Best Practices for Statistics in Your Own Projects

William F. Sensakovic, PhD, DABR, MRSC
Medical Physicist – Florida Hospital
Assoc. Prof. – University of Central Florida
Clin. Asst. Prof. – Florida State University
Adj. Prof. – Adventist University

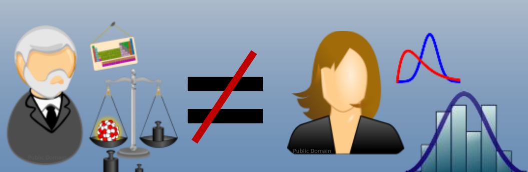
E-mail: wfsensak@gmail.com
Twitter: @wfsensak

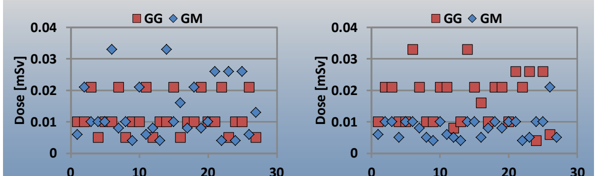
Relevant Conflicts of Interest

No relevant conflicts of interest…

but I would be happy to have someone change that

1. Use a Statistician . . . Stay Involved





P < 0.001

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Dose1 and Dose2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	.074	Retain the null hypothesis.

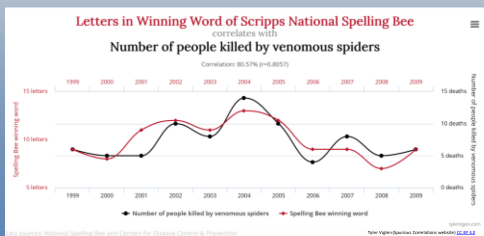
Asymptotic significances are displayed. The significance level is .05.

Sensakovic WF, et al. Image quality and dose differences caused by vendor-specific image processing of neonatal radiographs. Pediatric Radiology. 2016 Oct;46(11):1606-13.

Clinical Significance . . . Not Their Job

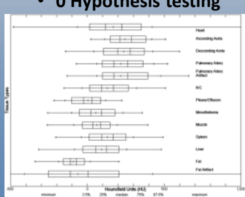
- 52% dose difference
  - Matters? . . . Depends on who you ask
- Could get better quality with same dose using different vendor's product
  - Matters? . . . More interesting

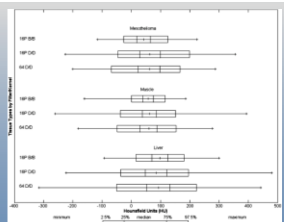
2. Multiple Comparisons



July 31, 2017 William F. Sensakovic, PhD 7

- 28 Patients
- 13 Tissue Types
- 3 Scanners
- 2 Kernels
- 0 Hypothesis testing

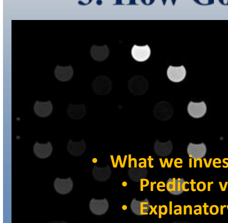


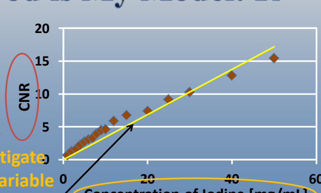


Corson N, Sensakovic WF, et al. Characterization of mesothelioma and tissues present in contrast-enhanced thoracic CT scans. Med Phys. 2011 Feb;38(2):942-7.

