Alliance for
Continuing Education
in the Health Professions

Alliance

2018 ANNUAL CONFERENCE

Destination: Patient Outcomes.
Our Journey to Improving Patient Care.

January 20-23, 2018
Grande Lakes, Orlando, Florida

Alliance for
Continuing Education
in the Health Professions

Your Activity Improved Outcomes, But Do You Know Why? How to Use Predictive Modeling to Maximize Educational Impact

Jamie Reiter, PhD
Director, Educational Outcomes
CME Outfitters, LLC

Before we start…

- Grab a cup of coffee
- If you would like to follow along with Excel when we review the examples, please do the following:
  - Make sure you have installed the Excel Data Analysis ToolPak (instructions available as pdf handout):
    - Mac: Tools menu, then Add-Ins
    - PC: Office logo, then Excel Options, then Add-Ins
  - Download Excel file provided with slides
- Additional handouts:
  - Terminology
  - How to install Data Analysis ToolPak

Brief Bio

- Previous:
  - PhD from Dept. Cognitive Sciences, UC Irvine
    - Emphasis statistical analysis and mathematical modeling of cognitive/neuropsych data
  - Biostatistics course developer and instructor (intro and intermediate), UC San Diego Extension
  - Statistical Consultant: biotech/pharma, medical societies, academia, contract research organizations
  - Senior Statistician:
    - UCSD Alzheimer's Disease Research Center
    - Dart Neuroscience
  - Director of Biostatistics and Research, CME LLC
- Current: Director, Educational Outcomes, CME Outfitters, LLC

Learning Objectives

- Upon completion, participants will be able to…
  - identify activities that would benefit from a predictive modeling analysis
  - determine variables most relevant for inclusion in their own predictive modeling analysis
  - conduct a basic predictive modeling analysis

Outline

- Why predictive modeling
- Types of predictive models, including:
  - Simple linear regression, with example
  - Multiple linear regression, with brief example
  - Simple logistic regression, with example
  - Multiple logistic regression, with brief example
- Conclusions

Why Predictive Modeling in Medical Education?

- Inform design for future activities, ie, determine factors that influence clinician knowledge, confidence, competence, and/or behavior
  - If demographics influence learning or behavior, can target subgroups
  - If knowledge for particular topics influences behavior, can offer additional education on those topics
  - If confidence influences behavior, can implement methods for improving confidence

Alliance for Continuing Education in the Health Professions

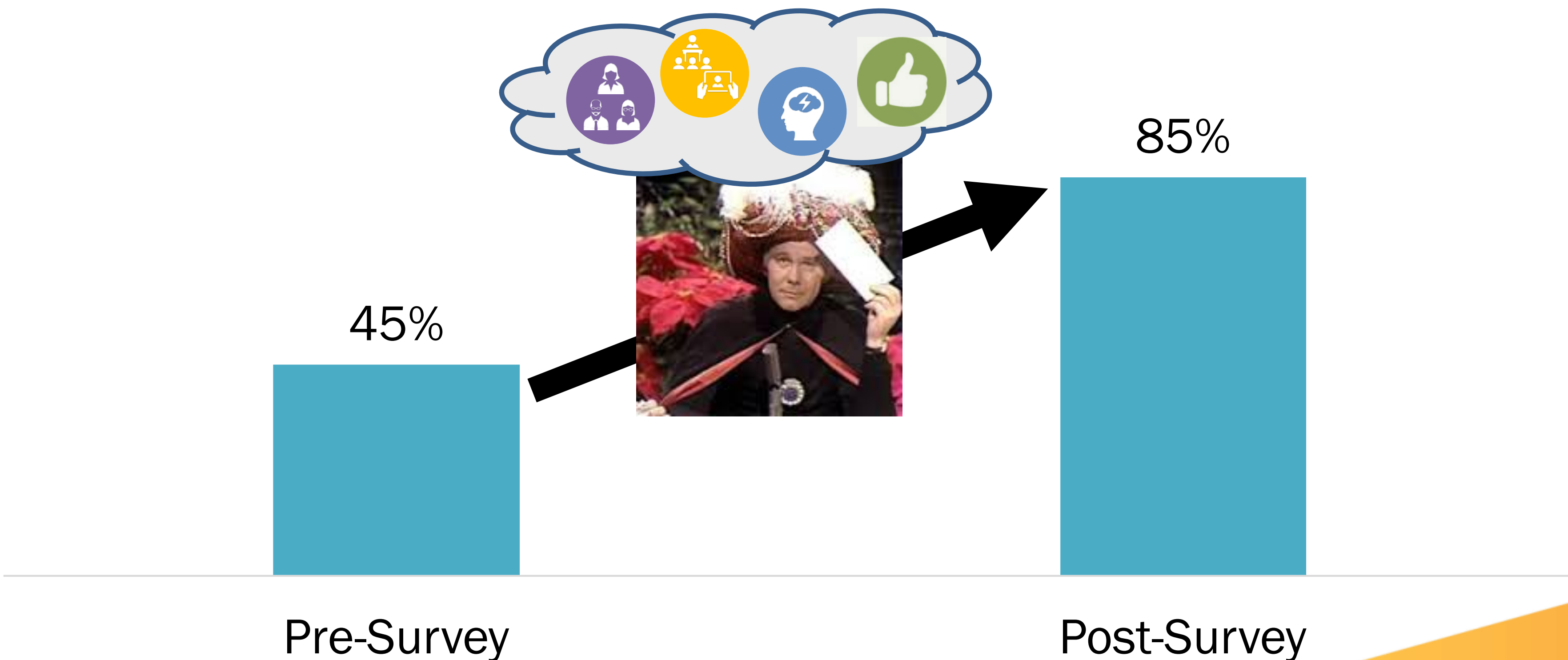
CME Outfitters' Experiences in Predictive Modeling

