Apply Statistical Analysis

Michael Dang

University of Missouri - Kansas City

Feburary 20, 2025

Assignment 1

Solutions are to be due on Feburary 20th.

Note: $(*)$ indicate that the calculation is done through SAS.
The source code can be found at: here
The data can be found at: here

1. Solutions:

   (a) The distribution of the test statistics, T, is:

   $$\frac{\overline{Y} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} = \frac{\overline{Y} - \mu}{\frac{S}{\sqrt{n}}} \sim t_9$$

   (b)

   $$T = \frac{\overline{Y} - \mu}{\frac{S}{\sqrt{n}}} = \frac{28.74 - 25}{\frac{4.27}{\sqrt{10}}} = 2.769$$

   (c)
   - Calculate the p-value, for the one-tail test (since we are testing, $\mu > 25$):

     $$p = P(T \geq 2.769) = 1 - P(T < 2.769) = 0.0109 \qquad (*)$$

   - Since $p - value < \alpha, \quad 0.0109 < 0.05$ respectively, we reject the null hypothesis H_0 and conclude that there is a significant evidence to support $\mu > 25$.

   □

2. Solutions:

```
/*2a.*/
/*Upload data*/
PROC IMPORT DATAFILE='/home/u64145784/HCL_data.txt'
    OUT=infection_rate
    DBMS=dlm
    REPLACE;
    DELIMITER=' ';
    GETNAMES=NO;
RUN;

/*Name the dataset*/
DATA infection_rate;
    SET infection_rate;
    RENAME VAR1=rate;
RUN;

/*90% CI for mu*/
PROC MEANS DATA=infection_rate MEAN STDDEV CLM ALPHA=0.10;
    VAR rate;
RUN;
```

Figure 1: SAS Code for 90% Confidence Interval for μ

The MEANS Procedure

Analysis Variable : rate			
Mean	Std Dev	Lower 90% CL for Mean	Upper 90% CL for Mean
1.9582857	0.3517309	1.8577546	2.0588169

Figure 2: SAS Output Result for 90% Confidence Interval for μ

(a)
- Figure of the source code and output:

- We are 90% confident that the true mean infection rate among the responding patients lies between $(1.0577, 2.0588)$.

(b)
- Figure of the source code and output:

```
/*2b.*/
PROC TTEST DATA=infection_rate H0=2.5 ALPHA=0.01 SIDES=L;
    VAR rate;
RUN;
```

Figure 3: SAS Code statistical test with $\alpha = 0.01$ for $\mu < 2.5$

The TTEST Procedure

Variable: rate

N	Mean	Std Dev	Std Err	Minimum	Maximum
35	1.9583	0.3517	0.0595	1.1700	2.9500

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
1.9583	-Infty	2.1034	0.3517	0.2671	0.5049

DF	t Value	Pr < t
34	-9.11	<.0001

Figure 4: SAS Output Result for statistical test with $\alpha = 0.01$ for $\mu < 2.5$

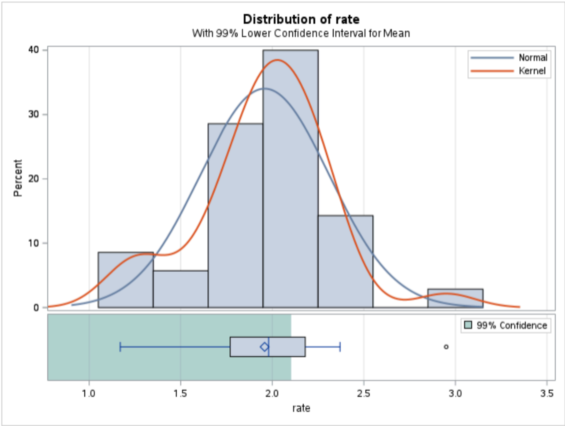

Figure 5: The distribution of rate approximately normal

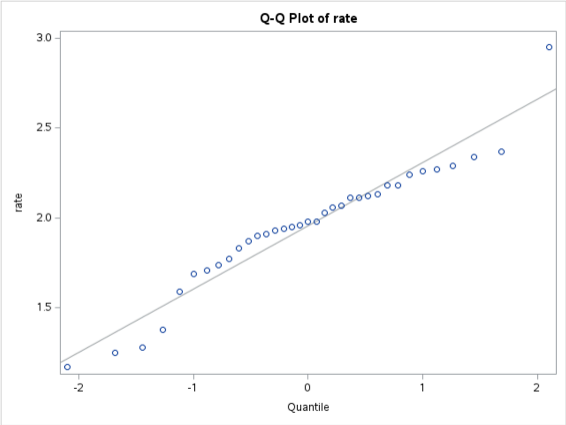

Figure 6: The qq-plot show the point line approximately on the line

(c) Based on Figure 5 and Figure 6, the normality assumption is not violated and we can assume that the observations are independently randomly selected.

