0307-614-01. Term 2205 Prof. Parody

Exam 2. 3/29/21. 100 points Name : Mrinal Satish Chaudhari

1. A study was run to assess patient outcomes to a new treatment of a respiratory disorder. The expectation was that the treatment yielded a better response than the placebo. A randomized double blind experiment was set-up to test this theory. Assume that 16 out of 64 people in the placebo group had a favorable outcome. For the treatment group, 40 out of 60 people had a favorable outcome.

a. Give the null and alternative hypotheses for this case

b. Calculate the test statistic (Show work)

c. Using a=0.05, make a decision based on the hypotheses

d. Are the assumptions for the test that you chose valid? (Explain why or why not)

e. Draw conclusions based on the real-world question

Solution:
From the question we get values,
X1=16, n1=64
x2=40, n2=60,α=0.05

The null and alternative hypothesis,
H_0: π_1-π_2=0
H_1:π_1-π_2<0

We will find,
p1=X1/n1=16/64=0.25;
p2=x2/n2=40/60=0.66

For sample 1,

S1= n1*p1=64*0.25=16
F1=N1(1-P1)= 64(1-0.25)=48

S2=N2*P2=60*0.66= 39.6
F2=N2(1-P2)= 60(1-0.66)=20.4

We have met the requirements to use the normal approximation to the binomial for both samples.

Now since under the null we assume that the population proportions are equal, we can use a pooled sample proportion in the test statistic. The pooled sample proportion p is:

p=X1+X2/N1+N2
p=16+40/64+60
p=0.4516

Now the test statistic becomes,

$Z_{stat}=(P1 - P2 - \delta 0)/[\sqrt{p}(1 - p) * (1/n1 + 1/n2)]$
$Z_{stat}=(0.25 - 0.66)/[\sqrt{0.4516}(1 - 0.4516) * (1/64 + 1/60)]$
$Z_{stat}=- 0.41/[\sqrt{0.007974}$
$Z_{stat}= -0.41/0.0893$
$Z_{stat}= -4.5912$

To find p-value,
p-value= $2*P(Z\geq|Z_{stat}|)$
p-value= 0.0000016
Since p-value <α
We reject the null hypothesis,
There is enough evidence to support alternative hypothesis,
Therefore, we can conclude that the expectation that the treatment yielded a better response than the placebo holds good.

$Z_{crit}=Z_{0.975}=1.96$

95% confidence interval,
$(\pi 1-\pi 2)\in(-0.5763, -0.2571)$

Hence, we can conclude the true mean difference is between -0.5763 and -0.2571

2. Suppose we want to determine if a new drug was effective at causing weight loss. To determine this, we randomly selected 10 subjects, and weighed them before and after treatment. The resulting summary data for the differences is given below. Assume that the differences come from a normal distribution and the data below is based on before-after.

Sample Size	**10**
Sample Differences Mean	**6.25**
Sample Differences Standard Deviation	**7.52**

Solution:

a. Explain why we consider this paired data

The distribution of the differences come from a normal distribution with unknown population standard deviation of the differences. Since 10 subjects are different in weights and measurements, we are not changing subjects across the samples. This indicates pairing so we will use paired t-test.

b. Give the null and alternative hypotheses

The null hypothesis in our case will be:

H_0: $\mu_1 - \mu_2 = 6.25$

The alternate hypothesis, in this case, will be:

H_0: $\mu_1 - \mu_2 > 6.25$

c. Create a 95% confidence interval and use it to make a decision based on the hypotheses

n_d=10, X_d=6.25, S_d=7.52, α=0.05

Now we will calculate test statistic,

t_{stat}= $(Xd - \Delta 0)/(Sd\sqrt{nd})$

t_{stat}= $6.25/(7.52\sqrt{10})$

t_{stat}= $6.25/(7.52/3.1622)$

t_{stat}= $6.25/2.378$

t_{stat}= 2.628

Now, we will calculate p-value.

Degree of freedom,

df=n_d-1

df=10-1=9

p-value=P($t_{df} \leq t_{stat}$)

p-value= P($t_9 \leq 2.628$)

p-value= 0.027449

Decision rule based on p-value,

Reject if , p-value ≤ α

Since p-value ≤ α,we reject the null hypothesis. There is enough evidence to support alternative hypotheses.

$t_{crit}=t_{0.95,9}$= 1.833

Lower Bound= $(X1 - X2) - tcrit * sd/\sqrt{nd}$=1.891
Upper Bound=$(X1 - X2) + tcrit * sd/\sqrt{nd}$=10.609

$\mu_1 - \mu_2 \in$ (1.891,10.609)

d. Draw conclusions based on the real-world question

We reject the null hypothesis. There is enough evidence to support alternative hypotheses. Hence we can conclude that, new drug was effective at weight loss.

