Solution week 1

1. Let $Z \sim \mathcal{N}(0, 1)$ and $Z_1 \sim \mathcal{N}(1, 1)$. Use tables to

   (a) find the following probabilities:
   
   \begin{align*}
   &P(Z \leq 1.25); \quad P(0.5 < Z_1 \leq 1.23); \\
   &P(t_3 \leq 2.353); \quad P(|t_5| \leq 1.476).
   \end{align*}

   (b) find $c$ such that
   
   \begin{align*}
   &P(t_8 > c) = 0.05; \quad P(|t_7| \geq c) = 0.50; \\
   &P(Z < c) = 0.05; \quad P(|Z| \geq c) = 0.05.
   \end{align*}

Solution. Recall the following facts:

\begin{align*}
&P(Z \leq x) = \Phi(x); \quad \Phi(-x) = 1 - \Phi(x), \text{ for } x \geq 0; \\
&P(y \leq Z_1 \leq x) = \Phi(x - 1) - \Phi(y - 1), \text{ for any } x > y; \\
&P(t_v \leq -x) = P(t_v \geq x); \quad P(|t_v| \leq x) = 2P(t_v \leq x) - 1.
\end{align*}

   (a)

\begin{align*}
&P(Z \leq 1.25) = 0.8943; \quad P(0.5 < Z_1 \leq 1.23) = \Phi(0.23) - \Phi(-0.5) = 0.2825; \\
&P(t_3 \leq 2.353) = 1 - .05 = 0.95; \quad P(|t_5| \leq 1.476) = 1 - 2 \times .1 = 0.8.
\end{align*}

   (b) $1.86, 0.711, -1.65, 1.96$

2. A chemical process has produced, on the average, 800 tons of chemical per day. The daily yields for the past week are 785, 805, 790, 793, and 802 tons. We wish to determine whether the data indicate that the average yield is less than 800 tons and hence that something is wrong with the process.

   (i) State the null and alternative hypothesis.

   (ii) What assumptions must be satisfied in order to process your testing?

   (iii) Find the test statistic and calculate the $p$-value.

   (iv) State your conclusions.
Solution.

(i) The hypothesis to be tested is

\[ H : \mu = 800 \quad \text{vs} \quad H_A : \mu < 800. \]

(ii) There are only five measurements on which to base the test. Let \( X_j, 1 \leq j \leq 5 \) denote the random variables that generate the observed values 785, 805, 790, 793, and 802. We have to assume that \( X_1, X_2, \ldots, X_5 \) are iid normal random variables.

(iii) A \( t \)-test must be employed. The test statistic is

\[ T_n = \sqrt{5}(\bar{X} - 800)/S. \]

The calculation of the \( p \)-value is as follows:

\[ \bar{x} = \frac{785 + 805 + 790 + 793 + 802}{5} = 795; \quad s^2 = \frac{\sum x_j^2 - n\bar{x}^2}{n - 1} = 69.5 \]

\[ t = \frac{\sqrt{5}(\bar{x} - 800)}{s} = -1.341 \]

\[ p = P_H(T_n \leq t) = P(t_4 \leq -1.341) = 0.1255026. \]

(vi) The data is consistent with the null hypothesis, that is, there is no indication that the average yield is less than 800 tons.

3. A USA supreme court case in 1965 concerned a panel of 100 potential jurors, of whom 8 were black. The panel was selected from a population of whom about 26% were black. Do you think the selection is random? (State formally the hypothesis being tested here and the evidence against it in terms of a \( p \)-value)

Solution:

(a) Let \( \theta \) be the proportion of black jurors selected in a panel. If the selection is random, then the \( \theta \) should be 0.26. The hypothesis to be tested is

\[ H : \theta = 0.26 \quad \text{vs} \quad H_A : \theta \neq 0.26. \]

Let \( Y \) be the number of black jurors in the panel. Then \( Y \sim B(100, \theta) \). \( n \) is large. We can use the test statistic:

\[ T_n = \frac{Y - 100 \times 0.26}{\sqrt{100 \times 0.26 \times 0.74}} \]

The observed value is

\[ t = \frac{8 - 100 \times 0.26}{\sqrt{100 \times 0.26 \times 0.74}} = -4.412613, \]

and hence the \( p \)-value

\[ p \approx 2(1 - \Phi(|t|)) = 1.021319e-05. \]

There is very strong evidence that the selection is not random.
Computer Exercises solution week 1

Note: pnorm, pt give distribution functions; qnorm, qt give quantiles. For more info on how to use an Splus function type ?name at the prompt, where name is the name of the corresponding function e.g. ?qnorm

1. Let $Z \sim \mathcal{N}(0,1)$ and $Z_1 \sim \mathcal{N}(1,1)$. Use Splus to

(a) find the following probabilities:

\[
\begin{align*}
    & P(Z \leq 1.25); \quad P(0.5 < Z_1 \leq 1.23); \\
    & P(t_3 \leq 2.353); \quad P(|t_5| \leq 1.476).
\end{align*}
\]

(b) find $c$ such that

\[
\begin{align*}
    & P(t_8 > c) = 0.05; \quad P(|t_7| \geq c) = 0.50; \\
    & P(Z < c) = 0.05; \quad P(|Z| \geq c) = 0.05.
\end{align*}
\]

Solution.

(a) In Splus

\[
\begin{align*}
    & > \text{pnorm}(1.25); \quad > \text{pnorm}(0.23)-\text{pnorm}(-0.5); \\
    & > \text{pt}(2.353,3); \quad > 2*\text{pt}(1.476,5)-1.
\end{align*}
\]

(b) In Splus

\[
\begin{align*}
    & > \text{qt}(0.95,8); \quad > \text{qt}(0.75,7); \\
    & > \text{qnorm}(0.05); \quad > \text{qnorm}(0.975).
\end{align*}
\]

2. The Splus dataset amp are the results of measurements made using two methods on 15 pairs of tablets to determine the dosage of aspicillin. Analyze the data to determine if there is a systematic difference between two methods.

(a) Type amp to have a look the dataset.

(b) State the null and alternative hypothesis.

(c) Create two vectors method.A, method.B whose entries correspond to two columns of the data.

(d) Calculate the mean difference $\bar{d}$ and its standard deviation $s_d$.

(e) Calculate the observed value of paired $t$-test statistic (use length(method.A) to get the sample size $n$) and then the $p$-value.

(f) Use Splus built in command t.test to confirm the $p$-value.

(g) Conclude your findings, and explain in words what assumptions must be satisfied in order for the procedure you used to analyse the data to be valid.
Solution.

In Splus

# (a)
> amp

# (b)
> # H: \mu_d=0 \ vs \ \mu_d \ not = 0

# (c)
> method.A<-amp[,1]
> method.B<-amp[,2]

# (d)
> bard<-mean(d)
> sd<-sqrt(var(d))

# (e)
> n<-length(method.A)
> t<-sqrt(n)*abs(bard)/sd
> p<-2-2*pt(t,n-1)
> p
[1] 0.718379

# (f)
> t.test(method.A, method.B, alternative="two.sided",mu=0, paired=T)

# (g) The data is consistent with the null hypothesis.
# Let X_j and Y_j denote the random variables that generate the observed
# values of method.A and method.B respectively. Because the sample size
# (15) is less than 20, in order for the procedure you used to analyze
# this data to be valid, you have to assume that X_j-Y_j are iid normal
# random variables.