ALSM: Chapter 4

Simultaneous Inferences and Other Topics in Regression Analysis

Darshan Patel

12/27/2020

```
library(tidyverse)
library(latex2exp)
library(gridExtra)
library(wesanderson)
library(broom)
theme_set(theme_minimal())
```

Problem 1:

When joint confidence intervals for β_0 and β_1 are developed by the Bonferroni method with a family confidence coefficient of 90 percent, does this imply that 10 percent of the time the confidence interval for β_0 will be incorrect? That 5 percent of the time the confidence interval for β_0 will be incorrect and 5 percent of the time that for β_1 will be incorrect? Discuss.

Answer: When joint confidence intervals for β_0 and β_1 are developed by the Bonferroni method with a family confidence coefficient of 90 percent, it implies that both intervals based on the same sample are correct 95 percent of the time. In the other 5 percent of times, one or both of the interval estimates would be incorrect. This is derived from the Bonferroni inequality. If A_1 and A_2 are two events, say estimate β_0 and β_1, then

$$P(\overline{A}_1 \cap \overline{A}_2) \geq 1 - 2\alpha$$

Problem 2:

Refer to Problem 2.1. Suppose the student combines the two confidence intervals into a confidence set. What can you say about the family confidence coefficient for this set?

Answer: The family confidence coefficient of this set is 90 percent. The α values for estimating β_0 and β_1 are .05 each and so

$$P(\overline{A}_1 \cap \overline{A}_2) \geq 1 - 2\alpha = 1 - 2(.05) = .9$$

Problem 3:

Refer to *Copier maintenance* Problem 1.20.

(a) Will $b_0 $ and b_1 tend to err in the same direction or in opposite directions here? Explain.

Answer: The covariance between b_0 and b_1 is given by

$$\sigma[b_0, b_1] = -\overline{X}^2 \sigma^2[b_1]$$

Hence if $\overline{X}$ is positive, then b_0 and $b1$ will err in opposite directions.

```r
copier = read.csv('CH01PR20.txt', sep = '', header = FALSE,
                  col.names = c('y', 'x'),
                  colClasses = c('numeric', 'numeric'))

mean(copier$x)
```

```
## [1] 5.111111
```

Since the mean of X is positive, b_0 and b_1 will tend to err in the opposite direction here, meaning they're negatively correlated.

(b) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

Answer:

```r
lm.fit_manual = function(X, Y){
  b1 = sum((X - mean(X))*(Y - mean(Y))) / (sum((X - mean(X))^2))
  b0 = mean(Y) - b1*mean(X)
  return(c(b0, b1))
}

bonferroni.joint.ci = function(data, alpha){

  model = lm.fit_manual(data$x, data$y)
  B = qt(1 - (alpha/4), nrow(data)-2)

  pred = model[1] + model[2]*data$x
  MSE = sum((data$y - pred)^2)/(nrow(data)-2)

  s2_b1 = MSE / sum((data$x - mean(data$x))^2)
  s2_b0 = MSE * ((1/nrow(data)) + (mean(data$x)^2/(sum((data$x - mean(data$x))^2))))


  b0_lower = round(model[1] - (B*sqrt(s2_b0)), 3)
  b0_upper = round(model[1] + (B*sqrt(s2_b0)), 3)

  b1_lower = round(model[2] - (B*sqrt(s2_b1)), 3)
  b1_upper = round(model[2] + (B*sqrt(s2_b1)), 3)

  paste("At the alpha level of", alpha, "the Bonferroni joint confidence intervals are:",
        b0_lower, "<= b0 <=", b0_upper, "and",
        b1_lower, "<= b1 <=", b1_upper, sep = ' ')
}

bonferroni.joint.ci(copier, 0.05)
```

```
## [1] "At the alpha level of 0.05 the Bonferroni joint confidence intervals are: -7.093 <= b0 <= 5.932
```

(c) A consultant has suggested that β_0 should be 0 and β_1 should equal 14.0. Do your joint confidence intervals in part (b) support this view?

Answer: The joint confidence intervals in part (b) support this view. Both β_0 and β_1 fall in the two interval estimates.

Problem 4:

Refer to *Airfreight breakage* Problem 1.21.

(a) Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

Answer:

```
airfreight = read.csv('CH01PR21.txt', sep = '', header = FALSE,
                      col.names = c('y', 'x'),
                      colClasses = c('numeric', 'numeric'))

mean(airfreight$x)
```

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

Since the mean of X is positive, b_0 and b_1 will tend to err in the opposite directions as they're negatively correlated.

(b) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 99 percent family confidence coefficient. Interpret your confidence intervals.

