Statistical Adjustment
Elementary Statistics & Inference (STAT:1020; Bognar)

Lets look at a little example regarding statistical adjustment.

We have salary data on white and minority employees at a large company. The years of experience x and salary y (in thousands) of 7 minority employees and 6 white employees is

Minority	$x =$ years:	3	3	5	6	8	8	10
	$y =$ salary:	17	19	20	21	22	24	25

White	$x =$ years:	1	2	3	3	4	5
	$y =$ salary:	18.0	19.3	21.6	19.6	21.9	23.2

The summary statistics are

$$\textit{Minority} \quad \begin{array}{ll} \bar{x} = 6.143 & s_x = 2.672 \\ \bar{y} = 21.143 & s_y = 2.795 \end{array} \quad r = 0.956$$

$$\textit{White} \quad \begin{array}{ll} \bar{x} = 3.0 & s_x = 1.414 \\ \bar{y} = 20.6 & s_y = 1.944 \end{array} \quad r = 0.946$$

Based upon the mean salaries (i.e. the $\bar{y}$'s), minority employees make more than white employees. However, minority employees have more years of experience! This makes a comparison of these "un-adjusted" mean salaries unfair. We want to adjust/account for years of experience before making salary comparisons.

1. Determine the least squares regression line for each group.

$$\begin{array}{rrcl} \textit{Minority}: & \hat{\beta}_1 & = & r\frac{s_y}{s_x} = 0.956\frac{2.795}{2.672} = 1.0 \\ & \hat{\beta}_0 & = & \bar{y} - \hat{\beta}_1\bar{x} = 21.143 - 1.0(6.143) = 15.0 \\ & \hat{y} & = & 15.0 + 1.0x \end{array}$$

$$\begin{array}{rrcl} \textit{White}: & \hat{\beta}_1 & = & r\frac{s_y}{s_x} = 0.946\frac{1.944}{1.414} = 1.3 \\ & \hat{\beta}_0 & = & \bar{y} - \hat{\beta}_1\bar{x} = 20.6 - 1.3(3.0) = 16.7 \\ & \hat{y} & = & 16.7 + 1.3x \end{array}$$

   The R output for the two regression analyses is at the end of this document. Scatterplots and the least squares regression lines for each group is shown in Figure 1.

2. Do white or minority employees have a higher mean (average) starting (i.e. no years of experience) salary?

$$\begin{array}{rcl} \textit{Minority} & \rightarrow & \hat{y} = 15.0 + 1.0(0) = 15.0 \\ \textit{White} & \rightarrow & \hat{y} = 16.7 + 1.3(0) = 16.7 \end{array}$$

   The mean starting pay of white nurses is approximately $1700 more than minority nurses.

3. Do white or minority employees get pay increases at a faster rate?

$$\begin{array}{rcl} \textit{Minority} & \rightarrow & \hat{\beta}_1 = 1.0 \\ \textit{White} & \rightarrow & \hat{\beta}_1 = 1.3 \end{array}$$

   White nurses get pay raises at a faster rate. On average, minority nurses get approximately $1000 more for every extra year worked, while white nurses get $1300.

4. After 5 years, do white or minority employees have a higher mean salary?

$$\begin{array}{rcl} \textit{Minority} & \rightarrow & \hat{y} = 15.0 + 1.0(5) = 20.0 \\ \textit{White} & \rightarrow & \hat{y} = 16.7 + 1.3(5) = 23.2 \end{array}$$

   After 5 years white nurses are making approximately $3200 more than minority nurses, on average.

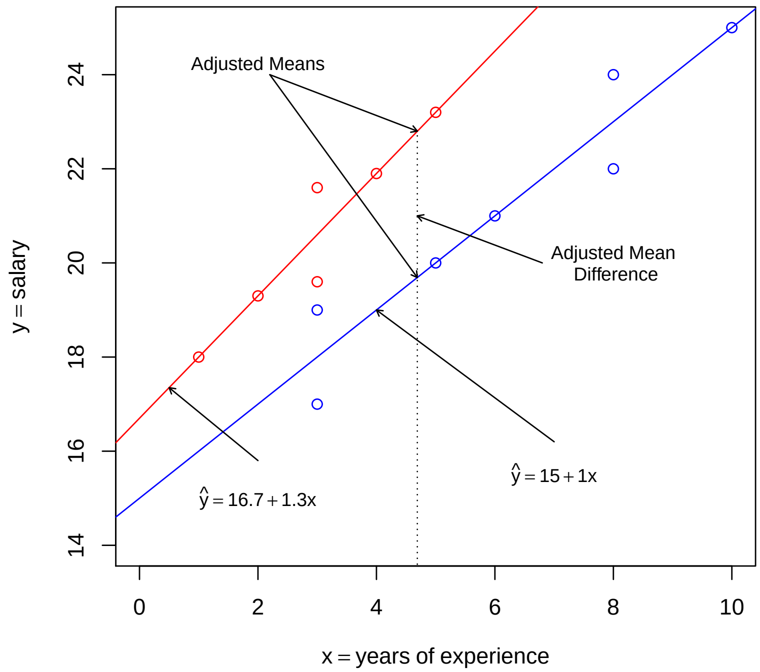

Figure 1: Nursing Salaries: Scatterplot and least squares regression line for the Minority nurses (in blue) and White nurses (in red).

5. Determine the *adjusted mean salaries.*

    Overall, the average amount of experience is

$$\frac{(3+3+5+6+8+8+10)+(1+2+3+3+4+5)}{13} = 4.69$$

    Therefore, the adjusted mean salaries are

$$\begin{aligned} \textit{Minority} &\rightarrow \hat{y} = 15.0 + 1.0(4.69) = 19.69 \\ \textit{White} &\rightarrow \hat{y} = 16.7 + 1.3(4.69) = 22.80 \end{aligned}$$

    See Figure 1.

6. What is the *adjusted mean difference*?

    The adjusted mean difference is the difference between the adjusted means: $22.80 - 19.69 = 3.11$. *Hence, the adjusted mean difference is \$3110. See Figure 1.*

7. In summary, after adjusting for years of experience, does there appear to be salary discrimination?

    Yes. After adjusting for years of experience, white nurses are making approximately \$3110 more than minority nurses, on average.

To learn how to assess statistical significance in this example, take another class.

```
\textcolor{blue}{Minority Employees}
\textcolor{blue}{==================}
> x <- c(3,3,5,6,8,8,10)
> y <- c(17,19,20,21,22,24,25)
> m.results <- lm(y~x)
> summary(m.results)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.0000     0.9048  16.578 1.46e-05
x             1.0000     0.1366   7.319 0.000746

Residual standard error: 0.8944 on 5 degrees of freedom
Multiple R-squared:  0.9146,Adjusted R-squared:  0.8976
F-statistic: 53.57 on 1 and 5 DF,  p-value: 0.0007461

> anova(m.results)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 42.857  42.857  53.571 0.0007461
Residuals  5  4.000   0.800

\textcolor{blue}{White Employees}
\textcolor{blue}{===============}
> x <- c(1,2,3,3,4,5)
> y <- c(18.0,19.3,21.6,19.6,21.9,23.2)
> w.results <- lm(y~x)
> summary(w.results)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.7000     0.7303  22.867 2.17e-05
x             1.3000     0.2236   5.814  0.00436

Residual standard error: 0.7071 on 4 degrees of freedom
Multiple R-squared:  0.8942,Adjusted R-squared:  0.8677
F-statistic:  33.8 on 1 and 4 DF,  p-value: 0.004357

> anova(w.results)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x          1   16.9    16.9    33.8 0.004357
Residuals  4    2.0     0.5
```