Table of contents

1 Introduction

2 Descriptive statistics

3 Inferential statistics
History of SPSS

- Originally developed by graduate students (Nie, Bent & Hull) at Stanford University.
- First version released in 1968.
- Acquired by IBM in 2009.
Datasets

- Data from the General Social Survey as used in Norušis (2006).
Descriptive statistics

- Counting responses.
- Computing descriptive statistics.
- Looking at distributions.
- Counting responses for combinations of variables ("Crosstabulations").
- Plotting data.
- Measuring Association.
Analyzing internet use . . .

GET

    FILE='C:\SPSS files\gssnet.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

FREQUENCIES VARIABLES=usenet netcat

/BARCHART FREQ

/ORDER=ANALYSIS.
Generate summary statistics for age and education ...

GET

    FILE='C:\SPSS files\gssnet.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

DESCRIPTIVES VARIABLES=age educ

/SAVE

/STATISTICS=MEAN STDDEV VARIANCE MIN MAX.
Comparing Groups

Age by internet use . . .

MEANS TABLES=age BY netcat

/CELLS MEAN COUNT STDDEV.

Plotting the means . . .

GRAPH

/BAR(SIMPLE)=MEAN(age) BY netcat.
Looking at distributions

Analyzing marathon completion times . . .

GET

FILE='C:\SPSS files\marathon.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

EXAMINE VARIABLES=hours BY sex

/ID=name

/PLOT BOXPLOT STEMLEAF HISTOGRAM NPLOT

/COMPARE GROUPS

/STATISTICS DESCRIPTIVES EXTREME

/CINTERVAL 95

/MISSING LISTWISE

/NOTOTAL.
Counting responses for combinations of variables

Crosstabulation of library use by degree . . .

GET

   FILE='C:\SPSS files\library.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

CROSSTABS
   /TABLES=libfreq BY degree
   /FORMAT=AVALUE TABLES
   /CELLS=COUNT
   /COUNT ROUND CELL.

Generate a clustered bar chart . . .

GRAPH
   /BAR(GROUPED)=COUNT BY degree BY libfreq.
Scatterplot of life expectancy with birthrate . . .

GET

    FILE='C:\SPSS files\country.sav'.
    DATASET NAME DataSet1 WINDOW=FRONT.

GRAPH

    /SCATTERPLOT(BIVAR)=birthrat WITH lifeexpf BY develop BY country (IDENTIFY)
    /MISSING=LISTWISE.
How strongly related are ratings of judges and ratings of the criminal justice system as a whole?
Treating these variables as nominal, compute the Proportional Reduction in Error (PRE), aka, “Lambda”.

```
GET
   FILE='C:\SPSS files\crimjust.sav'.

CROSSTABS
   /TABLES=judgrate BY cjsrate
   /FORMAT=AVALUE TABLES
   /STATISTICS=LAMBDA
   /CELLS=COUNT
   /COUNT ROUND CELL.
```
How strongly related are ratings of judges and ratings of the criminal justice system as a whole?
Treating these variables as ordinal, compute “Goodman and Kruskal’s gamma”.

GET
   FILE='C:\SPSS files\crimjust.sav'.

CROSSTABS
   /TABLES=judgrate BY cjsrate
   /FORMAT=AVALUE TABLES
   /STATISTICS=GAMMA
   /CELLS=COUNT
   /COUNT ROUND CELL.
How strongly correlated are ratings of judges and the ratings of the criminal justice system as a whole?
Treating these variables as interval/scale ...

GET
FILE='C:\SPSS files\crimjust.sav'.

CROSSTABS
/TABLES=judgrate BY cjsrate
/FORMAT=AVALUE TABLES
/STATISTICS=CORR
/CELLS=COUNT
/COUNT ROUND CELL.
Inferential statistics

- Evaluating results from samples (The Binomial Test).
- Testing a hypothesis about a single mean.
- Testing a hypothesis about two related means.
- Testing a hypothesis about two independent means.
- One-Way Analysis of Variance.
- Two-Way Analysis of Variance.
- Comparing Observed and Expected Counts.
- Nonparametric Tests.
- Linear Regression and Correlation.
- Analyzing residuals.
- Multiple Regression.
The Binomial Test

How unusual is an observed cure rate of 70% if the true cure rate is 50%?