Answer:

```
bonferroni.joint.ci(airfreight, .01)
```

```
## [1] "At the alpha level of 0.01 the Bonferroni joint confidence intervals are: 7.658 <= b0 <= 12.742
```

At a family confidence of 99 percent, β_0 is between 7.658 and 12.742, and β_1 is between 2.202 and 5.798.

Problem 5:

Refer to *Plastic hardness* Problem 1.22.

(a) Obtain Bonferroni join confidence intervals for β_0 and β_1, using a 90 percent family confidence coefficient. Interpret your confidence intervals.

Answer:

```
plastic = read.csv('CH01PR22.txt', sep = '', header = FALSE,
                   col.names = c('y', 'x'),
                   colClasses = c('numeric', 'numeric'))

bonferroni.joint.ci(plastic, .1)
```

```
## [1] "At the alpha level of 0.1 the Bonferroni joint confidence intervals are: 162.901 <= b0 <= 174.2
```

At a family confidence of 90 percent, β_0 is between 162.901 and 174.299, and β_1 is between 1.84 and 2.228.

(b) Are b_0 and b_1 positively or negatively correlated here? Is this reflected in your joint confidence intervals in part (a)?

Answer:

```
mean(plastic$x)
```

```
## [1] 28
```

b_0 and b_1 are negatively correlated here, as the mean of X is positive. This is reflected in the coefficient estimates. The estimates for β_0 are high compared to the estimates for β_1.

(c) What is the meaning of the family confidence coefficient in part (a)?

Answer: A family confidence of 90 percent implies that both estimates, b_0 and b_1, will fall within the intervals determined 90 percent of the time. In the other 10 percent of times, either one or both estimates will be outside of the confidence limits.

Problem 6:

Refer to *Muscle mass* Problem 1.27.

(a) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 99 percent family confidence coefficient. Interpret your confidence intervals.

Answer:

```
muscle = read.csv('CH01PR27.txt', sep = '', header = FALSE,
                  col.names = c('y', 'x'),
                  colClasses = c('numeric', 'numeric'))

bonferroni.joint.ci(muscle, .01)
```

```
## [1] "At the alpha level of 0.01 the Bonferroni joint confidence intervals are: 140.26 <= b0 <= 172.43
```

At a family confidence of 99 percent, β_0 is between 140.26 and 172.434, and β_1 is between -1.453 and -0.927.

(b) Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

Answer:

```
mean(muscle$x)
```

```
## [1] 59.98333
```

Since the mean of X is positive, b_0 and b_1 will err in opposite directions, as they are negatively correlated.

(c) A researcher has suggested that β_0 should equal approximately 160 and that β_1 should be between -1.9 and -1.5. Do the joint confidence intervals in part (a) support this expectation?

Answer: The joint confidence intervals in part (a) do not support this expectation.

Problem 7:

Refer to *Copier maintenance* Problem 1.20.

(a) Estimate the expected number of minutes spent when there are 3, 5, and 7 copiers to be serviced, respectively. Use interval estimates with a 90 percent family confidence coefficient based on the Working-Hotelling procedure.

Answer:

```
working.hotelling.reg.band = function(data, alpha, Xh){

  model = lm.fit_manual(data$x, data$y)
  preds = model[1] + model[2]*data$x
  MSE = sum((data$y - preds)^2) / (nrow(data)-2)

  W = sqrt(2 * qf(1 - alpha, 2, nrow(data)-2))
  Yh = model[1] + model[2]*Xh
  s2_Yh = MSE * ((1/nrow(data)) + (((Xh - mean(data$x))^2)/ sum((data$x - mean(data$x))^2)))

  Yh_lower = round(Yh - (W*sqrt(s2_Yh)), 3)
  Yh_upper = round(Yh + (W*sqrt(s2_Yh)), 3)

  paste("Using a", alpha*100, "percent alpha value and the Working-Hotelling procedure, Y_h is estimate
```

```
}

working.hotelling.reg.band(copier, .1, 3)
```

```
## [1] "Using a 10 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```
working.hotelling.reg.band(copier, .1, 5)
```

```
## [1] "Using a 10 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```
working.hotelling.reg.band(copier, .1, 7)
```

```
## [1] "Using a 10 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

(b) Two service calls for preventive maintenance are scheduled in which the numbers of copiers to be serviced are 4 and 7, respectively. A family of prediction intervals for the times to be spent on these calls is desired with a 90 percent family confidence coefficient. Which procedure, Scheffé or Bonferroni, will provide tighter prediction limits here?

Answer:

