# Statistics Overview

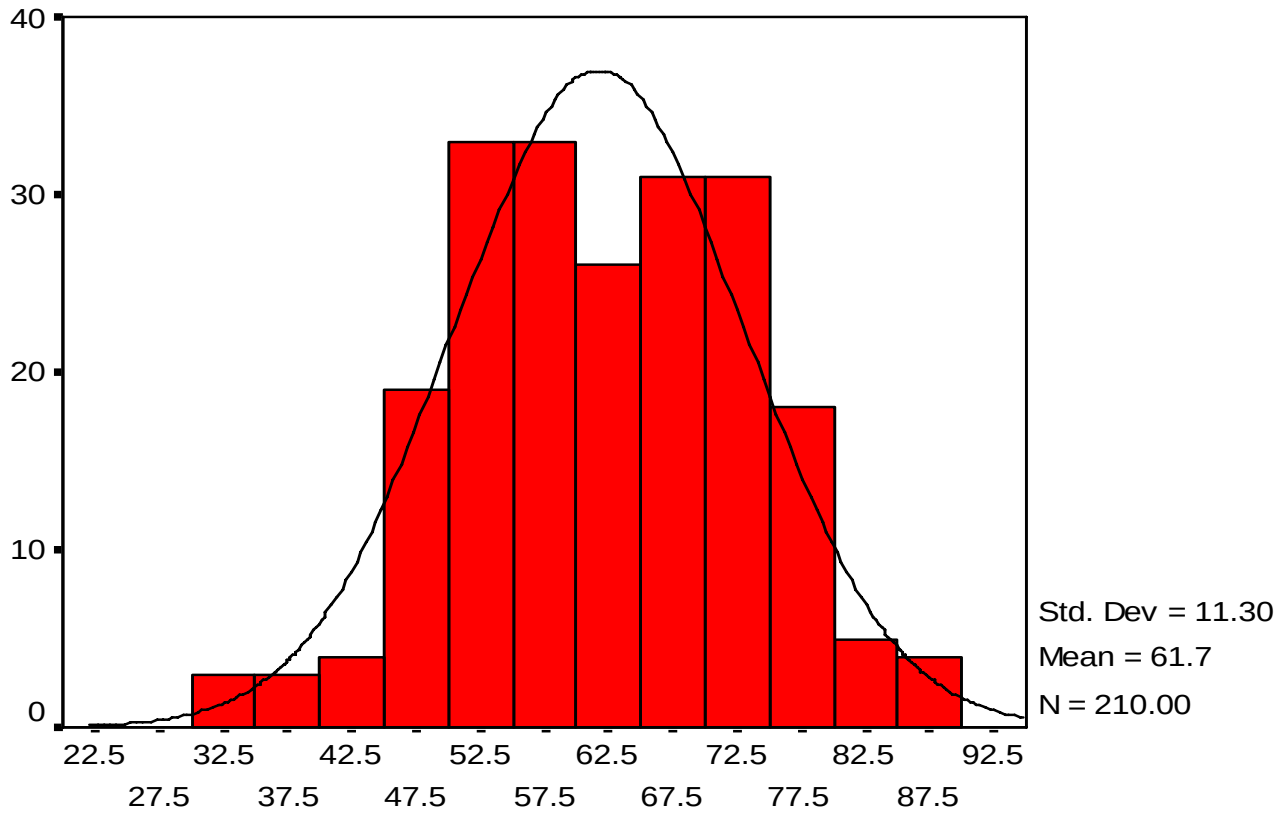
Lisa M. Pastore, PhD

OB/GYN Lecture

# Educational Objectives

- Learn that your first step with data analysis is to examine the data
- Understand how to interpret a relative risk
- Understand what a statistical test will tell you
- How to read selected SPSS output