3. Low-Back pain is a serious health problem in most industrial settings. An article reported the accompanying summary data on lateral range of motion (degrees) for a sample of workers with and without the malady. Assume that the data comes from normal distributions. Test to see if the mean motion is the same across both groups using a=0.05. To get full credit, you will need to justify what you are doing, do all tests needed and give the real-world answer.

Condition	**Sample Size**	**Sample Mean**	**Sample Std Dev**
No LBP	**14**	**91.5**	**2.5**
LBP	**16**	**88.3**	**4.5**

Solution:

Solution:

We have to check if variance is equal or not,

The null hypothesis,

$H_0: \sigma 1^2/\sigma 2^2 = 1$
$H_A: \sigma 1^2/\sigma 2^2 \neq 1$
The test statistic becomes,

F_{stat}= $(s1^2/s2^2) * (\sigma 1^2/\sigma 2^2)$
F_{stat}= $2.5^2/4.5^2$=0.3086
df1=n1-1=14-1=13
df2=n2-1=16-1=15

We find the p-value with the formula for a not equal to test,

p-value= 2* min[P($F_{df1,df2} \geq F_{stat}$),P($F_{df1,df2} \leq F_{stat}$)]
p-value= 2* min[1-P($F_{df1,df2} \leq F_{stat}$) , P($F_{df1,df2} \leq F_{stat}$)]
p-value=0.0395
Decision rule based on p-value,
Reject if p-value≤ α

Since p-value≤ α , we reject H_0,
There is enough evidence to state that variance assumption is equal.

Assume that the data is normal and the population variances are not equal from above calculations.

The null hypothesis,

H_0: $\mu_1-\mu_2=0$

The alternative hypothesis becomes,

H_A: $\mu_1-\mu_2 \neq 0$

Since the data comes from normal distribution with unknown population SD, we will uset-test.

Since, we have two different groups for Low-back pain, we will assume an independent sample.

We have,

$\bar{x}1$=91.5, s1=2.5, n1=14

$\bar{x}1$=88.3, s1=4.5, n1=16

The test statistic becomes,

tstat= $(\bar{x}1 - \bar{x}2 - \triangle 0)/[Sp(\sqrt{}1/n1 + 1/n2)]$

The formula for Sp is,

Sp= $\sqrt{}[(n1 - 1)S1^2 + (n2 - 1)S2^2/(n1 + n2 - 2)]$

Sp= $\sqrt{[(14-1)2.5^2+(16-1)4.5^2/(14+16-2)]}$
Sp= $\sqrt{[13*6.25+15*20.25/28]}$
Sp= $\sqrt{[81.25+303.75/28]}$
Sp= $\sqrt{[385/28]}$
Sp= $\sqrt{13.75}$
Sp= 3.7080

t_{stat}= $(\bar{x}1-\bar{x}2-\triangle 0)/[Sp(\sqrt{1/n1+1/n2})]$
Now, we will put this value in above equation,
t_{stat}= $(91.5-88.3)/[3.7080(\sqrt{1/14+1/16})]$
t_{stat}= $(3.2)/[3.7080(\sqrt{0.0714+0.0625})]$
t_{stat}= $(3.2)/[3.7080(\sqrt{0.1339})]$
t_{stat}= $(3.2)/[3.7080*0.3659]$
t_{stat}= $3.2/1.3567$
t_{stat}= 2.446

We can calculate p-value,
df=(n1+n2-2)
df=(14+16-2)
df=28

We will find p-value,
p-value=P(tdf≥tstat)
p-value=1-P(tdf≤tstat)
p-value= 1-P(t_{28} ≤ 2.3586)
p-value= 0.0218

Critical point:
t0.975,28= 2.06
Rejection rule based on p-value,
Reject if, p-value ≤ α

Since p-value ≤ α i.e 0.0256<0.05 we reject the null hypothesis. There is enough evidence to support alternate hypotheses.

We can conclude that there is strong evidence to state that mean motion is not the same across the group.

We can provide 95% confidence interval based on above calculation,
We are 95 % confidence that the mean difference value range from,
(μ1−μ2)∈(0.5052,5.895)

Therefore, we can conclude that the true mean difference between two groups lies between 0.5052 and 5.895

4. Two types of fish attractors, one made from vitrified clay pipes and the other from cement blocks and brush, were used during 16 different time periods spanning 4 years at

Lake Tohopekaliga, Florida. The following observations are of fish caught per fishing day.

Period

	1	2	3	4	5	6	7	8
Pipe	.00	1.80	4.86	.58	.37	.32	.11	.23
Brush	.48	2.33	5.38	.79	.32	.76	.52	.91

Period

	9	10	11	12	13	14	15
Pipe	.29	.85	.57	1.83	7.89	.63	.42
Brush	.75	1.61	.83	2.17	8.21	.56	.75

Does one attractor appear to be more effective on average than the other? Use .01 level of significance. To get full credit, you will need to justify what you are doing, do all tests needed and give the real-world answer

Solution:

Since fish attractors were used in all 16 different time periods, this will be a type of paired t test.