```
scheffe.coef = function(alpha, g, n){
  paste("The Scheffé procedure gives an estimate of",
        round(sqrt(g*qf(1 - alpha, g, n - 2)), 3),
        "for the multiple of the estimated standard deviation",
        sep = ' ')
}
bonferroni.coef = function(alpha, g, n){
  paste("The Bonferroni procedure gives an estimate of",
        round(qt(1 - (alpha/(2*g)), n-2), 3),
        "for the multiple of the estimated standard deviation",
        sep = ' ')
}

scheffe.coef(.1, 2, nrow(copier))
```

```
## [1] "The Scheffé procedure gives an estimate of 2.205 for the multiple of the estimated standard devi
```

```
bonferroni.coef(.1, 2, nrow(copier))
```

```
## [1] "The Bonferroni procedure gives an estimate of 2.017 for the multiple of the estimated standard d
```

The Bonferroni procedure provides the tighter prediction limits here.

(c) Obtain the family of prediction intervals required in part (b), using the most efficient procedure.

Answer:

```
family.pred.interval = function(data, alpha, g, Xh, method = c("Scheffe", "Bonferroni", "best")){

  model = lm.fit_manual(data$x, data$y)
  preds = model[1] + model[2]*data$x
  MSE = sum((data$y - preds)^2) / (nrow(data)-2)

  Yh = model[1] + model[2]*Xh
  s2_pred = MSE * (1 + (1/nrow(data)) +
                     (((Xh - mean(data$x))^2)/sum((data$x - mean(data$x))^2)))

  S = sqrt(g * qf(1 - alpha, g, nrow(data)-2))
```

```r
  B = qt(1 - (alpha / (2 * g)), nrow(data)-2)

  if(method == "Scheffe"){
    multiplier = S
  }
  else if(method == "Bonferroni"){
    multiplier = B
  }
  else{
    if(S > B){
      method = "Bonferroni (best of two)"
      multiplier = B
    }
    else{
      method = "Scheffe (best of two)"
      multiplier = S
    }
  }

  lower_bound = round(Yh - (multiplier*sqrt(s2_pred)), 3)
  upper_bound = round(Yh + (multiplier*sqrt(s2_pred)), 3)

  paste("Using a", alpha*100, "percent alpha value and the", method,
        "procedure, Y_h is estimated to lie between", lower_bound,
        "and", upper_bound, sep = ' ')
}

family.pred.interval(copier, .1, 2, 4, "Bonferroni")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 41
```

```r
family.pred.interval(copier, .1, 2, 7, "Bonferroni")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 86
```

Problem 8:

Refer to *Airfreight breakage* Problem 1.21.

(a) It is desired to obtain interval estimates of the mean number of broken ampules when there are 0, 1, and 2 transfers for a shipment, using a 95 percent family confidence coefficient. Obtain the desired confidence intervals, using the Working-Hotelling procedure.

Answer:

```r
working.hotelling.reg.band(airfreight, .05, 0)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```r
working.hotelling.reg.band(airfreight, .05, 1)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```r
working.hotelling.reg.band(airfreight, .05, 2)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

(b) Are the confidence intervals obtained in part (a) more efficient than Bonferroni intervals here? Explain.

Answer:

```
family.pred.interval(airfreight, .05, 3, 0, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 5.3
```

```
family.pred.interval(airfreight, .05, 3, 1, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 9.50
```

```
family.pred.interval(airfreight, .05, 3, 2, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 13.3
```

The confidence intervals obtained in part (a) are more efficient than Bonferroni intervals, since the width of the interval is tighter when using the Working-Hotelling procedure.

(c) The next three shipments will make 0, 1, and 2 transfers, respectively. Obtain prediction intervals for the number of broken ampules for each of these three shipments, using the Scheffé procedure and a 95 percent family confidence coefficient.

Answer:

```
family.pred.interval(airfreight, .05, 3, 0, "Scheffe")
```

```
## [1] "Using a 5 percent alpha value and the Scheffe procedure, Y_h is estimated to lie between 4.525 a
```

```
family.pred.interval(airfreight, .05, 3, 1, "Scheffe")
```

```
## [1] "Using a 5 percent alpha value and the Scheffe procedure, Y_h is estimated to lie between 8.767 a
```

```
family.pred.interval(airfreight, .05, 3, 2, "Scheffe")
```

```
## [1] "Using a 5 percent alpha value and the Scheffe procedure, Y_h is estimated to lie between 12.525
```

(d) Would the Bonferroni procedure have been more efficient in developing the prediction intervals in part (c)? Explain.

