Introduction to Statistics in Medicine

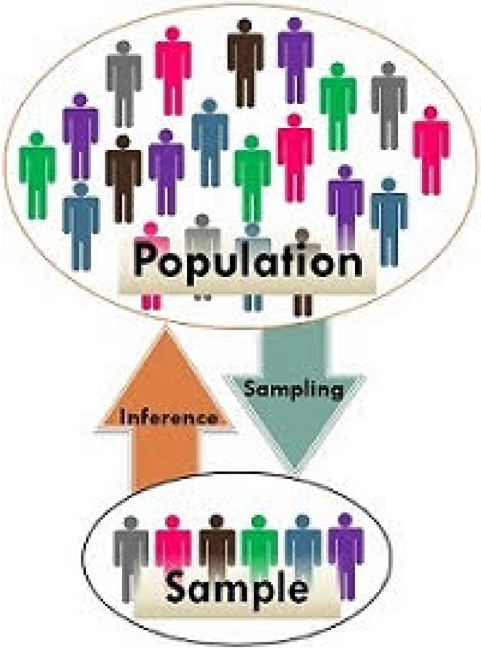
Dr. Maryna Yaskina

March 2019

women & children's health research institute
UNIVERSITY OF ALBERTA
Alberta Health Services
STOLLERY CHILDREN'S HOSPITAL FOUNDATION
LOIS HOLE HOSPITAL FOR WOMEN

Why do we need statistics?

A primary objective of statistics is to summarize the obtained data and make an inference about a population based on a sample from that population.



What statistics will do for us?

1. Measure and control uncertainty
2. Clarify our exact question
3. Identify the variable and the measure of that variable that will answer that question
4. Verify that the planned sample size is adequate
5. Answer the question asked, while limiting the risk of error in decision

What statistics will NOT do for us?

1. Statistics will not make uncertainty disappear
2. Statistics will not provide a credible conclusion from poor data

women & children's health research institute

Components of Medical Study

1. Clearly specify a question to be answered about an explicitly defined population
2. Identify a measurable variable capable of answering the question
3. Perform sample size calculations
4. Obtain observations on the variable from a sample that represents the population
5. Analyze the data with methods that provide an answer to the question

Components of Medical Study

It is very important to:

1. Consult a statistician at the design stage
2. Use EDC system (RedCap!!) for data collection, please do NOT use Excel

Descriptive Statistics

Types of Data

1. Continuous data is measured on a scale.
   Example: temperature, glucose level
2. Discrete data has gap between values, often is a result of counting.
   Example: age in years (or months),
   number of siblings
3. Categorical data are indicators of type or category. Can be ordinal (orderable) or nominal (non-orderable)
   Example: gender (nominal);
   obesity status: normal, overweight,
   obese (ordinal)

women & children's health research institute

Categorical data

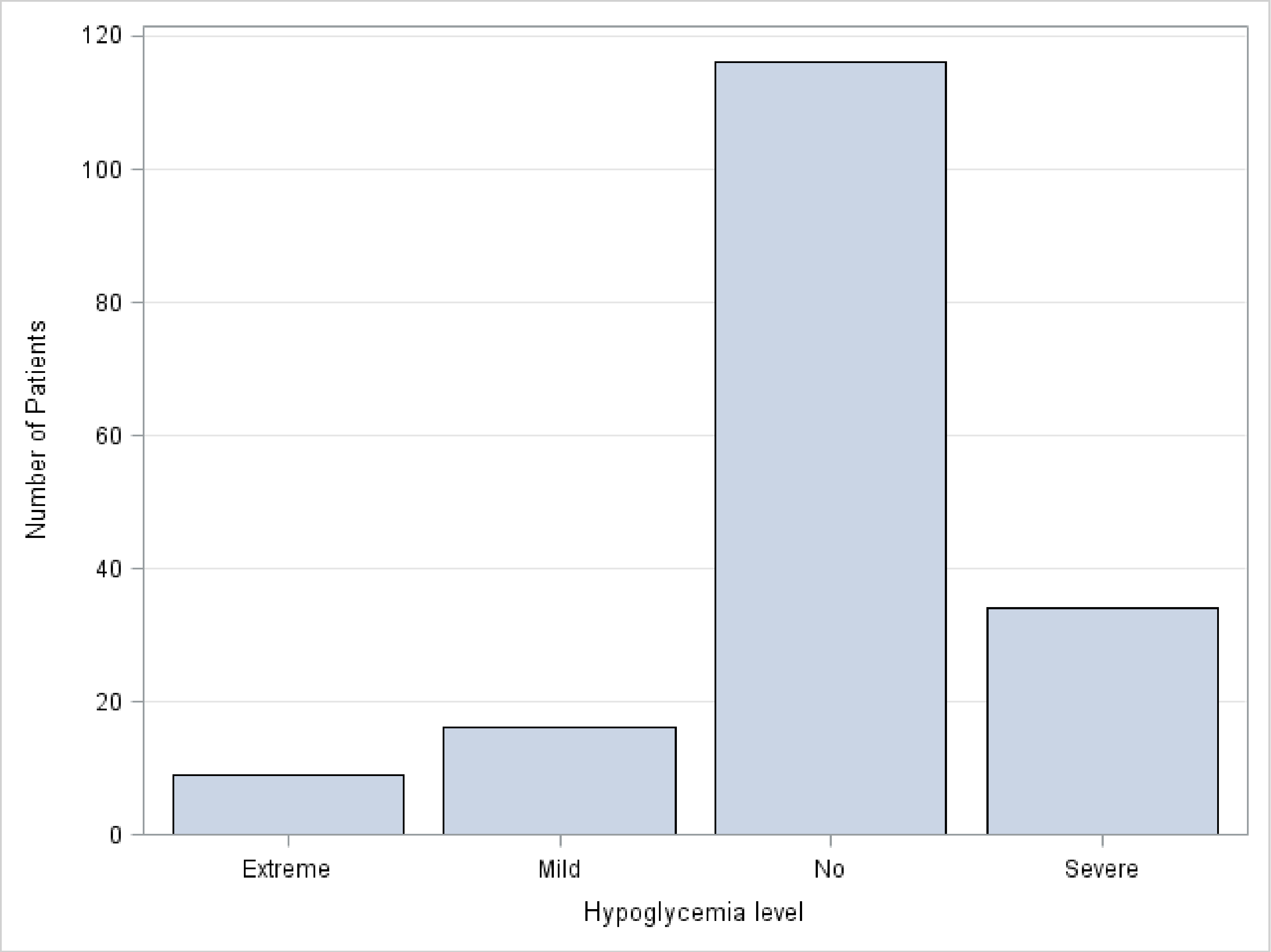
Usually is summarized by absolute (n) and relative frequencies (%)

women & children's health research institute

One-way Table

Hypoglycemia level		
	N	**%**
Extreme hypoglycemia: glucose level (0, 1.1]	9	5.1
Severe hypoglycemia: glucose level (1.1, 2.2]	34	19.4
Mild hypoglycemia: glucose level (2.2, 2.6)	16	9.1
No hypoglycemia: glucose level >= 2.6	116	66.3
Total	**175**	**100**

