Math 3080 § 1.	**Pregnancy Example:Ryan-Joiner**	Name: Example
Treibergs	**and Shapiro-Wilk tests for Normality**	April 5, 2014

This **R**© program explores two normality tests, the Ryan-Joiner Test discussed in the text and the Shapiro-Wilk test canned in **R**©.

This is the same data as in the "`M3082PregnancyEg`, Pregnancy Example: χ^2 Goodness of Fit Test for Continuous Variable with Known Parameters." It was taken from Larsen and Marx, *An Introduction to Mathematical Statistics and its Applications 4th ed.,* Prentice Hall, Upper Saddle River, NJ, 2006. The duration of pregnancy is thought to be a normal variable with mean $\mu = 266$ days and a standard deviation of $\sigma = 16$ days. The authors say that for the last 70 births at the Davidson County General Hospital in Nashville had durations

```
251  264  234  283  226  244  269  241  276  274
263  243  254  276  241  232  260  248  284  253
265  235  259  279  256  256  254  256  250  269
240  261  263  262  259  230  268  284  259  261
268  268  264  271  263  259  294  259  263  278
267  293  247  244  250  266  286  263  274  253
281  286  266  249  255  233  245  266  265  264
```

Is the data plausibly from a normal distribution?

The Ryan-Joiner Test works by measuring the linearity of points on the QQ-plot which is the one generated by **R**© doing `qqplot`. The sorted data vector $x_{(i)}$ is plotted against the theoretical quantiles $y_i = \Phi^{-1}\left(\dfrac{i - .375}{n + .25}\right)$. The linearity is measured by the sample correlation r. Since the y_i coordinates are not part of a bivariate normal distribution, the distribution of r in this case is not the same as from correlation tests. For one thing, both coordinates are increasing resulting in $r \geq 0$. Under the hypothesis that the sample is normal, the distribution of r is understood and the critical values are provided in Table A12.

The null and alternative hypotheses are

$$\mathcal{H}_0 : \text{the population distribution is normal;}$$
$$\mathcal{H}_a : \text{the population distribution is not normal.}$$

The correlation coefficient r of the points $(x_{(1)}, y_1), \ldots, (x_{(n)}, y_n)$ is the test statistic. The null hypothesis is rejected if $r \leq c_{\alpha,n}$, where $c_{\alpha,n}$ are given in table A12.

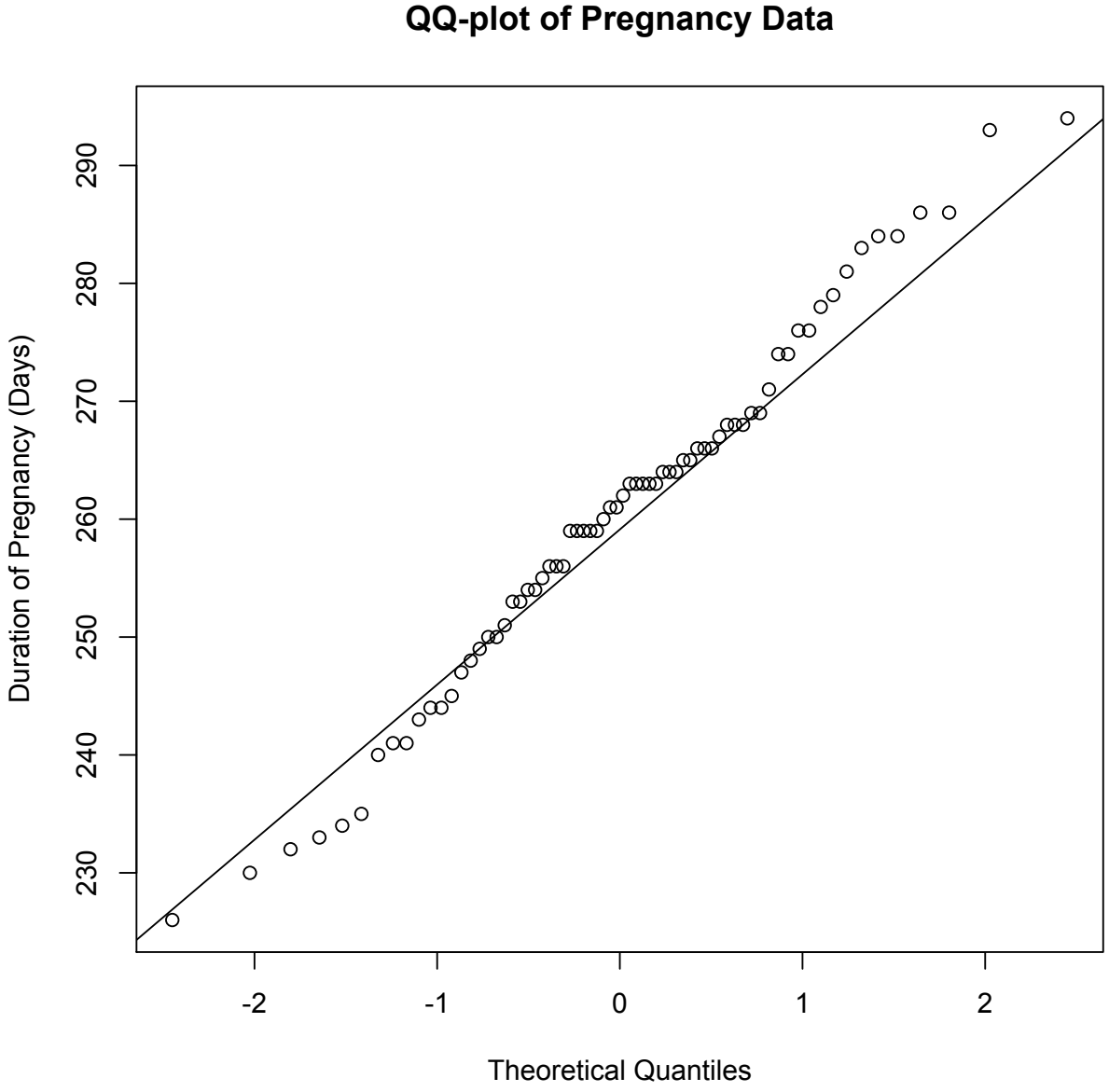
For the pregnancy data, $r = 0.995086$. The critical values for a significance level $\alpha = .10$ are $c_{.10,60} = .9835$ and $c_{.10,75} = .9865$ so interpolating, $c_{.10,70} = \frac{2}{3}c_{.10,75} + \frac{1}{3}c_{.10,60} = \frac{2}{3}(.9865) + \frac{1}{3}(.9835) = 0.9855$. In any case, $r > c_{.10,75}$ so we are unable to reject the null hypothesis even at the significance level as large as .10.

