web site
  - Nurse
  - Video

  - Adopted from http://www.ats.ucla.edu/stat/spss/output/SPSS_MANOVA_AO.htm

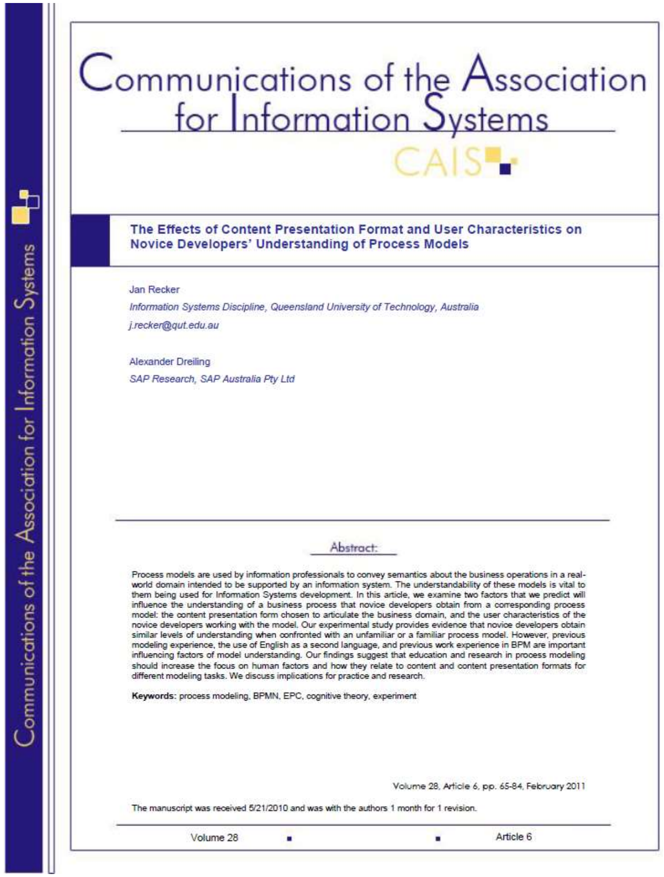
# Illustration: MANOVA

- Usefulness, difficulty and importance of dietary information from three sources
  - Web site
  - Nurse
  - Video

  - Adopted from http://www.ats.ucla.edu/stat/spss/output/SPSS_MANOVA_AO.htm

# Example: Analysis of Variance

- Model experiment:
- Which model (EPC or BPMN) do people understand better?
- Uses MANCOVA
- Read up at [http://eprints.qut.edu.au/40198/](http://eprints.qut.edu.au/40198/)




Communications of the Association for Information Systems

CAIS


The Effects of Content Presentation Format and User Characteristics on Novice Developers' Understanding of Process Models


Jan Recker

*Information Systems Discipline, Queensland University of Technology, Australia*

*j.recker@qut.edu.au*

Alexander Dreiling

*SAP Research, SAP Australia Pty Ltd*



Abstract:

Process models are used by information professionals to convey semantics about the business operations in a real-world domain intended to be supported by an information system. The understandability of these models is vital to them being used for Information Systems development. In this article, we examine two factors that we predict will influence the understanding of a business process that novice developers obtain from a corresponding process model: the content presentation form chosen to articulate the business domain, and the user characteristics of the novice developers working with the model. Our experimental study provides evidence that novice developers obtain similar levels of understanding when confronted with an unfamiliar or a familiar process model. However, previous modeling experience, the use of English as a second language, and previous work experience in BPM are important influencing factors of model understanding. Our findings suggest that education and research in process modeling should increase the focus on human factors and how they relate to content and content presentation formats for different modeling tasks. We discuss implications for practice and research.