- Industry recognition of value—in past 9 months:
  - Invited by ACEhp to conduct webinar on predictive modeling
  - Invited by BeaconLive to conduct webinar on predictive modeling
  - This workshop
- In the last year, CMEO has submitted 15 abstracts on predictive modeling to industry/scientific/medical meetings; 10 have been accepted and 3 are pending acceptance
- Findings:
  - Confidence has consistently been the strongest predictor of clinician behavior
  - Other predictors: knowledge, barriers to practice
- What we are doing about it:
  - Designed a "reinforcement" component to our activities that aim to improve confidence and subsequent behavior

PredictCME

ReinforceCME

Improvement



No Improvement

What happens if…

…your activity isn't successful (ie, behavior doesn't improve), and you keep repeating the model because you don't know what to fix? Or, you try different things until something works?

Conversely, what happens if…

…your activity was successful, and you decide you want to mix things up for the next time?

Alliance for Continuing Education in the Health Professions

Solution: Predictive Modeling

- Several types:
  - Regression
  - Neural networks
  - Naïve Bayes
  - CHAID (chi-square automatic interaction detection)
  - Others
- We will focus on the most common form, namely regression

What is Regression

- In a nutshell:
  - Regression is a way of describing the relationship between two or more variables
  - It takes actual data (eg, on a scatter plot), then finds a line (or curve) that best fits the data
  - This enables researchers to pick any value from one variable on the line/curve, and identify (or predict) the corresponding value for the other variable
  - The better the line/curve fits the data, the more accurate the prediction

Regression Terminology

- Response variable: the variable(s) to be predicted (aka dependent variable, criterion variable)
- Predictor variable: one or more variables that predict responses on the response variable (aka independent variable)
- Regression line/curve: best-fitting line/curve of the data (eg, through points in a scatter plot)
- Simple regression: single predictor and single response variable
- Multiple regression: multiple predictors and single response variable

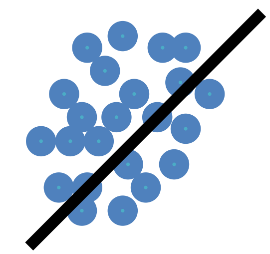
Possible Response Variables

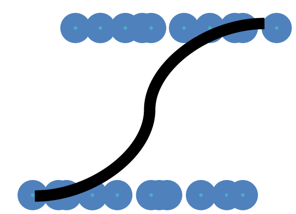
- Changes/improvements (eg, from pre to post)
- Single time point

