1. Calculate a 95% confidence interval for a statistic that indicates the strength of association between achievement and anxiety for children with the same level of locus of control.

There are two measures of the strength of association between achievement and anxiety for children with the same level of locus of control:

1. Squared semi-partial: $\rho^2_{Anx(Ach \bullet Loc)}$

2. Squared partial: $\rho^2_{Anx,Ach \bullet Loc}$

We do not know how to construct a CI for the squared semi-partial.

We do not know how to construct a CI for the squared partial, but we can construct a CI for partial correlation.

Compute $\rho_{Anx,Ach \bullet Loc}$. We can use the following formula

$$r_{Anx,Ach \bullet Loc} = \sqrt{\frac{t^2_{Ach}}{t^2_{Ach} + n - k - 1}}$$

We require t_{Ach}

$$t_{Ach} = \frac{b_{Ach}}{S_{b_{Ach}}} = \frac{-1.5}{.8} = -1.875$$

$$r_{Anx,Ach \bullet Loc} = \sqrt{\frac{(-1.875)^2}{(-1.875)^2 + 103 - 2 - 1}}$$
$$= -.184$$

The negative sign is because the partial regression coefficient for achievement is negative.

Confidence interval

Step 1

$$Z_{r_{YX_j \bullet ----}} \pm z_{\alpha/2}\sqrt{\frac{1}{n-k-2}}$$

Form Fisher's Z table

$$Z_{-.184} = -.1861$$

The critical value

$$z_{.05/2} = 1.96$$

$$-.1861 \pm 1.96\sqrt{\frac{1}{103-2-2}}$$

$$(-.3831,\ .0109)$$

Step 2

$$(-.3831,\ .0109) \approx (-.3826,\ .0110)$$

Using Fisher's Z table in reverse, we find

$$(-.365,\ .011)$$

2. Suppose a researcher believes that the correlation between achievement and anxiety for children homogeneous on locus of control has a population value somewhere in the range -.40 to -.50. What does the interval in number 1 indicate about the researcher's beliefs? Why?

It indicates the researcher's belief is not consistent with the data. The interval in 1 tell us we are 95% confident that the partial correlation coefficient $\rho_{Anx,Ach \bullet Loc}$ is in the interval $(-.365, .011)$. The values hypothesized by the researcher are entirely outside this interval.

3. What is the estimated mean difference in anxiety for children that are 10 points apart on locus of control but have the same achievement score?

$$b_{Loc} = 2.3$$

so the estimated mean difference in anxiety for children that are 1 points apart on locus of control but have the same achievement score is 2.3. Therefore for children that are 10 points apart the estimated mean difference is 23 points.

4. Using α = .05 for each test, set up and test hypotheses relevant to the question of which, if either, of locus of control and achievement adds to prediction of anxiety when the other variable is already in the model?

One question we want to answer is whether or not locus of control adds to prediction of anxiety, when achievement is already in the model. This is the same as asking whether or not locus of control is related to anxiety, with achievement controlled. So we test one of the following pairs:

$H_0 : \beta_{Loc} = 0$	$H_0 : \rho_{Anx,Loc \bullet Ach} = 0$	$H_0 : \rho^2_{Anx(Loc \bullet Ach)} = 0$
$H_A : \beta_{Loc} \neq 0$	$H_A : \rho_{Anx,Loc \bullet Ach} \neq 0$	$H_A : \rho^2_{Anx(Loc \bullet Ach)} > 0$

<u>Note</u>. $H_A : \rho^2_{Anx(Loc \bullet Ach)} > 0$ is used because a squared correlation coefficient cannot be less than zero. The alternative is still non-directional because either a positive or negative relationship between anxiety and locus of control could result in $\rho^2_{Anx(Loc \bullet Ach)} > 0$.

A non-directional hypothesis is used because locus of control contributes to prediction regardless of whether it is positively or negatively related to achievement.

Regardless of the selected form of the null and alternate hypothesis

$$t = \frac{b_{Loc}}{S_{b_{Loc}}} = \frac{2.30}{.92} = 2.5$$

and

$$\pm t_{\alpha/2, n-k-1} = \pm t_{.05/2, 103-2-1} = \pm 1.9837$$

We reject the null hypothesis and conclude that locus of control adds to prediction of anxiety when achievement is controlled

Another question we want to answer is whether or not achievement adds to prediction of anxiety, when locus of control is already in the model.

