Home Work 9

Sumeet Mishra

3/25/2019

1.

A company that manufactures light bulbs has advertised that its 75 watt bulbs burn an average of 800 hours before failing. In reaction to the company's advertising campaign, several dissatisfied customers have complained to a consumer watchdog organization that they believe the company's claim to be exaggerated. The consumer organization must decide whether or not to allocate some of its financial resources to countering the company's advertising campaign. So that it can make an informed decision, it begins by purchasing and testing 100 of the disputed light bulbs. In this experiment, the 100 light bulbs burned an average of $\bar{x} = 745.1$ hours before failing, with a sample standard deviation of s = 238.0 hours. Formulate null and alternative hypotheses that are appropriate for this situation. Calculate a significance probability. Do these results warrant rejecting the null hypothesis at a significance level of $\alpha = 0.05$?

$\mu = 800$
n=100 $\bar{x} = 745.1$
s=238.0

$H_0 : \mu >= 800$
$H_1 : \mu < 800$

$\alpha = 0.05$
z=$\frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$

```
z= (745.1 -800)/(238.0/sqrt(100))
z
```

```
## [1] -2.306723
```

```
p_value=pnorm(z)
p_value
```

```
## [1] 0.01053514
```

As the p_value<0.05, we can reject the null hypothesis.

2.

To study the effects of Alzheimer's disease (AD) on cognition, a scientist administers two batteries of neuropsychological tasks to 60 mildly demented AD patients. One battery is administered in the morning, the other in the afternoon. Each battery includes a task in which discourse is elicited by showing the patient a picture and asking the patient to describe it. The quality of the discourse is measured by counting the number of "information units" conveyed by the patient. The scientist wonders if asking a patient to describe Picture A in the morning is equivalent to asking the same patient to describe Picture B in the afternoon, after having described Picture A several hours earlier. To investigate, she computes the number of information units for Picture A minus the number of information units for Picture B for each patient. She finds an average difference of $\bar{x} = -0.1833$, with a sample standard deviation of s = 5.18633. Formulate null and alternative hypotheses that are appropriate for this situation. Calculate a significance probability. Do these results warrant rejecting the null hypothesis at a significance level of $\alpha = 0.05$?

n=60 $\bar{x} = -0.1833$
s=5.18633

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$
$\alpha = 0.05$

z=$\frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$ $\mu = 0$

```
z= ( -0.1833-0)/(5.18633/sqrt(60))
z
```

```
## [1] -0.273765
```

```
#As this is a 2 tailed test-
p_value=2*pnorm(z)
p_value
```

```
## [1] 0.7842652
```

As the p_value>0.05, we fail to reject the null hypothesis.

3.

Problem Set A

(1)

n=400 α=0.05

$H_0 : \mu <= 0$

$H_1 : \mu > 0$

$\bar{x} = 3.194887$

$s^2 = 104.0118$

a)

z=$\frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$

```
z= (3.194887-0)/(sqrt(104.0118/400))
z
```

```
## [1] 6.265333
```

b)

```
1-pnorm(1.253067)
```

```
## [1] 0.1050907
```

c) Given significance value, p = 0.03044555. As p_value < 0.05, we can reject Null hypothesis. Hence the statement is TRUE.

(2)

As Big > 255, let's assume- Big=256

```
val=c(251,238,249,256,243,248,229,256,235,244,254,251,252,244,230,222,224,246,256,239)
sorted_val=sort(val)
sorted_val
```

```
##  [1] 222 224 229 230 235 238 239 243 244 244 246 248 249 251 251 252 254
## [18] 256 256 256
```

```
n=length(sorted_val)
alpha=1-0.95
k=qbinom(alpha/2,n,0.5)
k
```

```
## [1] 6
```

```
1-2*pbinom(k,n,0.5)
```

```
## [1] 0.8846817
```

```
1-2*pbinom(k-1,n,0.5)
```

```
## [1] 0.9586105
```

By trial and error method, I found that decreasing k value by 1, i.e k=5 is giving around ~95%.

```
newk=k-1
newk
```

```
## [1] 5
```

```
print(paste(" The confidence interval is between",sorted_val[newk+1],"and",sorted_val[n-newk]))
```

```
## [1] " The confidence interval is between 238 and 251"
```

4.

Problem Set B

(2)

Confidence interval is 1-α=0.90

Significance level, is α=0.05

```
data=c(1.1402,-1.8658, 0.8520, -1.8251, 0.8530, -0.0589, -1.6554, -1.7599, -1.4330, -1.3853, 2.9794, 2.4
mu=0
sd=sqrt(mean(data^2)-mean(data)^2)
se=sd/sqrt(length(data))
t=(mean(data)-mu)/se
pvalue=1-pt(t,length(data)-1)
pvalue
```

```
## [1] 0.3599636
```

The pvalue is more that the significance value, α=0.05, hence we failed to reject Null hypothesis, H_0.

```
alpha=1-0.90
xbar=mean(data)
q=qt(1-alpha/2,length(data)-1)
upper=xbar+q*se
lower=xbar-q*se
print(paste("The confidence interval is between",lower,"and",upper))
```

```
## [1] "The confidence interval is between -0.495898831091488 and 0.759614620565172"
```

The confidence interval that we obtained here [-0.495, 0.759] is almost similar to the confidence interval we got in the case study.

