PSY3062 Research Design – READING NOTES

Week 1

<u>The Role of Music in Everyday Life: Current Directions in the Social Psychology of Music</u>

- People spend more than 15% of their waking hours with music playing

<u>Why Study Music?</u>

- Studies are beginning to establish robust connections between music and emotion, personality, self-identity, and relationships
- Findings suggest that people use music to regulate their moods and emotions
- situations can influence the styles of music people choose, which can affect behaviour in those situations
- individual differences in music preferences are linked to personality traits and values
- people use music as a vehicle for self-expression
- Similarity in music preferences is associated with attraction, closeness, and relationship satisfaction

<u>Frameworks for Studying Music in Everyday Life</u>

1. The media effects model
2. The uses and gratifications model

Media Effects

- The media effects model is one of the most influential frameworks for understanding mass media
- Based on belief that media have a direct impact on how people think, feel, and behave
- Assumption underlying research in this area is that exposure to music primes individuals to think/feel in ways congruent with the message of the music
- Participants exposed to tense sounding music with violent, as compared to non-violent, lyrics reported having more hostile feelings and aggressive thoughts – violent lyrics not the music itself triggered the aggressiveness
- Exposure to music with themes of violence, misogyny, and gender stereotypes can have harmful consequences
- Music can have positive effects - increasing prosocial behaviour, reducing prejudice
- Genre, tempo, social connotations of music in shops and restaurants can influence which products people buy, how quickly they move, how long they stay

Limitations

- Assumption that individuals are passive recipients of music, listening idly to whatever music they encounter
- Little attention is given to individual differences in preferences – eg not everyone prefers to listen to violent music

Uses and Gratifications

- The uses and gratifications model regards individuals as active agents who seek out or avoid particular content
- Assumptions underlying are that there are individual differences in media preferences and that people consume media to fulfil basic needs
- The uses and gratifications model is an individual differences approach
- It seeks to identify the motives and traits that underlie people's reasons for listening to music
- Research with the uses and gratifications model has examined people's motivations for listening to music and preferred music in particular

Most common self-reported reasons for listening to music are:

- Pass the time
- Regulate emotions
- Connect with peers
- Create an atmosphere
- Concentrate
- Increase physiological arousal
- Convey an image to others

<u>Current Directions in the Social Psychology of Music</u>

Mood and Emotion

- People are able to perceive emotions in music and that individuals generally perceive similar emotions in the same pieces of music
- Evidence that the emotions perceived in music are often the same emotion that the composer or performer intended to communicate
- Evidence that music does indeed elicit certain emotions and moods in listeners
- Limitations of this research is that it decontextualizes the music listening experience
- people experience music-induced emotions frequently + most of the emotions are positive
- People typically use music to achieve or maintain a positive affective state

How does music evoke emotional reactions in listeners?

Several mechanisms → autobiographical memory, emotional contagion, expectancy

Framework of psychological mechanisms responsible for evoking emotion from music:

1. cognitive appraisal
2. Brain stem reflexes
3. Evaluative conditioning
4. Emotional contagion
5. Visual imagery
6. Episodic memory
7. Musical expectancy

Personality and Individual Differences

Current work on individual differences in music preferences aims to identify links with explicit traits, values, and abilities

Assumption: individuals seek musical environments that reinforce and reflect aspects of their personalities, attitudes, and emotions

Music-Preference Dimensions

1. Mellow
2. Unpretentious
3. Sophisticated
4. Intense
5. Contemporary

- **Mellow** comprises soft rock, R & B, and adult contemporary - characterized as romantic, relaxing, slow, and quiet
- **Unpretentious** comprises country and folk -characterized as uncomplicated, relaxing, unaggressive, and acoustic
- **Sophisticated** comprises classical, opera, jazz, and world - characterized as inspiring, intelligent, complex, and dynamic
- **Intense** comprises rock, punk, and heavy metal - characterized as distorted, loud, aggressive, and not romantic, nor inspiring
- **Contemporary** comprises rap, electronica, and pop - characterized as percussive, electric, energetic, not sad

Music preferences associated with various internalizing and externalizing problem behaviours