Bar Chart



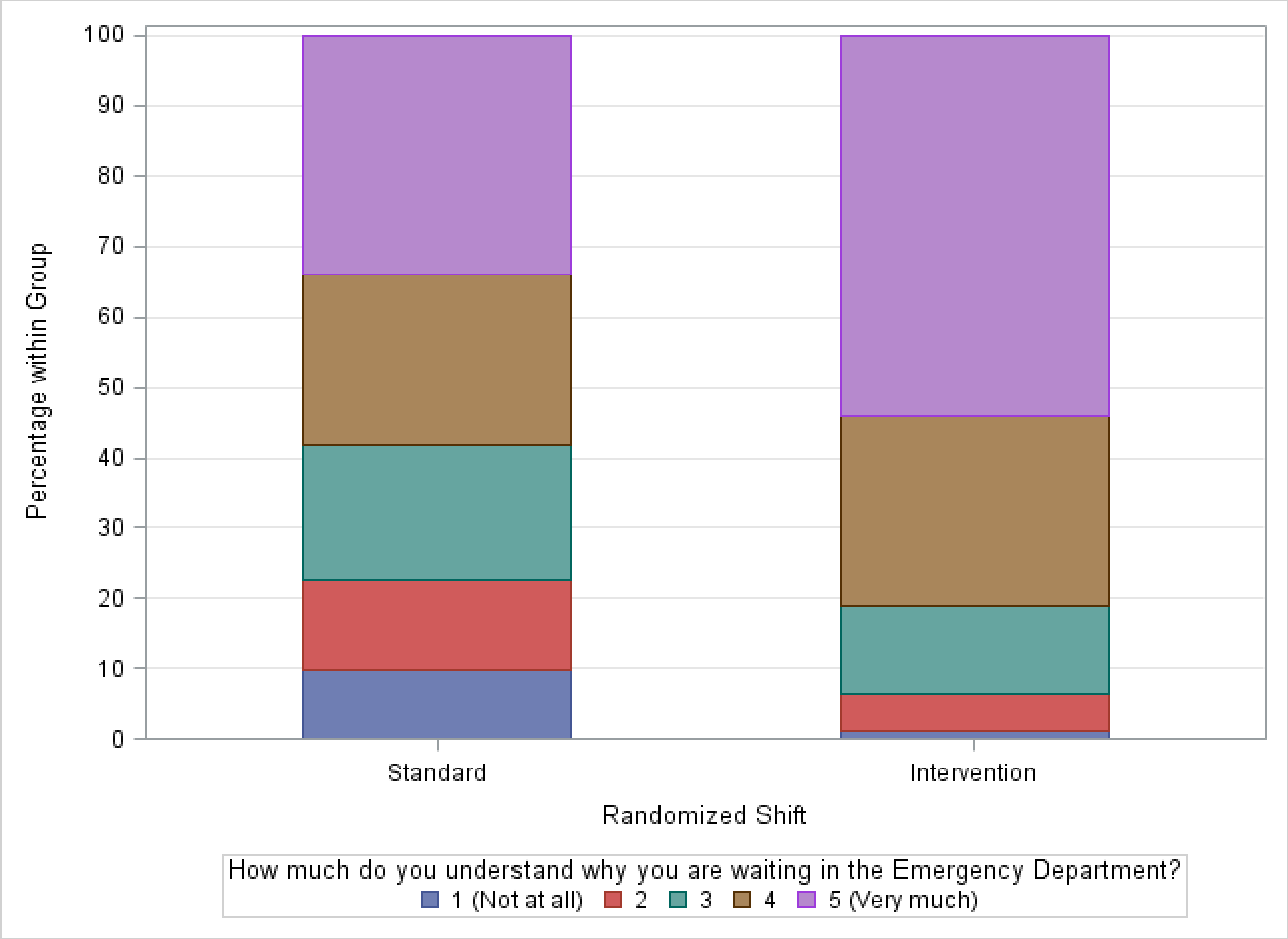
120
100
80
60
40
20
0
Number of Patients
Extreme
Mild
No
Severe
Hypoglycemia level

women & children's health research institute

Two-way Table

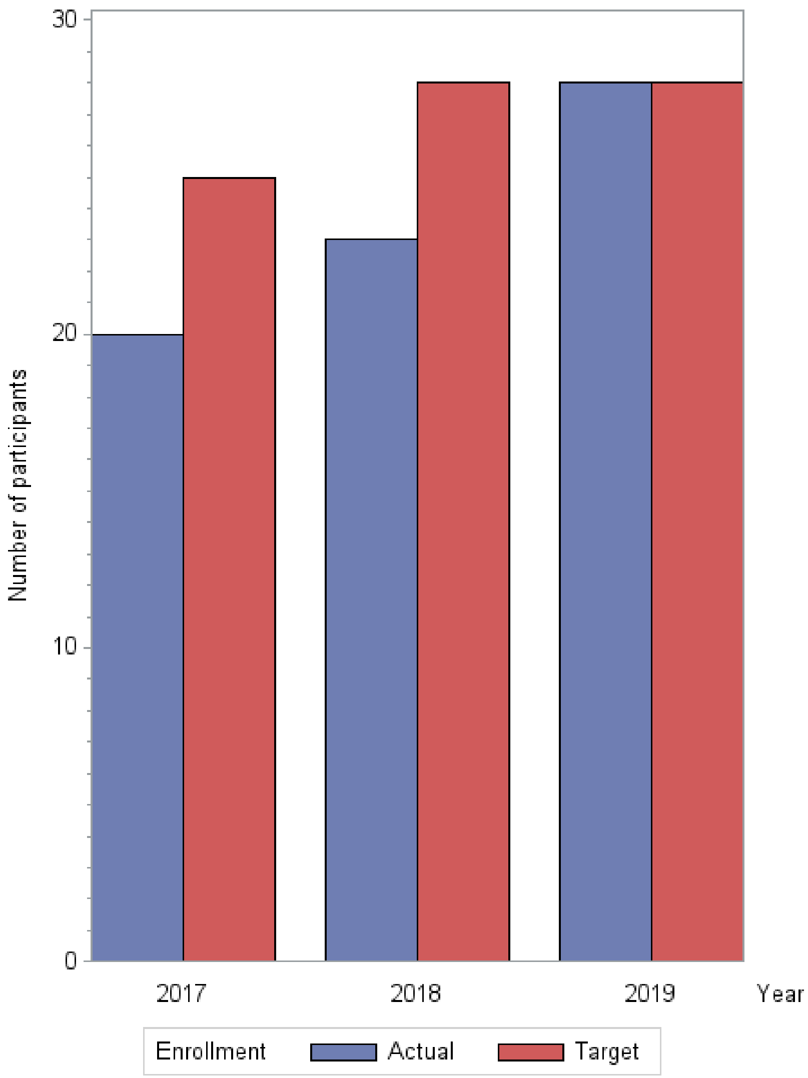
How satisfied are you with the amount of information you got about why you were waiting in this Emergency Department?	Randomized Shift		
	Standard	Intervention	Total
1 (Not at all)	71 20.7%	7 2.5%	**78**
2	68 19.8%	37 13.0%	**105**
3	100 29.2%	75 26.4%	**175**
4	57 16.6%	73 25.7%	**130**
5 (Very Much)	47 13.7%	92 32.4%	**139**
Total	**343**	**284**	**627**

