ps10.R

benjamin

2020-12-01

```r
## Ben Reichert
## Problem Set 10

# 1 Attached

# 2
# I think the simpler explanantion for decreased performance following praise is that performance varies
# what the pilots are told. A key topic in Chapter 15 is the regression effect, which is essentially th
# data to regress towards the mean in the case of pilots testing very well initially.

# 3
# a) The prediction of 98 wins again is likely too high because the average wins of an MLB team is 81 w
# deviation of 11.7, meaning getting anyhthing over 92-93 wins is extremely unlikely.

# b)
t <- 30
n <- 162
m <- 81
sd <- 11.7
corr <- 0.54
r <- 0.54

m1 <- 98
wins <- m*(1-r) + r*m1
cat("Predicted wins:",wins,"\n")
```

```
## Predicted wins: 90.18
```

```r
# c)
# In every set of data there are some outliers. It is almost expected that there will be some inconsist
# hence the 95% confidence interval mechanism. However, every team is still statistically expected to s
# the 95% of data. Outliers like winning 96 games are stastically very unlikely. Although it happens mos
# to at least one team, the individual probability of that happening to a given team is very low.

# 4.  Trosset chapter 15.7 exercise 8.
# a) I think it would be unfair to replace Jill's Test 2 score with her Test 1 score because Test 2 was
# than Test 1, so while an 80 on Test 1 is just above the average and within the standard deviation, an
# is way above the average and beyond the standard deviation. I would advise the professor gives Jill a
# because she did 1/2 of the standard deviation better than the ave on Test 1, so why not give her the
# Test 2.

# b) I like this suggestion because it is statistically consistent with how Jack would most likely perf
# to the rest of the class. I would advise the professor assigns the score of 85.
```

```r
# 5.  Trosset chapter 15.7 exercise 5, parts (a), (b), and (c).
x <- c(69,64,65,63,65,62,65,64,66,59,62)
y <- c(71,68,66,67,70,71,70,73,72,65,66)
n <- length(x)
sumx <- sum(x)
sumy <- sum(y)
sumxy <- sum(x*y)
sumx2 <- sum(x^2)
sumy2 <- sum(y^2)
# a)
pears <- ( (n*sumxy) - (sumx*sumy) )/sqrt( (n*sumx2 - sumx^2) * (n*sumy2 - sumy*sumy) )
cat("Coefficient of determination:",pears^2,"\n")
```

```
## Coefficient of determination: 0.3114251
```

```r
# b)
print(summary(lm(y~x)))
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5909 -1.2273 -0.9545  1.1136  4.0000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.1818    18.7584   1.662   0.1308
## x             0.5909     0.2929   2.018   0.0744 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.379 on 9 degrees of freedom
## Multiple R-squared:  0.3114, Adjusted R-squared:  0.2349
## F-statistic:  4.07 on 1 and 9 DF,  p-value: 0.07442
```

```r
# The P-value of 0.07442 > 0.05 thus we fail to reject the null hypothesis.
# We conclude that the data does not provide convincing evidence to conclude
# sister's height influences brother's.

# c)
m <- 0.5909
sd <- 0.2929
n <- 11
err <- qnorm(0.95)*sd/sqrt(n)
cat("90% confidence interval: (",m-err,",",m+err,")")
```

```
## 90% confidence interval: ( 0.4456386 , 0.7361614 )
```