

Comparing Confidence Intervals for a Difference in Adjusted Predicted Probabilities

Cralen Davis and Trevano Dean, Cytel, Inc.

PHUSE US Connect 2025

18-MAR-2025



The Introduction

- **The Origin Story**
- **The Idea**
- **The Data**
- **The Analysis**
- **The Results**
- **The Conclusion**



The Idea

- **Asymptotic Normal Approximation**
 - **Appropriate for C.I.?**
- **How to Check Validity and Performance?**
- **Empirical Comparator**
 - **Bootstrap sampling**
- **Bootstrapping**
 - **Sampling method with replacement**



The Data

- **Simulated dataset with 263 rows representing 263 unique subjects**
- **Subjects randomized 1:1 into 2 treatment groups**
 - **Group A**
 - **Group B**
- **There are 5 total variables in the dataset:**
 - **ID (unique numeric identifier)**
 - **TRT_GRP (Group A/Group B)**
 - **AGE (numeric)**
 - **SEX (M/F)**
 - **Severe_Event (0/1)**
 - **0 = No event**
 - **1 = Severe event occurred**



The TTEST Procedure

Analysis Variable : AGE						
TRT_GRP	N	Mean	Std Dev	Std Err	Minimum	Maximum
Group A	133	41.8	12.9	1.1	18.0	75.0
Group B	130	43.0	13.2	1.2	20.0	76.0

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	261.0	-0.79	0.4290
Satterthwaite	Unequal	260.5	-0.79	0.4291

Table of Severe_Event by TRT_GRP			
Severe_Event	TRT_GRP		
	Group A	Group B	Total
0	80 60.15	59 45.38	139
1	53 39.85	71 54.62	124
Total	133	130	263

Statistics for Table of Severe_Event by TRT_GRP

Statistic	DF	Value	Prob
Chi-Square	1	5.7521	0.0165

The Data (continued)

Table of SEX by TRT_GRP			
SEX(Sex)	TRT_GRP		
	Group A	Group B	Total
F	107 80.45	107 82.31	214
M	26 19.55	23 17.69	49
Total	133	130	263

Statistics for Table of SEX by TRT_GRP

Statistic	DF	Value	Prob
Chi-Square	1	0.1495	0.6990

The Analysis

➤ SAS 9.4

➤ Logistic Regression:

```
proc logistic order=data;  
class TRT_GRP SEX / order=data  
param=glm;  
model Severe_Event = AGE SEX TRT_GRP;  
lsmeans TRT_GRP / cl ilink diff;  
run;
```

➤ Difference in Mean APPs (p_{diff})

➤ $p_{\text{diff}} = p_b - p_a$

➤ $\text{Var}_{\text{pooled}} = \text{SE}_B^2 + \text{SE}_A^2$

➤ 95% C.I. by Asymtotic Normal Approximation

➤ $95\% \text{ CI} = p_{\text{diff}} \pm 1.96 \times \sqrt{\text{Var}_{\text{pooled}}}$



The Analysis (continued)

➤ Bootstrapping

- 3 Total rounds
- $k = 100, 500, \text{ and } 1000$

➤ Sampling with Replacement:

```
proc surveyselect data=tmp method=balboot outhits  
    reps=100 seed=1142 out=out100;  
run;
```

- tmp = original 263 record dataset
- balboot = bootstrapping
- outhits = retains all sampled repeats
- reps = k samples to draw
- seed = user provided integer for random sample processing
- out = output dataset (for $k = 100$; "out100" has 26300 total rows)

➤ Lots of Logistic Regression Models!



The Analysis (still going...)

- **Same PROC LOGISTIC Model**
 - Run k times
- **k Number of p_{diff} Values**
- **Empirical 95% C.I.**
 - Order p_{diff} values from smallest to larger
 - Select 2.5th %-tile for lower 95% C.I. bound
 - Select 97.5th %-tile for upper 95% C.I. bound
 - 95% C.I. = $(\{p_{\text{diff}2}/2 + p_{\text{diff}3}/2\}, \{p_{\text{diff}97}/2 + p_{\text{diff}98}/2\})$; for $k = 100$
 - 3 total CIs ($k = 100; 500; 1000$)



The Results

Table 1a.

