**(1)** Save the data to a file (excel or CSV file) and read it into R memory for analysis. (Q1 - 2 points)

**File read and saved with R commands:**

**setwd("/Users/anthony.valencia/met/cs555/assignments/03/")**

**fish <- read.csv("data.csv")**

**fish <- rename(fish, c("Number.of.meals.with.fish"="meals.w.fish", "Total.Mercury.in.mg.g"="total.hg"))**

**(2)** To get a sense of the data, generate a scatterplot (using an appropriate window, label the axes, and title the graph). Consciously decide which variable should be on the x-axis and which should be on the y-axis. Using the scatterplot, describe the form, direction, and strength of the association between the variables. (Q2 - 3 points)

**Produce plot with command:**

**plot(**

**fish$meals.w.fish,**

**fish$total.hg,**

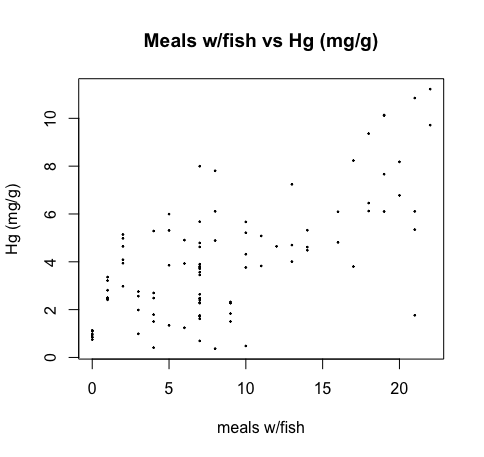
**main="Meals w/fish vs Hg (mg/g)",**

**xlab="meals w/fish",**

**ylab="Hg (mg/g)",**

**cex=0.2**

**)**



***Explanatory variable (x)* is the number of meals with fish, since we hypothesize this will explain the response**

***Response variable (y)* is total mercury (mg/g) which we assume is a response to the aforementioned explanatory variable**

***Form*: seems to trend upward. Almost linear, yet the seemingly scattered enough to almost seem random.**

***Direction*: the points appear to be positively associated. As the number of meals with fish increase the mercury content increases. There is certainly almost no mercury content when no fish is consumed, while there is the greatest mercury content when the most fish is consumed.**

***Strength of relationship*: the points show a loose association particularly between 1 and 10 meals with fish. The relationship is stronger between 10 and 15 meals with fish, but again is loosely related between 15 and 20 meals with fish.**

**(3)** Calculate the correlation coefficient. What does the correlation tell us? (Q3 - 2 points)

**The correlation coefficient 0.6991094 is calculated using the command:**

**cor(**

**fish$meals.w.fish,**

**fish$total.hg**

**)**

**If we hypothesize that**

**ρ=0 (H0: there is no linear association)**

**ρ≠0 (H1: there is a linear association).**

**Further analysis from command:**

**cor.test(**

**fish$meals.w.fish,**

**fish$total.hg,**

**alternative='two.sided'**

**)**

**Tells us for t = 9.6793, df = 98, p-value = 6.013e-16**

**alternative hypothesis: ρ≠0, so there IS a linear association**

**for a 95 percent confidence interval between 0.5827074 and 0.7874031**

**(4)** Find the equation of the least squares regression equation, and write out the equation. Add the regression line to the scatterplot you generated above. (Q4 - 4 points)

**is the equation for the simple linear regression line.**

**Where, is the expected value of y for a given value of x.**

**x is the explanatory variable.**

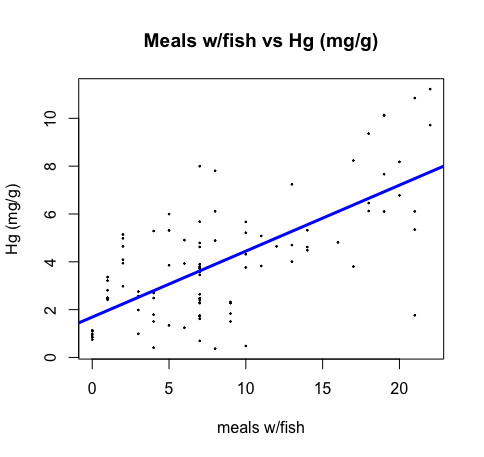
**is the least-squares estimates of the slope.**

**is the least-squares estimates of the intercept.**

**The regression line is added to the scatterplot from above via the commands:**

**lm.result <- lm(fish$total.hg~fish$meals.w.fish)**

**abline(lm.result, lty=1, lwd=3, col="blue")**



**(5)** What is the estimate for  beta\_1 ? How can we interpret this value? What is the estimate for beta\_0 ? What is the interpretation of this value? (Q5 - 4 points)

