[CC511] Homework 7 20204817 Federico Berto

November 1, 2020

1 Homework 7 - Federico Berto

[1]:
```python
# Useful libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
import statsmodels.stats.weightstats as sms
from scipy.stats import t
import math
```

1.1 8.1.12

We first find $z_{0.05}$

[2]:
```python
d = stats.norm()
print(("z = {:.4f}").format(d.ppf(0.95)))
```

```
z = 1.6449
```

Therefore c becomes:

$$c = \bar{x} - \frac{z_\alpha \sigma}{\sqrt{n}} = 11.80 - \frac{1.645 \times 2.0}{\sqrt{19}} = 11.045 \tag{1}$$

1.2 8.1.16

The interval can be written as: $(6.861 - 0.193, 6.861 + 0.193)$. We can get α from the t-distribution, given that:

$$t_{\alpha/2,n-1} = \frac{0.193\sqrt{n}}{S} = \frac{0.193\sqrt{16}}{0.440} = 1.7545 \tag{2}$$

[3]:
```python
from scipy.stats import t
print(("Alpha/2 = {:.4f}").format(t.sf(1.754, 15)))
```

```
Alpha/2 = 0.0499
```

So, the confidence level is $1 - 2(\alpha/2) = 1 - 2 \times 0.05 = 0.90$

1.3 8.2.4

The hypoteses are: - $H_0 : \mu \leq 70$ - $H_A : \mu > 70$ Test statistics:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{S} = \frac{\sqrt{25}(71.97 - 70)}{7.44} = 1.324 \quad (3)$$

Then we calculate the *p-value* using the survival function

[4]:
```
from scipy.stats import t
print(("P-value using sf (1-cdf) = {:.4f}").format(t.sf(1.324, 24)))
```

```
P-value using sf (1-cdf) = 0.0990
```

Given the significance level of 0.05, we can accept the hypotesis of the mean larger than 70 even though the p value is not too high

1.4 8.2.10

[5]:
```
d = stats.norm()
print(("Critical point when   = 0.10 = {:.4f}").format(d.ppf(1-0.10)))
print(("Critical point when   = 0.01 = {:.4f}").format(d.ppf(1-0.01)))
```

```
Critical point when   = 0.10 = 1.2816
Critical point when   = 0.01 = 2.3263
```

a) Null hypotesis is accepted when $z \geq -1.2816$

b) Null hypotesis is rejected when $z < -2.3263$

c) The test statistics becomes:

$$z = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} = \frac{\sqrt{29}(415.7 - 420)}{10} = -2.3156 \quad (4)$$

So the null hypotesis is rejected at size $\alpha = 0.10$ and accepted for $\alpha = 0.01$

[6]:
```
d = stats.norm()
print(("d) P-value = {:.4f}").format(d.cdf(-2.3156)))
```

```
d) P-value = 0.0103
```

1.5 8.2.18

[7]:
```
df=pd.read_excel('DS 6.1.7.xls')
print(df.describe())
```

```
       Paving Slab Weights
count           125.000000
mean              1.110536
std               0.052997
min               0.874000
25%               1.083000
```

```
50%                 1.110000
75%                 1.140000
max                 1.257000
```

We can consider the following hypoteses: - $H_0 : \mu = 1.1$ - $H_A : \mu \neq 1.1$ Test statistics:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{S} = \frac{\sqrt{125}(1.1105 - 1.1)}{0.0529} = 2.23 \tag{5}$$

[8]:
```
print(("P-value  = {:.4f}").format(2*t.sf(2.23, 124)))
```

```
P-value  = 0.0275
```

Given the significance level of 0.05, we should then reject the hypotesis and state that the manufacturing process probably needs some adjusting

1.6 8.2.28

[9]:
```
print(("Critical point when  /2 = 0.05 = {:.4f}").format(t.ppf(0.95, 19)))
print(("Critical point when  /2 = 0.005 = {:.4f}").format(t.ppf(0.995, 19)))
```

```
Critical point when  /2 = 0.05 = 1.7291
Critical point when  /2 = 0.005 = 2.8609
```

a) Null hypotesis is accepted when $|\, t \,| < 1.7291$

b) Null hypotesis is rejected when $|\, t \,| \geq 2.8609$

c) The test statistics becomes:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} = \frac{\sqrt{20}(436.5 - 430)}{11.90} = 2.4428 \tag{6}$$

So the null hypotesis is rejected at size $\alpha = 0.10$ and accepted for $\alpha = 0.01$

[10]:
```
print(("d) P-value = {:.4f}").format(2*t.sf(2.4428, 19)))
```

```
d) P-value = 0.0245
```