Time Point	Possible Response Variables
Pre-survey (participants)	Existing knowledge, confidence, competence, behavior
Post-survey (participants)	Immediate post-activity knowledge, confidence, competence, planned behavior
Follow-up survey (participants)	Longer-term knowledge, confidence, competence, behavior
Follow-up survey (controls)	Similar to pre-survey
Change scores (matched participants; pre/post, pre/follow-up, post/follow-up)	Changes in knowledge, confidence, competence, behavior

Main Types of Regression

- Linear
  - Continuous response variable (eg, %correct*)
- Logistic
  - Discrete/categorical data (eg, correct/incorrect, good/fair/poor, yes/no)





*Some statisticians argue that percentage is not continuous, but under certain conditions it is acceptable to treat it as such.

Alliance for
Continuing Education
in the Health Professions

Simple Linear Regression

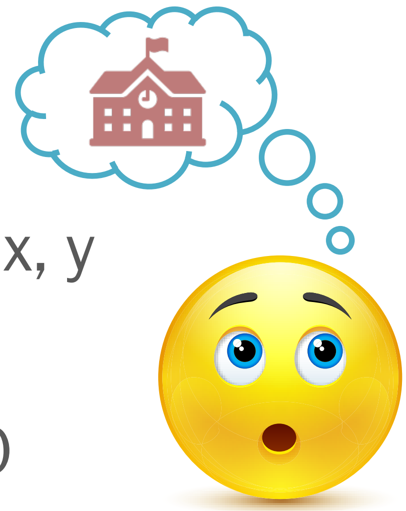
Linear Regression Assumptions

- Linear relationship between variables (check scatter plot)*
- Multivariate normality (plot of residuals, eg, QQ plot)
- No or little multicollinearity (independent variables are not correlated, eg, via Pearson bivariate)
- Observations are independent from one another (no autocorrelation)*
- Homoscedasticity (variation of residuals along horizontal axis is equal)
- Sample size rule of thumb: At least 20 cases per independent variable

*Most important assumptions

Linear Regression Equation

- $y = mx + b$
  - m = slope: for every unit increase in x, y increases by this amount
  - b = y-intercept: value of y when x = 0



Note: You may see the above equation with an error term (ε) added at the end—that describes the function for x and y, but the above equation refers to the best fit line.

Example: Linear Regression

- Standard educational activity/outcomes on Alzheimer's disease
- Research question: Does years of experience influence learning?
- Regression model:
  – **Response variable:** Relative* change in %correct knowledge (across 5 questions, post - pre)
  – **Predictor variable:** Years of experience

*Relative change = ((%post – %pre)/%pre)*100

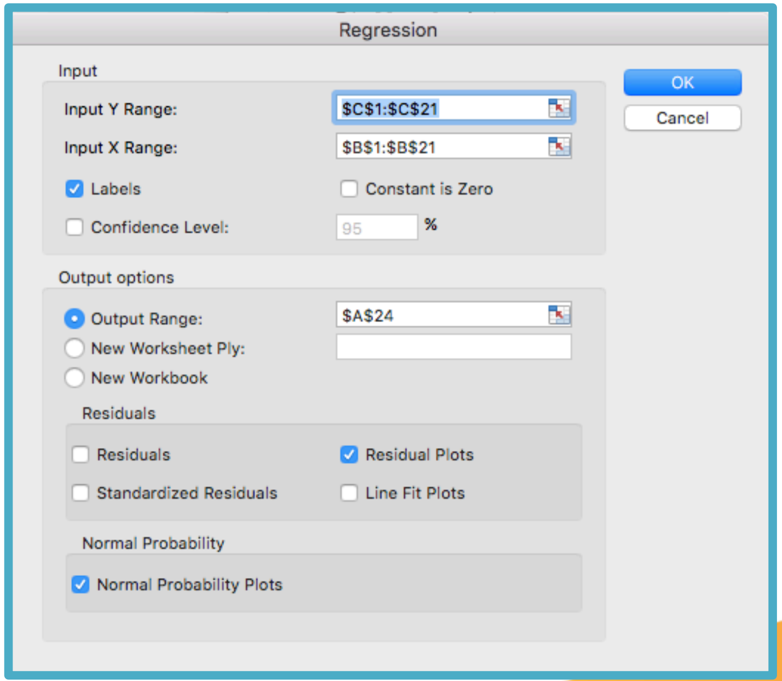
Example: Linear Regression (cont.)