# Look at the data: Graph!



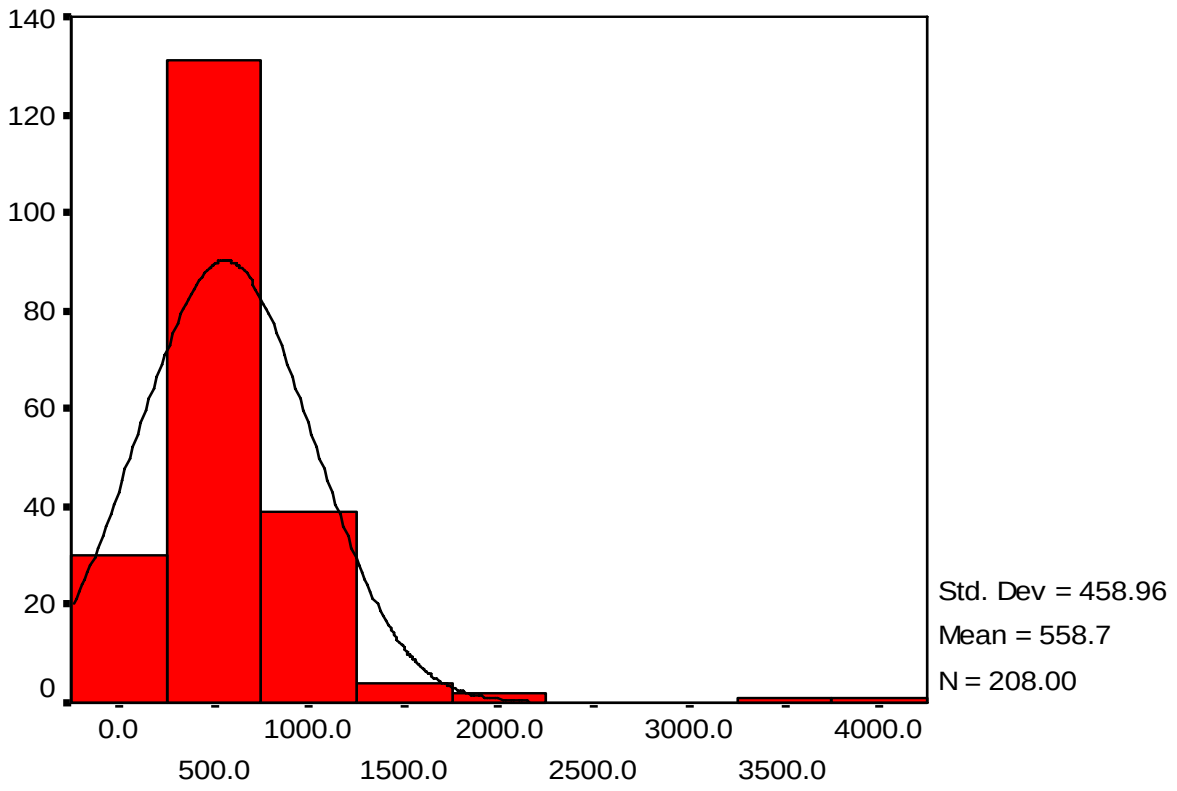
# Look at the data: Averages
## Mean, Median, Mode and Outliers