- **I**nternalizing behaviour, such as self-harm, was high among fans of heavy metal and rock music
- Externalizing behavior problems, including aggression and substance abuse, was comparatively high among fans of rock, heavy metal, and rap music
- Findings suggest music preferences and problem behavior may be manifestations of the same underlying dispositions

Crystallisation Hypothesis

- States that the music individuals enjoy in adolescence and early adulthood crystallizes and becomes the music they prefer throughout adulthood
- Music that people listen to in adolescence becomes psychologically and physiologically significant due to hormonal, emotional, physical changes etc
- The music that people listened to in adolescence is remembered across the lifespan and becomes a strong source of nostalgia

Self and Identity

- People, particularly young adults, place considerable importance on their music preferences.
- Music serves as a symbolic representation of self - individuals derive a sense of identity from the music they listen to
- individuals engage in a reflexive process of remembering and constructing their identities while listening to music - serve as a form of self-affirmation and discovery
- To the extent that individuals are attracted to a style of music - may align their self-image with the characteristics of that music
- people use music to achieve optimal distinctiveness
- individuals with preferences for styles of music with intermediate levels of popularity (optimally distinct) invested more commitment to their musical identities compared to people who preferred musical styles with limited or broad popularity

Social Perception

- people use their favourite music as an identity badge to broadcast information about themselves to others
- people believe their music preferences reveal information about who they are and can help them learn more about others
- Evidence that information about individuals' music preferences influences how they are perceived
- people have stereotypes about the psychological and social characteristics of most music fans – particularly fans of classical, rap, and heavy metal music
- older adults do not use music as an identity badge

Attraction and Social Bonding

- evidence that music does indeed affect attraction
- People are attracted to individuals who share their music preferences
- Underlying Assumption: people who enjoy the same music see and experience the world in similar ways and therefore agree about more things than do people with different preferences

Week 2

Chapter 4: Exploring Data with Graphs

Presenting Data

- Show the data
- Induce reader to think about data
- Avoid distorting the data
- Present many numbers with minimum ink
- Make large data sets coherent
- Encourage reader to compare different pieces of data
- Reveal underlying messages of data

Chartjunk: superfluous material that distracts from the data being displayed

- Vertical axis = y-axis (ordinate) of graph
- Horizontal axis = x-axis (abscissa) of graph
- Don't create false impressions of what data show by scaling y-axis in a weird way
- Abolish chartjunk – don't use 3D effects, shadows
- Avoid excess ink

Histograms

Simple Histogram → use when you want to see the frequencies of scores for a single variable

Stacked Histogram → if you had a grouping variable you could produce a histogram in which each bar is split by group – way to compare relative frequency or scores across groups

Frequency Polygon → displays the same data as the simple histogram except it uses a line instead of bars to show the frequency, and area below the line is shaded

Population Pyramid → like a stacked histogram, shows the relative frequency of scores in two populations

- Plots the variable on the vertical axis and the frequencies for each population on the horizontal axis: populations appear back to back on the graph
- If the bars either side of dividing line are equally long – the distributions have equal frequencies

Boxplots – aka box-whisker diagram

Centre of the plot is the median – which is surrounded by a box, the top and bottom of which are the limits within which the middle 50% of observations fall (interquartile range)

- Sticking out of the top and bottom are two whiskers which extend to the highest and lowest extreme scores (25% approx.)

Simple Boxplot → use this when you want to plot a boxplot of a single variable, but also want different boxplots for different categories

Clustered Boxplot → sample as simple, but can select a second categorical variable on which to split the data

- Boxplots for second variable are produced in different colours

1-D Boxplot → use this when you want to see a boxplot for a single variable – differs from simple boxplot in that no categorical variable is selected for the x-axis

Bar Charts and Error Bars

Bar Charts: graph in which a summary statistic (mean) is plotted on the y-axis against a categorical variable on the x-axis (categorical variable). Value of the mean for each category is shown by a bar

Simple Bar → use when you want to see the means of scores across different groups of cases

Clustered Bar → if you had a second grouping variable you could produce a simple bar but with bars in different colours for levels of a second grouping variable

Stacked Bar → same as clustered bar, except the different-coloured bars are stacked on top of each other rather than side by side

Simple 3D Bar → same as the clustered bar – except second grouping variable is displayed not by different-coloured bars, but by an additional axis

Clustered 3D Bar → like clustered bar, except you can add a third categorical variable on an extra axis