Answer:

```
bonferroni.coef(.05, 3, nrow(airfreight))
```

```
## [1] "The Bonferroni procedure gives an estimate of 3.016 for the multiple of the estimated standard d
```

```
scheffe.coef(.05, 3, nrow(airfreight))
```

```
## [1] "The Scheffé procedure gives an estimate of 3.493 for the multiple of the estimated standard devi
```

The Bonferroni procedure would have been more efficient in developing the prediction intervals, compared to the ones made in part (c). The widths of the limits are larger when using the Scheffé procedure.

Problem 9:

Refer to *Plastic hardness* Problem 1.22.

(a) Management wishes to obtain interval estimates of the mean hardness when the elapsed time is 20, 30, and 40 hours, respectively. Calculate the desired confidence intervals using the Bonferroni procedure and a 90 percent family confidence coefficient. What is the meaning of the family confidence coefficient here?

Answer:

```
family.pred.interval(plastic, .1, 3, 20, "Bonferroni")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 20
```

```r
family.pred.interval(plastic, .1, 3, 30, "Bonferroni")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 22
```

```r
family.pred.interval(plastic, .1, 3, 40, "Bonferroni")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 24
```

The family confidence coefficient here means that there is a 90 percent probability that Y will lie in these limits for $X = 20$, 30 and 40 hours, respectively.

(b) Is the Bonferroni procedure employed in part (a) the most efficient one that could be employed here? Explain.

Answer:

```r
working.hotelling.coef = function(alpha, n){
  paste("The Working-Hotelling procedure gives an estimate of",
        round(sqrt(2 * qf(1 - alpha, 2, n-2)), 3),
        "for the multiple of the estimated standard deviation",
        sep = ' ')
}
working.hotelling.coef(.1, nrow(plastic))
```

```
## [1] "The Working-Hotelling procedure gives an estimate of 2.335 for the multiple of the estimated st
```

```r
bonferroni.coef(.1, 3, nrow(plastic))
```

```
## [1] "The Bonferroni procedure gives an estimate of 2.36 for the multiple of the estimated standard d
```

The Bonferroni procedure is not the most efficient one to use in this scenario as it gives a wider width for the confidence band compared to the Working-Hotelling procedure.

(c) The next two test items will be measured after 30 and 40 hours of elapsed time, respectively. Predict the hardness for each of these two items, using the most efficient procedure and a 90 percent family confidence coefficient.

Answer:

```r
family.pred.interval(plastic, .1, 2, 30, "best")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni (best of two) procedure, Y_h is estimated to l
```

```r
family.pred.interval(plastic, .1, 2, 40, "best")
```

```
## [1] "Using a 10 percent alpha value and the Bonferroni (best of two) procedure, Y_h is estimated to l
```

Between the Bonferroni and Scheffe procedures, the Bonferroni procedure returns the more efficient interval estimates.

Problem 10:

Refer to *Muscle mass* Problem 1.27.

(a) The nutritionist is particularly interested in the mean muscle mass for women aged 45, 55 and 65. Obtain joint confidence intervals for the means of interest using the Working-Hotelling procedure and a 95 percent family confidence coefficient.

Answer:

```r
working.hotelling.reg.band(muscle, .05, 45)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```r
working.hotelling.reg.band(muscle, .05, 55)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

```r
working.hotelling.reg.band(muscle, .05, 65)
```

```
## [1] "Using a 5 percent alpha value and the Working-Hotelling procedure, Y_h is estimated to lie betw
```

(b) Is the Working-Hotelling procedure the most efficient one to be employed in part (a)? Explain.

Answer:

```r
working.hotelling.coef(.05, nrow(muscle))
```

```
## [1] "The Working-Hotelling procedure gives an estimate of 2.512 for the multiple of the estimated st
```

```r
bonferroni.coef(.05, 3, nrow(muscle))
```

```
## [1] "The Bonferroni procedure gives an estimate of 2.465 for the multiple of the estimated standard
```

The Bonferroni procedure provides a tighter width on the confidence interval than the Working-Hotelling procedure.

(c) Three additional women aged 48, 59 and 74 have contacted the nutritionist. Predict the muscle mass for each of these three women using the Bonferroni procedure and a 95 percent family confidence coefficient.

Answer:

```r
family.pred.interval(muscle, .05, 3, 48, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 78.
```

```r
family.pred.interval(muscle, .05, 3, 59, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 65.
```

