Psychology 205: Research Methods in Psychology
The problem of base rates

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN
UNIVERSITY

October, 2015

Outline

Hypothesis testing using inferential statistics

- How likely are the observed data given the hypothesis that an Independent Variable has no effect.
- Bayesian statistics compare the likelihood of the data given the hypothesis of no differences as contrasted to the likelihood of the data given competing hypotheses.
  - This takes into account our prior willingness to believe that the IV could have an effect.
  - Also takes into account our strength of belief in the hypothesis of no effect
- Conventional tests report the probability of the data given the "Null" hypothesis of no difference.
- The less likely the data are to be observed given the Null, the more we tend to discount the Null.
  - Three kinds of inferential errors: Type I, Type II and Type III
  - Type I is rejecting the Null when in fact it is true
  - Type II is failing to reject the Null when it is in fact not true
  - Type III is asking the wrong question

Hypothesis Testing

Table : The ways we can be mistaken

		State of the World	
		True	False
Scientists says	True	Valid Positive	Type I error
	False	Type II error	Valid rejection

Type III error is asking the wrong question!

Hypothesis Testing α and β

Table : The ways we can be mistaken

		State of the World	
		True	False
Scientists says	True	Valid Positive	(Type I error) $= \alpha$
	False	p(Type II error) $= \beta$	Valid rejection

Power as 1-β
We need to think about power.
Type III error is asking the wrong question!
Probability that a "significant effect" is a type I error $= \frac{\alpha * False}{\alpha * False + (1-\beta) * True}$

Consider a number of scenarios

1. 1000 studies all together
    - Consider $\alpha = .05, .01$
    - Consider $1 - \beta = .5, .8, .95, 1.0$
2. But, also consider the state of the world (What is the a priori likelihood of the outcome)
    - a 50-50 chance (boring result)
    - a 20-80 chance (interesting finding)
    - a 10-90 chance (very interesting finding)
    - a 1-99 chance (Wow, you found that!)
3. Probability that a "significant effect" is a type I error = $\frac{\alpha * \mathit{False}}{\alpha * \mathit{False} + (1-\beta) * \mathit{True}}$

Hypothesis Testing α and β

Table : The ways we can be mistaken

		State of the World	
		True	False
Scientists says	True	VP $= (1-\beta) * \textit{True}$	p(Type I)$= \alpha$ * False
	False	p(Type II) $= \beta * \textit{True}$	VR $= (1-\alpha) * \textit{False}$

Power as 1-β
We need to think about power.
probability that a "significant effect" is a type I error = $\frac{\alpha * \textit{False}}{\alpha * \textit{False} + (1-\beta) * \textit{True}}$

Hypothesis Testing α and β likely event, high power

Table : A 50 - 50 chance – high power

		State of the World		
		True	False	Total
Scientists says	True	475	25	500
	False	25	475	500
	Total	500	500	1000

p (False Positive — finding was significant) $= \frac{25}{25+475} = .05$

Hypothesis Testing α and β likely event, 50% power

Table : A 50 - 50 chance – low power

		State of the World		
		True	False	Total
Scientists says	True	250	25	275
	False	250	475	725
	Total	500	500	1000

p (False Positive — finding was significant) $= \frac{25}{25+250} = .09$

Hypothesis Testing α and β unlikely event, 50% power

Table : A 10 - 90 chance – low power

		State of the World		
		True	False	Total
Scientists says	True	50	45	95
	False	50	855	905
	Total	100	900	1000

p (False Positive — finding was significant) $= \frac{45}{45+50} = .47$

Hypothesis Testing α and β very unlikely event, 80% power

Table : A 1 - 99 chance – good power

		State of the World		
		True	False	Total
Scientists says	True	8	49.5	57.5
	False	2	940.5	942.5
	Total	10	990	1000

p (False Positive — finding was significant) $= \frac{49.5}{49.5+8} = .86$

Hypothesis Testing α and β very unlikely event, perfect power

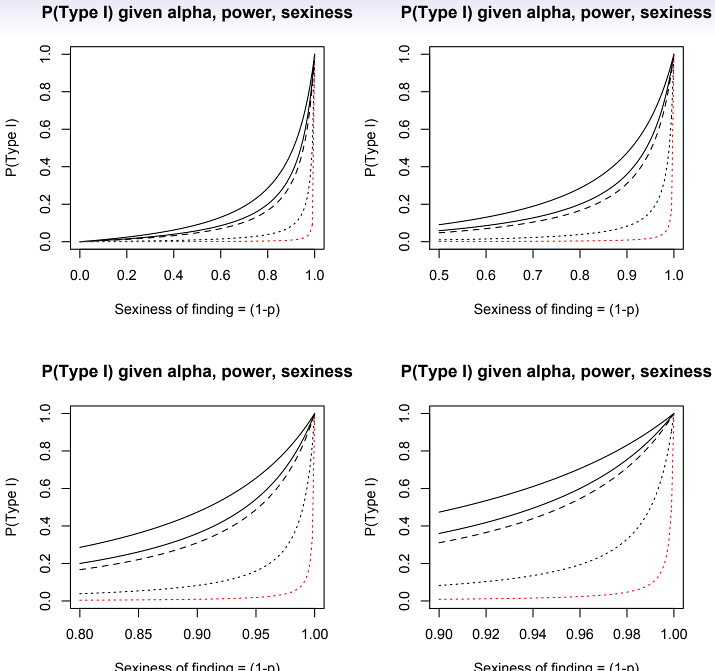
Table : A 1 - 99 chance – perfect power

		State of the World		
		True	False	Total
Scientists says	True	10	49.5	59.5
	False	0	940.5	94.5
	Total	10	990	1000

p (False Positive — finding was significant) $= \frac{49.5}{49.5+10} = .83$

Power, α level, and the excitement of the finding

1. There is a natural tendency to want to show the unlikely.
   - Showing your grandmother is right is not as interesting as showing she is wrong
   - Showing that what most people expect is wrong is very exciting
2. But, the less likely the effect is to be there, the more likely that a "significant effect" is actually a type I error.
   - Need to increase our Power and be sensitive to the replicability of our results.
3. The power of a good graphic to show the problem.
   - Five lines: $alpha = .05, .01, .001$
   - Power $= .8$ or 1


P(Type I) given alpha, power, sexiness
P(Type I)
Sexiness of finding = (1-p)
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
P(Type I) given alpha, power, sexiness
P(Type I)
Sexiness of finding = (1-p)
0.0 0.2 0.4 0.6 0.8 1.0
0.5 0.6 0.7 0.8 0.9 1.0
P(Type I) given alpha, power, sexiness
P(Type I)
Sexiness of finding = (1-p)
0.0 0.2 0.4 0.6 0.8 1.0
0.80 0.85 0.90 0.95 1.00
P(Type I) given alpha, power, sexiness
P(Type I)
Sexiness of finding = (1-p)
0.0 0.2 0.4 0.6 0.8 1.0
0.90 0.92 0.94 0.96 0.98 1.00