- Means will be impossible to read SO DON'T USE THIS

Stacked 3D Bar → same as clustered 3D – except the different-coloured stacked bars are stacked on top of each other instead of side by side

- Not a good graph for presenting data clearly

Simple Error Bar → same as simple bar chart – except instead of bars, the mean is represented by a dot, and a line represents the precision of the estimate of the mean

- Can add these error bars to a bar chart anyway

Clustered Error Bar → same as clustered bar chart – except the mean is displayed as a dot with an error bar around it

- Error bars can be added

Error Bar Chart: graphical representation of the mean of a set of observations that includes the 95% confidence interval of the mean
- The mean is usually represented as a circle, square or rectangle at the value of the mean
- The confidence interval is represented by a line protruding from the mean to a short horizontal line representing the limits of the confidence interval
- Error bars can be drawn using the standard error or standard deviation instead of the 95% confidence interval

<u>Line Charts</u>

Bar charts with lines instead of bars

Simple Line → use this option when you just want to see the means of scores across different groups of cases

Multiple Line → equivalent to the clustered bar chart – you can plots means of a particular variable but produce different-coloured lines for each level of a second variable

<u>Scatterplots</u>

Graph that plots each person's score on one variable against their score on another

Simple-Scatter → use when you want to plot values of one continuous variables against another

Grouped Scatter → like a simple scatterplot, except you ca display points belonging to different groups in different colours (or symbols)

Simple 3D Scatter → use this to plot values of one continuous variable against values of two others

Grouped 3D Scatter → use if you want to plot values of one continuous variable against two others, but differentiating groups of cases with different coloured dots

Summary Point Plot → same as a bar chart, but dots used instead of bars

Simple Dot Plot → aka density plot, similar to histogram – except rather than having a summary bar representing the frequency of scores, a density plot shows each individual scores as a dot
- Useful for looking at shape of distribution

Scatterplot Matrix → produces a grid of scatterplots showing the relationships between multiple pairs of variables

Drop-Line → produces a graph similar to a clustered bar chart, but with a dot representing a summary statistic (eg mean) instead of a bar, and with a line connecting means of different groups
- Can be useful for comparing statistics – eg mean – across different groups

----

Chapter 5: The Beast of Bias

What is Bias?
- When fitting a model to the data – we estimate the parameters and use the method of least squares
- We use the sample data to estimate the value of the parameters in the population
- When we estimate a parameter we compute an estimate of how well it represents the population – eg a standard error or confidence interval
- Can also test hypotheses about these parameters by computing test stats + their associated probabilities (p-values)

Bias considered within 3 contexts:
1. Things that bias the parameter estimates (eg effect sizes)
2. Things that bias standard errors (SE) and confidence intervals
3. Things that bias test statistics and p-values

Situations above are related
- If the standard error is biased → the confidence interval will also be biased because it is based on the standard error
- Test statistics are usually based on the standard error – so if the SE is biased test stats will be too
- If the test statistics is biased then so will the p-value

Sources of bias:
1. Outliers
2. Violations of assumptions

Assumptions
- An assumption is a condition that ensures that what you're attempting to do works
- If assumptions are not true (violated) then the test statistics and the p-value will be inaccurate and could lead to wrong conclusions

Regular assumptions
- Additivity and linearity
- Normality of something or other

- Homoscedasticity/homogeneity of variance
- independence

Parametric Tests: tests that require data from one of the large catalogue of distributions described by statisticians

- Parametric tests based on the normal distribution – which require 4 basic assumptions be met for the test to be accurate

Outliers

- ➔ Observation/s different from most others – outliers bias statistics (eg mean) and their standard errors and confidence intervals

Additivity and Linearity

- Assumption means that the outcome variable is linearly related to any predictors
- If you have several predictors then their combined effect is best described by adding their effects together
- This assumption is the most important because if it's not true – then even if other assumptions are met your model is invalid
- Invalid because you have described it incorrectly

Normally Distributed

Normal distribution is related to:

- **Parameter estimates –** mean is a parameter and extreme scores can bias it – illustrates that estimates of parameters are affected by non-normal distributions – parameter estimates differ in how much they're biased in a non-normal distribution