**The estimate for is 0.2759503 calculated from the commands:**

**x.bar <- mean(fish$meals.w.fish)**

**sx <- sd(fish$meals.w.fish)**

**y.bar <- mean(fish$total.hg)**

**sy <- sd(fish$total.hg)**

**r <- cor(**

**fish$meals.w.fish,**

**fish$total.hg**

**)**

**beta\_1 <- r\*sy/sx**

**Interpretation: this value is the slope of the regression line. This implies that the mercury content (mg/g) rises approximately 1 unit for every 4 meals with fish.**

**The estimate for is 1.687643 calculated from the command:**

**beta\_0 <- y.bar - beta1\*x.bar**

**Interpretation: this value is the y intercept of the regression line. This implies that the individual starts with about 1.69 mg/g of mercury content before even consuming any meals with fish.**

**(6)** Calculate the ANOVA table and the table which gives the standard error of  (hat beta 1) . Formally test the hypothesis that beta\_1 = 0 using either the F-test or the t-test at the alpha level a=0.10. Either way, present your results **using the 5-step procedure as in the course notes**. Within your conclusion, calculate the R2 (R squared) value and interpret this.

Also, calculate and interpret the 90% confidence interval for beta\_1 . (Q6 - 5 points)

**The ANOVA table:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SS | df | MS |
| Regression | 309.23933 | 1 | 309.2393349 |
| Residual | 323.47025 | 98 | 3.300716833 |
| Total | 632.70958 |  |  |

