Hw8公衛二 梁嫚芳 b07801003

1. (c)

從所繪圖中可看出，S0(t)和S1(t)有顯著的差別，S0(t)圖較高，且下降至x軸(存活率=0)的時間較晚，即曲線下所圍面積(E[T|X\*=0])較大，因此WBC對remission time有相當的影響，並以X\*=0 have better experience of remission time。

Question2. (Extra)

|  |
| --- |
| data<-read.csv(file.choose())  data$x<-ifelse(data$WBC<median(data$WBC),0,1)  data$t=ifelse(data$Remission.Time<median(data$Remission.Time),0,1) |

(a) χ2-test for T\* and X\*

|  |  |
| --- | --- |
| Code | chisq.test(data$t,data$x) |
| output |  |
| 假設 | H0∶ T\*與X\*獨立  H1∶ T\*與X\*不獨立 |
| 結論 | p-value = 0.0006775 < α (0.05)  在95%的信心水準下拒絕H0，T\*與X\*不獨立。即分組後的 remission time及分組後的 WBC之間具有統計顯著的相關性。 |

(b) Two-sample t-test for T|X\* = 0 and T|X\* = 1

|  |  |
| --- | --- |
| Code | var.test(data$Remission.Time~data$x)  t.test(data$Remission.Time~data$x,var.equal=F) |
| output |  |
| F-test假設 | H0∶ 兩母體變異數相同  H1∶ 兩母體變異數不同 |
| F-test結論 | p-value = 0.001568 < α (0.05)  在95%的信心水準下拒絕H0，兩母體變異數不同。 |
| t-test假設 | H0: E[T|X\*=1] = E[T|X\*=0]  H0: E[T|X\*=1] ≠ E[T|X\*=0] |
| t-test結論 | p-value = 2.623e-05 < α (0.05)  在95%的信心水準下拒絕H0，兩母體平均數不同。即在分組後的WBC下，兩組平均的 remission time具有統計顯著的差異。 |

(c) Pearson's correlation coefficient for T and X

|  |  |
| --- | --- |
| Code | cor.test(data$Remission.Time, data$WBC,method="pearson") |
| output |  |
| 假設 | H0∶ ρ(T,X)=0  H1∶ ρ(T,X)≠0 |
| 結論 | p-value = 4.443e-07 < α (0.05)  在 95%的信心水準下拒絕H0，T與X有關。即WBC及remission time具有統計顯著的負相關，相關係數= -0.689399。 |

(d) Linear regression for T and X

|  |  |  |
| --- | --- | --- |
| Code | model<-lm(data$Remission.Time ~ data$WBC)  summary(model) | |
| output |  | |
| Model | Y=E[Y|X]+ϵ ϵ~N(0,σ2)  =β0+β1X+ϵ  E^[Y|X]= -7.023X+33.461 | Y: T (Remission.Time)  X: WBC  β0: 模型的截距  β1: WBC對Remission.Time造成的斜率 |
| 假設 | H0∶ β1=0  H1∶ β1≠0 | |
| 結論 | p-value = 4.44e-07 < α (0.05)  在 95%的信心水準下拒絕H0，β1不為0，X對T具統計顯著的影響。即WBC對remission time具有統計顯著的影響，當WBC每增加一單位，remission time平均減少7.023。 | |

(e) Logistic regression for T\* and X

|  |  |  |
| --- | --- | --- |
| Code | model\_log<-glm(data$t ~ data$WBC, family=binomial)  summary(model\_log) | |
| output |  | |
| Model | Y|X~Bernoulli(px) px =P(Y=1|X)=E[Y|X]  logit(px)=ln = β0+β1X  logit(p^x)= -2.2701X+6.4876 | Y: T\*(分組後的remission time)  X: WBC  β0: 模型的截距  β1: WBC每增加一單位，T\*的OR增加eβ1倍 |
| 假設 | H0∶ β1=0  H1∶ β1≠0 | |
| 結論 | p-value = 0.00153 < α (0.05)  在95%信心水準下拒絕H0，β1不為0，X對T\*具統計顯著的影響。即WBC對remission time具有統計顯著的影響，WBC每增加一單位對分組後remission time的OR增加eβ1倍。 | |

(f) Poisson regression for T and X

|  |  |  |
| --- | --- | --- |
| Code | model\_poi=glm(data$Remission.Time~data$WBC,family=poisson)  summary(model\_poi) | |
| output |  | |
| Model | Y|X~Poisson(m·λx) m·λx=E[Y|X]  ln(λx)= β0+β1X  ln(λ^x)= -0.64563X+4.29404 | Y: T (remission time)  X: WBC  β0: 模型的截距  β1: WBC每增加一單位，risk ratio of remission time平均增加eβ1倍 |
| 假設 | H0∶ β1=0  H1∶ β1≠0 | |
| 結論 | p-value <2e-16 < α (0.05)  在95%的信心水準下拒絕H0，β1不為0，X對T具統計顯著的影響。即WBC對remission time具統計顯著的影響，WBC每增加一單位對risk ratio of remission time平均增加eβ1倍。 | |