Name:_______________________________________ Class:________________________ Date:_____________

Chapter 02

Multiple Choice

1. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of $Y = Time$ (expected trip time to hike the peak, in hours) on $X = Ascent$ (in feet). What is $SE_{\hat{\beta}}$?

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541 1.8661683   2.256  0.02909 *
Ascent      0.0020805 0.0005909   3.521  0.00101 **
---
Signif. codes: 0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1

Residual standard error: 2.496 on 44 degrees of freedom
Multiple R-squared: 0.2198,     Adjusted R-squared: 0.2021
F-statistic: 12.4 on 1 and 44 DF, p-value: 0.001014
```

a. 4.21005

b. 1.86617

c. 0.00208

d. 0.00059

e. 2.496

f. 0.2198

g. 0.2021

h. 0.001014

ANSWER: d

2. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of $Y = Time$ (expected trip time to hike the peak, in hours) on $X = Ascent$ (in feet). What is the coefficient of determination?

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541 1.8661683   2.256  0.02909 *
Ascent      0.0020805 0.0005909   3.521  0.00101 **
---
Signif. codes: 0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1

Residual standard error: 2.496 on 44 degrees of freedom
Multiple R-squared: 0.2198,     Adjusted R-squared: 0.2021
F-statistic: 12.4 on 1 and 44 DF, p-value: 0.001014
```

a. 4.21005

b. 1.86617

c. 0.00208
d. 0.00059
e. 2.496
f. 0.2198
g. 0.2021
h. 0.001014

ANSWER: f

3. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. The relationship between Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet) is positive. The linear regression model *Time* on *Ascent* results in r^2 = 21.98%. Determine the correlation coefficient.

a. 0.0483
b. 0.2198
c. 0.4688
d. 4.6883
e. Unable to determine

ANSWER: c

4. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet). Use this output to calculate the coefficient of determination.

```
> anova(time.lm)
Analysis of Variance Table
Response: Time
          Df  Sum Sq Mean Sq F value   Pr(>F)
Ascent     1  77.261  77.261  12.399 0.001014 **
Residuals 44 274.174   6.231
```

a. 0.2198
b. 0.2818
c. 12.399%
d. Unable to determine

ANSWER: a

5. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet).

Find the *P*-value for the test of the hypothesis that the correlation between *Time* and *Ascent* is 0.

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541 1.8661683   2.256  0.02909 *
Ascent      0.0020805 0.0005909   3.521  0.00101 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.496 on 44 degrees of freedom
Multiple R-squared: 0.2198,     Adjusted R-squared: 0.2021
F-statistic: 12.4 on 1 and 44 DF, p-value: 0.001014
```

a. 0.02909

b. 0.00101

c. Unable to determine

ANSWER: b

6. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of *Y* = *Time* (expected trip time to hike the peak, in hours) on *X* = *Ascent* (in feet).

Report a 95% confidence interval for the mean trip time when *Ascent* is 3000 feet.

```
> predict.lm(time.lm, newdata=data.frame(Ascent=3000),interval="confidence")
       fit      lwr      upr
1 10.45163 9.701043 11.20222
> predict.lm(time.lm, newdata=data.frame(Ascent=3000),interval="prediction")
       fit      lwr      upr
1 10.45163 5.365099 15.53816
```

a. (9.701 hours, 11.202 hours)

b. (5.365 hours, 15.538 hours)

c. 10.452 hours

d. Unable to determine

ANSWER: a

7. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of *Y* = *Time* (expected trip time to hike the peak, in hours) on *X* = *Ascent* (in feet).

Report a 95% prediction interval for the trip time when *Ascent* is 3000 feet. Then interpret the interval in context.