**Keywords**: process modeling, BPMN, EPC, cognitive theory, experiment

Volume 28, Article 6, pp. 65-84, February 2011

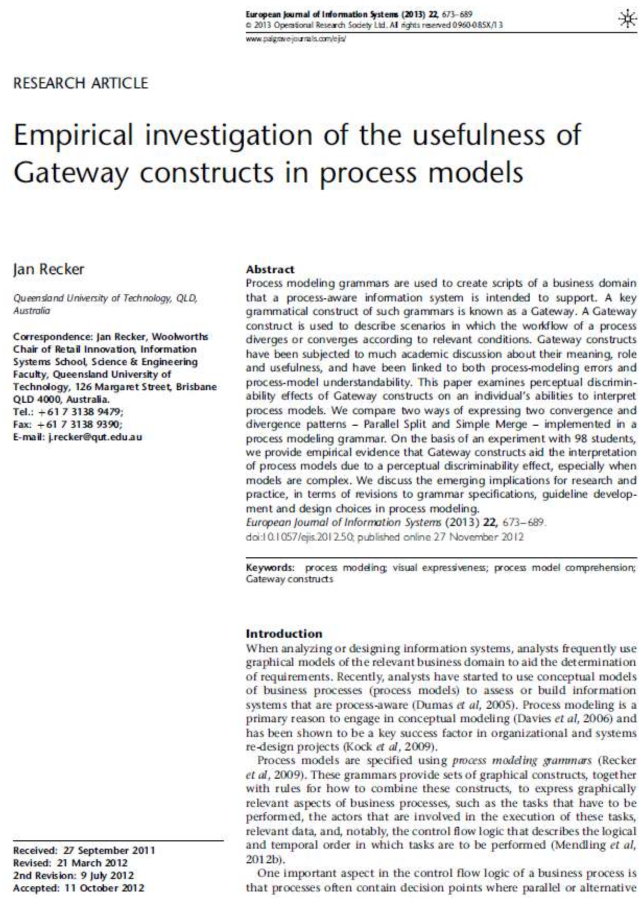
The manuscript was received 5/21/2010 and was with the authors 1 month for 1 revision.

Volume 28 ▪ Article 6

Communications of the Association for Information Systems

# Reporting (M)AN(C)OVA tests

- Example: [http://eprints.qut.edu.au/59428/](http://eprints.qut.edu.au/59428/)
- Reports on
- Repeated measures (M)ANCOVA



European Journal of Information Systems (2013) 22, 673–689
© 2013 Operational Research Society Ltd. All rights reserved 0960-085X/13
www.palgrave-journals.com/ejis/

RESEARCH ARTICLE

## Empirical investigation of the usefulness of Gateway constructs in process models

Jan Recker

*Queensland University of Technology, QLD, Australia*

**Correspondence: Jan Recker, Woolworths Chair of Retail Innovation, Information Systems School, Science & Engineering Faculty, Queensland University of Technology, 126 Margaret Street, Brisbane QLD 4000, Australia.**
**Tel.: +61 7 3138 9479;**
**Fax: +61 7 3138 9390;**
**E-mail: j.recker@qut.edu.au**

**Abstract**
Process modeling grammars are used to create scripts of a business domain that a process-aware information system is intended to support. A key grammatical construct of such grammars is known as a Gateway. A Gateway construct is used to describe scenarios in which the workflow of a process diverges or converges according to relevant conditions. Gateway constructs have been subjected to much academic discussion about their meaning, role and usefulness, and have been linked to both process-modeling errors and process-model understandability. This paper examines perceptual discriminability effects of Gateway constructs on an individual's abilities to interpret process models. We compare two ways of expressing two convergence and divergence patterns – Parallel Split and Simple Merge – implemented in a process modeling grammar. On the basis of an experiment with 98 students, we provide empirical evidence that Gateway constructs aid the interpretation of process models due to a perceptual discriminability effect, especially when models are complex. We discuss the emerging implications for research and practice, in terms of revisions to grammar specifications, guideline development and design choices in process modeling.
*European Journal of Information Systems* (2013) **22,** 673–689.
doi:10.1057/ejis.2012.50; published online 27 November 2012

**Keywords:** process modeling; visual expressiveness; process model comprehension; Gateway constructs

**Introduction**

When analyzing or designing information systems, analysts frequently use graphical models of the relevant business domain to aid the determination of requirements. Recently, analysts have started to use conceptual models of business processes (process models) to assess or build information systems that are process-aware (Dumas *et al*, 2005). Process modeling is a primary reason to engage in conceptual modeling (Davies *et al*, 2006) and has been shown to be a key success factor in organizational and systems re-design projects (Kock *et al*, 2009).

Process models are specified using *process modeling grammars* (Recker *et al*, 2009). These grammars provide sets of graphical constructs, together with rules for how to combine these constructs, to express graphically relevant aspects of business processes, such as the tasks that have to be performed, the actors that are involved in the execution of these tasks, relevant data, and, notably, the control flow logic that describes the logical and temporal order in which tasks are to be performed (Mendling *et al*, 2012b).

