**HUYNH DO**

**Module 3, Part 2/2**

1. **State the purpose of the exercise**

This exercise aimed to determine whether there is a statistically significant relationship between age and diabetes. If yes then what does the relationship look like for example strong, moderate, or weak?

**Individual Sub Analysis:**

1. **Chi-Square Test:** To investigate if there is a significant association between diabetes diagnosis and age across different age groups.
2. **Post-Hoc:** To perform additional tests on individual age groups to identify specific patterns or age ranges where the association is most significant.
3. **Control Factor**: The exercise also applies an additional control **PHYSACTS** factor (to both Chi-Square test and Post-Hoc Analysis) to see whether the ability to perform everyday physical activities plays any role in this relationship.
4. **State how to calculate the statistical technique**
* Chi-Square Test Example:
* **Step 1**: Calculate expected value for each age group



With:

* Total Yes (Column Total): 153
* Total No (Column Total): 1254
* Grand Total: 1407
* At age 23:
**E** “Yes”: (23 \* 153 ) / 1407 = 2.50
**E** “No”: (23 \* 1254)/1407 = 20.49
* At age 86:

**E** “Yes”: (2 \* 153) **/**1407 = 0.21

**E** “No”: (2 \* 1254) **/**1407 = 1.78

* **Step 2**: Calculate Chi-Square each data point
* At age 23:
Chi-Square “Yes”: ( 0 – 2.50 ) 2 / 2.50 = 2.50
Chi-Square “No”: (23 – 20.49) 2 / 20.49 = 0.30
* At age 86:

Chi-Square “Yes”: (1 – 0.22) 2 **/** 0.22 = 2.765

Chi-Square “No”: (1 – 1.78) 2 **/**1.78 = 0.342

* **Step 3**: Sum all Chi-Square Test



Where Oi​ is the observed frequency and Ei​ is the expected frequency.

Apply the same similar Chi-Square Test technique for each age from age 18 to 89:

χ2 = 2.50 + 0.307 + …. + 2.765 + 0.342 +... = **144.861**

* Standardized Residual Example:

Expect Counts **(E) = (**Row Total × Column Total for ’Diabetes: Yes’)**/**Grand Total



* O is the observed count for "Diabetes: Yes"
* E is the expected count
* Grand Total = 1407



****

1. Age 23:



**E** = (23 x 153) / 1407 = 2.5 |

**R** = (0 – 2.5) / Sqroot (2.5) = -1.58

1. Age 86
**
E** = (2 x 153) / 1407 = 0.2174

 **R** = (1 – 0.2174) / Sqroot( 0.2174 ) = 1.68

1. **Interpret the findings**

Overall, all data analysis confirm that there is a clear association between age and the likelihood of being told one has diabetes, especially starting from the middle age group, the trend starts to move upward, especially high among the elderly.

The research also found that there’s a corresponding relationship between the ability to perform everyday physical actives and the diabetes variable as well.