2.4.1: SAS - Simultaneous Inference and Regression Through Origin

Stat 5100: Dr. Bean

Example: Toluca dataset

```
# Input data
library(stat5100)
data(toluca)

# Create linear model and obtain 95% confidence interval of beta parameters,
# but this time use the simultaneous comparison adjustment
toluca_lm <- lm(workhours ~ lotsize, data = toluca)
stat5100::coefficient_confidence_lm(toluca_lm, confidence = 0.95, simul = TRUE)

## lower.est and upper.est are the 97.5% confidence limits.
## The Bonferroni adjustment for simultaneous confidence levels was made.
##              Estimate Std. Error   t value     Pr(>|t|)  lower.est  upper.est
## (Intercept) 62.365859 26.1774339  2.382428 2.585094e-02 -0.4043574 125.136075
## lotsize      3.570202  0.3469722 10.289592 4.448828e-10  2.7382061   4.402198
```

Simultaneous 90% interval estimation of mean workhours (Bonferroni and Working-Hotelling)

```
toluca_sites_A <- data.frame(lotsize = c(30, 65, 100))
stat5100::simul_mean_prediction_limits(toluca_lm, toluca_sites_A, confidence = 0.90)

##   lotsize     yhat   se_yhat WH_lower WH_upper  B_lower  B_upper
## 1      30 169.4719 16.969741 131.1542 207.7897 131.0570 207.8868
## 2      65 294.4290  9.917579 272.0351 316.8229 271.9783 316.8797
## 3     100 419.3861 14.272328 387.1591 451.6130 387.0774 451.6947
```

Simultaneous 95% prediction limits for two lots (Bonferroni and Scheffe)

```
toluca_sites_B <- data.frame(lotsize = c(80, 100))
stat5100::simul_prediction_limits(toluca_lm, toluca_sites_B, confidence = 0.95)

##   lotsize     yhat se_yhat_pred  S_lower  S_upper  B_lower  B_upper
## 1      80 347.9820     49.91095 217.4073 478.5568 228.3018 467.6622
## 2     100 419.3861     50.86664 286.3111 552.4610 297.4142 541.3579
```

Regression through origin example

This dataset is called "warehouse": a plumbing supplies company that is looking at the relation between work units (X) and labor costs (Y) at its 12 warehouses.

```r
data(warehouse)
head(warehouse)
```

```
##   work cost
## 1   20  114
## 2  196  921
## 3  115  560
## 4   50  245
## 5  122  575
## 6  100  475
```

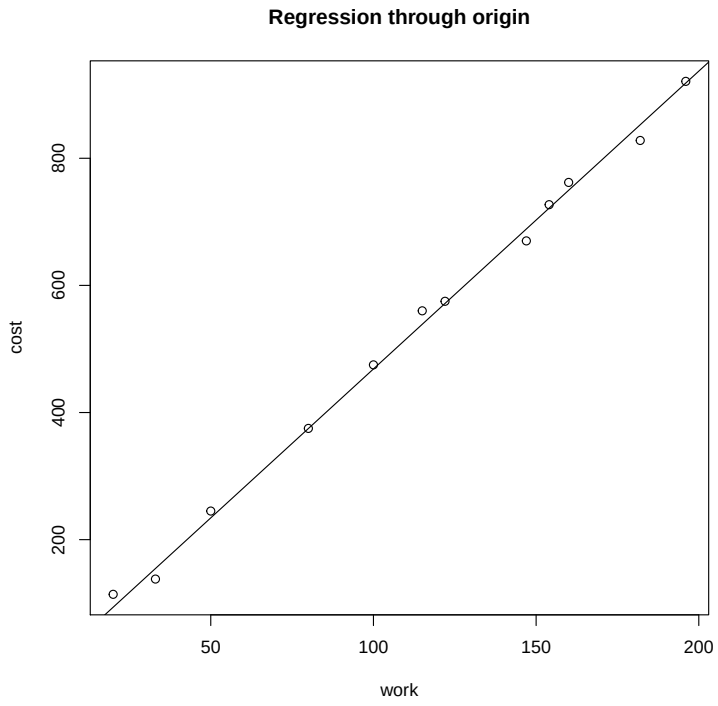
```r
# Fit a linear model through the origin
warehouse_lm_origin <- lm(cost ~ 0 + work, data = warehouse)

# Look at some statistics
anova(warehouse_lm_origin)
```

```
## Analysis of Variance Table
##
## Response: cost
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## work       1 4191980 4191980   18762 < 2.2e-16 ***
## Residuals 11    2458     223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(warehouse_lm_origin)
```

```
##
## Call:
## lm(formula = cost ~ 0 + work, data = warehouse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.720  -4.020   4.432  11.141  21.194
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## work  4.68527    0.03421     137   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.95 on 11 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9994,Adjusted R-squared:  0.9994
## F-statistic: 1.876e+04 on 1 and 11 DF,  p-value: < 2.2e-16
```

```r
# What does the fit look like?
# (R doesn't plot the origin inside of the graph, but only because the data
# doesn't stretch to the point (0,0). Understand that this plot does have the
# regression line going through the origin even though it is outside the graph)
stat5100::fit_plot(warehouse_lm_origin, xlab = "work", ylab = "cost",
                   main = "Regression through origin")
```


Regression through origin
cost
work
200
400
600
800
50
100
150
200