- Enter Data
  - First column: Participant ID/#
  - Second column: Predictor variable (Years Experience)
  - Third column: Response variable (Relative %Change in Knowledge)

Participant ID	Predictor Variable: Years Experience	Response Variable: Relative %Change in Knowledge
1	19	80
2	5	63
3	9	70
4	16	75
5	6	61
6	20	87
7	31	94
8	27	86
9	1	53
10	7	63
11	3	58
12	22	85
13	16	82
14	31	91
15	26	94
16	8	67
17	13	74
18	11	68
19	14	81
20	37	102

Example: Linear Regression (cont.)

- Select Regression in Data Analysis (from ToolPak)
  - Y-axis: Response variable (eg, Relative %Change)
  - X-axis: Predictor variable (eg, Years Experience)
  - Options to select:
    - Residual plots
    - Normal probability plots



Example: Linear Regression (cont.)

- Additional information re: Options to select:
  - Residual plots: determine whether data fit linearity and homoscedasticity assumptions
    - Linearity: same # points above and below line (ie, mean = 0)
    - Homoscedasticity: random pattern, vertical spread should be the same across x-axis (not conical)
  - Normal probability plots: determine whether data fit a normal distribution, ie, should be straight line

Example: Linear Regression (cont.)

Demo

Example: Linear Regression (cont.)

- Relevant Excel output terminology:
  - Multiple R: Same as Pearson correlation coefficient (r)
  - R Square: AKA Coefficient of Determination, proportion of points that fall on regression line, or proportion of values that fit the model.
    - Size doesn't always matter! Humans are harder to predict (eg, vs. physical processes)
  - Adjusted R Square: Best used with multiple predictors
  - Coefficients: Values of m and b in the regression equation
  - Lower and Upper 95%: Confidence interval (CI)

Alliance for Continuing Education in the Health Professions

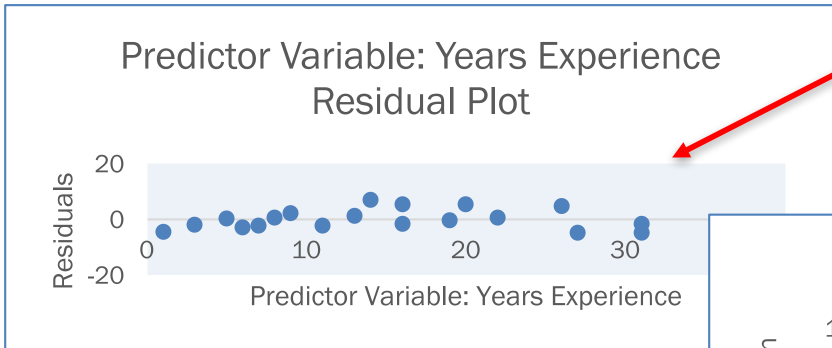
Example: Linear Regression (cont.)

Output:

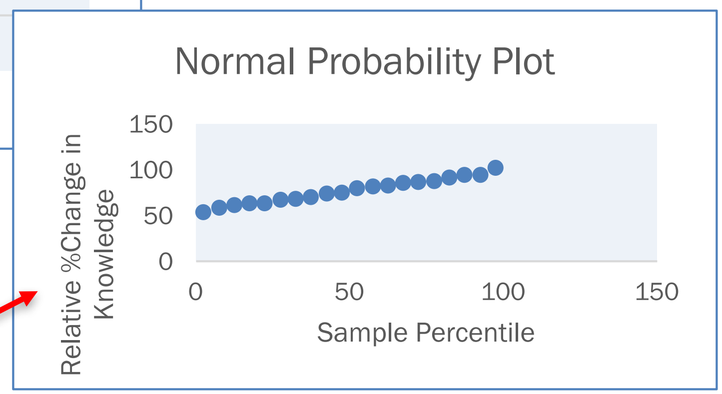
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.967043756							
R Square	0.935173626							
Adjusted R Squ	0.931572161							
Standard Error	3.560600778							
Observations	20							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	3291.998198	3291.9982	259.664766	3.86457E-12			
Residual	18	228.2018022	12.6778779					
Total	19	3520.2						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	56.14575701	1.503629898	37.3401441	1.6555E-18	52.98674781	59.3047662	52.9867478	59.3047662
Predictor Varia	1.276661056	0.07922625	16.114117	3.8646E-12	1.110212881	1.44310923	1.11021288	1.44310923