Segmented Bar Chart



women & children's health research institute

Side-by-side Bar Chart



Continuous Data
Measures of centre

Mean is the arithmetic average of all the values in the data set.

It is often used to approximate the central tendency of a set of data but is very susceptible to outliers.

Example: 3, 4, 2, 3, 5

mean = (3+4+2+3+5)/5=3.4

Example: 3, 4, 2, 3, 25

mean = (3+4+2+3+25)/5=6.8

Measures of centre

Median is the middle value of a data set when the set is ordered.

It is not influenced by outliers but does not give indication of the actual values in a data set; it is not an average.

Example: 3, 4, 2, 3, 5

median = 3

(re-order set : 2, 3, 3, 4, 5)

women & children's health research institute

Measures of spread

Standard deviation shows how far away from the mean an individual value lies

Percentile – a value of a sample below which that percent of the sample lies. For example, 50th percentile = median (value below which half of the sample value lies)

Often 25th and 75th percentiles are used (1st and 3rd quartiles)

IQR=75th percentile – 25th percentile (shows middle half of the data)

women & children's health research institute

A confidence interval (CI) is an interval about an estimate which expresses the confidence, or probability, that that interval contains the population statistics being estimated.

95% confidence means that “95% of samples of this size will produce confidence intervals that capture the true population statistics (mean or proportion)”

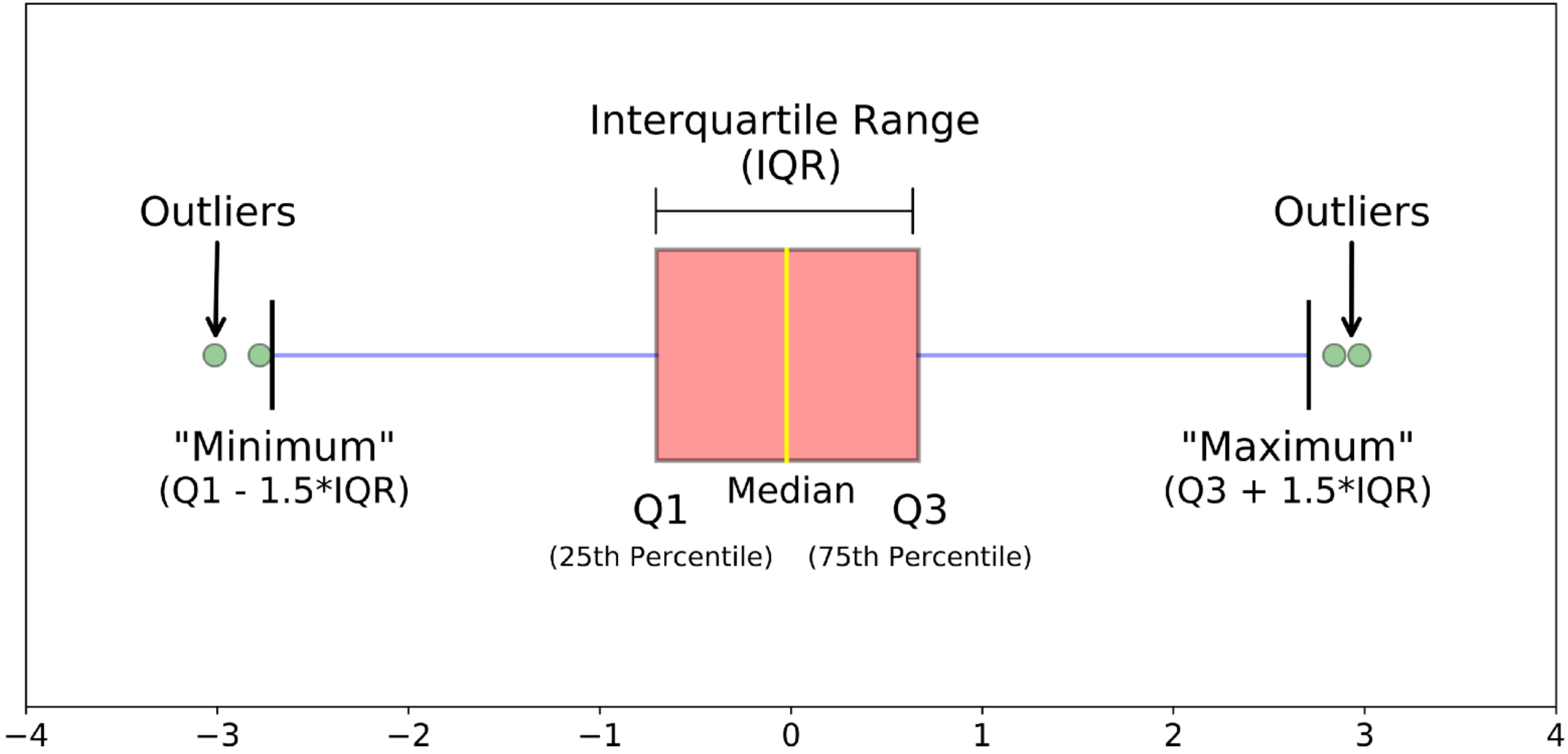
In short “we are 95% confident that the true population statistics (mean or proportion) lies in our interval”

Continuous data - descriptive summary

Variable	*N*	*Mean*	*Std Dev*	*Median*	*25th percentile*	*75th percentile*	*Minimum*	*Maximum*
Glucose level	175	3.20	1.45	3.00	2.30	3.90	0.60	9.90

