MAST90105: Lab and Workshop Problems for Week 12

The Lab and Workshop this week covers problems arising from Module 8.2 to the end of the course.

1 Lab

1. A company that manufactures brackets for an auto maker regularly selects brackets from the production line and performs a torque test. The goal is for the mean torque to equal 125. Let $X \sim N(\mu, \sigma^2)$ be the torque and suppose we take a random sample of size $n = 15$ to test $H_0 : \mu = 125$ against a two sided alternative. Suppose the following data are observed:

128	149	136	114	126	142	124	136
122	118	122	129	118	122	129	

   Use the t.test command to test the hypotheses and construct a 95% confidence interval for μ.

2. Let X be the weight in grams of a Low Fat Strawberry Kudo and Y ther weight of a Low Fat Blueberry Kudo. Assume $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. A random sample of 9 observations on X yielded
   21.7, 21.0, 21.2, 20.7, 20.4, 21.9, 20.2, 21.6, 20.6
   and a random sample of 13 observations on Y yielded
   21.5, 20.5, 20.3, 21.6, 21.7, 21.3, 23.0, 21.3, 18.9, 20.0, 20.4, 20.8, 20.3

   a. Test $H_0 : \mu_X = \mu_Y$ against a two-sided alternative. You may choose the significance level.

   b. Construct a box plot to support your conclusions.

3. Let $X_1, \ldots, X_{12}$ be a random sample from an exponential distribution with mean θ. Then $f(x;\theta) = \theta^{-1} e^{-x/\theta}, \quad 0 < x < \infty$. We shall test the null hypothesis $H_0 : \theta = 9$ against the alternative $H_1 : \theta \neq 9$.

   a. Recall that the moment generating function (m.g.f) of an exponential random variable with mean θ is $M(t) = (1 - \theta t)^{-1}$ and that of a $\chi^2(r)$ random variable is $M(t) = (1 - 2t)^{-r/2}$. Find the moment generating function of $(2/\theta) \sum_{=1}^{12} X_i$ and hence deduce that $(2/\theta) \sum_{i=1}^{12} X_i \sim \chi^2(24)$.

   b. Find constants a and b so that, under H_0,

$$P\left[\frac{2}{9}\sum_{i-1}^{12} X_i \leq a\right] = 0.05, \text{ and } P\left[\frac{2}{9}\sum_{i-1}^{12} X_i \geq b\right] = 0.05$$

c. Show that a critical region of size $\alpha = 0.10$ is

$$C = \left\{ (x_1, \ldots, x_{12}) : \sum_{i=1}^{12} x_i \leq \frac{9a}{2} \text{ or } \sum_{i=1}^{12} x_i \geq \frac{9b}{2} \right\}.$$

d. Define the power function for this test.

e. For $\theta = 1, 3, 5, \ldots, 19$, generate 100 random samples of size 12 from an exponential distribution with mean θ. For each θ determine the number of times H_0 was rejected.

f. Show that the theoretical and empirical power functions are the same by plotting them on the same figure.

4. Let X be the weight (in grams) of a grape flavoured jolly rancher. Denote the median of X by m. We shall test $H_0 : m = 5.900$ against $H_1 : m > 5.900$. A random sample of size $n = 25$ yielded the data:
5.625, 5.665, 5.697, 5.837, 5.863, 5.870, 5.878, 5.884, 5.908, 5.967, 6.019, 6.020, 6.029, 6.032, 6.037, 6.045, 6.049, 6.050, 6.079, 6.116, 6.159, 6.186, 6.199, 6.307, 6.387 Test the hypotheses using

   a. The sign test.

   b. The Wilcoxon test. (wilcox.test)

   c. The t test.

   d. Construct a box plot of the data.

   e. Compare your results and discuss.

5. Let X equal the amount of butterfat in pounds produced by 90 cows during a 305-day milk production period following their first calf. Test the hypothesis that the distribution of X is $N(\mu, \sigma^2)$ using the following data:
486 537 513 583 453 510 570 500 458 555 618 327 350 643 500 497 421 505 637 599 392 574 492 635 460 696 593 422 499 524 539 339 472 427 532 470 417 437 388 481 537 489 418 434 466 464 544 475 608 444 573 611 586 613 645 540 494 532 691 478 513 583 457 612 628 516 452 501 453 643 541 439 627 619 617 394 607 502 395 470 531 526 496 561 491 380 345 274 672 509 The data is available as `Butterfat.xls` in theLMS and in the MAST90105 folder in the Lab.