Example: Linear Regression (cont.)

Output (cont.):



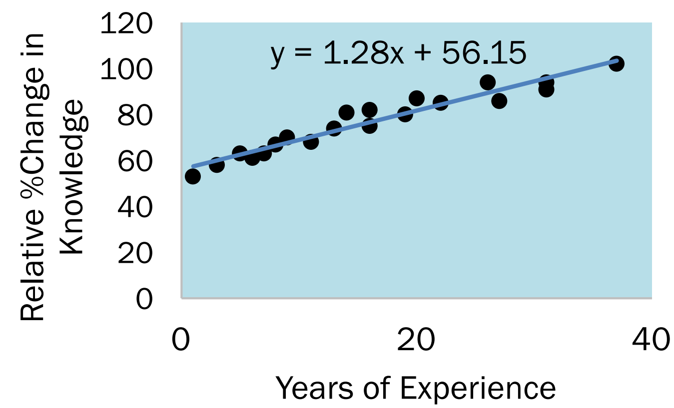
Shows linearity and homoscedasticity



Shows data follow normal distribution

Example: Linear Regression (cont.)

- Interpreting results:
  - Years of experience predictor *p*-value < .001
  - Regression equation: y = 1.28x + 56.15
    - 1.28 = slope: for every year of experience, change in percentage correct increases by 1.28 (95% CI: 1.11 – 1.44)
    - 56.15 = y-intercept: when years of experience = 0, change in percentage correct = 56.15 (often not interpretable/utilized)



Alliance for Continuing Education in the Health Professions

Example: Linear Regression (cont.)

- What to do with this information:
  - Years of experience was a significant predictor of change in knowledge
  - Can be interpreted in two ways:
    - If you had a clinician with several years' of experience, you would expect a greater increase in knowledge. Therefore future activities in Alzheimer's should target clinicians with more experience to maximize learning.
    - A clinician with less experience would have a smaller increase in knowledge for some reason, therefore further exploration would be warranted to determine why.

Alliance for
Continuing Education
in the Health Professions

Multiple Linear Regression

Multiple Linear Regression

- Same as linear regression, but with more than one predictor
- Equation (2 predictors):
  - $y = m_1x_1 + m_2x_2 + b$
- Our example:
  - Add a predictor for Years of Education

Partic. ID	Predictor Variable: Years Experience	Predictor Variable: Years of Education	Relative %Change in Knowledge
1	19	20	80
2	5	18	63
3	9	21	70
4	16	16	75
5	6	16	61
6	20	19	87
7	31	15	94
8	27	17	86
9	1	19	53
10	7	21	63
11	3	20	58
12	22	19	85
13	16	16	82
14	31	18	91
15	26	17	94
16	8	20	67
17	13	21	74
18	11	17	68
19	14	19	81
20	37	20	102

Multiple Linear Regression (cont.)

- Same assumptions as simple linear regression
- Similar output:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.967398467
R Square	0.935859794
Adjusted R Square	0.928313888
Standard Error	3.644386114
Observations	20

ANOVA

	df	SS	MS	F	Significance F
Regression	2	3294.413647	1647.206824	124.022182	7.2545E-11
Residual	17	225.7863526	13.28155015		
Total	19	3520.2			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	52.34719304	9.039253206	5.791097101	2.1705E-05	33.2760358	71.4183503	33.2760358	71.4183503
Predictor Variable: Years Experience	1.286422836	0.084259414	15.2674078	2.3383E-11	1.10865101	1.46419466	1.10865101	1.46419466
Predictor Variable: Years of Educatic	0.197365816	0.462804239	0.426456369	0.67512656	-0.7790658	1.17379741	-0.7790658	1.17379741

Multiple Linear Regression (cont.)