GET
    FILE='C:\SPSS files\simul.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.

NPAR TESTS
    /BINOMIAL (0.50)=binom10
    /MISSING ANALYSIS.

OR

NPTESTS
    /ONESAMPLE TEST (binom10) BINOMIAL(TESTVALUE=0.5)
    /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE.
Testing a hypothesis about a single mean

Do college graduates work longer than the standard 40 hours per week?
Perform a “one-sample T test” ...

GET
   FILE='C:\SPSS files\gssft.sav'.

SELECT IF DEGREE >= 3.
T-TEST
   TESTVAL=40
   MISSING=ANALYSIS
   VARIABLES=HRS1
   CRITERIA=CI(.95).
Testing a hypothesis about two related means

Do endorphin levels change during a run?
Can either do a one-sample T test on the difference with a test value of zero or perform a “paired-sample T test” ...

GET
   FILE='C:\SPSS files\endorph.sav'.

T-TEST PAIRS=before WITH after (PAIRED)
   /CRITERIA=CI(.9500)
   /MISSING=ANALYSIS.
Is time spent watching television different for internet users?
Perform a “two-independent-samples T test”...

GET
   FILE='C:\SPSS files\gss.sav'.

T-TEST GROUPS=usenet(0 1)
   /MISSING=ANALYSIS
   /VARIABLES=tvhours
   /CRITERIA=CI(.95).
One-Way Analysis of Variance

Assumptions for ANOVA:
- Independence
- Normality
- Equality of variance.
Are weekly hours worked different for those of different educational backgrounds (gender)?

Perform a “one-way anova” ...

GET
   FILE='C:\SPSS files\gssft.sav'.

GRAPH
   /ERRORBAR(CI 95)=hrs1 BY degree.

ONEWAY hrs1 BY degree
   /STATISTICS HOMOGENEITY
   /MISSING ANALYSIS
   /POSTHOC=BONFERRONI ALPHA(0.05).
Are weekly hours worked different by degree and gender?
Perform a “two-way anova” ...

GET
   FILE='C:\SPSS files\gssft.sav'.

GRAPH
   /BAR(GROUPED)=MEAN(hrs1) BY degree BY sex.

UNIANOVA hrs1 BY degree sex
   /METHOD=SSTYPE(3)
   /INTERCEPT=INCLUDE
   /PLOT=PROFILE(degree*sex)
   /CRITERIA=ALPHA(0.05)
   /DESIGN=degree sex degree*sex.
Do women behave more politely than men?

Perform a “chi-square test” ...

GET
   FILE='C:\SPSS files\manners.sav'.

CROSSTABS
   /TABLES=sex BY opendoor
   /FORMAT=AVALUE TABLES
   /STATISTICS=CHISQ
   /CELLS=COUNT EXPECTED ROW RESID
   /COUNT ROUND CELL.
Do two measures agree on whether a person used the Internet yesterday?
Perform a “McNemar’s test” …

GET
    FILE='C:\SPSS files\siqss.sav'.

NPAR TESTS
    /MCNEMAR=intdiary WITH intfolup (PAIRED)
    /MISSING ANALYSIS.
Do the reported times spent on the internet for each measure differ?

Perform a “Sign test” ...

GET
    FILE='C:\SPSS files\siqss.sav'.

SELECT if intdiary=1 & intfolup=1.

NPAR TESTS
    /SIGN=netimfol WITH netimdry (PAIRED)
    /MISSING ANALYSIS.

OR

NPTESTS
    /RELATED TEST(netimfol netimdry) SIGN
    /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
    /CRITERIA ALPHA=0.05 CILEVEL=95.
Do the reported times spent on the internet for each measure differ?

Perform a “Wilcoxon test” ...

GET
   FILE='C:\SPSS files\siqss.sav'.

SELECT IF INTDIARY=1 & INTFOLUP=1.

NPAR TESTS
   /WILCOXON=NETIMFOL WITH NETIMDRY (PAIRED)
   /MISSING ANALYSIS.