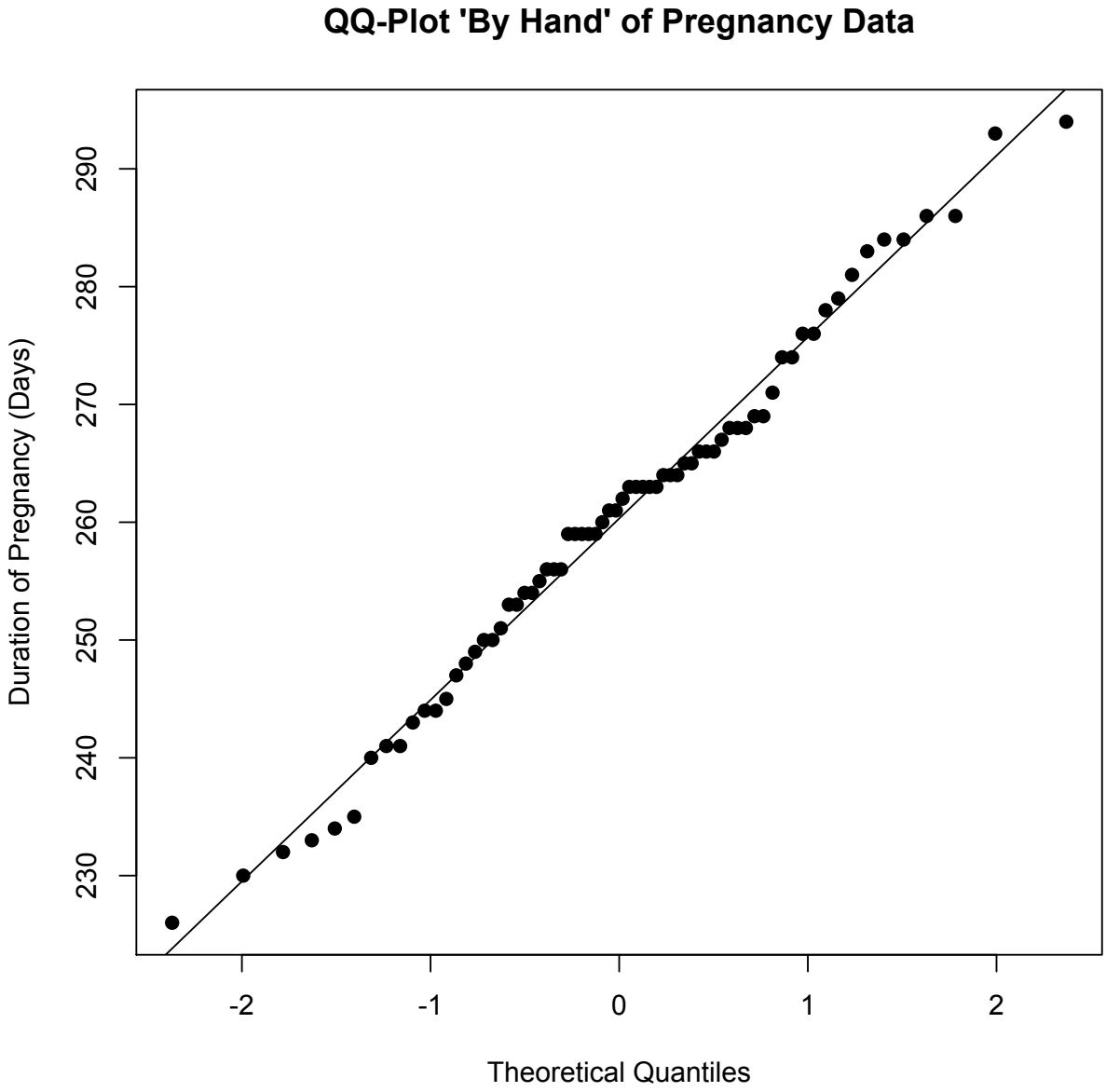
The Shapiro-Wilk test is more powerful that the Ryan Joiner test as well as the other normality tests. It is based on a different statistic W, but tests the same hypotheses and is otherwise similar. Running the Shapiro-Wilk test gives a p-value of 0.688 for the pregnancy data. We are also unable to reject the null hypothesis even at the significance level .10.

