IMPORTANT: The answers are provided with the view to deepen your understanding of the topic and to stimulate more in depth discussions. They should not be shared with others or put on the web. The questions and answers may be revised in the next course delivery.

Answers to Tutorial 9

Qn 1

Since each I_j is a Bernoulli variable with $p = \frac{1}{2}$

$$E[I_j] = p = \frac{1}{2}$$

$$Var[I_j] = p(1-p) = \frac{1}{4}$$

Hence

$$E[T] = E\left[\sum_{j=1}^{n} jI_j\right] = \sum_{j=1}^{n} \frac{j}{2} = \frac{n(n+1)}{4}$$

$$Var(T) = Var\left(\sum_{j=1}^{n} jI_j\right) = \sum_{j=1}^{n} Var(jI_j) = \sum_{j=1}^{n} j^2 Var(I_j) = \sum_{j=1}^{n} \frac{j^2}{4} = \frac{n(n+1)(2n+1)}{24}$$

Qn 2

a) Z test: the distribution is normal and the variance is known

b) t test: the distribution is normal but the variance is unknown

c) testing of equality of means of two normal populations with known variance:

the two distributions are both normal, and they are independent of each other, and both variance are known.

d) testing of equality of means of two normal populations with unknown variance

the two distributions are both normal, and they are independent of each other, both variance are unknown but assumed to be equal to each other

e) paired t test

the two distributions are both normal with unknown variance (not necessarily equal to each other), and they are independent of each other. The two variables are generated in pairs.

f) signed test

test the hypothesis about the median of an unknown distribution. (The unknown distribution may or may not be normal)

g) signed rank test

test the hypothesis about the median of an unknown distribution, but make the additional assumption that the unknown distribution is approximately symmetrical about the median

h) u test

test whether two unknown distribution is the same. It is the two sample version of the signed test.

Qn 3

Subject	1	2	3	4	5	6	7	8
Pulse Rate before	74	86	98	102	78	84	79	70
Pulse Rate After	70	85	90	110	71	80	69	74
Difference	-4	-1	-8	8	-7	-4	-10	4

Since the data comes in pairs from the same subject, a paired test should be used.

As we assume that both the distributions of "pulse rate before" and "pulse rate after" are normal and independent of each other, a paired t-test can be used.

Since we wish to test whether a reduction in pulse rate actually occur, we test the null hypothesis that there is no reduction in pulse rate against the alternative hypothesis that the pulse rate does decrease.

H_0: $\mu_0 = 0$

H_1: $\mu_0 < 0$

$\bar{X} = -2.75$ $S = 6.158618$ $n = 8$

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{-2.75 - 0}{6.158618/\sqrt{8}} = -1.26297403$$

At level of significance $\alpha = 0.05$,

$T_{0.05,7} = 1.895$

Since

$T > -T_{0.05,7}$

the hypothesis is not rejected. There is not enough supporting evidence that jogging reduces pulse rate.

Note: the sample size is quite small. It would be more prudent to conduct a larger scale study.

Qn 4

a)

$T = 13$

p-value = $2\ P\{T \geq 13\} = 2\left(\sum_{i=13}^{18} \binom{18}{i} (0.5)^{18}\right)$

Using online binomial distribution calculator, e.g. http://stattrek.com/online-calculator/binomial.aspx

p-value = $2(0.048126220703125) \approx 0.096$

Since p-value > 0.05, we cannot reject the hypothesis, thus we cannot conclude that the medicine has an effect on blood pressure.

b)

$$Z = \frac{T - np}{\sqrt{np(1-p)}} = \frac{13 - 18(0.5)}{\sqrt{(18)(0.5)(1-0.5)}} = 1.885618083$$

p-value = $P\{|Z| > 1.89\} = 2(1 - P\{Z \leq 1.89\}) = 2(1 - 0.9706) = 0.0588$

Since p-value > 0.05, we cannot reject the hypothesis.

c)

Difference	Rank	Sign Rank of –ve values
-1	1.5	1.5
+1	1.5	
+2	3	
+4	4	
-5	6	6
-5	6	6
+5	6	
-8	8.5	8.5

+8	8.5	
-9	10	10
-11	11	11
-12	12	12
-15	13	13
-16	14	14
-18	15	15
-21	16	16
-22	17	17
-25	18	18
		T=148

$n = 18$

$$E[T] = \frac{n(n+1)}{4} = 85.5$$

$$Var[T] = \frac{n(n+1)(2n+1)}{24} = 527.25$$

$\sigma = \sqrt{527.25} = 22.961925$

$$Z = \frac{T - 85.5}{22.961925} = 2.721897228$$

p-value = $2P\{Z > 2.72\} = 2(1 - 0.9967) = 0.0066$

Since p-value < 0.05, the hypothesis is rejected, the medicine has an effect on blood pressure.

Note: b) and c) use the normal approximation. Thus the result cannot be directly compared with the result in a).

Qn 5

a)

Hypothesize that the median $m_0 = -4.5$

X_i	$X_i - m_0$
-3	1.5
-6	-1.5
5	9.5

4	8.5
-2	2.5
-1	3.5

$n = 6, \quad T = 1$

p-value = 2 $P\{Bin\ (6, 0.5) \le 1\} = 2\left(\binom{6}{0}(0.5)^6 + \binom{6}{1}(0.5)^6\right) = 0.21875$

Since p-value > 0.05, the null hypothesis is accepted.

b)

$T = 1$

$$Z = \frac{T - np}{\sqrt{np(1-p)}} = \frac{1 - 6(0.5)}{\sqrt{(6)(0.5)(1-0.5)}} = -1.632993162$$

p-value = $2P\{Z > 1.63\} = 2(1 - 0.9484) = 0.1032$

Since p-value > 0.05, the hypothesis is accepted.

c)

Hypothesize that the median $m_0 = -4.5$

X_i	$X_i - m_0$	$\lvert X_i - m_0 \rvert$	Rank	Signed Rank of –ve value
-3	1.5	1.5	1.5	
-6	-1.5	1.5	1.5	1.5
5	9.5	9.5	6	
4	8.5	8.5	5	
-2	2.5	2.5	3	
-1	3.5	3.5	4	
				T=1.5

$n = 6$

$$E[T] = \frac{n(n+1)}{4} = 10.5$$

$$Var[T] = \frac{n(n+1)(2n+1)}{24} = 22.75$$

$\sigma = \sqrt{22.75} = 4.769696007$

$$Z = \frac{T - 10.5}{4.769696007} = -1.886912706$$

p-value = $2P\{Z > 1.89\} = 2(1 - 0.9706) = 0.0588$

Since p-value > 0.05, the hypothesis is accepted.