

MATH 204: PRINCIPLES OF STATISTICS 2
WINTER 2026
MULTIPLE TESTING CORRECTIONS IN REGRESSION

In a general multiple regression model

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i$$

We can fit each of these models easily using least-squares to obtain estimates $\hat{\beta}_0$ and

$$(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$$

where $\hat{\beta}_j$ can be interpreted as the *amount of increase* in response y when x_j increases by one unit when the other predictors are held fixed. We can test the hypothesis

$$H_0 : \beta_j = 0$$

$$H_0 : \beta_j \neq 0$$

using the usual hypothesis testing approach. The test statistic is

$$T_{obs,j} = \frac{\hat{\beta}_j}{\text{e.s.e.}(\hat{\beta}_j)} = \frac{\text{ESTIMATE}}{\text{STANDARD ERROR}}$$

If H_0 is true,

$$T_{obs,j} \sim t_{n-k-1}$$

as we are estimating $k + 1$ parameters overall.

In multiple regression, when testing for each coefficient, we should strictly use a *multiple testing correction* (as in post-hoc tests in ANOVA). However, the precise correction to make is not completely clear, as the tests are in general dependent.

Example: Suppose we have three predictors, and the underlying model is

$$Y_i = 2 + \epsilon_i$$

where $\epsilon_i \sim \text{Normal}(0, 1)$, and yet we fit the model

$$E[Y_i|x_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

That is, in such a model we have in reality

$$\beta_0 = 2 \quad \beta_1 = \beta_2 = \beta_3 = 0.$$

We examine the behaviour of the estimates in a simulation study of $N = 100000$ replicates for a sample size of $n = 500$.

```
> set.seed(29)
> n<-500
> x1<-rnorm(n, 10, 1)
> x2<-rnorm(n, -0.2*x1, 1)
> x3<-rnorm(n, x1+x2, 1)
> X<-cbind(x1, x2, x3)
> Y<-2+rnorm(n)
```

We run the replicated analyses, storing the estimates, t-statistics and p-values.

```

> N<-100000
> ests<-tstats<-pvals<-rejectH0<-matrix(0,nrow=N,ncol=4)
> alpha<-0.05
> for(irep in 1:N){
+   Y<-2+rnorm(n)
+   fit0<-lm(Y~x1+x2+x3)
+   ctab<-summary(fit0)$coef
+   ests[irep,]<-ctab[,1]
+   tstats[irep,]<-ctab[,3]
+   pvals[irep,]<-ctab[,4]
+   rejectH0[irep,]<-as.numeric(ctab[,4] < alpha)
+ }

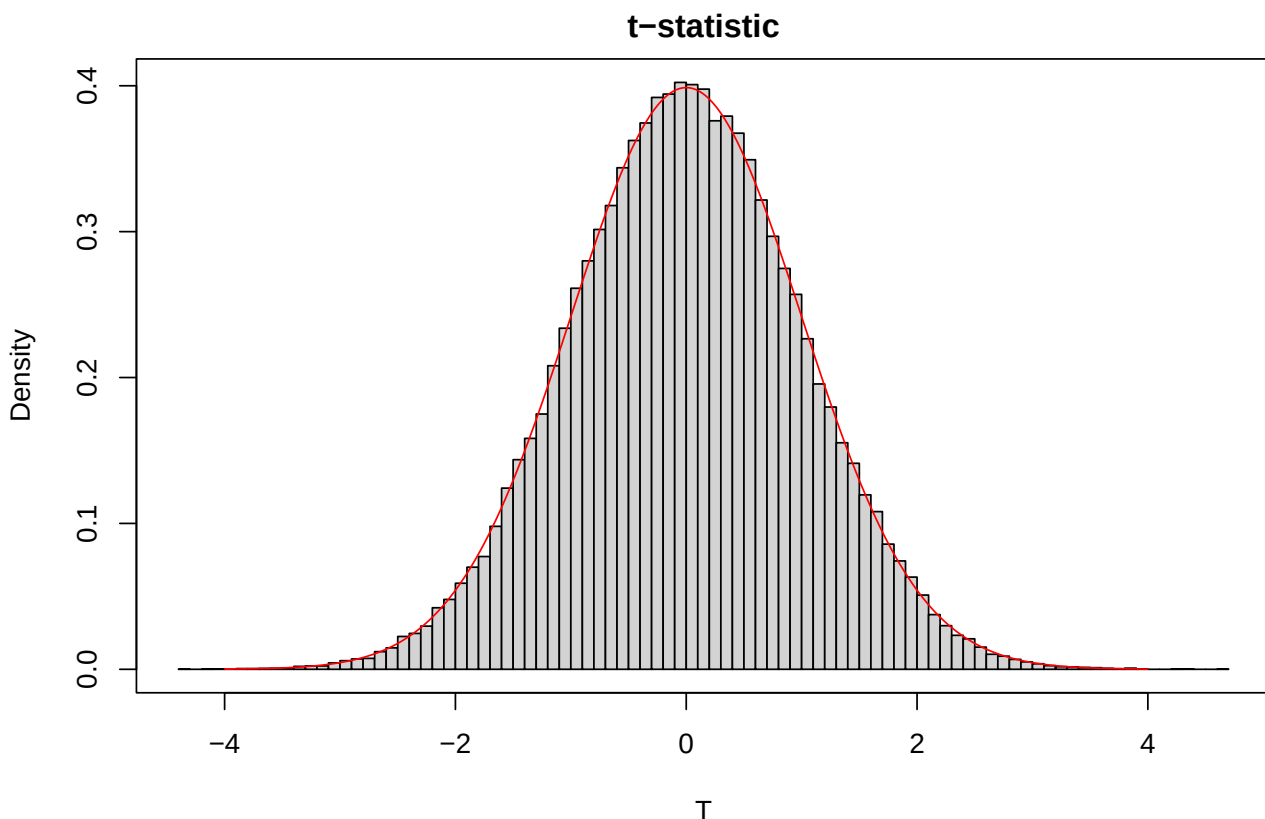
```