- **Confidence intervals –** we use values of the standard normal distribution to compute the confidence interval around a parameter estimate – using values of the standard normal distribution makes sense only if the parameter estimates actually come from one

- **Null hypothesis significance testing:** to test a hypo about a model we assume that the parameter estimates have a normal distribution because the test stats have distributions related to the normal distribution – if our parameter estimate is normally distributed then these test stats and p-values will be accurate

- **Errors –** can calculate the error for each case of data – if residuals are normally distributed in the population then using the method of least squares to estimate the parameters will produce better estimates than other methods

Normality

Assumption of normality means different things in different contexts

1) For confidence intervals around a parameter estimate to be accurate – estimate must come from a normal distribution
2) For significance tests of models to be accurate – the *sampling distribution* of what's being tested must be normal (sampling distribution of means – differences between means – must be normal)
3) For the estimates of the parameters that define a model to be optimal – the residuals (error) must be normally distributed

Central Limit Theorem (CLT)

The central limit theorem means that *there are a variety of situations in which we can assume normality regardless of the shape of our sample data*

1. **For confidence intervals around a parameter estimate to be accurate – that estimate must come from a normal distribution (ND)**
   - The central limit theorem tells us that in large samples the estimate will have come from a ND
   - If you want to compute confidence intervals – don't need to worry about assumption of normality if sample is large enough

2. **For significance tests of models to be accurate, the sampling distribution of what's being tested must be normal**
   - CLT tells us that in large samples this will be true no matter the shape of the population
   - So shape of data shouldn't affect significance tests provided sample is large enough

3. **For the estimates of model parameters to be optimal the residuals in the population must be normally distributed**
   - Method of least squares will always give you an estimate of the model parameters that minimises error
   - So you don't need to assume normality of anything to fit a linear model and estimate the parameters that define it

Homoscedasticity/Homogeneity of Variance

- Assumption means that each of these samples comes from populations with the same variance
- In correlational designs – assumption means that the variance of the outcome variable should be stable at all levels of the predictor variable

Homogeneity of Variance: assumption that the variance of one variable is stable (rel. similar) at all levels of another variable

Homoscedasticity: assumption in regression analysis that the residuals at each level of the predictor variable/s have similar variances

- At each point along any predictor variable, the spread of residuals should be constant

Heterogeneity of Variance: means that the variance of one variable varies (is different) across levels of another variable

Heteroscedasticity: occurs when the residuals at each level of the predictor variable/s have unequal variances

- At each point along any predictor variable – the spread of residuals is different

- ➔ Method of least squares will produce unbiased estimates of parameters even when homogeneity of variance can't be assumed
- ➔ Better estimates can be achieved using different methods – by using weighted least squares
- ➔ **Weighted least squares** – each case is weighted by a function of its variance

- Unequal variances/heteroscedasticity creates a bias + inconsistency in the estimate of the standard error ass. with the parameter estimates in your model
- Then your confidence intervals + significance tests for the parameter estimates will we biased (because they're computed using the standard error)
- Confidence intervals can be inaccurate when homogeneity of variance/homoscedasticity cannot be assumed

Independence

- Assumption means that the errors in your model are not related to each other
- Equation used to estimate the standard error is valid only if observations are independent
- If you violate assumption of independence – confidence intervals + significance tests will be invalid
- If using method of least squares – model parameter estimates will still be valid but not optimal

Spotting Bias

- When isolated outliers are easier to spot using graphs like histograms + boxplots
- Can use Z-scores to spot outliers

Spotting Normality

P-P Plot: probability-probability plot, a graph plotting the cumulative probability of a variable against the cumulative probability of a particular distribution (a normal distribution)

- Like a Q-Q plot its values fall on the diagonal of the plot then the variable shares the same distribution as the one specified
- Deviations from the diagonal show deviations from the distributions of interest

- P-P plots used to check normality
- Data are ranked and sorted then for each rank the corresponding z-score is calculated to create an expected value that the score should have in a normal distribution
- Then the score is converted to a z-score
- Actual z-score is plotted against the expected z-core
- If the data are normally distributed – the actual z-score will be the same as the expected z-score

Kolmogorov-Smirnov Test and Shapiro-Wilk Test – both compare scores in the sample to a normally distributed set of scores with the same mean and standard deviation

