Statistical and Time Series Analysis of Climate Variables

Self Assessment Quiz

1. The quiz is both looking back and looking forward. The looking back part is to let you assess if you have studied the basic statistical tools to allow you to enjoy the course fully. Note that they are taken from an exam of a second year engineering modelling course.

   (a) I was asked to construct a linear regression model for the amount of light penetrating 6 metres into an office when a special louvre reflector was fitted to the window. After some investigations, it was determined that the possible explanatory variables are the amount of solar radiation on a horizontal surface (Radiation), the angle the sun makes with the horizontal (Altitude), and the angle the sun is away from due North (Azimuth). Use the Minitab output below to determine

      (i) whether there is a significant relationship between the variables,

      (ii) whether the y-intercept is significantly different from zero,

      (iii) which explanatory variables have significant regression coefficients.

      (iv) what the R^2 value is and what it means.

      (v) **Note that in the exam, the students did not have access to software so they could not perform any additional analysis. Given the results of the tests performed, what extra analysis, if any, would you undertake?**

Regression Analysis: Light versus Radiation, Altitude, Azimuth

```
The regression equation is
Light = 2.34 + 36.3 Radiation + 2.38 Altitude - 0.156 Azimuth

Predictor        Coef     SE Coef          T        P
Constant        2.343       1.678       1.40    0.166
Radiation       36.26       13.80       2.63    0.010
Altitude       2.3826      0.4006       5.95    0.000
Azimuth      -0.15585     0.09055      -1.72    0.088

S = 4.931       R-Sq = 99.2%     R-Sq(adj) = 99.2%

Analysis of Variance

Source             DF          SS          MS        F        P
Regression          3      295107       98369   4045.71    0.000
Residual Error     96        2334          24
Total              99      297441

Source       DF      Seq SS
Radiation     1      294127
Altitude      1         908
Azimuth       1          72
```

(b) Find below the descriptive statistics and histograms for the winter daily total solar radiation and summer daily total radiation (in $MJ/m.^2$) for Adelaide airport, as well as the graph of the empirical distribution of the summer daily totals.

(i) What statistical measures would you use to describe the centre and spread of the data in each case? Give reasons.

(ii) If X denotes the daily winter solar total, determine the value x such that $P(X < x) = 0.05$. **Hint**: From the answers in the previous question, you should be able to make an assumption about the distribution of this variable.

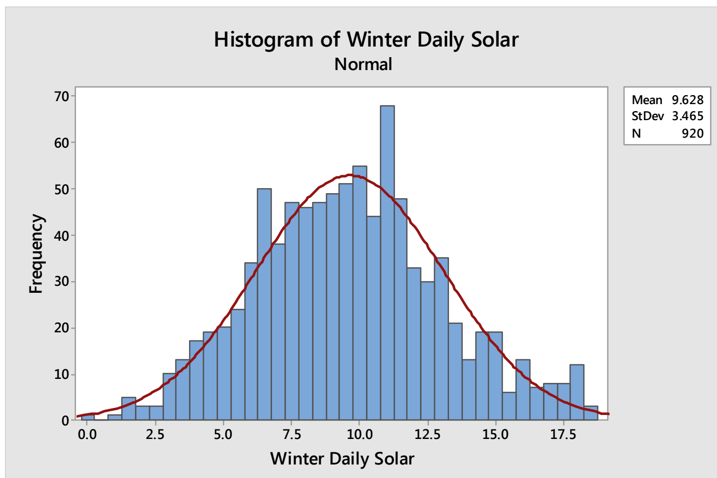
(iii) If Y denotes the summer daily total solar radiation, given the graph of the empirical cumulative distribution function (CDF) of the summer daily totals, determine y such that $P(Y < y) = 0.05$.