The t-statistics have the correct distribution, in this case the t_{n-4} distribution, which is close to the standard Normal: the following histogram is for the β_1 estimator.

```

> par(mar=c(4,4,2,2))
> hist(tstats[,2], freq=FALSE,
+   nclass=100, main='t-statistic', xlab='T');box()
> xv<-seq(-4,4,by=0.01); fv<-dt(xv,n-3-1)
> lines(xv,fv,col='red')

```



If we examine the test characteristics, we note that each individual test has a Type 1 error rate of $\alpha = 0.05$. We can compute the empirical Type 1 error rate from the test results, and they are close to 0.05.

```
> apply(rejectH0[, -1], 2, sum) / N
+ [1] 0.04881 0.04958 0.04994
```

However, the test results are positively correlated: we can inspect the correlation between the binary variables recording the rejection/non-rejection of the null.

```
> cor(rejectH0[, 2:4])
+           [,1]      [,2]      [,3]
+ [1,] 1.000000 0.199686 0.334138
+ [2,] 0.199686 1.000000 0.317107
+ [3,] 0.334138 0.317107 1.000000
```

indicating that if one null hypothesis is rejected, the other two are more likely to be rejected. This means that the usual Familywise Type 1 error rate given by the formula $\alpha_F = 1 - (1 - \alpha)^3$ does not match the actual rate.

```
> sum(apply(rejectH0[, -1], 1, sum) != 0) / N
+ [1] 0.109
> 1 - (1 - alpha) ^ 3
+ [1] 0.142625
```

