METHODS OF QUANTITATIVE DATA ANALYIS
MSR Course, 2011-2012

Harry B.G. Ganzeboom

Lecture 1.2: Inferential Statistics

April 3 2012

IS and DS lingo

- Standard error
- Confidence level
- Confidence interval
- Probability level
- Significance (alpha) level
- Statistically significant
- T-test / t-value
- F-test
- Chi-squared test
- Alpha / beta (type I/II) error

- Mean, median, mode
- Standard deviation
- Percentile scores
- Z-scores
- Association, correlation
- Regression
- Variance, explained variance

IS

- IS is about DS.
- IS adds to descriptive statistics (such as (difference in) means, standard deviation, regression- and correlation coefficients) on sample data, how these statistics would be in the population.
- Note that population statistics are fixed (but unknown), the uncertainty is in the sample.

What if there is no (good) sample?

- IS assumes (simple) random sampling.
- If there is no random sample, IS quantities have no litteral interpretation, they are only ‘metaphors’.
  - What could we have said about the population if the observations would have constituted a simple random sample?
- I still find this a very valuable point of view. Others disagree.
- NOTE: If there is no simple random sample, is does not mean that there is no sampling error. You just cannot model it exactly.

Benchmark: simple random sampling (SRS)

- Random: lottery determines inclusion in the sample.
- Simple: there is only one single draw.
- Almost all IS quantities are calculated assuming SRS.
- In practical sampling SRS is hardly ever used.
- Techniques for handling complex random sampling: multi-level analyses, robust estimation, weights.

Random, but not simple

- Systematic sampling with random begin
- Multistage clustered sampling
- Stratified sampling
- PPS sampling
- Evaluation of these sampling schemes requires specialized statistical programs (STATA).

Non-random sampling

- Quota sampling
- Snowball sampling
- Convenience sampling
- Purposive sampling
- IS for these procedures is at best metaphorical

Sampling distributions

- Take any DS quantity: P, M, SD, R, B or whatever.
- Assume a (fixed) population value.
- Draw an infinite number of samples (SRS) of size N and calculate the sample quantity: p, m, sd, r, b or whatever.
- The frequency distribution of all the sample quantities together is called the <u>sampling distribution</u>.

Sampling distributions

- Have a normal (symmetric, bell-shaped) form
  - If population distribution is normal, or
  - If N is large (> 30)
  - Note that normality arises with larger N in the sample, irrespective of the nature of the distribution in the population.
- Have a t-distribution, when:
  - Sample is small / distribution is very non-normal.
- The (expected) mean of the sampling distribution is the population value of the DS quantity of interest.
- The (expected) standard deviation (SD) of the sampling distribution can be derived mathematically. It always contains an element that resembles $1/\sqrt{N}$: The higher N, the smaller the SD of the sampling distribution.

SE

- The SD of the sampling distribution is called the Standard Error (SE) and is often printed in computer outputs.
- SE denotes the variation of a statistic when many samples (of size N) would be drawn from the population.
- Important SE’s: P, M, M1-M2, R, B.
- Most simple SE is that of the pearson correlation: $1/\sqrt{N}$.

SE's in regression

- It is useful to study the formula's for SE's of simple and partial regression coefficients:
- SE(B) = SQRT(SS(Y-^Y)) / SS(X)*(N-2))
- SE(B) = SQRT(SS(Y-^Y)) / SS(X)*(1-Rx^2)*(N-k-1))
- Elements:
  - SS(Y-^Y): residual variation in Y
  - SS(X): Variation of X
  - (1-Rx^2): VIF = explained variance in Xi by remaining Xk
  - (N-k-1): degrees of freedom

Implications

- Sampling variability goes up:
  - With smaller N
  - With more X variables
  - When X-variables a stronger correlated (='collinearity')
  - When X-variable explain less variance in Y.
- These are all fundamental lessons in research design.

Confidence intervals (CI)

- If we know the (normal) shape of the sampling distribution and SE, we can estimate the population value S of the statistic from the sample.
- The best estimate of the statistic S is its value s in the sample.
- The uncertainty in the estimate is formulated as a (95%) confidence interval: $S = s \pm 2*SE$.
- Again: the variation is among the samples, not in the population value. Each sample will give you a different estimate and in the long run 95% of the CI's will contain the population value.

Significance testing

- (Significance) testing is a (binary) test whether a sample statistics could have been produced by a hypothetical population value, called the null-hypothesis (H0).
- H0 usually assumes that a population statistic is 0, but this is not necessary.
- Statistical testing is the decision whether the sample statistic lies inside or outside the confidence interval of the population value specified in H0.

Confidence level 1-α

- Confidence levels are chosen by the researcher. α is the chosen probability that the conclusion (population value is in CI) will be wrong.
- This is called the type I error: rejecting H0 while is it correct.
- If we increase 1-α, then we will less often be wrong, but we will also be less informative.
- 1- α is conventionally very often chosen at 95%. This is arbitrary, but it could never be, say, 50%.

α-errors

- A α-error (or type I error) occurs when the H0 is true, but we reject the H0 (i.e. decide that the sample statistic is significant).
- We make α-errors in α% (say: 5%) of all decisions. This is NOT influenced by the size of the sample.
- The only way to avoid making so much α-errors is by choosing α at a lower level.
- In practice, researchers (should) care very little about this type of error.

Probability values

- Using the sampling distribution of the H0, we can calculate the probability that the sample statistic would arise.
- This probability (p) is calculated by computer programs and printed in the output. The SPSS header for this is “Sig.”.
- If the probability is <u>lower</u> than the significance level α, we reject the H0 and call the sample result “statistically significant”.
- We will be wrong in α% of the case and commit an alpha-error (type I error): reject the H0, while it is correct.
- We know in advance <u>how often</u> we will make an α-error, but NOT <u>when</u> we make it.

T-test

- We can compare the probability to the significance levels, or alternatively calculate:
- Effect / SE = t
- If $t > 2$ or $t < -2$, we call the result statistically significant.
- Exact evaluation of t depends upon degrees of freedom, but this matters only for small samples.
- Most computer programs print both p and t.

One- and two-tailed

- Computer programs usually calculate two tailed probability levels: what is the probability that the sample statistic <u>or its negative counterpart</u> would arise if H0 is true?
- This is so because computer programs cannot know the expected direction of a result.
- However, researchers usually have a one-tailed interest: they have an directional alternative hypothesis in mind.
- One-tailed probabilities are just half the two-tailed ones.
- The decision about one/two tailed is in the mind of the researcher.

β-errors

- A β-error (or type II error) occurs when the H0 is not true (but some alternative hypotheses H1), but we accept ('do not reject') the H0 (i.e. decide that the sample statistic is not significant)
- If we do not have a fixed H1, the size of β cannot be calculated. However, we do know some circumstances in which β becomes smaller / larger.
- The ability to avoid β-errors is called the statistical power of a test *[onderscheidingsvermogen]*. Power = 1- β.
- β-errors are usually of greater concern than α -errors.

When is β smaller?

- β is smaller:
  - With larger sample size N
  - With more extreme H1
  - With <u>higher</u> alpha
  - In one-sided problems
  - With better measurement, also higher level of measurement
  - With stronger designs: matching, repeated observations / panel, paired observations, higher explained variance
  - With better samples (SRS or better).

Problems with significance testing

- Significance testing leads to a binary decision (yes/no), but our research often requires nuances.
- Significance level is arbitrarily set (at, say, 5%).
- It all depends on SRS assumptions.
- Statistical significance is not relevance.
- In the end, we know more about H0 than about H1 – and we know very little about β.