```r
family.pred.interval(muscle, .05, 3, 74, "Bonferroni")
```

```
## [1] "Using a 5 percent alpha value and the Bonferroni procedure, Y_h is estimated to lie between 47.
```

(d) Subsequently, the nutritionist wishes to predict the muscle mass for a fourth woman aged 64, with a family confidence coefficient of 95 percent for the four predictions. Will the three prediction intervals in part (c) have to be recalculated? Would this also be true if the Scheffé procedure had been used in constructing the prediction intervals?

Answer: Under this new situation, the three prediction intervals in part (b) will have to be recalculated as $X = 64$ is not one of the X_hs predicted for. In addition, if the Scheffé procedure is to be used, the prediction intervals in part (c) would have to be recalculated as well, since the multiplier for the standard deviation is from a different distribution.

Problem 11:

A behavioral scientist said, "I am never sure whether the regression line goes through the origin. Hence I will not use such a model." Comment.

Answer: If a regression line is destined to go through the origin, the model will attempt to fixate the b_0 estimate to be as close to the origin as could be possible, given the other points in the data.

Problem 12:

Typographical errors. A firm specializing in technical manuscripts has a random sample of recent orders, where X is the number of galleys for a manuscript and Y is the total dollar cost of correcting typographical errors.

Since Y involves variable costs only, an analyst wished to determine whether regression-through-the-origin model is appropriate for studying the relationship between the two variables.

(a) Fit the regression model prescribed above and state the estimated regression function.

Answer:

```
typo_errors = read.csv('CH04PR12.txt', sep = '', header = FALSE,
                       col.names = c('y', 'x'),
                       colClasses = c('numeric', 'numeric'))

lm.fit.origin_manual = function(x, y){
  b1 = sum(x*y)/sum(x^2)
  return(b1)
}

model = round(lm.fit.origin_manual(typo_errors$x, typo_errors$y), 3)
paste("Y = ", model, "x", sep = '')
```

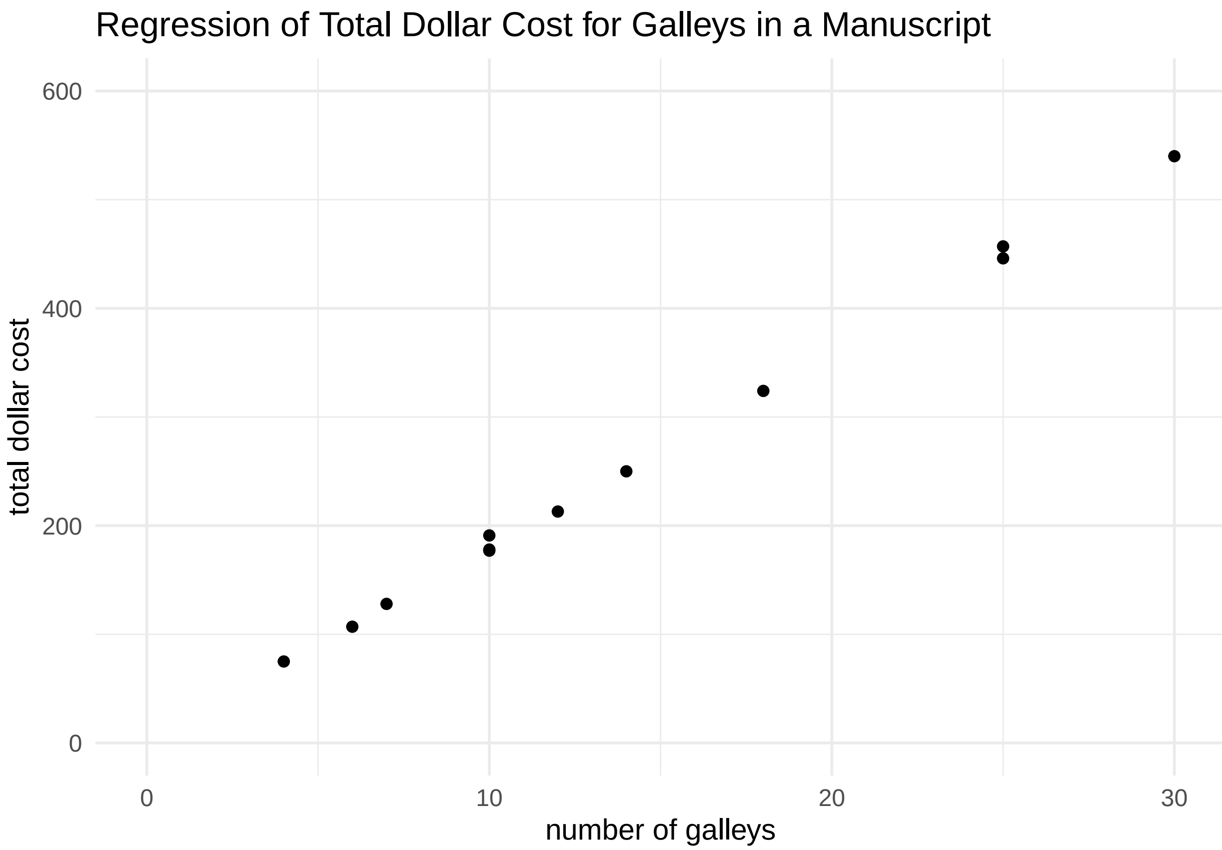
```
## [1] "Y = 18.028x"
```

(b) Plot the estimated regression function and the data. Does a linear regression function thought the origin appear to provide a good fit here? Comment.

Answer:

