MTH372 - Ststistical Inference
Assignment 2 Report

Aditya Chetan (2016217)

April 21, 2019

1 Question 1

1.1 Hypothesis

We want to see if the mean response time of the sample of pigs changes from alcohol.

Thus if μ_0 is the true response time, then our hypothesis is:

$$H_0 : \mu = \mu_0 \text{ Vs. } H_1 : \mu \neq \mu_0$$

Here, $\mu_0 = 0.8$

1.2 Assumptions

We work on the following assumptions:

- The sample can be assumed to be a simple random sample.
- Since the sample size, n 30, we can assume the sample to be normally distributed.

1.3 Test Statistic

We have a case of hypothesis testing on a normal distribution with unknown mean, and unknown variance. Hence, the test statistic is given by,

$$T_{test} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where, $\bar{x}$ and s are the sample mean and sample standard deviation respectively. Also, $T_{test} \sim T(n-1)$

1.4 Calculations

We have,
$n = 28$, $s = 0.3$, $\bar{X} = 1.0$, $\mu_0 = 0.8$ and $\sigma = 0.05$
First, we calculate the value of T_{test}. Let this be t. The `p-value` is given by,

$$P[T > |t|] = 2P[T > t]$$

.

The output of the code is:

```
Null Hypothesis: Alcohol does not affect mean response time (H_0: mu = mu_0)
Alternate Hypothesis: Alcohol affects mean response time (H_0: mu != mu_0)

Value of test statistic: 3.527668
P-value: 0.001521206
Is P-value < Significance Level? TRUE
Null Hypothesis is rejected.
```

1.5 Inference

Our `p-value` is 0.001521206 and significance level, α is 0.05. Clearly, `p-value` $< \alpha$. Hence, we **reject** the Null Hypothesis.

Therefore, we infer that alcohol does have an influence on the mean response time of pigs.

2 Question 2

2.1 Hypothesis

We want to see if the average life of the batteries is at least 240 hours.

Thus, if we take μ_0 as the true average life, then the hypothesis is,

$$H_0 : \mu > \mu_0 \text{ Vs. } H_1 : \mu \leq \mu_0$$

Here, $\mu_0 = 240$

2.2 Assumptions

- We assume the sample to be a simple random sample.
- The data is given to be approximately normally distributed. Hence, we can use that.

2.3 Test Statistic

We have a case of hypothesis testing on a normal distribution with unknown mean, and unknown variance. Hence, the test statistic is given by,

$$T_{test} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where, $\bar{x}$ and s are the sample mean and sample standard deviation respectively. Also, $T_{test} \sim T(n-1)$

2.4 Calculation

First, we calculate the value of T_{test}. Let this be t. The `p-value` is given by, $P[T \leq t]$.

The output of the code is:

```
P-value: 0.1417525
Is P-value < Significance Level? FALSE
Null Hypothesis is accepted.
```

2.5 Inference

Our `p-value` is 0.1417525 and significance level, α is 0.05. Clearly, `p-value` $> \alpha$. Hence, we **accept** the Null Hypothesis.

Therefore, we infer that the batteries do have a mean lifetime of at least 240 hours.

3 Question 3

3.1 Hypothesis

If σ_0 is the true standard deviation of the weights, then our hypothesis is:

$$H_0 : \sigma < \sigma_0 \text{ Vs. } H_1 : \sigma \geq \sigma_0$$

Here, $\sigma_0 = 0.4$

3.2 Assumptions

- We assume the sample to be a simple random sample.
- We can assume the weights to be normally distributed, since no other information is given.

3.3 Test Statistic

This is the case of hypothesis test on standard deviation of a normal distribution with unknown mean and unknown variance. Thus, the test statistic is,

$$\chi_{test} = \frac{(n-1)S^2}{\sigma_0^2}$$

where S is the sample standard deviation. Also $\chi_{test} \sim \chi^2_{n-1}$.

3.4 Calculations

After calculating χ_{test}, we calculate the `p-value` for the test, using $P[\chi^2_{n-1} \geq t]$, where t is the value for the test statistic (This is done because of the one-sided test).

The output of the code gives,

```
Null Hypothesis: Standard deviation of weights falls below 0.4 (H_0: sigma < s_0)
Alternate Hypothesis: Standard deviation of weights remains above 0.4 (H_1: sigma >= s_0)

P-value: 1
Is P-value < Significance Level? FALSE
Null Hypothesis is accepted.
```

3.5 Inference

Since the `p-value` is less than the significance level, hence we **accept** the Null Hypothesis.
Hence, we infer than by adopting the new strategy, the standard deviation in the weights will go below 0.4.

