**DIABETES INFERANCE  BASED ON THE  
 UNITED STATES REPORT (2022)**Diabetes is one of the most prevalent non-communicable diseases worldwide. Three types of diabetes can occur – Type 1, Type 2, and Type 3 . These diseases occur due to an insufficient production of the hormone insulin or due to an inability of the body to use insulin effectively. Insulin is critical for the body’s cells to uptake glucose for use as energy . Diabetes can lead to many complications such as heart disease, renal failure, blindness, and stroke .It has become a major health concern due to the increase in new diabetic patients and premature deaths due to diabetes. In 2013, half of diabetes related deaths were in patients under age 60 .

* Diabetes is one of the leading causes of morbidity and mortality and it can result in several complications such as kidney failure, heart failure, stroke, and blindness making it a major medical and public health concern in the United States. Statistical methods are important to detect risk factors and identify the best sampling plan to determine predictive bounds for diabetic patients’ data. The main objective of this paper is to identify the best fit bootstrapping sampling method and to draw the predictive bound considering diabetes patient data

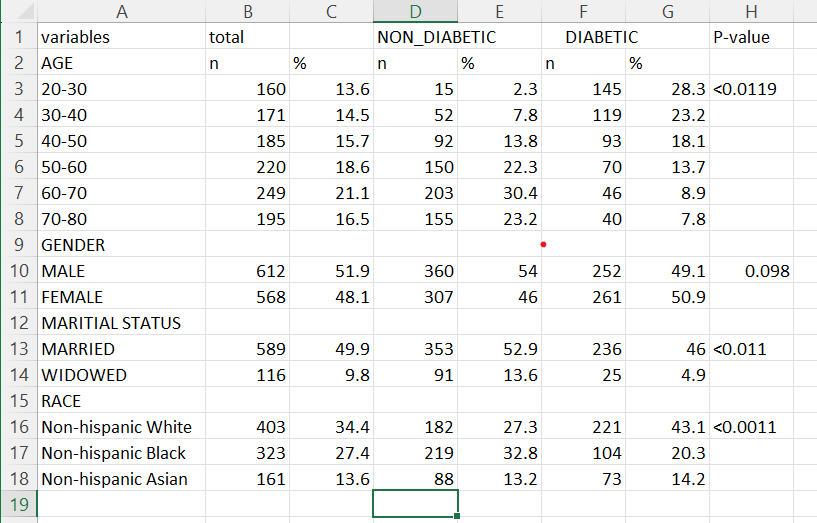
**RISK FACTORS**

DIABETES can happen at any age to any person, regardless to age gender or food consumption.In our table we can consider

1.AGE

2.GENDER

3.MARITAL STATUS

4.RACE .

**AGE :**

Age is an important factor in the development of diabetes and though there has been an increase in the prevalence of diabetes among adolescents, research over the last 20 years indicates that some risk factors increase with age.

P=T.TEST(array1,array2,tail)

A pooled analysis that studied the effects of age on various risk factors related to the development of diabetes found that the effect of body mass index (BMI) on diabetes increased with age, especially among western cultures . However, the effects of age and the development of diabetes and other comorbidities can be slowed by other factors, such as marital status. A separate study found that for married patients already diagnosed with diabetes their quality of life declined at a much slower rate than unmarried diabetes patients .

We divide the entire table into small parts to understand each factors effect.

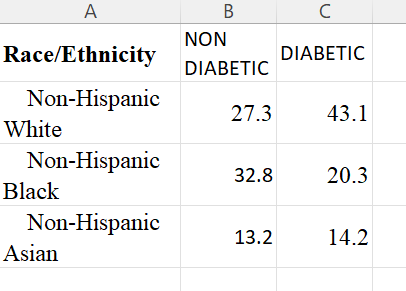
From this age chart we can clearly identify that as age increaser the rate of diabetes increases might be due to food consumption and obesity ,stress level etc.

Among the patients, the highest number of diabetic patients (28.3%) was observed in the age group 20–30 and the lowest number of diabetic patients was observed in the age group 70–80

***RACE :***

This data is from U.S and in U.S people reside from many different parts of world .in which three main categories are White , Black and Asian ,this concept is considered because each community have their own eating habits ,and these eating habits can effect the BMI of an human .

In Whites the chances of diabetes are more as per the diagram and data. As their diet contains less protein and obesity due to low physical workouts.



**MARITAIAL STAUTS AND GENDER**:

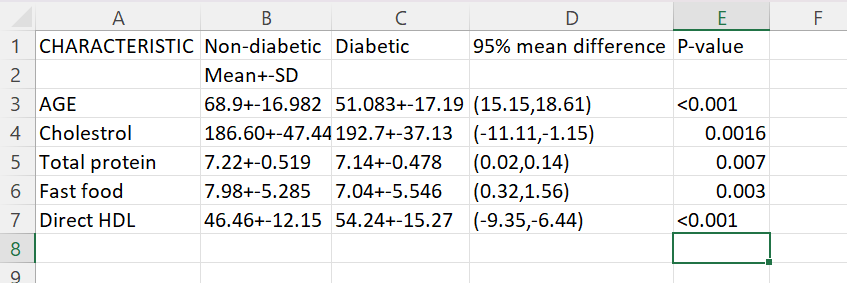
Here we can see that married humans are much effects or have diabetes as compared to unmarried or widowed.

