1. (a)  Use Table A to find the proportion of observations from a standard normal distribution that satisfies each of the following statements.  In each case, sketch a standard normal curve and shade the area under the curve that corresponds to the computed proportion.

(i)  $Z \leq 1.96$

(ii) $Z \geq -0.90$

(iii)    $0.90 \leq Z \leq 1.96$

(iv) $0 \leq Z \leq 1.96$

(v) $-1.96 \leq Z \leq -1.65$

(vi) $-2.52 \leq Z \leq 0$

(vii)   $Z \geq 4.90$

(viii)  $Z \leq -5.30$

(b)  Use Table A of your textbook to find the value of z from the standard normal distribution, i.e., the normal distribution with mean $\mu = 0$, and standard deviation $\sigma = 1$, that satisfy each of

the following conditions.  Use the value of z from Table A that comes closest to satisfying the condition.  <u>In each case, sketch a standard normal curve with your value of z marked on the axis.</u>

(i)  The point z with 90% of the observations falling below it

(ii) The point z with 0.3% of the observations falling below it

(iii)The point z with 20% of the observations falling above it

(iv)The point z with  80% of the observations falling above it

(v) The point z with 60% of the observations falling above it.

(v)   Find the quartiles and the IQR  for  the standard normal distribution.

(vi)   The point z with 60% of the observations falling  below it.

(vii) Find the 30<sup>th</sup> percentile and the 70<sup>th</sup> percentile of the standad normal distribution.

2. A university professor keeps records of his travel time while he is driving between his home and the university. Over a long period of time, he has found that his morning travel times are approximately normally distributed with a mean of 31 minutes and a standard deviation of 3 minutes. Use the 68-95-99.7 rule to answer the following questions.

(i) What percent of his morning journeys are longer than 37 minutes?

(ii) Between what values do the middle 95% of his morning journey times fall?

(iii) What percent of his morning journeys are between 31 and 34 minutes?

Solution:

Let X be the random variable that corresponds to the morning travel times of the university professor in question. We will assume that X ~ N(31,3) for this problem.

(a)     What percent of morning journeys are longer than 37 minutes?

37 minutes corresponds to 2 standard deviations above the mean. From the empirical rule, we know that 95% of the observations fall between 25 and 37 minutes, so 5% of his trips are either faster than 25 minutes or longer than 37 minutes. Since the normal distribution is symmetric, the probability of each event is half of this 5%; hence, 2.5% of his morning journeys are longer than 37 minutes.
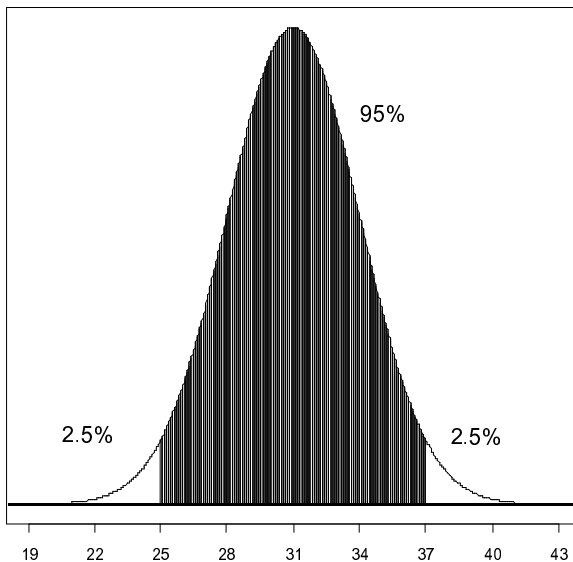
(b)    Between what values do the middle 95% of his morning journey times fall?

Using the figure from the previous page, it follows that the middle 95% of journey times fall between 25 and 37 minutes; i.e., within + or − 2 standard deviations of the mean.

(c)    What percent of his morning journeys are between 31 and 34 minutes?

34 minutes is 1 standard deviation above the mean. The rule tells us that 68% of the travel times are within one standard deviation of the mean; i.e., between 28 and 34 minutes. In other words, $P(28 < X < 34) = 0.68$, approximately. By the symmetry of the normal distribution, it follows that half of this probability corresponds to the event $31 < X < 34$. Therefore, $P(31 < X < 34) = 0.34$, at least approximately. (If this isn't clear, draw a picture!)

3. Text book, page 63, exercise 1.60 (a) to (c). Also answer the following questions:

(d) What percent of people aged between 20 and 34 have IQ scores between 95 and 120?

(e) What percent of people aged 20 and 34 have IQ scores less than 105

4. Exercise 1.68, page 65. Answer parts a, b, and c from the text. Also answer the following questions:

(d) How long do the shortest 25% of pregnancies last?

(e) How long do the longest 5% of pregnancies last?

(f) Find the proportion of pregnancies that last within the intervals (250, 282), (234, 298) and (218, 314) respectively

5. **<u>NOTE:  Do this problem using JMPIN</u>**

The following data came from an investigation of the completion times experienced by users of a multiuser computer system at various levels of total workload for the system.  The variables measured are TIME PER TASK in seconds and the NUMBER OF TERMINALS running the task.  For a given data point all the terminals initiated the same task.  Given the data,

| Number of Terminals | Time per Task |
|---|---|
| 40 | 9.9 |
| 50 | 17.8 |
| 60 | 18.4 |
| 45 | 16.5 |
| 40 | 11.9 |
| 10 | 5.5 |
| 30 | 11.0 |
| 20 | 8.1 |

| 50 | 15.1 |
|---|---|
| 30 | 13.3 |
| 65 | 21.8 |
| 40 | 13.8 |
| 65 | 18.6 |
| 65 | 19.8 |

(i)  Identify the explanatory variable and response variable.