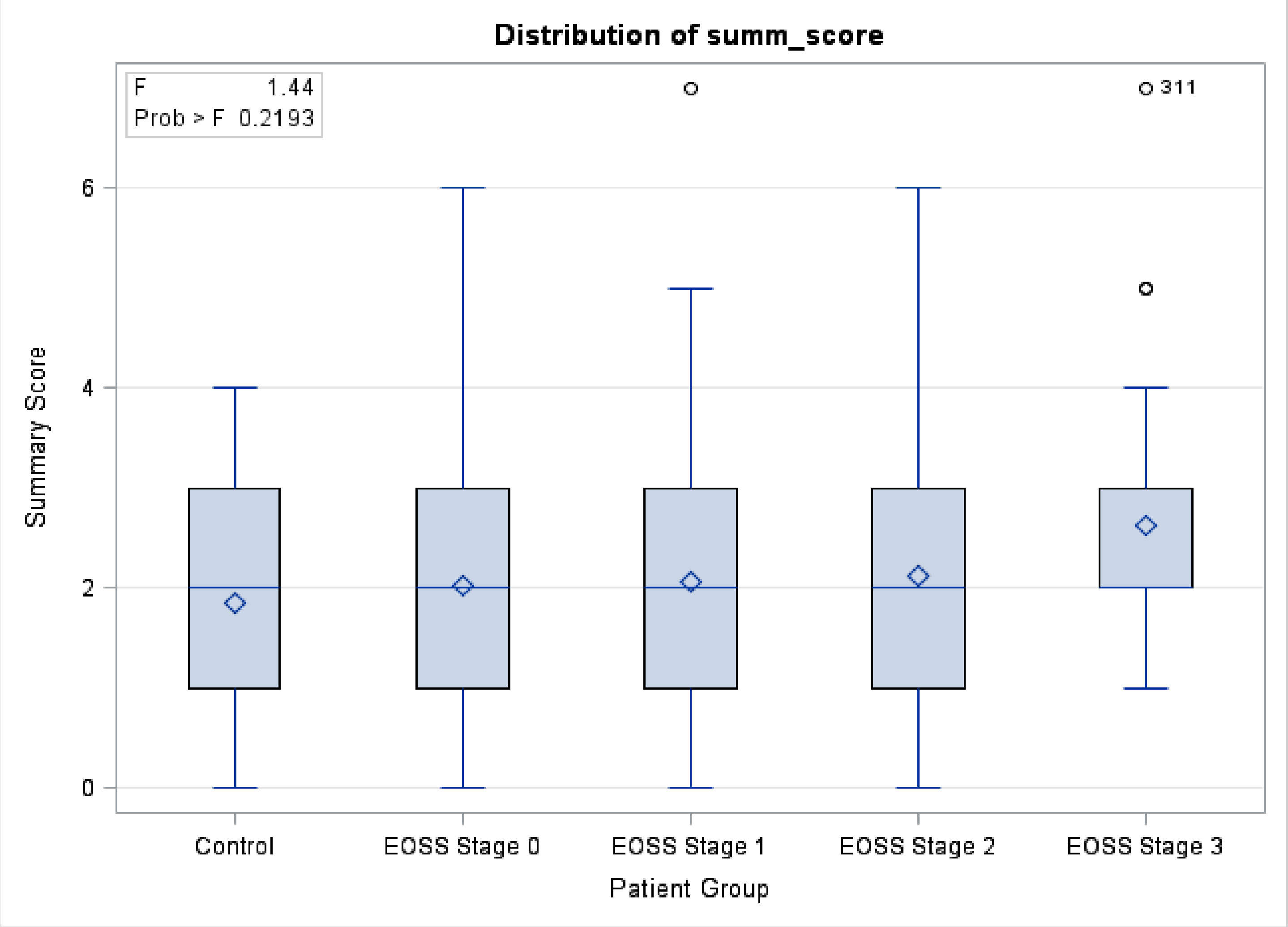
Glucose level 95% confidence interval (2.98, 3.41)