One important aspect in the control flow logic of a business process is that processes often contain decision points where parallel or alternative

**Received: 27 September 2011**
**Revised: 21 March 2012**
**2nd Revision: 9 July 2012**
**Accepted: 11 October 2012**

# Reporting (M)AN(C)OVA tests

- Descriptive statistics
  - Mean plus standard deviation
  - Per group
  - Per repeated measures

- Any assumption tests and eventual corrections

Usefulness of Gateway constructs   Jan Recker   681

task scores and the knowledge of control flow logic as a covariate on the individual samples. For all three cases of missing task completion times for the low, average and high complexity model comprehension times, we found the differences in model comprehension task scores and knowledge of control flow logic to be insignificant. *P*-values for the differences in model comprehension task scores were 0.31 (low complexity model comprehension score), 0.31 (average complexity model comprehension score) and 0.61 (high complexity model comprehension score). *P*-values for the differences in knowledge of control flow logic scores were 0.76, 0.59 and 0.95, respectively. These results indicate that bias from missing entries is not significant.

Finally, we examined guessing as a potential response strategy. We tried to minimise learning effects and experiment fatigue bias by randomising the sequence of model comprehension tasks. Yet, participants may have still relied on guessing as an answer strategy. For instance, by relying on random chance, participants would have been able to score on average half of the comprehension questions. We performed one-sample *t*-tests of the model comprehension task scores against the value '2' to examine this potential source of bias. The average scores (see Table 2) were in all cases significantly different from the value '2' (with *P*-values ranging from 0.00 to 0.03). Next, we compared good performers with bad performers in terms of task completion times to examine whether good performance resulted from guessing the right answers, which would be evident from lower task completion times. We created a binary dummy variable based on a median split of the total model comprehension score for all three model cases, and conducted t-tests for each of the three task completion times on the individual samples. While well-performing participants (total comprehension task score>8) were significantly faster in completing the low complexity model task ($t=2.09$, $P=0.02$), they were not significantly faster in completing the average ($t=0.89$, $P=0.38$) and high complexity model tasks ($t=1.36$, $P=0.18$). These results suggest that good comprehension scores were comparable in terms of the time investments into the tasks. Finally, we compared whether participants that received the low complexity model comprehension task prior to a high complexity comprehension task completed their comprehension tasks faster, and vice versa, which would indicate a form of experiment fatigue in which participants seek to quickly select answers only to complete the study. Forty-six participants received a low complexity model prior to receiving the high complexity model, and 52 participants vice versa. Independent samples t-tests between the groups showed that task completion times for the low and high complexity model were not significantly different across these two groups, although an effect for the high complexity model can be noted ($t=0.46$, $P=0.64$, and $t=1.70$, $P=0.07$, respectively). Overall, we posit that response bias is minimal in our study.

**Hypotheses tests**

Data associated with interpretational fidelity – measured through comprehension task scores – were analyzed using a repeated measures Analysis of Covariance (ANCOVA) test, with the between-subject factor treatment (with two levels) and the within-subject factor complexity (with three levels), and using prior control flow knowledge as a covariate. The tests were computed using IBM SPSS Statistics Version 19.0.

Mauchly's test of sphericity was significant ($\chi^2=7.87$, $P=0.02$), suggesting the use of Greenhouse-Geisser correction for sphericity of 0.93 (Hair *et al*, 2010). Table 3 shows average scores across all participants (mean) and standard deviations (std. deviation) and Table 4 describes the results from the repeated measures ANCOVA test, including the degrees of freedom (df), the results from the F-test (F), the resulting significance value *P* (sig.) and the effect size (partial eta squared). Table 4 also report the corrected degrees of freedom associated with the model error term (error) as per reporting guidelines in Hair *et al* (2010).

To examine differences in interpretational efficiency – measured through comprehension task completion times scores, we repeated the data analysis, viz., we again used a repeated measures ANCOVA test, with the same independent factors treatment and complexity, and using prior control flow knowledge as a covariate. As a dependent factor we considered the comprehension task completion times scores. Again, Mauchly's test of sphericity was significant ($\chi^2=9.15$, $P=0.01$), and thus we again used a Greenhouse-Geisser correction for sphericity of 0.93 (Hair *et al*, 2010). Table 5 shows mean values and standard deviations and Table 6 gives the results from the repeated measures ANCOVA test.