July 31, 2017 William F. Sensakovic, PhD 8

3. How Good is My Model: R^2





- What we investigate
  - Predictor variable
  - Explanatory variable
  - Independent variable

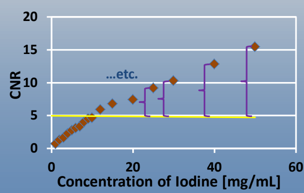
determine the best fit line to the data

ZE Labby, WF Sensakovic, etc. SU-C-134-05 CT Contrast Media: Impact of Scanner Parameters On Enhancement and Detectability Medical Physics, Vol. 40, No. 6, pg 96 2013

July 31, 2017 William F. Sensakovic, PhD

Coefficient of Determination

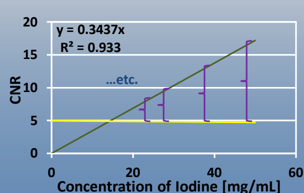
- Total Sum of Squares
  - How much does the data differ from the mean?
  - Variation in the data from all sources



July 31, 2017 William F. Sensakovic, PhD

Coefficient of Determination

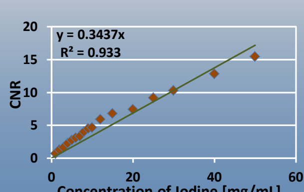
- Regression Sum of Squares
  - How much do model-predicted values differ from the mean?
  - Variation of the model



July 31, 2017 William F. Sensakovic, PhD

Coefficient of Determination

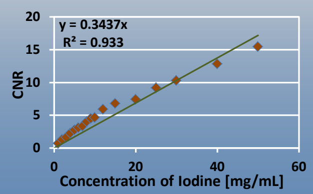
- How much variation does the model explain?
  - $R^2 = \frac{Model\ Variation}{Total\ Data\ Variation}$
    - 0: accounts for no variation
    - 1: perfect fit



July 31, 2017 William F. Sensakovic, PhD

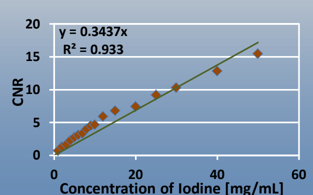
Coefficient of Determination

- How much variation does the model explain?
  - $R^2 = 0.933$ or ~93%
    - So not bad, but does that mean it is the correct model . . . Maybe



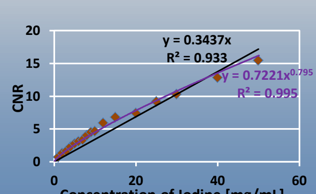
Coefficient of Determination

- What about the other 7%?
  - Random?
    - Measurement error
  - Other predictor?
  - Wrong model?



Coefficient of Determination

- Linear model
  - $R^2 = 0.933$
- Power model
  - $R^2 = 0.995$
- Is power model the correct fit?
  - . . . maybe
- R^2 just tells you how much variation the model accounts for, nothing more



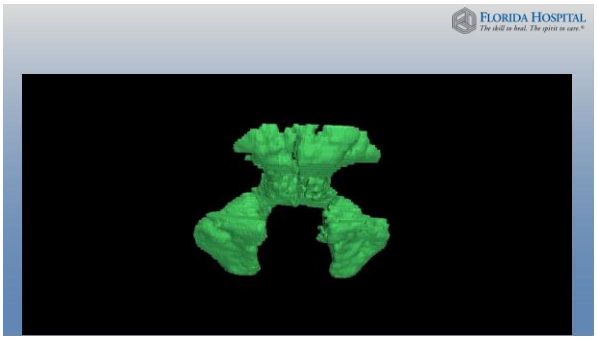
4. Bland-Altman Analysis

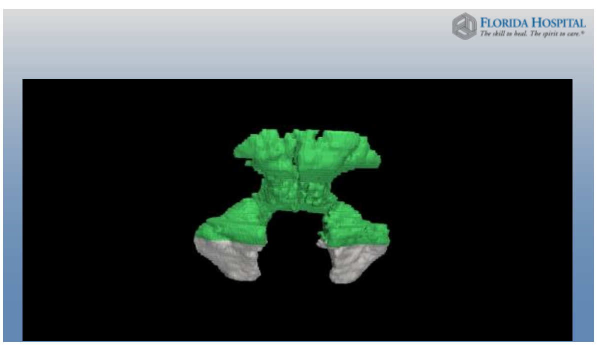
- Pearson Correlation
  - Linear relationship
- Spearman Correlation
  - Non-linear relationship
- Best if dependent and independent variables are different categories
- Not an indication of accuracy!

