Stat 415 Regression: Classwork/Lab 2

Frankie Tyndall

2022-06-29

## Study of Airfreight breakage

A substance used in biological and medical research is shipped by airfreight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y).Assume that first-order regression model is appropriate.

x = The number of times the carton was transferred from one aircraft to another over the shipment route

y = The number of ampules found to be broken upon arrival

airfreight\_shipments <- tribble(~Xi, ~Yi,  
 1 , 16,  
 0 , 9,  
 2 , 17,  
 0 , 12,  
 3 , 22,  
 1 , 13,  
 0 , 8,  
 1 , 15,  
 2 , 19,  
 0 , 11  
 )   
  
airfreight\_shipments

## # A tibble: 10 × 2  
## Xi Yi  
## <dbl> <dbl>  
## 1 1 16  
## 2 0 9  
## 3 2 17  
## 4 0 12  
## 5 3 22  
## 6 1 13  
## 7 0 8  
## 8 1 15  
## 9 2 19  
## 10 0 11

## Hypothesis for Simple Linear Regression for Airfreight breakage Study

H0: β = 0 null hypothesis is that the population slope is equal to 0 HA: β ≠ 0 alternative hypothesis is that the population slope does not equal 0

## Obtain the estimated regression function for the bivariate data given in the table above.Using R coding, find the 95% confidence interval for your slope coefficient.

airfreight\_shipments\_data <- lm(airfreight\_shipments$Yi ~ airfreight\_shipments$Xi)  
airfreight\_shipments\_data

##   
## Call:  
## lm(formula = airfreight\_shipments$Yi ~ airfreight\_shipments$Xi)  
##   
## Coefficients:  
## (Intercept) airfreight\_shipments$Xi   
## 10.2 4.0

airfreight\_shipments\_data1 <- tidy(airfreight\_shipments\_data, conf.int = TRUE)  
select(airfreight\_shipments\_data1, term, estimate, p.value, conf.low, conf.high)

## # A tibble: 2 × 5  
## term estimate p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 10.2 0.000000318 8.67 11.7   
## 2 airfreight\_shipments$Xi 4 0.0000275 2.92 5.08

The estimated regression function: y(hat) = 10.2 + 4Xi

Interpretation: We are 95% confident that the true population slope falls between 2.918388 and 5.081612

## Slope and y-intercept

slope: 4

y-intercept: 10.2

The slope indicates that for every increase by 1 transfer (Xi), the number of ampules found to be broken upon arrival (Yi) increases by 4 transfers on average.

## Using 4 or 5 sentences, explain the difference between a prediction interval and a confidence interval.

The difference between a prediction interval and a confidence interval is that a predictor interval is usually more wider than the confidence interval because there is more uncertainty. Confidence interval pertains to the slope estimated from multiple values while trying to estimate the population.The predictor interval expresses uncertainty in a specific data point as well as sample. Confidence interval of points on the line, while predictor interval is for estimating the value of a single observational unit and interval for points above or below the regression line.

## Find the prediction interval for a response value generated by inputting 19 for X into the model you generated in part a

airfreight\_data <- lm(Yi ~ Xi, data = airfreight\_shipments)  
airfreight\_data

##   
## Call:  
## lm(formula = Yi ~ Xi, data = airfreight\_shipments)  
##   
## Coefficients:  
## (Intercept) Xi   
## 10.2 4.0

newdf <- data.frame(Xi = 19)  
newdf

## Xi  
## 1 19

predict(object = airfreight\_data, newdata = newdf, interval = "prediction") %>% cbind(newdf)

## fit lwr upr Xi  
## 1 86.2 66.40325 105.9967 19

## Find the confidence interval for a response value generated by inputting 19 for X into the model you generated in part a

airfreight\_data <- lm(Yi ~ Xi, data = airfreight\_shipments)  
airfreight\_data

##   
## Call:  
## lm(formula = Yi ~ Xi, data = airfreight\_shipments)  
##   
## Coefficients:  
## (Intercept) Xi   
## 10.2 4.0

newdf <- data.frame(Xi = 19)  
newdf

## Xi  
## 1 19

predict(object = airfreight\_data, newdata = newdf, interval = "confidence") %>% cbind(newdf)

## fit lwr upr Xi  
## 1 86.2 66.70097 105.699 19