**Discussion**

**Summary of results**

Our empirical study set out to test four hypotheses about the effects of representational forms for convergence/divergence and complexity of process models on the

**Table 3 Means (standard deviations) for comprehension task scores**

| *Type* | *Mean* | *Std. deviation* |
|---|---|---|
| *Low complexity model* | 3.57 | 0.73 |
| with use of connectors | 3.77 | 0.51 |
| without use of connectors | 3.35 | 0.88 |
| *Average complexity model* | 2.78 | 1.06 |
| with use of connectors | 2.92 | 1.12 |
| without use of connectors | 2.61 | 0.98 |
| *High complexity model* | 2.27 | 1.18 |
| with use of connectors | 2.31 | 1.15 |
| without use of connectors | 2.02 | 1.18 |

European Journal of Information Systems

# Reporting (M)AN(C)OVA tests (ctd.)

- Each analysis results
  - Df
  - F
  - Sig.
  - Partial Eta Squared
  - Error



**Table 4 Results of the repeated-measures ANCOVA for comprehension task scores**

| *Factor* | *df* | *F* | *Sig.* | *Partial eta squared* |
|---|---|---|---|---|
| *Between-subjects* | | | | |
| Treatment | 1 | 3.78 | 0.05 | 0.03 |
| Control flow knowledge [covariate] | 1 | 9.12 | 0.00 | 0.09 |
| Error | 95 | | | |
| *Within-subjects* | | | | |
| Complexity | 1.85 | 22.76 | 0.00 | 0.19 |
| Complexity × treatment | 1.85 | 4.85 | 0.03 | 0.05 |
| Complexity × control flow knowledge [covariate] | 1.85 | 8.21 | 0.01 | 0.08 |
| Error | 175.87 | | | |

**Table 5 Means (standard deviations) for comprehension task completion times**

| *Type* | *Mean* | *Std. deviation* |
|---|---|---|
| *Low complexity model* | 178.70 | 141.00 |
| with use of connectors | 150.13 | 98.16 |
| without use of connectors | 210.00 | 179.19 |
| *Average complexity model* | 173.26 | 111.00 |
| with use of connectors | 202.48 | 226.35 |
| without use of connectors | 145.36 | 118.13 |
| *High complexity model* | 121.54 | 146.90 |
| with use of connectors | 120.50 | 173.16 |
| without use of connectors | 121.54 | 146.90 |

**Table 6 Results of the repeated-measures ANCOVA for comprehension task completion times**

| *Factor* | *df* | *F* | *Sig.* | *Partial eta squared* |
|---|---|---|---|---|
| *Between-subjects* | | | | |
| Treatment | 1 | 0.31 | 0.58 | 0.00 |
| Control flow knowledge [covariate] | 1 | 0.07 | 0.80 | 0.00 |
| Error | 77 | | | |
| *Within-subjects* | | | | |
| Complexity | 1.88 | 0.45 | 0.63 | 0.01 |
| Complexity × treatment | 1.88 | 2.47 | 0.09 | 0.03 |
| Complexity × control flow knowledge [covariate] | 1.88 | 0.19 | 0.81 | 0.00 |
| Error | 144.48 | | | |

interpretability of differently complex models in terms of their interpretational fidelity and efficiency.

In hypothesis H1 we speculated that the use of Gateway constructs will have a significant positive effect on interpretational fidelity (measured through comprehension task scores). Table 4 shows that the treatment variable (the use *vs* non-use of Gateway constructs) had a consistently significant effect on the comprehension task performance ($F=3.78$, $P=0.05$). The mean comprehension task scores shown in Table 3 further show that indeed in all cases interpretational fidelity was increased when Gateway constructs were used in the model. These results support hypothesis H1.

In hypothesis H3 we then speculated that the positive effects of Gateway constructs on model interpretational fidelity increase when model complexity is increased. The data displayed in Tables 3 and 4 shows that, first, interpretational fidelity decreased significantly ($F=22.76$, $P=0.00$) when model complexity was increased, from an average comprehension task score of 3.57 (low complexity model) to 2.78 (average complexity model) and 2.27 (high complexity model). Table 4 further shows that the interaction effect between model complexity and treatment was significant ($F=4.85$, $P=0.03$), showing that the treatment effect increased when model complexity was increased. These results support hypothesis H3.

In hypothesis H2 we speculated that the use of Gateway constructs will have a significant positive effect on interpretational efficiency (measured by task completion time). The data in Table 5, however, show mixed results. For low complexity models, average task completion times were lower when Gateway constructs were used in the model (mean = 139.79 *vs* mean = 188.26), but for the average complexity models, the effect was reversed (mean = 180.94 *vs* mean = 134.52). For the high complexity models, differences were virtually non-existent (mean = 111.09 *vs* mean = 111.05). Table 6 confirms that the treatment effect was insignificant ($F=0.05$, $P=0.95$). These results are contrary to hypothesis H2.

