**Explaining the “standard deviation” formulas for CI and HT.**

Two Independent Means:

Recall that for independent random variables that . Earlier we have used (and in HW25 will give evidence of this) that the standard deviation of calculating a sample mean of X was . This is actually called the standard error of the mean instead of the standard deviation of the mean since it is an approximation (the standard deviation of the mean is actually . Squaring gives the variance which is . Using the formula above we get  +. To get the standard deviation (again it’s actually the standard error because we are approximating the population variances with the sample variances), take the square root to get

.

Binomial: Recall the formula  from early on. Suppose we have 1 trial with chance of success *p*. Then the mean is (1)(*p*) = *p*. What is the variance in the number of successes? The number of success is either 1 (with probability *p*) or 0 (with probability *q*) , so = 

But we also have  which if used *n* times with *X* =1 binomial trial we get 

One Proportion:

Recall the binomial. *X* = how many successes, we are interested in *p’* = proportion of successes. The difference is that *p’* is *X/n*, so using the rules for means and variances, the mean of *p’* is that of *X* divided by *n*, and the variance of *p’* is that of *X* divided by  . Recall  and.

|  |  |  |
| --- | --- | --- |
|  | *X* = binomial = how many successes | *p’* = sample proportion of successes |
| mean | *np* | *p* |
| variance | *npq* | *pq/n* |
| Standard deviation |  |  |
| Normal (*z)* is a good approximation if | *np* and *nq* >10 | *np* and *nq* > 10 |

Two Proportions:

Recall that for independent random variables that . Earlier we have seen that the standard deviation of calculating a sample proportion was . Squaring gives the variance which is. Using the formula above we get  . To get the standard deviation take the square root to get . Usually the subscripts are written with 1’s and 2’s instead such as .

**Explaining the details of “O and E stuff”.**

We need to recall some probability facts.

Probability Flashback 1: The mean of a binomial is *np*, like the example if 100 fair coins are tossed on average we expect to see 100(.5) = *np* heads.

Probability Flashback 2: If A and B are independent then P(A|B)=P(A)

Probability Flashback 3: P(A|B)=P(A and B)/ P(B)

Probability FB 2 and 3 combined yield that if A and B are independent then P(A and B)=P(A)P(B)

Example of finding an E:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Col 1 | Col 2 | totals |
| Row 1 |  |  | R1 tot |
| Row 2 |  |  | R2 tot |
| Row 3 |  | E | R3 tot |
| totals | C1 tot | C2 tot | *n* = total |

E = *n*(P(R3 and C2)) by Probability FB 1

= *n*P(R3)P(C2) by Probability FB 2 and 3

 combined (see above)

= *n*((R3 tot)/*n*)((C2 tot)/*n*) because our best estimate of

 a probability is how many

 times it occurs out of how

 many trials

= (R3 tot)(C2 tot)/*n* by canceling an *n*