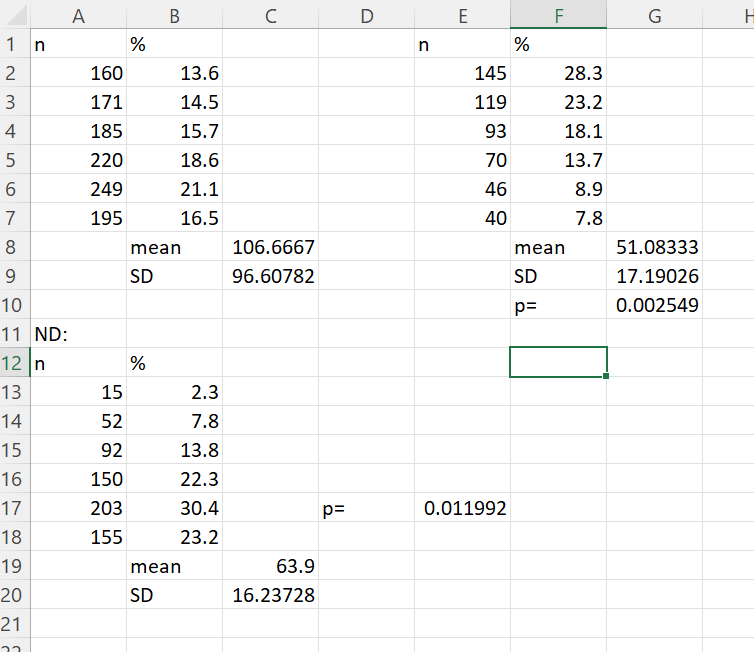
Whereas if we remove marital status and only look at gender regardless of age , its quiet prominent to see that diabetes in male prevails and by a certain age the count goes down but in female the count slowly increases by age and exceeds the count of male diabetes. This can also be because usually male are said to have diabetes through type 3 that’s genetically and female gain diabetes after child birth as per some medical believes.

A blue pie chart with white text

Description automatically generated

The following TABLE shows the summary of statistics for the continuous variables, age, cholesterol, total protein, fast food, and direct HDL. The total number of diabetic patients was 513 and non-diabetic patients 667. There was a statistical significance difference between the means of age for the diabetic and non-diabetic patients (p<0.0001). Similarly, there was a statistical significance difference between the means of cholesterol, total protein, fast food, and direct HDL for the diabetic and non-diabetic patients .





A logistic regression model was used to calculate the estimated coefficients; gender age, ethnicity, marital status, cholesterol (mg/dL), total protein (mg/dL), time spent eating fast food (in hours), and direct HDL (mg/dL). The results indicate that gender, ethnicity, marital status, and cholesterol were not statistically significant at alpha 0.05 (α=0.05). In some instances, when no significant effect is found, it may be due to a lack of sufficient data rather than no correlation between the variables. In this study, the sample size is large; while this may be a desirable attribute, it may also hinder the differences between covariates. The remaining covariates, age, total protein (g/dL), time spent eating fast food (in hours), and direct HDL (mg/dL) were found to be significant at α = 0.05.

[Table](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5429869/table/T3/) 2 shows the summary statistics (estimates, standard error, z-statistic, and p-value) for the risk factors using a logistic regression model. From the table, four of the eight variables were statistically significant at α=0.05: age, total protein, fast food time (in hours), and direct HDL. The remaining variables; gender, ethnicity, marital status, and cholesterol were not significant at α = 0.05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Standard Error** | **z-statistic** | **P-value\*** |
| (Intercept) | −8.1276563 | 1.2389582 | −6.56 | <.001 |
| Gender | 0.1642385 | 0.1520694 | 1.08 | 0.2801 |
| Age | 0.0837226 | 0.0054062 | 15.486 | <.001 |
| Ethnicity | 0.0834528 | 0.0489055 | 1.706 | 0.0879 |
| Marital Status | 0.0094733 | 0.0321231 | 0.295 | 0.7681 |
| Cholesterol (mg/DL) | 0.0009353 | 0.0017872 | 0.523 | 0.6008 |
| Total Protein (g/dL) | 0.867148 | 0.1576945 | 5.499 | <.001 |
| Fast food (Hours) | 0.0295454 | 0.0139276 | 2.121 | 0.0339 |
| Direct HDL (mg/dL) | −0.61729 | 0.0060564 | −10.221 | <.001 |

**Conclusions and future directions**

We found there was a statistical significance difference between age group, marital status, and race/ethnicity and diabetic status. There was a statistical significance difference between the means of age, means of cholesterol, mean total protein, mean fast food, and mean direct HDL for the diabetic and non-diabetic patients. The 95% confidence interval for each of the mean differences are reported.

Our findings suggest that four variables, (age, total protein (g/dL), time spent eating fast food (in hours), and direct HDL (mg/dL) of the tested covariates were statistically significant using the logistic regression model. The summary results for estimated coefficients for each covariate were determined by using the logistic regression model. From this table we are able to see that age, total protein, time in hours spent eating fast food, and direct HDL are significant predictors for diabetes at α =0.05 level of significance.