Methodological skills

rMA linguistics, week 5

Tamás Biró
ACLC
University of Amsterdam
`t.s.biro@uva.nl`

Measures of success in computational linguistics

RE: Catia Cucchiarini, Ambra Neri, Helmer Strik 'Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback', presented last week by Elisabetta.

- Goal: find/accept what must be found/accepted, and do not return/reject what must be ignored/rejected.

- $CA =$ correctly accepted (a.k.a. $TP = true\ positives$).
  $CR =$ correctly rejected (a.k.a. $TN = true\ negatives$).
  $FA =$ falsely accepted (a.k.a. $FP = false\ positives$).
  $FR =$ falsely rejected (a.k.a. $FN = false\ negatives$).

Measures of success in computational linguistics

	Found true	Found false
Is true	*true positives*	*false negative*
Is false	*false positives*	*true negative*

$$\begin{aligned}
\text{Scoring Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
\text{Precision} &= \frac{TP}{TP+FP} \\
\text{Recall} &= \frac{TP}{TP+FN} \\
F\text{-score} &= 2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}
\end{aligned}$$

(F-score: harmonic mean of prec. and recall, closer to the lesser of the two values.)

- Tomasello, M., D. Stahl, 'Sampling children's spontaneous speech: how much is enough?' (2004), in: *Journal of child language*, 31, 101-121.

- presented by Caitlin.

- Student projects
- Basic structure of a research paper

(See pdf of week 4.)

Types of the explanatory variables

$\times$ type of the dependent variable

Scale of the dependent variable is	categorical (nominal, ordinal)	quantitative (interval, ratio, log)
Explanatory variable(s) with categorical scale	*crosstabs*	*logistic regression*
Explanatory variable(s) with quantitative scale	*t-test, ANOVA*	*correlation, regression*

Throwing a die

Population: the outcomes of all throws in NL in 2012.

Distribution of the population:

(approximately) uniform distribution. No mode.

Parameters:

- Population size $N =$ "very large".
- Population min $= 1$. Population max $= 6$.

  Population range $= 5$.

- Population median $= 3.5$ (or 3 or 4).

  Population 1st quartile $= 2$. Population 3rd quartile $= 5$.

  Population 1th percentile $= 1$. 95th percentile $= 6$, etc.

  Population interquartile range $= 3$. Semi-IQR $= 1.5$.

- Population mean $\mu = 3.5$.

  Population variance (N): $\sigma_N^2 = \frac{8.75}{3} \approx 2.91667$.

  Population standard deviation (N): $\sigma_N = \sqrt{\frac{8.75}{3}} \approx 1.708$.

Throwing a die

Sample of size $n = 1$. Value obtained: x_1.

Statistics of this sample:

- Mean of the sample: $\overline{x} = x_1$.

  Sample median $= x_1$. Min. $= x_1$. Max. $= x_1$, etc.

- Range, IQR $= 0$.

  Sample variance and standard deviation with n: $s_n = 0$.

  Sample variance and st. dev. with $n - 1$ is not defined.

Throwing a die

Sampling distribution of the mean for samples of size $n = 1$:

Same as the population:
an (approximately) uniform distribution. No mode.

- Min of the sampling distribution $= 1$. Max $= 6$. Range $= 5$.
- Median of the sampling distribution $= 3.5$. Etc.
- Mean of the sampling distribution $= 3.5$. Standard deviation (N) of the sampling distribution: $\sqrt{\frac{8.75}{3}} \approx 1.708$.

Throwing a die

Sample of size $n = 2$. Values obtained: x_1 and x_2.

Statistics of this sample:

- Sample size: $n = 2$.
- Mean of the sample: $\overline{x} = \frac{x_1+x_2}{2}$.

  Sample median $= \frac{x_1+x_2}{2}$.

  Sample min: the lower one of x_1 and x_2.

  Sample max: the larger one of x_1 and x_2.

- Range: $|x_1 - x_2|$.