$H_0 : \beta_{Ach} = 0$	$H_0 : \rho_{Anx,Ach \bullet Loc} = 0$	$H_0 : \rho^2_{Anx(Ach \bullet Loc)} = 0$
$H_A : \beta_{Ach} \neq 0$	$H_A : \rho_{Anx,Ach \bullet Loc} \neq 0$	$H_A : \rho^2_{Anx(Ach \bullet Loc)} > 0$

<u>Note</u>. $H_A : \rho^2_{Anx(Ach \bullet Loc)} > 0$ is used because a squared correlation coefficient cannot be less than zero. The alternative is still non-directional because either a positive or negative relationship between anxiety and locus of control could result in $\rho^2_{Anx(Ach \bullet Loc)} > 0$.

Regardless of the selected form of the null and alternate hypothesis

$$t = \frac{b_{Ach}}{S_{b_{Ach}}} = \frac{-1.5}{.8} = -1.875$$

and

$$\pm t_{\alpha/2,n-k-1} = \pm t_{.05/2,103-2-1} = \pm 1.9837$$

We fail to reject the null hypothesis and cannot conclude that achievement adds to prediction of anxiety when locus of control is already in the model.

5. Calculate a 95% confidence interval on β_1 and interpret the interval.

$$b_j \pm t_{\alpha/2,n-k-1} S_{b_j}$$

$$2.30 \pm t_{.05/2,103-2-1}(.92)$$

$$2.30 \pm 1.9837(.92)$$

$$(.47,\ 4.13)$$

We are 95% confident that the population slope for locus of control is in the interval .47 to 4.13. An equivalent interpretation is we are 95% confident that for two groups of

students, whose members have the same achievement score, that differ by one point on the locus of control scale, the mean difference in anxiety will between .47 and 4.13 points.

6. What proportion of the total anxiety variance is uniquely associated with locus of control?

This requires the squared semi-partial correlation coefficient:

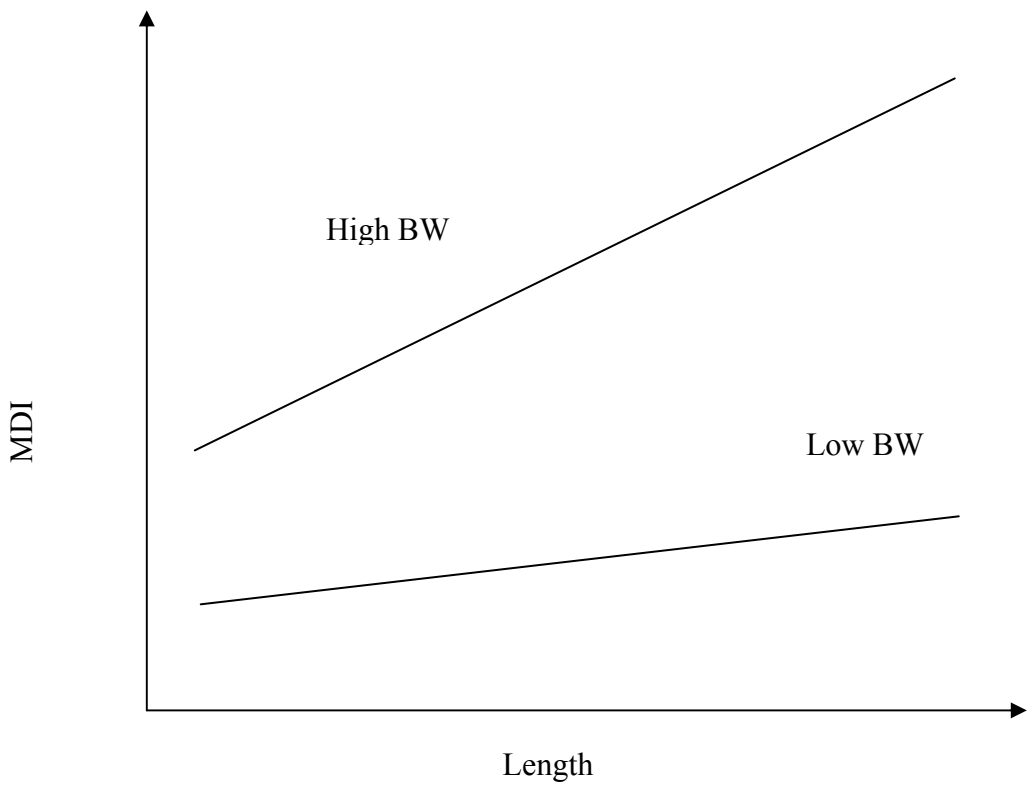
$$r^2_{Anx(Loc \bullet Ach)} = R^2_{inc,Loc}$$