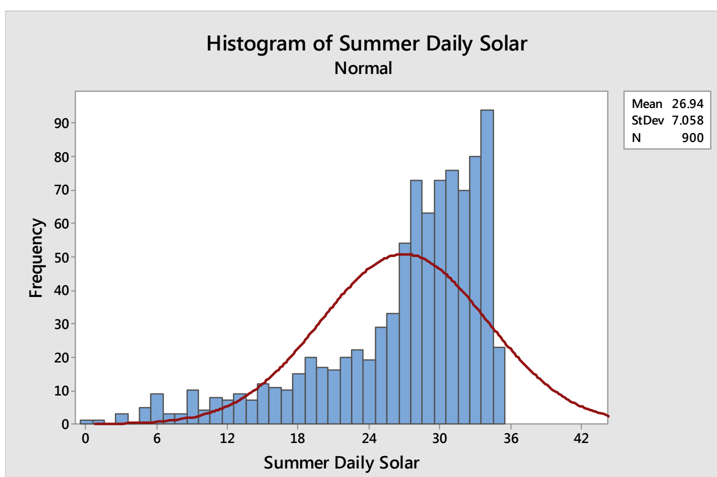
Descriptive Statistics: Winter Daily Solar

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Daily Solar	920	9.62	3.46	0.071	7.13	9.53	11.72	18.74

Descriptive Statistics: Summer Daily Solar

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Daily Solar	900	26.93	7.05	0.32	24.11	28.92	32.15	35.41





(c) Define homoscedasticity and heteroscedasticity and tell what significance they have for linear regression.

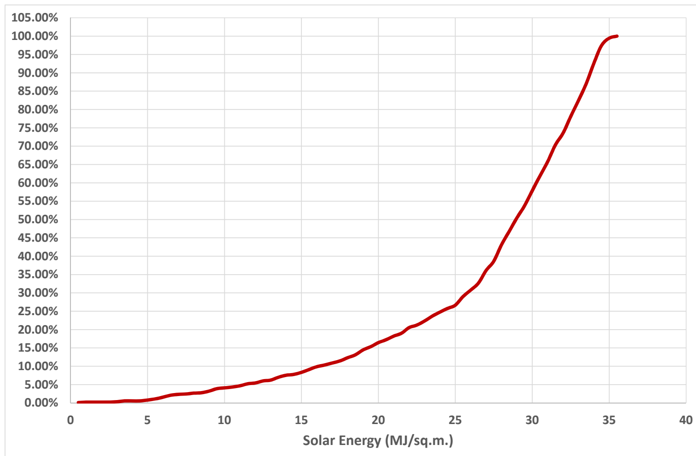


Figure 1: Summer Solar CDF

2. This part is not only for your benefit, but also mine. It will help me ensure that we can cover topics that are of use to you. What do you want to get out of this course for your future study or career? Please be specific and not give a general answer. What tools do you expect to gain knowledge of and how do you think that will help you in future?

Statistical and Time Series Analysis of Climate Variables

Self Assessment Quiz

1. The quiz is both looking back and looking forward. The looking back part is to let you assess if you have studied the basic statistical tools to allow you to enjoy the course fully. Note that they are taken from an exam of a second year engineering modelling course.

   (a) I was asked to construct a linear regression model for the amount of light penetrating 6 metres into an office when a special louvre reflector was fitted to the window. After some investigations, it was determined that the possible explanatory variables are the amount of solar radiation on a horizontal surface (Radiation), the angle the sun makes with the horizontal (Altitude), and the angle the sun is away from due North (Azimuth). Use the Minitab output below to determine

   (i) whether there is a significant relationship between the variables,

$$\begin{aligned} H_0 &: \text{there is no relationship between the light penetrating and the predictors} \\ H_a &: \text{there is} \\ \alpha &= 0.05 \end{aligned}$$

   Since the p-value of $0.000 < 0.05$, we reject H_0 and conclude there is a relationship.

   (ii) whether the y-intercept is significantly different from zero,

$$\begin{aligned} H_0 &: \text{the y-intercept} = 0 \\ H_a &: \text{it is significantly different from } 0 \\ \alpha &= 0.05 \end{aligned}$$

   Since the p-value of $0.166 > 0.05$, we do not reject H_0 and conclude that the y-intercept is zero.

   (iii) which explanatory variables have significant regression coefficients.

$$\begin{aligned} H_0 &: \text{radiation coefficient} = 0 \\ H_a &: \text{it is significantly different from } 0 \\ \alpha &= 0.05 \end{aligned}$$

Since the p-value of 0.01 $<$ 0.05, we reject H_0 and conclude that the radiation coefficient is significantly different from zero.

$$\begin{aligned} H_0 &: \text{altitude coefficient } = 0 \\ H_a &: \text{it is significantly different from } 0 \\ \alpha &= 0.05 \end{aligned}$$

Since the p-value of 0.000 $<$ 0.05, we reject H_0 and conclude that the altitude coefficient is significantly different from zero

$$\begin{aligned} H_0 &: \text{azimuth coefficient } = 0 \\ H_a &: \text{it is significantly different from } 0 \\ \alpha &= 0.05 \end{aligned}$$

Since the p-value of 0.088 $>$ 0.05, we do not reject H_0 and conclude that the azimuth coeffcient is zero.

(iv) what the R^2 value is and what it means.
$R^2 = 99.2\%$. This means that 99.2% of the variation in the light data is due to the connection with the predictor variables.

(v) **Note that in the exam, the students did not have access to software so they could not perform any additional analysis. Given the results of the tests performed, what extra analysis, if any, would you undertake?**
I would re-estimate the model only with radiation and altitude as predictor variables and force the y-intercept to zero.

Regression Analysis: Light versus Radiation, Altitude, Azimuth

```
The regression equation is
Light = 2.34 + 36.3 Radiation + 2.38 Altitude - 0.156 Azimuth

Predictor        Coef     SE Coef         T        P
Constant        2.343       1.678      1.40    0.166
Radiation       36.26       13.80      2.63    0.010
Altitude       2.3826      0.4006      5.95    0.000
Azimuth      -0.15585     0.09055     -1.72    0.088

S = 4.931       R-Sq = 99.2%     R-Sq(adj) = 99.2%

Analysis of Variance

Source            DF          SS          MS        F       P
Regression         3      295107       98369  4045.71   0.000
Residual Error    96        2334          24
Total             99      297441

Source       DF      Seq SS
Radiation     1      294127
Altitude      1         908
Azimuth       1          72
```

(b) Find below the descriptive statistics and histograms for the winter daily total solar radiation and summer daily total radiation (in $MJ/m.^2$) for Adelaide airport, as well as the graph of the empirical distribution of the summer daily totals.

(i) What statistical measures would you use to describe the centre and spread of the data in each case? Give reasons.

For the winter solar radiation, it is close to symmetric and therefore I would use the mean and standard deviation as indicative of the centre and spread of the data. On the other hand, the summer solar radiation is highly skewed to the left so I would use the median and interquartile range.

(ii) If X denotes the daily winter solar total, determine the value x such that $P(X < x) = 0.05$. **Hint**: From the answers in the previous question, you should be able to make an assumption about the distribution of this variable.

We work on the assumption that the winter solar radiation is Normally distributed and thus we want to know x such that $P(X < x) = 0.05 = P(Z < \frac{x-9.62}{3.46}) = 0.05$. From the Normal distribution table, $P(Z < -1.645) = 0.05$. Therefore, $\frac{x-9.62}{3.46} = -1.645$, and then $x = 3.93 Mj./sq.m.$.

(iii) If Y denotes the summer daily total solar radiation, given the graph of the empirical cumulative distribution function (CDF) of the summer daily totals, determine y such that $P(Y < y) = 0.05$.

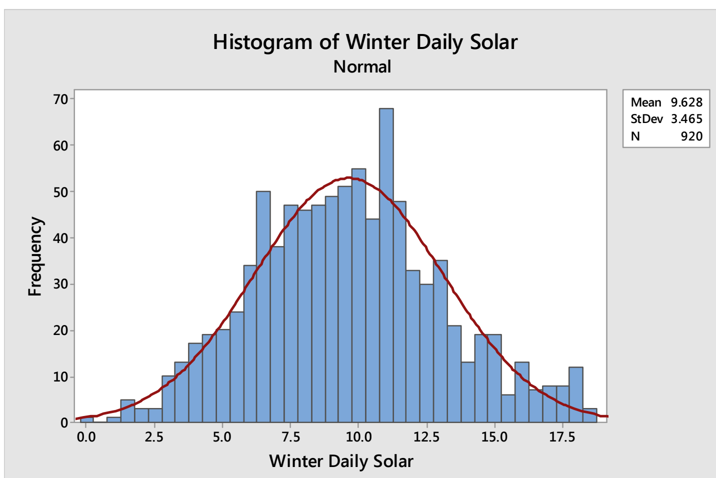
Using the empirical CDF and looking across from 5% on the y-axis, the cumulative probability, we see that the corresponding solar radiation value on the x-axis is $10 Mj./sq.m.$ approximately.

