Started on	Tuesday, 15 August 2017, 10:11 AM
State	Finished
Completed on	Tuesday, 15 August 2017, 10:53 AM
Time taken	42 mins 8 secs
Grade	**9.0** out of 10.0 (**90**%)
Feedback	Congratulations! You have passed the quiz and mastered the simple linear regression.

Question 1

Correct

Mark 0.5 out of 0.5

Please select the correct statement, from A, B, C, D, E in the following.

Select one:

- A. Simple linear regression models the mean of response as a function of an explanatory variable. ✓
- B. Simple linear regression models the response as a function of an explanatory variable.
- C. Simple linear regression models the response as a function of the mean of an explanatory variable.
- D. Simple linear regression models the mean of response as a function of the mean of an explanatory variable.
- E. None of the above.

Your answer is correct.

The correct answer is: Simple linear regression models the mean of response as a function of an explanatory variable.

Question 2

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

Suppose we are interested in using a person’s height to predict his/her weight by employing simple linear regression model. The data consist of the heights and weights for 1,000 persons. Simple linear regression assumptions allow the existence of the height and weight data for a father and his son in the regression procedure.

Select one:

- True
- False ✓

Please see Lecture Notes 1.

The correct answer is 'False'.

Question 3

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

In the simple linear regression assumptions, sometimes the normality assumption can be violated since the least squares estimates are robust to some non-normal distributions, e.g., the heavy-tailed distribution or the skewed distribution.

Select one:

- True ✓
- False

Please see Lecture Notes 3.

The correct answer is 'True'.

Question 4

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

The parameters in the simple linear regression model are unknown for real data. But they are known if we use the simple linear regression model to generate random samples for simulation.

Select one:

- True ✓
- False

Please see Lecture Notes 2.

The correct answer is 'True'.

Question 5

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

If a Q-Q plot for the residuals in a simple linear regression model shows a heavy-tailed case, then the response variable follows a *t* distribution.

Select one:

- True
- False ✓

Please see Lecture Notes 3.

The correct answer is 'False'.

Question 6

Incorrect

Mark 0.0 out of 0.5

Is the following statement “True” or “False”?

The sampling distributions of the least squares estimates do not depend on the simple linear regression model assumptions.

Select one:

- True ✗
- False

Please see Lecture Notes 3.

The correct answer is 'False'.

Question 7

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

The standard error of the least squares estimate is the standard deviation of the least squares estimate.

Select one:

- True
- False ✓

Please see Lecture Notes 2.

The correct answer is 'False'.

Question 8

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

The least squares estimates are unbiased means that the mean of a least squares estimator equals to the true value of the corresponding regression parameter.

Select one:

- True ✓
- False

Please see Lecture Notes 2.

The correct answer is 'True'.

Question 9

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

We can use both the response versus explanatory variable plot and the residuals versus fitted values plot to detect the violation for the linearity assumption in simple linear regression models.

Select one:

- True ✓
- False

Please see Lecture Notes 3.

The correct answer is 'True'.

Question 10

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

Sometime, the reason why we do not have enough evidence to reject the null hypothesis is that the sample size of the data is too small.

Select one:

- True ✓
- False

Please see Lecture Notes 2.

The correct answer is 'True'.

Question 11

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

In the *t*-test for whether or not the regression coefficient of the explanatory variable is 0, if we reject the null hypothesis, we can conclude that the *t*-test is significant and the explanatory variable is significant in the simple linear regression model.

Select one:

- True ✓
- False

Please see Lecture Notes 2.

The correct answer is 'True'.

Information

Effectiveness of Measles Vaccine. The data consists of the number of measles cases reported in the United States for each year from 1950 to 2008. (The dataset was from the Centers for Disease Control, and was retrieved on July 23, 2009.) The dataset is stored in the object "data" in R. The first 6 rows of the dataset are shown using the following R code:

```
> head(data)
  Year  Cases
1 1950 319124
2 1951 530118
3 1952 683077
4 1953 449146
5 1954 682720
6 1955 555156
```

We are interested in how strong the evidence is, that the time changed the number of measles cases. Hence we include “Year” in a regression model to predict the logarithm of “Cases” in the dataset.