```
> predict.lm(time.lm, newdata=data.frame(Ascent=3000),interval="confidence")
       fit      lwr      upr
1 10.45163 9.701043 11.20222
> predict.lm(time.lm, newdata=data.frame(Ascent=3000),interval="prediction")
       fit      lwr      upr
1 10.45163 5.365099 15.53816
```

a. (9.701 hours, 11.202 hours)

b. (5.365 hours, 15.538 hours)

c. 10.452 hours

d. Unable to determine

ANSWER: b

8. Consider the following statement: "When predicting the value of Y at a value $X = x*$, the 90% confidence interval for the mean response is wider than the 90% prediction interval for a particular response."

Is this statement always true, sometimes true, or never true?

a. Always true
b. Sometimes true
c. Never true

ANSWER: c

9. Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol, which results in r^2 = 37.5%. From this information, which of the following are true? (You may select more than one.)

a. 37.5% of the variability in HDL is explained by the linear relationship with total cholesterol.
b. 37.5% of the variability in total cholesterol is explained by the linear relationship with HDL.
c. The relationship between HDL and total cholesterol is important.
d. The relationship between HDL and total cholesterol is strong.
e. The evidence for an association between HDL and total cholesterol is strong.
f. The relationship between HDL and total cholesterol is positive.

ANSWER: a

10. Cholesterol levels are measured on a sample of 21 volunteers. Our response variable is HDL (high-density lipoprotein, or "good" cholesterol). Below are the correlations between HDL and three different potential predictors: total cholesterol (*Chol*), total triglycerides (*Triglyc*), and the presence (1) or absence (0) of a sticky substance called sinking pre-beta (*SPB*).

Based on the information in the correlations, which of the potential predictors (*Chol*, *Triglyc*, *SPB*) is the *weakest* predictor (on its own) of the HDL response variable?

```
> cor(HDL)
              HDL      Chol   Triglyc       SPB
HDL     1.0000000 0.6123659 0.7236147 0.6698262
Chol    0.6123659 1.0000000 0.6972721 0.3102052
Triglyc 0.7236147 0.6972721 1.0000000 0.4154681
SPB     0.6698262 0.3102052 0.4154681 1.0000000
```

a. *Chol*
b. *Triglyc*
c. *SPB*
d. Unable to determine

ANSWER: a

Multiple Response

11. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations

near or above 4000 feet. *Y* = *Time* (expected trip time to hike the peak, in hours) is regressed on *X* = *Ascent* (in feet). The *t* test for the slope results in a *P*-value of 0.001. Based *only* on this information, which of the following is true? Assume all the conditions for the model are met. (You may select more than one.)

a. The probability that there is no linear relationship is 0.001.
b. The probability that there is a linear relationship is 0.001.
c. If there is no linear relationship between *Time* and *Ascent*, the probability of getting results like ours is about 0.001.
d. If there is a linear relationship between *Time* and *Ascent*, the probability of getting results like ours is about 0.001.
e. The relationship between *Time* and *Ascent* is important.
f. The relationship between *Time* and *Ascent* is strong.
g. The evidence for an association between *Time* and *Ascent* is strong.

ANSWER: c, g

12. Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol. An ANOVA *F* test is performed and returns a *P*-value of 0.003. Assuming all the conditions for the model are met, which of the following is true? (You may select more than one.)

a. The probability that there is no linear relationship is 0.003.
b. The probability that there is a linear relationship is 0.003.
c. If there is no linear relationship between HDL and total cholesterol, the probability of getting results like ours is about 0.003.
d. If there is a linear relationship between HDL and total cholesterol, the probability of getting results like ours is about 0.003.
e. The relationship between HDL and total cholesterol is important.
f. The relationship between HDL and total cholesterol is strong.
g. The evidence for an association between HDL and total cholesterol is strong.
h. The relationship between HDL and total cholesterol is positive.

ANSWER: c, g

Essay

13. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of *Y* = *Time* (expected trip time to hike the peak, in hours) on *X* = *Ascent* (in feet).

Use this output to test the null hypothesis that *Ascent* is not linearly related to *Time*. Provide the *P*-value and state the conclusion.

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541  1.8661683   2.256  0.02909 *
```

```
Ascent      0.0020805  0.0005909   3.521  0.00101 **
---
Signif. codes: 0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1
```

ANSWER: *P*-value = 0.001. Assuming all conditions for the linear model are met, there is strong evidence that *Ascent* and *Time* have a linear relationship.

14. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet).

We believe that *Ascent* and *Time* will have a positive relationship. Use this output to test this hypothesis. Provide the *P*-value and state the conclusion.

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541  1.8661683   2.256  0.02909 *
Ascent      0.0020805  0.0005909   3.521  0.00101 **
---
Signif. codes: 0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘’1
```

ANSWER: *P*-value = 0.001/2 = 0.0005. Assuming all conditions for the linear model are met, there is strong evidence that *Ascent* and *Time* have a positive linear relationship.

15. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet).

Use this output to test the null hypothesis that *Ascent* is not linearly related to *Time*. Provide the *P*-value and state the conclusion.