**Statistics**

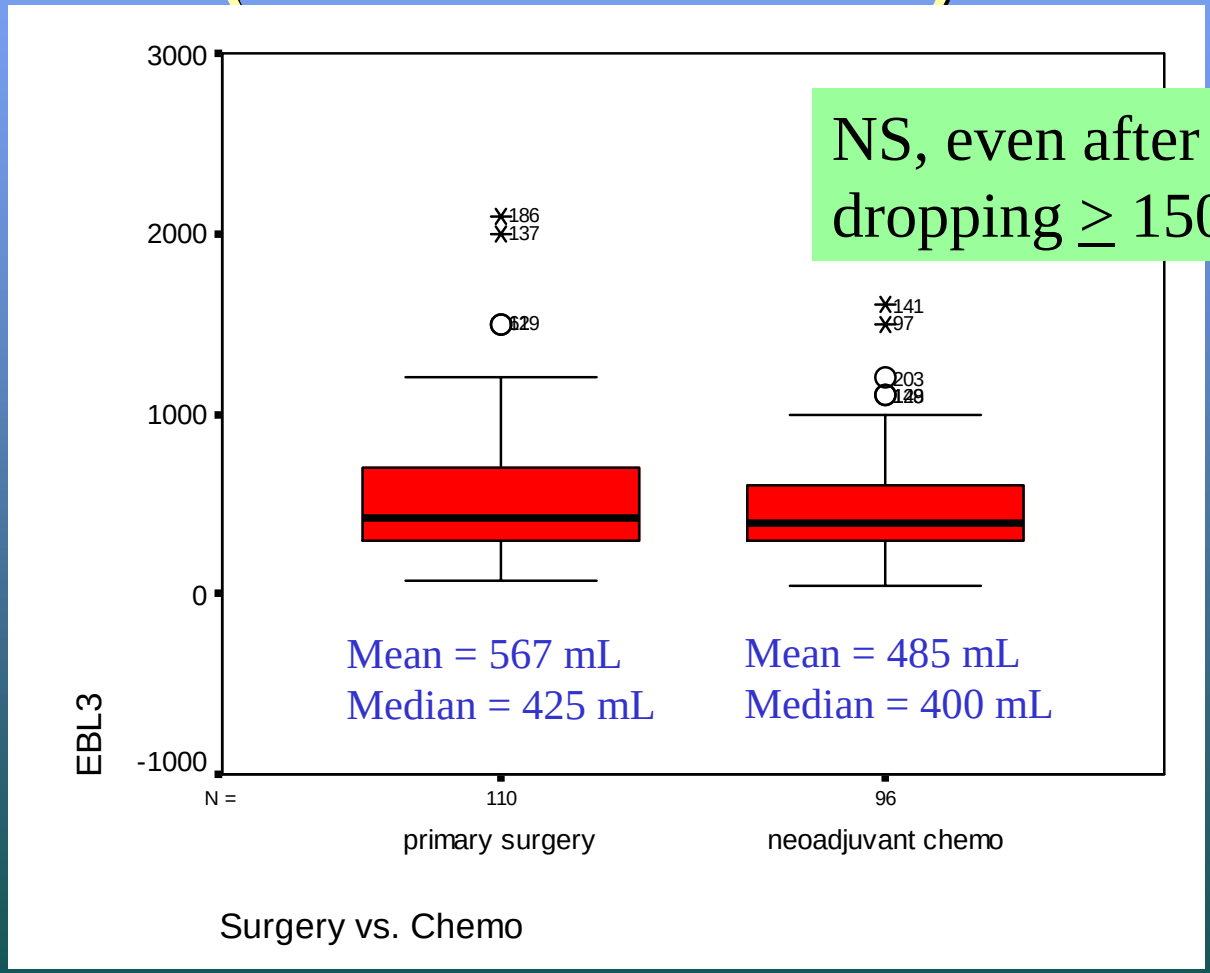
Estimated blood loss in milliliters

| | | |
|---|---|---|
| N | Valid | 208 |
| | Missing | 2 |
| Mean | | 558.65 |
| Median | | 400.00 |
| Mode | | 400 |
| Std. Deviation | | 458.959 |
| Variance | | 210643.3 |

# Let's Try Graphing the Data – Outliers!



# Remove Outliers, Re-Graph
## (data from E Everett)



# Step 2
# Basic 2 x 2 Table

| | “Diseased” | Not “diseased” |
|---|---|---|
| “Exposed” | a | b |
| Not “exposed” | c | d |

# Two by Two Table from SPSS
## Comorbidities by Surgery Method (E Everett data)

**Surgery vs. Chemo * Comorbidities Crosstabulation**

| | | | Comorbidities | | |
|---|---|---|---|---|---|
| | | | No | Yes | Total |
| Surgery vs. Chemo | primary surgery | Count | 52 | 60 | 112 |
| | | % within Surgery vs. Chemo | 46.4% | 53.6% | 100.0% |
| | | % within Comorbidities | 61.2% | 48.0% | 53.3% |
| | neoadjuvant chemo | Count | 33 | 65 | 98 |
| | | % within Surgery vs. Chemo | 33.7% | 66.3% | 100.0% |
| | | % within Comorbidities | 38.8% | 52.0% | 46.7% |
| Total | | Count | 85 | 125 | 210 |
| | | % within Surgery vs. Chemo | 40.5% | 59.5% | 100.0% |
| | | % within Comorbidities | 100.0% | 100.0% | 100.0% |