$$R^2_{inc,Loc} = \frac{MSE}{SST} t^2_{Loc} = \frac{20}{2160} 2.5^2$$
$$= .057$$

5.7% of the total anxiety variance is associated uniquely with locus of control

7. Suppose the researcher believes that the average MDI increases less rapidly with changes in length for children with low birth weights than it does for children with high birth weights. Write a regression equation that could be used to investigate the researcher's beliefs.

The following graph depicts the researcher's beliefs



The appropriate regression equation is:

$$MDI = \alpha + \beta_1 (Length) + \beta_2 (Weight) + \beta_3 (Length)(Weight) + \varepsilon$$

8. How can the researcher determine whether the sample slope of the relationship between MDI and length is positive for children weighing 1800 grams at birth?

Calculate the conditional slope relating MDI to length with BW controlled. The formula for the conditional slope (aka the simple slopes) is coefficient for the variable of interest plus the coefficient of the product term multiplies by the other independent variable:

Coefficient for variable of interest + coefficient of product term $\times$ other independent variable

The variable of interests is length and its coefficient is b_1. The coefficient of the product term is b_3. The other variable is birth weight. So the conditional slope is

$$b_1 + b_3\left(BW\right)$$

To find the simple slope for children weighing 1800 grams at birth, substitute 1800 for birth weight:

$$b_1 + b_3\left(BW\right) = b_1 + b_3\left(1800\right)$$

9. What are the critical values for testing each of the following hypotheses against a two-tailed alternative if $n = 25$ and $\alpha = .05$?

Null Hypothesis	Critical Value	Null Hypothesis	Critical Value
$H_0 : \beta_4 = 0$	$\pm t_{\alpha/2,n-k-1}$ $= \pm t_{.05/2,25-4-1}$ $= \pm 2.0860$	$H_0 : \beta_2 = 0$	$\pm t_{\alpha/2,n-k-1}$ $= \pm t_{.05/2,25-2-1}$ $= \pm 2.0739$
$H_0 : \beta_3 = 0$	$\pm t_{\alpha/2,n-k-1}$ $= \pm t_{.05/2,25-3-1}$ $= \pm 2.0796$	$H_0 : \beta_1 = 0$	$\pm t_{\alpha/2,n-k-1}$ $= \pm t_{.05/2,25-1-1}$ $= \pm 2.0687$

10. Based on the available results what polynomial adequately fits the data? Why?
In each model we test the coefficient of the most complex term. For $H_0 : \beta_4 = 0$ in the quartic $t\left(20\right) = 1.0$ and is not more extreme than ± 2.086. So $H_0 : \beta_4 = 0$ is not rejected.

For $H_0 : \beta_3 = 0$ in the cubic, $t(21) = 2.5$ is more extreme than ± 2.08. So $H_0 : \beta_3 = 0$ is rejected. This tells us that we do not need the complexity of a quartic model, but do require the complexity of a cubic model.

11. For the quadratic model what proportion of the total variance in *Y* is uniquely associated with the quadratic term?

$$\begin{aligned} R^2_{inc,X^2} &= R^2(Quadratic) - R^2(Linear) \\ &= .49 - .20 = .29 \end{aligned}$$

12. Based on the results, which degree polynomial would seem to be required to adequately fit the data? (Do not calculate.)

Quadratic, the more complicated models do not increase R^2 appreciably.