```
typo_errors %>%
  ggplot(aes(x,y)) +
  geom_point() +
  geom_abline(intercept = model[1], slope = model[2],
              color = "cadetblue") +
  scale_x_continuous(limits = c(0, 30)) +
  scale_y_continuous(limits = c(0, 600)) +
  labs(x = "number of galleys",
       y = "total dollar cost",
       title = "Regression of Total Dollar Cost for Galleys in a Manuscript")
```

```
## Warning: Removed 1 rows containing missing values (geom_abline).
```



By showing the origin on the plot, it is apparent that a linear regression function through the origin provides a good fit to the data.

(c) In estimating costs of handling prospective orders, management has used a standard of \$17.50 per galley for the cost of correcting typographical errors. Test whether or not this standard should be revised; use $\alpha = .02$. State the alternatives, decision rules and conclusion.

Answer: Let the null hypothesis be that $\beta_1 = \$17.50$ and the alternative hypothesis be that $\beta_1 \neq \$17.50$. Then

```
b1.origin.test = function(data, beta1 = 0, alpha = .01){

  model = lm.fit.origin_manual(data$x, data$y)
  pred = model*data$x
  MSE = sum((data$y - pred)^2) / (nrow(data)-1)
  s2_b1 = MSE / sum((data$x)^2)
  t_ast = round((beta1 - model) /sqrt(s2_b1), 3)
  t = qt(1 - (alpha/2), nrow(data)-1)

  if(abs(t_ast) > t){
    paste("At the alpha level of", alpha, "the test statistic is", round(t_ast, 3),
          "and the decision is to reject H_0.", sep = ' ')
  }
  else{
    paste("At the alpha level of", alpha, "the test statistic is", round(t_ast, 3),
          "and the decision is to fail to reject H_0.", sep = ' ')
  }
}
```

```r
b1.origin.test(typo_errors, beta1 = 17.50, alpha = .02)
```

```
## [1] "At the alpha level of 0.02 the test statistic is -6.647 and the decision is to reject H_0."
```

It can be said with 98 percent confidence that b_1 is not \$17.50.

(d) Obtain a prediction interval for the correction cost on a forthcoming job involving 10 galleys. Use a confidence coefficient of 98 percent.

Answer:

```r
Yhat_pred_interval_origin = function(data, Xhat, alpha = 0.01){

  model = lm.fit.origin_manual(data$x, data$y)
  pred = model*data$x
  Yhat = model*Xhat
  MSE = sum((data$y - pred)^2) / (nrow(data)-1)
  s2_Yhat_new = MSE * (1 + (Xhat^2/sum(data$x^2)))
  t = qt(1 - (alpha/2), nrow(data)-1)
  error = t*sqrt(s2_Yhat_new)
  lower = round(Yhat - error, 3)
  upper = round(Yhat + error, 3)
  return(c(lower, upper))
}