Boxplots

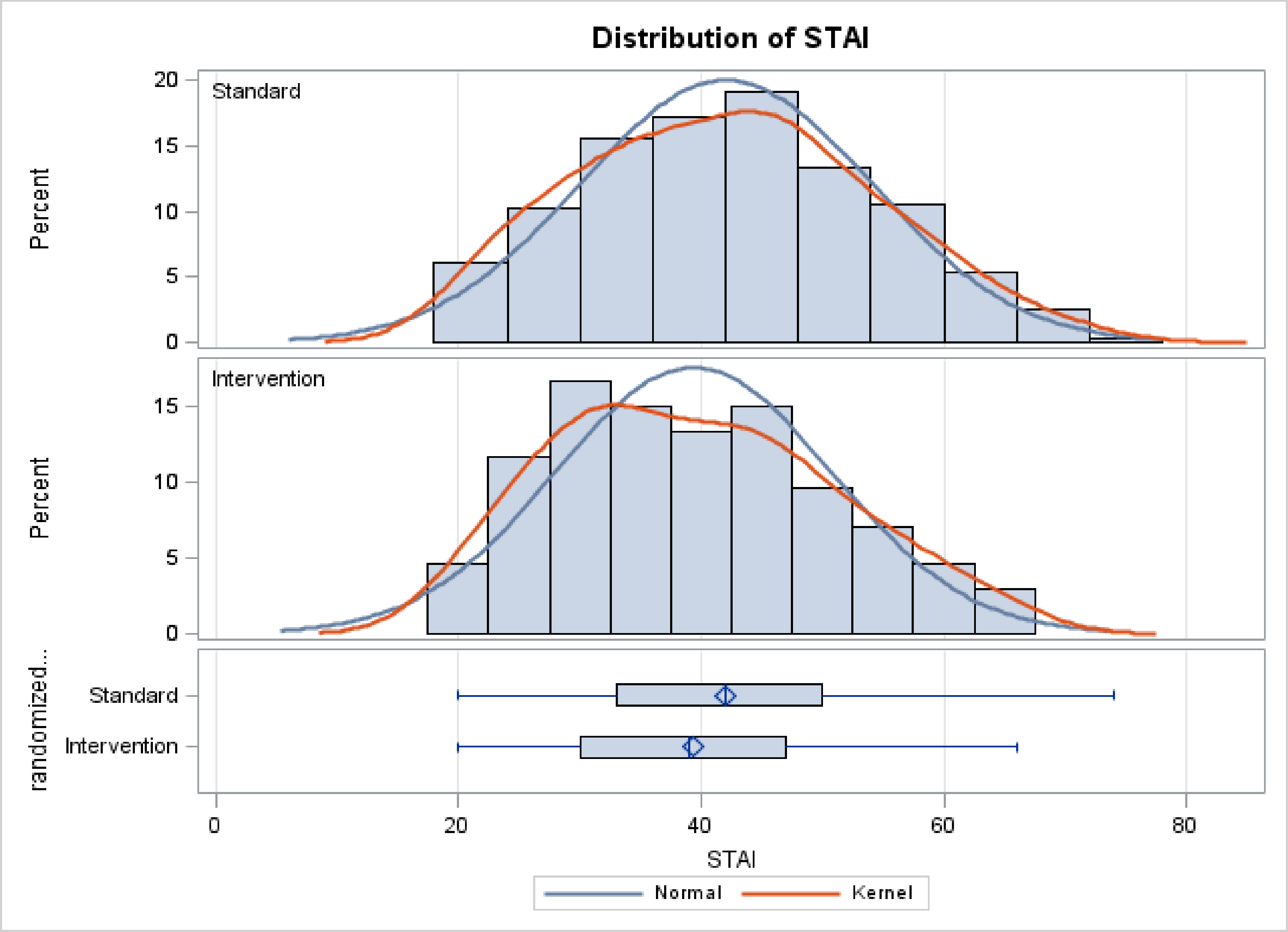

Interquartile Range
(IQR)
Outliers
Outliers
"Minimum"
(Q1 - 1.5*IQR)
Q1
(25th Percentile)
Median
Q3
(75th Percentile)
"Maximum"
(Q3 + 1.5*IQR)
−4
−3
−2
−1
0
1
2
3
4