Source	Group A p	$\sqrt{Var}$	Median
Unadjusted Data	0.3985	0.0425	--
Logistic Modeling	0.4291	0.0495	--
100 Samples WR	0.4295	0.0496	0.4333
500 Samples WR	0.4287	0.0500	0.4289
1000 Samples WR	0.4292	0.0513	0.4301

Table 1b.

Source	Group B p	$\sqrt{Var}$	Median
Unadjusted Data	0.5462	0.0437	--
Logistic Modeling	0.5847	0.0505	--
100 Samples WR	0.5858	0.0556	0.5906
500 Samples WR	0.5856	0.0509	0.5876
1000 Samples WR	0.5857	0.0543	0.5867

The Results (yeah! more numbers...)

Table 2.

Source	p_{diff} ($p_B - p_A$)	$\sqrt{Var}$	Median
Unadjusted Data	0.1477	0.0609	--
Logistic Modeling	0.1556	0.0707	--
100 Samples WR	0.1563	0.0647	0.1492
500 Samples WR	0.1569	0.0622	0.1591
1000 Samples WR	0.1565	0.0618	0.1583

Table 3.

Source	95% C.I. LCL	95% C.I. UCL	C.I. Width
Unadjusted Data	0.0283	0.2670	0.2387
Logistic Modeling	0.0170	0.2942	0.2773
100 Samples WR	0.0098	0.2730	0.2632
500 Samples WR	0.0384	0.2903	0.2519
1000 Samples WR	0.0323	0.2755	0.2432

