Satoshi Ido
34788706
Professor. Craig
19 February, 2024

STAT 582

Task3

Problem 1

Suppose an animal study is being planned as a completely randomized design involving 6 treatment groups. Past studies indicate the standard deviation of a response measurement is 7.0 units. Using the 5% significance level,

a. What sample size is needed for the power of the overall F test to be at least 80% given that the researchers want to detect any pairwise difference in means of 8.0 units?

---

21 samples are needed in a conservative assessment of sample size given information (sigma = 0.7 and least 80% power, difference in means of 8.0 units).
The output are below:

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	-4 0 0 0 0 4
Standard Deviation	7
Nominal Power	0.8

Computed N per Group	
Actual Power	N per Group
0.810	21

b. What sample size per group is needed to detect a difference between Trt 1 and Trt 2 of 8.0 units with 80% power?

---

11 samples are needed in each group based on two sample t-tests.
Code and output are below:

The POWER Procedure
Two-Sample t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	1
Null Difference	0
Alpha	0.05
Mean Difference	8
Standard Deviation	7
Nominal Power	0.8

Computed N per Group	
Actual Power	N per Group
0.827	11

c. Explain as if you were talking to a client why these two calculations result in different sample sizes.

For the overall F test in an ANOVA, think of it as trying to determine if there's any evidence at all among six treatment groups. We need more evidence, which means a larger sample size, to be confident that there are certain differences in means. Especially because the overall F test involves comparing all six groups simultaneously, it must account for the increased variability. We often require a larger sample size to maintain the power to detect smaller effects in the presence of this variability.

In contrast, the pairwise comparison using a two-sample t-test is focused on detecting a difference between just two specific groups (Trt 1 and Trt 2). This test does not need to account for variability among all six groups, only the variability within and between the two groups being compared. As such, the sample size required to achieve the same power is typically smaller because the test is more targeted and less susceptible to the overall variability present in the ANOVA.

Problem 2

2. Refer to the scenario of #1b. What sample size per group is needed if you gauged significance using Tukey's multiple comparison adjustment?

We need at least **22** sample sizes per group given Tukey's adjusted multiple comparison. The SAS output is below (For simplicity, extracted necessary data)

Obs	n	df	power
1	15	84	0.58578
2	16	90	0.62597
3	17	96	0.66362
4	18	102	0.69864
5	19	108	0.73102
6	20	114	0.76075
7	21	120	0.78792
8	22	126	0.81260
9	23	132	0.83492
10	24	138	0.85501
11	25	144	0.87301

Problem 3

Suppose that in the process of running the animal study described in #1 you learn that the animals were from three different suppliers. A reanalysis of the data suggest that supplier explains 60% of the variance of the response measurements. If the researcher were to rerun this study of six treatments but now using supplier as a blocking factor, how many animals per block/treatment would be needed for at least 80% power?

The researcher needs **6** replicates per treatment per treatment per supplier (block factor). Given the standard deviation = 7.0 and block factor explains 60% of the variance of the response measurements, we calculate variances between and within the block.

- Variance of block = 7^2 * 0.6 = 29.4
- Variance within block = 7^2 * 0.4 = 19.6

Assuming the overall mean = 50 and experimented with the impact of replication on power analysis, below are the output of two cases. The first output assumes 3 replicates of animals per treatment per block, and the second output assumes 6 replicates of animals per treatment per block. As we can see from the output below, the power for the case with 3 replicates is very low (0.488) while the power for the case with 6 replicates is over 0.8. Because I evenly assigned numbers of samples to block and treatment, I cannot choose cases with 4 or 5. Hence, I will choose **6** replicates per treatment per supplier (block) as the statistically reasonable number.

Obs	alpha	noncent	crit	NumDF	DenDF	Power
1	0.05	11.2902	3.32583	5	10	0.48845

Obs	alpha	noncent	crit	NumDF	DenDF	Power
1	0.05	28.9108	2.55813	5	28	0.97945

Problem 4

A researcher is planning a three-arm clinical trial with a time-to-event endpoint measured in days. In an earlier study involving the current standard-of-care treatment, the natural logarithm of time-to-event was approximately Normal with mean 3.4 and standard deviation 0.5. The researcher thinks two new treatments might increase the average time-to-event by at least 14 days and wants to make sure he can detect this. How many patients per group do you recommend he consider for 80% power?

Surprisingly, we only need to have **2** samples per group to meet 80% power requirement.
The researcher believes that two new treatments might increase the average time-to-event. This belief suggests a direction in the change, namely an increase. Therefore, it seems that the interest is in testing whether the new treatments are better than the standard-of-care treatment, which could indeed imply a one-sided test. It is reasonable to compare each treatment to a control rather than making all pairwise comparisons in this problem setting. Hence, we use Dunnet's adjustment.

Given the information, delta = ln(14) = 2.64, var=0.5*0.5 = 0.25, nc = d/sqrt(var*2/n), I calculate power using a non-central t-distribution, which gives me the output below.

Obs	n	df	crit	power
1	2	3	3.86634	0.82953
2	3	6	2.86275	0.99750
3	4	9	2.61395	0.99998
4	5	12	2.50241	1.00000
5	6	15	2.43926	1.00000
6	7	18	2.39865	1.00000
7	8	21	2.37035	1.00000
8	9	24	2.34941	1.00000
9	10	27	2.33341	1.00000

Problem 5

An animal science researcher is planning a three-period crossover design to compare diets on milk production. Previous studies have suggested milk production to be Normally distributed with a mean of 2100 liters and a standard deviation of 48 liters and that the intraclass correlation of milkings within a cow is 0.50. How many cows are needed if the researcher wants to detect differences in means of 20 liters (80% power)?

