

Lesson 03

Christopher A. Swenson (chris@cswenson.com)

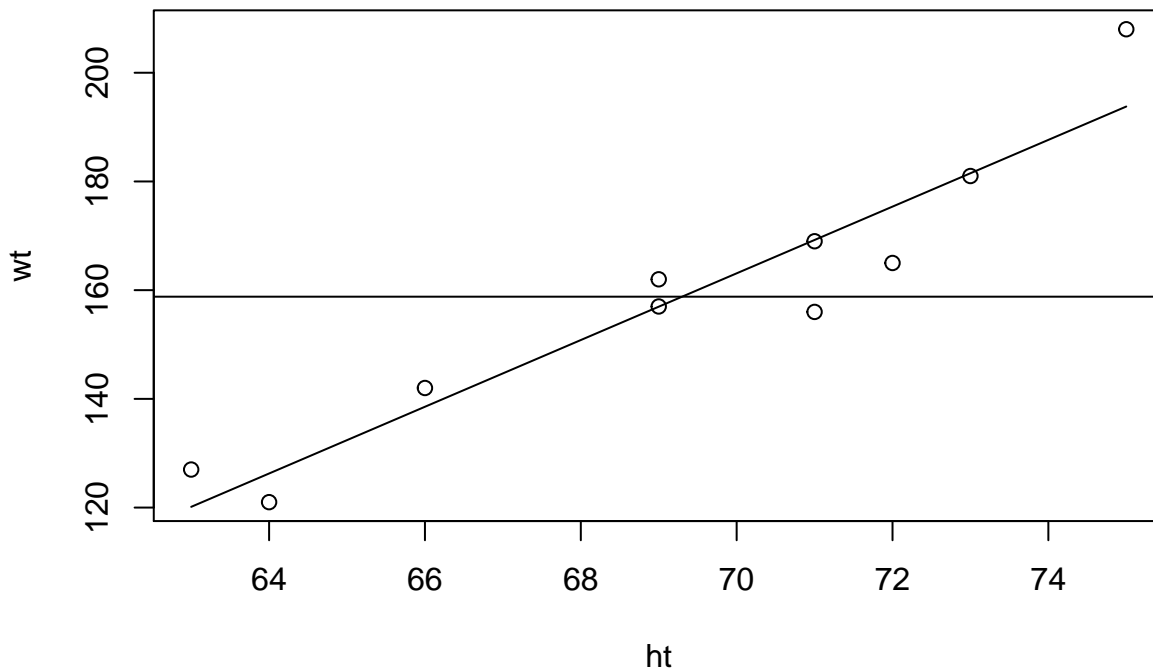
11/27/2021

Student heights and weights

Load the heightweight data. Fit a simple linear regression model with $y = \text{wt}$ and $x = \text{ht}$. Display a scatterplot of the data with the simple linear regression line and a horizontal line at the mean weight. Use the model to predict weight for height = 64.

```
heightweight <- read.table("./Data/student_height_weight.txt", header=T)
attach(heightweight)

model <- lm(wt ~ ht)
plot(x=ht, y=wt,
     panel.last = c(lines(sort(ht), fitted(model)[order(ht)]),
                    abline(h=mean(wt))))
```



```
mean(wt) # 158.8
```

```
## [1] 158.8
```

```
predict(model, newdata=data.frame(ht=64)) # 126.2708
```

```
##      1
```

```
## 126.2708
```

```
detach(heightweight)
```