women & children's health research institute

Boxplots

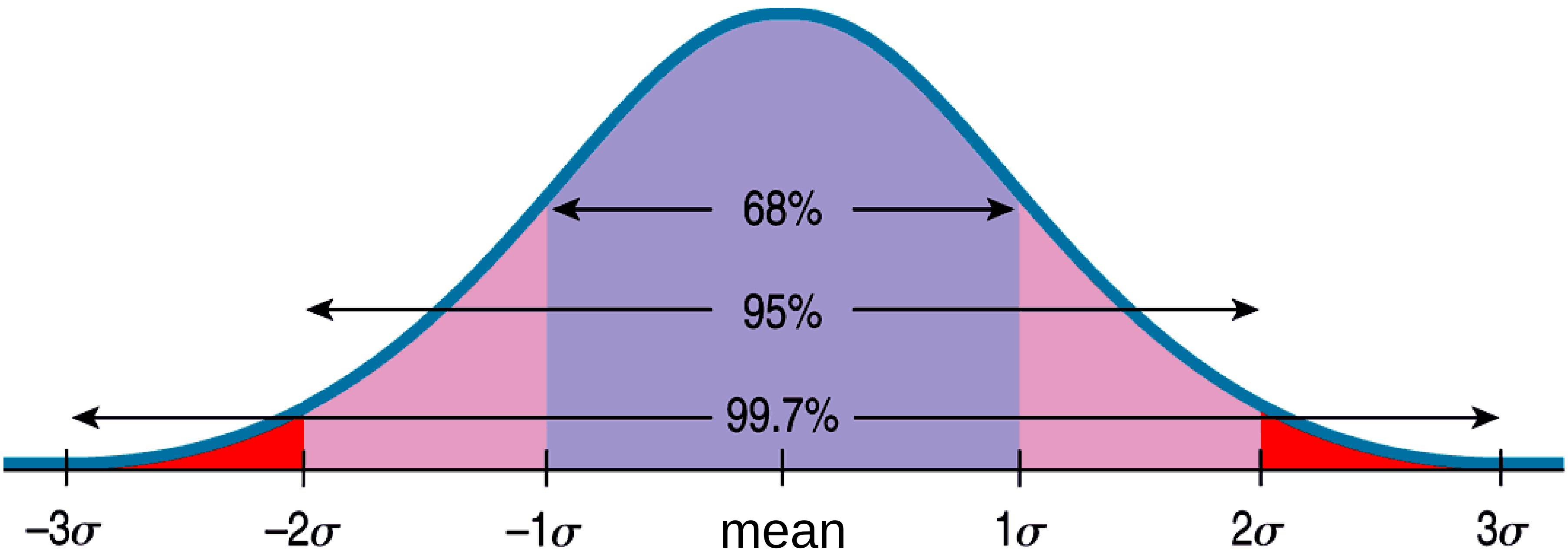


Histograms

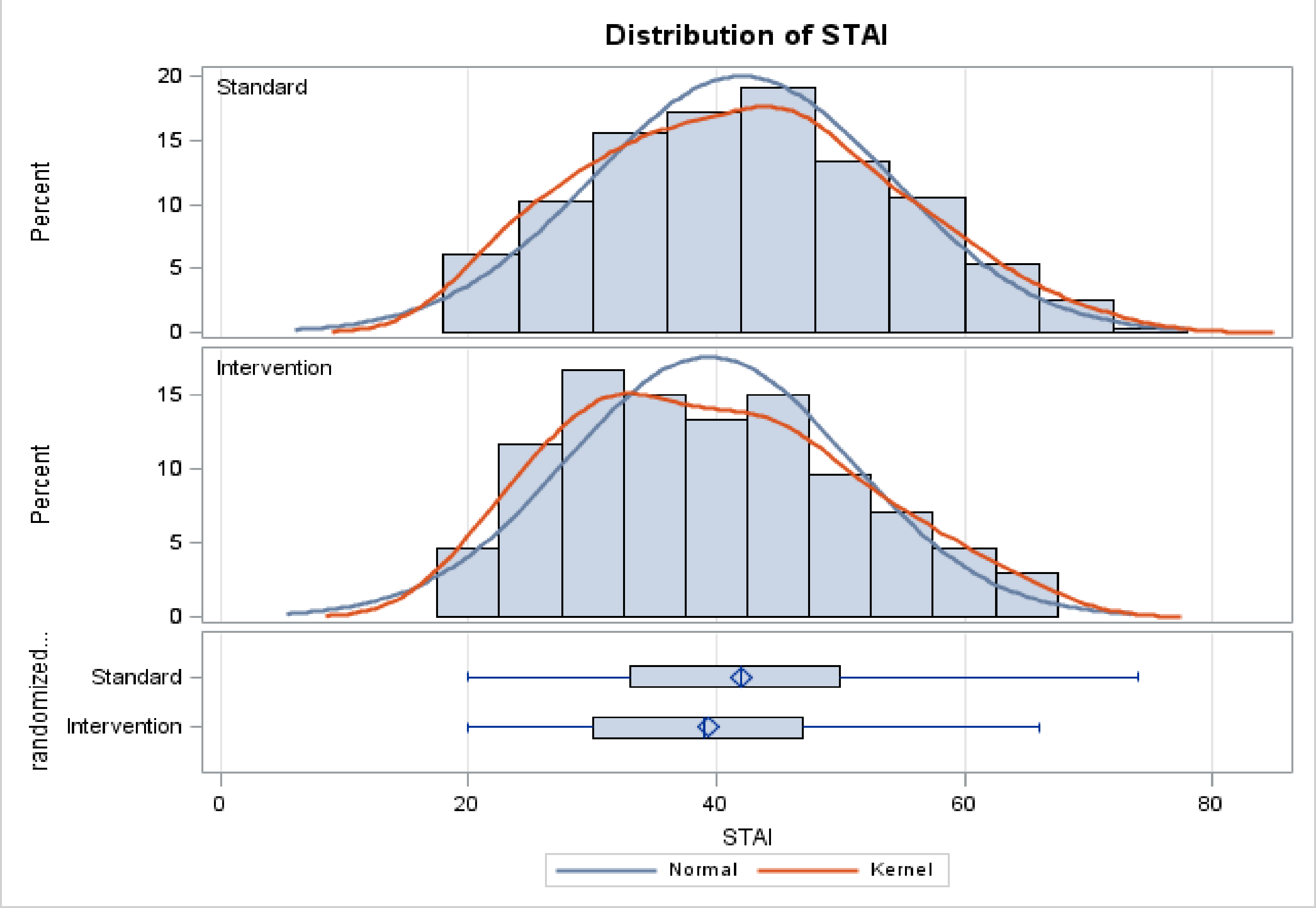


Distribution of STAI
Standard
Intervention
Percent
Percent
randomized...
Standard
Intervention
0
5
10
15
20
0
20
40
60
80
STAI
Normal
Kernel

