SCRIPT `MOD3S2A`: HYPOTHESIS TESTING IN LEAST SQUARES ESTIMATION

INSTRUCTOR: KLAUS MOELTNER

1. Load data & run OLS

This is our wage data set from script `mod1s2b`. We start by re-estimating this model via OLS.

```
R> load("c:/klaus/AAEC5126/R/data/wage1000.rda")
```

Recall the variable definitions:

Table 1. Variable description

pos.	variable	description
1	wage	hourly wage (1995 dollars)
2	gender	(1= worker = female)
3	race	(1= worker = non-white)
4	union	(1 = worker = unionized)
5	education	years of education
6	experience	years of work experience
7	age	age in years

Table 2. OLS output for wage model

variable	estimate	s.e.	t
constant	-8.579	1.161	-7.388
gender	-3.099	0.424	-7.313
race	-1.607	0.603	-2.664
unionmember	0.821	0.583	1.408
education	1.498	0.075	19.948
experience	0.170	0.018	9.197

2. Linear Hypothesis Tests

2.1. Example 1.

H_A: "A female worker earns \$3 less than male worker."

```
R> Rmat<-matrix(c(0, 1, 0, 0, 0, 0),nrow=1)
R> q<- -3
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 0.054. The corresponding p-value is 0.816.

2.2. **Example 2.** H_A: "One year of education is worth eight years of experience."

```
R> Rmat<-matrix(c(0, 0, 0, 0, 1, -8),nrow=1)
R> q<- 0
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 0.84. The corresponding p-value is 0.36.

2.3. **Example 3.** H_A: "Both previous hypotheses hold."

```
R> Rmat1<-matrix(c(0, 1, 0, 0, 0, 0),nrow=1)
R> Rmat2<-matrix(c(0, 0, 0, 0, 1, -8),nrow=1)
R> Rmat<-rbind(Rmat1,Rmat2)
R> q<- matrix(c(-3,0),nrow=2)
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 0.446. The corresponding p-value is 0.64.

2.4. **Example 4.** H_A: All of the above, plus "A unionized worker earns \$1 more than a non-unionized worker".

```
R> Rmat1<-matrix(c(0, 1, 0, 0, 0, 0),nrow=1)
R> Rmat2<-matrix(c(0, 0, 0, 0, 1, -8),nrow=1)
R> Rmat3<-matrix(c(0, 0, 0, 1, 0, 0),nrow=1)
R> Rmat<-rbind(Rmat1,Rmat2,Rmat3)
R> q<- matrix(c(-3,0,1),nrow=3)
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 0.364. The corresponding p-value is 0.779.

2.5. **Example 5a: Full Chow Test.** H_A: "At least one of the marginal effects of the remaining regressors is different for white vs. nonwhite workers". We will build a larger data set that includes all the original variables, plus all possible interactions of "white" with the remaining regressors. We can then test if all of the differences between corresponding interactions are jointly zero.

```
R> #prepare data
R> #declare nonwhite to be of matrix form (so we can replicate it below),
R> #and create a matrix form of the opposite ethinicity variable.
R> nonwhite<-matrix(race)
R> white<-matrix(1-race)
R> #replicate these ethnicity vectors 4 times and element-by-element multiply
R> #with the remaining regressors to genenrate the required interaction terms.
```

```
R> int1<-matrix(rep(nonwhite,4),nrow=n)*cbind(gender,unionmember,education,experience)
R> int2<-matrix(rep(white,4),nrow=n)*cbind(gender,unionmember,education,experience)
R> #define the augmented X matrix
R> X<-cbind(nonwhite,int1,white,int2)
R> k<-ncol(X)
R> #run OLS on augmented model
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)# compute OLS estimator
R> e<-y-X%*%bols # Get residuals.
R> SSR<-(t(e)%*%e)#sum of squared residuals - should be minimized
R> s2<-(t(e)%*%e)/(n-k) #get the regression error (estimated variance of "eps").
R> Vb<-s2[1,1]*solve((t(X))%*%X) # get the estimated variance-covariance matrix of bols
R> se=sqrt(diag(Vb)) # get the standard erros for your coefficients;
R> tval=bols/se # get your t-values.
R> tt<-data.frame(col1=c("intercept nw","gender*nw","unionmember*nw","education*nw","experience*nw",
 "intercept w","gender*w","unionmember*w","education*w","experience*w"),
                 col2=bols,
                 col3=se,
                 col4=tval)
R> colnames(tt)<-c("variable","estimate","s.e.","t")
```