(3)

Confidence interval is 1-α=0.90

Significance level, is α=0.05

```
y=sum(data<0)
pvalue=1-pbinom(y-1,length(data),0.5)
pvalue
```

```
## [1] 0.5
```

Since, p-value > significance level, we failed to reject null hypothesis.

```
sorted_data=sort(data)
sorted_data
```

```
##  [1] -1.8658 -1.8251 -1.7599 -1.6554 -1.4330 -1.3853 -0.8365 -0.7164
##  [9] -0.5479 -0.0589  0.6462  0.8520  0.8530  1.1402  1.1997  2.1601
## [17]  2.2670  2.4919  2.9794
```

```
l=length(sorted_data)
l
```

```
## [1] 19
```

```
alpha=1-0.90
data_k=qbinom(alpha/2,l,0.5)
data_k
```

```
## [1] 6
```

```
1-2*pbinom(data_k,l,0.5)
```

```
## [1] 0.8329315
```

```r
1-2*pbinom(data_k-1,l,0.5)
```

```
## [1] 0.9364319
```

By trial and error method, I found that decreasing k value by 1, i.e k=5 is giving around ~90%.

```r
newkval=data_k-1
print(paste("The confidence interval is between",sorted_data[newkval+1],"and",sorted_data[l-newkval]))
```

```
## [1] "The confidence interval is between -1.3853 and 1.1402"
```

The confidence interval we obtained here [-1.3853, 1.1402] does not match with the one in the case study.

5.

Problem Set D

(3)

Confidence interval - 90%

```r
log_data=c(0.693,0.662,0.690,0.606,0.570,0.749,0.672,0.628,0.609,0.844,0.654,0.615,0.668,0.601,0.576,0.
sorted_log_data=sort(log_data)
sorted_log_data
```

```
##  [1] 0.553 0.570 0.576 0.601 0.606 0.606 0.609 0.611 0.615 0.628 0.654
## [12] 0.662 0.668 0.670 0.672 0.690 0.693 0.749 0.844 0.933
```

```r
l1=length(sorted_log_data)
l1
```

```
## [1] 20
```

```r
alpha=1-0.90
log_data_k=qbinom(alpha/2,l1,0.5)
log_data_k
```

```
## [1] 6
```

```r
1-2*pbinom(log_data_k,l1,0.5)
```

```
## [1] 0.8846817
```

```r
1-2*pbinom(log_data_k-1,l1,0.5)
```

```
## [1] 0.9586105
```

By trial and error method, I found that for k value = 6, the confidence interval is coming to around ~90%.

```r
print(paste("The confidence interval is between",sorted_log_data[log_data_k+1],"and",sorted_log_data[l1
```

```
## [1] "The confidence interval is between 0.609 and 0.67"
```

So the confidence interval is [0.609, 0.670].

6.

H_0 - There is no change in the students' math test scores with live reggae music => $\mu <= 0$

H_1 - There is a positive change in the students' math test scores with live reggae music => $\mu > 0$

As population variance is unknown and sample size more than 30, we use z test.

```
n=61
xbar=6.5
s=12

z = (xbar - 0)/(s/sqrt(n))
alpha=0.05

pval=1-pnorm(z)
pval
```

```
## [1] 1.165593e-05
```

Assuming, α=0.05(by default) and pval<0.05, we can reject Null hypothesis, H_0.

```
cv=qnorm(1-alpha/2)
stder=s/sqrt(n)
lo=xbar-cv*stder
up=xbar+cv*stder
print(paste("The confidence interval is between",lo,"and",up))
```

```
## [1] "The confidence interval is between 3.48862791966374 and 9.51137208033626"
```

Since the interval of the confidence interval is wide enough, we can say that there can be improvements in the students test scores.

7.

Problem Set B

(3)

(a) The experimental unit is a person.

(b) The experimental units drawn from one population, i.e. aerobic students.

(c) Two measurements were taken on each experimental unit:

(1) Number of watts expended during protocol S (30-minute ride on the first week)

(2) Number of watts expended during protocol D (30-minute ride on the second week)

(d)

Let S_i be the score on protocol S for student i, and let D_i denote score on protocol D for student i. Then, $X_i = D_i - S_i$ is the random variable of interest. We are interested on drawing inferences about μ

(e)

$\mu > 0$ iff $D_i > S_i$. Thus, to test the theory in favor of dynamic streches we might want to test $H_0 : \mu <= 0$ vs $H_1 : \mu > 0$.