```
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Residual standard error: 2.496 on 44 degrees of freedom
Multiple R-squared: 0.2198,     Adjusted R-squared: 0.2021
F-statistic: 12.4 on 1 and 44 DF, p-value: 0.001014
```

ANSWER: *P*-value = 0.001. Assuming all conditions for the linear model are met, there is strong evidence that *Ascent* and *Time* have a linear relationship.

16. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet).

Use this output to test the null hypothesis that *Ascent* is not linearly related to *Time*. Provide the *P*-value and state the conclusion.

```
> anova(time.lm)
Analysis of Variance Table
```

Chapter 02

```
Response: Time
          Df  Sum Sq Mean Sq F value   Pr(>F)
Ascent     1  77.261  77.261  12.399 0.001014 **
Residuals 44 274.174   6.231
```

ANSWER: *P*-value = 0.001. Assuming all conditions for the linear model are met, there is strong evidence that *Ascent* and *Time* have a linear relationship.

17. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. We fit a linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet). The 95% confidence interval for β_1, the coefficient of *Ascent*, is (0.00089, 0.00327). Interpret this interval in the context of the problem.

ANSWER: We are 95% confident that for each additional foot of ascent, the trip should take between 0.00089 and 0.00327 additional hours. Or, if we rephrase to have more pleasant units: We are 95% confident that for each additional 1000 feet of ascent, the trip should take between 0.89 and 3.27 additional hours.

18. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. The linear regression model of Y = *Time* (expected trip time to hike the peak, in hours) on X = *Ascent* (in feet) results in r^2 = 21.98%. Interpret this value in context.

ANSWER: About 22% of the variability in trip *Time* is explained by the linear relationship with *Ascent*.

19. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Y = *Time* (expected trip time to hike the peak, in hours) is regressed on X = *Ascent* (in feet). When *Ascent* is 3000 feet, the 95% confidence interval for the mean is (9.701, 11.202). Interpret this interval in the context of this problem.

ANSWER: We are 95% confident that the average trip time for all ascents of 3000 feet is between 9.7 hours and 11.2 hours.

20. Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol. An ANOVA *F* test is performed and returns a *P*-value of 0.003. Write down the hypotheses being tested.

ANSWER: We are testing if the linear model is effective. that is, whether the slope of the linear model is 0. In symbols,
$H_0{:}\beta_1 = 0$ vs. $H_0{:}\beta_1 \neq 0$

21. Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol. An ANOVA *F* test is performed and returns a *P*-value of 0.003. Make a conclusion based on this *P*-value. (You may assume all conditions are met.)

ANSWER: We have strong evidence that there is a linear relationship between HDL and total cholesterol.

22. Cholesterol levels are measured on a sample of 21 volunteers. Our response variable is HDL (high-density lipoprotein, or "good" cholesterol). We are interested in the correlations between HDL and three different potential predictors: total cholesterol (*Chol*), total triglycerides (*Triglyc*), and the presence (1) or absence (0) of a sticky substance called sinking pre-beta (*SPB*).

Below is some R output. Write down the hypotheses being tested here.

Chapter 02

```
> cor.test(x=HDL$Triglyc,y=HDL$HDL, conf.level=0.9)
Pearson's product-moment correlation
data: HDL$Triglyc and HDL$HDL
t = 4.5699, df = 19, p-value = 0.0002093
alternative hypothesis: true correlation is not equal to 0
90 percent confidence interval:
0.4834646 0.8624646
sample estimates:
     cor
0.7236147
```

ANSWER: We are testing if there is significant correlation between HDL and total triglycerides. In symbols, $H_0{:}\rho = 0$ vs. $H_0{:}\rho \neq 0$

23. Cholesterol levels are measured on a sample of 21 volunteers. The correlation between HDL (high-density lipoprotein, or "good" cholesterol) and triglycerides is 0.723, with an associated 90% confidence interval of (0.4835, 0.8625). Interpret this interval in the context of this problem.

ANSWER: Based on this sample of 21, we are 90% confident that the true correlation between HDL and triglycerides is between 0.4835 and 0.8635.

24. Cholesterol levels are measured on a sample of 21 volunteers. HDL (high-density lipoprotein, or "good" cholesterol) is regressed on total cholesterol. We want to find an interval that would be 95% certain to contain the actual HDL level for a patient with a total cholesterol level of 280 mg/dl. Do we want to construct a confidence interval or a prediction interval?

ANSWER: prediction interval

25. Below is a partial ANOVA table from a simple linear regression model. Using only the information given, fill in the missing values (A, B, and C).

```
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value  Pr(>F)
X          1       A  791.45    11.4 0.00317
Residuals  C 1319.12       B
```

ANSWER: A = 791.45; B = 69.43; C = 19