OR

NPTESTS
   /RELATED TEST(NETIMFOL NETIMDRY) WILCOXON
   /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
   /CRITERIA ALPHA=0.05 CILEVEL=95.
Do people who use the internet spend less time with family?

Perform a “Mann-Whitney test” . . .

GET
   FILE='C:\SPSS files\siqss.sav'.

NPAR TESTS
   /M-W= etafam BY intdiary(0 1)
   /MISSING ANALYSIS.

OR

NPTESTS
   /INDEPENDENT TEST (etafam) GROUP (intdiary) MANN_WHITNEY
   /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
   /CRITERIA ALPHA=0.05 CILEVEL=95.
Nonparametric Tests

Does time spent on the internet mean less time with family?
Perform a “Kruskal-Wallis test” ...

GET
  FILE='C:\SPSS files\siqss.sav'.

select if intdiary=1.

NPAR TESTS
  /K-W=etafam BY donlncat(1 3)
  /MISSING ANALYSIS.

OR

NPTESTS
  /INDEPENDENT TEST (etafam) GROUP (donlncat) KRUSKAL_WALLIS(COMPARE=PAIRWISE)
  /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
  /CRITERIA ALPHA=0.05  CILEVEL=95.
Is internet usage constant throughout the day?
Perform a “Friedman test” …

GET
   FILE='C:\SPSS files\siqss.sav'.

select if intdiary=1.

NPAR TESTS
   /FRIEDMAN=tonline1 tonline2 tonline3 tonline4 tonline5 tonline6
   /MISSING LISTWISE.

OR

NPTESTS
   /RELATED TEST(tonline1 tonline2 tonline3 tonline4 tonline5 tonline6) FRIEDMAN(COMpare=PAIRWISE)
   /MISSING SCOPE=ANALYSIS USERMISSING=EXCLUDE
   /CRITERIA ALPHA=0.05   CILEVEL=95.
Does birth rate explain female life expectancy?

GET
   FILE='C:\SPSS files\cntry15.sav'.

GRAPH
   /SCATTERPLOT(BIVAR)=birthrat WITH lifeexpf BY country (NAME)
   /MISSING=LISTWISE.

REGRESSION
   /DESCRIPTIVES MEAN STDDEV CORR SIG N
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS CI(95) R ANOVA
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT lifeexpf
   /METHOD=ENTER birthrat
   /RESIDUALS ID(country).
“If the assumptions required for a regression analysis are met, the residuals should have the following characteristics” (Norušis 2006, 498):

- They should be approximately normally distributed.
- Their variance should be the same for all values of the independent variable.
- They should show no pattern when plotted against the predicted values.
- Successive residuals should be approximately independent.
To generate the residuals . . .

GET
  FILE='C:\SPSS files\cntry15.sav'.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT lifeexpf
  /METHOD=ENTER birthrat
  /SAVE PRED RESID ZRESID SRESID.
Generate a Q-Q plot to examine the normality of the standardized residuals . . .

EXAMINE VARIABLES=ZRE_1
    /ID=country
    /PLOT BOXPLOT STEMLEAF NPLOT
    /COMPARE GROUPS
    /STATISTICS DESCRIPTIVES
    /CINTERVAL 95
    /MISSING LISTWISE
    /NOTOTAL.
Analyzing residuals - contd.

Plot studentized residuals against predicted values check for constant variance . . .

GRAPH
   /SCATTERPLOT(BIVAR)=PRE_1 WITH SRE_1 BY country (IDENTIFY)
   /MISSING=LISTWISE.
Plot studentized residuals against order of observations to check for independence ... 

GRAPH

/SCATTERPLOT(BIVAR)=sequence WITH SRE_1 BY country (IDENTIFY)
/MISSING=LISTWISE.
Predicting life expectancy . . .

GET
    FILE='C:\SPSS files\country.sav'.

REGRESSION
    /MISSING LISTWISE
    /STATISTICS COEFF OUTS R ANOVA
    /CRITERIA=PIN(.05) POUT(.10)
    /NOORIGIN
    /DEPENDENT lifeexpf
    /METHOD=ENTER urban lnbeds lnradio lndocs lngdp
    /RESIDUALS ID(country).