University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Exam 2 - practice problems

Answer the following questions:

a. Show that the F test for the overall significance of the model is equal to $\frac{R^2}{1-R^2}\frac{n-k-1}{k}$.

b. Refer to question (a). Find the distribution of R^2.
Hint 1: For easier notation let $R^2 = W$, so $F = \frac{W}{1-W}\frac{n-k-1}{k}$. Solve for W, and use the method of CDF to find the distribution of W.
Hint 2: Let $X \sim F_{n_1,n_2}$. The pdf of the F distribution is

$$f(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{1}{2}(n_1+n_2)}, \quad 0 < x < \infty.$$

Hint 3: As a reminder the beta distribution has the following pdf. Let $X \sim Beta(\alpha, \beta)$ then

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad \alpha > 0,\ \beta > 0,\ 0 \le x \le 1, \text{ where } B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \text{ (beta function).}$$

c. Refer to question (b). Find the expected value and variance of R^2.
Note: The mean and variance of the beta distribution are: $E(X) = \frac{\alpha}{\alpha+\beta}$ and $var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

d. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$. The Gauss-Markov conditions hold and also $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Construct a prediction interval for the average of m new observations of Y for a given new case (not one of the rows of $\mathbf{X}$), $x_0' = (1, x_{01}, x_{02}, \ldots, x_{0k})'$.

e. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The Gauss-Markov conditions hold and suppose that $\mathbf{1}'\mathbf{X}_{(\mathbf{0})} = \mathbf{0}$. Find expressions for $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}_{(\mathbf{0})}$, and the variance covariance matrix of $\begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{(\mathbf{0})} \end{pmatrix}$.

f. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ subject to a set of linear constraints of the form $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbf{C}$ is $m \times (k+1)$ matrix. The Gauss-Markov conditions hold and also $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Are the quadratic expressions $\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}$ and $\mathbf{Y}'(\mathbf{H}-\mathbf{H_1})\mathbf{Y}$ independent? ($\mathbf{H}$ and $\mathbf{H_1}$ are the hat matrices of the full and reduced model respectively.)

g. Refer to question (f). Transform the model into its canonical form and find the distribution of $\frac{\hat{\boldsymbol{\beta}}_{\mathbf{1r}}'\mathbf{X}_{\mathbf{1r}}'\mathbf{X}_{\mathbf{1r}}\hat{\boldsymbol{\beta}}_{\mathbf{1r}}}{\sigma^2}$.

h. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$. The Gauss-Markov conditions hold and also $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Assume $n = 359, k = 5$. For this regression problem it is given that $\mathbf{y}'\mathbf{y} = 1463631$, $\mathbf{y}'(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{y} = 392179$, and $\mathbf{y}'\mathbf{H}\mathbf{y} = 1345608$. Test the overall significance of the model.

i. Consider a multiple regression problem with $n = 155$ and $k = 3$ predictors. We obtained $\hat{\boldsymbol{\beta}} = \begin{pmatrix} 7.20 \\ -14.18 \\ -0.19 \\ 0.43 \end{pmatrix}$,
$s_e^2 = 665$, and $\mathbf{X}'\mathbf{X}, (\mathbf{X}'\mathbf{X})^{-1}$ as follows:

```
> t(X) %*% X
       155.0    503.10    6249.0    72806.0
       503.1   3545.15   32175.4   418819.5
      6249.0  32175.40  338293.0  4151102.0
     72806.0 418819.50 4151102.0 54948600.0


> round(solve(t(X) %*% X),7)
      0.0449434  0.0089640 -0.0015572 -1.02e-05
      0.0089640  0.0047180 -0.0003741 -1.96e-05
     -0.0015572 -0.0003741  0.0000958 -2.30e-06
     -0.0000102 -0.0000196 -0.0000023  4.00e-07
```

Test the hypothesis that $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$.

j. Refer to question (i). After we dropped predictor x_1 from the model we have obtained a new value $s_e^2 = 941$. Use the extra sum of squares method to test the hypothesis that $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$. Compare the test statistic with the one in question (b).

k. Refer to question (i). Test the hypothesis $H_0 : \beta_1 + \beta_2 + 15 = 0$ against $H_a : \beta_1 + \beta_2 + 15 \neq 0$.