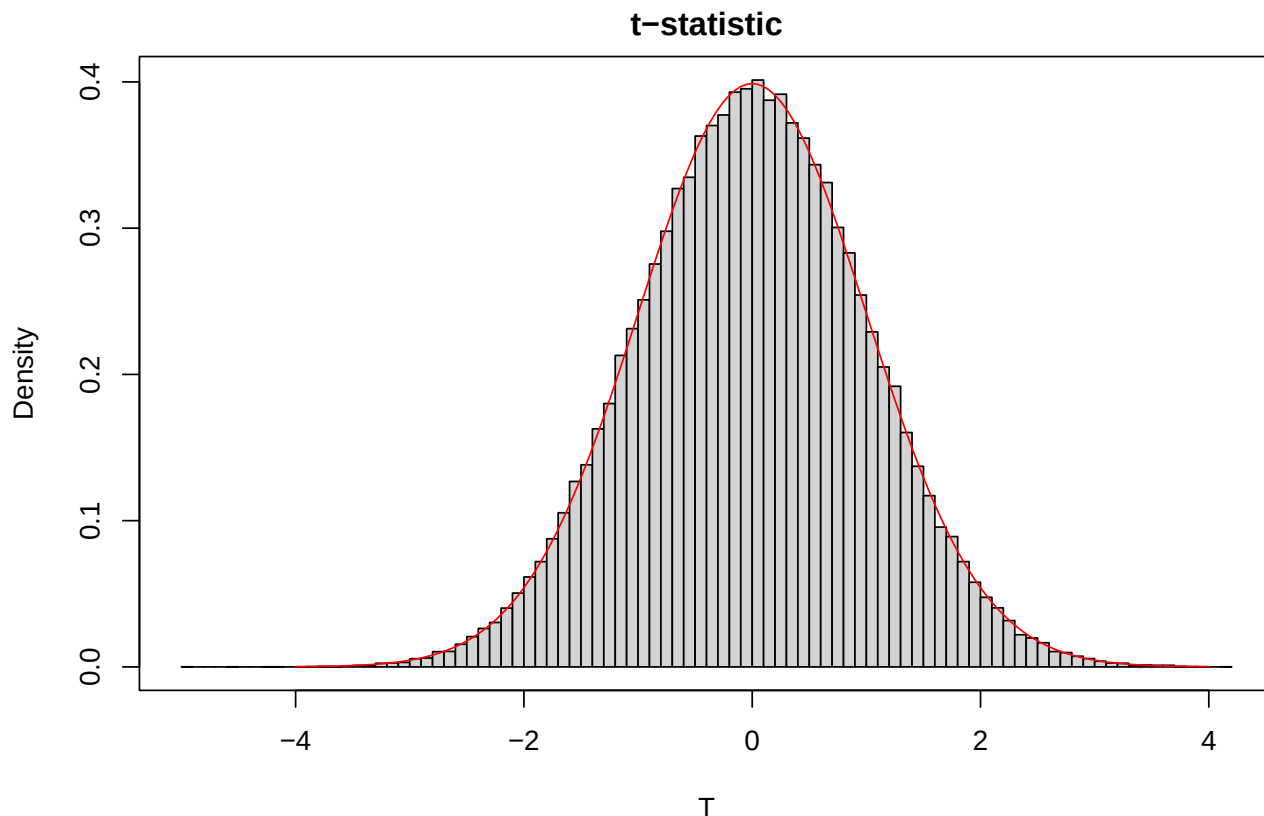
The results above are due to the dependence between the predictors. We may repeat the analysis using completely independent predictors.

```
> set.seed(29)
> n<-500
> x1<-rnorm(n, 10, 1)
> x2<-rnorm(n, -8, 1)
> x3<-rnorm(n, 2, 1)
> X<-cbind(x1, x2, x3)
> Y<-2+rnorm(n)
```

We run the replicated analyses, storing the estimates, t-statistics and p-values.

```
> N<-100000
> ests<-tstats<-pvals<-rejectH0<-matrix(0, nrow=N, ncol=4)
> alpha<-0.05
> for(irep in 1:N){
+   Y<-2+rnorm(n)
+   fit0<-lm(Y~x1+x2+x3)
+   ctab<-summary(fit0)$coef
+   ests[irep,]<-ctab[, 1]
+   tstats[irep,]<-ctab[, 3]
+   pvals[irep,]<-ctab[, 4]
+   rejectH0[irep,]<-as.numeric(ctab[, 4] < alpha)
+ }
```

The t-statistics again have the correct distribution:.



If we examine the test characteristics, we note that each individual test has a Type 1 error rate of $\alpha = 0.05$. We can compute the empirical Type 1 error rate from the test results, and they are close to 0.05.

```
> apply(rejectH0[, -1], 2, sum) / N
+ [1] 0.04952 0.04961 0.04994
```

The test results are now essentially uncorrelated

```
> cor(rejectH0[, 2:4])
+           [,1]      [,2]      [,3]
+ [1,] 1.00000000 0.01110872 0.00395652
+ [2,] 0.01110872 1.00000000 0.00216665
+ [3,] 0.00395652 0.00216665 1.00000000
```

and Familywise Type 1 error rate given by the formula matches the actual rate.

```
> sum(apply(rejectH0[, -1], 1, sum) != 0) / N
+ [1] 0.14106
> 1 - (1 - alpha)^3
+ [1] 0.142625
```