(ii) Make a scatter plot and comment on the association of the two variables.

(iii)Find coefficient of linear correlation r. Compare your result with your findings in part (ii).

(iv)Fit a least-squares regression line.  Interpret the slope of the regression line 'b' in the context of this situation.

(v) Predict the time per task when the number of terminals used is 55.

(vi) Use the least-squares regression line obtained in (iv) to predict the average time per task when the number of terminals used is 100. Is this result reasonable? Why or why not?

Solution:

This problem concerns completion times experienced by users of a multiuser computer system at various levels of total workload for the system. The variables in the data set are time per task (sec.) and the number of terminals in use.

(a) Identify the explanatory variable (X) and the response variable (Y).

The explanatory variable in this problem is the number of terminals in use by the system, while the response variable is the time per task. Thus,

we expect that the time required to complete a task depends upon (or is related to) the number of terminals in use by the system. In particular, we should anticipate that adding more terminals to the system workload will slow down performance, i.e., increase the (average) time per task.
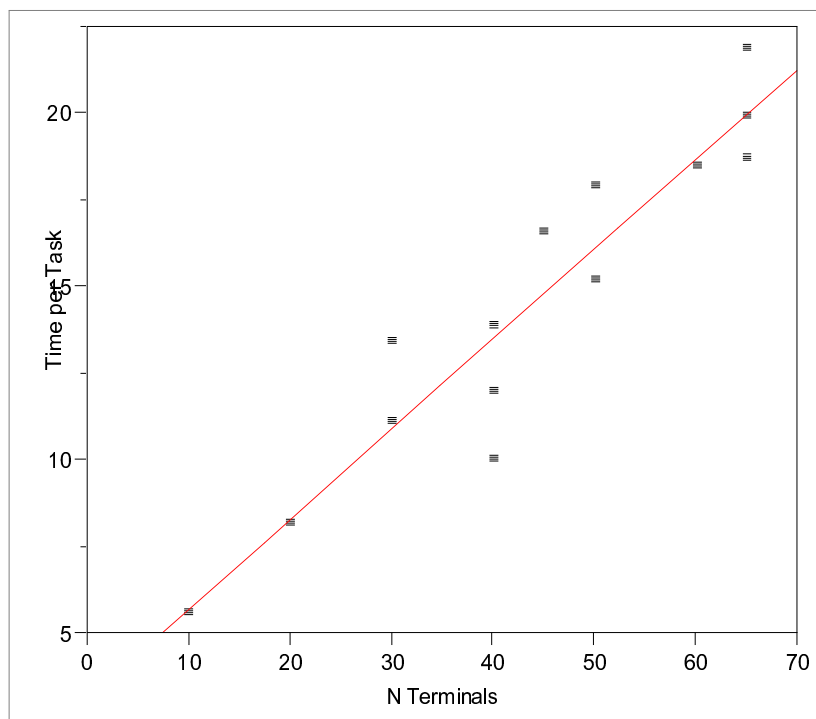


Linear Fit

## Linear Fit

Time per Task = 3.0496262 + 0.2603364 N Terminals

## Summary of Fit

| | |
|---|---|
| RSquare | 0.88 |
| | 5313 |
| RSquare Adj | 0.87 |
| | 5755 |
| Root Mean | 1.67 |
| Square Error | 2117 |
| Mean of | 14.3 |
| Response | 9286 |
| Observations | 14 |
| (or Sum | |
| Wgts) | |

(b)   Comment on the association between the two variables.

The scatter plot is given above. It shows that there is a strong positive association between

the number of terminals in use and the time per task, which closely follows a linear trend.

(c) Find the coefficient of linear correlation r. Compare with part (b).

Since the association is positive, we can find r by taking the square root of the coefficient of determination $R^2$. Thus, r = sqrt(0.88531) = 0.941, which suggests that there is a strong positive correlation between the number of terminals in use and the time per task. The difference between correlation and regression is that in correlation, we do not assign roles to the variables, but in regression we do – i.e., one identifies an explanatory variable and a response.

(d)   Fit a least squares regression line. Interpret the slope $b$ in this situation.

From the output above, the fitted least squares line is

$$Yhat = 3.05 + 0.26\ X.$$

The slope is interpretable as the average increase in processing time per task for each additional terminal in use.

(e)   Predict the time per task when the number of terminals in use is 55.

The predicted time per task with 55 terminals in use is $3.05 + 0.26*55 =$ **17.35** seconds.

(f)   Use the least squares regression line to predict the average time per task when the number of terminals in use is 100. Is this result reasonable?  Why or why not?

The predicted time per task with 100 terminals is 3.05 + 0.26*100 = **29.05** seconds.  This prediction is made outside the range of observed X values (the largest number of terminals in the data set was 65), so this is an instance of extrapolation. Unlike the prediction in (e), which was made at an X value within the observable range, this prediction is not a safe one, and so is 'unreasonable' in this context.