**Is calculated from the large table at the end.**

**R2 = RegSS/TotalSS = 309.23933/632.70958 = 0.488753991**

**48.9% of the variability in the mercury content (mg/g) can be explained by the number of meals with fish.**

**f-test:**

1. **Set up the hypotheses and select the alpha level**

**H0: (there is no linear association)**

**H1: (there is a linear association)**

**a=0.10**

1. **Select the appropriate test statistic**

**With 1 and n-2 = 98 degrees of freedom**

1. **State the decision rule**

**F-distribution with 1, 98 degrees of freedom and association with a = 0.10.**

**qf(0.9,df1=1, df2=98)**

**F1,98,0.10 = 2.75743**

**Decision Rule: Reject H0 if**

**Otherwise, do not reject H0**

1. **Compute the test statistic using the command:**

**anova(lm.result)**

***Analysis of Variance Table***

***Response: fish$total.hg***

***Df Sum Sq Mean Sq F value Pr(>F)***

***fish$meals.w.fish 1 309.24 309.239 93.689 6.013e-16 \*\*\****

***Residuals 98 323.47 3.301***

***---***

***Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1***

**So, F is 93.689. Note that all of these values match those calculated by hand above.**

1. Conclusion
2. Since **93.689 = F and , we reject H0. We have significant evidence at a=0.10 level that . There is evidence of a significant linear association between mercury content and meals with fish (here, p < 0.001 as calculated).**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| meals with fish | Hg in mg/g |  |  |  |  |  |
| x | y | y\_hat | y\_hat - y\_mean | (y\_hat - y\_mean)^2 | y - y\_hat | (y - y\_hat)^2 |
| 14 | 4.484 | 5.551 | 1.573 | 2.474 | -1.067 | 1.138 |
| 7 | 4.789 | 3.619 | -0.359 | 0.129 | 1.170 | 1.368 |
| 5 | 3.856 | 3.067 | -0.911 | 0.829 | 0.789 | 0.622 |
| 8 | 4.888 | 3.895 | -0.083 | 0.007 | 0.993 | 0.986 |
| 21 | 10.849 | 7.483 | 3.505 | 12.282 | 3.366 | 11.333 |
| 18 | 6.457 | 6.655 | 2.677 | 7.165 | -0.198 | 0.039 |
| 22 | 11.222 | 7.759 | 3.781 | 14.292 | 3.463 | 11.995 |
| 6 | 4.908 | 3.343 | -0.635 | 0.403 | 1.565 | 2.448 |
| 19 | 10.116 | 6.931 | 2.953 | 8.718 | 3.185 | 10.146 |
| 7 | 3.567 | 3.619 | -0.359 | 0.129 | -0.052 | 0.003 |
| 16 | 6.092 | 6.103 | 2.125 | 4.515 | -0.011 | 0.000 |
| 17 | 3.799 | 6.379 | 2.401 | 5.764 | -2.580 | 6.655 |
| 20 | 6.781 | 7.207 | 3.229 | 10.424 | -0.426 | 0.181 |
| 5 | 5.995 | 3.067 | -0.911 | 0.829 | 2.928 | 8.571 |
| 7 | 1.717 | 3.619 | -0.359 | 0.129 | -1.902 | 3.619 |
| 14 | 4.615 | 5.551 | 1.573 | 2.474 | -0.936 | 0.876 |
| 1 | 3.362 | 1.964 | -2.014 | 4.058 | 1.398 | 1.956 |
| 6 | 3.928 | 3.343 | -0.635 | 0.403 | 0.585 | 0.342 |
| 9 | 1.833 | 4.171 | 0.193 | 0.037 | -2.338 | 5.467 |
| 10 | 5.668 | 4.447 | 0.469 | 0.220 | 1.221 | 1.490 |
| 13 | 4.700 | 5.275 | 1.297 | 1.682 | -0.575 | 0.331 |
| 9 | 2.272 | 4.171 | 0.193 | 0.037 | -1.899 | 3.607 |
| 16 | 4.812 | 6.103 | 2.125 | 4.515 | -1.291 | 1.666 |
| 5 | 1.342 | 3.067 | -0.911 | 0.829 | -1.725 | 2.977 |
| 18 | 6.123 | 6.655 | 2.677 | 7.165 | -0.532 | 0.283 |
| 7 | 4.622 | 3.619 | -0.359 | 0.129 | 1.003 | 1.005 |
| 8 | 7.805 | 3.895 | -0.083 | 0.007 | 3.910 | 15.286 |
| 7 | 2.643 | 3.619 | -0.359 | 0.129 | -0.976 | 0.953 |
| 8 | 6.111 | 3.895 | -0.083 | 0.007 | 2.216 | 4.910 |
| 7 | 2.476 | 3.619 | -0.359 | 0.129 | -1.143 | 1.307 |
| 10 | 4.317 | 4.447 | 0.469 | 0.220 | -0.130 | 0.017 |
| 4 | 1.789 | 2.791 | -1.187 | 1.408 | -1.002 | 1.005 |
| 4 | 2.484 | 2.791 | -1.187 | 1.408 | -0.307 | 0.095 |
| 7 | 1.757 | 3.619 | -0.359 | 0.129 | -1.862 | 3.468 |
| 6 | 1.239 | 3.343 | -0.635 | 0.403 | -2.104 | 4.428 |
| 5 | 5.311 | 3.067 | -0.911 | 0.829 | 2.244 | 5.034 |