□

3. Solutions:

(a) • Figure of the source code and output:
Denote: $\sigma^2 :=$ variance of the lengths of the ears of corn

```
/*3a.*/
/*Upload data*/
PROC IMPORT DATAFILE='/home/u64145784/Earlenth.dat'
    OUT=length_ear_corns
    DBMS=dlm
    REPLACE;
    DELIMITER=' ';
    GETNAMES=NO;
RUN;

/*Name the dataset*/
DATA length_ear_corns;
    SET length_ear_corns;
    RENAME VAR1=length;
RUN;

/*95% CI for mu*/
ods select BasicIntervals;
PROC UNIVARIATE DATA=length_ear_corns CIBASIC(ALPHA=0.05);
   VAR length;
RUN;
```

Figure 7: SAS Code for 95% Confidence Interval for σ^2

The UNIVARIATE Procedure
Variable: length

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	15.16000	13.80081	16.51919
Std Deviation	2.45438	1.79692	3.87081
Variance	6.02400	3.22892	14.98314

Figure 8: SAS Code Output Result for 95% Confidence Interval for σ^2

- With 95% confidence, we can say that the true variance of the lengths of the ears of the corn lies between $(3.228, 14.983)$.

(b) Figure of the source code and output:

```
/* 3b: Chi-Square test for variance */
PROC IML;
    /* Load data */
    USE length_ear_corns;
    READ ALL VAR {length} into x;
    n = nrow(x);
    s2 = VAR(x); /* Sample variance */
    sigma0 = 7.5; /* Hypothesized variance */
    df = n - 1;

    /* Compute Chi-Square test statistic */
    chi_sq_stat = df * s2 / sigma0;

    /* Compute p-value for one-tailed test (variance > 7.5) */
    p_value = 1 - PROBCHI(chi_sq_stat, df);

    /* Print results */
    PRINT n s2 chi_sq_stat p_value;
QUIT;
```

Figure 9: SAS Output Result for statistical test with $\alpha = 0.01$ for σ^2

n	s2	chi_sq_stat	p_value
15	6.024	11.2448	0.6667069

Figure 10: SAS Code Output Result for 95% Confidence Interval for σ^2

- Since $p - value > \alpha, \quad 0.666 > 0.01$ respectively, we fail to reject the null hypothesis H_0 and conclude that we don't have enough evidence to support $\sigma^2 > 7.5$.

(c) The data for ear length is assumed to follow a normal distribution and the sample of ear lengths is assumed to be independent randomly selected.

(d) Figure of the source code and output:

```
/* 3d: Compute Power when actual variance = 9 */
PROC IML;
    n = 15;
    df = n - 1;
    alpha = 0.01;
    sigma0 = 7.5;     /* Null hypothesis variance */
    sigma_a = 9;      /* Actual variance */

    /* Critical value for the Chi-Square test under H0 */
    crit_value = CINV(1 - alpha/2, df);

    /* Correct adjusted critical value under Ha */
    test_statistic = crit_value * (sigma0**2 / sigma_a**2);
    print test_statistic;

    /* Power calculation */
    power = 1 - PROBCHI(test_statistic, df);

    /* Print power */
    PRINT power;
QUIT;
```

Figure 11: SAS code for compute the power

test_statistic
21.749548

power
0.0839091

Figure 12: SAS Result Output for the power

$\square$

4. Solutions:

   (a) • 95% C.I. for σ^2:

   $$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L} \qquad (1)$$

   •
   $$\overline{X} = \sum_{i=1}^{15} \frac{X_i}{n} = \frac{18.1 + 15.3 + ... + 14.7 + 13.3}{15} = 15.16$$

   •
   $$s^2 = \frac{\sum_{i=1}^{15}(X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{15}(X_i - 15.16)^2}{14} = 6.024$$

   •
   $$\chi^2_{R,df} = \chi^2_{1-\frac{\alpha}{2},df} = \chi^2_{.975,14} = 5.628 \qquad (*)$$

   •
   $$\chi^2_{L,df} = \chi^2_{\frac{\alpha}{2},df} = \chi^2_{.025,14} = 26.118 \qquad (*)$$

   • From (1), we have:

   $$\frac{14 \cdot 6.024}{5.628} < \sigma^2 < \frac{14 \cdot 6.024}{26.118}$$

   $$= (14.983, 3.228)$$

- Hence with 95% confidence we can say, the population variance of length of ears of corn have the variance between $(14.983, 3.228)$