Yhat_pred_interval_origin(typo_errors, 10, .02)
```

```
## [1] 167.844 192.722
```

Problem 13:

Refer to *Typographical errors* Problem 4.12.

(a) Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

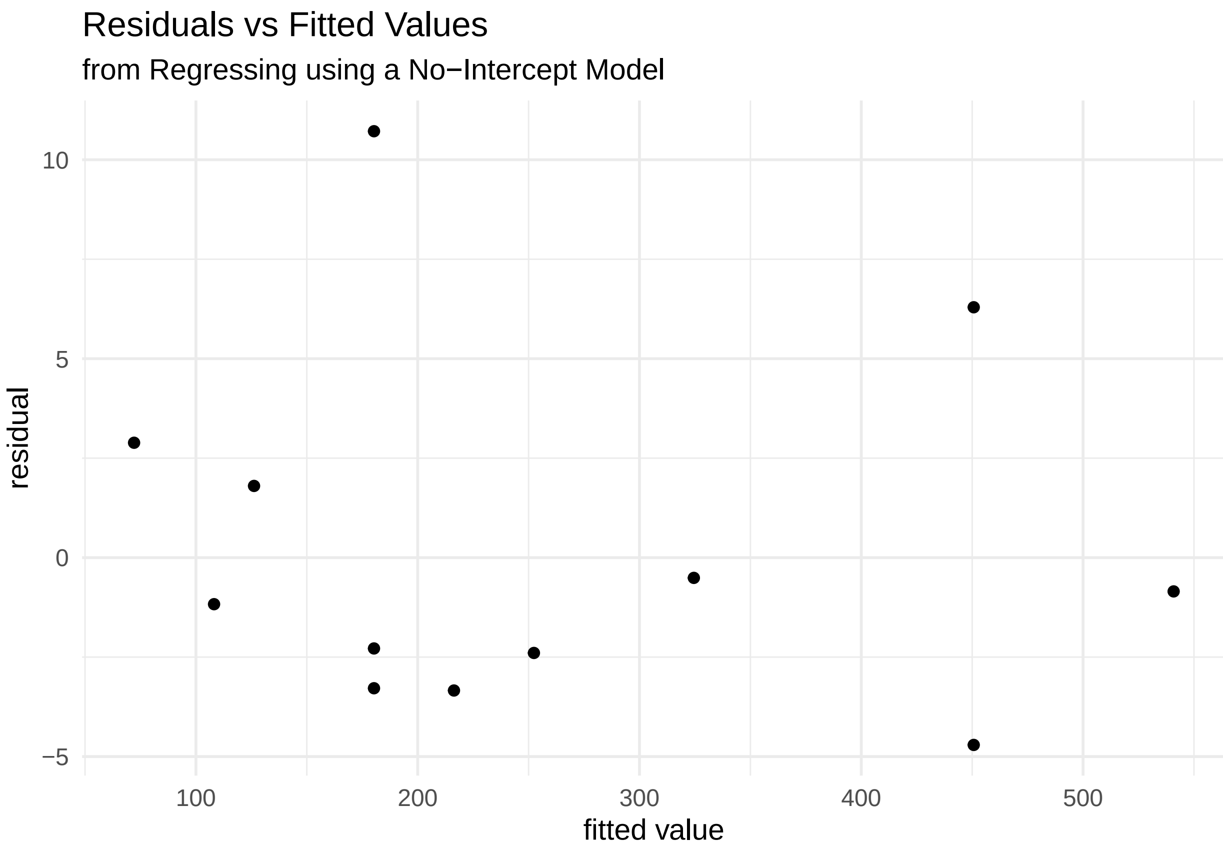
Answer:

```r
model = lm.fit.origin_manual(typo_errors$x, typo_errors$y)
typo_errors_zero_model = typo_errors %>%
  mutate(preds = model*x,
         resids = y - preds)
typo_errors_zero_model$resids %>% sum
```

```
## [1] 3.159876
```

The residuals do not sum to zero.

```r
typo_errors_zero_model %>%
  ggplot(aes(preds, resids)) +
  geom_point() +
  labs(x = "fitted value",
       y = "residual",
       title = "Residuals vs Fitted Values",
       subtitle = "from Regressing using a No-Intercept Model")
```



Using a zero intercept model creates a U-shaped pattern in the fitted values vs residual plot. It may not be the best fit for the data.

(b) Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .01$. State the alternatives, decision rule and conclusion. What the P-value of the test?

Answer:

Problem 14:

Refer to *Grade point average* Problem 1.19. Assume that linear regression through the origin model is appropriate.

(a) Fit regression model and state the estimated regression function.

Answer:

```
gpa = read.csv('CH01PR19.txt', sep = '', header = FALSE,
               col.names = c('y', 'x'),
               colClasses = c('numeric', 'numeric'))
```

(b) Estimate β_1 with a 95 percent confidence interval. Interpret your interval estimate.

Answer:

(c) Estimate the mean freshman GPA for students whose ACT test score is 30. Use a 95 percent confidence interval.

Answer:

Problem 15:

Refer to *Grade point average* Problem 4.14.

(a) Plot the fitted regression line and the data. Does the linear regression function through the origin appear to be a good fit here?

Answer:

(b) Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

Answer:

(c) Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .005$. State the alternatives, decision rule and conclusion. What is the P-value of the test?

Answer:

Problem 16:

Refer to *Copier maintenance* Problem 1.20. Assume that linear regression through the origin mdodel is appropriate.

(a) Obtain the estimated regression function.

Answer:

(b) Estimate β_1 with a 90 percent confidence interval. Interpret your interval estimate.

Answer:

(c) Predict the service time on a new call in which six copiers are to be serviced. Use a 90 percent prediction interval.

Answer:

Problem 17:

Refer to *Copier maintenance* Problem 4.16.

(a) Plot the fitted regression line and the data. Does the linear regression through the origin appear to be a good fit here?

Answer:

(b) Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

Answer:

(c) Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .01$. State the alternatives, decision rule and conclusion. What is the P-value of the test?

Answer:

Problem 18:

Refer to *Plastic hardness* Problem 1.22. Suppose that errors arrise in X because the laboratory technician is instructed to measure the hardness of the ith specimen (Y_i) at a prerecorded elapsed time (X_i), but the timing is imperfect so the true elapsed time varies at random from the prerecorded elapsed time. Will ordinary least squares estimates be biased here? Discuss.

Answer:

Problem 19:

Refer to *Grade point average* Problem 1.19. A new student earned a grade point average of 3.4 in the freshman year.

(a) Obtain a 90 percent confidence interavl for the students' ACT test score. Interpret your confidence interval.

Answer:

(b) Is criterion (4.33) as to the appropriateness of the approximate confidence interval met here?

Answer:

Problem 20:

Refer to *Plastic hardness* Problem 1.22. The measurement of a new test item showed 238 Brinell units of hardness.

(a) Obtain a 99 percent confidence interval for the elapsed time before the hardness was measured. Interpret your confidence interval.

Answer:

(b) Is criterion (4.33) as to the appropriateness of the approximate confidence interval met here?

Answer:

Problem 21:

When the predictor variable is coded that $\overline{X} = 0$ and the normal error regression model applies, are b_0 and b_1 independent? Are the joint confidence intervals for β_0 and β_1 then independent?

Answer:

Problem 22:

Derive an extension of the Bonferroni inequality (4.2a) for the case of three statements, each with statement confidence coefficients $1 - \alpha$.

Answer:

Problem 23:

Show that for the fitted least squares regression line through the origin (4.15), $\sum X_i e_i = 0$.

Answer:

Problem 24:

Show that $\hat{Y}$ as defined in (4.15) for linear regression through the origin is an unbiased estimator of $E[Y]$.

Answer:

Problem 25:

Derive the formula for $s^2[\hat{Y}_h]$ for linear regression through the origin.

Answer:

Problem 26:

Refer to the *CDI* dataset in Appendix C.2 and Project 1.43. Consider the regression relation of number of active physicians to total population.

(a) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

Answer:

```
cdi_cols = c('ID', 'county', 'state', 'area', 'total_pop', 'perc_pop_18to34',
             'perc_pop_65plus', 'num_physicians', 'num_hosp_beds', 'total_crimes',
             'perc_hs_grads', 'perc_bach', 'perc_below_poverty', 'perc_unemploy',
             'per_capita_income', 'total_personal_income', 'geographic_region')
cdi_colclasses = c('integer', 'character', 'character', rep('numeric', 13), 'factor')
cdi = read.csv('APPENC02.txt', sep = '', header = FALSE,
                  col.names = cdi_cols,
                  colClasses = cdi_colclasses)
```

(b) An investigator has suggested that β_0 should be -100 and β_1 should b .0028. Do the joint confidence intervals in part (a) support this view? Discuss.

Answer:

(c) It is desired to estimate the expected number of active physicians for countries with total population of $X = 500,\ 1,000,\ 5,000$ thousands with family confidence coefficient .90. Which procedure, the Working-Hotelling or the Bonferroni, is more efficient here?

Answer:

(d) Obtain the family of interval estimates required in part (c), using the more efficient procedure. Interpret your confidence intervals.

Answer:

Problem 26:

Refer to the *SENIC* dataset in Appendix C.1 and Project 1.45. Consider the regression relation of average length of stay to infection risk.

(a) Obtain Bonferroni join confidence intervals for β_0 and β_1, using a 90 percent family confidence coefficient.

Answer:

```
senic_cols = c('ID', 'length_stay', 'age', 'infection_risk', 'culturing_ratio',
               'chest_xray_ratio', 'num_beds', 'med_school_aff', 'region',
               'avg_daily_census', 'num_nurses', 'available_facilities')
senic_colclasses = c(rep('numeric', 6), rep('factor', 2), rep('numeric', 3))
senic = read.csv('APPENC01.txt', sep = '', header = FALSE,
                  col.names = senic_cols,
                  colClasses = senic_colclasses)
```

(b) A researcher suggested that β_0 should be approximately 7 and β_1 should be approximately 1. Do the joint intervals in part (a) support this expectation? Discuss.

Answer:

(c) It is desired to estimate the expected hospital stay for persons with infection risks $X = 2, 3, 4, 5$ with family confidence coefficient .95. Which procedure, the Working-Hotelling or the Bonferroni, is more efficient here?

Answer:

(d) Obtain the family of interval estimates required in part (c), using the more efficient procedure. Interpret your confidence intervals.

Answer: