Statistical Analysis Techniques in Serious Games Research

Thomas Hainey
Prof. Thomas Connolly

Structure of Presentation

- Basics of experimental design and statistics
- Descriptive statistics
- Inferential statistics
- Simple experimental designs
- Single factorial designs
- Correlation and regression

Basics of experimental design and statistics

- Experimental designs are based on *predictions*
- *Prediction* – "statement that a change in one thing (the **independent variable** or IV) will produce a change in another thing (the **dependent variable** or DV)"
- Independent variable will relate to a change in the *conditions* governing behaviour and the dependent variable will respond to some measure of the subject's behaviour or *performance* under those conditions.

Basics of experimental design and statistics

- ***Research Design*** – the formulation of a plan for collecting relevant data
- Scientists are in general agreement that the most effective means of testing a prediction is deliberately to manipulate the independent variable and to observe the consequential changes in the dependent variable.
- It is the only method of collecting data – the *experimental* method that has the power to reveal cause-and-effect relationships in an unambiguous way.

Basics of experimental design and statistics

- **Experiment –** a means of collecting evidence to show the effect of one variable upon another. In the ideal case the experimenter manipulates the IV, holds all other variables constant, and then observes the changes in the DV
- **Irrelevant variables** - Since we are only interested in the effects of IV then all other factors are referred to as *irrelevant variables.* Complete control over all irrelevant variables can never be achieved. This is either physically impossible or practically impossible
- Irrelevant variables can lead to the *confounding factor*

Basics of experimental design and statistics

- **Subject variables** – these are irrelevant variables associated with the subject such as intelligence, motivation, personality etc.
- **Situational variables –** these are the variables associated with the conditions under which the experiment is conducted such as background noise, instructions and experimenter’s tone of voice.

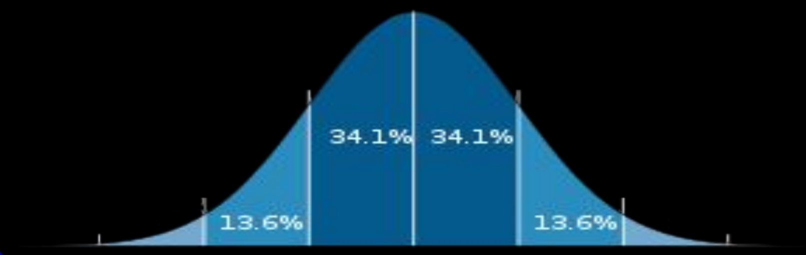
Basics of experimental design and statistics

Subject variable control ranking:

- Repeated measures - same subjects in each group
- Matched subjects - mimics the repeated measures but requires pairs of very similar individuals
- Independent groups - different groups
- Randomisation ensures no *systematic bias* in favour of one group and controls situational variables

Descriptive statistics

- Presents numerical descriptions in a manageable form
- *Distribution* – summary of frequencies or individual values or ranges for a variable
- *Normal/Gaussian distribution* – is the shape of a bell curve



- Can be defined by a function with only two parameters (the mean and the standard deviation)

Descriptive statistics

Central Tendency (middle)

- Mean
  - Probably the most often used
  - Sum of observations / num of observations
- Median
  - Divides a distribution in half where one half of the scores is <= median and one half of scores is >= median
  - Odd 4, 7, 8, 11, 15, Even 4, 4, 5, 6, 8, 14
  - 1, 3, 3, 4, 5, 5, 5, 5, 7, 8
- Mode
  - The value of the most frequently occurring observation
  - 4, 5, 9, 3, 6, 1, 3, 5, 3, 7, 1, 5, 2, 7, 5, 4, 6, 3, 2, 8
  - 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 8, 9

Descriptive statistics

- Which Central Tendency Measure?
  - If the distribution is completely symmetrical then mode, median and mean are equal
  - If the distribution is skewed then it's better to use the median than the mean
  - The mean is the most preferred – it's a good indication of the balance point for most normally distributed data and it has numerical properties that allow it to be manipulated in useful ways
  - The mode is more suitable for qualitative data

Descriptive statistics

- Dispersion – spread of values within a distribution. A common measurement is the range however a measurement with more utility describes the relationship between the spread of scores and the mean of those scores
- Each score in the sample differs from the mean (0, <, >) – difference of each score is known as a *deviation*
- We calculate each deviation and have a new set of scores some of which will be negative
- To deal with this we square each deviation and add them together to get the *sum of squares*

Descriptive statistics

- To derive the *variance* we divide the sum of squares by n (number of scores in our sample) – 1
- *Standard deviation* is the square root of the variance. It is a more accurate and detailed estimate because an *outlier* value can effect the range
  - Find the mean, find the distance between each score and the mean, square each distance, add all the squared distances, divide the sum by the number of scores – 1, find the square root of the result

Descriptive statistics

- 15, 30, 33, 34
- (15+30+33+34) / 4 = 28
- Differences from the mean
  - 15 – 28 = -13
  - 30 – 28 = +2
  - 33 – 28 = +5
  - 34 – 28 = +6
- Square each deviation
  - -13 x -13 = 169
  - +2 x +2 = 4
  - +5 x +5 = 25
  - +6 x +6 = 36
- 234 / number of scores – 1 = 78
- √78 = 8.83

Descriptive statistics

How good a description of the set of scores is contained in the mean and the SD?

- *Standard error of the mean* – if we were to measure a variable from not just one sample but many – take the mean of the means and the SD of all of the means
- *Z-score* – indicates the relative position of an individuals score in the distribution. Score – Mean / SD – IQ as an example
- 2 disruptions to normality – Skew and Kurtosis

Inferential statistics

- We also use statistics in a much more fundamental way to draw *inferences* from the results of an experiment. We do not simply want to report the scores of our two groups of subjects or to give the mean and standard deviation and leave it at that. We want to use the data to test the original prediction – to decide whether the independent variable is having the effect we supposed, or whether, perhaps, there is no real difference between the performance of the two groups.
- The problem for statistical inference is to decide whether the actual differences are caused by chance or whether these differences are so large that we can ascribe them at least in part, to the effects of the independent variable
- Whether the change is *significant*

Inferential statistics

- What is significant? - most experimental psychologists choose a *significance level* of 0.05 or 1/20 - This means that a difference between two groups will be assumed to reflect chance factors unless the results could only arise by chance one time in twenty, or less.
- Randomness and uncertainty is modelled within statistical theory by *probability*
- P(x) = number of possibilities favouring the occurrence of x / total number of possibilities

Inferential statistics

- Hypothesis Testing
  - *Null and alternate hypothesis* - We have represented the process of statistical inference as deciding between two competing explanations of the difference between our two test scores. Either the differences arise because of purely chance fluctuations in the two groups of scores – NULL Hypothesis or the differences are caused at least in part, by the independent variable – ALTERNATE Hypothesis.

Inferential statistics

- State the null and alternative hypothesis.
- Caclulate the probability, p, that our results could have been obtained under the null hypothesis. This involves:
  - Selecting the right statistical test.
  - Executing the test.
- Reject null hypothesis if $p \leq 0.05$ (i.e. accept the alternate hypothesis that the difference between the groups is cause by the independent variable). Do not reject the null hypothesis is $p > 0.05$.
- Type I error we reject a true null hypothesis (false positive)
- Type II we fail to reject a false null hypothesis (false negative)

Inferential statistics

Selecting an appropriate statistical test

- Two major factors determine the correct test for any particular set of experimental results. 1) the research design and 2) the nature of the dependent variable, that is the actual data.

1) Research Design – One-sample design, two-sample designs, k-sample designs, correlations

Inferential statistics

2) Parametric

- Normality of distribution – Shapio-Wilk's W test is a formal test for normality. Kilmorgorov-Smirnov (K-S) test or K-S Lillefors test, chi-square test
- Homogeneity of variances – Levene's test, Bartlett's test
- Interval level data

Simple experimental designs

"An unsophisticated researcher uses statistics as a drunken man uses lamp-posts – for support rather than illumination." (Andrew Lang)

- The simplest experimental design compares results from two conditions; however there are different variations even of this simple design:
  - *Repeated measures/related measures within-subjects design* – measurements made within a group of people
  - *Independent measures/unrelated samples/between-subjects design* – measurements made between two groups of people

Simple experimental designs

Analysing data from experiments with two conditions

- *t-test* can be performed. Assess whether the means of two sets of data are statistically different from each other
- We should be confident that the conditions are statistically different because of something we have done to one of the conditions
- *t* = difference between the group means / variability of groups
- If the *t* value is positive then the first mean is larger than the second
- Bottom line: If $p<0.05$ you can say that the data sets are significantly different. If $p>0.05$ they are not. The lowest the p-value, the highest your confidence that data sets are different

Simple experimental designs

Analysis of Data in Within-Subjects Design

- Here we are only concerned with the difference between the two groups so we can treat this as one sample of data described as difference (D) – D / variability of groups
- *Wilcoxon test* – Non-parametric equivalent used for smaller samples in Within-Subject Designs

Simple experimental designs

- Analysis of Data in Between-Subjects Design
- An independent sample *t*-test can be performed
- *Mann-Whitney U test* – Non-parametric equivalent used in Between-Subjects Design
- One- and two-tailed *t*-tests – directional hypothesis

Single Factor Designs

- Experimental designs in which we set up a comparison between two sets of data may be somewhat simplistic – Perfetti (1974) suggests 7 levels

ANOVA – Analysis of Variance – analyses the relation between a dependent variable and one or more factors

- A one-way ANOVA compares means of two or more levels of a variable taking into account that variance amongst the groups of data
- Why not carry out numerous *t*-tests? Each test has a certain significance level indicating the probability of making a Type I error. If we carry out more tests with the associated alpha then it becomes compounded and Type I error probability rises

Single Factor Designs

ANOVA

- An omnibus test - the equality of any number of population means in a single F-test
- Greater power than a collection of t-tests and much less effort involved
- Compares the variation (measured by items related to the variance) between the levels of the factor with the variation within the groups
- Parametric test

Single Factor Designs

- If we can see that means are different it is not simply enough to say this – we need to take into account the variation in the data – we need the SS within the groups and between the groups and the total SS
- 1. Calculates the total sum of squares
- 2. Estimate total variability of scores around each group mean (within groups sum of squares)
- 3. Difference between 1 and 2 is how much of the total variability is due to the group effect (between groups sum of squares)

Single Factor Designs

- 4. Calculate the mean square by dividing each sum of squares by the relevant degrees of freedom
- 5. Calculate the F ratio (between group MS to within group MS). If F value is small then the samples belong to the same population, If F value is large then different populations.
- Fortunately we have programs to do this for us!

Correlation and Regression

- Correlation – measures the “degree of association” between two variables. The associations can either be positive or negative
- Correlation is commonly measured with Pearson’s correlation coefficient (r) – r will take values between -1 (perfect negative association) and +1 (perfect positive association)
- The correlation coefficient doesn’t give us the value of the slope. When r is +ve then it slopes upwards left to right. When r is -ve it slopes downwards
- Evidence of association does not imply cause and effect

Correlation and Regression

- Regression – evidence of a linear relationship between two or more variables
- One variable has the role of predictor and the other response
- Formulation of the equation of a straight line: $y = bx + a$
- Where there are two or more predictor variables this is termed multiple regression