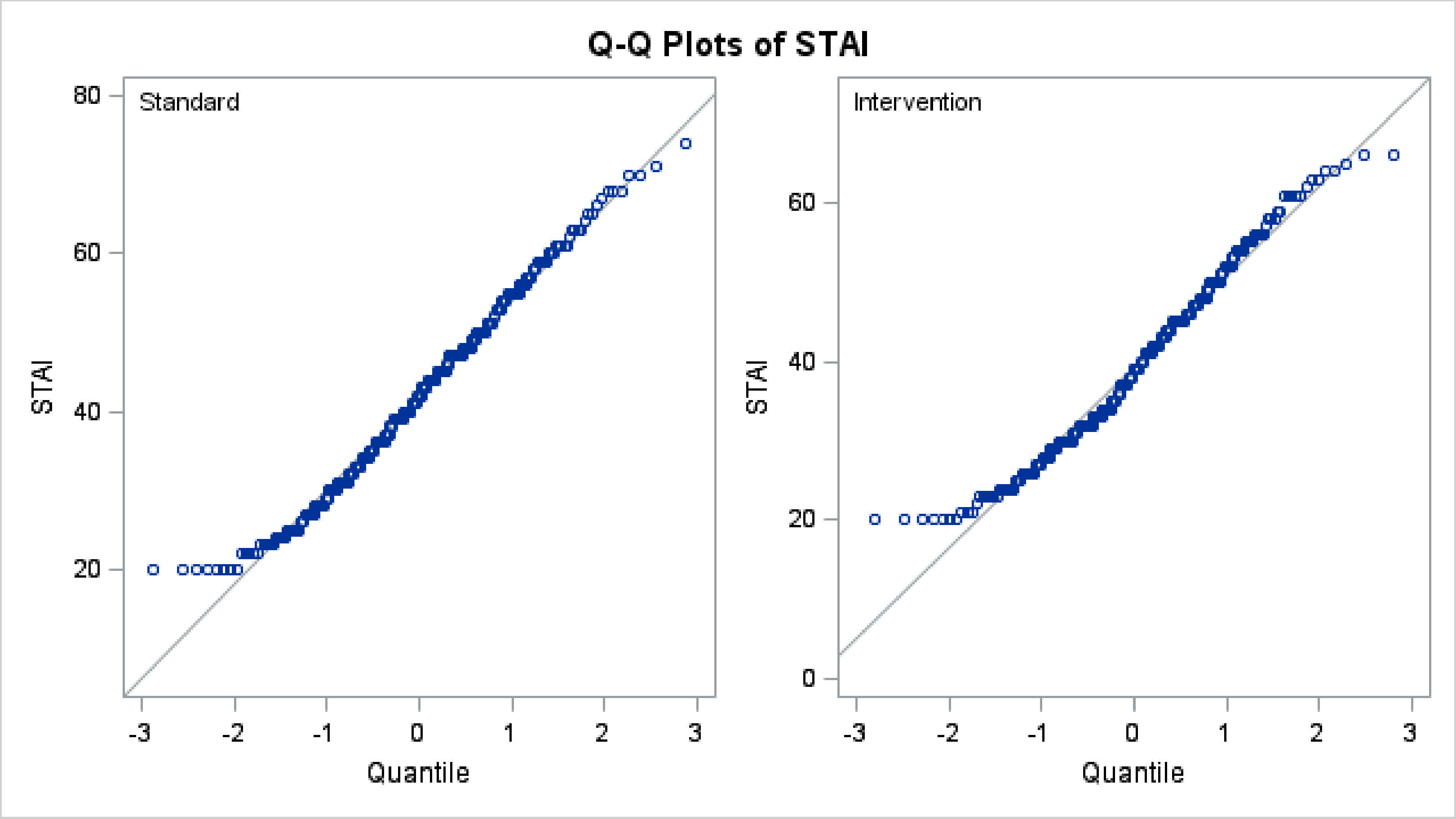
Normal Distribution



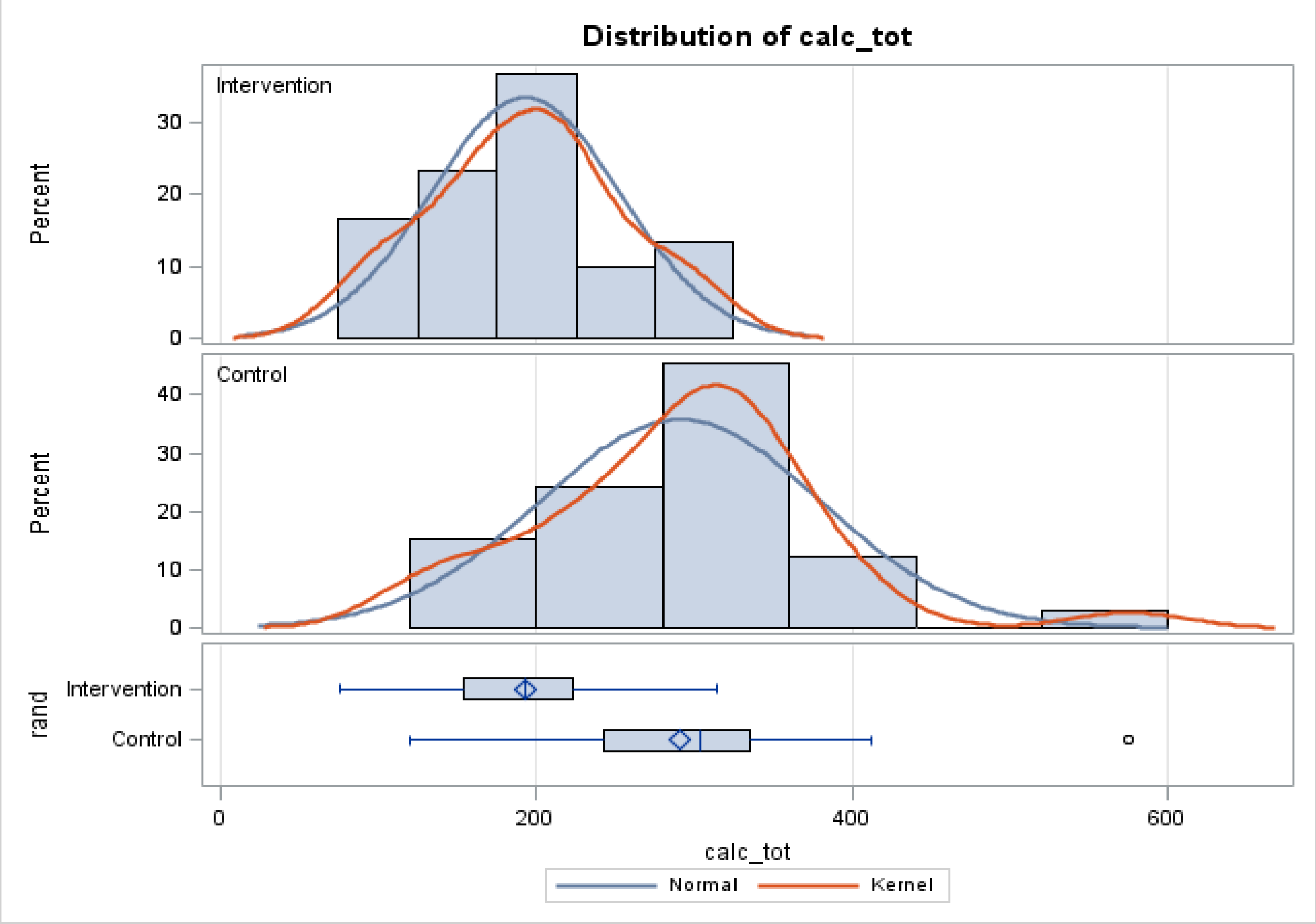
Examples of Normal and Non-Normal Distributions



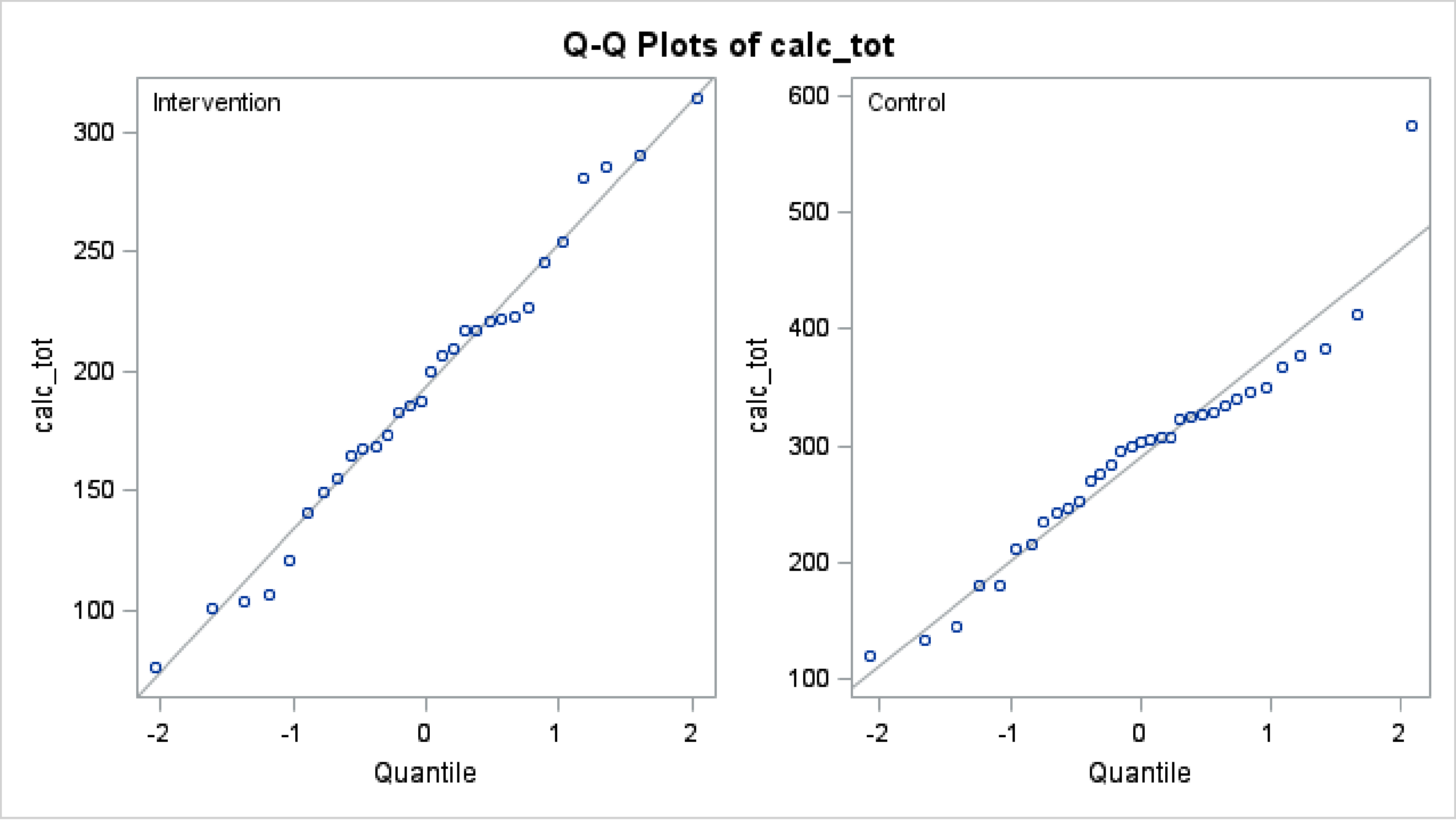
Examples of Normal and Non-Normal Distributions



Examples of Normal and Non-Normal Distributions



Examples of Normal and Non-Normal Distributions



Statistical Testing

Hypotheses

Hypotheses are working models that we adopt temporarily.

Our starting hypothesis is called the null hypothesis (H_0). In most instances H_0= our sample is no different from known information.

Hypothesis we want to prove – alternative hypothesis.