4 Question 4

4.1 Hypothesis

This is a case of hypothesis testing on two populations sampled from normal distributions. If μ_1, μ_2 are the true means of the two populations, then our hypothesis is:

$$H_0 : \mu_1 = \mu_2 \text{ Vs. } H_1 : \mu_1 \neq \mu_2$$

4.2 Assumptions

- σ_1, σ_2 are both unknown and unequal.
- Since the populations are randomly selected, hence both the samples are simple random samples.
- We can assume the populations to be from normal distribution as no other information is given.

4.3 Test Statistic

The test statistic is given by,

$$Z_{test} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Since, we are testing for $\mu_1 = \mu_2, \mu_1 - \mu_2 = 0$. Z_{test} follows T distribution with degree of freedom as,

$$df = \frac{(A+B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}}$$

where, $A = \frac{s_1^2}{n_1}$ and $B = \frac{s_2^2}{n_2}$

4.4 Calculations

After calculating the value of test statistic, t, we find the `p-value` for the test, from $P[T \geq |t|] = 2P[T \geq t]$.

The output of the code gives:

```
Null Hypothesis: Smoking does not affect blood pressure (H_0: mu_1 = mu_2)
Alternate Hypothesis: Smoking affects smoking (H_0: mu_1 != mu_2)

P-value: 0.01946976
Is P-value < Significance Level? TRUE
Null Hypothesis is rejected.
```

4.5 Inference

Since the `p-value` is less than the significance level, we **reject** the Null Hypothesis.

We infer that smoking does affect the blood pressure of smokers.

5 Question 5

5.1 Hypothesis

We have 3 different populations undergoing three different treatments. Let μ_1, μ_2, μ_3 be the true means of the three populations. Hence, our hypothesis is,

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ Vs. } H_1 : \textsc{At least one of the means is not equal to the others}$$

5.2 Assumptions

- We can assume the populations to be independent of each other.
- We can also assume the populations to have normal distribution.
- We assume the variances of all the 3 samples to be equal.
- Samples have to be simple random samples.
- We consider only one factor characterization.

5.3 Test Statistic

The test statistic for one-way ANOVA is given by:

$$F_{test} = \frac{\frac{1}{k-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y_{i,.}} - y_{.,.})}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y_{i,.}})^2}$$

This follows the F distribution.

5.4 Calculations

We need to calculate the `p-value` for the test. To do this, I used the `aov` function in R.

The out put was:

```
            Df    Sum Sq   Mean Sq F value Pr(>F)
class        2 0.0000145 7.270e-06   0.049  0.953
Residuals   12 0.0017892 1.491e-04
```

The `p-value` is given by the last column, i.e. 0.953.

5.5 Inference

Our `p-value` is 0.953 and significance level, α is 0.05. Clearly, `p-value` $> \alpha$. Hence, we **accept** the Null Hypothesis.

Therefore, we infer that all the 3 categories of resins have the same efficiency.

6 Question 6

6.1 Hypothesis

The hypothesis is that we want to see if the sample median is equal to a given true median or not:

$$H_0 : m = m_0 \text{ Vs. } H_1 : m \neq m_0$$

6.2 Assumptions

- The two samples are simple random samples.
- The samples are dependant on each other

6.3 Test Statistic

For the sign test, we choose the test statistic to be $min(n_+, n_-)$ where $n_+ = |\{i : x_1i - x_2i > 0\}|$ and $n_- = |\{i : x_1i - x_2i < 0\}|$

For the Wilcoxon sign test, we choose the test statistic to be $U = min(U_1, U_2)$, where,

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

and,

$$U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where, n_1, n_2 are the sizes of the two populations and R_1, R_2 are the sum of ranks in the respecting samples.

6.4 Calculations

I used the `binom.test` function in R for the sign test calculations. I fed it the value of the test statistic as well as the fact that the test was two-sided and confidence level was 0.95.

For the Wilcoxon test, I used the `wilcox.test` function in R. I fed it the matched pair of populations and the fact that the test was two-sided and confidence level was 0.95.

The output is shown below:

```
Null Hypothesis: Additive does not affect mileage (H_0: m_1 = m_2)
Alternate Hypothesis: Additive affects mileage (H_0: m_1 != m_2)


 Exact binomial test

data:  x and n
number of successes = 2, number of trials = 8, p-value = 0.2891
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.03185403 0.65085579
sample estimates:
probability of success
                  0.25


P-value: 0.2890625
Is P-value < Significance Level? FALSE
Null Hypothesis is accepted.

 Wilcoxon signed rank test

data:  car_with_additive and car_wo_additive
V = 32, p-value = 0.05469
alternative hypothesis: true location shift is not equal to 0
```

```
P-value: 0.0546875
Is P-value < Significance Level? FALSE
Null Hypothesis is accepted.
```

6.5 Inference

In both the tests, the outcome shows that the `p-value` is greater than the significance level. Hence, we accept the Null Hypothesis.

We infer that the additive did not affect the mileage.