- If the test is non-significant ($p > .05$) it means that the distribution of the sample is not significantly different from a normal distribution (it's normal)
- If the test is significant – then the distribution is significantly different from the ND (It's non-normal)

Skewness and Kurtosis

- To check that the distribution of scores is normal – need to look at skewness + kurtosis
- Positive values of skewness indicate too many low scores – negative values indicate a build-up of high scores
- Positive values of kurtosis indicate a pointy and heavy-tailed distribution – negative values indicate a flat and light-tailed distribution
- The further the value is from zero – the more likely the data is not normally distributed
- You can convert these scores to z-scores by dividing by their standard error – if the resulting score is greater than 1.96 then it is significant ($p < .05$)
- Significance tests of skew + kurtosis should not be used in large samples – because they are likely to be significant even when skew + kurtosis are not too different from normal

Q-Q Plot: similar to P-P plot but instead plots the quantiles of the data instead of every individual score in the data

- If values fall on the diagonal of the plot – then the variable shares the same distribution as the one specified
- Deviations from the diagonal show deviations from the distribution of interest
- Kurtosis is shown up by the dots sagging above or below the line
- Skew is shown up by the dots snaking around the line in an 'S' shape

Reducing Bias – TWAT

1. **Trim the data –** delete a certain amount of scores from the extremes
2. **Winsorizing –** substitute outliers with the highest value that isn't an outlier
3. **Analyse with robust methods –** involves a technique called bootstrapping
4. **Transform the data –** involves applying a mathematical function to scores to correct any problems with them

Trimming the Data

- Trimming the data means deleting some scores from the extremes
- Simplest form = deleting the data from the person who contributed the outlier
- Should be done only if you have good reason to believe that this case is not from the sample population
- Often – trimming involves removing extreme scores using 1 of 2 rules
  1. A percentage-based rule
  2. A standard deviation based rule

The mean in a trimmed sample = the trimmed mean

Trimmed Mean: statistic used in robust tests, mean calculated after a certain percentage of the distribution has been removed at the extremes

- The mean depends on a symmetrical distribution to be accurate
- But a trimmed mean produces accurate results even when the distribution is not symmetrical
- There are more complex examples of robust methods – eg bootstrapping

M-Estimator: robust measure of location

- Example is the median
- In some cases it is a measure of location computed after outliers have been removed
- Unlike trimmed mean – the amount of trimming used to remove outliers is determined empirically

Winsorizing

- Winsorizing involves replacing outliers with the next highest score that is *not* an outlier
- Done manually in SOAA

Robust Methods

- Best option with poor data is to use a test that is robust to violations of assumptions and outliers
- First set of tests are ones that don't rely on the assumption of normally distributed data – non-parametric tests
- But non-parametric tests are only designed for specific situations

1. **Trimmed Mean and M-Estimators**
2. **Bootstrap**

Bootstrap: technique from which the sampling distribution of a statistic is estimated by taking repeated samples (with replacement) from the data set

- Treats the data as a population from which smaller samples are taken
- The statistic of interest (eg mean) is calculated for each sample – sampling distribution estimated from this
- The standard error of the stat is estimated as the standard deviation of the sampling distribution created from the bootstrap samples
- From this – confidence intervals and significance tests can be computed

Lack of normality prevents us from knowing the shape of the sampling distribution unless we have big samples

➔ Bootstrapping gets around this by estimating the properties of the sampling distribution from the sample data
➔ Bootstrapping is based on random samples from the data so the estimates you get will be slightly different every time

Transforming Data

- To combat problems with normality and linearity you can transform the data
- Idea is that you so something to every score to correct for distributional problems, outliers, lack of linearity, unequal variances
- Transforming the data changes the form of the relationships between variables
- But the relative differences between people for a given variable stay the same
- You can just transform the problematic variable – but if you're looking at differences between variables then you need to transform all of those variables