  Sample variance with n: $s_n^2 = \frac{(x_1-x_2)^2}{4}$

  Sample variance with $n-1$: $s_{n-1}^2 = \frac{(x_1-x_2)^2}{2}$

  Sample standard deviation with n: $s_n = \frac{(x_1-x_2)}{2}$

  Sample standard deviation with $n-1$: $s_{n-1} = \frac{(x_1-x_2)}{\sqrt{2}}$

Throwing a die

Sampling distribution of the mean for samples of size $n = 2$:

1 1	2 1	3 1	4 1	5 1	6 1
1 2	2 2	3 2	4 2	5 2	6 2
1 3	2 3	3 3	4 3	5 3	6 3
1 4	2 4	3 4	4 4	5 4	6 4
1 5	2 5	3 5	4 5	5 5	6 5
1 6	2 6	3 6	4 6	5 6	6 6

Triangular distribution: $P(\overline{x} = 1) = \frac{1}{36}$, $P(\overline{x} = 1.5) = \frac{2}{36}$,
$P(\overline{x} = 2) = \frac{3}{36}$, $P(\overline{x} = 2.5) = \frac{4}{36}$, $P(\overline{x} = 3) = \frac{5}{36}$,
$P(\overline{x} = 3.5) = \frac{6}{36}$, $P(\overline{x} = 4) = \frac{5}{36}$, $P(\overline{x} = 4.5) = \frac{4}{36}$,
$P(\overline{x} = 5) = \frac{3}{36}$, $P(\overline{x} = 5.5) = \frac{2}{36}$, $P(\overline{x} = 6) = \frac{1}{36}$.

Throwing a die

Sampling distribution of the mean for samples of size $n = 2$:

- Min of the sampling distribution $= 1$. Max $= 6$. Range $= 5$.
- Median of the sampling distribution $= 3.5$.
  Mode of the sampling distribution $= 3.5$.
- Mean of the sampling distribution $= 3.5$.

  Variance (N) of the sampling distribution: $\frac{52.5}{36} \approx 1.4583$.
  Standard deviation (N) of the sampling distribution: $\sqrt{\frac{52.5}{36}} \approx 1.208$.

Throwing a die

Sample of size $n = 3$. Values obtained: x_1, x_2 and x_3.

Sample mean: $\overline{x} = \frac{x_1+x_2+x_3}{3}$.

Sampling distribution of the mean has:

- Mean $= 3.5$.
- Standard deviation: even lower.

Excel experiment.

Remember from last week:

Reliability and validity

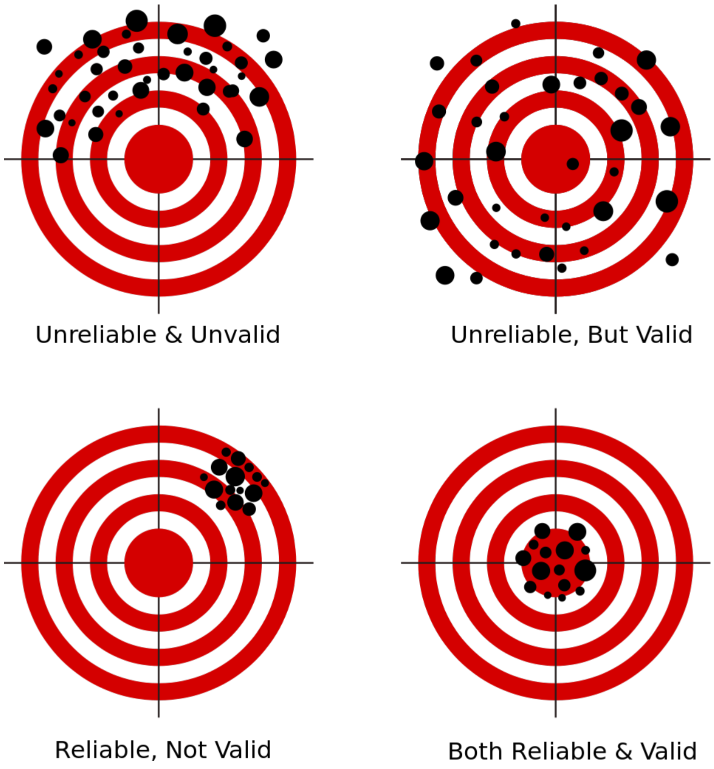
- **Reliability** of the procedure:

  Do we get different results if we repeat the experiment?