We need at least **9** sample sizes to detect differences in means of 20 liters. This experiment is a three-period cross design to compare diets on milk production and holds intraclass correlation of milkings within a cow is 0.50. Given the prior knowledge of milk production to be Normally distributed with a mean of 2100 liters and a standard deviation of 48 liters, I generate a dataset and run a mixed model power calculations. The first output below is the one with the sample size=8 and the second output below is the one with the sample size=9. We can see from these outputs, to meet 80% power, we need at least 9 sample sizes.

Obs	alpha	noncent	crit	NumDF	DenDF	Power
1	0.05	8.16097	3.73889	2	14	0.62502

Obs	alpha	noncent	crit	NumDF	DenDF	Power
1	0.05	12.0555	3.63372	2	16	0.81269

Appendix

<Problem 1a>

```
proc power;
    onewayanova alpha=.05 test=overall
    groupmeans=(-4 0 0 0 0 4)
    npergroup=.
    stddev=7.0
    power=.80;
run;
```

<Problem 1b>

```
proc power;
    twosamplemeans
    alpha=.05
    nulldiff=0
    sides=1
    meandiff=8.0
    npergroup=.
    stddev=7.0
    power=.80;
run;
```

<Problem 2>

```
data prob2;
    a=6; alpha=.05; var=49; d=8;
    do n=15 to 25;
        /* df for the error term */
        df = a*(n-1);
        /* non-centrality */
        nc = d/sqrt(var*2/n);
        crit = probmc("range",.,1-alpha,df,a)/sqrt(2);
        /* Tukey */
        power=1-probt(crit,df,nc)+probt(-crit,df,nc);
    output;
end;
```

<Problem 3>

```
* 1. Create a data set of means. with 2 replicate per treatment per block;
data animal_study2;
  do block = 1 to 3;                /* Blocking factor with 3 levels */
    do trt = 1 to 6;                   /* Treatment factor with 6 levels */
      do rep = 1 to 2;              /* 2 replicates per treatment per block */
        y = 50 /* Overall mean */
            + (trt=1)*4 /* Effect of treatment 1 */
            - (trt=6)*4 /* Effect of treatment 6 */
            + rand('normal', 0, 7); /* Random normal variability with SD=7 */
        output;
      end;
    end;
  end;
run;
proc print data=animal_study2; run;
*2. Run MIXED with several additions;
Proc Mixed noprofile data=animal_study2;
   class block trt;
   model y = trt;
   random block;
   parms 29.4 19.6 / Noiter;
   ODS Output Tests3=ANOVAtest2;
run;

data power2;
    set ANOVAtest2;
    alpha = 0.05;
    noncent = numdf*Fvalue;
    crit = Finv(1-alpha,numdf,dendf,0);
    Power = 1-Probf(crit,numdf,dendf,noncent);
run;
proc print data=power2;
    var alpha noncent crit numdf dendf power;
run;
```

<Problem 4>

```
data prob4;
  d=2.64; alpha=.05; a=3; var=0.5*0.5;
  do n=2 to 10;
  df = a*(n-1);
  nc = d/sqrt(var*2/n);
  crit = probmc("dunnett2",.,1-alpha,df,a-1); *Dunnett adjustment;
  power=1-probt(crit,df,nc); *one-sided test;
  output;
end;
proc print;
  var n df crit power;
run;
```

<Problem 5>

```
/* generate the data */
%let mean = 2100; /* mean milk production */
%let stddev = 48; /* standard deviation of milk production */
%let icc = 0.50; /* intraclass correlation */
%let diff = 20; /* desired detectable difference */
%let power = 0.8; /* desired power */
%let alpha = 0.05; /* significance level */
%let max_sample_size = 100; /* maximum sample size to test */
%let min_sample_size = 10; /* minimum sample size to test */
%let step = 10; /* step size for sample size increments */
%let num_simulations = 1000; /* number of simulations to perform per sample size */


data milk;
call streaminit(123); /* set the seed for reproducibility */
do cow = 1 to 8; /* loop over cows */

    /* generate a cow-specific random effect based on the ICC */
    cow_effect = rand('NORMAL', 0, &stddev * sqrt(&icc));

    do period = 1 to 3; /* Loop over periods */
        /* apply the treatment effect in period 2 */
        if period = 2 then treatment_effect = &diff;
        else treatment_effect = 0;

        /* simulate milk production measurements with cow-specific random effect and
        milk_production = &mean + treatment_effect + cow_effect +
                          rand('NORMAL', 0, &stddev * sqrt(1 - &icc));
        output;
    end;
end;
run;


proc mixed data=milk method=ml;
    class cow period;
    model milk_production=period;
    random intercept / subject=cow type=cs;
    estimate 'Treatment Effect' | period 1 / subject 1 -1;
    ODS Output Tests3=ANOVAtest2;
run;

data milk2;
    set ANOVAtest2;
    alpha = 0.05;
    noncent = numdf*Fvalue;
    crit = Finv(1-alpha,numdf,dendf,0);
    Power = 1-Probf(crit,numdf,dendf,noncent);
run;
proc print data=milk2;
    var alpha noncent crit numdf dendf power;
run;
```