Table 3. OLS output for augmented wage model

variable	estimate	s.e.	t
intercept nw	-5.999	2.971	-2.019
gender*nw	-1.645	1.115	-1.476
unionmember*nw	2.101	1.388	1.514
education*nw	1.110	0.198	5.607
experience*nw	0.151	0.050	3.034
intercept w	-9.354	1.245	-7.516
gender*w	-3.375	0.457	-7.378
unionmember*w	0.497	0.644	0.772
education*w	1.570	0.081	19.368
experience*w	0.170	0.020	8.559

Now for the full Chow test:

```
R> Rmat1<-matrix(c(1, 0, 0, 0, 0, -1, 0, 0, 0, 0),nrow=1)
R> Rmat2<-matrix(c(0, 1, 0, 0, 0, 0, -1, 0, 0, 0),nrow=1)
R> Rmat3<-matrix(c(0, 0, 1, 0, 0, 0, 0, -1, 0, 0),nrow=1)
R> Rmat4<-matrix(c(0, 0, 0, 1, 0, 0, 0, 0, -1, 0),nrow=1)
R> Rmat5<-matrix(c(0, 0, 0, 0, 1, 0, 0, 0, 0, -1),nrow=1)
R> Rmat<-rbind(Rmat1,Rmat2,Rmat3,Rmat4,Rmat5)
R> q<- matrix(c(0,0,0,0,0),nrow=5)
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 3.082. The corresponding p-value is 0.009.

2.6. **Example 5b: Partial Chow Test.** H_0: "The marginal effects of "gender" and "education" are the same for white and nonwhite workers".

```
R> Rmat1<-matrix(c(0, 1, 0, 0, 0, 0, -1, 0, 0, 0),nrow=1)
R> Rmat2<-matrix(c(0, 0, 0, 1, 0, 0, 0, 0, -1, 0),nrow=1)
R> Rmat<-rbind(Rmat1,Rmat2)
R> q<- matrix(c(0,0),nrow=2)
R> J<-nrow(Rmat)
R> b<-bols
R> Fstat<-(1/J)* t(Rmat %*% b-q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*% (Rmat%*%b-q)
R> pval<-1-pf(Fstat,J,n-k)
```

The F-statistic for this test is 3.595. The corresponding p-value is 0.028.

3. Testing Nonlinear Restrictions

Assume you want to test the hypothesis that the ratio of the marginal effects of "education" over "experience" equals 10 times the marginal effect of "union membership", i.e. $\beta_5/\beta_6 = 10 * \beta_4$, or, equivalently, $\beta_5/(\beta_6 * \beta_4) = 10$.

```
R> #Re-estimate original OLS model
R> X<-cbind(rep(1,n),gender,race,unionmember,education,experience)
R> k<-ncol(X)
R> bols<-solve((t(X)) %*% X) %*% (t(X) %*% y)# compute OLS estimator
R> e<-y-X%*%bols # Get residuals.
R> SSR<-(t(e)%*%e)#sum of squared residuals - should be minimized
R> s2<-(t(e)%*%e)/(n-k) #get the regression error (estimated variance of "eps").
R> Vb<-s2[1,1]*solve((t(X))%*%X) # get the estimated variance-covariance matrix of bols
R> se=sqrt(diag(Vb)) # get the standard erros for your coefficients;
R> tval=bols/se # get your t-values.
R> #label relevant elements of b, define c(b) and q, and compute all relevant derivatives
R> b<-bols
R> b4<-b[4]
R> b5<-b[5]
R> b6<-b[6]
R> cb<-b5/(b6*b4)
R> q<-10
R> J<-1
R> delb4<--b5*b6*(b6*b4)^(-2)
R> delb5<-1/(b6*b4)
R> delb6<--b5*b4*(b6*b4)^(-2)
R> #Compose C, V(c(b)), and the Wald statistic
R> C<-matrix(c(0,0,0, delb4, delb5, delb6),nrow=1)
R> Vcb<-C %*% Vb %*% t(C)
R> W<-t(cb-q) %*% solve(Vcb) %*% (cb-q)
R> pval<-1-pchisq(W,J)
```

The Wald-statistic for this test is 0.01. The corresponding p-value is 0.92.

```
R> proc.time()-tic
   user  system elapsed
   0.17    0.11    0.28
```