(b)

$$H_0 : \sigma^2 = 7.5 \quad vs \quad H_A : \sigma^2 > 7.5$$

- Test statisitcs:

$$T = \frac{(n-1)s^2}{\sigma^2} = \frac{14 \cdot 6.024}{7.5} = 11.245$$

- Critical value:

$$\chi^2_{0.99,14} = 29.141 \qquad (*)$$

- Since $T < \chi^2_{0.99,14}$, $\quad 11.245 < 29.141$ respectively, we can say with 99% confidence, we fail to reject H_0. Hence we can conclude that we don't have enough evidence to support $\sigma^2 > 7.5$.

(c) 3d. Compute the power:

- Given the actual variance, $\sigma_a^2 = 9$

$$\begin{aligned} Power &= 1 - \beta \\ &= 1 - P(\text{Fail to reject } H_0 \text{ — } H_0 \text{ is false}) \\ &= 1 - P(\text{Fail to reject } H_0 \text{ — } \sigma_a^2 = 9) \\ &= 1 - P(\text{test stat} < \chi_R^2 \text{ — } \sigma_a^2 = 9) \\ &= 1 - P\Big(\frac{(n-1)s^2}{\sigma_0^2} < \chi_R^2 \text{ — } \sigma_a^2 = 9\Big) \\ &= 1 - P((n-1)s^2 < \chi_R^2\sigma_0^2 \text{ — } \sigma_a^2 = 9) \\ &= 1 - P\Big(\frac{(n-1)s^2}{\sigma_a^2} < \frac{\chi_R^2\sigma_0^2}{\sigma_a^2} \text{ — } \sigma_a^2 = 9\Big) \\ &= 1 - P\Big(\chi^2 < \frac{\chi_R^2\sigma_0^2}{\sigma_a^2} \text{ — } \sigma_a^2 = 9\Big) \\ &= 1 - P\Big(\chi^2 < \frac{\chi_R^2\sigma_0^2}{\sigma_a^2} \text{ — } \sigma_a^2 = 9\Big) \\ &= 1 - P\Big(\chi^2 < \frac{\chi_R^2\sigma_0^2}{\sigma_a^2}\Big) \quad (*) \\ &= 1 - P\Big(\chi^2 < \frac{26.118 \cdot (7.5)}{9}\Big) \\ &= 0.0839 \end{aligned}$$

$\square$

5. Solutions:

- Assume $\alpha = 0.05$

$$H_0 : \mu \neq 5 \quad vs \quad H_A : \mu = 5$$

- Test statistic:

$$T = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.027 - 5}{\frac{0.1}{\sqrt{100}}} = 2.7$$

- Since this is a two-tail test, hence at 5% significant level, the critical value is:

$$z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} = \pm 1.96$$

- Since $|T| > Z_{0.025}$, $2.7 > 1.96$ respectively, hence we reject H_0. With 95% confidence, we have enough evidence to support the assumption of the engineer about the population mean.

$\square$

6. Solutions:

- We compare each observation to the hypothesized median (1.8 hours):

Observation (hours)	**Comparison to 1.8**	**Sign**
1.5	< 1.8	-
2.2	> 1.8	+
0.9	< 1.8	-
1.3	< 1.8	-
2.0	> 1.8	+
1.6	< 1.8	-
1.8	$= 1.8$	0 (excluded)
1.5	< 1.8	-
2.0	> 1.8	+
1.2	< 1.8	-
1.7	< 1.8	-

- Hypothesis test:

$$H_0 : m = m_0 \quad vs \quad H_A : m \neq m_0$$

Note: $X \sim Bin(10, 0.5)$

- Calculate the $p - value$:

$$\begin{aligned}
p - value &= 2 \cdot P(X \leq 3) \\
&= 2 \cdot \sum_{k=0}^{3} \binom{10}{k} (0.5)^{10} \\
&= 2 \cdot \left[\binom{10}{0} (0.05^{10} + \binom{10}{1} (0.5)^{10} + \binom{10}{2} (0.5)^{10} + \binom{10}{3} (0.5)^{10} \right] \\
&= 2 \cdot \left[\frac{1}{1024} + \frac{10}{1024} + \frac{45}{1024} + \frac{120}{1024} \right] \\
&= 2 \cdot -.1719 \\
&= 0.3438
\end{aligned}$$

- Since the $p-value > \alpha$, $\quad 0.3438 > 0.05$ respectively, we fail to reject H_0. Hence with 95% confidence, we don't have enough evidence to support the median operating time differs from 1.8 hours.

$\square$