4. Bland-Altman Analysis

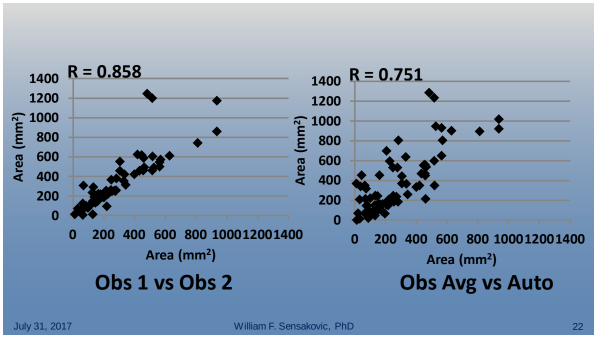
- Best choice when measuring the same quantity with different methods!
  - Bias
  - Variance
  - Trends

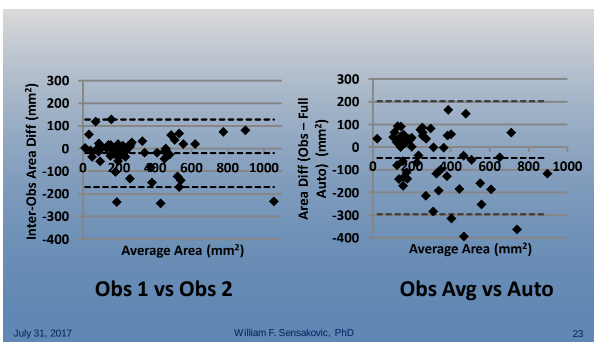
FLORIDA HOSPITAL
The skill to heal. The spirit to care.®

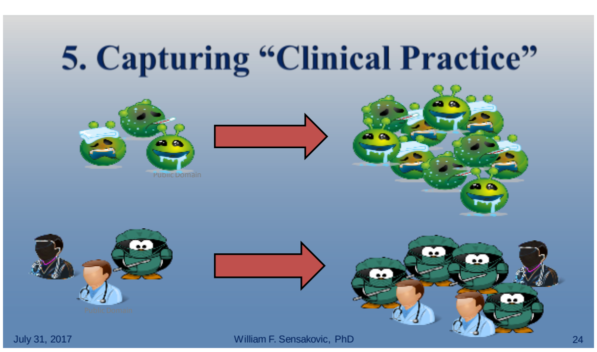

Florida Hospital
The skill to heal. The spirit to care.


Florida Hospital
The skill to heal. The spirit to care.


Florida Hospital
The skill to heal. The spirit to care.


R = 0.858
1400
1200
1000
800
600
400
200
0
Area (mm²)
0 200 400 600 800 1000 1200 1400
Area (mm²)
Obs 1 vs Obs 2
R = 0.751
1400
1200
1000
800
600
400
200
0
Area (mm²)
0 200 400 600 800 1000 1200 1400
Area (mm²)
Obs Avg vs Auto
July 31, 2017
William F. Sensakovic, PhD
22


300
200
100
0
-100
-200
-300
-400
Inter-Obs Area Diff (mm²)
0 200 400 600 800 1000
Average Area (mm²)
Obs 1 vs Obs 2
300
200
100
0
-100
-200
-300
-400
Area Diff (Obs – Full Auto) (mm²)
0 200 400 600 800 1000
Average Area (mm²)
Obs Avg vs Auto
July 31, 2017
William F. Sensakovic, PhD
23


5. Capturing "Clinical Practice"
July 31, 2017
William F. Sensakovic, PhD
24

Case Selection

- Comparisons should be on same cases
  - Sensitivity 25%-100% depending on case selection
    - Nishikawa RM, et al. Effect of case selection on the performance of computer-aided detection schemes. Med Phys 21, 265 (1994)
- The normal case subtlety must be considered to ensure sufficient number of false-positive responses
  - Rockette, et al. Selection of subtle cases for observer-performance studies: The importance of knowing the true diagnosis (1998).
- Study disease prevalence does not need to match disease population prevalence
  - ROC AUC stable between 2%-28% study prevalence, but small increases in observer ratings are seen with low prevalence
    - Gur D, et al. Prevalence effect in a laboratory environment. Radiology 228:10 (2003).
    - Gur D, et al. The Prevalence Effect in a Laboratory Environment: Changing the Confidence Ratings. Acad Radiol 14:49 (2007).

