Name Jerry

Student Number

STA 312 f2023 Quiz 7

1. (4 points) Let the continuous random variable X have density $f_x(x) = \exp\{x - e^x\}$, for all real x. Find the density of $Y = e^X$. Show your work. Please be sure to specify where the density of Y is non-zero, and where it is zero.

For $y > 0$,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \le y) = \frac{d}{dy} P(e^X \le y)$$

$$= \frac{d}{dy} P(X \le \log y) = \frac{d}{dy} F_X(\log y)$$

$$= f_X(\log y) \cdot \frac{1}{y} = \text{Exp}\{\log y - e^{\log y}\} \cdot \frac{1}{y}$$

$$= \text{Exp}\{\log y - y\} \cdot \frac{1}{y} = e^{\log y} e^{-y} \frac{1}{y}$$

$$= y e^{-y} \frac{1}{y} = e^{-y}, \text{ so}$$

$$f_Y(y) = \begin{cases} e^{-y} & \text{for } y > 0 \\ 0 & \text{for } y \le 0 \end{cases}$$

2. (3 points) For the regression equation $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \epsilon_i$, the response variable y_i is the cumulative grade point average (GPA) of High School student i, $x_{i,1}$ is family income, $x_{i,2}$ is mother's education, and $x_{i,3}$ is father's education. Give the null hypothesis you would test to answer the following questions. A null hypothesis is a statement about the model parameters.

(a) Once you correct for parents' education, is GPA still related to family income?

$H_0: \beta_1 = 0$

(b) Allowing for family income, is GPA related to parents' education?

$H_0: \beta_2 = \beta_3 = 0$ (No marks for just $\beta_2 = \beta_3$)

(c) Once you control for family income and mother's education, is GPA related to father's education?

$H_0: \beta_3 = 0$

3. In Question 8 of Assignment 7, you analyzed data from a sample of pigs. Consider this question: Adjusting for mother's weight and father's weight, which drug helps the average pig gain more weight, Drug 1 or Drug 2?

(a) (1/2 point) In the space below, write the F or t statistic, a number from your printout. On your printout, circle the number and write "Question 3a" beside it.

$F = 9.251$

(b) (1/2 point) In the space below, write the p-value, a number from your printout. On your printout, circle the number and write "Question 3b" beside it.

$p = 0.0033$

(c) (2 points) In plain, non-statistical language, what do you conclude from this test? Your answer should be guided by the $\alpha = 0.05$ significance level, but you must not mention that, because this is a plain language answer. You cannot get any marks for this part unless the test statistic and p-value are correct.

Allowing for mother's and father's weight (can be omitted)

Drug 1 results in higher average weight than Drug 2.

(One mark off for failure to draw a directional conclusion, for example just saying the drugs have different effects)

Please attach the printout(s) with your answers to Question 8 of the assignment, showing all the input and output. **Make sure your name and student number are written on the printout.**

```
> # Pig weight data
> rm(list=ls())
> options(scipen=999) # To avoid scientific notation
>
> pigs = read.table("http://www.utstat.utoronto.ca/~brunner/data/legal/pigweight.data.txt")
> head(pigs); attach(pigs)
  Drug Momweight Dadweight Pigweight
1    1    133.55    172.97     71.99
2    1    143.65    183.32     76.76
3    1    130.27    186.53     72.22
4    1    128.14    174.55     69.56
5    1    128.21    182.79     73.48
6    1    130.49    182.73     69.85
>
> n = length(Drug); n
[1] 75
> # Make dummy variables
> d1=d2=d3 = numeric(n)
> d1[Drug==1] = 1; d2[Drug==2] = 1; d3[Drug==3] = 1
> fullmodel = lm(Pigweight ~ d1+d2 + Momweight + Dadweight)
> summary(fullmodel)

Call:
lm(formula = Pigweight ~ d1 + d2 + Momweight + Dadweight)

Residuals:
   Min     1Q Median     3Q    Max
-3.905 -1.174  0.187  1.351  3.657

Coefficients:
            Estimate Std. Error t value    Pr(>|t|)
(Intercept)  6.77683    9.09308   0.745       0.4586
d1           0.70480    0.52871   1.333       0.1868
d2          -0.90077    0.53916  -1.671       0.0992 .
Momweight    0.26363    0.04727   5.578 0.000000428 ***
Dadweight    0.17442    0.03465   5.034 0.000003580 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.855 on 70 degrees of freedom
Multiple R-squared:  0.4561,    Adjusted R-squared:  0.425
F-statistic: 14.67 on 4 and 70 DF,  p-value: 0.000000009393

>
> # 8d: Predict the dressed weight of a pig getting Drug 2, whose mother weighed 140 pounds,
and whose father weighed 185 pounds.
>
> b = coefficients(fullmodel); b
(Intercept)          d1          d2   Momweight   Dadweight
  6.7768313   0.7048000  -0.9007653   0.2636323   0.1744219
> sum(b*c(1,0,1,140,185)) # 75.05263
[1] 75.05263
> porky = data.frame(d1=0,d2=1,Momweight=140,Dadweight=185)
> predict(fullmodel,newdata=porky) # 75.05263
       1
75.05263
>
> # 8e: Give a 95% confidence interval for the difference in expected weight between drug
treatments 2 and 3. (That's just beta2.)
>
> tcrit = qt(0.975,70); me = tcrit*0.53916
> c(-0.90077-me, -0.90077+me) # -1.9760907  0.1745507
    -1.9760907  0.1745507
    Now do it the long way. (predict is out because of the intercept.)
> ell = c(0,0,1,0,0)
> tcrit*sqrt(t(ell) %*% vcov(fullmodel) %*% ell); me # same of course.
         [,1]
[1,] 1.075316
```

```
[1] 1.075321
>
> # Test hypotheses in Question 8g
> source("http://www.utstat.toronto.edu/brunner/Rfunctions/ftest.txt")
>
> #i) Type of drug controlling for parents
> redmod = lm(Pigweight ~ Momweight + Dadweight)
> anova(redmod,fullmodel)
Analysis of Variance Table

Model 1: Pigweight ~ Momweight + Dadweight
Model 2: Pigweight ~ d1 + d2 + Momweight + Dadweight
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     72 272.70
2     70 240.81  2    31.894 4.6356 0.01286 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> #ii) Drug 1 vs 2
> C2 = rbind(c(0,1,-1,0,0))
> ftest(fullmodel,C2) # Cross-checked by new ref cat.
         F          df1          df2      p-value
9.251054028 1.000000000 70.000000000 0.003311102
>
> #iii) 1vs3 on printout
> #iv)  2vs3 on printout
>
> #v)   momslope vs dadslope
> C5 = rbind(c(0,0,0,1,-1))
> ftest(fullmodel,C5)
         F        df1        df2    p-value
 2.5354089  1.0000000 70.0000000  0.1158241
```

Question 3b

Question 3a