   a. Compute the sample mean $\bar{x}$ and standard deviation s_x.

   b. Use the commands

```
b <- c(0, seq(374, 624, 50), 1000)
(T <- table(cut(X, breaks = b)))
O <- as.numeric(T)
```

   to compute observed frequencies in the given cells.

c. Compute expected frequencies using

```
p <- rep(0, 7)
p[1] <- pnorm(b[2], mu, s) - pnorm(b[1], mu, s)
# ...
p[7] <- pnorm(b[8], mu, s) - pnorm(b[8], mu, s)
```

To speed this up, you can use

```
for (k in 1:7) {
    p[k] = pnorm(b[k + 1], mu, s) - pnorm(b[k], mu,
        s)
}
E <- p * length(X)
C <- sum((O - E)^2/E)
d1 <- length(T)
qchisq(0.95, d1 - 3)
1 - pchisq(C, d1 - 3)
1 - pchisq(C, d1 - 3)  #p-value
cbind(O, E)
```

d. If you conduct the chisquare test using

```
chisq.test(O, p = p)
```

remember to adjust your degrees of freedom.

e. Use the `qqnorm` command to illustrate your result.

6. The Mathematica notebook, Lab12.nb, is available on the LMS and also on the server. It contains instructions on exploring the distribution of the Wilcoxon signed rank statistic. At the end of working through it you will have computed the exact p-value that was obtained in lectures, seen how to get the distribution symbolically (and thus exactly) as well as computing and plotting the distribution, including getting an expression for the moment generating function as the sample size varies.

7. Develop a function to simulate the distribution of the Wilcoxon two sample statistic based on sample sizes n, m by drawing random samples, using the R command `sample`.

```
f <- function(x) {
    sum(sample(x[1] + x[2], size = x[2]))
}
W <- function(x, r) {
    t <- matrix(rep(x, r), byrow = TRUE, nrow = r,
        ncol = 2)
    apply(t, 1, f)
```

```
}
# This can be tested out on the sample sizes of 8
# and 8, used in the cinammon packet filling
# example as follows.  The code produces 10,000
# values of the W statistic It checks the mean and
# variance of the simulated samples It then finds
# the emprical probability that W is at least 87
w <- W(c(8, 8), 10000)
# compare to theoretical values
c(mean(w), 8 * 17/2)

## [1] 67.9886 68.0000

c(sd(w), sqrt(8 * 8 * 17/12))

## [1] 9.423638 9.521905

# empirical p-value compared to normal
# approximation
c(sum(w >= 87)/10000, 1 - pnorm((87 - 4 * 17)/sqrt(64 *
    17/12)))

## [1] 0.02370000 0.02299968

# try again with a larger number of repititions
w <- W(c(8, 8), 1e+06)
# compare to theoretical values
c(mean(w), 8 * 17/2)

## [1] 67.9935 68.0000

c(sd(w), sqrt(8 * 8 * 17/12))

## [1] 9.522140 9.521905

# empirical p-value
c(sum(w >= 87)/1e+06, 1 - pnorm((87 - 4 * 17)/sqrt(64 *
    17/12)))

## [1] 0.02503100 0.02299968
```

2 Workshop

8. Vitamin B_6 is one of the vitamins in a multiple vitamin pill manufactured by a pharmaceutical company. The pills are produced with a mean of 50 milligrams of vitamin B_6 per pill. The company believes there is a deterioration of 1 milligram per month, so that after 3 months they expect that $\mu = 47$. A consumer group suspects that $\mu < 47$ after 3 months.

   a. Define a critical region to test $H_0 : \mu = 47$ against $H_1 : \mu < 47$ at the $\alpha = 0.05$ significance level based on a random sample of of size $n = 20$. ($t_{0.05}(19) = 1.729$, $t_{0.05}(20) = 1.724$, $t_{0.025}(19) = 2.093$, $t_{0.025}(20) = 2.086$).

   b. If the 20 pills yielded a mean of $\bar{x} = 46.94$ with standard deviation of $s = 0.15$, what is your conclusion?

   c. What is the approximate p-value of this test?