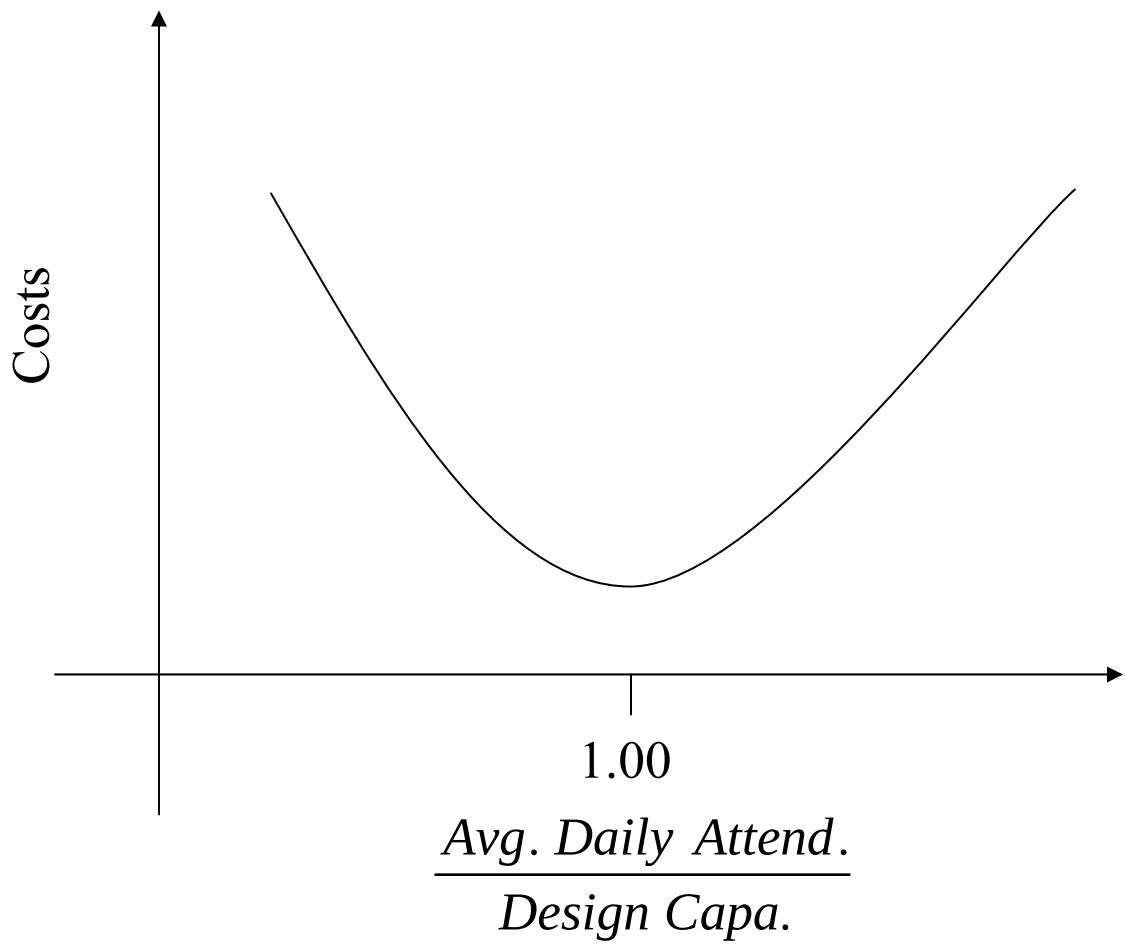
13. A superintendent of schools in a large city commissions the study of the following variables for elementary schools:

X = Average daily attendance divided by design capacity of each elementary school

Y = Yearly supply and maintenance cost for each school

The superintendent feels that a school that has an appropriate number of students for its design capacity will tend to have the lowest supply and maintenance costs. Schools with too few or too many students will be more costly to run. What kind of polynomial equation best represents his hypothesis about the X - Y relationship?

