Assignment 7

(1) (a) $E(T^k) = \int_0^\infty t^k \alpha\lambda(\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha}\, dt$

$u = (\lambda t)^\alpha \quad du = \alpha(\lambda t)^{\alpha-1}\lambda\, dt$

t	u
∞	∞
0	0

And $u = \lambda^\alpha t^\alpha \iff t^\alpha = \frac{u}{\lambda^\alpha} \iff t = \left(\frac{u}{\lambda^\alpha}\right)^{\frac{1}{\alpha}} = \frac{u^{1/\alpha}}{\lambda}$,

so

$$E(T^k) = \int_0^\infty \left[\frac{u^{1/\alpha}}{\lambda}\right]^k e^{-u}\, du = \frac{1}{\lambda^k}\underbrace{\int_0^\infty e^{-u} u^{(\frac{k}{\alpha}+1)-1}\, du}_{\text{Gamma function}}$$

$$= \frac{\Gamma(\frac{k}{\alpha}+1)}{\lambda^k}$$

(b) $S(t) = \int_t^\infty \alpha\lambda(\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}\, dx$

Same $u = (\lambda x)^\alpha$

x	u
∞	∞
t	$(\lambda t)^\alpha$

$$= \int_{(\lambda t)^\alpha}^\infty e^{-u}\, du = e^{-(\lambda t)^\alpha}$$

So median $= m$ means

$e^{-(\lambda m)^\alpha} = \frac{1}{2} \iff -(\lambda m)^\alpha = \log(\frac{1}{2}) = \log(2^{-1}) = -\log 2$

$\iff \lambda^\alpha m^\alpha = \log 2 \iff m^\alpha = \frac{\log 2}{\lambda^\alpha} \iff m = \left(\frac{\log 2}{\lambda^\alpha}\right)^{1/\alpha}$

$\iff m = \frac{(\log 2)^{1/\alpha}}{\lambda}$

(1c)

$$h(t) = \frac{f(t)}{S(t)} = \frac{\alpha \lambda^\alpha t^{\alpha-1} \cancel{e^{-(\lambda t)^\alpha}}}{\cancel{e^{-(\lambda t)^\alpha}}}$$

$$= \alpha \lambda^\alpha t^{\alpha - 1}$$

(2) $T \sim Exp(1)$, $Y = \log(T)$ note $Y \in (-\infty, \infty)$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \le y) = \frac{d}{dy} P(\log(T) \le y)$$

$$= \frac{d}{dy} P(T \le e^y) = \frac{d}{dy} F_T(e^y) = f_T(e^y) e^y$$

$$= e^y e^{-e^y} = e^{y - e^y} \quad \text{for } -\infty < y < \infty$$

(3) (a) $M_Z(t) = E(e^{Zt}) = \int_0^\infty e^{zt} e^z e^{-e^z} dz$

$u = e^z \quad du = e^z dz$

$$\begin{array}{c|c} z & u = e^z \\ \infty & \infty \\ -\infty & 0 \end{array}$$

$$= \int_0^\infty u^t e^{-u} du$$

$$= \int_0^\infty e^{-u} u^{(t+1)-1} du = \Gamma(t+1)$$

(3b) $M_Z(t) = \Gamma(t+1)$, so

$$E(Z) = \lim_{t\to 0} \Gamma'(t+1) \overset{\text{cont}}{=} \Gamma'\left(\lim_{t\to 0} t+1\right) = \Gamma'(1)$$

(c) If $T \sim EXP(1)$, $S(m) = \frac{1}{2} = e^{-m}$

$\iff -\log 2 = -m \iff m = \log 2$, and

$$\frac{1}{2} = P(T \le \log 2) = P(\log T \le \log(\log 2))$$

$$= P(Z \le \log(\log 2))$$

Median

(d) $\log f_Z(z) = \log\left(e^{z - e^z}\right)$

$= z - e^z$, and

$$\frac{d}{dz} \log f_Z(z) = 1 - e^z \overset{set}{=} 0 \Rightarrow e^z = 1$$

$$\Rightarrow z = \log(1) = 0 = \text{Mode}$$

Just checking 2nd derivative,

$$\frac{d^2}{dz^2} \log f_Z(z) = \frac{d}{dz}(1 - e^z) = -e^z > 0 \text{ neg}$$

CCD $\cap$ MAX

4 (a) $Y = \sigma Z + \mu$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \le y) = \frac{d}{dy} P(\sigma Z + \mu \le y)$$

$$= \frac{d}{dy} P\left(Z \le \frac{y-\mu}{\sigma}\right) = \frac{d}{dy} F_Z\left(\frac{y-\mu}{\sigma}\right)$$

$$= f_Z\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sigma} \operatorname{Exp}\left\{\left(\frac{y-\mu}{\sigma}\right) - e^{\left(\frac{y-\mu}{\sigma}\right)}\right\}$$

for $-\infty < y < \infty$

(b) $E(Y) = E(\sigma Z + \mu) = \sigma E(Z) + \mu$

$$= \sigma \Gamma'(1) + \mu$$

(c) $Var(Y) = Var(\sigma Z + \mu) = Var(\sigma Z)$

$$= \sigma^2 Var(Z) = \sigma^2 \frac{\pi^2}{6}$$

(3c)