We will first check for equal variance test,

For this we need variance for two attractor that we will get using stathelper,

Descriptive Statistics

	Sample 1	Sample 2
Sample Size	15.00	15.00
Sample Variance	4.72	4.82
Standard Deviation	2.17	2.20
Degrees of Freedom	14.00	14.00

S_{pipe}=2.17. S_{brush}=2.20, N_{pipe}=15, N_{brush}=15, α=0.01

The null hypothesis,
H_0:$\sigma1^2/\sigma2^2 = 1$
H_A:$\sigma1^2/\sigma2^2 \neq 1$
The test statistic becomes,

F_{stat}= $(s1^2/s2^2) * (\sigma1^2/\sigma2^2)$
F_{stat}= $(2.17/2.20)$=0.98

df $_{brush,pipe}$=15-1=14
We find the p-value with the formula for a not equal to test,

p-value= 2* min[$P(F_{df1,df2} \geq F_{stat})$,$P(F_{df1,df2} \leq F_{stat})$]
p-value= 2* min[1-$P(F_{df1,df2} \leq F_{stat})$, $P(F_{df1,df2} \leq F_{stat})$]
p-value= 0.9699
Decision rule based on p-value,
Reject if p-value≤ α
Since p-value>α, we failed to reject the null hypothesis.
There is evidence to state that the variances are equal.

Results for the Normality Test

For the Shapiro-Wilk normality test, we assume that:

H_0 : The data comes from a normal distribution

H_A : The data does not comes from a normal distribution

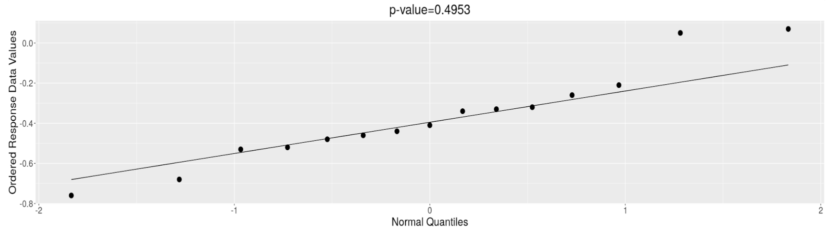
The p-value for this test is 0.4953

Since 0.4953> 0.05 we fail to reject the null hypothesis.

This means that we should assume that the given data comes from a normal distribution.

This is reflected in the normal probability plot below

The plot has points that closely follow the line.



From the above graph we can say that data comes from normal distribution.

The null and alternative hypothesis,
$H_0: \mu_{pipe} - \mu_{brush} = 0$
$H_A: \mu_{pipe} - \mu_{brush} \neq 0$

Statistic	*Value*
Mean	*-0.3747*
Median	*-0.41*
Sample Size	*15*
Q1	*-0.52*
Q3	*-0.26*
Min	*-0.76*
Max	*0.07*
IQR	*0.26*
Range	*0.83*
Variance	*0.0525*
Standard Deviation	*0.2292*
Lower Fence	*-0.91*
Upper Fence	*0.13*

Pipe	Brush	Diff
0	0.48	-0.48
1.8	2.33	-0.53
4.86	5.38	-0.52
0.58	0.79	-0.21
0.37	0.32	0.05
0.32	0.76	-0.44
0.11	0.52	-0.41
0.23	0.91	-0.68
0.29	0.75	-0.46
0.85	1.61	-0.76
0.57	0.83	-0.26
1.83	2.17	-0.34
7.89	8.21	-0.32
0.63	0.56	0.07
0.42	0.75	-0.33

$S_{diff}=0.2292, N_{diff}=15, \bar{x}_{diff}= -0.3747, \alpha=0.01$

$t_{stat}= (\bar{x}diff - \Delta 0)/(Sdiff/\sqrt{}ndiff)$

$t_{stat}= (-\ 0.3747)/(0.2292/\sqrt{}15)$

$t_{stat}= (-\ 0.3747)/(0.2292/3.8729)$

$t_{stat}= (-\ 0.3747)/(0.05918)$

$t_{stat}= -6.3315$

df=n-1=15-1=14

Now we will calculate p-value,

p-value=$P(t_{df} \leq t_{stat})$

p-value= 0.000019

$t_{crit}=t_{0.995,14}=2.977$

Decision rule based on p-value,

Reject if , p-value ≤ α

Since p-value ≤ α, we reject the null hypothesis.

There is enough evidence to support alternative hypotheses.

Hence we can conclude, there is enough evidence to state that one fish attractor seems to be more effective than another and they both are unequal.

Lower Bound= $\bar{x}diff - tcrit * sd/\sqrt{nd}$=-0.5301
Upper Bound=$\bar{x}diff + tcrit * sd/\sqrt{nd}$= -0.2193

$\mu_1 - \mu_2 \in (-0.5301,-0.2193)$

We are 99% confident that the true mean difference lies between -0.5301 and -0.2193.

We can also state that the brush attractor is more effective than the pipe attractor.