R Session:

```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-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 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrejstreibergs/.Rapp.history]

> x=scan()
1:  251 264 234 283 226 244 269 241 276 274
11: 263 243 254 276 241 232 260 248 284 253
21: 265 235 259 279 256 256 254 256 250 269
31: 240 261 263 262 259 230 268 284 259 261
41: 268 268 264 271 263 259 294 259 263 278
51: 267 293 247 244 250 266 286 263 274 253
61: 281 286 266 249 255 233 245 266 265 264
71:
Read 70 items

> #############  QQ-PLOT OF PREGNANCY DATA  ###############
> qqnorm(x,main="QQ-plot of Pregnancy Data",
         ylab="Duration of Pregnancy (Days)")
> qqline(x)
>
> ############  SORT THE DATA  ##############################
> sortx=sort(x)
> n=length(x);n
[1] 70
>
```

```
> ############  THE QUANTILES ARE BASED ON PROBABILITIES ##
> ############  (i-.375)/(n + .25)  WHCH ARE CANNED IN R ##
> ############  AS  ppoints(n,a=.375)                    ##
> (1:5-.375)/(5+.25)
[1] 0.1190476 0.3095238 0.5000000 0.6904762 0.8809524
> ppoints(5,a=.375)
[1] 0.1190476 0.3095238 0.5000000 0.6904762 0.8809524
>
> ############  COMPUTE THE THEORETICAL QUANTILES  ########
> y=qnorm(ppoints(n,a=.375))
>
> ############   QQ-PLOT "BY HAND" FOR PREGNANCY DATA #####
>
> plot(y, sortx, pch=19, main="QQ-Plot 'By Hand' of Pregnancy Data",
  ylab="Duration of Pregnancy (Days)",xlab="Theoretical Quantiles")
> abline(lm(sortx~y))
>
> ############  COMPUTE THE RYAN-JOINER  r   ###############
> r=cor(sortx,y); r
[1] 0.995086
>
> ############  INTERPOLATE THE CRITICAL TABLE VALUES  ####
> crit60=.9835
> crit75=.9865
> 75*2/3+60/3
[1] 70
> crit70=crit75*2/3+crit60/3; crit70
[1] 0.9855
>
> ############  RUN THE SHAPIRO-WILK TEST  ##################
>
> shapiro.test(x)

Shapiro-Wilk normality test

data:  x
W = 0.987, p-value = 0.688
```


QQ-plot of Pregnancy Data
Duration of Pregnancy (Days)
230
240
250
260
270
280
290
-2
-1
0
1
2
Theoretical Quantiles


QQ-Plot 'By Hand' of Pregnancy Data
Duration of Pregnancy (Days)
230
240
250
260
270
280
290
-2
-1
0
1
2
Theoretical Quantiles