# Relative Risk

"Relative" = comparison of two groups, usually with vs. without the exposure

- Odds ratio (OR) for case-control studies
- Risk ratio (RR) for cohort studies

# Basic Interpretation

Relative risk

- > 1.0 implies the exposure may cause disease
- = 1.0 implies no relationship
- < 1.0 implies the exposure may prevent disease

# Three Examples from the June 2003 Green Journal

2nd trimester placental growth factor (mom's blood) predicts severe, early-onset preeclampsia, OR=4.2 (1.4-13.1) Polliotti. Obstet Gynecol 2003;101:1266-74

Alcohol consumption in general does not influence ovarian cancer rates, with non-signif OR ranging from 0.7-1.2 Goodman. Obstet Gynecol 2003;101:1221-8

Risk for post-c-section endometritis with extended antibiotic prophylaxis vs. placebo = 0.7 (0.4-0.98) Andrews. Obstet Gynecol 2003;101:1183-9

# How Would You Communicate this to Your Patients?

2nd trimester placental growth factor (mom's blood) predicts severe, early-onset preeclampsia, OR=4.2 (1.4-13.1)

Explain as a relative risk

Risk for post-c-section endometritis with extended antibiotic prophylaxis vs. placebo = 0.7 (0.4-0.98)

Explain as a relative risk

# Relative Risks as a Percentage Change

OR = 2.0 is equivalent to 100 % increase in risk of disease with the exposure

OR = 0.50 is equivalent to 50 % decrease in risk of disease with the exposure

OR = 4.0 is equivalent to 300 % increase in risk of disease with the exposure

OR = 0.75 is equivalent to 25 % decrease in risk of disease with the exposure

# Absolute Risk

Actual number of health problems that occur (or are prevented) because of the exposure

- "X number of extra cases in YYY number of people over ZZ years"
- "Gale model" for cancer risk

# Absolute Risk vs. Relative Risk

- Can only calculate absolute risk if one has the prevalence of the conditions/exposure in the population
- Insufficient data in reference articles for me to calculate absolute risk
- WHI: invasive breast cancer risk associated with HRT use
  - Absolute risk: 8 extra cases per 10,000 person years (38 cases/10,000 on HRT vs 30 cases/10,000 not on HRT)
  - Relative risk: 1.26

# Step 3: What does “running a statistical test” mean?

- We presume there is no relationship (association) between two factors.

  $H_0 : \mu_0 = \mu_1$

  THIS IS WHAT THE STATISTICS PRESUME IS TRUE.

- If the statistics say “this is not true because it is so improbable (unlikely) that there would be no relationship given the data”, then

  We report that there is an association.

  $H_A : \mu_0 \neq \mu_1$

# Error “Types”

If we were able to compare our study results to the Truth, there could be two (four) answers:

1. Our study found the True answer:
   a. No association
   b. There is an association

2. Our study did not find the True answer:
   a. We said there is an association but there is Truly none = “Type I Error”
   b. We said there is no association but there Truly is one = “Type II Error”

# How you use the error types without even knowing it

- Type I error: p-value < 0.05 means the relationship found is not a true difference but is instead due to “chance”; if study was repeated 100 times, 5% would find a relationship simply due to chance
- Type II error: statistical power (or having enough people in the study) addresses this; 80% power is standard = missing a true relationship 1 out of 5 times

# Stat Test from SPSS (Example 1)
# Comorbidities by Surgery Method

**Surgery vs. Chemo * Comorbidities Crosstabulation**

| | | | Comorbidities | | |
|---|---|---|---|---|---|
| | | | No | Yes | Total |
| Surgery vs. Chemo | primary surgery | Count | 52 | 60 | 112 |
| | | % within Surgery vs. Chemo | 46.4% | 53.6% | 100.0% |
| | | % within Comorbidities | 61.2% | 48.0% | 53.3% |
| | neoadjuvant chemo | Count | 33 | 65 | 98 |
| | | % within Surgery vs. Chemo | 33.7% | 66.3% | 100.0% |
| | | % within Comorbidities | 38.8% | 52.0% | 46.7% |
| Total | | Count | 85 | 125 | 210 |
| | | % within Surgery vs. Chemo | 40.5% | 59.5% | 100.0% |
| | | % within Comorbidities | 100.0% | 100.0% | 100.0% |

