UNIVERSITY OF NEW SOUTH WALES
DEPARTMENT OF STATISTICS

MATH5856 Introduction to Statistics and Statistical Computations

Tutorial Problems week 11 (CI and HT)

1. Based on the observed value of a single random variable X with probability function $f(x) = P(X = x)$ and possible values $1, 2, \ldots, 6$, it is desired to test the hypothesis $H_0 : f = f_0$ versus $H_1 : f = f_1$ where f_0 and f_1 are as follows:

x	1	2	3	4	5	6
$f_0(x)$	0.03	0.02	0.50	0.40	0.03	0.02
$f_1(x)$	0.20	0.10	0.05	0.05	0.40	0.20

   (a) Calculate the size and power of the test with rejection region $x \leq 2$.

   (b) Find the rejection region and the power of the most powerful size 0.05 test of H_0 versus H_1.

2. (a) Let $X_1, \ldots, X_n$ be a random sample from the population having density function

$$f_X(x; \theta) = \frac{1}{2\theta} e^{-|x|/\theta}.$$

   We wish to test

$$H_0 : \theta = 1 \quad \text{versus} \quad H_1 : \theta = 4.$$

   Show that the most powerful test rejects H_0 if and only if

$$\frac{1}{n} \sum_{i=1}^{n} |X_i| \geq c$$

   for some constant $c > 0$.

(b) In the case $n = 1$ find an expression for c in terms of α, the significance level of the test.

3. Let $X_1, \ldots, X_n$ be a random sample from the density

$$f_X(x) = e^{-(x-\theta)} \ , \ x > \theta; \ \theta > 0.$$

The parameter θ is unknown. Let $Y = \min(X_1, \ldots, X_n)$. Then Y is a statistic with density function

$$f_Y(y; \theta) = ne^{-n(y-\theta)} \ , \ y > \theta.$$

(a) Show that $W = e^{(Y-\theta)}$ is a pivotal function, i.e., the density function of W does not depend on any unknown parameters.

(b) Find a constant $b > 0$ such that $P(W < b) = 0.95$.

(c) Calculate a 95% lower confidence bound for θ when $n = 10$ and Y is observed to be 5.

4. The article 'The influence of Corrosion Inhibitor and Surface Abrasion on the Failure of Aluminium-Wired Twist-on Connections' (IEEE Trans. Components, Hybrids and Manuf. Tech., 1984, pp. 20-25) reported the following summary data on potential drop measurements for a sample of connectors wired with alloy aluminium and for another sample wired with EC aluminium:

type	sample size	sample mean	sample st. dev.
Alloy	21	17.5	0.55
EC	21	16.9	0.49

Assuming normality, calculate

(a) a 95% confidence interval for the difference of the means of potential drops, assuming common variance,

(b) a 99% confidence interval for the ratio of the standard deviations of the potential drops.

5. Toxaphene is an insecticide which has been identified as a pollutant. To investigate the effect of toxaphene exposure on animals, a group of rats were given a dose of toxaphene in their diet. The article 'Reproduction Study of Toxaphene in the Rat' in the 1988 issue of the Journal of Environmental Science and Health reported weight gains (in grams) for rats given toxaphene and for control rats whose diet did not include the insecticide. The sample standard deviation of the weight gains for 25 control rats was 32 grams and for 21 exposed rats the sample standard deviation was 54 grams. Do these data suggest there is more variability in weight gain among exposed rats than among control rats?

   Answer this question by setting up appropriate hypotheses, carrying out a statistical test and reaching a conclusion from the results.

6. For a random sample of 25 high-strength magnetic alloy steel rings used in turbine generators the sample mean ultimate tensile strength was 152.7 ksi (kips per square inch), while the sample standard deviation was 4.6 ksi. Assuming normality, calculate

   (a) a 95% confidence interval for the mean ultimate tensile strength,

   (b) a 99% confidence interval for the variance of the ultimate tensile strength.

7. The article 'The Effect of Doping on the Voltage Holdoff Performance of Alumina Insulators in Vacuum' in IEEE Transactions, 1985, reported a random sample of 15 alumina ceramic insulators doped in a certain

manner yielded a sample mean holdoff voltage (in kV) of 110 and sample standard deviation of 24. A sample of 76 plain alumina ceramic insulators gave a sample mean of 101 and sample standard deviation of 22. Assuming normality and common variance, carry out a statistical test of the hypothesis that the mean voltages for doped and plain insulators are the same.

8. The College and Personnel Association of the U.S. lists the median salaries for administrators in private and public universities in 1985 as shown in the table below.

