

STA 100 Homework 4 Solution

Yidong Zhou

1. Here $n = 51, \bar{Y} = 355, s = 42, \alpha = 0.1$. Since $n = 51 > 30$, the Central Limit Theorem applies.
 - (a) $H_0 : \mu \leq 350$ v.s. $H_A : \mu > 350$.
 - (b) Test statistics is $T = \frac{\bar{Y}-350}{s/\sqrt{n}} = \frac{355-350}{42/\sqrt{51}} = 0.8501$.
The critical value is $t_{50}(0.1) = 1.299$.
 - (c) The rejection region is $\{T : T > t_{50}(0.1)\}$. Since $T = 0.8501 < t_{50}(0.1)$, we fail to reject H_0 .
 - (d) We don't have enough evidence to support the claim that the average height of a Redwood tree in California is greater than 350 ft.
2. Taking the difference between each pair yields the table below.

	Score
Pre-test	70, 75, 82, 68, 72, 80, 77, 73, 78, 85, 67, 75, 88, 71, 74, 79, 76, 81, 84, 69
Post-test	78, 80, 86, 74, 79, 83, 82, 77, 81, 89, 76, 81, 91, 78, 81, 83, 79, 84, 87, 75
Difference	-8, -5, -4, -6, -7, -3, -5, -4, -3, -4, -9, -6, -3, -7, -7, -4, -3, -3, -3, -6

Here $n = 20, \bar{D} = -5, s = 1.8918, \alpha = 0.05$.

- (a) $H_0 : \mu_D = 0$ v.s. $H_A : \mu_D \neq 0$.
 - (b) Test statistics is $T = \frac{\bar{D}-0}{s/\sqrt{n}} = \frac{-5-0}{1.8918/\sqrt{19}} = -11.5205$.
The critical value is $t_{19}(0.025) = 2.093$.
 - (c) The rejection region is $\{T : |T| > t_{19}(0.025)\}$. Since $|T| = 11.5205 > t_{19}(0.025)$, we would reject H_0 .
 - (d) There is a significant difference in the mean test scores before and after the teaching method.
 - (e) The 95% confidence interval is $\bar{D} \pm t_{n-1}(\alpha/2) \times \frac{s}{\sqrt{n}} = -5 \pm 2.093 \times 1.8918/\sqrt{19} = (-5.91, -4.09)$.
Since 0 is not included in the interval, it is consistent with the conclusion in (d).
3. $\tilde{p} = \frac{Y+2}{n+4} = \frac{68+2}{340+4} = 0.2035, \alpha = 1 - 0.95 = 0.05$.
 - (a) The confidence interval is $\tilde{p} \pm z_{\alpha/2} \times \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.2035 \pm 1.96 \times \sqrt{\frac{0.2035(1-0.2035)}{340+4}} = (0.161, 0.246)$.
 - (b) We are 95% confident that the true proportion of infants with an adverse reaction is between 0.161 and 0.246.
 - (c) Since the confidence interval bounds is less than the value 0.25, it does support the claim that under 25% have an adverse reaction.
 - (d) By the definition of a confidence interval, we would expect $(1 - \alpha)100\%$ of them (i.e., 95%) to cover the true proportion.
4. $\tilde{p} = \frac{Y+2}{n+4} = \frac{14+2}{71+4} = 0.2133, \alpha = 1 - 0.95 = 0.05$.
 - (a) The confidence interval is $\tilde{p} \pm z_{\alpha/2} \times \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.2133 \pm 1.96 \times \sqrt{\frac{0.2133(1-0.2133)}{71+4}} = (0.1206, 0.3060)$.
 - (b) We are 95% confident that the true proportion of orangutans with type B blood is between 0.1206 and 0.3060.

- (c) Since the confidence interval contains the value 0.25, it does support the claim that 25% has type B blood.
- (d) By the definition of the confidence interval, we would expect 5% not to cover the true proportion.
5. (a) $H_0 : P(A) = 0.5, P(B) = 0.25, P(C) = 0.25$.
 H_A : At least two of the hypothesized proportions differ from the null.
- (b) The chi-squared test statistic is:

$$\begin{aligned} T &= \sum_{i=1}^3 \frac{(o_i - e_i)^2}{e_i} \\ &= \frac{(220 - 227)^2}{227} + \frac{(129 - 113.5)^2}{113.5} + \frac{(105 - 113.5)^2}{113.5} \\ &= 0.2159 + 2.1167 + 0.6366 \\ &= 2.9692. \end{aligned}$$

- (c) Because there are three categories, the degrees of freedom for the null distribution are calculated as $df = 3 - 1 = 2$. From χ^2 Table with $df = 2$ we find that $\chi_2^2(0.2) = 3.22 > T$. Thus, the range for our p -value is: $p\text{-value} > 0.2$.
- (d) Since the p -value is greater than $\alpha = 0.05$, we fail to reject H_0 at the 0.05 level of significance.
- (e) We conclude that there is enough evidence to support the genetic model.
6. (a) $H_0 : P(\text{Weekend}) = \frac{2}{7}, P(\text{Weekday}) = \frac{5}{7}$.
 $H_A : P(\text{Weekend}) \neq \frac{2}{7}, P(\text{Weekday}) \neq \frac{5}{7}$.
- (b) The chi-squared test-statistic is

$$\begin{aligned} T &= \sum_{i=1}^2 \frac{(o_i - e_i)^2}{e_i} \\ &= \frac{(216 - 266.286)^2}{266.286} + \frac{(716 - 665.714)^2}{665.714} \\ &= 9.4961 + 3.7985 \\ &= 13.2946. \end{aligned}$$