Demo

Multiple Linear Regression (cont.)

- A few burning questions…

Multiple Linear Regression (cont.)

- Did adding another predictor improve the model?
  - Compare Adjusted R Square between simple and multiple regression outputs.
    - Yes: Adjusted R Square is larger in multiple regression
    - No: Adjusted R Square is not larger in multiple regression
    - Our examples: Both Adjusted R Squares were ~.93, so adding the predictor did not improve the model

Simple Regression

Regression Statistics	
Multiple R	0.967043756
R Square	0.935173626
Adjusted R Square	0.931572161
Standard Error	3.566600778
Observations	20

Multiple Regression

Regression Statistics	
Multiple R	0.967398467
R Square	0.935859794
Adjusted R Square	0.928313888
Standard Error	3.644386114
Observations	20

Multiple Linear Regression (cont.)

- Why wasn't education a significant predictor of knowledge?
  - Most likely a violation of regression assumptions:
    - Linear relationship between variables
    - Multivariate normality
    - No or little multicollinearity
    - No auto-correlation
    - Homoscedasticity
  - Or, it just wasn't

Multiple Linear Regression (cont.)

- How do you know if predictor variables are correlated?
  - VIF (variance inflation factor)
  - Easier: correlation
    - Note: Excel calculates correlation, but doesn't automatically provide associated p-values
- What do you do if variables are correlated?
  - Drop one or more

Multiple Linear Regression (cont.)

- If more than one predictor is significant, which one has the most influence?
  - 3 Methods:
    - Standardize the independent variables prior to analysis; variable with highest regression coefficient (absolute value) has the largest effect.
    - Rerun the regression removing one independent variable from the model and compare Adjusted R Squares (run one regression per independent variable). The predictor variable associated with the greatest increase in Adjusted R Square is the strongest.
    - Easiest: Look at t-statistics in output; largest value corresponds to stronger predictor

Multiple Linear Regression (cont.)

- What can we do with this information?
  - Since years of experience predicts relative change in knowledge, future educational activities can target a subset of clinicians based on experience
  - Education did not predict change in knowledge, so future activities do not need to target clinicians based on this variable

Alliance for
Continuing Education
in the Health Professions

Logistic Regression

Logistic Regression

- Used when response variable is categorical (eg, correct/incorrect, good/fair/poor, yes/no)
- Calculates the probability* of something happening (eg, answering correctly = p in the equation) based on a set of predictors
- Provides odds ratio as output

*Strictly speaking, probability does not equal odds or odds ratio, which is what we're actually measuring; see later slides.

Logistic Regression (cont.)

- Formulas for the math nerds, like me:
  - Recall linear regression equation:

y = mx + b

More formally written as:

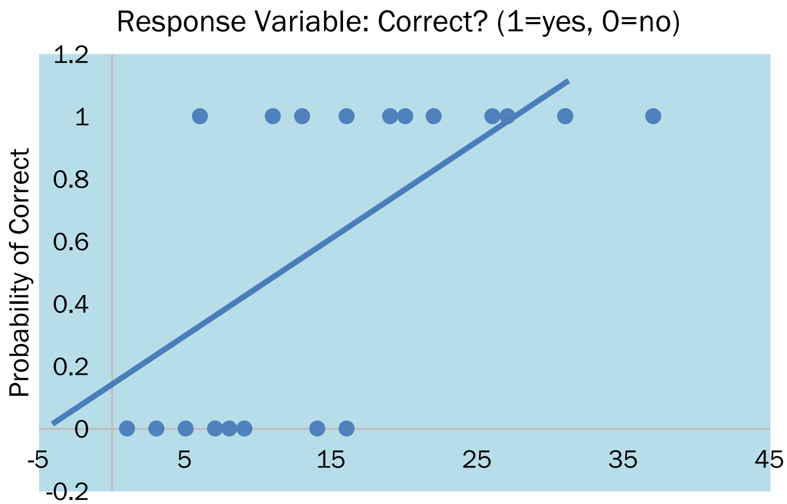
$$y = \alpha + \beta x$$

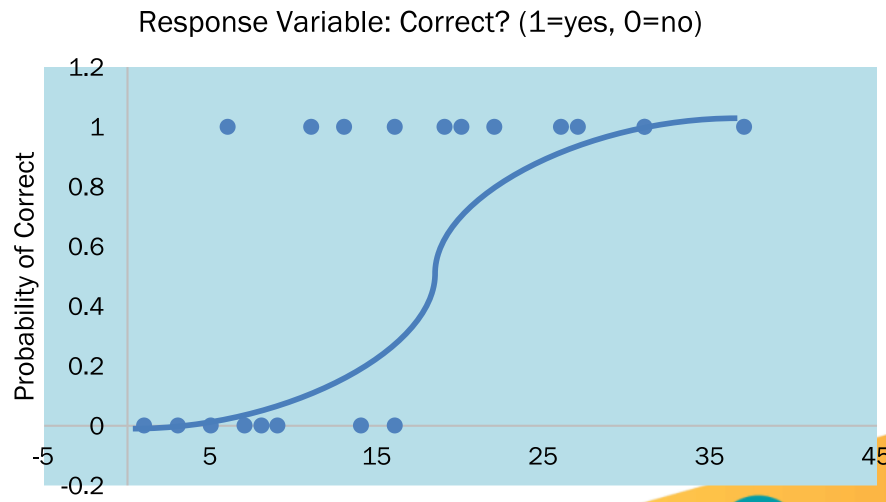
  - Logistic regression equation:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Logistic Regression (cont.)

Graphs of data





Logistic Regression (cont.)

- Probability versus odds versus odds ratio
  - Probability: #times an event occurs/#total events
  - Odds: probability that the event occurs divided by the probability that the event does not occur
  - Odds ratio: ratio of two odds

Logistic Regression (cont.)

- Probability versus odds versus odds ratio (cont.)
  - Example: coin flip
    - Probability (heads): #heads/#flips
      - Fair coin: 5/10 = .5
      - Unfair coin: 7/10 = .7
    - Odds: (#heads/#flips) / (#tails/#flips)
      - Fair coin: .5/.5 = 1
      - Unfair coin: .7/.3 = 2.33
    - Odds ratio (fair coin:unfair coin): odds(fair)/odds(unfair)
      - 1/2.33 = .43

Logistic Regression Assumptions

- Response variable is dichotomous (eg, yes/no, correct/incorrect)
- Response variable has mutually exclusive and exhaustive categories (each data point falls into only one category, all data points accounted for)
- Predictor variable is categorical or continuous
- Observations are independent
- Continuous predictor variables are linearly related to log odds (this is different from linearity assumption for linear regression)
  - Check using Box-Tidwell test
- Sample size rule of thumb: At least 10 (some say 30) cases per independent variable

Example: Simple Logistic Regression

- Standard educational activity/outcomes on Alzheimer's disease
- Research question: Does years of experience predict whether or not participants perform a behavior (eg, screen patients over age 60 for cognitive deficits) at least 50% of the time
  - Response options: 0% of the time, 1%-25%, 26%-50%, 51%-75%, 76%-100%
  - Coded as 1 for "51%-75%" and "76%-100%"; coded as 0 otherwise
  - Pre-test data only
- Regression model:
  - Response variable: Yes/no (perform or not perform behavior)
  - Predictor variable: Years of experience

Alliance for Continuing Education in the Health Professions

Example: Logistic Regression (cont.)

– Enter Data

- First column: Participant ID/#
- Second column: Predictor variable (Years Experience)
- Third column: Response variable (perform behavior, coded as 1s and 0s)

Participant ID	Predictor Variable: Years Experience	Response Variable: Perform Behavior >50% of the Time? (1=yes, 0=no)
1	19	1
2	5	0
3	9	0
4	16	1
5	6	1
6	20	1
7	31	1
8	27	1
9	1	0
10	7	0
11	3	0
12	22	1
13	16	0
14	31	1
15	26	1
16	8	0
17	13	1
18	11	1
19	14	0
20	37	1

Example: Logistic Regression (cont.)

