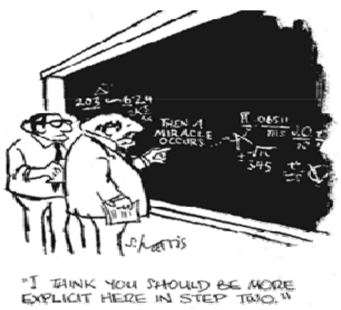
Welcome to

Statistical Secrets Revealed!

Beth Luman, PhD
US Centers for Disease Control and Prevention
African Journal of Laboratory Medicine

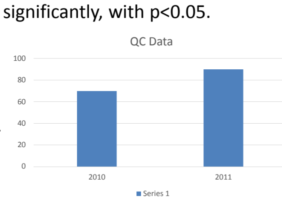


- 12 Examples of common mistakes
- Basic Statistics: n, mean, median, range, IQR, sensitivity, specificity, negative predictive value, positive predictive value
- Frequently Asked Questions
- Presentation of Data

12 Examples

Example 1) A laboratory assesses its QC results. In 2010 it received a score of 70% based on 10 binary criteria. In 2011, it received a score of 90% based on the same 10 criteria. You used a z-test of proportions and found that the score improved significantly, with $p<0.05$.

- **Common mistake**: One laboratory is not a sample
- **What makes it a mistake**: Statistical tests require estimates of variability to calculate p-values. In this example, n=1. Also, z-test assumes simple random sampling, independent samples, and a large underlying population. You don't have any of those.
- **What to consider:** Just report the results

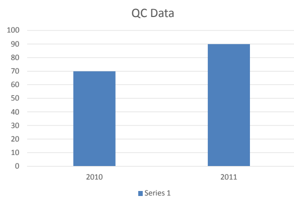


Tip…

- Statistical testing is not always necessary. In this case, the score changed dramatically – it's up to the laboratory to decide if that improvement is sufficient.

By the way…

- Did the laboratory improve by 20% or 20 percentage points?
  - Percent is a relative change (relative to where it started).
  - Percentage points is absolute change.
  - An increase from 10% to 20% is a 100% improvement (the score doubled), but also an improvement of 10 percentage points



Example 2) 5 laboratories participate in the SLMTA program. Baseline and exit audits are completed, with scores for each of 12 QSE's. A total of 5x12x2=120 data points are entered into a statistical package to see if scores at exit are higher than those at baseline.

- **Common mistake**: Incorrect Sample Size
- **What makes it a mistake**: Correlated data, repeat measures
- **What to consider:** Your sample size is really 5, not 120!

Tips…

- Statistical packages will give you an answer to the question you asked, which is not necessarily the correct question
- Parametric tests rely on large sample sizes and assumption of normality. If you don't have this, you have to use nonparametric tests.

Example 3) Measles vaccination coverage in district A is 52%. Coverage in district B is 54%. P=.001.

- **Common mistake**: Clinical vs Statistical Differences
- **What makes it a mistake**: While the difference may be statistically significant, it is not clinically relevant. Either way, vaccination coverage is way too low to prevent outbreaks, and well below the 80% GAVI goals.
- **What to consider:** Statistics can only tell you so much. You also need to think about what the results mean.

Example 4) Water treatment practices increased from 30.362% (s.e. 3.613) before the cholera outbreak to 73.923% (s.e. 2.946) after the outbreak.

- **Common mistake**: Implied precision
- **What makes it a mistake**: Your estimate is only precise to the whole number at best. Giving results to 3 decimal places implies that you are more sure about your result than you really are.
- **What to consider:** Limit results to whole numbers or tenths, unless warranted.

ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

By the way…

- Your computer may spit out more numbers than you need to present!

  P=0.02468367289384762820 9
- When rounding, round 5 and above up, less than 5 downward.
- What would you round this to?

  P=0.02

ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

Example 5) Patients with diabetes were more likely to have "mobility disabilities", such as inability to walk up a flight of stairs, than those without diabetes.

