

> # Ex5

```
> set.seed(10000)
> x1 = runif(10000,1,3)
> x2 = rgamma(10000,3,scale=2)
> x3 = rbinom(10000,1,0.3)
> noise = rnorm(10000,2,1)
> y = 0.5 + 1.2*x1-0.9+0.1*x3 + noise
> ydum = c()
> ybar = mean(y)
> i = 1
> for(val in y) {
+   if (val> ybar) {
+     ydum[i] =1
+   }
+   else {
+     ydum[i] = 0
+   }
+   i = i+1;
+ }
```

> # Ex6

> ## correlation

```
> cor(y, x1)
[1] 0.5732947
> cor(y, x1) - 1.2
[1] -0.6267053
```

Significant.

```
> x_all = cbind(rep(10000,1),x1,x2,x3)
```

> ## coefficients

```
> beta_hat = solve(t(x_all) %*% x_all) %*% t(x_all) %*% y
> print(beta_hat)
```

[,1]

0.0001584527

x1 1.2067044540

x2 0.0011340067

x3 0.0922800481

> ## standard error

```
> s2 = sum((y-x_all%*%beta_hat)^2)/(10000-4)
> var = s2*solve(t(x_all)%*%x_all)
> se = diag(sqrt(var))
> print(se)
```

x1

x2

x3

4.033776e-06 1.723276e-02 2.854180e-03 2.167972e-02

```
> # Ex7
> ## logit
> logit = glm(ydum~x1+x2+x3,family=binomial(link = "logit"))
> summary(logit)
```

Call:

```
glm(formula = ydum ~ x1 + x2 + x3, family = binomial(link = "logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0929	-0.8857	0.4923	0.8889	2.0832

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.063777	0.102477	-39.656	<2e-16 ***
x1	1.996247	0.044795	44.564	<2e-16 ***
x2	0.003127	0.006566	0.476	0.6339
x3	0.123977	0.049980	2.481	0.0131 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 13863 on 9999 degrees of freedom
 Residual deviance: 11297 on 9996 degrees of freedom
 AIC: 11305

Number of Fisher Scoring iterations: 4

```
> ## probit
> probit = glm(ydum~x1+x2+x3,family=binomial(link = "probit"))
> summary(probit)
```

Call:

```
glm(formula = ydum ~ x1 + x2 + x3, family = binomial(link = "probit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1213	-0.8919	0.4777	0.8969	2.1120

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.458112	0.058705	-41.873	<2e-16 ***
x1	1.206640	0.025429	47.451	<2e-16 ***

x2	0.001868	0.003909	0.478	0.6327
x3	0.076200	0.029729	2.563	0.0104 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 13863 on 9999 degrees of freedom
 Residual deviance: 11298 on 9996 degrees of freedom
 AIC: 11306

Number of Fisher Scoring iterations: 4

```
> ## linear
> linear = lm(ydum~x1+x2+x3)
> summary(linear)
```

Call:
 lm(formula = ydum ~ x1 + x2 + x3)

Residuals:

Min	1Q	Median	3Q	Max
-0.93539	-0.34494	0.06901	0.34662	0.93042

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.355601	0.017731	-20.055	<2e-16 ***
x1	0.421245	0.007575	55.610	<2e-16 ***
x2	0.000566	0.001255	0.451	0.6519
x3	0.023270	0.009530	2.442	0.0146 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4369 on 9996 degrees of freedom
 Multiple R-squared: 0.2366, Adjusted R-squared: 0.2364
 F-statistic: 1033 on 3 and 9996 DF, p-value: < 2.2e-16

Significance

All coefficients from 3 models are statistically significant.

```
> # Ex8
> ## logit
> library("margins")
> x = glm(ydum~x1+x2+x3,family=binomial(link = "logit"))
```

```

> m = margins(x)
> summary(m)

```

factor	AME	SE	z	p	lower	upper
x1	0.3804	0.0047	80.5766	0.0000	0.3711	0.3896
x2	0.0006	0.0013	0.4763	0.6339	-0.0019	0.0030
x3	0.0236	0.0095	2.4829	0.0130	0.0050	0.0423

```

> ## probit
> library("margins")
> x = glm(ydum~x1+x2+x3,family=binomial(link = "probit"))
> m = margins(x)
> summary(m)

```

factor	AME	SE	z	p	lower	upper
x1	0.3862	0.0049	78.7609	0.0000	0.3766	0.3958
x2	0.0006	0.0013	0.4779	0.6327	-0.0019	0.0030
x3	0.0244	0.0095	2.5653	0.0103	0.0058	0.0430