	Public	Private
	52.5	52.0
	51.9	43.8
	34.8	28.0
	28.6	24.9
	33.3	25.0
	33.1	30.3
	38.4	31.5
	30.5	24.0
	51.2	50.0
	58.2	64.8
	49.0	48.0
	76.7	78.7
	70.4	78.5
sample mean	46.8	44.6
sample variance	237	389

Assuming normality and independence of the samples, compute a 90% confidence interval for

$$\frac{\text{variance of private university salaries}}{\text{variance of public university salaries}}.$$

9. The glucose levels (in mg/100 ml) are measured at two hours of life for five "normal" infants and ten infants small-for-gestational-age (SGA). The data are as follows:

Normal Infants	SGA Infants
66	53
57	58
52	62
47	47
64	48
$\overline{x} = 57.2$	31
$s_1^2 = 63.7$	41
	30
	37
	28
	$\overline{y} = 43.5$
	$s_2^2 = 144.72$

(a) Assuming the glucose levels to be normally distributed, find a 95% confidence interval for the mean glucose level of SGA infants.

(b) Assuming the glucose levels to be normally distributed and that the variances are equal, test at level 0.05 the null hypothesis that there is no difference in the mean glucose levels of the two groups of infants versus the alternative hypothesis that the mean glucose level is lower for SGA infants.

10. The speeds (km/h) of 23 cars travelling along Anzac Parade are recorded, leading to the following set of data:

$$57, 65, 65, 55, 58, 57, 71, 60, 57, 65, 62, 60, 54, 70, 57, 52, 49, 65, 68, 60, 65, 68, 44$$

The sample mean and standard deviation are, respectively, 60.174 km/h and 6.807 km/h. Graphical inspection of the data shows that the normality assumption can be reasonably assumed.

(a) Calculate a 90% confidence interval for the mean speed of cars travelling along Anzac Parade.

(b) Calculate a 95% confidence interval for the corresponding standard deviation.

11. For a random sample of 25 high-strength magnetic alloy steel rings used in turbine generators the sample mean ultimate tensile strength was 152.7 ksi (kips per square inch), while the sample standard deviation was 4.6 ksi. Assuming normality, obtain a 99% confidence interval for the variance of the ultimate tensile strength.

R excercises

1. Generate a random sample of 100 Normal random samples from $N(\mu = 2, \sigma = 4)$ distribution, use the *rnorm()* command and assign the values to a vector x

   (a) Using the *plot()* command to make a scatterplot of the values in x

   (b) You can control the labelling on the $x-$ and $y-$ axis by using the *xlab="", ylab=" "* options inside the *plot()* function, now try changing the xlabel to a by adding *xlab="a"*, similarly, change the ylabel to b.

   (c) You can control the limits on the $x-$ and $y-$ axes by using the *xlim=c(x1,x2), ylim=c(y1,y2)* options inside the plot command, where $x1, x2, y1, y2$ are the lower and upper limits of the desired x and y values. Now change the $y-$ limits to be between -10 and 15, and the $x-$ limits to be between 10 and 80.

   (d) The option *main* in the plot command allows you to add a title to your graph, now add a title using the *main* option.

   (e) The *lines()* command allows you to add lines to the current plot, try *lines(x)* command after you have already made the scatterplot of x

   (f) There are many options in plotting graphs in R, *help(par)* gives more details on the various things you can do. For example, you can change the color of the lines that you add by typing *lines(x, col=3)*, change the width of the lines by adding *lines(x, col=3, lwd=5)*. You can also add margins using the *text()* command, try typing *text(10,10, "margins")* on a new line.

2. The ocean swell produces spectacular eruptions of water through a hole in the cliff at Kiama Blowhole, about 120km south of Sydney. Jim Irish (Faculty of Engineering, University of Technology, Sydney) observed the duration of 65 successive eruptions of the blowhole starting at 1.40pm on 12 July, 1998. The first twenty of these observations are shown in the table below:

   (a) Enter the durations in R as a vector `kiama` using the concatenate function `c()`.

Observation number	Duration	Observation number	Duration
1	83	11	18
2	51	12	16
3	87	13	29
4	60	14	54
5	28	15	91
6	95	16	8
7	8	17	17
8	27	18	55
9	15	19	10
10	10	20	35

(b) Produce a summary of the data by typing `summary(kiama)`.

(c) Produce a histogram of the data by typing `hist(kiama)`. Read the R help to read more about options to control the appearance of the histogram (type `help(hist)`).

(d) Produce a boxplot of the data by typing `boxplot(kiama)`. Read the R help to learn more about options to control the appearance of the boxplot (type `help(boxplot)`).

3. Write an R function, called `mymax`, which takes input values (a, b, c), and returns the maximum of the three values.