In hypothesis H4 we speculated that the positive perceptual discriminability effects of Gateway constructs on interpretation efficiency will increase for complex models. The data in Table 5 show, however, that comprehension task completion times decreased when model complexity was increased (from mean = 162.66 to 159.03 and 111.07). The differences, however, were not significant ($F=0.85$, $P=0.43$). Likewise, the interaction effect Complexity × treatment was not yielding significant differences ($F=2.47$, $P=0.09$). These results are contrary to hypothesis H4.

Finally, we note that control flow knowledge was a significant covariate for explaining comprehension task performance but not for explaining comprehension task completion times. These results are largely in line with prior studies (Mendling *et al*, 2012b).

**Discussion**

With respect to interpretational fidelity, our results show that both our hypotheses (H1 and H3) are fully supported from the data. Specifically, we found that a visually explicit representation form chosen to express convergence and divergence has a significant positive

# A Detailed Look

**Table 3** Means (standard deviations) for comprehension task scores

| *Type* | *Mean* | *Std. deviation* |
|---|---|---|
| *Low complexity model* | 3.57 | 0.73 |
| with use of connectors | 3.77 | 0.51 |
| without use of connectors | 3.35 | 0.88 |
| *Average complexity model* | 2.78 | 1.06 |
| with use of connectors | 2.92 | 1.12 |
| without use of connectors | 2.61 | 0.98 |
| *High complexity model* | 2.27 | 1.18 |
| with use of connectors | 2.31 | 1.15 |
| without use of connectors | 2.02 | 1.18 |

**Table 4** Results of the repeated-measures ANCOVA for comprehension task scores

| *Factor* | *df* | *F* | *Sig.* | *Partial eta squared* |
|---|---|---|---|---|
| *Between-subjects* | | | | |
| Treatment | 1 | 3.78 | 0.05 | 0.03 |
| Control flow knowledge [covariate] | 1 | 9.12 | 0.00 | 0.09 |
| Error | 95 | | | |
| *Within-subjects* | | | | |
| Complexity | 1.85 | 22.76 | 0.00 | 0.19 |
| Complexity × treatment | 1.85 | 4.85 | 0.03 | 0.05 |
| Complexity × control flow knowledge [covariate] | 1.85 | 8.21 | 0.01 | 0.08 |
| Error | 175.87 | | | |