Demo

Example: Logistic Regression (cont.)

Computing Demo (http://statpages.info/logistic.html)

Example: Logistic Regression (cont.)

Descriptives...
8 cases have Y=0; 12 cases have Y=1.

Variable	Avg	SD
1	16.1000	10.0494

Iteration History...
-2 Log Likelihood = 26.9205 (Null Model)
-2 Log Likelihood = 25.2230
-2 Log Likelihood = 22.5520
-2 Log Likelihood = 19.8437
-2 Log Likelihood = 17.6423
-2 Log Likelihood = 16.1194
-2 Log Likelihood = 15.2531
-2 Log Likelihood = 14.8964
-2 Log Likelihood = 14.8126
-2 Log Likelihood = 14.8049
-2 Log Likelihood = 14.8048
-2 Log Likelihood = 14.8048 (Converged)

Overall Model Fit...
Chi Square= 12.1157; df=1; p= 0.0005

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	-- High
1	0.2696	0.1202	0.0249	1.3094	1.0346	1.6572
Intercept	-3.2217	1.5736	0.0406			

Example: Logistic Regression (cont.)

- Output interpretation (relevant sections, circled on previous slide):
  - **Variable 1**: Years of experience
  - **Coefficient:** in log odds unit, so not interpretable without transformation
  - **p**: *p*-value for predictor (= .025, so significant)
  - **Odds Ratio (O.R.):** For every year increase in experience, the odds of performing the behavior increased by a factor of 1.31 (or by approx. 31%)
  - **Low – High**: 95% CI

Overall Model Fit...

Chi Square= 12.1157; df=1; p= 0.0005

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	– High
1	0.2696	0.1202	0.0249	1.3094	1.0346	1.6572
Intercept	-3.2217	1.5736	0.0406			

Example: Logistic Regression (cont.)

- What to do with this information:
  - Years of experience was a significant predictor of behavior
  - Possible next steps:
    - If only interested in this one question, then next steps would be similar to those for linear regression example
    - Alternatively, determine predictor(s) for other behavior questions to see if results are consistent
      - If consistent, next steps would be similar to those for linear regression
      - If not consistent, develop hypotheses for why not, then use the information to guide future education

Multiple Logistic Regression

- As with linear regression, multiple logistic regression has more than one predictor
- Use same method as for simple logistic regression, except:
  - Copy over 3+ columns of data rather than two
  - Specify 2+ predictor variables in online calculator
- Output interpreted similarly as well
- Quick data and output demo:
  - Same data as simple logistic, but added self-rating of knowledge as another predictor

Multiple Logistic Regression (cont.)

- What can we do with this information?
  - Since years of experience predicts behavior, future educational activities can target a subset of clinicians based on experience
  - Self-rating of knowledge did not predict behavior, so future activities do not need to target clinicians based on this variable

Alliance for Continuing Education in the Health Professions

Other Types of Logistic Regression

- Multinomial:
  - Used when the response variable has 3 or more unordered categories (eg, MD/PA/NP)
- Ordinal:
  - Used when response variable has 3 or more ordered categories (eg, good/fair/poor)
- Details are beyond the limits of this webinar…but it can be done using statistical software or online calculators

Alliance for Continuing Education in the Health Professions

Two Final Burning Questions

- How do you decide which activities on which to conduct predictive modeling?
  - If you can't do all, suggest starting with those that are less successful in terms of highest level of outcomes measured
- How do you decide which variables to include as predictors?
  - Personal preference
  - Research precedent
  - "Hypothesized" supporter interest
  - Quality of data
  - Non-correlated variables

Conclusions

- Maximizing educational impact depends on factors influencing success
- Regression is one method for determining these factors
- Selecting which type of regression depends on several factors, including type of data and model assumptions
- Information learned from the regression model can help in developing future activities

Thank you

Questions?

Contact info:

Jamie Reiter, PhD
jreiter@cmeoutfitters.com
Office: 614-328-4527
Mobile: 951-587-1508
https://www.linkedin.com/in/jamiereiterphd/