Outline of Solution to Homework 3

Problem 1 Suppose the observation Z and the true value $S > 0$ are linked through *multiplicative* measurement error by $Z = S \cdot (\sigma\epsilon)$, where random variables S and ϵ are independent and ϵ has probability density function (pdf) $f^{(\epsilon)}$. Derive the conditional pdf $f^{(Z|S)}(z|s)$.

Solution. We have

$$\begin{aligned} F^{(Z|S)}(z \mid s) &= \Pr(Z \le z \mid S = s) = \Pr(s \cdot (\sigma\epsilon) \le z \mid S = s) = \Pr(\epsilon \le z/(s\sigma) \mid S = s) \\ &= F^{(\epsilon)}(z/(s\sigma)). \end{aligned}$$

Therefore

$$\begin{aligned} f^{(Z|S)}(z \mid s) &= \frac{\partial}{\partial z} F^{(Z|S)}(z \mid s) = \frac{\partial}{\partial z} F^{(\epsilon)}(z/(s\sigma)) \\ &= f^{(\epsilon)}(z/(s\sigma))/(s\sigma). \end{aligned}$$

Problem 2 Consider a random variable X with mean μ_X and variance σ_X^2, and let $Y = g(X)$, where $g(\cdot)$ is a twice differentiable function. Using a Taylor series expansion, it can be shown that μ_Y and σ_Y^2 can be approximated by

$$\begin{aligned} \mu_Y &\simeq g(\mu_X) + g''(\mu_X)\sigma_X^2/2 \\ \sigma_Y^2 &\simeq (g'(\mu_X))^2 \sigma_X^2, \end{aligned}$$

where g' and g'' are, respectively, the first and second derivatives of g. [This approach to finding an approximate relationship between moments of X and moments of Y is known as the δ-method.]

(i) Suppose $X \sim \text{Bin}(n, p)$, the binomial distribution with parameters n and p. Let $Y = \text{logit}(X/n)$. Find approximate expressions for μ_Y, σ_Y^2.

Solution. Let $\hat{p} \equiv X/n$. Then $E(\hat{p}) = p$ and $\text{var}(\hat{p}) = p(1-p)/n$. Therefore, $Y = \text{logit}(\hat{p}) = g(\hat{p})$, with $g(\hat{p}) \equiv \log(\hat{p}/(1-\hat{p})) = \log(\hat{p}) - \log(1-\hat{p})$. Then

$$g'(\hat{p}) = \frac{1}{\hat{p}} - \frac{1}{1-\hat{p}}(-1) = \frac{1}{\hat{p}(1-\hat{p})}$$

and

$$g''(\hat{p}) = -\left(\hat{p}(1-\hat{p})\right)^{-2}(1-2\hat{p}) = \frac{2\hat{p}-1}{\left(\hat{p}(1-\hat{p})\right)^2}.$$

Hence,

$$\mu_Y \simeq \text{logit}(p) + \frac{1}{2}\,\frac{2p-1}{\left(p(1-p)\right)^2}\,\{p(1-p)/n\} = \text{logit}(p) + \frac{2p-1}{2np(1-p)}$$

and

$$\sigma_Y^2 \simeq \left(\frac{1}{p(1-p)}\right)^2 \left(\frac{p(1-p)}{n}\right) = \frac{1}{np(1-p)}.$$

Suppose σ_X is in fact a function of μ_X. Write $\sigma_X \equiv \sigma(\mu_X)$.

(ii) Use the δ-method to show that the transformation g in $Y = g(X)$, for which $\sigma_Y \simeq$ constant, satisfies $g'(x) = \text{constant}/\sigma(x)$. [The resulting g is known as a variance-stabilizing transformation.]

Solution. Using the δ-method, we get

$$\sigma_Y^2 \simeq \left(g'(\mu_X)\right)^2 \sigma(\mu_X)^2 = \left(g'(\mu_X)\sigma(\mu_X)\right)^2 = \text{constant}.$$

This holds only if $g'(x) = c/\sigma(x)$, where c is some constant.

(iii) Suppose X is a positive random variable and $\sigma_X = a(\mu_X)^b$. Show that $g(x) = c_1 x^{1-b} + c_2$ is a variance-stabilizing transformation.

Solution. Now, $g'(x) = c_1(1-b)x^{-b}$. Therefore,

$$\begin{aligned}\sigma_Y^2 &\simeq \left(g'(\mu_X)\sigma(\mu_X)\right)^2 = \left(a(\mu_X)^b \cdot c_1(1-b)(\mu_X)^{-b}\right)^2 = (ac_1(1-b))^2 \\ &= \text{constant},\end{aligned}$$

since a, c_1, and b are constants.

Problem 3. The `data.frame`, air, in `S-PLUS` contains 111 observations (rows), and 4 variables (columns), taken from an environmental study that measured the four variables, ozone, solar radiation, temperature, and wind speed, for 111 consecutive days in New York. Ozone is the only variable of interest in (a)–(d).

(a) Obtain summary information on the data by using the `summary()` function.

(b) Make a scatter plot of ozone versus temperature and label the x-axis with "Temperature (F)", the y-axis with "Ozone (ppm)" and put on the title "The NY ozone data". Don't print this plot.