http://eprints.qut.edu.au/59428/, pg. 681-682

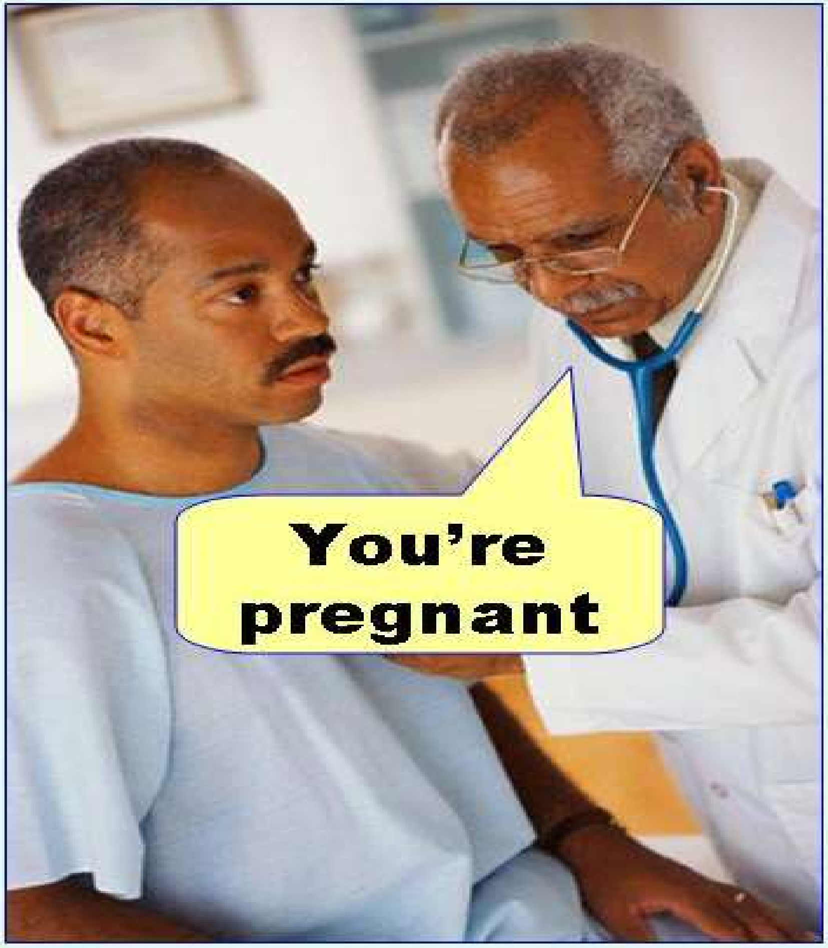
# Some Tips

- Conduct Power Analysis
  - Upfront to understand sample size requirements of your chosen design
  - http://www.gpower.hhu.de/en.html
- Contrasts are very powerful tools in conjunction with post-hoc tests
  - If you have categorical instead of binary IV
- Always test different ANOVA models
  - With/without covariates
  - With/without interaction effects
- Beware of type-1/type-2 errors!
  - Very prevalent in (M)ANOVAs!

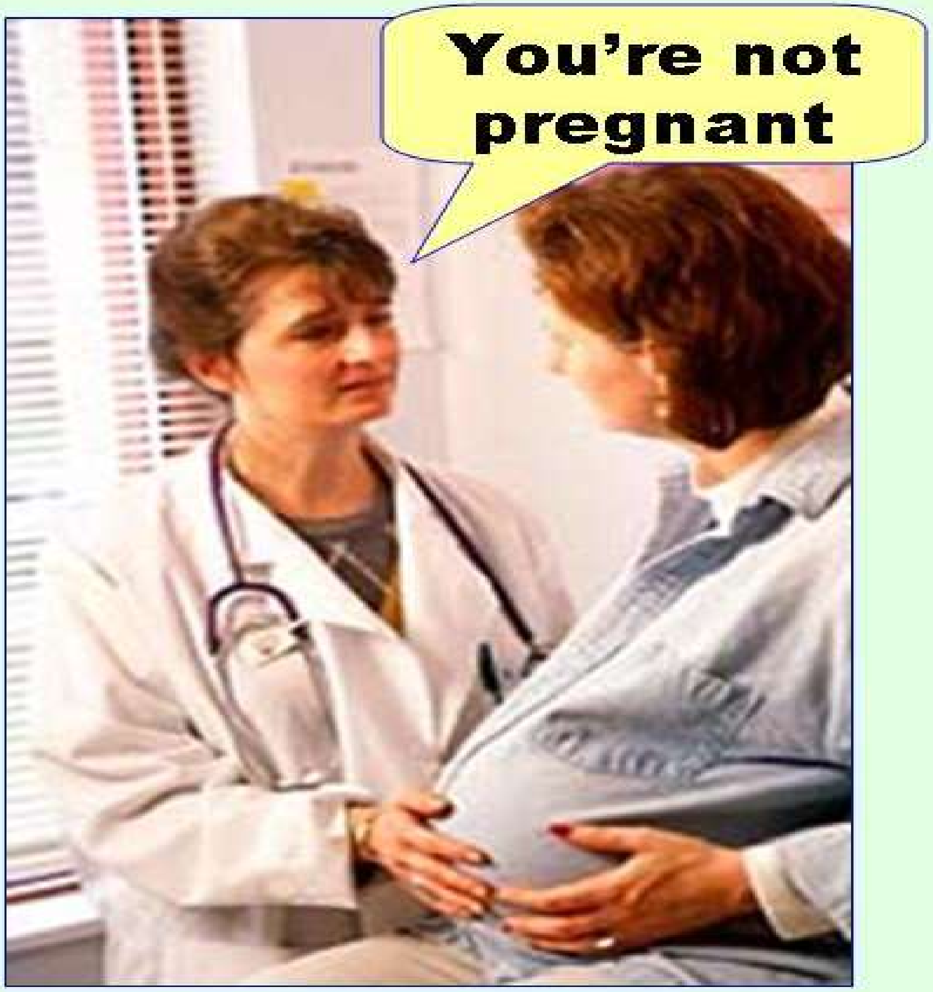
# Type-1 and Type-2 Errors

| | | Reality | |
|---|---|---|---|
| | | True | False |
| Measured / Perceived | True | | |
| | False | | |

**Type I error**
(false positive)




**Type II error**
(false negative)



# End of Part 2 and 3