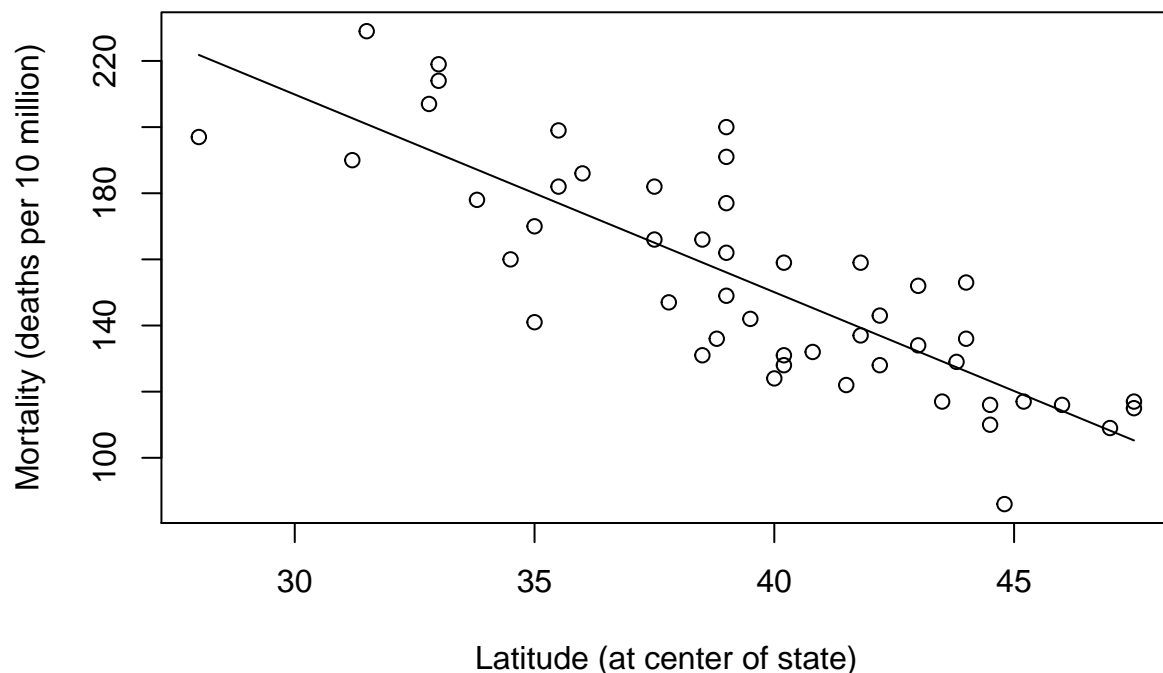
Skin cancer mortality

Load the skin cancer data. Fit a simple linear regression model with $y = \text{Mort}$ and $x = \text{Lat}$. Display a scatterplot of the data with the simple linear regression line. Use the model to calculate 95% confidence intervals for $E(\text{Mort})$ at $\text{Lat} = 40$ and 28. Calculate $\text{mean}(\text{Lat})$. Use the model to calculate 95% prediction intervals for Mort at $\text{Lat} = 40$. Display a scatterplot of the data with the simple linear regression line, confidence interval bounds, and prediction interval bounds.

```
skincancer <- read.table("./Data/skincancer.txt", header=T)
attach(skincancer)

model <- lm(Mort ~ Lat)
plot(x=Lat, y=Mort,
     xlab="Latitude (at center of state)", ylab="Mortality (deaths per 10 million)",
     main="Skin Cancer Mortality versus State Latitude",
     panel.last = lines(sort(Lat), fitted(model)[order(Lat)]))
```