➤ **The Origin Story**

➤ **The Idea**

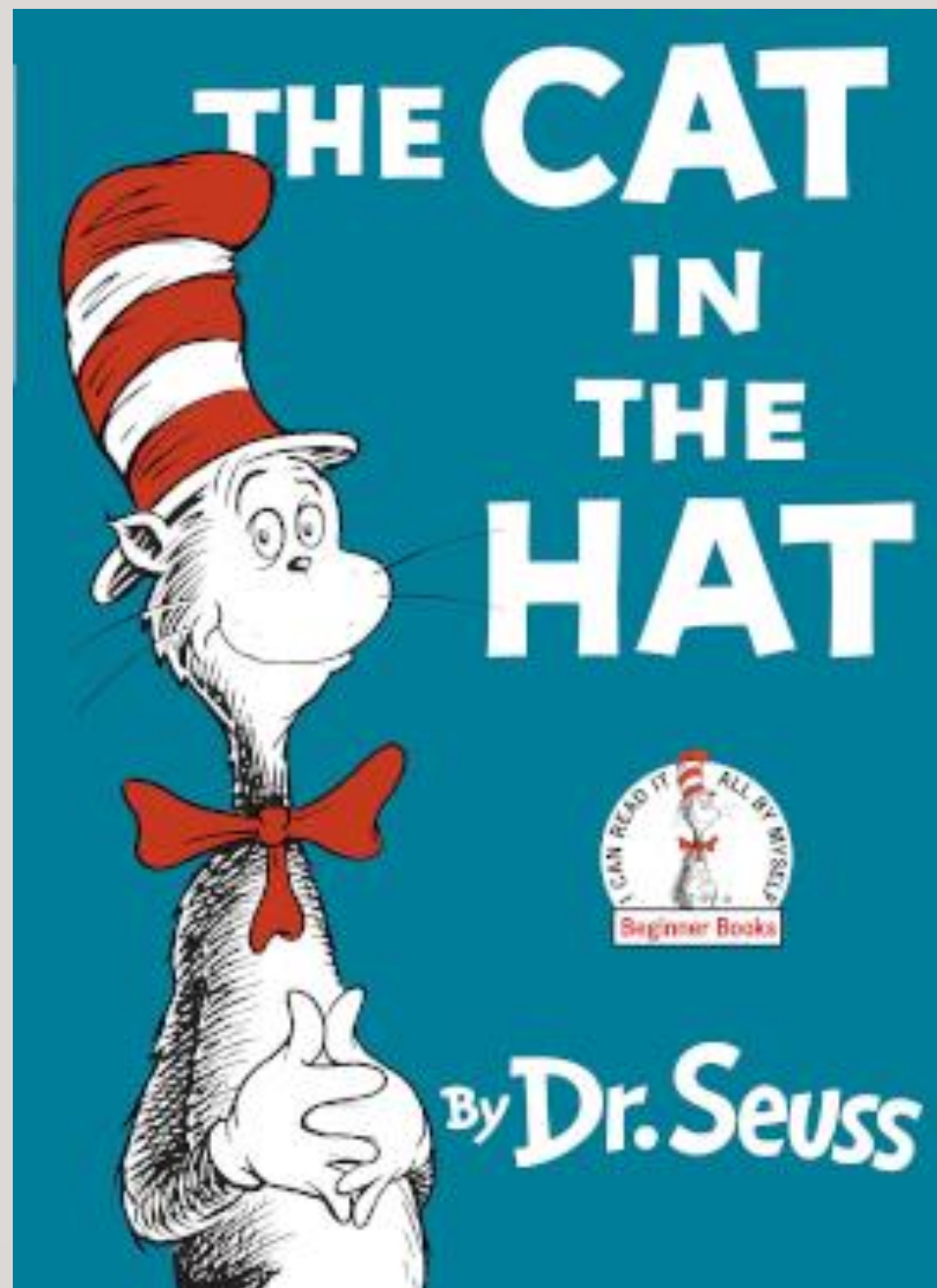
➤ **The Data**

➤ **The Analysis**

➤ **The Results**

➤ **The Recap**





The Conclusion

- **Generating adjusted predicted probability estimates is straight forward**
- **Practical to use logistic model estimates**
 - **Difference in mean APPs**
 - **Asymptotic Normal Approximated C.I.**
- **Valid approach with adequate performance**
 - **Bootstrap sample-based C.I. comparator**
- **Important note: KNOW your data**
 - **Factors to consider: sample size; covariates; missing data; data imbalance; etc.**



The Acknowledgments

- **Cytel Leadership and Colleagues**
- **PHUSE 2025**
 - **Organizers**
 - **Stream Chairs**
 - **Volunteers**
- **Last, but certainly not least...**



Thank All of You for Attending!

Questions?

Break In Case Of Certain Questions

TRT_GRP	Estimate	Standard Error	95% C.I. LCL	95% C.I. UCL	Logit	$\sqrt{Var}$	95% C.I. LCL	95% C.I. UCL
Group B	0.3419	0.2080	-0.0657	0.7495	0.5847	0.0505	0.4836	0.6791
Group A	-0.2857	0.2022	-0.6819	0.1105	0.4291	0.0495	0.3358	0.5276

Model	Estimate	$\sqrt{Var}$	95% C.I. LCL	95% C.I. UCL	Logit	$\sqrt{Var}$	95% C.I. LCL	95% C.I. UCL
Group B - A	0.6276	0.2523	0.1332	1.1221	0.6520	N/A	0.5332	0.7544

Normal Approximation	<u>Probit B_p</u>	<u>Probit A_p</u>			Logit Difference	$\sqrt{Var}$ pooled	95% C.I. LCL	95% C.I. UCL
Group B - A	0.5847	0.4291			0.1556	0.0707	0.0170	0.2942

$$\text{Logit} = \log\left(\frac{p}{1-p}\right) \Rightarrow p = \left(\frac{\exp(\beta)}{1 + \exp(\beta)}\right)$$