Statistics, Assignment 2

BM Corser

April 23, 2017

1. (i) If the die is fair, and that its rolls follow a normal distribution, the population mean will be $\mu_0 = \frac{7}{2}$. We can use a one sample t-test in which our null hypothesis is that the die is fair and our alternative hypothesis is that the die is not fair. As such $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$. Our alternative hypothesis is two-sided. Our test statistic is $t = \left|\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}\right|$ and we will reject our null hypothesis H_0 at the $100\alpha\%$ significance level if the percentage point $t_{n-1}(50\alpha) < t$.

   (ii) $\bar{x} = \frac{17}{5}$, $s = \sqrt{\frac{544112}{999}}$,

$$
\begin{aligned}
t &= \left|\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}\right| \\
&= \left|\frac{\frac{17}{5}-\frac{7}{2}}{\sqrt{\frac{\frac{544112}{999}}{1000}}}\right| \\
&= 3\frac{\sqrt{\frac{555}{68014}}}{2} \\
&= 1.81668...
\end{aligned}
$$

   Using Table 10. from NCST to find $t_\nu(P)$ with $\nu = \infty$ and $P = 2.5$ (because our H_1 is two sided) as an estimate for $t_{999}(2.5)$ we find that $t_\infty(2.5) = 1.96$ which is greater than our test statistic, t, above. Therefore we don't reject H_0 at the 5% significance level, and the die is probably fair.

(iii) R input

```
scores = c(
    rep(1, times = 212),
    rep(2, times = 140),
    rep(3, times = 156),
    rep(4, times = 170),
    rep(5, times = 172),
    rep(6, times = 150)
)
t.test(scores, mu=3.5, alternative="two.sided")
```

R output

```
        One Sample t-test

data:  scores
t = -1.814, df = 999, p-value = 0.06998
alternative hypothesis: true mean is not equal to 3.5
95 percent confidence interval:
 3.291821 3.508179
sample estimates:
mean of x
      3.4
```

2. (i) Assuming the proportion of individuals having a particular characteristic follows a binomial distribution, we can test the null hypothesis $H_0 : p = 0.5$ against the one-sided alternative hypothesis $H_1 : p > 0.5$ by using the test statistic $z = \frac{\hat{p}-0.5}{\sqrt{\frac{(0.5)^2}{5008}}}$. We reject H_0 at the $100\alpha\%$ significance level if the percentage point of a normal distribution $x(100\alpha) < z$.

(ii) $\hat{p} = \frac{2524+896}{5008} = \frac{3420}{5008}$, $z = \frac{\frac{3420}{5008}-0.5}{\sqrt{\frac{0.25}{5008}}} = 24.877$, $\alpha = 0.05$. Since $x(100\alpha) = 1.6449 < z$, we reject H_0 at the 5% significance level. There is strong evidence that there is an association between eye colour of parents and children.

(iii) R input

```
prop.test(3420,5008,alternative="greater",correct=FALSE)
```

R output

```
        1-sample proportions test without continuity correction

data:  3420 out of 5008, null probability 0.5
X-squared = 670.17, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.671995 1.000000
```

```
sample estimates:
        p
0.6829073
```

3. (i) Assuming the difference follows a normal distribution, $D \sim N(\mu_D, \sigma_D^2)$ can use a paired comparison with null hypothesis $H_0 : \mu_D = 0$ and the one-sided alternative hypothesis $H_1 : \mu_D < 0$. We can use the test statistic $t = \frac{\bar{d}}{\frac{s_D}{\sqrt{n}}}$ where $n = 9$, $\bar{d} = 8.443333$, $s_D^2 = 981.8365$. Then $t = 0.80838$ and the percentage point for a t-distribution with $\nu = 9$ at 5 is $t_9(5) = 1.833$. Because $t < t_9(5)$ we do not reject H_0 at the 5% significance level and as such there is no strong evidence that the drug is effective at reducing cholesterol.

   R input

```
NoDrug = c(206.13, 203.62, 226.43, 139.81, 137.40, 131.80, 145.41, 141.64, 216.86)
Drug = c(166.81, 181.14, 211.65, 96.99, 141.41, 166.91, 101.25, 169.05, 237.90)
D = NoDrug - Drug
t.test(D)
```

   R output

```
        One Sample t-test

data:  D
t = 0.80838, df = 8, p-value = 0.4422
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -15.64232  32.52899
sample estimates:
mean of x
 8.443333
```

   (ii) Observing our variable `D`, we can see that $n = 9$ (since no difference is 0) and there are four negative differences.

   R output

```
> D
[1]  39.32  22.48  14.78  42.82  -4.01 -35.11  44.16 -27.41 -21.04
```

   We can then set up a sign test against a null hypothesis $H_0 : \eta = 0$ and one sided alternative hypothesis $H_1 : \eta < 0$

   Referring to Table 1 with $n = 9$, $r = 4$ and $p = 0.5$ we can find a p-value st. $p = \Pr(X \leq 4) = F_3 = 0.5$. Because $p > 0.1$ we do not reject H_0 at the 10% significance level.

R input

```
binom.test(sum(D < 0), sum(D != 0), alternative="less")
```

R output

```
        Exact binomial test

data:  sum(D < 0) and sum(D != 0)
number of successes = 4, number of trials = 9, p-value = 0.5
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.7486324
sample estimates:
probability of success
             0.4444444
```

(iii) $H_0 : \eta = 0$, $H_1 : \eta < 0$

d_i	signed rank
39.32	7
22.48	4
14.78	2
42.82	8
-4.01	-1
-35.11	-6
44.16	9
-27.41	-5
-21.04	-3

$T^+ = 30$, $x(P) = 8$ and since $x(P) < T^+$, we do not reject H_0 at a 5% significance level; there is some evidence that the drug is effective at reducing cholesterol.

```
wilcox.test(D)
```

R output

```
        Wilcoxon signed rank test

data:  D
V = 30, p-value = 0.4258
alternative hypothesis: true location is not equal to 0
```