Skin Cancer Mortality versus State Latitude



```
predict(model, interval="confidence", se.fit=T,
        newdata=data.frame(Lat=c(40, 28)))
```

```
## $fit
##      fit      lwr      upr
## 1 150.0839 144.5617 155.6061
## 2 221.8156 206.8855 236.7456
##
## $se.fit
##      1      2
## 2.745000 7.421459
```

```
##
## $df
## [1] 47
##
## $residual.scale
## [1] 19.11503

# $fit
#      fit      lwr      upr
# 1 150.0839 144.5617 155.6061
# 2 221.8156 206.8855 236.7456
#
# $se.fit
# 1      2
# 2.745000 7.421459

mean(Lat) # 39.53265

## [1] 39.53265

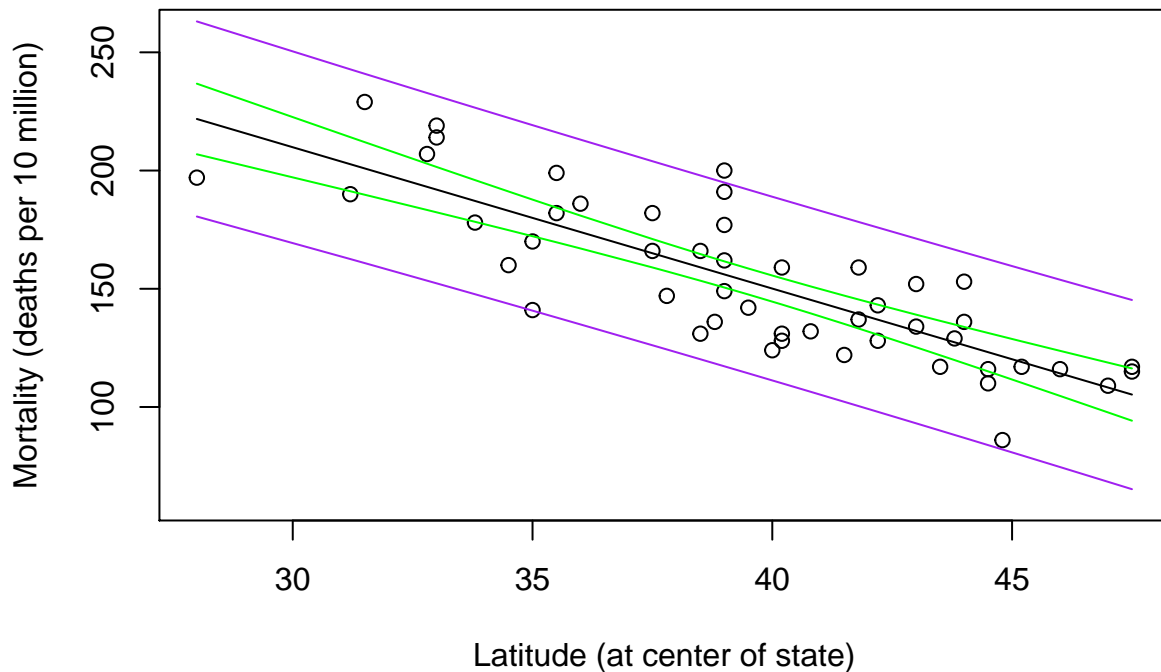
predict(model, interval="prediction",
        newdata=data.frame(Lat=40))

##      fit      lwr      upr
## 1 150.0839 111.235 188.9329

#      fit      lwr      upr
# 1 150.0839 111.235 188.9329

plot(x=Lat, y=Mort,
     xlab="Latitude (at center of state)", ylab="Mortality (deaths per 10 million)",
     ylim=c(60, 260),
     panel.last = c(lines(sort(Lat), fitted(model)[order(Lat)]),
                    lines(sort(Lat),
                          predict(model,
                                interval="confidence")[order(Lat), 2], col="green"),
                    lines(sort(Lat),
                          predict(model,
                                interval="confidence")[order(Lat), 3], col="green"),
                    lines(sort(Lat),
                          predict(model,
                                interval="prediction")[order(Lat), 2], col="purple"),
                    lines(sort(Lat),
                          predict(model,
                                interval="prediction")[order(Lat), 3], col="purple"))))

## Warning in predict.lm(model, interval = "prediction"): predictions on current data refer to _future_
## Warning in predict.lm(model, interval = "prediction"): predictions on current data refer to _future_
```

