Homework #5/Project (200pts total) Name\_\_\_Yuxuan Chen\_\_\_\_\_\_\_\_\_

Due 4/10 1pm

1a.(50 pts) The primary objective of this study is to show that chronic treatment with drug A in Cystic Fibrosis (CF) subjects with high risk prediabetes prevents the conversion to diabetes. The primary outcome variable is the conversion from high risk prediabetes to overt diabetes. Forty percent of CF patients in the study age range are expected to have high risk prediabetes. Hence, for our sample size calculation, we assume that 20% of the CF adults with high risk prediabetes will develop CFRD each year. In addition, based on preliminary data, we estimate the one-year conversion rate for the treatment group is between 2-8%.

The sample sizes are equal in the two treatment groups. Table 1 presents the total sample is required to achieve 80% power when a log-rank test is used to compare the two groups with a one-year enrollment, one-year follow-up, assuming a 0% dropout rate. Complete the table below of the sample sizes needed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | One-year conversion rate in the control group | One year conversion rate in the treatment group | | | | 2% | 5% | 8% | | 25% | N=36 | N=58 | N=94 | | 20% | N=50 | N=91 | N=170 | | 17% | N=64 | N=131 | N=279 | |

*R code:*

*library(gsDesign)*

*control.convert = c(0.25, 0.2, 0.17)*

*treat.convert = c(0.02, 0.05, 0.08)*

*lambda.c = -log(1-control.convert)*

*lambda.t = -log(1-treat.convert)*

*for(i in lambda.c) for(j in lambda.t){*

*cat(i, j,'\n')*

*samp = nSurvival(lambda1=j, lambda2=i,*

*eta = 0, # drop out rate*

*Ts = 2, # enrollment + follow up*

*Tr = 1, # enrollment*

*sided = 2, alpha = 0.05, beta=0.2)$n*

*cat("sample", samp,"\n")*

*}*

1b. (10pts) The secondary objective from the study above in part (a) is exploratory and will compare some baseline lab biomarkers by randomized arm. From the table above, choose one of the sample sizes that you calculated. Calculate the effect size that you will be able to detect with 80% statistical power, assuming the biomarker is continuous and using a two sample t test.

If the sample size per group is 91, the effect size that is able to be detected with 80% statistical power is 0.4176 given that the standard deviation of the biomarker is 1 and significant level is 0.05.

*R code:*

*power.t.test(n=91, # sample size per group*

*power=0.8, sig.level = 0.05, type="two.sample")*

1c.**Bonus** (10 pts): Assume the lab biomarker is to be a repeated measure throughout the study. If a positive correlation between measurements at different visits from the same individual exists, would you expect the power to increase, decrease or stay the same? Why?

The power would decrease if all the repeated measurements were used for two-sample t test. Because the two-sample t test assumes the observations are independent.

2. (10 pts) A randomized trial is investigating the reduction in mortality attributable to reduced dietary fat (i.e. randomize patients to reducing their fat intake vs eat normally). Investigators choose alpha=.05 and beta=.1 for the trial design. Discuss the appropriateness of these choices.

The choice of power might be set too high. This is because mortality attributable to reduced dietary fat should be very rare. Therefore, the sample size needed would be very high if the power is 90%.

3.(10 pts) Suppose the study in question #2 is studying the use of a synthetic fat substitute instead of dietary modification ((i.e. randomize patients to synthetic fat substitute vs eat normally). Does your opinion of alpha=.05 and beta=.1 change? Why or why not?

Yes. Because synthetic fat is known to be an important risk factor to CVD and thus mortality in the literature. Therefore, 90% power is reasonable.

4. (20 pts) A randomized trial will compare the means of the two equally sized treatment groups using a t-test. How many subjects are required to detect a difference of .5 standard deviations using alpha=.05 (two sided) and beta=.2? How does the sample size change if twice as many subjects are randomized to one arm vs the other?

64 subjects per group is needed for difference of .5 and standard deviation of 1.

If twice as many subjects are randomized to one arm vs the other, the sample size will be larger.

*R code:*

*power.t.test(delta=0.5, power=0.8, sig.level = 0.05, type="two.sample")*

5a. (10pts )Suppose that investigators wish to detect the difference between the proportion of successes of p1=.4 and p2=.55 using equal treatment groups with alpha=.05 (two sided) and beta=.1. How large must the trial be?

231 subjects per group should be recruited.

*R code:*

*power.prop.test(power=0.9, p1=0.4, p2=0.55, sig.level=0.05)*

5b.(10pts) Suppose that the allocation ratio is 1.5:1 instead of 1:1. Then what is the new sample size?

The sample size should be 204 and 305.

*R code:*

*library(EnvStats)*

*propTestN(p.or.p1 = 0.4, p0.or.p2 = 0.55, alpha=0.05, power=0.9, ratio = 1.5, sample.type = "two.sample", alternative = "two.sided")*

6. (10pts) Investigators will compare survival on two treatments using the logrank statistic. What sample size is required to detect a hazard ratio of 1.5, assuming no censoring?

The sample size would be 209 with 90% power and significance level=0.05.

*R code:*

*alpha=0.05*

*beta = 0.1*

*hr = 1.5*

*zalpha=qnorm(alpha)*

*zbeta=qnorm(beta)*

*num=4\*(zalpha + zbeta)^2*

*denom = (log(hr))^2*

*num/denom*

7. (20pts) Five-year survival on standard treatment is 40%. If a new treatment can improve 5 yr-survival to 60% what size trial would be required to demonstrate it using alpha=.05 and beta=.1?

The sample size required is 101.

*R code:*

*alpha=0.05*

*beta = 0.1*

*hr = (-log(0.4)/5)/(-log(0.6)/5) # hazard ratio*

*zalpha=qnorm(alpha)*

*zbeta=qnorm(beta)*

*num=4\*(zalpha + zbeta)^2*

*denom = (log(hr))^2*

*num/denom*

8.Researchers are interested in the relationship between the use of oral contraceptives (OC) and systolic blood pressure (SBP) in women 35-44 years of age. Assume that the true SBP distributions of OC and non-OC users are both normal with a common standard deviation of 17 mmHg.

(a)  (10pts) Researchers wish to detect a difference of least 5 mmHg in SBP between the mean SBP of OC users and the mean SBP of non-OC users, with OC users having the higher mean SBP. Determine the number of women needed in each group to achieve 90% power if a two-sided test will be used with a 1% significance level. Assume equal allocation.

346 subject in each group is needed.

*R code:*

*power.t.test(delta=5, sd=17, power=0.9, sig.level = 0.01, type="two.sample")*

(b)  (10pts) Comment on the magnitude of the sample size. If the researchers cannot afford a sample of this size, what can you suggest?

The sample size might be too large to recruit.

I would suggest tuning down the power, or use an unequal allocation with more subjects in the control group, so that the recruitment is easier.

(c)  (10pts) Suppose 100 OC users and 100 non-OC users are available for study and a true difference in mean SBP of 5 mmHg is anticipated, with OC users having the higher mean SBP. Using α = 0.01, how much power will the study have?

The power would be 0.3038.

*R code:*

*power.t.test(n=100, delta=5, sd=17, sig.level = 0.01, type="two.sample")*

9. Click on the link below to download a paper by Hoenig and Heisey (2001) published in *The American Statistician*

<https://www.vims.edu/people/hoenig_jm/pubs/hoenig2.pdf>

Answer the questions below briefly.

(a) (10pts) Study Figure 1. What does this figure suggest? (b) What is meant by posthoc power calculation?

Figure 1 suggests that there is a one-to-one relationship between p values and observed power, and that nonsignificant p values always correspond to low observed power (smaller than 50%).

Posthoc power calculation means to calculate the statistical power from the observed value of the test statistic after the experiment given that a statistically nonsignificant results was obtained.

(b) (10pts) Why do posthoc power calculations not help in interpreting the study results?

Because the posthoc power will always be low for nonsignificant results, it could never lead to a statement such as “because the null hypothesis could not be rejected and the observed power was high, the data support the null hypothesis.”