

Reading Assignment: Collett, Chapters 1, 2 and 5

Written Assignment: Due Tuesday, September 27, in class

1. The following data were taken from Table 3.11 on page 73 in the book, Categorical Data Analysis, by Alan Agresti.

	Cancer Controlled	Cancer Not Controlled
Surgery	21	2
Radiation Therapy	15	3

Assume that the 41 larynx cancer patients were randomly assigned to the two treatments. Use Fisher's exact test to test the null hypothesis that the two treatments are equally effective in controlling cancer against the alternative that the treatments are not equally effective. Report a p-value and state your conclusion.

2. In a study of the effects of treating multiple sclerosis patients with human fibroblast interferon (IFN-B) (reported by Jacobs, O'Malley, Freeman, and Ekes (1981), Science, 214, pp. 1026-1028), 20 multiple sclerosis patients were randomly divided into a group of 10 IFN-B recipients and a group of 10 controls. At the beginning of the study the severity of each patient's symptoms was evaluated and at the end of the study each patient was reevaluated and classified as either improved, unchanged, or worsened. The data are given in the following table.

	Result of Treatment			TOTALS
	Improved	Unchanged	Worsened	
Treated with IFN-B	5	4	1	10
Controls	1	4	5	10

Perform an "exact" randomization test of the null hypothesis that the IFN-B treatment produces the same results as the treatment given to the controls against the null hypothesis that the IFN-B treatment gives better results. Using whatever criterion you think is best to order the tables, report the possible tables that are less consistent with the null hypothesis than the observed table. Compute the p-value for your test and state your conclusion.

3. In a clinical trial to study the possible effects of soy isoflavones on preventing bone loss in post-menopausal women, the plan is to enroll 240 women in the trial. Each woman who volunteers for the study will be examined to determine if her medical history satisfies the criterion for entry into the study. These criteria are not described here, but one example is that women with a history of diabetes will be excluded from the study. Women will have to satisfy other criteria such as time from onset of menopause that are also not discussed here. Women who satisfy the entry criteria and agree to participate will be randomly assigned to one of three treatment groups. Women in one treatment group will each day take a pill containing 20 mg of the isoflavones. Women in a second treatment group will each day take a pill containing 10 mg of isoflavones, and women in the third group will each day take a pill containing no isoflavones (a placebo). Bone density measurements will be taken at entry into the study (baseline) and at three-month intervals for four years. The women in this study will be recruited during a two-year period. Starting just a few women each week during the recruitment period will keep the cost of the study low by allowing the work of measuring bone density, analyzing blood samples, and taking other measurements to be accomplished by a relatively small group of medical professionals and one set of equipment. The researchers are concerned, however, that seasonal variation in weather and other conditions could affect some of the measurements. Furthermore, 120 women will be recruited and measured in central Iowa and the other 120 women will be recruited and measured in central California. These two populations will differ with respect to climate and environment and other factors such as racial composition and consumption of soy.
 - A. Suppose the researchers come to you for advice on how to randomly assign women to the three treatment groups. Describe the randomization protocol that you would recommend. Give enough detail, so the researchers would know how to carry out your plan. Comment on the strengths and weakness of your protocol.
 - B. Some participants in a study may not completely adhere to the treatment instructions. For example, they may simply forget to take some pills, or they may stop taking pills because they feel ill. In this study a woman was classified as adherent if she took at least 80 percent of the pills issued to her over the course of the study. Otherwise, she was classified as nonadherent. Adherence was checked by requiring participating women to bring any unused pills to their check ups. It was found that 34 women were nonadherent. To avoid complications with the issue of dropouts, assume the nonadherent women did come in for all of their scheduled check ups, so you have data on their response to treatment. Should those 34 women be included in the analysis of the data? Describe potential advantages and disadvantages of using an “intent to treat” policy for which all nonadherent women are included in the analysis. Describe possible advantages and disadvantages of excluding nonadherent women for the analysis of the data.
 - C. Suppose it was found that 22 women who did not meet the criteria for participation in the study had been enrolled in the study. This could happen, for example, because researchers are careless in examining results of baseline

measures or randomization to treatments is done before some baseline information is obtained from the lab or incorrect information was given to the researchers at baseline. Suppose the women want to stay in the study and the physicians on the safety monitoring board rule that those women are not subject to undo risk of serious harm to their health and they can stay in the study. Should you keep those women in the study or exclude them? As in part B, comment of potential advantages and disadvantages.

- D. Consider the situation in part C, but assume that the physicians on the safety monitoring board are concerned by increased risk of potential adverse health effects and order you to immediately stop treating those 22 women (take away any unused pills) who violated the criteria for entry into the study. As a safety measure, you are also ordered to continue to have those women come in for their scheduled check ups under the study protocol. These women will not be adherent because they are no longer being treated, but you will have measurements on the outcome variable, bone-loss. Should you include the bone loss information from those women in the analysis of the effects of treatments on bone loss?

4. This problem asks you to run some simulations to examine the use of a randomization test and an approximate F-test to make inferences about differences in three treatment means for a randomized clinical trial, where randomization to the four treatments is done within blocks of four subjects. You will be asked to consider different numbers of blocks and examine how well an F-test can approximate a randomization test as the number of blocks (sample size) increases. Note that the responses are not assumed to follow a normal distribution or any other distribution. The justification for using an F-test follows solely from how well the randomization distribution of the F statistic is approximated by the distribution of a central F random variable.

- A. Using the code posted as `randomhw.sas` or `randomhw.R` to perform the simulations, your first task is to generate sets of 10000 F-values for different possible random assignments of subjects to treatments within blocks, for different number of blocks. In this simulation the null hypothesis of no treatment effects is true and you are to examine how well the 95-th percentiles of central F-distributions approximate the 95-th percentiles of the randomization distribution. Each student should run his or her own simulations using unique seeds to initialize the random number generator. Each line of the data matrix used in these programs represents the outcomes for the subjects in one block. Run this simulation using just 2 blocks (only the first two lines of data), just 4 blocks (only the first four lines of data), just 6 blocks, just 8 blocks, and all 10 blocks. Complete the following table, where b is the number of blocks:

Number of blocks	95-th Percentile of the Simulated F-Values	95-th Percentile of a Central F Distribution with (3, 3(b-1)) df.	Ratio: Column 2 divided by Column 3	Proportion of Simulated F-values that exceed the value in Column 3
2				
4				
6				
8				
10				

Is there a pattern in this table? If so, describe it. What did you expect to see? What are your conclusions?

- An important step for planning and determining the cost of a study is the determination of the number of subjects to use in the study. As an example, consider planning a randomized clinical trial for evaluating the effectiveness of a drug for lowering cholesterol in patients with coronary heart disease (CDH). Cholesterol is the main lipid associated with arteriosclerotic vascular disease, and lowering cholesterol levels may reduce the risk of arteriosclerotic heart disease. The liver metabolizes cholesterol to its free form, which is transported in the bloodstream by lipoproteins. Approximately 75% of the cholesterol is bound to low density lipoproteins (LDLs) and the remaining 25% is bound to high density lipoproteins (HDLs). Therefore, cholesterol is the main component of LDLs and a smaller component of HDLs and very low density lipoproteins, and LDL is strongly associated with increased risk of CHD.

A pharmaceutical company is interested in conducting a randomized clinical trial to compare a new cholesterol lowering drug to a widely used drug (gold standard) for lowering cholesterol in patients with CDH. In this study, n_1 patients will be randomly assigned to the gold standard and n_2 patients will be randomly assigned to the new drug. For this study it is reasonable to assume that each patient responds independently of any other patient. The response of primary interest is reduction in plasma LDL level. Let μ_1 and μ_2 denote the population means and let σ_1^2 and σ_2^2 denote the population variances of reductions in LDL levels for patients treated with the gold standard and the new drug, respectively. Further assume that the variances are homogeneous, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Let \bar{X}_1 and \bar{X}_2 denote the sample means and let s_1^2 and s_2^2 denote the sample variances for the reductions in LDL

levels for patients treated with the gold standard and the new drug, respectively.

Let $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$ denote the pooled estimate of the common variance.

- A. Assuming the reductions in LDL levels are approximately normally distributed for each drug, the null hypotheses of equal mean LDL reductions ($H_0 : \mu_1 = \mu_2$) and be tested against the one-sided alternative that the new drug provides a larger mean reduction ($H_a : \mu_1 < \mu_2$) by rejecting the null

hypotheses if $t = \frac{\bar{X}_2 - \bar{X}_1}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{(n_1+n_2-2),\alpha}$, where $t_{(n_1+n_2-2),\alpha}$ is the

$(1 - \alpha) \times 100$ th percentile of the central t-distribution with $n_1 + n_2 - 2$ degrees of freedom. Under the alternative where the actual positive difference in response means of $\delta = \mu_2 - \mu_1$, the power of the test is

$$\begin{aligned}
 \text{power} = 1 - \beta &= \mathbb{P} \left(\frac{\bar{X}_2 - \bar{X}_1}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{(n_1+n_2-2),\alpha} \right) \\
 &= \mathbb{P} \left(\bar{X}_2 - \bar{X}_1 > t_{(n_1+n_2-2),\alpha} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \\
 &= \mathbb{P} \left((\bar{X}_2 - \bar{X}_1) - \delta > t_{(n_1+n_2-2),\alpha} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} - \delta \right) \\
 &= \mathbb{P} \left(\frac{(\bar{X}_2 - \bar{X}_1) - \delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{t_{(n_1+n_2-2),\alpha} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} - \delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) \\
 &= \mathbb{P} \left(t_{(n_1+n_2-2)} > t_{(n_1+n_2-2),\alpha} - \frac{\delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right)
 \end{aligned}$$

where $t_{(n_1+n_2-2)}$ denotes a random variable with a central t-distribution with

$n_1 + n_2 - 2$ degrees of freedom. For a particular choice of λ , set $n_1 = \lambda n_2$ and find the value of n_2 that corresponds to specific values of α and $\beta=1$ -power. You can use a computer to do this (see code posted on the assignment section of the course web page as `power2t.sas` or `power2t.R`), or you can perform the calculations with a calculator by first finding the approximate solution

$$\tilde{n}_2 = \frac{(z_\alpha + z_\beta)^2 s^2 (1 + \lambda^{-1})}{\delta^2}$$

where z_α and z_β are percentiles of the standard normal distribution. Then computing the required sample sizes as

$$n_2 = \frac{(t_{(\lambda\tilde{n}_2 + \tilde{n}_2 - 2), \alpha} + t_{(\lambda\tilde{n}_2 + \tilde{n}_2 - 2), \beta})^2 s^2 (1 + \lambda^{-1})}{\delta^2} \quad \text{and} \quad n_1 = \lambda n_2.$$

Using $n_1 = n_2 = n$ and $\alpha = .05$, find the value of n that provides power=0.8 for rejecting the null hypothesis when the alternative is $\delta = \mu_2 - \mu_1 = 5$ and the standard deviation is approximately 2δ .

- B. Consider the same study, assume that the researcher wants to show that the two drugs yield different mean reductions in LDL, that is, the researcher wants to test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the two-sided alternative $H_a : \mu_1 \neq \mu_2$. Using $n_1 = n_2 = n$ and $\alpha = .05$, find the value of n that provides power=0.8 for rejecting the null hypothesis when the alternative is $\delta = 5$ and the standard deviation is approximately 2δ .
- C. Sometimes the objective of a study is not to show that a new drug is superior to the best drug in current use (a gold standard), but to show that the new drug is equivalent to the gold standard in effectiveness. This would be sufficient justification for introducing the new drug if the new drug was cheaper or had fewer serious side effects than the gold standard. A traditional test hypothesis test, like those considered in parts A and B, are not appropriate for this purpose. Failing to reject the null hypothesis does not necessarily justify the conclusion that the drugs are equally effective. Suppose the new drug in this study is considered to be essentially equivalent in effectiveness to the gold standard if the new drug provides a mean reduction in LDL within θ of the mean reduction of LDL provided by the

gold standard. This leads to a test of the null hypothesis $H_0 : |\mu_1 - \mu_2| \geq \theta$, the drugs are not equivalent, against the alternative $H_a : |\mu_1 - \mu_2| < \theta$, the drugs are equivalent. Assuming independent observations, homogeneous variances and approximate normality for the distribution of LDL reductions under each drug, derive a formula for the power of the test when $n_1 = \lambda n_2$, $\alpha = .05$, the actual difference in the response means is $\delta = \mu_2 - \mu_1$, and the common standard deviation is σ .

- D. Using your result in part C with $n_1 = n_2 = n$ and $\alpha = .05$, find the value of n that provides power=0.8 for rejecting the null hypothesis of non-equivalence when the actual difference in the response means is $\delta = \mu_2 - \mu_1 = \theta/2$ and the standard deviation is approximately $2\delta = \theta$.
- E. Sometimes one is more interested in accurately estimating the difference in the response means for two drugs. Then the sample size can be determined by specifying the size of the desired standard error for the estimate of the difference in the response means or by specifying the desired length of a confidence interval for the difference in response means. Find the sample size needed so that the half length of a 95% confidence interval for the difference in the means will be about 0.5 when the standard deviation is 5.0.
6. Data are available from a pilot study that compared the effectiveness of a new compound and a standard therapy for preventing relapse in patients with schizophrenia and schizoaffective disorders. In that study, 24 of the 80 subjects who were randomly assigned to treatment with the new compound experienced relapse and 30 of the 80 subjects who were randomly assigned to the standard therapy experienced relapse. Use this information to determine the sample size required for a larger study to compare relapse probabilities for the new compound and the standard treatment. Assume that $n_1 = n_2 = n$ patients will be treated with the new compound and the standard therapy, respectively and that each test will be performed with a Type I error level of $\alpha = 0.05$. Let π_1 denote the proportion of this population of patients with schizophrenia and schizoaffective disorders who would experience relapse when treated with the new compound, and let π_2 denote the proportion of this population who would experience relapse under the standard therapy.
- A. Use the information from the pilot study to estimate π_2 / π_1 , the relative risk of relapse for the standard therapy relative to treatment with the new compound. Also construct an approximate 95% confidence interval for the relative risk of relapse.

- B. Consider a new study in which n patients will be randomly assigned to treatment with the new compound and the other n patients will receive the standard therapy. The objective is to demonstrate that the new compound reduces the incidence of relapse. Consequently, the researchers want to test $H_0 : \pi_1 = \pi_2$ against the alternative $H_a : \pi_1 < \pi_2$. Determine the sample size needed in order to achieve power of at least 0.80 when the relative risk (π_2 / π_1) is 1.20. If you can, present a formula in addition to a numerical value for the sample size.
- C. Suppose that the new compound and the standard therapy are considered to be equivalent if the absolute value of the logarithm of the relative risk does not exceed 0.1. Determine the sample size needed in order to achieve power of at least 0.8 for establishing equivalence when $\log(\pi_2 / \pi_1) = 0.05$. In making this determination, assume that the relapse rate for the standard therapy is equal to the observed rate in the pilot study ($30/80 = 0.375$). Present a formula in addition to a numerical value for the sample size.
- D. Suppose the objective is to construct a 95% confidence interval for the relative risk of relapse for the new therapy versus the standard treatment, such that the length of the confidence does not exceed 5% of the actual value of the relative risk. If it is possible to do this, determine the required sample size. Assume $n_1 = n_2 = n$.

(Note: Answers to parts A, B, C and D will depend on assumptions for the values of π_1 and π_2 that are used to evaluate variances. For example, the variance of the difference of two proportions will be largest, leading to conservative sample size determination, when π_1 and π_2 are both taken to be close to 0.5. Useful code is posted as `power2p.sas`, `power2e.sas`, `power2p.R` and `power2e.R` on the assignment section of the course web page.)