

18.443 Problem Set 7 Spring 2015
Statistics for Applications
Due Date: 4/17/2015
prior to 3:00pm

Problems from John A. Rice, Third Edition. [*Chapter.Section.Problem*]
For full credit, explain your reasoning and detail any calculations.

1. 10.9.6 (See script in RProject7)

Solution: See the script/html file "Rproject7_5_beeswax_solution.r"

(a). The data appears normal with some deviations at the extreme tails. These deviations can happen with normal samples, but occur about 1 in 15 times. (It is possible to formally test for normality using the Kolmogorov-Smirnov Test, which gives a P -value of about 0.06).

(b). There are several ways to answer this question. The R script evaluates the power of level-0.05 test of H_0 (no dilution with distribution that is normal with mean and variance equal to those in the sample) versus H_1 (dilution increasing the mean hydrocarbon level). The analysis finds the power for detecting the alternative is 0.231 for dilution of 1%, 0.859 for dilution of 2% and 0.998 for dilution of 3%

An alternative way is to construct a 95% confidence interval for the mean hydrocarbon level of beeswax using the sample data. Then, evaluate whether the means of the diluted beeswax distributions are in the confidence interval. For those dilutions which have mean values outside the interval, we would expect the power of the test to be greater than 50%. The first approach provides some insight into how the power varies at different dilution levels.

2. 10.9.7 (See script in Rproject7)

Solution: See the script/html file "Rproject7_6_lifetimes_solution.r".

3. 10.9.8

Consider a sample of size 100 from an exponential distribution with $\lambda = 1$ (rate).

(a). Sketch the approximate standard deviation of the empirical log survival function, $\log S_n(t)$, as a function of t .

Solution: In lecture we derived the following expression for the variance of the survival function (see p. 384 or Rice)

$$\text{Var}(\log[S_n(t)]) = \text{Var}(\log[1 - F_n(t)]) = \frac{1}{n} \left(\frac{F(t)}{1-F(t)} \right)$$

where $F(t) = 1 - e^{-t} = P(X \leq t)$, the cdf of an *exponential*($\lambda = 1$) random variable.

The R script “Problem10_9_8.r” graphs $\sqrt{\text{Var}(\log[S_n(t)])}$ versus t .

4. 10.9.35

Solution: For a normal sample, the proportion of observations expected to be marked by an asterisk on a boxplot is equal to the percentage of a normal sample that falls more than 1.5 IQRs above the third quartile and lower than 1.5 IQRs below the lower quartile. The answer is 0.000697 (less than 1%).

```
> # For boxplot, an asterisk on the high side is the upper-probability
> # of a N(0,1) distribution corresponding to a critical value
> #which is 1.5 IQRs above the third quartile
> z0=qnorm(.75)+(qnorm(.75)-qnorm(.25))*1.5
>
> 1-pnorm(z0)
[1] 0.003488302
>
> 2*(1-pnorm(z0))
[1] 0.006976603
```

5. 10.9.39 Read in the data and generate the boxplot by executing the following code snippet in Rproject7

```
windspeed=read.table(file="Rice 3e Datasets/ASCII Comma/Chapter 10/windspeed.txt",
  sep=",",stringsAsFactors = FALSE)
windspeed.0=t(windspeed[,-1])
dimnames(windspeed.0)<-list(c(1:nrow(windspeed.0)), windspeed[,1])
dim(windspeed.0)
head(windspeed.0); tail(windspeed.0)
boxplot(windspeed.0)
```

Also, using your answer to 10.9.35, comment on the data points displayed as asterisks in the boxplots. is their frequency consistent with

a normal distribution? Would you expect these data to be consistent with a normal distribution?

Soluton: See Script "RProject8_1_windspeed.r". There are many asterisk points in the boxplots which is not consistent with normal distributions for the small samples with sizes 37. The extreme wind speeds are postively skewed distributions that differ significantly from the normal model.

6. 10.9.43 (See script RProject7_7_kevlar.r)

From the output, for the higher stresses, there is a faster rate of failure – the slopes become stronger (negative).

At 70% stress, the ESF is mostly linear but steepens at the end (indicating a faster failure rate).

This is more pronounced in the 80% stress case.

At 90% stress, the ESF is mostly linear.

MIT OpenCourseWare
<http://ocw.mit.edu>

18.443 Statistics for Applications
Spring 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.