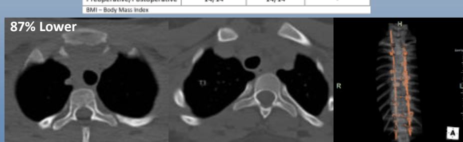
Observer Selection

- Observer Experience
  - Sp 0.9:
    - Se - 0.76 (high volume mammographers)
    - Se - 0.65 (low volume mammographers)
    - Esserman L, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 6;94(5):369 (2002)

Table 1. Patient and Scan Demographics

Demographic	Standard Cohort	Low-Dose Cohort	Significance (p-value)
Patients	17	17	-
Age (years)	14.3±1.9	14.5±2.0	0.79
Weight (kg)	49.8±18.2	51.4±9.4	0.77
Lateral Extent (cm)	27.4±5.4	27.4±4.1	0.99
Scan Length (cm)	43.1±7.7	49.8±10.2	<0.005
BMI (<25/25-35)	10/7	9/8	-
Sex (M/F)	3/14	2/15	-
Preoperative/Postoperative	14/14	14/14	-

BMI – Body Mass Index



WF Sensakovic et al. CT Radiation Dose Reduction in Robot-Assisted Pediatric Spinal Surgery Spine (Phila Pa 1976) 42 (7), E417-E424. 2017

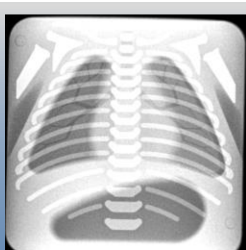
6. Choosing the Correct Test

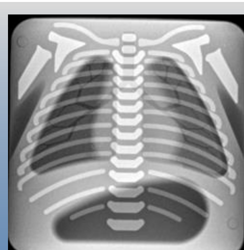
Continuous Dependent Variables	Independent Variables	Independent?	Test
1 normal	1 categorical	Yes	t-test
1 normal	1 categorical	No	paired t-test
1 non-normal	1 categorical	Yes	rank sum
1 non-normal	1 categorical	No	signed rank
1 normal	1 normal continuous	-	Pearson
1 non-normal	1 non-normal continuous	-	Spearman
1 normal	> 1 categorical	Yes	ANOVA
1 non-normal	> 1 categorical	No	Kruskal-Wallis

6. Choosing the Correct Test

- **Parametric**
  - **Non-normal & 15-20 samples per category**
  - **Mean describes the data**
- **Non-parametric**
  - **Deals with outliers better**
  - **Median describes the data**

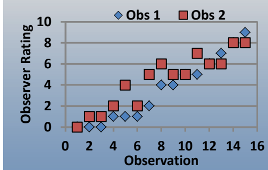
July 31, 2017 William F. Sensakovic, PhD 29





Observers rate "image quality" at different doses

July 31, 2017 William F. Sensakovic, PhD 30



Student's t-test
paired: 0.017
unpaired: 0.41

Wilcoxon
signed rank: 0.026
rank sum: 0.34

Sensakovic WF, et al. Image quality and dose differences caused by vendor-specific image processing of neonatal radiographs. Pediatr Radiol. 2016 Oct;46(11):1606-13.

July 31, 2017 William F. Sensakovic, PhD 31

Cite This Talk/Handout

Sensakovic, WF. MO-F-201-3: Best Practices for Statistics in Your Own Projects. Med Phys. 44 (6), 3714 (2017); 3091-3092 DOI: 10.1002/mp.12304

July 31, 2017 William F. Sensakovic, PhD

Thank You!