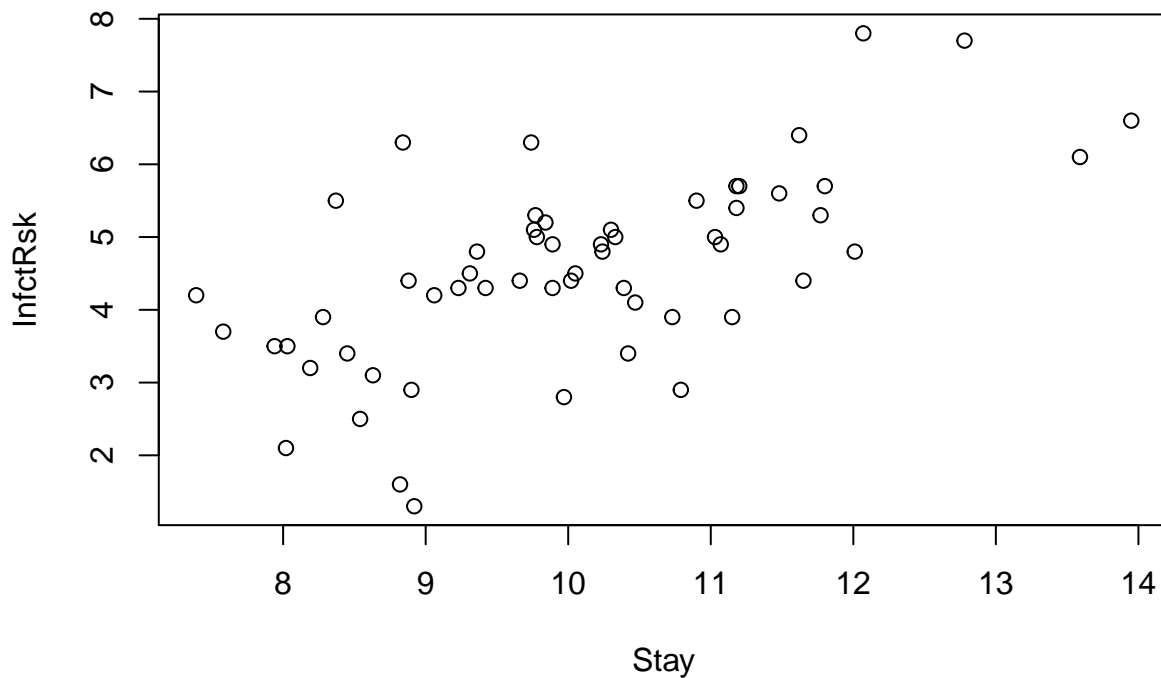


```
detach(skincancer)
```

Hospital infection risk

Load the infectionrisk data. Select only hospitals in regions 1 or 2. Display a scatterplot of Stay versus InfctRsk. Select only hospitals with Stay < 16 (i.e., remove the two hospitals with extreme values of Stay). Fit a simple linear regression model with $y = \text{InfctRsk}$ and $x = \text{Stay}$. Use the model to calculate 95% confidence intervals for $E(\text{InfctRsk})$ at Stay = 10. Use the model to calculate 95% prediction intervals for InfctRsk at Stay = 10. Display a scatterplot of the data with the simple linear regression line, confidence interval bounds, and prediction interval bounds.

```
infectionrisk <- read.table("./Data/infectionrisk.txt", header=T)
infectionrisk <- infectionrisk[infectionrisk$Region==1 | infectionrisk$Region==2, ]
attach(infectionrisk)
plot(x=Stay, y=InfctRsk)
detach(infectionrisk)
infectionrisk <- infectionrisk[infectionrisk$Stay<16, ]
attach(infectionrisk)
plot(x=Stay, y=InfctRsk)
```



```
model <- lm(InfctRsk ~ Stay)
```

```
predict(model, interval="confidence",
        newdata=data.frame(Stay=10))
```

```
##          fit      lwr      upr
## 1 4.528846 4.259205 4.798486
```

```
#          fit      lwr      upr
# 1 4.528846 4.259205 4.798486
```

```
predict(model, interval="prediction",
        newdata=data.frame(Stay=10))
```

```
##          fit      lwr      upr
## 1 4.528846 2.45891 6.598781
```

