What factors effect diabetes diagnosis and to what degree based on this sample?

Hypothesis testing:

Due to the Central Limit Theorem and that the distribution approaches more of a normal distribution in samples over 50 I used Z test rather than T test to check if each parameters has an effect on the diagnosis of diabetes or not.

Using a significance level = .05 (95 CI)

The results are as follows for each parameter:

Number of Pregnancies Effect on Diabetes

Null: Number of Pregnancies does not have an effect on diabetes

Alternative Hypothesis: Number of Pregnancies does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$Pregnancies and yes\_diabetes$Pregnancies

z = -5.907, p-value = 3.485e-09

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.087834 -1.047509

sample estimates:

mean of x mean of y

3.298000 4.865672

Result: According to the small p value less than .05, having a more pregnancies does in fact give a higher level of frequencies

Glucose Level Effect on Diabetes

Null: glucose level does not have an effect on diabetes

Alternative Hypothesis: higher or lower glucose levels does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$Glucose and yes\_diabetes$Glucose

z = -13.752, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-35.73534 -26.81958

sample estimates:

mean of x mean of y

109.9800 141.2575

Result: The glucose level higher or lower than average effects the occurrence of diabetes.

Blood Pressure Effect on Diabetes

Null: blood pressure does not have an effect on diabetes

Alternative Hypothesis: higher or lower blood pressure does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$BloodPressure and yes\_diabetes$BloodPressure

z = -1.7131, p-value = 0.0867

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-5.6618014 0.3805476

sample estimates:

mean of x mean of y

68.18400 70.82463

Result: Reject alternative hypotheses. Blood pressure does not effect the occurrence of diabetes.

Skin Thickness Effect on Diabetes

Null: Skin thickness does not have an effect on diabetes

Alternative Hypothesis: higher or lower skin thickness does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$SkinThickness and yes\_diabetes$SkinThickness

z = -1.9706, p-value = 0.04877

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-4.98689009 -0.01346812

sample estimates:

mean of x mean of y

19.66400 22.16418

Result: The p-value is .048 and barely lower than the confidence interval of 95 to accept the alternative hypotheses and reject the null hypotheses. A thicker skin is more likely to result in the occurrence of diabetes.

Insulin Levels Effect on Diabetes

Null: Insulin level does not have an effect on diabetes

Alternative Hypothesis: higher or lower insulin level does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$Insulin and yes\_diabetes$Insulin

z = -3.3009, p-value = 0.0009638

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-50.27352 -12.81413

sample estimates:

mean of x mean of y

68.7920 100.3358

Result: Reject the Null hypotheses and accept the alternative hypotheses. Saying that insulin level does have an effect on Diabetes. Showing that a higher level of insulin shows more likely the occurrence of diabetes.

BMI Effect on Diabetes

Null: BMI does not have an effect on diabetes

Alternative Hypothesis: higher or lower BMI does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$BMI and yes\_diabetes$BMI

z = -8.6193, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-5.938537 -3.738138

sample estimates:

mean of x mean of y

30.30420 35.14254

Result: The null hypotheses is rejected and the alternative hypotheses is accepted. A larger BMI increases the occurrences of diabetes.

The Diabetes Pedigree Function Effect on Diabetes

Note: The diabetes pedigree function is a function in the instance that is dependent upon the patients relationship with other people with diabetes. Here it is not said what function in particular they used to collect the data.

Null: The Diabetes Pedagree Function does not have an effect on diabetes

Alternative Hypothesis: higher or lower Diabetes Pedigree Function does does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$DiabetesPedigreeFunction and yes\_diabetes$DiabetesPedigreeFunction

z = -4.5768, p-value = 4.721e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.17248257 -0.06904943

sample estimates:

mean of x mean of y

0.429734 0.550500

Result: The Diabetes Pedigree Function does have an effect on the number of occurrences of diabetes.

Age Effect on Diabetes

Null: Age does not have an effect on diabetes

Alternative Hypothesis: higher or lower Age does have an effect on diabetes

Accept Alternative Hypotheses if p-value is less than .05

Two-sample z-Test

data: no\_diabetes$Age and yes\_diabetes$Age

z = -6.9207, p-value = 4.493e-12

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-7.541586 -4.212742

sample estimates:

mean of x mean of y

31.19000 37.06716

Result: The null hypotheses is rejected and the alternative hypothesis is accepted. An increase in age has an effect on diabetes diagnosis.