a)

it basically represents the odds that an outcome will occur given a particular exposure.

|  |
| --- |
| > logistic.display(logistic)  Logistic regression predicting test    crude OR(95%CI) adj. OR(95%CI) P(Wald's test) P(LR-test)  age (cont. var.) 1.04 (1.03,1.06) 1.01 (1,1.03) 0.111 0.112    bmi (cont. var.) 1.1 (1.07,1.12) 1.09 (1.06,1.13) < 0.001 < 0.001    diastolic (cont. var.) 1.01 (1,1.02) 0.99 (0.98,1) 0.011 0.011    diabetes (cont. var.) 2.95 (1.87,4.67) 2.57 (1.43,4.63) 0.002 0.001    glucose (cont. var.) 1.04 (1.03,1.05) 1.04 (1.03,1.04) < 0.001 < 0.001    insulin (cont. var.) 1.0023 (1.001,1.0036) 0.9988 (0.997,1.0006) 0.186 0.187    pregnant (cont. var.) 1.15 (1.1,1.2) 1.13 (1.06,1.2) < 0.001 < 0.001    triceps (cont. var.) 1.0099 (1.0005,1.0194) 1.0006 (0.9872,1.0142) 0.929 0.929    Log-likelihood = -361.7227  No. of observations = 768  AIC value = 741.4454 |
|  |
| |  | | --- | |  | |

We can see that, the p value for diastolic is less than 0.05, and the OR is also close to 0. We can also say that, the exposure is associated with lower odds of outcome

b)

Looking at the output,

|  |
| --- |
| > logistic2 <- glm(test ~ age + triceps + insulin,family=binomial(logit),data=pima)  > summary(logistic2)  Call:  glm(formula = test ~ age + triceps + insulin, family = binomial(logit),  data = pima)  Deviance Residuals:  Min 1Q Median 3Q Max  -1.8627 -0.8802 -0.7246 1.2443 1.8638  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -2.4909109 0.2822754 -8.824 < 2e-16 \*\*\*  age 0.0451287 0.0067562 6.680 2.4e-11 \*\*\*  triceps 0.0073476 0.0055525 1.323 0.18573  insulin 0.0022292 0.0007602 2.932 0.00336 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 993.48 on 767 degrees of freedom  Residual deviance: 933.26 on 764 degrees of freedom  AIC: 941.26  Number of Fisher Scoring iterations: 4 |
|  |
| |  | | --- | | > | |

We can say that age and insulin are essential variables, while triceps can be dropped. We look at the p values and see if it is less than 0.05. In our case, the p value for both insulin and age is low and hence, they are essential, where as p value of triceps is higher than others and thus, it can be dropped

c) Looking at the output,

|  |
| --- |
| > logistic3 <- glm(test ~ insulin + diastolic,family=binomial(logit),data=pima)  > summary(logistic3)  Call:  glm(formula = test ~ insulin + diastolic, family = binomial(logit),  data = pima)  Deviance Residuals:  Min 1Q Median 3Q Max  -1.5500 -0.9156 -0.8597 1.3684 1.7356  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -1.2554389 0.3077945 -4.079 4.53e-05 \*\*\*  insulin 0.0022158 0.0006544 3.386 0.000709 \*\*\*  diastolic 0.0064330 0.0042346 1.519 0.128724  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 993.48 on 767 degrees of freedom  Residual deviance: 978.41 on 765 degrees of freedom  AIC: 984.41  Number of Fisher Scoring iterations: 4 |
|  |
| |  | | --- | | > | |

We can say that they are not equal