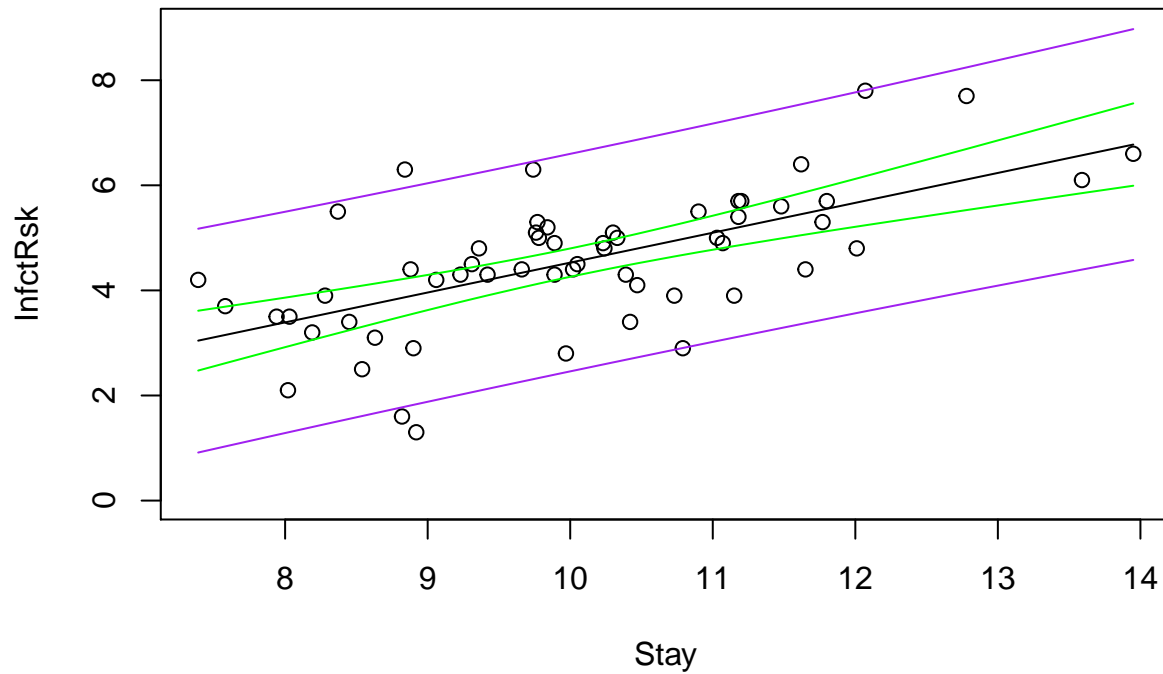
```
#          fit      lwr      upr
# 1 4.528846 2.45891 6.598781
```

```
plot(x=Stay, y=InfctRsk,
     ylim=c(0, 9),
     panel.last = c(lines(sort(Stay), fitted(model)[order(Stay)]),
                    lines(sort(Stay),
                          predict(model,
                                interval="confidence")[order(Stay), 2], col="green"),
                    lines(sort(Stay),
                          predict(model,
                                interval="confidence")[order(Stay), 3], col="green"),
                    lines(sort(Stay),
                          predict(model,
                                interval="prediction")[order(Stay), 2], col="purple"),
                    lines(sort(Stay),
                          predict(model,
```

```
interval="prediction")[order(Stay), 3], col="purple"))
```

```
## Warning in predict.lm(model, interval = "prediction"): predictions on current data refer to _future_
```

```
## Warning in predict.lm(model, interval = "prediction"): predictions on current data refer to _future_
```



```
detach(infectionrisk)
```