140.652 - Lab 13

HW5 Problem 11

Problem 11. Suppose that 18 obese subjects were randomized, 9 each, to a new diet pill and a placebo. Subjects' body mass indices (BMIs) were measured at a baseline and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the baseline (followup - baseline) was $-3\ kg/m^2$ for the treated group and $1\ kg/m^2$ for the placebo group. The corresponding standard deviations of the differences was $1.5\ kg/m^2$ for the treatment group and $1.8\ kg/m^2$ for the placebo group. Does the change in BMI over the two year period appear to differ between the treated and placebo groups? Perform the relevant test and interpret. Give a P-value. Assume normality and a common variance.

Assumptions

We let D_T , D_C to be the four week difference in BMI in the treatment and control groups, respectively. Assume:

- $D_T \sim N(\mu_T, \sigma^2)$
- $D_C \sim N(\mu_C, \sigma^2)$

Above states we assume common variance in distribution of D_T and D_C .

Test formulation

Our null and alternative hypotheses are then:

$$H_0 : \mu_T = \mu_C$$
$$H_a : \mu_T \neq \mu_C$$

Since we assume common variance, we use the pooled variance estimate for calcualting our T statistic.

Background: test statistic in t-test for difference between two groups, assuming equal variance

Recall (Lecture 10 slide 5) the formula for pooled variance estimator of variance parameter σ^2 :

$$S_p^2 = \left\{ (n_x - 1)\, S_x^2 + (n_y - 1)\, S_y^2 \right\} / \left(n_x + n_y - 2 \right)$$

Then from the lecture (Lecture 10 slide 8) that the test statistic

$$TS = \frac{\frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{\sigma \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}}}{\sqrt{\frac{(n_x + n_y - 2) S_p^2}{(n_x + n_y - 2)\sigma^2}}} == \frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}}$$

is a standard normal divided by the square root of an independent Chi-squared divided by its degrees of freedom, hence, it follows t distribution with $n_x + n_y - 2$ degrees of freedom.

- Under the null hypothesis $\mu_T = \mu_C$ we have $\mu_T - \mu_C = 0$ and

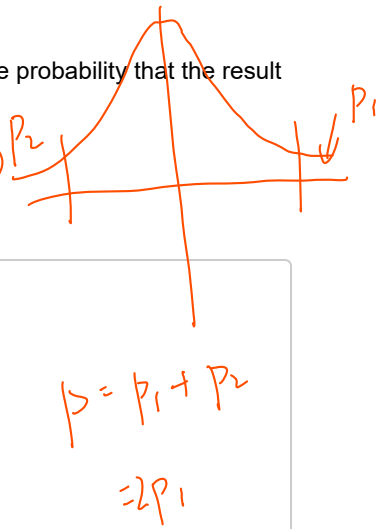
$$TS = \frac{\overline{Y} - \overline{X}}{S_p^{1/2}}$$

- % confidence interval for $\mu_y - \mu_x$ is $(1-\alpha) \times 100$

$$\overline{Y} - \overline{X} \pm t_{n_x+n_y-2,\ 1-\frac{\alpha}{2}} \cdot S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- Denote t - observed value of test statistic TS . Given we work with two-side t -test, the probability that the result is at least as extreme as the one we observe is

$$P(TS \geq |t|) + P(TS \leq -|t|) = 2 \cdot P(TS \geq |t|).$$



Result

```
alpha <- 0.05

n_T <- 9
n_C <- 9

mu_T <- -3
mu_C <- 1

sd_T <- 1.8
sd_C <- 1.5

## pooled variance
Sp2 <- ((n_C-1) * sd_C^2 + (n_T-1) * sd_T^2) / (n_T + n_C - 2)
Sp <- sqrt(Sp2)

## CI
(mu_C - mu_T) + c(-1, 1) * qt(1 - alpha/2, n_T + n_C - 2) * Sp * sqrt(1/n_T + 1/n_C)
```

```
## [1] 2.344301 5.655699
```

```
## observed value of test statistic TS
ts <- (mu_C - mu_T) / (Sp * sqrt(1/n_T + 1/n_C))
ts
```

```
## [1] 5.121475
```

```
## p-value
2 * pt(abs(ts), n_T + n_C - 2, lower.tail = FALSE)
```

```
## [1] 0.0001025174
```

- We performed two-sided t -test verifying $H_0 : \mu_T = \mu_C$ aganist $H_a : \mu_T \neq \mu_C$. Test statistic was distributed with t distribution with 16 degrees of freedom. We obtained observed test statistic value equal 5.1214752.
- We obtained p-value < 0.05 , thus at significance level $\alpha = 0.05$ we reject the null hypothesis that change in BMI is the same in the two groups.
- We obtained 95 % confidence interval $[2.344,\ 5.656]$ which does not contain 0 .
- Note that the sign of ts and the sign of limits of 95 % confidence interval suggesting the difference of BMI change (control BMI change - treatment BMI change) is positive.

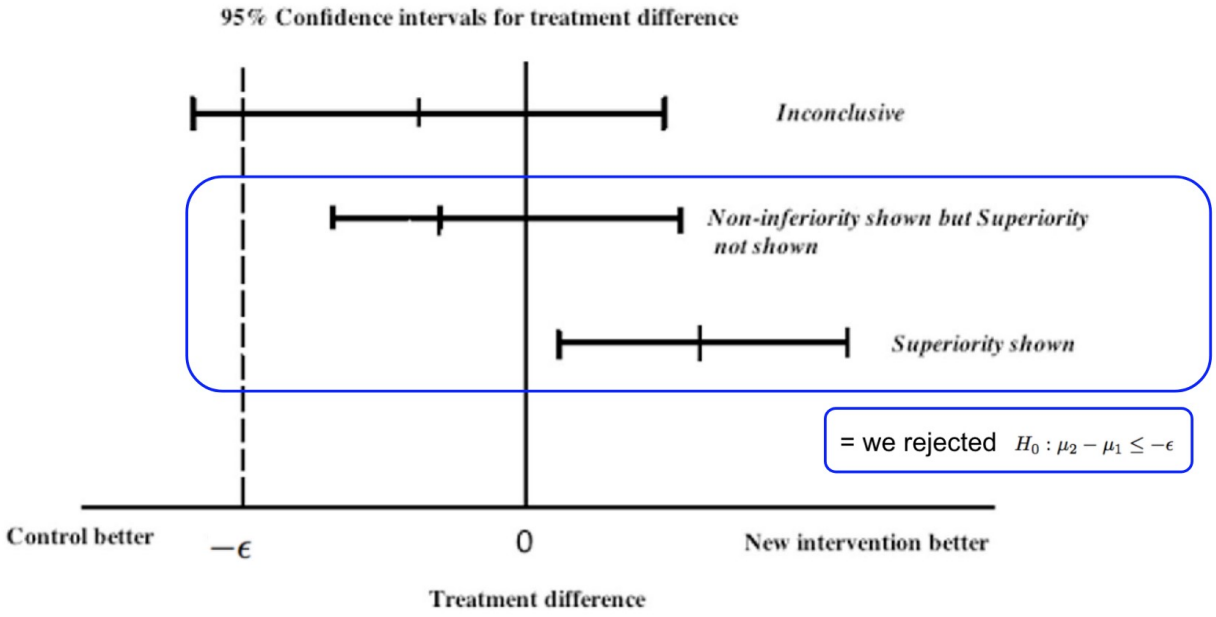
Book p 472 Problem 7

Problem 7. Sometimes proving superiority of a new drug may not be possible. Therefore, in practice one sometimes uses a test for the null hypothesis of non-inferiority. For the two-sample the non-inferiority hypothesis is

$$H_0 : \mu_2 - \mu_1 \leq -\epsilon \quad \text{versus} \quad H_A : \mu_2 - \mu_1 > -\epsilon \,,$$

where ϵ is a small positive constant. The idea is that under the null the mean of group 2, μ_2, is inferior to the mean in group 1, μ_1 by a margin ϵ. That is $\mu_2 \leq \mu_1 - \epsilon$. For an α level and power $1 - \beta$ derive the sample size necessary to detect an alternative. Use the Normal test and t-test for equality of means.

Illustration (source link (https://www.semanticscholar.org/paper/Understanding-the-null-hypothesis-(H0)-in-trials-Mallat/3320fb1ebe8314cddf64e2d2db68a31e72854e2b/figure/0), after modifications):



Note: solution is provided based on article Methods for Equivalence and Noninferiority Testing (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2701110/).

More on interpretation

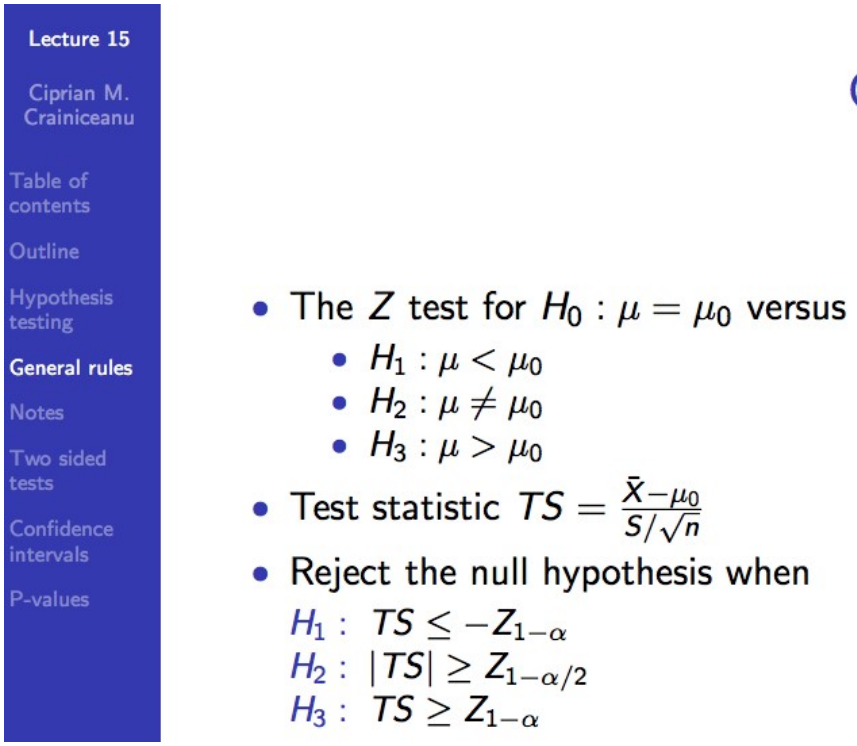
- $\epsilon > 0$ is the **noninferiority margin of clinical interest**.
- Here, the null hypothesis is stating that the new drug treatment group mean is at least an amount ϵ worse than the control mean, whereas the alternative hypothesis that we want to prove is stating that the treatment is noninferior.
- Noninferiority here means that the treatment group mean μ_2 can be no worse than ϵ lower than the control group mean μ_1 .
- Note that **noninferiority hypotheses require that one define a noninferiority margin** $\epsilon > 0$. This noninferiority margin should be based on clinical judgment and it has the interpretation of how close the new treatment must be to the control to be considered noninferior. The choice of the margin ϵ is a critical issue and a difficult task.
- The noninferiority hypothesis above apply when we define the means μ_1 and μ_2 as a “good outcome” (meaning: positive value is “good”).

More on formulation

- Note here H_0 is stated as **inequality**: $H_0 : \mu_2 - \mu_1 \leq -\epsilon$. It is different than **equality** we use in the classical approach, where it necessary to have parameter assumed at a particular value (e.g. $H_0 : \mu_2 - \mu_1 = -\epsilon$) to work out the test statistic distribution and probabilities.
- However, one can rephrase the above H_0 onto $H_0 : \mu_2 - \mu_1 = -\epsilon$; it is because using $\mu_2 - \mu_1 = -\epsilon$ for the null hypothesis is a "hardest case scenario" to reject out of different cases within $\mu_2 - \mu_1 \leq -\epsilon$; that is, if you reject $H_0 : \mu_2 - \mu_1 = \mu_0$ for some μ_0 and $H_a : \mu_2 - \mu_1 > \mu_0$, **then you would also** reject $H_0 : \mu_2 - \mu_1 = \mu_0^*$ for some $\mu_0^* < \mu_0$.

Solution: test statistic and rejection rule

- Recall general rule:



- Recall we reprhased our test onto

$$H_0 : \mu_2 - \mu_1 = -\epsilon$$
$$H_A : \mu_2 - \mu_1 > -\epsilon$$

- Consider Normal test. Assume common error variance in control and treatment group is questionable (do not use pooled variance). Then we infer that the Normal test statistic is of a form

$$z = \frac{\overline{X}_2 - \overline{X}_1 + \epsilon}{\sqrt{S_2^2/n_2 + S_1^2/n_1}} \sim N(0,1)$$

We reject H_0 at significance level α (noninferiority is concluded at significance level α) if $z > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$ -percentile of a standard normal distribution.

Solution: sample size computation

- Recall: for "classic" two-sample Normal test, if the sample sizes are equal, i.e., $n_1 = n_2 = n$, we compute sample size using the formula

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 2 \cdot \sigma^2}{\Delta^2}$$

where $1 - \beta$ is the power, and Δ is the "true difference" one would like to detect, is σ^2 is the population variance.

- Power has a **slightly different interpretation** when we are doing noninferiority testing than when we are performing a traditional hypothesis test.
    - In the classical hypothesis testing framework, power typically refers to the probability that we correctly conclude the treatment and control are different (in two-side test), or that we correctly conclude the treatment is better than control (in one-side test).
    - In the noninferiority setting, power typically refers to the probability that we correctly conclude the treatment is noninferior, when it really is noninferior.

- Here, power depends both on the noninferiority margin ϵ and the true difference in proportions Δ , between the treatment and control arms.
- To evaluate power when the treatment is considered noninferior, **we usually assume that the treatment and control have truly the same outcomes, so that** $\Delta = 0$.
- One can show that in the balanced case with $n_1 = n_2 = n$, the sample size required to have a prespecified power $1 - \beta$ is given by

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 2 \cdot \sigma^2}{\epsilon^2}$$

(compare with derivations for the usual testing problem).

- Note that this formula looks similar to the sample size formula for the usual testing problem, except that the true difference, Δ used in the usual formulation, is replaced by the noninferiority margin ϵ . This is an important difference:
    - Typically, a margin ϵ for considering a treatment noninferior might be much smaller than true difference Δ . This narrower margin results in the generally larger sample sizes often associated with noninferiority testing.

Exercise: Book p 472 Problem 8

Problem 8. Another type of testing is that of equivalence. In this scenario we still have two groups, say of the same size, n, but we are interested in testing the equivalence of two treatments/drugs. In this case we are testing

$$H_0 : |\mu_2 - \mu_1| \geq \epsilon \quad \text{versus} \quad H_A : |\mu_2 - \mu_1| < \epsilon \, .$$

The idea is that under the null the mean of group 2, μ_2, is substantially different from the mean in group 1, μ_1, by a margin ϵ. The alternative is that the difference between the two means is small, that is the two groups/treatments are equivalent. For an α level and power $1 - \beta$ derive the sample size necessary to detect an alternative. Use the Normal test and t-test for equality of means.