Contents

PSY3062 – RESEARCH METHODS & DESIGN

Week 2 readings

Chapter 4 – Exploring data with graphs

What makes a good graph:

- Show data
- Induce reader to think about data being presented
    - Chartjunk: SPSS's ability to add pointless fluff to graphs (Tufte)
    - Less patterns
    - No 3D
    - No effects
    - Label axes properly
- Avoid distorting data
- Present many numbers with minimum ink
- Make large data sets coherent
- Encourage reader to compare diff pieces of data
- Reveal underlying message of data

Draw graphs in SPSS using Chart Builder

- Gallery: for each type of graph → gallery of possible variants is shown
- Variable list: variables in data editor listed here
    - Can be dragged into drop zones to specify
- Canvas: main area in dialog box where preview of graph is displayed
- Drop zones: designated with blue dotted lines
- 2 ways to draw graph:
    1. Using gallery of predefined graphs (default option)
    2. Building graph on element-by-element basis

Histograms (frequency distributions)

- Simple histogram: use when you just want to see frequencies of scores for single variable

- Stacked histogram: if you have grouping variable → can produce histogram where each bar's split by group
    - Good way to compare relative frequency of scores across groups
- Frequency polygon: displays same data as simple → uses line instead of bars & area below line is shaded
- Population pyramid: like stacked → shows relative frequency of 2 populations' scores
    - Plots variable on vertical axis
    - Frequencies for each population on horizontal

Boxplots

- Median at centre of plot → limits within middle 50% of observations
- Whiskers: top & bottom 24% of scores
- 3 types:

1. Simple boxplot: use when plotting boxplot of single variable but want diff boxplots produced for diff categories in data
2. Clustered boxplot: same as simple → can select second categorical variable to split data
3. 1D boxplot: use for boxplot of single variable → no categorical variable on x-axis

Bar charts

- How they're displayed usually depends on how data was collected → eg. means come from independent cases or same cases
- 8 types:
    1. Simple bar: use when you just want to see means of scores across diff groups of cases
    2. Clustered bar: if you have second grouping variable → bars produced in diff colours for levels
    3. Stacked bar: same as clustered → diff coloured bars on top not next to
    4. Simple 3D: same as clustered → second grouping variable is displayed by additional axis (not diff coloured bars)
       - don't use it → can obscure data
    5. Clustered 3D bar: like clustered bar → can add third categorical variable on extra axis
    6. Stacked 3D bar: same as clustered 3D → but diff coloured bars stacked on top not next to each other
    7. Simple error bar: same as simple bar → but instead of bars mean is represented by dot & line represents precision of estimation
    8. Clustered error bar: same as clustered bar → but mean is displayed as dot with error bar around it

Line charts

- Bar charts with lines instead of bars
- 2 types:
    1. Simple line: use when you just want to see means of scores across diff groups
    2. Multiple line: same as clustered bar chart

Scatterplot

- Graph that plots each person's score on 1 variable against score on another
- Tells us whether there's relationship between variables & what kind of relationship
- 8 types:
  1. Simple scatter: use when you want to plot values of 1 continuous variable against another
  2. Grouped scatter: like simple → but can display points belonging to diff groups in diff colours/symbols
  3. Simple 3D: use to plot values of 1 continuous variable against 2 others
  4. Grouped 3D: use when you want to plot values of 1 continuous variable against 2 others → but differentiating groups with diff coloured dots
  5. Summary point plot: same as bar chart → but line instead of bar
  6. Simple dot plot (density plot): similar to histogram → shows each individual score as dot instead of summary bar representing frequency of scores
     - useful for looking at shape of distribution
  7. Scatterplot matrix: produces grid of scatterplots showing relationships between multiple pairs of variables
  8. Drop-line: produces graph similar to clustered bar chart → but dot representing summary statistic instead of bar & line connecting means of diff groups
     - useful for comparing statistics across diff groups

Chapter 5 – The beast of bias

- Think about bias within 3 contexts:
  - Things that bias parameter estimates (incl. effect sizes)
  - Things that bias standard errors & confidence intervals
  - Things that bias test statistics & p-values

Assumptions

- Additivity & linearity
  - Outcome variable is linearly related to any predictors
- Normality
  - Parameter estimates: mean is parameter → extreme scores can bias it → shows that parameters affected by non-normal distributions
  - Confidence intervals: values of standard normal distribution to compute confidence interval around parameter estimate → using these values only make sense if parameter estimates come from 1
  - Null hypothesis significance testing

→ Sampling Distribution

- Sample distribution: frequency distribution of data in sample
- Sampling distribution: frequency distribution of means of random samples taken from population
- Assessing sampling distribution → central limit theorem:

        - If our N is large, then sampling distribution will be approx. normal even if our sample distribution is non-normal
            - Large N > 30
        - If our N is small, we rely on sample
        - If sample distribution is normal, then sampling distribution will also probably be normal
- Homoscedasticty/homogeneity of variance
    - Spread of scores is same at each level of variable
    - Heterogeneity of variance: some levels of variable variance of sores is diff than at other levels
- Independence

Outliers

- Score very diff from rest of data