Computing in Epidemiology and Biostatistics

Modern statistical computing in R: Monte-Carlo simulations (Type I error rates and power)

Wan-Yu Lin

Ex: If sleep duration (*Y*) is normally distributed, and

$$Y_i = 7 - 0.05 \cdot Age_i + \beta \cdot Watchpc_i + e_i$$

where Age_i is the age of the ith subject, $Watchpc_i$ is the time for watching television of the ith subject, ε_i is the random error term following the standard normal distribution.

$Age \sim N(\mu = 35, \sigma = 5)$, and $Watchpc \sim N(\mu = 4, \sigma = 1)$. Suppose the sample size is 50. Number of replications = 1000, and seed numbers from 1 to 1000, respectively. Please use simulations to evaluate the type I error rate and power of the Wald's test for $H_0: \beta = 0$ *vs.* $H_1: \beta \neq 0$, where $\beta = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5$, respectively. Please make a figure to describe both type I error rate and power.

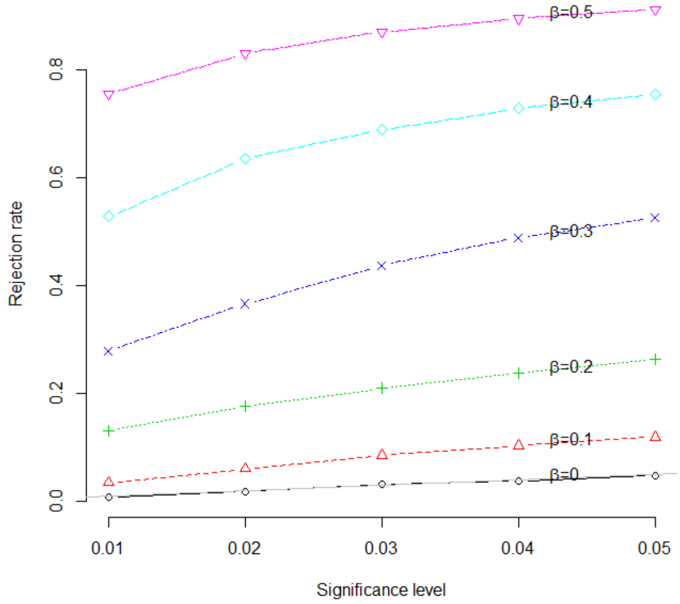
Wald's test: https://en.wikipedia.org/wiki/Wald_test

rej.rate:

	sig.level=0.01	sig.level=0.02	sig.level=0.03	sig.level=0.04	sig.level=0.05
$\beta = 0$					
$\beta = 0.1$					
$\beta = 0.2$					
$\beta = 0.3$					
$\beta = 0.4$					
$\beta = 0.5$					

```
betawatchpc <- seq(0,0.5,0.1)
sig <- seq(0.01,0.05,0.01)      # significance level
n <- 50
Y <- c()
no.rep <- 1000
rej.rate <- matrix(NA,length(betawatchpc),length(sig))      # rejection rate
for(betaloop in 1:length(betawatchpc)){      # the 1st loop
pvalue <- c()
```

```
pvalueLR <- c()
for(i in 1:no.rep){        # the 2nd loop
   set.seed(i)
   age <- rnorm(n,35,5)
   watchpc <- rnorm(n,4,1)
   random.error <- rnorm(n,0,1)
   Y <- 7-0.05*age+betawatchpc[betaloop]*watchpc+random.error
# above: data generation process
# below: data analysis process
   X <- cbind(rep(1,length(Y)),age,watchpc)
   MLE <- solve(t(X)%*%X)%*%t(X)%*%Y
   MSE <- sum((Y-X %*% MLE)^2)/(n-3)      # residual
   seMLE <- sqrt(diag(MSE * solve(t(X)%*%X)))        # variance-covariance of betahat
   pvalue[i] <- ((1-pt(abs(MLE/seMLE),n-3))*2)[3]      # Wald's test statistic
   pvaluelm <- summary(lm(Y~age+watchpc))$coef[3,4]
   if(abs(pvalue[i]-pvaluelm) > 1e-4){        # is my answer the same as the result from lm?
      cat('Error\n')
      break
   }
}
for(k in 1:length(sig)){
   rej.rate[betaloop,k] <- sum(pvalue<sig[k])/no.rep       # calculate rejection rate
}
}
rej.rate
matplot(sig,t(rej.rate),col=c(1:length(betawatchpc)),pch=c(1:length(betawatchpc)),lty=c(1:length(betawatchpc)),type="b",frame=F,xlab="Significance level",ylab="Rejection rate")
abline(a=0,b=1,col=8)
legend(0.04,rej.rate[1,4]+0.05,expression(paste(beta,'=0')),bty="n")
legend(0.04,rej.rate[2,4]+0.05,expression(paste(beta,'=0.1')),bty="n")
legend(0.04,rej.rate[3,4]+0.05,expression(paste(beta,'=0.2')),bty="n")
legend(0.04,rej.rate[4,4]+0.05,expression(paste(beta,'=0.3')),bty="n")
legend(0.04,rej.rate[5,4]+0.05,expression(paste(beta,'=0.4')),bty="n")
legend(0.04,rej.rate[6,4]+0.05,expression(paste(beta,'=0.5')),bty="n")
```



$H_0 : \beta = 0$ *vs.* $H_1 : \beta \neq 0$

Recall:

Hypothesis	**Do not reject H_0**	**Reject H_0**
H_0 is true ($\beta = 0$)	Correct	Type I error
H_1 is true ($\beta \neq 0$)	Type II error	Correct

Type I error rate = $\Pr(\textit{reject } H_0 \mid H_0 \textit{ is true})$

Power = $\Pr(\textit{reject } H_0 \mid H_1 \textit{ is true})$

Ex 23: Please compare the above test (the Wald's test) with the likelihood ratio test (so-called LR test). Which one is more powerful? Please put the rejection rates of the LR test on the above figure.

Likelihood-ratio test: https://en.wikipedia.org/wiki/Likelihood-ratio_test

$$\lambda_{\mathrm{LR}} = -2\left[\,\ell(\theta_0) - \ell(\hat{\theta})\,\right] = 2\left[l(H_1) - l(H_0)\right] \sim \chi^2_{df=1}$$

$H_1: Y_i = 7 - 0.05 \cdot Age_i + \beta \cdot Watchpc_i + e_i$
$H_0: Y_i = 7 - 0.05 \cdot Age_i + e_i$

Homework (8 points, please pay attention to all the words in this orange box)

Ex 23-1: If the probability of getting an admission is π_i for the ith subject, which might be related to his/her gpa and gre scores. Considering the model

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -6 + \beta \cdot gpa_i + 0.005 gre_i$$

$gpa \sim N(\mu = 3.1, \sigma = 0.3)$ and $gre \sim N(\mu = 580, \sigma = 80)$. Suppose the sample size is 1000.

Number of replications = 100, and seed numbers from 1 to 100, respectively. Please use simulations to evaluate the type I error rate and power of the Wald's test for $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$, where $\beta = 0,\ 0.5,\ 0.8$, respectively. Please make a figure to describe both type I error rate and power.

Hint:

1. Generate gpa and gre scores for n subjects.
2. Given gpa and gre, claculate π_i.
3. sample(c(0,1),1,c(1-π_i, π_i),replace=F)
4. Use the Newton-Raphson method to find MLEs, variance-covariance matrix, and the Wald's test statistic.

Please note: Using the R built-in function "glm(admit~gpa+gre,family=binomial)**" to answer this homework will be scored as 0, although you may use it to check your own answers.**