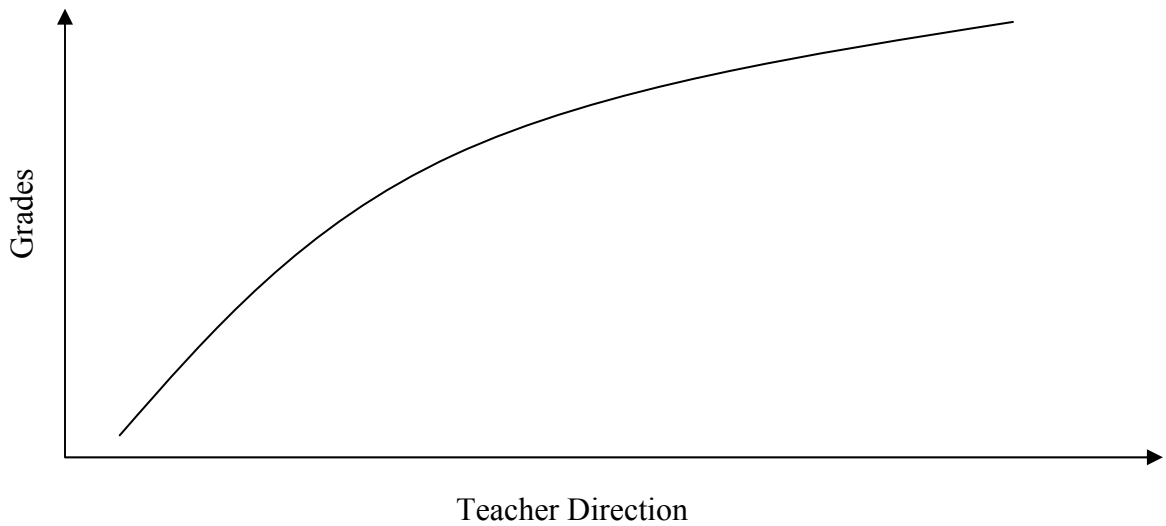
The following diagram depicts the principal's beliefs



To investigate the hypothesis we would use a quadratic model.

$$Costs = \alpha + \beta_1 Ratio + \beta_2 Ratio^2 + \varepsilon$$

14. A researcher believes that the relationship between teacher direction and grades looks like the following graph:

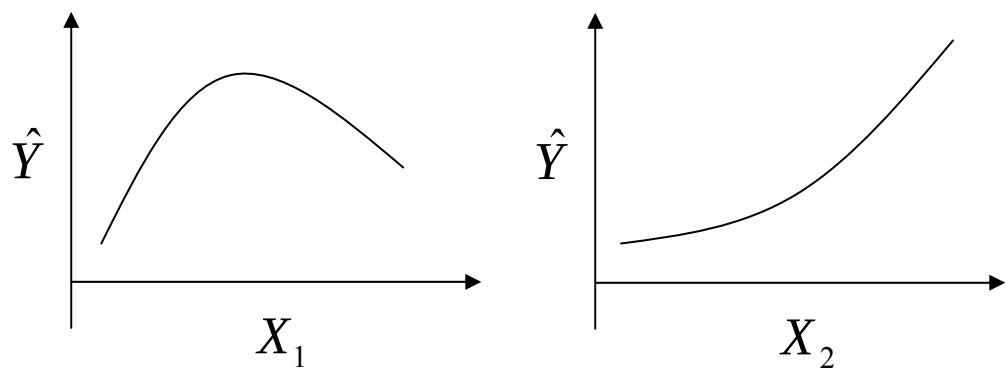


The researcher fits a quadratic model $\left(Y = \alpha + \beta_1 X + \beta_2 X^2 + \varepsilon\right)$ to data on the two variables and finds $b_2 = 1.5$ and that the statistic is significant in a non-directional test. Does this result support the researcher's hypothesis? Why?

No because $b_2 = 1.5$ implies a cup-up curve and the principal hypothesizes a cup-down curve.

15. Based on the numbers in the ESTIMATE column, on the graphs below, for each of $\hat{Y}$ and X_1 and $\hat{Y}$ and X_2, draw a graph indicating whether the relationships is cup-up or cup-down.

The coefficient for the squared term tells us whether the curve is cup-up or cup-down. For corrective feedback, the coefficient of the squared term $\left(X_1^2\right)$ is negative, so for corrective feedback the curve is cup down. For easy difficulty level, the coefficient of the squared term $\left(X_2^2\right)$ is positive, so the curve is cup-up:



16. What is the numeric value of the standard error of estimate for this model?
The standard error is reported as ROOT MSE, so the standard error of estimate is1.932. Alternatively, the conditional variance is reported as the mean square for error. The standard error of estimate is the square root of the mean square for error. So the standard error of estimate is $\sqrt{3.734} = 1.932$

17. What do the results imply about the nature of the relationship between Y and X_1 for people who have the same score on X_2? Support your answer with the appropriate statistical test(s).

We start by testing

$$H_0 : \beta_2 = 0$$

$$H_A : \beta_2 \neq 0$$

The test statistic is $t(24) = -.50$ and the critical value is

$\pm t_{\alpha/2,n-k-1} = \pm t_{.05/2,29-4-1} = \pm 2.0639$; so we fail to reject and conclude that the squared term for X_1 is not necessary.