| 19 | 6.103 | 6.931 | 2.953 | 8.718 | -0.828 | 0.685 |
| 3 | 1.984 | 2.515 | -1.463 | 2.139 | -0.531 | 0.282 |
| 4 | 2.697 | 2.791 | -1.187 | 1.408 | -0.094 | 0.009 |
| 7 | 0.692 | 3.619 | -0.359 | 0.129 | -2.927 | 8.569 |
| 7 | 2.404 | 3.619 | -0.359 | 0.129 | -1.215 | 1.477 |
| 9 | 1.503 | 4.171 | 0.193 | 0.037 | -2.668 | 7.119 |
| 17 | 8.231 | 6.379 | 2.401 | 5.764 | 1.852 | 3.431 |
| 14 | 5.321 | 5.551 | 1.573 | 2.474 | -0.230 | 0.053 |
| 7 | 3.810 | 3.619 | -0.359 | 0.129 | 0.191 | 0.036 |
| 21 | 1.765 | 7.483 | 3.505 | 12.282 | -5.718 | 32.691 |
| 4 | 0.408 | 2.791 | -1.187 | 1.408 | -2.383 | 5.681 |
| 7 | 3.901 | 3.619 | -0.359 | 0.129 | 0.282 | 0.079 |
| 10 | 0.480 | 4.447 | 0.469 | 0.220 | -3.967 | 15.738 |
| 11 | 3.826 | 4.723 | 0.745 | 0.555 | -0.897 | 0.805 |
| 7 | 3.451 | 3.619 | -0.359 | 0.129 | -0.168 | 0.028 |
| 9 | 2.320 | 4.171 | 0.193 | 0.037 | -1.851 | 3.427 |
| 2 | 4.086 | 2.240 | -1.738 | 3.022 | 1.846 | 3.409 |
| 7 | 2.272 | 3.619 | -0.359 | 0.129 | -1.347 | 1.815 |
| 3 | 2.564 | 2.515 | -1.463 | 2.139 | 0.049 | 0.002 |
| 7 | 7.998 | 3.619 | -0.359 | 0.129 | 4.379 | 19.173 |
| 11 | 5.081 | 4.723 | 0.745 | 0.555 | 0.358 | 0.128 |
| 8 | 0.366 | 3.895 | -0.083 | 0.007 | -3.529 | 12.456 |
| 7 | 2.477 | 3.619 | -0.359 | 0.129 | -1.142 | 1.305 |
| 4 | 5.288 | 2.791 | -1.187 | 1.408 | 2.497 | 6.233 |
| 7 | 5.676 | 3.619 | -0.359 | 0.129 | 2.057 | 4.230 |
| 7 | 2.296 | 3.619 | -0.359 | 0.129 | -1.323 | 1.751 |
| 21 | 6.110 | 7.483 | 3.505 | 12.282 | -1.373 | 1.884 |
| 4 | 1.502 | 2.791 | -1.187 | 1.408 | -1.289 | 1.663 |
| 7 | 3.710 | 3.619 | -0.359 | 0.129 | 0.091 | 0.008 |
| 3 | 2.752 | 2.515 | -1.463 | 2.139 | 0.237 | 0.056 |
| 3 | 0.987 | 2.515 | -1.463 | 2.139 | -1.528 | 2.336 |
| 19 | 10.140 | 6.931 | 2.953 | 8.718 | 3.209 | 10.300 |
| 7 | 1.616 | 3.619 | -0.359 | 0.129 | -2.003 | 4.013 |
| 12 | 4.650 | 4.999 | 1.021 | 1.042 | -0.349 | 0.122 |
| 13 | 7.241 | 5.275 | 1.297 | 1.682 | 1.966 | 3.865 |
| 18 | 9.360 | 6.655 | 2.677 | 7.165 | 2.705 | 7.318 |
| 7 | 3.753 | 3.619 | -0.359 | 0.129 | 0.134 | 0.018 |
| 13 | 4.008 | 5.275 | 1.297 | 1.682 | -1.267 | 1.605 |
| 21 | 5.345 | 7.483 | 3.505 | 12.282 | -2.138 | 4.569 |
| 1 | 2.455 | 1.964 | -2.014 | 4.058 | 0.491 | 0.241 |
| 0 | 0.941 | 1.688 | -2.290 | 5.246 | -0.747 | 0.557 |
| 1 | 2.478 | 1.964 | -2.014 | 4.058 | 0.514 | 0.265 |
| 1 | 3.212 | 1.964 | -2.014 | 4.058 | 1.248 | 1.559 |
| 10 | 5.214 | 4.447 | 0.469 | 0.220 | 0.767 | 0.588 |
| 0 | 1.120 | 1.688 | -2.290 | 5.246 | -0.568 | 0.322 |
| 0 | 0.745 | 1.688 | -2.290 | 5.246 | -0.943 | 0.889 |
| 2 | 4.645 | 2.240 | -1.738 | 3.022 | 2.405 | 5.786 |
| 2 | 4.981 | 2.240 | -1.738 | 3.022 | 2.741 | 7.516 |
| 1 | 2.812 | 1.964 | -2.014 | 4.058 | 0.848 | 0.720 |
| 0 | 0.846 | 1.688 | -2.290 | 5.246 | -0.842 | 0.708 |
| 2 | 5.142 | 2.240 | -1.738 | 3.022 | 2.902 | 8.424 |
| 0 | 1.111 | 1.688 | -2.290 | 5.246 | -0.577 | 0.333 |
| 0 | 1.094 | 1.688 | -2.290 | 5.246 | -0.594 | 0.352 |
| 2 | 2.978 | 2.240 | -1.738 | 3.022 | 0.738 | 0.545 |
| 2 | 3.942 | 2.240 | -1.738 | 3.022 | 1.702 | 2.898 |
| 0 | 1.131 | 1.688 | -2.290 | 5.246 | -0.557 | 0.310 |
| 0 | 0.979 | 1.688 | -2.290 | 5.246 | -0.709 | 0.502 |
| 0 | 0.844 | 1.688 | -2.290 | 5.246 | -0.844 | 0.712 |
| 1 | 2.411 | 1.964 | -2.014 | 4.058 | 0.447 | 0.200 |
| 1 | 2.497 | 1.964 | -2.014 | 4.058 | 0.533 | 0.285 |
| 10 | 3.764 | 4.447 | 0.469 | 0.220 | -0.683 | 0.467 |
| 20 | 8.178 | 7.207 | 3.229 | 10.424 | 0.971 | 0.944 |
| 19 | 7.664 | 6.931 | 2.953 | 8.718 | 0.733 | 0.538 |
| 22 | 9.716 | 7.759 | 3.781 | 14.292 | 1.957 | 3.832 |
|  |  |  |  | 309.239 |  | 323.470 |