We perform the following simple linear regression analysis

```
> Cases=data$Cases
> Year=data$Year
> fit=lm(log(Cases)~Year)
> summary(fit)

Call:
lm(formula = log(Cases) ~ Year)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8083 -0.8006 -0.1031  0.5677  3.0315

Coefficients:
             Estimate          Std. Error   t value   Pr(>|t|)
(Intercept) 369.425780  14.423963   25.61   <2e-16 ***
Year         -0.182023       0.007288      A       <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.9533 on 57 degrees of freedom

Multiple R-squared: 0.9163, Adjusted R-squared: 0.9148

F-statistic: 623.7 on 1 and 57 DF, p-value: < 2.2e-16

Based on the R code and R output, please answer the following Questions 12-19.

Question 12

Correct

Mark 0.5 out of 0.5

What is the value of "**A**" in the aforementioned "summary()" R output? Please round your final answer to four decimal places. But in the middle the computation process, if some figures in the R output are used, please take all the decimal places shown in the R output.

Answer: -24.9757 ✓

Please see Lecture Notes 2.

The correct answer is: -24.9757

Question 13

Correct

Mark 0.5 out of 0.5

What is the interpretation of the coefficient estimate "-0.182023" from the aforementioned R output?

Select one:

- A. The log(Cases) will decrease 0.182023 unit for this year compared to that for the previous year.
- B. The mean of log(Cases) will decrease 0.182023 unit for this year compared to that for the previous year.
- C. The estimated mean of log(Cases) will decrease 0.182023 unit for this year compared to that for the previous year. ✓
- D. None of the above.

Your answer is correct.

The correct answer is: The estimated mean of log(Cases) will decrease 0.182023 unit for this year compared to that for the previous year.

Question 14

Correct

Mark 0.5 out of 0.5

What is the prediction of “Cases” for year 1956 based on the established simple linear regression model? Please round your final answer to four decimal places. But in the middle the computation process, if some figures in the R output are used, please take all the decimal places shown in the R output.

Answer: 652647.2087

Please see Lecture Notes 2.

The correct answer is: 652647.2087

Question 15

Correct

Mark 0.5 out of 0.5

Based on the R code and R output, please select the correct statement, from A, B, C, D, E in the following.

Select one:

- A. The regression coefficient (parameter) for the variable “Year” is numerically different from 0.
- B. The regression coefficient (parameter) for the variable “Year” is significantly different from 0.
- C. We accept the null hypothesis that the regression coefficient (parameter) for the variable “Year” is 0.
- D. We do not have enough evidence to reject the null hypothesis that the regression coefficient (parameter) for the variable “Year” is 0.
- E. None of the above.

Your answer is correct.

The correct answer is: The regression coefficient (parameter) for the variable “Year” is significantly different from 0.

Question 16

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

Based on the R code and R output, an increase in “Year” causes a decrease in “Cases”.

Select one:

- True
- False ✓

Please see Lecture Notes 2.

The correct answer is 'False'.

Question 17

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

The 90% prediction interval for a future response at “Year” 2018 has a larger range compared to the 90% confidence interval for the mean of response at “Year” 2018.

Select one:

- True ✓
- False

Please see Lecture Notes 2.

The correct answer is 'True'.

Question 18

Correct

Mark 0.5 out of 0.5

Is the following statement “True” or “False”?

Based on the R code and R output, those *p*-values saved in the aforementioned “summary(fit)” can still be used for hypothesis testing if the response versus explanatory variable plot for the real data shows the violation of the constant variance assumption.

Select one:

- True
- False ✓

Please see Lecture Notes 3.

The correct answer is 'False'.

Question 19

Correct

Mark 0.5 out of 0.5

In addition, consider the following R code and R output.

```
> X=data.frame(Year=2018)
> predict(fit,X,interval='confidence',level=0.95)
       fit      lwr      upr
1 2.104142 1.483063 2.725221
```

What is the interpretation of 1.483063 and 2.725221?

Select one:

- A. We have 95% confidence that the mean of “log(Cases)” at “Year” 2018 lies in [1.483063, 2.725221]. ✓
- B. The probability that the mean of “log(Cases)” at “Year” 2018 lies in [1.483063, 2.725221] is 95%.
- C. We have 95% confidence that the response “log(Cases)” at “Year” 2018 lies in [1.483063, 2.725221].
- D. The probability that the response “log(Cases)” at “Year” 2018 lies in [1.483063, 2.725221] is 95%.
- E. None of the above.

Your answer is correct.

The correct answer is: We have 95% confidence that the mean of “log(Cases)” at “Year” 2018 lies in [1.483063, 2.725221].

Question 20

Incorrect

Mark 0.0 out of 0.5

We have the following simulation experiment

```
> beta0=2;beta1=-1
>
> n=5
> R=1000
>
> hatbeta1=rep(0,R)
> for(r in 1:R) {
+   X=1:n
+   errors=rnorm( n )
+   Y=beta0+beta1*X+errors
+   SLRfit=lm(Y~X)
+   hatbeta1[r]=SLRfit$coef[2]
+ }
>
> sd(hatbeta1)
```

What is the most appropriate guess for the R output of “sd(hatbeta1)”? Please round your final answer to four decimal places. Please do not round in the middle the computation process.

Answer: 0.3305 ✗

Please see Lecture Notes 2.

The correct answer is: 0.3162