Probability Rules and Applications

Karen Grace-Martin

THE ANALYSIS FACTOR

1

Empirical Probability

Empirical probability is the relative frequency of a frequency distribution based upon observation

P(Event) = Frequency of Event / Number of Trials

Presence of suspected water source	**Presence of diarrhea**		**Total**
	Yes (1)	No (2)	
Yes (1)	78 (a)	1,422 (b)	1,500
No (2)	50 (c)	950 (d)	1,000
Total	128 (a+c)	2,372 (b+d)	2,500

2

Probability Rules

1. All probabilities are between 0 and 1 inclusive

The probability of an impossible event is 0

The probability of a sure event is 1

2. The sum of all the probabilities for all possible events is equal to one

Presence of suspected water source	Presence of diarrhea		Total
	Yes (1)	No (2)	
Yes (1)	78 (a)	1,422 (b)	1,500
No (2)	50 (c)	950 (d)	1,000
Total	128 (a+c)	2,372 (b+d)	2,500

4

3. The probability of an event not occurring is one minus the probability of it occurring.

$$P(A) = 1-P(\text{not } A)$$

5

4. Addition Rule

General Addition Rule

P(A or B) = P(A) + P(B) - P(A and B)

Presence of suspected water source	Presence of diarrhea		Total
	Yes (1)	No (2)	
Yes (1)	78 (a)	1,422 (b)	1,500
No (2)	50 (c)	950 (d)	1,000
Total	128 (a+c)	2,372 (b+d)	2,500

6

Mutually Exclusive Events

Two events are mutually exclusive if they cannot occur at the same time.

P(A and B) = 0

Specific Addition Rule

Only valid when the events are mutually exclusive.

P(A or B) = P(A) + P(B)

7

Conditional Probability

The probability that event B occurs, given that event A has already occurred is

P(B|A) =

Presence of suspected water source	Presence of diarrhea		Total
	Yes (1)	No (2)	
Yes (1)	78 (a)	1,422 (b)	1,500
No (2)	50 (c)	950 (d)	1,000
Total	128 (a+c)	2,372 (b+d)	2,500

5. Multiplication Rule

General Multiplication Rule

P(A and B) = P(A) * P(B|A) and

P(A and B) = P(B) * P(A|B)

Presence of suspected water source	Presence of diarrhea		Total
	Yes (1)	No (2)	
Yes (1)	78 (a)	1,422 (b)	1,500
No (2)	50 (c)	950 (d)	1,000
Total	128 (a+c)	2,372 (b+d)	2,500

9

Independent Events

Two events are independent if one occurring does not affect or influence the probability of the other occurring

Specific Multiplication Rule

If A and B are Independent:

$$P(A \text{ and } B) = P(A) * P(B)$$

Independence Revisited

The following four statements are equivalent:

- A and B are independent events
- P(A and B) = P(A) * P(B)
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$

11

Chi-Square Test of Independence

			Experienced Joint Pain		Total
			No	**Yes**	
Runs more that 25km/week	**No**	Count	215	75	290
		% of Non-runners	74%	**26%**	100%
	Yes	Count	785	380	1165
		% of Runners	67%	**33%**	100%
Total		Count	1000	455	1455

H_0: Joint Pain and Running are Independent

A = Experienced Joint Pain

B = Runs a lot

If A and B are independent, then P(A|B) = P(A)

P(A|B) =

12

Logistic Regression

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

$$Ln\left(\frac{\hat{P}}{1-\hat{P}}\right) = 4.32 - 6.59AL - 6.18FL - 7.11LA - 6.14MS$$

$$Ln\left(\frac{\hat{P}}{1-\hat{P}} \mid AL\right) = 4.32 - 6.59(1) - 6.18(0) - 7.11(0) - 6.14(0)$$

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	4.3197	0.4109	110.4913	<.0001
State	AL	1	-6.5884	0.5107	166.4152	<.0001
State	FL	1	-6.1804	0.5779	114.3788	<.0001
State	LA	1	-7.1100	0.5661	157.7236	<.0001
State	MS	1	-6.1382	0.4426	192.3230	<.0001