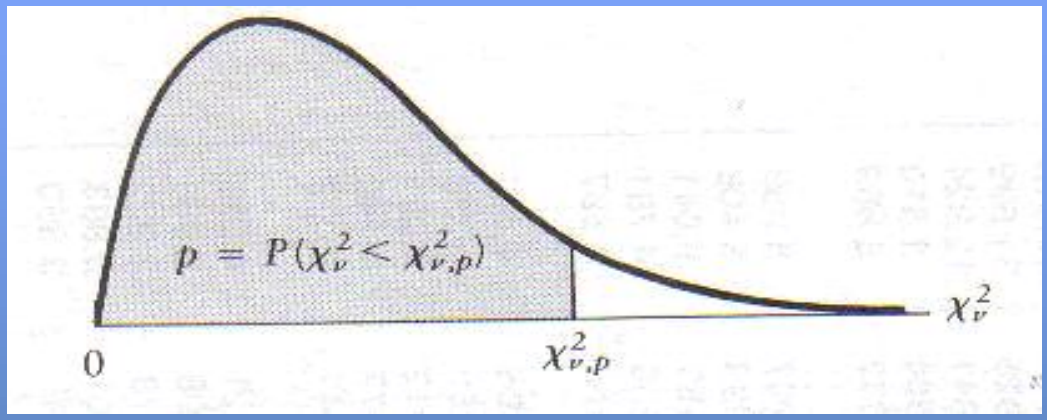
| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| **Pearson Chi-Square** | **3.529(b)** | **1** | **.060** |

# Stat Test from SPSS (Example 2)
## Chemo Regimen by Surgery Method

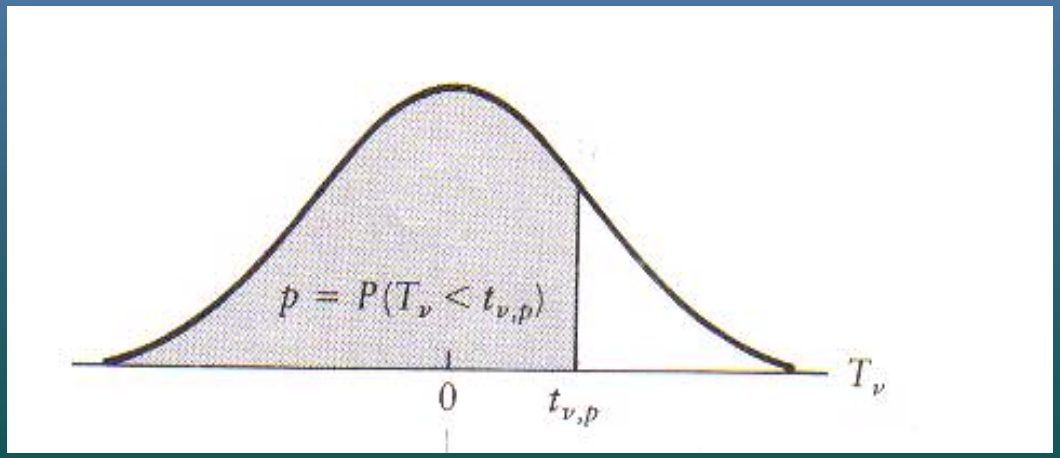
| | | Other | Cispl/Taxol or Carbo/Taxol | Total |
|---|---|---|---|---|
| primary surgery | | 10 | 95 | 105 |
| | | 9.5% | 90.5% | 100.0% |
| neoadjuvant chemo | | 3 | 95 | 98 |
| | | 3.1% | 96.9% | 100.0% |
| Total | | 13 | 190 | 203 |
| | | 6.4% | 93.6% | 100.0% |

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 3.532(b) | 1 | .060 | | |
| Fisher's Exact Test | | | | .084 | .054 |

# Chi-Square Distribution




# Normal Distribution



# Stat Test from SPSS (Example 3)
## Student's t-Test
## Operating Time by Surgery Method

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | |
|---|---|---|---|---|---|---|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference |
| Equal variances assumed | 5.408 | .021 | -2.765 | 203 | .006 | -26.7149 |
| Equal variances not assumed | | | -2.730 | 181.93 | .007 | -26.7149 |