JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Lecture 5b: Practice Problem Solutions

John McGready
Johns Hopkins University

Practice Problems

1. In a high school in the United States, dietary counseling is being tested to measure the program's long-term impact on student's fat intake. Of the three hundred students at the school, 150 are randomized to receive five one-hour sessions of dietary counseling; the other 150 students receive no counseling.
   - Six months after the last counseling sessions, all students are asked to keep a food diary for one week
   - Each student's average fat intake in grams is calculated at the end of this week
   - The results of this exercise are as follows:

Practice Problems

- Intervention group
  - $\bar{x}_1$ = 54.8 grams , s1 = 28.1 grams, n1 = 146
- Control group
  - $\bar{x}_2$ = 62.8 grams, s2 = 34.7 grams, n2 = 142
  - (Please note—follow up sample sizes differ slightly from initial sample size because of loss to follow up)
- The public-health question of interest is whether there is a difference in mean fat intake between the two groups, six months after the intervention ended. You are going to help answer this question:
  - Compute a p-value for testing the null of no association between counseling and average fat intake. Is this consistent with the confidence interval estimated in the section A problems?

Practice Problems

- To create a p-value for testing
  - $H_o : \mu_2 - \mu_1 = 0$
  - $H_A : \mu_2 - \mu_1 \neq 0$

- We need to measure how far the sample mean difference of $\bar{x}_2 - \bar{x}_1 = 8\ grams$ is from the null value of 0 in terms of standard errors

- Recall: $$\hat{SE}(\bar{x}_2 - \bar{x}_1) = \sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}$$

$$= \sqrt{\frac{(28.1)^2}{146} + \frac{(34.7)^2}{142}} \approx 3.7\ grams$$

Practice Problems

- So the distance measure t is given by:

$$t = \frac{8\ grams}{3.7\ grams} \approx 2.2$$

- So we have a result that is 2.2 standard errors above the mean of 0 on a normal curve (the sampling distribution is normal by the CLT)
  - This would result in a p-value such that $0.01 < p < 0.05$

- So the result is statistically significant at the α=.05 level, which is consistent with the 95% CI not including 0 from the last section problems

- To get an exact p-value we can appeal to the ttesti command in Stata

Practice Problems

- Stata results:

```
. ttesti 142 62.8 34.7  146 54.8 28.1

Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     142        62.8    2.911959        34.7    57.04326    68.55674
       y |     146        54.8    2.325573        28.1     50.2036     59.3964
---------+--------------------------------------------------------------------
combined |     288    58.74444    1.869474    31.72603    55.06483    62.42406
---------+--------------------------------------------------------------------
    diff |                   8    3.715851                .6861162    15.31388
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =   2.1529
Ho: diff = 0                                     degrees of freedom =      286

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9839         Pr(|T| > |t|) = 0.0322          Pr(T > t) = 0.0161
```

Practice Problems

- Notice, direction of comparison is arbitrary: control vs. intervention in Stata gives same results, in opposite direction

```
. ttesti 146 54.8 28.1 142 62.8 34.7

Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     146        54.8    2.325573        28.1     50.2036     59.3964
       y |     142        62.8    2.911959        34.7    57.04326    68.55674
---------+--------------------------------------------------------------------
combined |     288    58.74444    1.869474    31.72603    55.06483    62.42406
---------+--------------------------------------------------------------------
    diff |                  -8    3.715851               -15.31388   -.6861162
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =  -2.1529
Ho: diff = 0                                     degrees of freedom =      286

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0161         Pr(|T| > |t|) = 0.0322          Pr(T > t) = 0.9839
```