Two types of error: false positive (Type I error - α) and false negative (Type II error – β, or power of a test).

women & children's health research institute

Steps in Setting up a Test

1. Specify clearly and completely the question you will be asking of your data
2. Identify, specify and plan how to measure the variable(s) to answer that question
3. Review your definitions of sample and population, and verify the appropriateness of generalization
4. Review the sampling scheme to obtain your data
5. Specify exactly the null and alternative hypothesis
6. Select risks of a false-positive result (Type I error α) and of a false-negative result (Type II error β)

Steps in Setting up a Test

7. Choose the right test
8. Verify that your sample size is adequate to achieve the proposed power (power calculation)
9. Obtain your data. Use EDC system (RedCap) with data checks for clean data collection
10. Identify and test possible bias
11. Carry out analysis

Information Required to Choose a Test

1. What is your independent variable? What it your dependent variable?
2. What type of data do you have (categorical, rank-order, continuous)?
3. Are your measurements independent (e.g. on patients in treatment vs control group) or paired/ matched (e.g. pre-treatment, post-treatment and 6-months follow up on same patient)?

Multiple Testing

Multiple comparisons or **multiple testing problem** occurs when one considers a set of statistical inferences simultaneously.

The more inferences are made, the more likely false positive result to occur.

Bonferroni correction: significance= α/n (n = number of tests)

Common Tests for Categorical Data

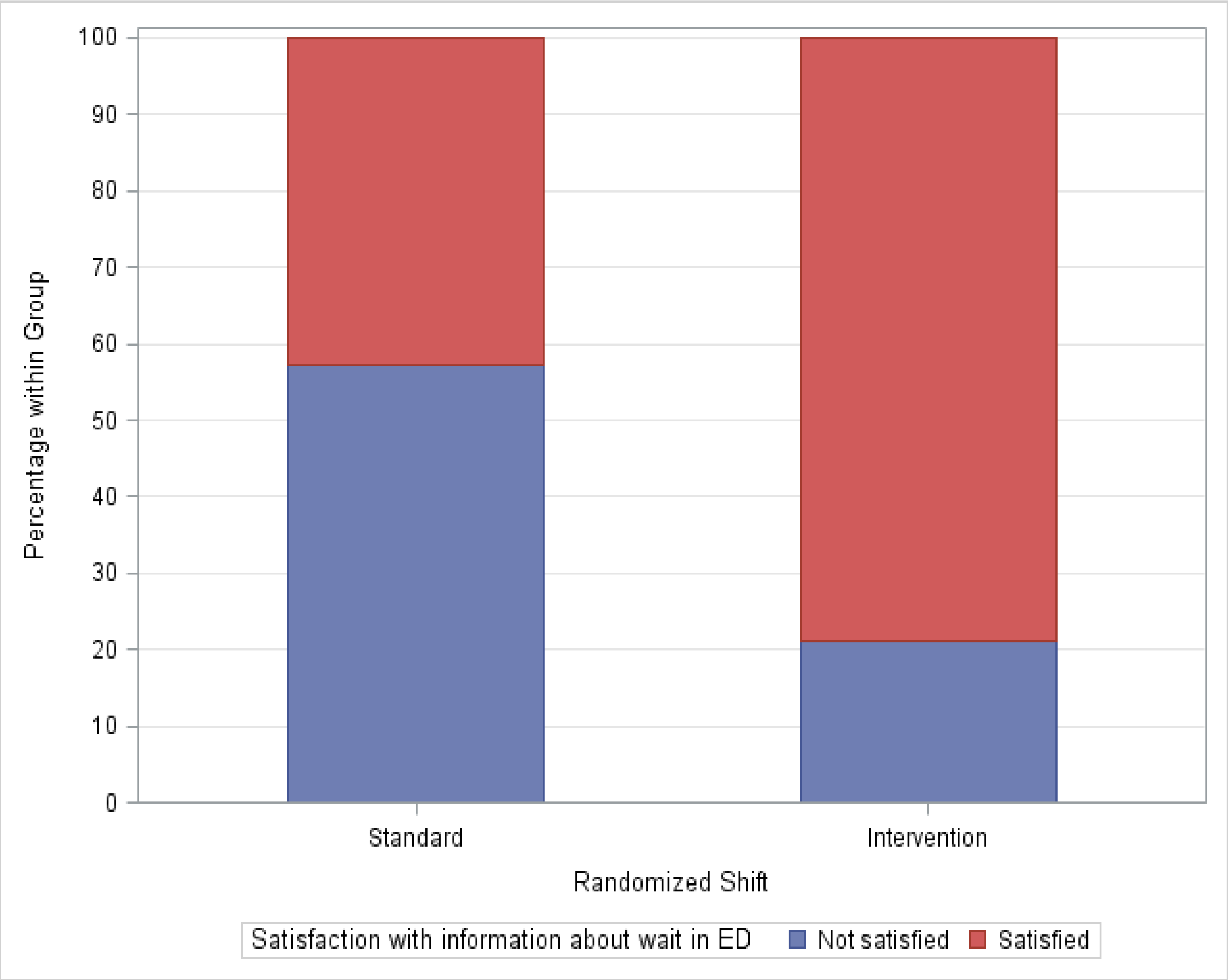
- Chi-square test
  - One sample (goodness-of-fit)
  - Two samples (homogeneity or independence)
- Fisher’s exact test
  - Small samples
- McNemar’s test
  - Paired data

Chi-square test

Satisfaction with information about reasons of wait in ED	Randomized Shift		
	Standard	Intervention	Total
Satisfied (4 and 5)	104 42.8%	165 78.9%	**269**
Not satisfied (1 and 2)	139 57.2%	44 21.1%	**183**
Total	**243**	**209**	**452**

Sample size is large, each cell count ≥ 5.

The difference in proportions of satisfied and not satisfied participants is statistically significant between standard and intervention arms, p<0.0001 (chi-square test).


100
90
80
70
60
50
40
30
20
10
0
Percentage within Group
Standard
Intervention
Randomized Shift
Satisfaction with information about wait in ED
Not satisfied
Satisfied

Common Tests for Continuous Data

- T-test (independent or paired data, 2 groups)
- ANOVA (3 or more independent groups)

Both work on normal data, ANOVA has several assumptions

Non-parametric tests:

- Mann-Whitney U test (independent)
- Wilcoxon signed rank (paired)
- Kruskal-Wallis test (ANOVA analogue)

T-test

State-Trait Anxiety Inventory (STAI) questionnaire

Randomized Shift	N	Mean	Std Dev	95% CI for Mean		Median	25th percentile	75th percentile	Min	Max
				Lower	Upper					
Standard	315	42.02	11.99	40.82	43.34	42.00	33.00	50.00	20.00	74.00
Intervention	241	39.38	11.38	37.93	40.69	39.00	30.00	47.00	20.00	66.00

Sample size is large, distributions are approximately normal, no outliers.

Mean STAI score was statistically significantly higher for standard arm than for the intervention arm, $p = 0.01$ (independent-samples t-test).

women & children's health research institute