- (c) Because there are two categories, the degrees of freedom for the null distribution are calculated as $df = 2 - 1 = 1$. From χ^2 Table with $df = 1$ we find that $\chi_1^2(0.001) = 10.83$ and $\chi_1^2(0.0001) = 15.14$. Thus, the range for our p -value is: $(0.0001, 0.001)$.
- (d) If in reality there were the same proportion of births on the weekend that we would expect by chance, we would observe our data or more extreme with a probability between 0.0001 and 0.001.
- (e) Since the p -value is less than $\alpha = 0.01$, we reject H_0 at the 0.01 level of significance.
- (f) We conclude there is no enough evidence to suggest that the same proportion of births occur on the weekend as what we would expect by chance.
7. (a) $H_0 : P(\text{brown}) = 1/3, P(\text{black}) = 1/3, P(\text{white}) = 1/3$.
 H_A : At least two of the hypothesize proportions is different than the null.
- (b) The chi-squared test statistic is:

$$\begin{aligned} T &= \sum_{i=1}^3 \frac{(o_i - e_i)^2}{e_i} \\ &= \frac{(40 - 47)^2}{47} + \frac{(59 - 47)^2}{47} + \frac{(42 - 47)^2}{47} \\ &= 1.0426 + 3.0638 + 0.5319 \\ &= 4.6383. \end{aligned}$$

- (c) Because there are three categories, the degrees of freedom for the null distribution are calculated as $df = 3 - 1 = 2$. From χ^2 Table with $df = 2$ we find that $\chi_2^2(0.1) = 4.61$ and $\chi_2^2(0.05) = 5.99$. Thus, the range for our p -value is: $(0.05, 0.1)$.
- (d) Since the p -value is less than $\alpha = 0.1$, we reject H_0 at the 0.1 level of significance.
- (e) We conclude that there is evidence to suggest the true proportion of Mongolian Gerbils are not equally likely to be brown, black, or white.

Problem 8

In R, you can use the `t.test()` function to perform a paired-sample t test and calculate the confidence interval for the mean difference.

```
# Paired-Sample T Test
pre_growth <- c(2.1, 1.8, 2.4, 1.6, 2.0, 2.3, 1.7, 2.2, 2.5, 2.0)
post_growth <- c(2.6, 2.3, 2.8, 2.1, 2.4, 2.7, 2.0, 2.5, 2.9, 2.3)

# Perform paired-sample t test
result <- t.test(pre_growth, post_growth, paired = TRUE, conf.level = 0.95)
test_statistic <- result$statistic
p_value <- result$p.value

# Calculate confidence interval
confidence_interval <- result$conf.int

# Print results
cat("Paired-Sample T Test:\n")
```

```
## Paired-Sample T Test:
```

```
cat("test statistic:", test_statistic, "\n")
```

```
## test statistic: -15.49193
```

```
cat("p-value:", p_value, "\n")
```

```
## p-value: 8.521135e-08
```

```
cat("Confidence Interval:", confidence_interval, "\n")
```

```
## Confidence Interval: -0.4584086 -0.3415914
```

Based on the results, if the significance level is set at 0.05, we can conclude that there is a significant difference in the growth rates of plants before and after the treatment. The 95% confidence interval for the mean difference in growth rates is approximately 0.0531 to 0.3469 mm/day.

Problem 9

First, you need to transform the data into the table we usually have.

```
data.ori = read.csv("beans.csv")

plant <- table(data.ori$plant)
plant
```

```
##
##   Cowpea   Navy Northern   Pinto
##     155     155     192     144
```

The null hypothesis is $p_1 = p_2 = p_3 = p_4 = 1/4$.

```
res <- chisq.test(plant, p = c(1/4, 1/4, 1/4, 1/4))
res
```

```
##
## Chi-squared test for given probabilities
##
## data: plant
## X-squared = 8.1796, df = 3, p-value = 0.04244
```

(a) - (c)

The test statistic is 8.1796

The p -value is 0.04244.

Since the p -value is less than 0.05, we reject H_0 at the 0.05 level of significance.

(d)

Just compare the observation and expectation value.

```
res$observed[which(res$observed > res$expected)]
```

```
## Northern
##      192
```

The weevils prefer Northern more than expected if the null was true.

(e)

```
# (oi - ei)^2/ei
(res$observed - res$expected)^2/res$expected
```

```
##
## Cowpea      Navy Northern      Pinto
## 0.2616099 0.2616099 5.7600619 1.8962848
```

Northern contributed most to the value of the test-statistic.

Problem 10

```
data1 = read.csv("DRP.csv")
```

(a)

```
twoside = t.test(Response ~ Treatment, data = data1,
                 conf.level = 0.95, alternative = "two.sided")
# Calculate confidence interval
confidence_interval <- twoside$conf.int
cat("Confidence Interval:", confidence_interval, "\n")
```

```
## Confidence Interval: -18.67588 -1.23302
```

(b - c)

```
oneside = t.test(Response ~ Treatment, data = data1, alternative = "less")
test_statistic <- oneside$statistic
p_value <- oneside$p.value
cat("test statistic:", test_statistic, "\n")
```

```
## test statistic: -2.310889
```

```
cat("p-value:", p_value, "\n")
```

```
## p-value: 0.01319121
```

(d)

Since p -value is less than $\alpha = 0.05$, we reject H_0 at the 0.05 level of significance.