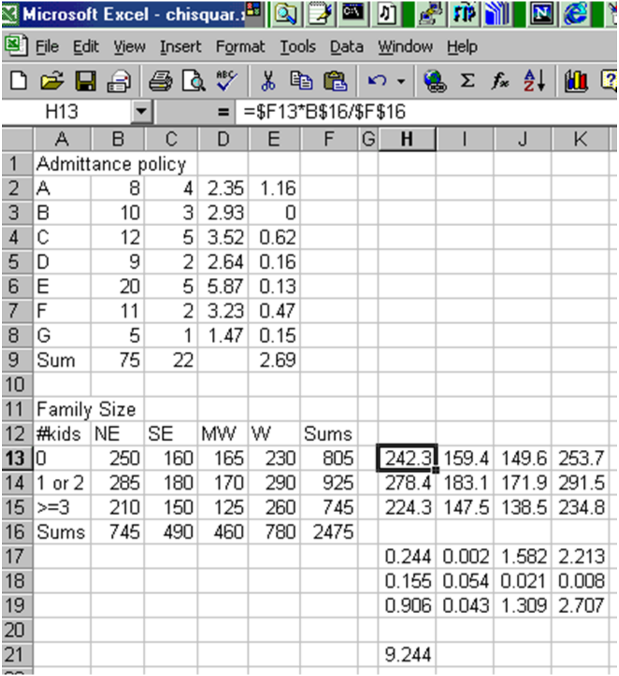
Unit 11: The Chi-Squared Test



Microsoft Excel - chisquar.x

File Edit View Insert Format Tools Data Window Help

H13 = =$F13*B$16/F16

	A	B	C	D	E	F	G	H	I	J	K
1	Admittance policy										
2	A	8	4	2.35	1.16						
3	B	10	3	2.93	0						
4	C	12	5	3.52	0.62						
5	D	9	2	2.64	0.16						
6	E	20	5	5.87	0.13						
7	F	11	2	3.23	0.47						
8	G	5	1	1.47	0.15						
9	Sum	75	22		2.69						
10											
11	Family Size										
12	#kids	NE	SE	MW	W	Sums					
13	0	250	160	165	230	805		242.3	159.4	149.6	253.7
14	1 or 2	285	180	170	290	925		278.4	183.1	171.9	291.5
15	>=3	210	150	125	260	745		224.3	147.5	138.5	234.8
16	Sums	745	490	460	780	2475					
17								0.244	0.002	1.582	2.213
18								0.155	0.054	0.021	0.008
19								0.906	0.043	1.309	2.707
20											
21								9.244			

Type I and type II errors

A "type I" error in a sig test is to reject H_0 when it is true.
Many "discoveries" are type I errors.
By def, α-value is P(type I error — H_0 is true)
A "type II" error is to fail to reject H_0 when it is false.
We never "accept H_0" — no science is exactly right. (E.g., Einstein corrected Newton.)

Example: α-level and type I or II

Flip a coin 10 times, get 8 heads. With an H_0 of a fair coin,

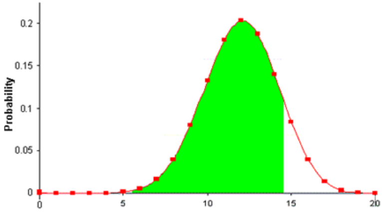
$$P(\text{count} \geq 8) = \frac{C(10,8) + C(10,9) + C(10,10)}{210} \approx 5.5\%$$

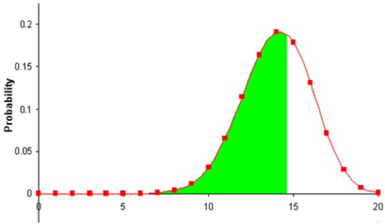
So with $\alpha = 10\%$, we reject null, while
with $\alpha = 5\%$, we do not reject null.
Therefore, if the coin is fair, 10% makes a type I error, and 5% yields the correct answer,
while if the coin is unfair, 10% yields the correct answer, and 5% makes a type II error.

The "power" of a test is $P(\text{no type II error} \mid H_0 \text{ is false})$ $= 1 - P(\text{type II} \mid H_0 \text{ false})$, but ...

... we can't compute the power because it depends on <u>how</u> false H_0 is, and usually we don't even know whether it is false.

Ex: Test a coin for fairness ($H_0 : P(\text{head}) = 0.5$) with 20 flips. With $\alpha = 5\%$, test says coin unfair if #heads ≤ 5 or ≥ 15. Sample (binomial) distributions when $P(\text{head}) = 0.6$ and 0.7: $P(\text{type II error}) = 87\%$ and 58% respectively.





Another example: An oracle speaks

Valesky vs. Brown for State Senate, interviewing 400 people. H_0: $p = 50\%$ of voters for Val; H_a: $p > 50\%$

EV of (sample) % = .5, SE of % $= \sqrt{\sqrt{.5(.5)}400} = .025$

For sig, need $z = ((\#/400) - .5)/.025 \geq 1.96$, i.e., $\# \geq 220$

Oracle reveals true $p = 52\%$: EV = .52, SE $= \sqrt{.52(.48)}400 = .025$: $P(\text{Type II error}) = P(z \leq ((220/400) - .52)/.025 = 1.2) = 89\%$

Oracle reveals true $p = 55\%$: EV = .55, SE $= \sqrt{\frac{.55(.45}{400}} = .025$: $P(\text{Type II error}) = P(z \leq ((220/400) - .55)/.025 = 0) = 50\%$

Still, . . .

As P(Type I error) [rejecting H_0 when you shouldn't] goes up — maybe by picking a larger α or using a larger sample — P(Type II error) [failing to reject H_0 when you should] goes down, and v.v.