18. Assume the residual plots indicate there are no violations of assumptions. Based on the hypothesis tests that should be conducted for the model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \varepsilon$$

what model, if any, would you estimate next? For which coefficient(s) of this new model would you test null hypotheses?

The evidence in 17 indicated we do not need the squared term for X_1. Our next model is

$$Y = \alpha + \beta_1 X_1 + \cancel{\beta_2 X_1^2} + \beta_3 X_2 + \beta_4 X_2^2 + \varepsilon$$

$$Y = \alpha + \beta_1 X_1 + \beta_3 X_2 + \beta_4 X_2^2 + \varepsilon$$

We would want to test $H_0 : \beta_1 = 0$ to determine if there is a linear trend to the relationship between Y and X_1.

We do not need to test $H_0 : \beta_4 = 0$ because it is rejected in the original model $t(24) = 3.00$.

We would not test $H_0 : \beta_3 = 0$ in the new model (or in the original model) because it refers to the direction of change at $X_2 = 0$ parameter.

19. Estimate the squared multiple correlation.

We start by calculating R^2

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{16499.47}{36182.40}$$
$$= .543$$

Next we calculated adjusted R^2

$$
\begin{aligned}
R_c^2 &= R^2 - \frac{k}{n-k-1}\left(1-R^2\right) \\
&= .543 - \frac{3}{50-3-1}\left(1-.543\right) \\
&= .513
\end{aligned}
$$

20. Set up and test a hypothesis relevant to the question of whether the model is useful in predicting electricity bills? Use α = .05

The model is useful if any of the variables predict electricity usage. So we can test any of the following equivalent hypotheses (recall that equivalent means that if one is true the others are true and if one is false the others are false):

$$H_0 : \rho^2_{Y(X_1X_2X_3)} = 0$$

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_0 : \rho_{YX_1 \bullet X_2X_3} = \rho_{YX_2 \bullet X_1X_3} = \rho_{YX_3 \bullet X_1X_2} = 0$$

$$H_0 : \rho^2_{Y(X_1 \bullet X_2X_3)} = \rho^2_{Y(X_2 \bullet X_1X_3)} = \rho^2_{Y(X_2 \bullet X_1X_2)} = 0$$

$$H_0 : \rho^2_{inc,X_1} = \rho^2_{inc,X_2} = \rho^2_{inc,X_3} = 0$$

$$H_0 : \rho_{YX_1} = \rho_{YX_2} = \rho_{YX_3} = 0$$

Regardless of the form of the hypothesis, the test statistic is

$$
\begin{aligned}
F &= \frac{n-k-1}{k}\frac{R^2}{1-R^2} \\
&= \frac{50-3-1}{3}\frac{.543}{1-.543} = 18.22
\end{aligned}
$$

You could also complete the summary ANOVA table

SV	DF	SS	**MS**	**F**
MODEL	**3**	19682.93	**6560.976**	**18.29**
ERROR	**46**	16499.47	**358.684**	
TOTAL		36182.40		

In either case the critical value is

$$F_{\alpha,k,n-k-1} = F_{.05,3,50-3-1} = 2.8068$$

We reject the null hypothesis and conclude that the model is useful in predicting electric usage

21. What, if anything, would failure to reject the null hypothesis in question 20 imply about ρ_{YX_1}, ρ_{YX_2}, and ρ_{YX_3}?

It would mean we cannot conclude that any of the three zero-order correlation are not equal to zero.

22. How much does R^2 increase when AC usage is added to the model with washing machine and dishwasher usage already in it?

The squared semi-partial correlation coefficient is interpreted as the increase in R^2 when the variable of interest is added to a model that already includes the other variables. So we calculate squared semi-partial correlation coefficient for AC usage:

$$R^2_{inc,AC} = \frac{MSE}{SST} t^2_{AC}$$

$$MSE = \frac{SSE}{n-k-1} = \frac{16499.47}{50-3-1} = 358.684$$

$$t_{AC} = \frac{b_{AC}}{S_{b_{AC}}} = \frac{0.330}{0.069}$$
$$= 4.78$$

$$R^2_{inc,AC} = \frac{358.684}{36182.40} 4.78^2$$
$$= .226$$

23. b

24. a

25 d.

26. a

27. a