1.7 8.2.34

We will use a computer package to calculate the statistics directly given that $t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$

[11]:
```
n = 39
x_bar = 5532
mu_0 = 5680
s = 287.9
t_stat = math.sqrt(n)*(x_bar - mu_0)/s
print(("a) t = {:.4f}").format(t_stat))
print(("   P-value = {:.4f}").format(2*(t.sf(-t_stat, n-1))))
```

```
a) t = -3.2103
   P-value = 0.0027
```

[12]:
```python
n = 39
x_bar = 5532
mu_0 = 5450
s = 287.9
t_stat = math.sqrt(n)*(x_bar - mu_0)/s
print(("b) t = {:.4f}").format(t_stat))
print(("   P-value = {:.4f}").format(t.sf(t_stat, n-1)))
```

```
b) t = 1.7787
   P-value = 0.0416
```

1.8 8.6.5

[13]:
```python
df=pd.read_excel('DS 6.6.4.xls')
print(df.describe())
```

```
         Heights
count  60.000000
mean   69.618334
std     1.523098
min    66.900002
25%    68.599998
50%    69.500000
75%    70.224998
max    75.900002
```

[14]:
```python
n = 60
x_bar = 69.6184
mu_0 = 70
s = 1.5231
alpha_1 = 0.1
alpha_2 = 0.05
alpha_3 = 0.01
crit = t.ppf(1-alpha_1/2, n-1)
wing_span = crit*s/(math.sqrt(n))
print(("Confidence interval with 90% confidence level: ({:.4f})({:.4f})").
 ↪format(x_bar - wing_span, x_bar+wing_span))
crit = t.ppf(1-alpha_2/2, n-1)
wing_span = crit*s/(math.sqrt(n))
print(("Confidence interval with 95% confidence level: ({:.4f})({:.4f})").
 ↪format(x_bar - wing_span, x_bar+wing_span))
crit = t.ppf(1-alpha_3/2, n-1)
wing_span = crit*s/(math.sqrt(n))
print(("Confidence interval with 99% confidence level: ({:.4f})({:.4f})").
 ↪format(x_bar - wing_span, x_bar+wing_span))
```

```
Confidence interval with 90% confidence level: (69.2898)(69.9470)
Confidence interval with 95% confidence level: (69.2249)(70.0119)
```

```
Confidence interval with 99% confidence level: (69.0950)(70.1418)
```

Since 70 inches is included in the 95% confidence level, there is no strong evidence this value is not a plausible one for the mean

1.9 8.6.18

Given that $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{S}$, we can use the following hypotesis test: - $H_0 : \mu = 385$ - $H_A : \mu \neq 385$

[15]:
```
n = 33
x_bar = 382.97
mu_0 = 385
s = 3.81
alpha = 0.01
t_stat = math.sqrt(n)*(x_bar - mu_0)/s
print(("t-statistics t = {:.4f}").format(t_stat))
print(("P-value = {:.4f}").format(2*t.cdf(t_stat, n-1)))
```

```
t-statistics t = -3.0608
P-value = 0.0044
```

Given the P-value we can establish there is sufficient evidence the population mean is not 385

[16]:
```
crit = t.ppf(1-alpha/2, n-1)
wing_span = crit*s/(math.sqrt(n))
print(("Confidence interval with 99% confidence level: ({:.4f})({:.4f})").
 ↪format(x_bar - wing_span, x_bar+wing_span))
```

```
Confidence interval with 99% confidence level: (381.1537)(384.7863)
```

1.10 8.6.48

Given that $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{S}$, we can use the following hypotesis test: - $H_0 : \mu \geq 10$ - $H_A : \mu < 10$

[17]:
```
n = 40
x_bar = 9.39
mu_0 = 10
s = 1.041
alpha = 0.01
t_stat = math.sqrt(n)*(x_bar - mu_0)/s
print(("t-statistics t = {:.4f}").format(t_stat))
print(("P-value = {:.4f}").format(t.cdf(t_stat, n-1)))
```

```
t-statistics t = -3.7060
P-value = 0.0003
```

Given the extremely low P-value, we can safely reject the null hypotesis and assert that the phone surveys will last less than 10 minutes each on average

[18]:
```
crit = t.ppf(1-alpha, n-1)
wing_span = crit*s/(math.sqrt(n))
print(("Confidence interval with 99% confidence level: (-∞)({:.4f})").
 ↪format(x_bar+wing_span))
```

```
Confidence interval with 99% confidence level: (-∞)(9.7893)
```