(c) Repeat the same plot command used in **(b)** but with `type='n'` (plot nothing). Then (by using the `points()` function to add to the plotting device):

(i) Plot ozone versus temperature using plot symbol number 1 (`pch=1`), but only use the observations where wind is less than its median.

(ii) Plot ozone versus temperature using the plot symbol number 2 (`pch=2`), but now use the observations where wind is greater, or equal, than its median.

(d) Add a horizontal line to the plot showing ozone's median (`abline()`) and a vertical line showing temperature's median. Use line type 2.

Bonus: before you print out your plot, execute these two commands:

```
> legend(57,5.518,c('low wind','high wind'),marks=c(1,2))
> lines(lowess(air$te,air$oz),lty=3)  ## smooth trend line
```

Hand in the final plot made and the commands used to create it.

Solution. The S-PLUS commands used for parts (a)–(d) are:

```
## for (a):
summary(air)

attach(air)

## for (b):
plot(temperature,ozone,xlab='Temperature (F)',
     ylab='Ozone (ppm)',main='NY ozone data')

## for (c)

## first, for creating a hard-copy of plot...
trellis.device(postscript,file='myplot.ps',horizontal=T)
```

```
## then start the plotting commands...
plot(temperature,ozone,type='n',xlab='Temperature (F)',
     ylab='Ozone (ppm)',main='NY ozone data')
ind <- wind < median(wind)
points(temperature[ind],ozone[ind],pch=1)
points(temperature[!ind],ozone[!ind],pch=2)

## for (d)
abline(v=median(temperature),h=median(ozone),lty=2)

## bonus:
legend(57,5.518,c(`low wind',`high wind'),marks=c(1,2))
lines(lowess(air$te,air$oz),lty=3)  ## smooth trend line

## close the plotting device...
dev.off()

detach(air)
```

This creates a postscript file `myplot.ps`. Of course, the plotting commands were first tried out on a 'regular' plotting device before using the postscript device.

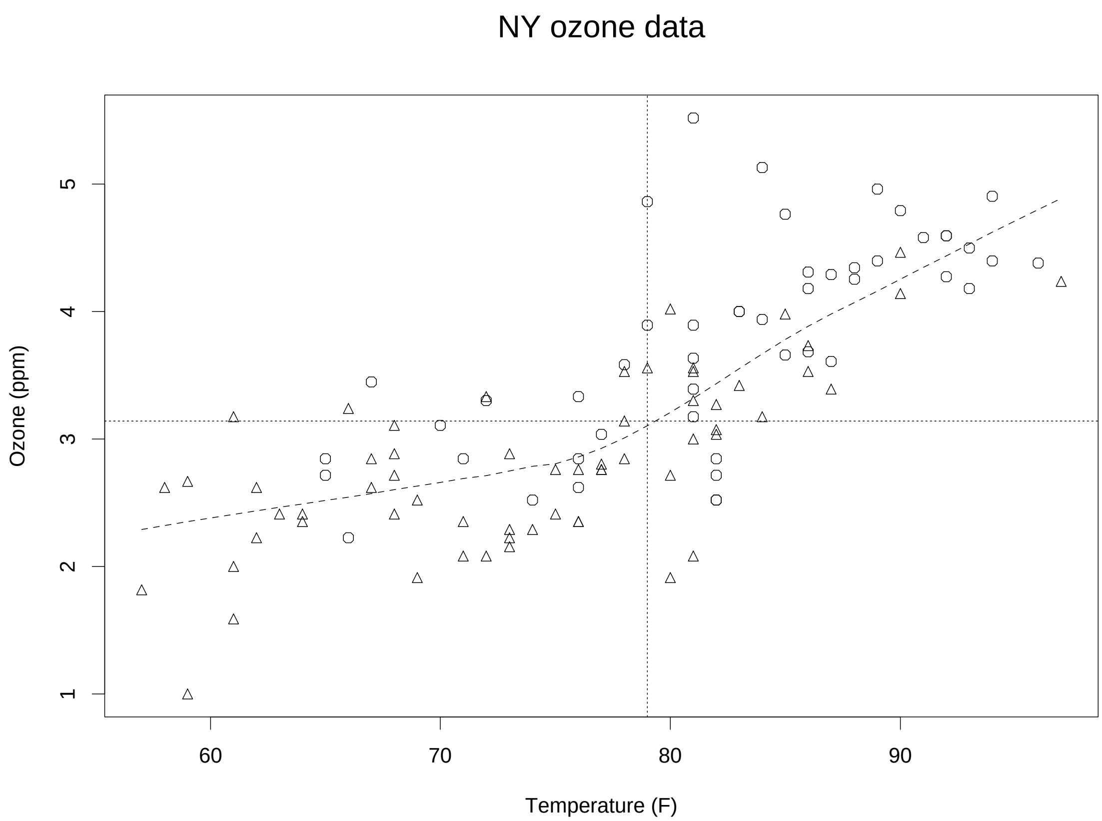

Figure 1: The resulting plot from part (b)–(d), plus the bonus part, in problem 3.