MA684 HW3

Jiayuan Shi

1A. r=0.60

This high school GPA is a good predictor of freshman GPA. Because the correlation coefficient is 0.6, freshman GPA and high school GPA have a moderate positive association, and freshman GPA expected to be high as high school GPA is high.

1B.

We can reject H0, and conclude that there is significant evidence of an association between high school and freshman year GPA.

1C. > r.con(0.60,135,p=.95,twotailed=TRUE)

[1] 0.4797 0.6982

The 95% confidence interval for this correlation coefficient is (0.4797, 0.6982).

2A.

> cor.test(AGE[SEXMALE==1],SYSBP[SEXMALE==1])

Pearson's product-moment correlation

data: AGE[SEXMALE == 1] and SYSBP[SEXMALE == 1]

t = 2, df = 220, p-value = 0.05

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.0008948 0.2608282

sample estimates:

cor

0.1332

> cor.test(AGE[SEXMALE==0],SYSBP[SEXMALE==0])

Pearson's product-moment correlation

data: AGE[SEXMALE == 0] and SYSBP[SEXMALE == 0]

t = 9.7, df = 280, p-value <2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.4101 0.5858

sample estimates:

cor

0.5031

For males, the correlation coefficient between age and systolic blood pressure is 0.1332, and so age and systolic blood pressure have a positive, very weak association. This means males’ systolic blood pressure expected to increase a little as age increases.

For females, the correlation coefficient between age and systolic blood pressure is 0.5031, and so age and systolic blood pressure have a positive, moderate association. This means females’ systolic blood pressure expected to increase moderately as age increases.

Because correlation coefficient for males is smaller than that for females, we can say that, the association between age and systolic blood pressure is weaker for males than that for females.

2B. r1=0.1332, r2=0.5031

> fisherz(0.1332)

[1] 0.134

> fisherz(0.5031)

[1] 0.5534

> table(SEXMALE)

SEXMALE

0 1

280 220

H0 ρ1 = ρ2 ,

p-value<0.001

We can reject the H0, and say that there is significant evidence of the association between age and systolic blood pressure is different in males from in females.

3A.



From the above formula, we can get the results,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SS | d.f. | MS | F |
| Model  Error | 13750  87090 | 1  198 | 13750  439.8 | 31.26 |
| Total | 100840 | 199 | 506.7 |  |

p-value=7.418e-08<0.001

We can reject H0(r2 pop=0), and conclude that there is significant evidence of an association between weight and blood pressure in a sample of n=200 adults.

3B. R2 = SSModel / SSTotal = 13750/100840 = 0.1364

The proportion of variation in systolic blood pressure explained by weight is 13.64 %.

4A.

> cor.test(educyears,qol)

Pearson's product-moment correlation

data: educyears and qol

t = 2.7, df = 98, p-value = 0.009

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.06612 0.43352

sample estimates:

cor

0.2592

> cor.test(income,qol)

Pearson's product-moment correlation

data: income and qol

t = 1.1, df = 98, p-value = 0.3

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.08945 0.29891

sample estimates:

cor

0.1089

> cor.test(genhealth ,qol)

Pearson's product-moment correlation

data: genhealth and qol

t = -3.9, df = 98, p-value = 2e-04

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.5277 -0.1867

sample estimates:

cor

-0.3696

The two variables education and general health are significantly associated with quality of life, because the p-values for the hypothesis tests(H0: true correlation is equal to 0) is 0.009 and 2e-04, respectively. These p-values are smaller than 0.01, so we can reject H0, and conclude that there is significant evidence of an association between education and quality of life, and general health and quality of life.

r(education)=0.2592, r(income)=0.1089, r(general health)=-0.3696

The association between education and quality of life is a weak positive association, which means, quality of life expected to increase a little as education increases. The association between general health and quality of life is a moderate negative association, which means, quality of life expected to decrease moderately as general health increases. The association between income and quality of life is a very weak positive association, which means, quality of life expected to increase very little as general income.

General health is the most strongly associated with quality of life.

4B.

> reg <- lm(qol ~ educyears)

> summary(reg)

Call:

lm(formula = qol ~ educyears)

Residuals:

Min 1Q Median 3Q Max

-207.19 -62.44 -7.15 59.06 182.81

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 110.90 43.16 2.57 0.0117 \*

educyears 7.77 2.92 2.66 0.0092 \*\*

---

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

Residual standard error: 82.2 on 98 degrees of freedom

Multiple R-squared: 0.0672, Adjusted R-squared: 0.0577

F-statistic: 7.06 on 1 and 98 DF, p-value: 0.00922

> summary.aov(reg)

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

educyears 1 47731 47731 7.06 0.0092 \*\*

Residuals 98 662864 6764

---

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SS | d.f. | MS | F |
| Model  Error | 47731  662864 | 1  98 | 47731  6764 | 7.06 |
| Total | 710595 | 99 | 7178 |  |

F= MSReg/MSE= 47731/6764=7.06, p-value=0.0092 <0.05

We can reject H0(r2 pop=0), and conclude that there is significant evidence of an association between years of education and quality of life.

R2 = SSModel / SSTotal = 47731/710595= 0.0672

Years of education explains 6.717% of the variability in quality of life.

4C.

- verify that the R-square for the regression is calculated from the Sums of Squares from the ANOVA table

R2 = 0.0672 for the regression, which is equal to SSModel / SSTotal = 47731/710595= 0.0672, so years of education explains only 6.72% of the variability in quality of life. The association between education and quality of life is very weak.

- verify that the R-square for the regression is the square of the correlation coefficient

- verify that the F-statistic from the ANOVA for the regression equals the square of the t-statistic for the slope from the regression, and that the p-values are equal

- verify that the t-statistic for the correlation coefficient matches the t-statistic for the slope

- verify that the Mean Square Error from the ANOVA table relates to s(y|x) – that the square root of the Mean Square Error is equal to s(y|x).