Sampling

Simple Random Sampling: Each individual has an equal probability of being sampled

P(sampled) = 1/N

Two-Stage Cluster Sampling: N_c Clusters are sampled, then samples of size n are randomly taken from each sample

Sampling Weight = 1/P(sampled)

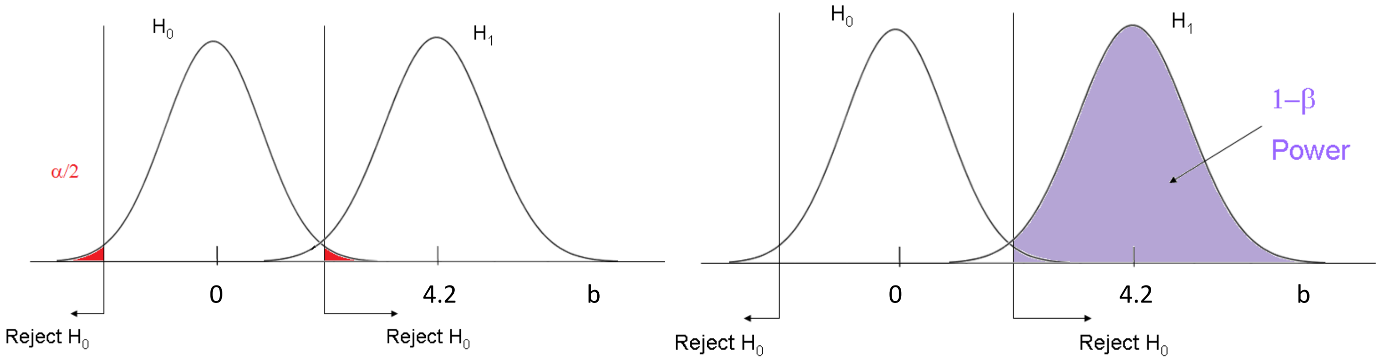
P(1st person in cluster 1 is sampled)

= P(sampled|cluster is sampled)*P(cluster is sampled)

= $(1/n)*(1/N_c)$

14

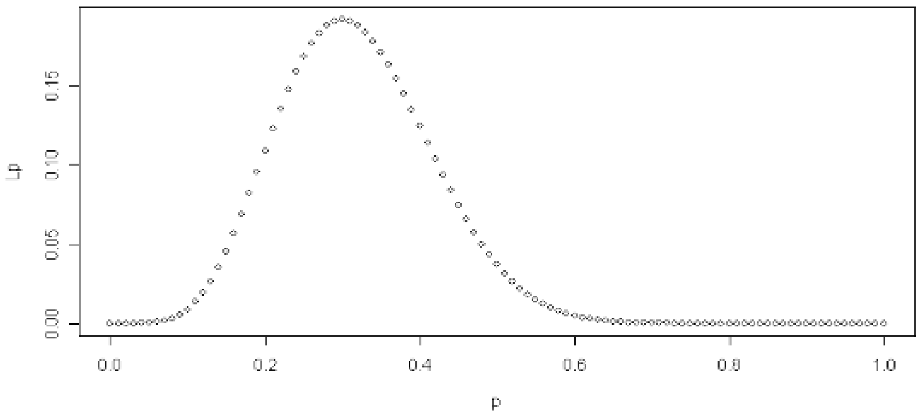
P values and Power



		Truth	
		H_0 True	H_0 False
Decision	Reject H_0	Type I error	**Correct**
	Accept H_0	Correct	Type II error

15

Maximum Likelihood Estimation



$$L(p) = P(p \mid x_i)$$

$$f(x) = \left(\frac{n!}{x!(n-x)!}\right) p^x (1-p)^{n-x}$$

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left(\frac{n!}{x_i!(n-x_i)!}\right) p^{x_i} (1-p)^{n-x_i}$$

16