Choosing a Transformation

Data Transformation	Can Correct For
Log Transformation (log X) ▪ Taking the logarithm of a set of numbers squashes the right tail of the distribution. ▪ Good way to reduce positive skew – also useful if you have problems with linearity ▪ Can't get a log value of zero or negative numbers – so if data tend to zero or negative numbers you need to add a constant to all the data before transforming	Positive skew, positive kurtosis, unequal variances, lack of linearity
Square Root Transformation ▪ Taking the square root of large values has more of an effect than taking the square root of small values ▪ Taking the square root of each of your scores will bring any large scores closer to the centre – like the log transformation ▪ Can reduce positive skew, but same problem with negatives and sero	Positive skew, positive kurtosis, unequal variances, lack of linearity
Reciprocal Transformation ▪ Dividing 1 by each score also reduces the impact of large scores ▪ The transformed variable will have a lower limit of 0 ▪ This transformation reverses the scores – scores that were originally large become small (close to 0) ▪ Scores that were originally small become large	Positive skew, positive kurtosis, unequal variances
Reverse Score Transformations ▪ Above transformations can correct negatively skewed data but first you have to reverse the scores ▪ To do this – subtract each score from the highest score obtained ▪ Reverse scores back afterwards	Negative skew

---

Week 3

Chapter 11 – Comparing Several Means: ANOVA (GLM 1)

- ANOVA is a way of comparing the ration of systematic variance to unsystematic variance in an experimental study
- Ratio of these variances is the F-ratio
- F-ratio also a way to assess how well a regression model can predict an outcome compared to the error within that model
- When we test differences between means we are fitting a regression model + using F to see how well it fits the data
- ANOVA can be represented by the multiple regression equation where a number of predictors is one less than the number of categories of the IV
- ANOVA tells us that using group means to predict scores is better than using the overall mean – in other words the groups are significantly different

Logic of the F-Ratio

- F-ratio tests the overall fit of a regression model to a set of observed data
- It is the ratio of how good the model is compared to how bad it is (error)
- If the group means are the same – ability to predict the observed data will be poor (F will be small)
- If the means differ we will be able to better discriminate b/w different groups (F will be large)

Grand Mean: mean of an entire set of observations

Logic of ANOVA

➔ The simplest model we can fit to a set of data is the grand mean (mean of the outcome variable) – this basic model represents 'no effect' or 'no relationship' between the predictor variable + the outcome

➔ Can fit a different model to the data that represents hypotheses – if this model fits the data well it must be better than using the grand mean

➔ The intercept and one or more parameters (*b)* describe the model

➔ The parameters determine the shape of the model that we have fitted – the bigger the coefficients, the greater the deviation between the model and the grand mean

➔ The parameters (b) represent the differences between group means – the bigger the differences between group means, the greater the difference between the model + the grand mean

➔ If the differences b/w group means are large enough – resulting model will be a better fit of the data than the grand mean

➔ Can then infer that our model (predicting scores from group means) is better than not using a model (predicting scores from the grand mean) – means groups are significantly different

Total Sum of Squares (SST)

To find the total amount of variation within the data – calculate the difference between each observed data point and the grand mean

Then square these differences and add them together to give total sum of squares (SST)

Grand Variance: the variance within an entire set of observations

Model Sum of Squares (SSM)

The model sum squares tells us how much of the total variation can be explained by the fact that different data points come from different groups

- Calculated by taking the difference between the values predicted by the model and the grand mean
- In ANOVA – the values predicted by the model are the group means
- Model sum of squared error = sum of the squared distances between what the model predicts for each data point
- For each participant – the value predicted by the model is the mean for the group to which the participant belongs
- Model sum of squares requires calculating the differences between each participant's predicted value + the grand mean

How to calculate SSM:

1. Calculate the difference between the mean of each group + the grand mean
2. Square each of these differences
3. Multiply each result by the number of participants within that group (nk)
4. Add the values for each group together

Residual Sum of Squares (SSR)

Residual sum of squares tells us how much of the variation cannot be explained by the model

- This value is the amount of variation caused by extraneous factors – eg individual differences
- Way to calculate SSR is to subtract SSM from SST
- SSR is calculated by looking at the difference between the score obtained by a person + the mean of the group to which the person belongs

Mean Squares (MS)

- Mean squares is the average sum of squares – which is the sum of squares divided by the degrees of freedom
- Divide by degrees of freedom rather than number of parameters used to calculate SS
- Do this because we are trying to extrapolate to a population and so some parameters within that population will be held constant
- MSM represents the average amount of variation explained by the model (systematic variation)

- MSR is the gauge of the average amount of variation explained by extraneous variables (unsystematic variation)