9. Let X be the forced vital capacity (FVC) in liters for a female college student. Assume that $X \sim N(\mu, \sigma^2)$ approximately. Suppose it is known that $\mu = 3.4$ litres. A volleyball coach claims the FVC of volleyball players is greater than 3.4. She plans to test this using a random sample of size $n = 9$.

   a. Define the null hypothesis.

   b. Define the alternative hypothesis.

   c. Define a critical region for which $\alpha = 0.05$. lllustrate this on a figure. ($t_{0.025}(8) = 2.306$, $t_{0.05}(8) = 1.859$, $t_{0.01}(8) = 2.896$)

   d. Calculate the value of the test statistic if $\bar{x} = 3.556$ and $s = 0.167$.

   e. What is your conclusion?

   f. What is the approximate p-value of this test?

10. Among the data collected for the World Health Organisation air quality monitoring project is a measure of suspended particles in $\mu g/m^3$. Let X and Y equal the concentration of suspended particles in the city centres of Melbourne and Houston. Using $n = 13$ observations of X and $m = 16$ observations of Y, we shall test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$.

    a. Define the test statistic and the critical region assuming the variances are equal. Let $\alpha = 0.05$

    b. If $\bar{x} = 72.9$, $s_x = 25.6$, $\bar{y} = 81.7$ and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.
    ($t_{0.025}(27) = 2.052$, $t_{0.05}(27) = 1.703$, ($t_{0.1}(27) = 1.314$, ($t_{0.25}(27) = 0.684$)

    c. Give limits for the p-value of this test.

11. It is claimed that the median weight m of certain loads of candy is 40,000 pounds.

a. Use the following data and the Wilcoxon test statistic at an approximate significance level of $\alpha = 0.05$ to test the null hypothesis $H_0 : m = 40,000$ against $H_1 : m < 40,000$.

41195, 39485, 41229, 36840, 38050, 40890, 38345, 34930, 39245, 31031, 40780, 38050, 30906

It may help to complete the following table. Ties are assigned the average rank.

	X	$X - m$	Rank	Sign
1	41195			
2	39485			
3	41229			
4	36840			
5	38050			
6	40890			
7	38345			
8	34930			
9	39245			
10	31031			
11	40780			
12	38050			
13	30906			

b. What is the approximate p-value?

```
qnorm(seq(0.9, 0.975, 0.025))

## [1] 1.281552 1.439531 1.644854 1.959964
```

c. Use the sign test to test the same hypothesis.

```r
pbinom(6:12, 13, 0.5)

## [1] 0.5000000 0.7094727 0.8665771 0.9538574
## [5] 0.9887695 0.9982910 0.9998779
```

d. Compare the results of the two tests.

12. A 1-pound bag of candy-coated chocolate covered peanuts contained 224 pieces of candy coloured brown, orange, green and yellow. Test the null hypothesis that the machine filling these bags treats the four colours of candy equally likely. That is test

$$H_0 : p_B = p_O = p_G = p_Y = \frac{1}{4}.$$

The observed values were 42 brown, 64 orange, 53 green, and 65 yellow. You may select the significance level or give an appropriate p-value.
$(\chi^2_{0.025}(3) = 9.348$, $\chi^2_{0.05}(3) = 7.815$, $\chi^2_{0.10}(3) = 6.251$.

13. In a biology laboratory the mating of two red eye fruit flies yielded $n = 432$ offspring among which 254 were red-eyed, 69 were brown-eyed, 87 were scarlet-eyed, and 22 were white-eyed. Use these data to test, with $\alpha = 0.05$, the hypothesis that the ratio among the offspring would be 9:3:3:1 respectively.
$(\chi^2_{0.025}(3) = 9.348$, $\chi^2_{0.05}(3) = 7.815$, $\chi^2_{0.10}(3) = 6.251)$.

14. We wish to determine if two groups of nurses distribute their time in six different categories about the same way. That is, the hypothesis under consideration is $H_0 : p_{i1} = p_{i2}$, $i = 1\ldots,6$. To test this, nurses are observed at random throughout several days, each observation resulting in a mark in one of the six categories. The summary data is given in the following frequency table

	Category						
	1	2	3	4	5	6	Total
Group I	95	36	71	21	45	32	300
Group II	53	26	43	18	32	28	200

Use a chi-square test with $\alpha = 0.05$.

```r
qchisq(seq(0.9, 0.975, 0.025), 5)

## [1]  9.236357 10.008315 11.070498 12.832502
```

15. A random sample of 1000 individuals from a rural area had 620 in favour of the election of a certain candidate, whilst a random sample of 1000 individuals from an urban area had 550 in favour of the same candidate. At the 5% level, test the hypothesis that area and opinion about the candidate are independent.