Solutions for the Questions Related to Session #5
Regression and Correlation

Study

Forty four males and 44 females were randomly assigned to treatmill workouts which lasted from 306 to 976 seconds. VO_2 Max (maximum O_2 consumption normalized by body weight (ml/kg/min)) was the outcome measure.

Regression Model 1

The following common slope multiple linear regression model was estimated by least squares.

$E(VO_2\ Max_i | X) = \beta_o + \beta_1(\text{exercise duration}_i) + \beta_2(z_{2,i})$

where $z_{2,i}$ = 1 if the *ith* participant was male, and 0 if *ith* participant was female.

Regression Analysis Summary

Table 1. The regression ANOVA table from the multiple regression analysis.

Parameter	DF	SS	MS	$F_{observed}$
Regression	2	6044.03	3022.02	190.06
Error	84	1335.60	15.90	
Total	86	7379.62		

Table 2. The regression parameter estimates.

Parameter	Estimate b	SE b	$t_{observed}$
Intercept	-1.360	2.220	
Duration	0.059	0.004	14.75
Gender=male	3.396	1.016	3.34

1) Utilize the information from Table 1, to compute the mean square regression (MSR), the mean square error (MSE) and the F-statistic ($F_{observed}$).

(see Table 1.)

2) For a two-sided test with significance level α=0.05 we reject Ho: $\beta_1=\beta_2=0$, if $F_{observed} \geq F_{(2,84,.95)} = 3.105$. Do we reject?.

$F_{observed}=190.06 \geq F_{(2,84,.95)} = 3.105$ so we should reject Ho: $\beta_1=\beta_2=0$.

2) Utilizing the information from Table 1, compute the value of the coefficient of determination (R^2) and give a simple interpretation for the value of R^2 that you calculated.

R^2 = SSR/SST = 6044.3/7379.62 = 0.82.

The R^2 value of 0.82 suggests that 82.0% of the total variation in VO_2 Max was explained by the independent variables; gender, and treadmill exercise duration (s).

4) Utilizing the information from Table 2, compute the *t*-statistics ($t_{observed}$) for the regression parameter related to exercise duration and for the regression parameter related to gender=male.

(see Table 2.)

5) For a two-sided test with significance level 0.05 we reject Ho: $\beta_1=0$, and Ho: $\beta_2=0$ if $|t_{observed}| \geq t_{(84,.975)}=1.990$. Do we reject Ho: $\beta_1=0$?. Do we reject Ho: $\beta_2=0$?.

For the null hypothesis Ho: $\beta_1=0$, $|t_{observed}| = 14.75 \geq t_{(84,.975)}=1.990$, so we should reject Ho: $\beta_1=0$

For the null hypothesis Ho: $\beta_2=0$, $|t_{observed}| = 3.75 \geq t_{(84,.975)}=1.990$, so we should reject Ho: $\beta_2=0$

6) For females and males, write out the estimated regression equation for predicting VO_2 Max as a linear function of exercise duration.

<u>Female</u>

$E(VO_2\ Max_i\,|X) = -1.360 + 0.059$ ml/kg/min/s (exercise $duration_i$ (s))

<u>Male</u>

$E(VO_2\ Max_i\,|X) = 2.036 + 0.059$ ml/kg/min/s (exercise $duration_i$ (s))

7) What is the expected value of VO_2 Max for a female who spent 450 seconds on the treadmill?. What is the expected value of VO_2 max for a male who spent 450 seconds on the treadmill.

<u>Female</u>

$E(VO_2\ Max_i\,|X) = -1.360 + 0.059\ (450) = 25.19$ (ml/kg/min)

<u>Males</u>

$E(VO_2\ Max_i\,|X) = 2.036 + 0.059\ (450) = 28.57$ (ml/kg/min)

Regression Model 2

The following separate slopes multiple linear regression model was fit to the same data by least squares.

$E(VO_2\ Max_i \mid X) = \beta_o + \beta_1(\text{exercise duration}_i) + \beta_2(z_{2,i}) + \beta_3(z_{2,i} \text{ x exercise duration}_i)$

where $z_{2,i}$ = 1 if the *ith* participant was male, and 0 if *ith* participant was female.

Regression Analysis Summary

Table 3. The regression ANOVA table from the multiple regression analysis.

Parameter	DF	SS	MS	$F_{observed}$
Regression	3	6089.35	2029.12	130.57
Error	83	1290.27	15.54	
Total	86	7379.62		

8) Utilize the information in Table 1 and Table 3 to compute the extra-sum of squares F-test for the null hypothesis Ho: $\beta_3 = 0$.

$$F^* = \frac{\dfrac{SSE_R - SSE_F}{df_R - df_F}}{\dfrac{SSE_F}{df_F}} = \frac{\dfrac{1335.60 - 1290.27}{84 - 83}}{\dfrac{1290.27}{83}} = 2.916$$

9) For a two-sided test with significance level 0.05 we reject Ho: β_3=0, if $F_{observed} \geq F_{(1,83,.95)}$ = 3.955. Do we reject?.

$F_{observed}$ 2.916 $< F_{(1,83,.95)}$ = 3.955, so we should not reject Ho: β_3=0.

10) We observe from a sample of 44 paired measurements a sample correlation r=0.35. Based on this information compute the value of the one-sample *t*-Test ($t_{observed}$) for testing the the null hypothesis that $\rho = 0$.

$$t_{observed} = \frac{r(n-2)^{1/2}}{(1-r^2)^{1/2}} = \frac{0.35(44-2)^{1/2}}{(1-0.35^2)^{1/2}} = 2.42$$

11) For a two-sided test with significance level $\alpha = 0.05$ we reject Ho: $\rho = 0$ if $|t_{observed}| \geq t_{(42,0.975)}$= 2.08. Do we reject Ho?.

$t_{observed} = 2.42 \geq t_{(42,0.975)}$= 2.08 so we should reject Ho: $\rho = 0$.