(d) $\frac{1}{2} = P(Z \le \log(\log 2)) = P(\sigma Z + \mu \le \sigma \log(\log 2) + \mu)$

So $med(Y) = \sigma \log(\log 2) + \mu$

(e) $$\frac{d}{dy} \log f_Y(y) = \frac{d}{dy}\left(\log\frac{1}{\sigma} + \left(\frac{y-\mu}{\sigma}\right) - e^{\left(\frac{y-\mu}{\sigma}\right)}\right)$$

$$= \frac{1}{\sigma} - e^{\left(\frac{y-\mu}{\sigma}\right)} \cdot \frac{1}{\sigma} = \frac{1}{\sigma}\left(1 - e^{\left(\frac{y-\mu}{\sigma}\right)}\right) \overset{\text{set}}{=} 0$$

$$\Rightarrow e^{\left(\frac{y-\mu}{\sigma}\right)} = 1 \Rightarrow \frac{y-\mu}{\sigma} = 0 \Rightarrow y = \mu$$

And Mode $= \mu$

(5) (a) $T \sim \text{Weibull}(\alpha, \lambda)$ & $Y = \log(T)$

$$f_T(t) = \alpha\lambda(\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha} I(t \geq 0)$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(\log T \leq y)$$

$$= \frac{d}{dy} P(T \leq e^y) = \frac{d}{dy} F_T(e^y) = f_T(e^y) \cdot e^y$$

$$= \alpha\lambda(\lambda e^y)^{\alpha-1} \operatorname{Exp}\{-(\lambda e^y)^\alpha\} e^y$$

$$= \alpha\lambda\lambda^{\alpha-1}(e^y)^{\alpha-1+1} \operatorname{Exp}\{-\lambda^\alpha e^{\alpha y}\}$$

$$= \alpha\lambda^\alpha e^{\alpha y} \operatorname{Exp}\{-\lambda^\alpha e^{\alpha y}\}$$

(b) Setting $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{-\mu}$,

$$f_Y(y) = \frac{1}{\sigma} e^{-\mu/\sigma} e^{y/\sigma} \operatorname{Exp}\{-e^{-\mu/\sigma} e^{y/\sigma}\}$$

$$= \frac{1}{\sigma} e^{\left(\frac{y-\mu}{\sigma}\right)} \operatorname{Exp}\{-e^{\left(\frac{y-\mu}{\sigma}\right)}\}$$

(c) $= \frac{1}{\sigma} \operatorname{Exp}\{\left(\frac{y-\mu}{\sigma}\right) - e^{\left(\frac{y-\mu}{\sigma}\right)}\}$

As in Q4(a)

6

	x_1	$E(y \mid x)$
Disc	1	$\beta_0 + \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$
Mem	0	$\beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$

(a) $H_0: \beta_1 = 0$

(b) $H_0: \beta_4 = \beta_5 = 0$

(c) $H_0: \beta_2 = \beta_3 = 0$

(d) $H_0: \beta_3 = 0$

7 (a)

	d_1	d_2	d_3	$E(Y)$
A	1	0	0	$\beta_0 + \beta_2 + \beta_1 x$
B	0	1	0	$\beta_0 + \beta_3 + \beta_1 x$
C	0	0	1	$\beta_0 + \beta_4 + \beta_1 x$
wait L	0	0	0	$\beta_0 + \beta_1 x$

(b) $y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \varepsilon$

(c) See above

(d) $H_0: \beta_2 = \beta_3 = \beta_4$

(e) $$\underbrace{\begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}}_{L} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(f) $H_0: \beta_3 = 0$

(g) $H_0: \beta_2 = \beta_3$

(h) $\hat{\beta}_2 - \hat{\beta}_4$

(i) Yes, because of random assignment.

8 (a) $y_i = \beta_0 + \beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 x_{i1} + \beta_4 x_{i2} + \varepsilon_i$

(b)

	d_1	d_2	$E(y\|x)$
1	1	0	$\beta_0 + \beta_1 \quad + \beta_3 x_3 + \beta_4 x_4$
2	0	1	$\beta_0 + \beta_2 \quad + \beta_3 x_3 + \beta_4 x_4$
3	0	0	$\beta_0 \quad + \beta_3 x_3 + \beta_4 x_4$

(c) See printout. $R^2 = 0.4561$

(d) 75.05

(e) $(-1.976, 0.175)$

(f) (i) $H_0: \beta_1 = \beta_2 = 0$
(ii) $H_0: \beta_1 = \beta_2$
(iii) $H_0: \beta_1 = 0$
(iv) $H_0: \beta_2 = 0$

(g) (i) $F = 4.6356$, $P = 0.01286$
(ii) $F = 9.25$, $P = 0.0033$
(iii) $t = 1.333$, $P = 0.1868$
(iv) $t = -1.671$, $P = 0.0992$
(v) $F = 2.535$, $P = 0.1158$

(h) Drug One is prefered over 2. It's better to have pigs with big parents.

```
R version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.79 (8198) x86_64-apple-darwin17.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

> # 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] -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
>
>
```