The F-Ratio
- The F-ratio is a measure of the ratio of the variation explained by the model + the variation explained by unsystematic factors
- It's the ratio of how good the model is against how bad it is (how much error there is)
- Calculated by dividing the model mean squares by the residual mean squares

$F = \frac{MSM}{MSR}$

F-ratio is a measure of the ratio of systematic variation to unsystematic variation

- If its value is less than 1 – it must represent a non-significant effect
- If the F-ratio is less than 1 it means that MSR is greater than MSM – means there is more unsystematic than systematic variance

<u>Assumptions of ANOVA</u>

Homogeneity of Variance
- Assumption that the variance of the outcome is steady as the predictor changes
- Assumption can be tested using Levene's test – which tests the null hypo that the variances of the groups are the same
- It's an ANOVA test conducted on the absolute differences between the observed scores + the mean/median of the group from which each score came
- If Levene's test is significant (p less than .05) → variances are significant – means we have to rectify this
- Can adjust F-test to correct problem with Brown-Forsythe F and Welch's F

Brown-Forsythe F: version of the F-ratio designed to be accurate when the assumption of homogeneity of variance is violated

Welch's F: version of the F-ratio designed to be accurate when the assumption of homogeneity of variance is violated

Is ANOVA Robust?
- F controls the Type I error rate well under conditions of skew, kurtosis and non-normality
- Skewed distributions have little effect on the error rate + power for two-tailed tests
- For kurtosis → leptokurtic distributions make the Type I error rate too low (too many null effects are significant) – means the power is too high
- Platykurtic distributions make the Type I error rate too high – means power is too low
- Power of F is relatively unaffected by non-normality
- Suggests that *when group sizes are equal* → F-stat can be robust to violations of normality
- When group sizes are not equal – the accuracy of F is affected by skew + non-normality affects the power of F
- For homogeneity of variance → ANOVA is robust in terms of error rate when sample sizes are equal
- When sample sizes are unequal – ANOVA is not robust to violations of homogeneity of variance
- When homogeneity of variance is broken → Type I error rate is inflated

What to do When Assumptions are Violated

<u>Planned Contrasts</u>

Set of comparisons between group means that are constructed before any data are collected

- These are theory-led comparisons
- Based on the idea of partitioning the variance created by the overall effect of group differences into gradually smaller portions of variance
- Have more power than post hoc tests
- Planned comparisons done when you have specific hypotheses that you want to test
- Post hoc tests done when you have no specific hypos

Choosing Contrasts

Planned comparisons break down the variation due to the experiment into component parts

Rules
1. If we have a control group this is usually because we want to compare it against the other groups
2. Each contrast must compare only two 'chunks' of variation
3. Once a group has been singled out in a contrast it can't be used in another contrast

Defining Contrasts Using Weights

Weights: number by which something (usually a variable) is multiplied
- The weight assigned to a variable determines the influence that variable has within a mathematical equation
- Large weights give the variable a lot of influence

To get SPSS to carry out planned comparisons – need to tell it which groups we would like to compare – can be complex

- When carrying out contrasts we assign values to certain variables in the regression model
- To carry out contrasts we assign certain values to the dummy variables in the regression model
- Resulting coefficients in the regression model represent the comparisons
- The values assigned to the dummy variables are known as weights

Rules

Rule 1: choose sensible comparisons – you only want to compare two chunks of variation – if a group is singled out in comparison that group should be excluded from subsequent contrasts

Rule 2: groups coded with positive weights will be compared against groups coded with negative weights – assign one chunk of variation positive weights and the opposite negative weights

Rule 3: the sum of weights for a comparison should be zero – if you add up the weights the result should be zero

Rule 4: if a group is not involved in a comparison – automatically assign it a weight of zero – if we give a group a weight of zero this eliminates that group from all calculations

Rule 5: for a given contrast, the weights assigned to the group(s) in one chunk should be equal to the number of groups in the opposite chunk of variation

Non-Orthogonal Comparisons

Non-orthogonal contrasts are comparisons that are in some way related

- Best way to get them is to disobey rule 1
- Must be careful interpreting results
- With non-orthogonal contrasts – the comparisons are related and so the resulting tests statistics and p-values will be correlated to some extent
- Should use a more conservative probability level to accept that a given contrast is statistically meaningful