Descriptive Statistics: Winter Daily Solar

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Daily Solar	920	9.62	3.46	0.071	7.13	9.53	11.72	18.74

Descriptive Statistics: Summer Daily Solar

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Daily Solar	900	26.93	7.05	0.32	24.11	28.92	32.15	35.41



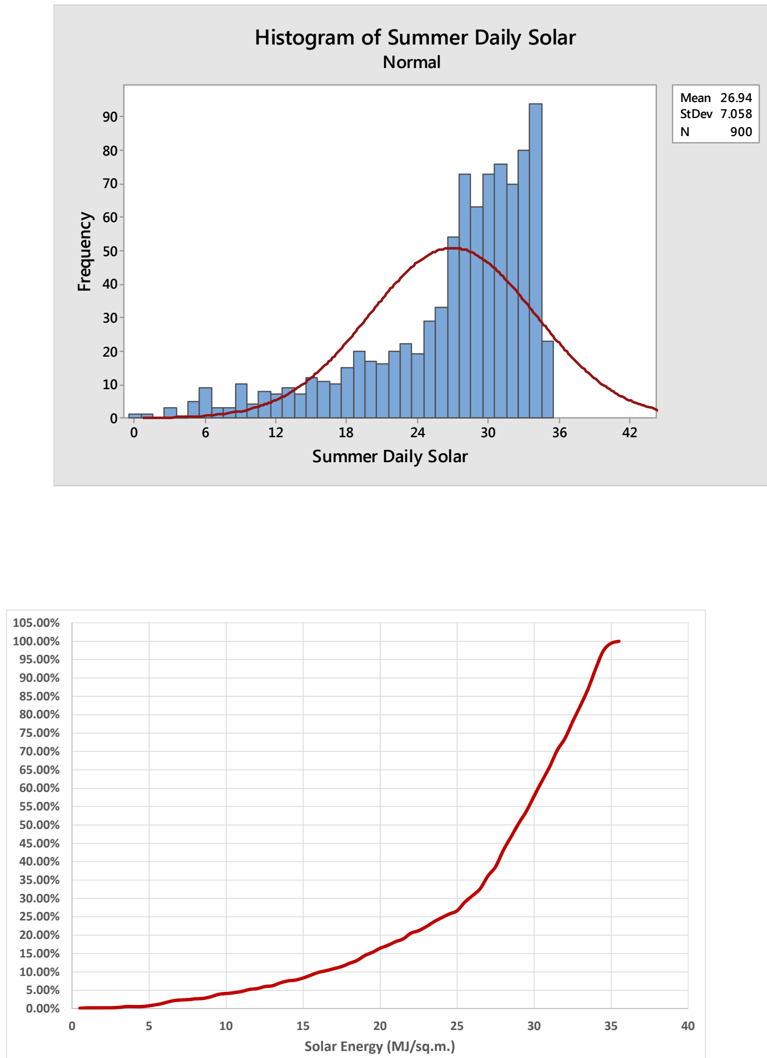


Figure 1: Summer Solar CDF

(c) Define homoscedasticity and heteroscedasticity and tell what significance they have for linear regression.

In statistics, a sequence (or a vector) of random variables is homoscedastic if all its random variables have the same finite variance. This is also known as homogeneity of variance. The complementary notion is called heteroscedasticity. In linear regression, it is assumed that for every value of the predictor variable, there is a random variable with constant variance and distribution. In other words, we assume homoscedasticity of the response variable across the predictor variable. If, on the other hand, the response variable is heteroscedastic, ie. changing variance, we have to deal with that eventuality. This may entail either transformations or even bootstrapping.

2. This part is not only for your benefit, but also mine. It will help me ensure that we can cover topics that are of use to you. What do you want to get out of this course for your future study or career? Please be specific and not give a general answer. What tools do you expect to gain knowledge of and how do you think that will help you in future?

Please email John Boland at john.boland@unisa.edu.au with your response to this question