

Reproducible Reporting

Edward Bosko

2024-10-07

Reproducible Reporting

```
dat <- read.csv('C:/Users/Eddie/OneDrive/Desktop/BIOSTAT/BIOS 6621/Week 6/lead-iq-01.csv')
outlier <- which(dat$IQ==999)
dat$IQ[outlier] <- 99
```

Make graph

```
boxplot(IQ ~ Smelter, data = dat, main = 'IQ by Distance from Smelter', ylab = 'IQ')
```



Make table

```

mymod <- lm(IQ ~ as.factor(Smelter), data = dat)
summary(mymod)

##
## Call:
## lm(formula = IQ ~ as.factor(Smelter), data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -46.687  -9.193  -0.687   7.313  48.313
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      92.687      1.754  52.847  <2e-16 ***
## as.factor(Smelter)Near  -3.494      2.587  -1.351   0.179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.36 on 122 degrees of freedom
## Multiple R-squared:  0.01473,    Adjusted R-squared:  0.006654
## F-statistic: 1.824 on 1 and 122 DF,  p-value: 0.1793

df <- data.frame(
  estimate = c(92.687, -3.494),
  SE = c(1.754, 2.587),
  teststat <- c(52.847, -1.351),
  pval <- c('<0.0001', '0.179')
)
colnames(df) = c('Estimate', 'SE', 'Test Statistic', 'P-Value')

df %>% kable(caption = "Model Output of mymod")

```

Table 1: Model Output of mymod

Estimate	SE	Test Statistic	P-Value
92.687	1.754	52.847	<0.0001
-3.494	2.587	-1.351	0.179

We clearly see from our boxplot that the outlier is gone. From our simple linear regression model, we see that children who are within 1 mile of the smelter are expected to have an IQ 3.494 points fewer when compared to children who live further than 1 mile from the smelter, although this difference is not statistically significant ($p=0.179$).

```
mean(dat$IQ)
```

```
## [1] 91.08065
```