- **Common mistake**: Confounding
- **What makes it a mistake**: Obesity causes both diabetes and mobility disability. The results reported above suggest that diabetes somehow prevents a person from walking up stairs, when it may be that they are both caused by obesity.
- **What to consider:** What are some other potential causes for the results you found?

By the way…

- Randomized controlled studies can help avoid confounding.
- But… you can't always (ethically) do this

Example 6) The average weight of men in the study was 198 pounds.

- **Common mistake**: Relying on point estimates only
- **What makes it a mistake**: Point estimates ignore the uncertainty
- **What to consider:** Confidence intervals or standard errors will help describe the level of certainty of the point estimate. A 95% CI of (108, 288) has a greatly different interpretation than (195, 201).

Example 7) My confidence intervals didn't overlap. Are the 2 groups different?

- **Common mistake**: Hypothesis testing using confidence intervals
- **What makes it a mistake**: If confidence intervals do not overlap, the 2 estimates will be significantly different. If they do overlap, they may be different.

  Why? Statistical testing uses the square root of the sum of squared standard errors. Confidence intervals use the sum of the standard errors. The square root of the sum of the squares of 2 numbers will always be less than the sum of the 2 numbers.
- **What to consider:**
  - Calculate the CI of the difference between the 2 groups. If it doesn't include 0 then the groups are significantly different.
  - Perform an appropriate hypothesis test

By the way…

- What does the 95% Confidence Interval mean?
- There is a 95% chance that the true population mean is within the CI of your sample (pretty much)
- On repeated samples from a population, the 95% CI of your samples will contain the true mean value 95% of the time (best answer)

Example 8) We tested viral load on 10 patients at 6 weeks and 10 weeks, and compared using a t-test

- **Common mis** … pendent
- **What makes** … one group of patients at … ect variability.
- **What to cons** …
  - Use method … t-test.

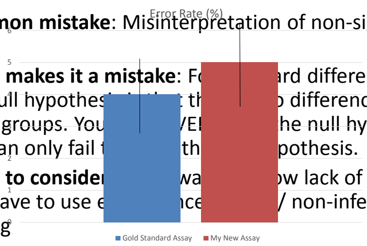
	Viral Load		
Patient	6 weeks	10 weeks	Difference
1	1235	1124	-111
2	54	34	-20
3	3478	3077	-401
4	256	102	-154
5	63589	63189	-400
6	84	134	50
7	1198	1068	-130
8	57	17	-40
9	1668	1638	-30
10	576	376	-200
Mean	7219.5	7075.9	-143.6
Two-Sample T-test	0.987231		
Paired T-test	0.016161		

By the way…

- If the sample size is small (<~30), use non-parametric tests (student's t-test, Wilcoxin matched pair, Fisher's exact, etc.)

Example 9) The p-value comparing our new assay to the gold standard was >0.05. Therefore, our new assay is just as good as the old one.

- **Common mistake**: Misinterpretation of non-significant p-value
- **What makes it a mistake**: Fo[illegible]ard difference testing, the null hypothes[illegible] th[illegible] difference between the 2 groups. You [illegible]/E[illegible] the null hypothesis. You can only fail t[illegible] th[illegible]pothesis.
- **What to conside**[illegible] wa[illegible]ow lack of difference, you have to use e[illegible]nce[illegible] / non-inferiority testing



By the way…

- The null hypothesis for equivalence testing is that the difference between 2 groups is at least a set amount, C.
- If you reject the null, you conclude that there is not more than a difference of C between groups.
- You have to decide what difference is too small to be clinically relevant.
- This can sometimes be politically challenging (such as when you want to show equity of service delivery)
- It can also be tricky to analyze and require large sample sizes.

Tip…

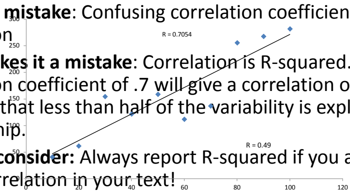
- If it gets too complicated, call in a statistician!

Example 10) A new assay is compared to an old assay to see if they agree when measuring viral load of 100 patients. Correlation was very high, at 0.98 with a p-value <0.05. You claim that the assays agree.

- **Common mistake**: Correlation only measures strength of relationship, not agreement
- **What makes it a mistake**: 100x is perfectly correlated with x. But 100x ≠ x!
- **What to consider:** A Bland-Altman plot can assess agreement. Always check to make sure your results make sense.

Example 11) You do a linear regression to see if an assay measuring CD4 is predictive of an assay that measures viral load. R=.7, with p=0.04, so you conclude that they are fairly well-correlated at .7.

- **Common mistake**: Confusing correlation coefficient with correlation
- **What makes it a mistake**: Correlation is R-squared. So a correlation coefficient of .7 will give a correlation of .7*.7=.49. Meaning that less than half of the variability is explained by the relationship.
- **What to consider:** Always report R-squared if you are talking about correlation in your text!



Example 12) You do a study of prevalence of diabetes in your clinic in Napahala township, and find that 10% of the 100 people tested have diabetes. You conclude that diabetes is an important problem in Africa, and that countries should do routine screening to identify cases.

- **Common mistake**: Extrapolating beyond your study population.
- **What makes it a mistake**: A sample is only representable of the population from which you sampled. Hospital-based samples are especially uncharacteristic of the general population. Plus you can't make a sweeping statement based on one small sample in one place.
- **What to consider:** Don't overstate your results

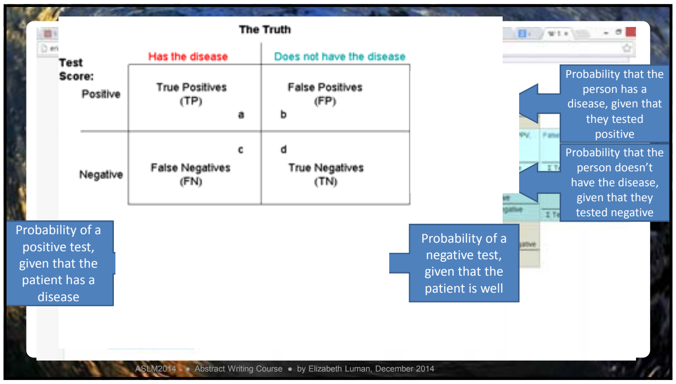
Basic Statistics

Basic Statistics

- **n = sample size**
- **s.e. = standard error (measure of variability)**
- **Mean (sum / n)**
- **Median**
- **Range**
- **Inter Quartile Range**

14	**3**
3	**3**
3	**3**
8	**4**
3	**4**
4	**4**
4	**8**
32	**9**
9	**12**
4	**12**
12	**14**
12	**20**

n=12
sum=108
Mean=9
Median=6
Range=(3,20)
IQR=(3.5,12)



DIY Statistics…

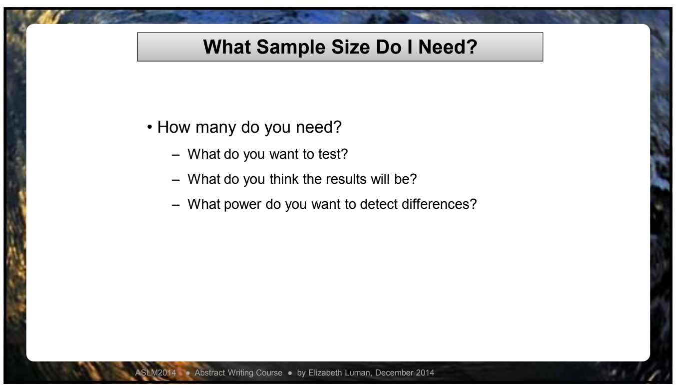
Just because you can plug numbers into a program doesn't change the fact that if you don't know what you're doing, you're almost guaranteed to get meaningless results -- if not dangerously misleading ones. Statistics really is like rocket science; it isn't easy, even to us who have studied it for a long time. Anybody who think it's easy surely lacks a deep enough knowledge to understand why it isn't! If your scientific integrity matters, and statistics is a mystery to you, then you need expert help. Find a statistician in your company or at a nearby university, and talk to her face-to-face if possible. It may well cost money. It's worth it.

R. V. Lenth *http://homepage.stat.uiowa.edu/~rlenth/Power/*

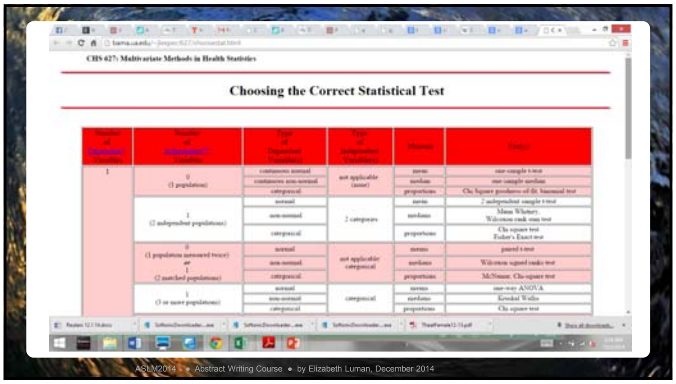
Tips…

- Involve a statistician early – from the concept phase of the study onwards
- Develop data collection and analysis methods together
- Consider power and magnitude of study when determining statistical methods
- Data analysis is *constantly ongoing* throughout the study, manuscript preparation, and writing


Frequently Asked Questions
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


What Sample Size Do I Need?
• How many do you need?
– What do you want to test?
– What do you think the results will be?
– What power do you want to detect differences?
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Raosoft
Sample size calculator
Alternate scenarios
More information
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014



What Does P-value *Really* Mean?

Examples of what P-value does *not* mean…

- P = 0.05 does not mean there is only a 5% chance that the null hypothesis is true.
- P = 0.05 does not mean there is a 5% chance of a Type I error (i.e. false positive).
- P = 0.05 does not mean there is a 95% chance that the results would replicate if the study were repeated.
- P > 0.05 does not mean there is no difference between groups.
- P < 0.05 does not mean you have proved your experimental hypothesis.

P-value means: ***The probability of getting the results you did (or more extreme results) given that the null hypothesis is true.***

http://labstats.net/articles/pvalue.html

Is this an appropriate conclusion?

~~• There was no difference between mentored and non-mentored laboratories.~~
• The difference between mentored and non-mentored laboratories did not reach statistical significance at the p=.05 level.

• ~~Our study proves that mentorship is feasible and effective.~~
• We found that providing mentorship was feasible in our setting, and appeared to contribute to the success of the SLMTA program.

What's wrong with this conclusion?

Bottom line…

The SCIENCE should always come FIRST

- Don't get too caught up in statistical significance
- Use inputs that are ***clinically relevant*** based on consideration of the underlying ***scientific goals***
- A statistical plan should meet the needs of scientific goals, not the other way around

Presentation of Data

Tables

Rule of thumb- Use tables to present data that are detailed and important and to highlight individual values

- Tables are "expensive" in terms of space
- Consider using text instead if data are:
  - Not detailed; use sentences to describe data
  - Not important; all data do not need to be presented -- a summary can be given

Tables Do's and Don'ts

- DO include some measure of statistical precision (p-values or confidence intervals) if appropriate
- DO use appropriate and consistent number of decimal places (usually to the 10ths)
- DO double-check all numbers. Then check them again.
- DON'T use grid lines
- DO present results that you don't like as well as those you do!

Class Exercise

What's wrong with this table…

		2003	2013
	13,000	87.635	93.887
North	5010	85	90
South	4187	50.1	80.3
East	2110	78.35	76.86
West	4992	83.6	89.4

Measles vaccination coverage among children aged 19-35 months for regions of Oahu. National Immunization Survey, 2003 vs. 2013

Region	Population aged 19-35 months	Vaccination Coverage		P-value
		2003	2013	
Total	13,000	87.6	93.9	0.15
North	5,010	85.0	90.0	0.08
South	4,187	50.1	80.3	0.03
East	3,110	78.4	76.9	0.67
West	693	83.6	89.4	0.04

Figures

Rule of thumb- Use figures to:

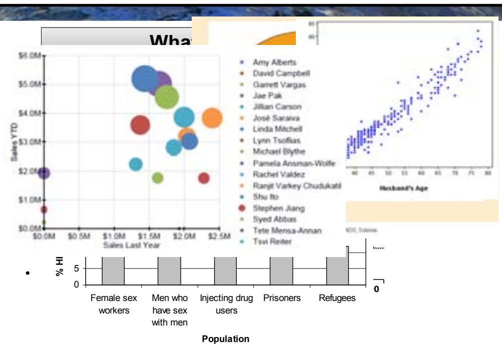
- Show trends or relationships in data (as graphs)
- Present schematic diagrams, images, photographs, and maps

Note: Use color only when necessary. Make sure figures are still readable if printed in black-and-white.

Rules for Figures

1) Make sure all parts of the figure are labeled clearly
2) Independent variable on the x axis, dependent variable on the y axis
3) Figures must stand on their own – use complete titles and footnotes, write out acronyms
4) Don't use 3-d figures for 2-d data
5) Choose colors that will print in black and white
6) Don't use line graph for categorical data
7) What is the scale?
8) Does the graph show a full picture of the data?
9) Percentage graphs should (usually) go to 100%

Wha



Parts of a Graph

Class Exercise

What are the parts of a graph?

- Title
- Legend
- Footnotes
  - Define all acronyms or abbreviations
  - Data Source
  - Caveats
- Y axis, x axis (bar graph, line graph or scatter plot)
- Value labels

More about Graphs…

Independent vs. Dependent Variables

- Independent variable – isn't manipulated. Ex: time, age. The variable that causes change.

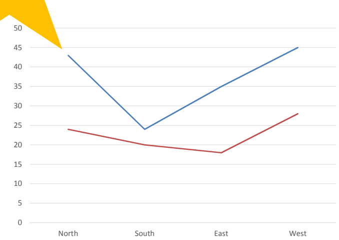
- Dependent variable – the variable that you are trying to measure, to see if it is affected by the independent variable. The variable that is changed.

- Example
  - Test score
  - Time spent studying

ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

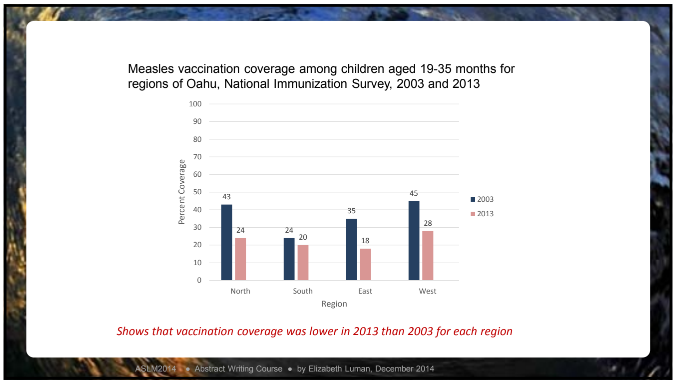
What are some problems with this graph??

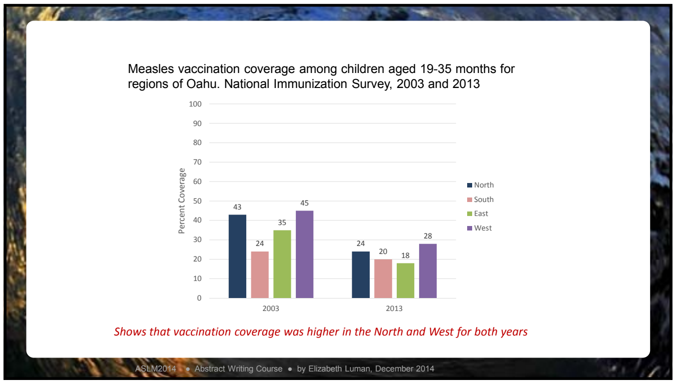
Class Exercise

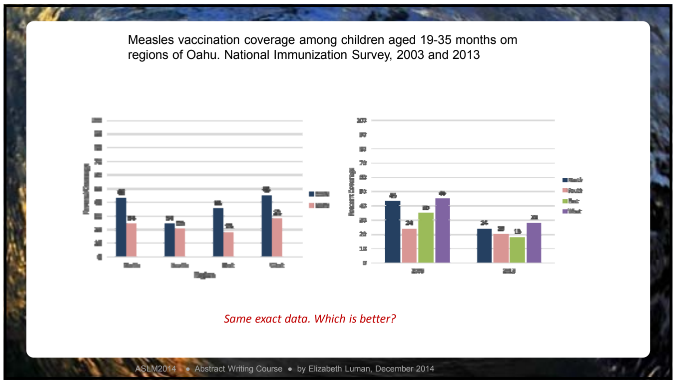


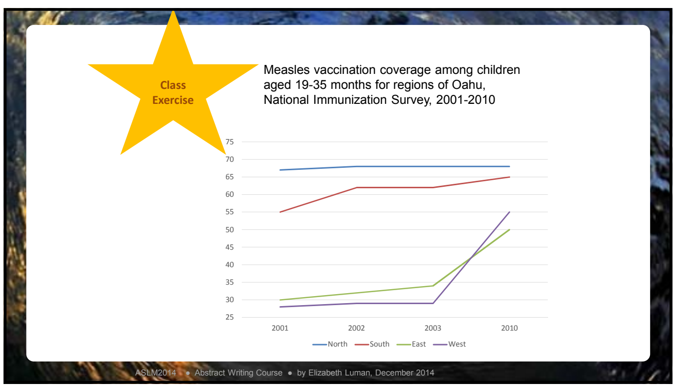
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

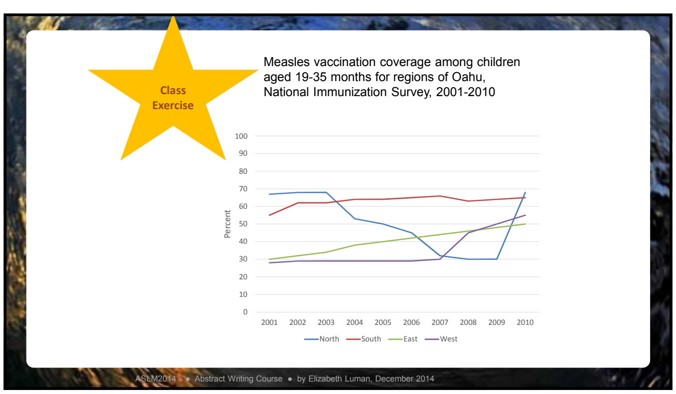

What are some problems with this graph??
Class Exercise
40
35
30
25
20
15
10
5
0
North
South
East
West
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Measles vaccination coverage among children aged 19-35 months for regions of Oahu, National Immunization Survey, 2003 and 2013
Percent Coverage
100
90
80
70
60
50
40
30
20
10
0
43
24
24
20
35
18
45
28
North
South
East
West
Region
2003
2013
Shows that vaccination coverage was lower in 2013 than 2003 for each region
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Measles vaccination coverage among children aged 19-35 months for regions of Oahu. National Immunization Survey, 2003 and 2013
Percent Coverage
100
90
80
70
60
50
40
30
20
10
0
43
24
35
45
24
20
18
28
2003
2013
North
South
East
West
Shows that vaccination coverage was higher in the North and West for both years
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Measles vaccination coverage among children aged 19-35 months om
regions of Oahu. National Immunization Survey, 2003 and 2013
Same exact data. Which is better?
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Class
Exercise
Measles vaccination coverage among children
aged 19-35 months for regions of Oahu,
National Immunization Survey, 2001-2010
75
70
65
60
55
50
45
40
35
30
25
2001
2002
2003
2010
North
South
East
West
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Class
Exercise
Measles vaccination coverage among children
aged 19-35 months for regions of Oahu,
National Immunization Survey, 2001-2010
100
90
80
70
60
50
40
30
20
10
0
Percent
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
North
South
East
West
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

Things to Consider

- What am I trying to say with my data?
- What information is necessary to support my message? NO FLUFF!
- Do I need a table, graph, or figure?
- Have I been consistent, concise, and clear with all supplemental material?

Class Exercise

Example – SLMTA Data Paper

SLMTA Program Improvement – Star Levels

a) 321 laboratories in 45 countries have completed the Strengthening Laboratory Management Toward Accreditation Program (SLMTA)
b) They are audited before (baseline) and after (exit) the program using a checklist. Scores are broken down into star levels, from 0 to 5 stars
c) What is the best way to show the improvement of star levels?

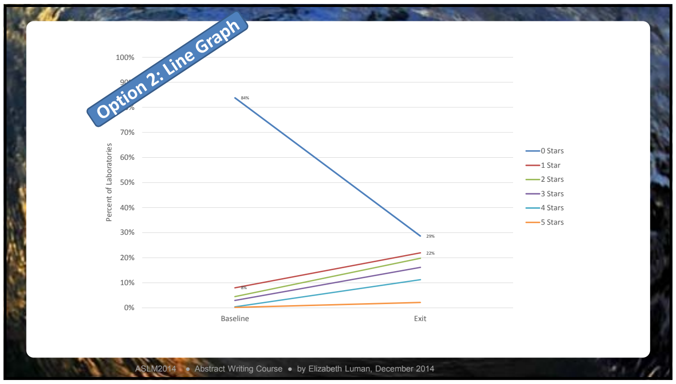
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

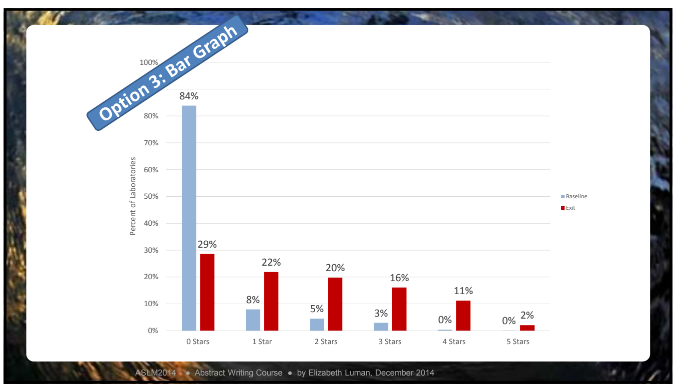
Option 1: Table

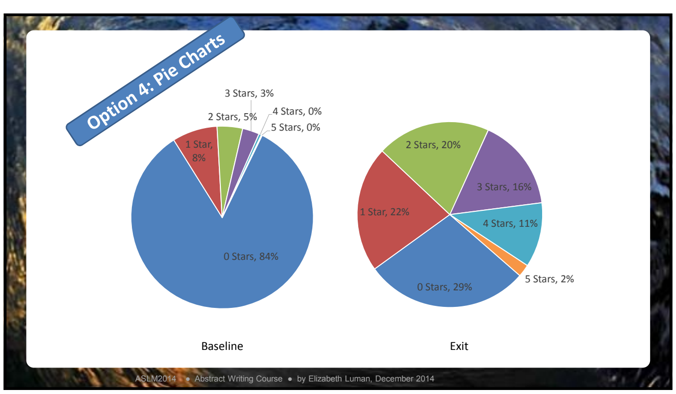
		Exit						
		0	**1**	**2**	**3**	**4**	**5**	**Total**
Baseline	**0**	90	60	52	39	30	3	262
	1	2	21	2	1	1	0	27
	2	1	0	7	5	3	2	18
	3	0	0	2	6	2	1	11
	4	0	0	0	0	1	1	2
	5	0	0	0	1	0	0	1
	Total	93	69	63	52	37	7	321

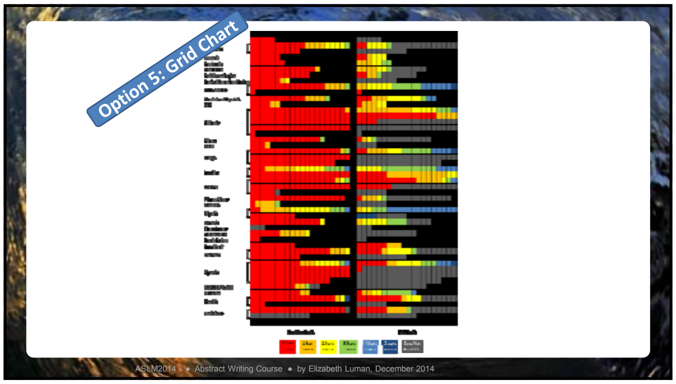
Table 2: Number of laboratories by quality stars at baseline and exit audits, based on the Stepwise Laboratory Quality Improvement Process Towards Accreditation checklist. 2010-2013

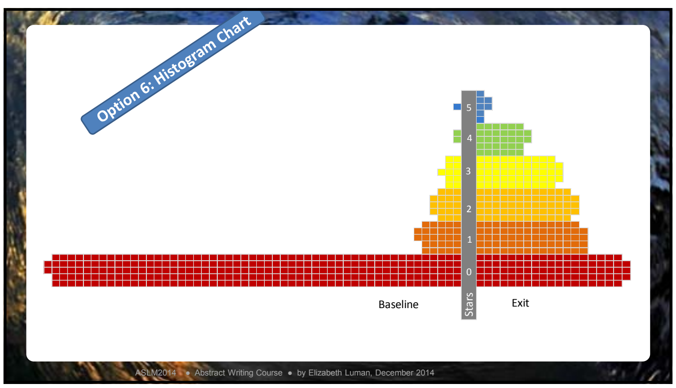
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014

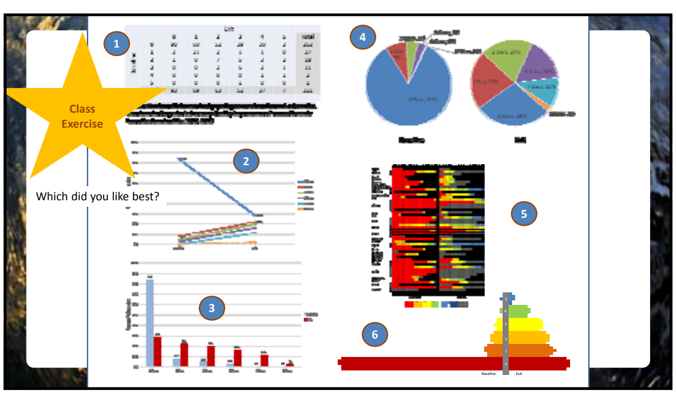

Option 2: Line Graph
100%
70%
60%
50%
40%
30%
20%
10%
0%
Percent of Laboratories
Baseline
Exit
0 Stars
1 Star
2 Stars
3 Stars
4 Stars
5 Stars
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Option 3: Bar Graph
100%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent of Laboratories
84%
29%
8%
22%
5%
20%
3%
16%
0%
11%
0%
2%
0 Stars
1 Star
2 Stars
3 Stars
4 Stars
5 Stars
Baseline
Exit
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Option 4: Pie Charts
3 Stars, 3%
2 Stars, 5%
4 Stars, 0%
5 Stars, 0%
1 Star, 8%
0 Stars, 84%
2 Stars, 20%
3 Stars, 16%
1 Star, 22%
4 Stars, 11%
5 Stars, 2%
0 Stars, 29%
Baseline
Exit
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Option 5: Grid Chart
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Option 6: Histogram Chart
5
4
3
2
1
0
Stars
Baseline
Exit
ASLM2014 • Abstract Writing Course • by Elizabeth Luman, December 2014


Class Exercise
Which did you like best?
1
2
3
4
5
6