- **Validity** of the procedure:

  Does the outcome of the experiment target what we would like to know?

Reliability and validity



http://en.wikipedia.org/wiki/File:Reliability_and_validity.svg

Reliability and validity

- **Reliability** of the procedure:

  Procedure is *reliable* if sampling distribution has small spread — given our procedure.

- **Validity** of the procedure:

  Unbiased statistic: if mean of sampling distribution is targeted parameter — given our procedure.

  Is repeated throwing of dice valid and reliable?

GOOD NEWS!

Sampling distribution

- To reduce bias, achieve validity:

  use random sampling!

- To reduce variability, achieve reliability :

  use larger sample! $\text{Central Limit Theorem!}$

- Population size N (if much larger than sample size n) does not matter.

Central Limit Theorem

- Given population with any distribution.

  Population mean is μ. Population standard deviation is σ.

- Draw a *simple random sample* (SRS) of size n.

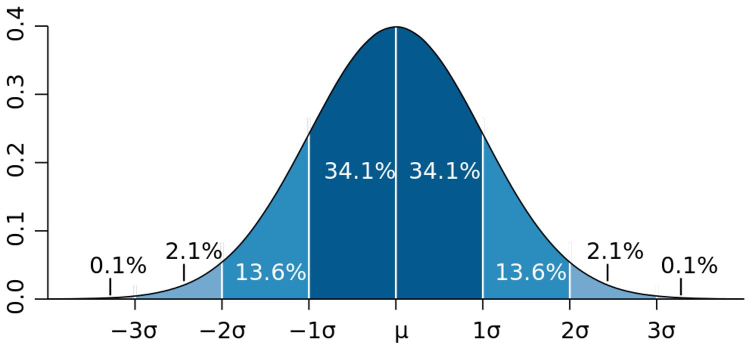
  Calculate sample mean $\bar{x}$.

- **Central Limit Theorem**:

  sampling distribution of $\bar{x}$ is (approx.) Normal: $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Normal (Gaussian) distribution

$$N(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

- $e = 2.7182...$. Mean: μ. Standard deviation: σ.
- Area under curve is 1.

Central Limit Theorem

- Even if we do not know the distribution of some variable in the entire population,
- we know how the empirical mean $\bar{x}$ of any large random sample behaves:
- Sampling distribution of the mean is distributed around the mean μ of the population, and
- follows a Normal distribution of mean μ and st. dev. $\frac{\sigma}{\sqrt{n}}$.

  The larger the sample size n, the narrower the distribution.

Central Limit Theorem

- **Central Limit Theorem** (version 1):

  sampling distribution of $\bar{x}$ is Normal: $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

- This theorem is only approximately true if original population is not Normal, but n is large. (Not true if n is small.)

- **Central Limit Theorem** (version 2):

  The sum (and, hence, the mean) of *independent* random variables X_1, X_2,...,X_n approaches a Normal distribution, as n grows large.

- Therefore: many statistical procedures require:
  - Independence of the cases in the sample.
  - Normality of the population, or
  - close to Normal distribution and larger sample size, or
  - very large sample size (if Normality does not hold).

  Additionally:

"Normality of the population" can be replaced by
"Normality of the sample".

  Testing Normality of the sample: Normal quantile plots!

Next week

Finally getting to

inferences!

To prepare for next week:

Think about paper structure and data types:

- Intro: General problem
  $\rightarrow$ *anecdotal evidence* and *available data*.
- Precise research question:
  Hypothesis to be tested/rejected (H_0 and H_a).
- How to proceed?
  Sample survey (observation) or experiment (intervention)?
- Pilot vs. "the real stuff".

To prepare for next week:

Define research question, in terms of what is your:

- Motivation? General problem? Operationalized research question?
- Population?

  Parameter(s) of the population that interests you?
- Units?

  Sample and sampling method?
- Explanatory variables, response/dependent variables?

  Levels of the variables?

Send 1-page summary (ppt or pdf).
Prepare for 1-minute presentations.

Subsequently:

- What to measure on the sample?
- Statistic to be calculated?
- Best visualization?
- How to draw conclusions in order to answer research question?
- How to draw conclusions in order to contribute to general problem?

See you next week!