MT2-Fill-in-Blank

Thursday, April 15, 2021 2:43 PM

Exam 2
Practice Fil…

Problem Information:

Sample Mean Difference (f vs. m): 12.03-15.90 = -3.87
Female Smoker Sample Size: 234
Male Smoker Sample Size: 187
Test Statistic -4.11
P-value: 0.000059

Reference the Population Mean Difference Part of the Exam 2 Practice Test to Fill in the Blanks Below

A. Problem Statement

We would like to test the following hypotheses:

$$H_0: \mu_F - \mu_M = 0$$
$$H_A: \mu_F - \mu_M \neq 0,$$

Where:

- μ_F is the average number of weekend cigarettes ALL female smokers in the U.K. smoke, and
- μ_M is the average number of weekend cigarettes ALL male smokers in the U.K. smoke

In order to test this, we need to know more about the sampling distribution of

(1) Sample mean differences.

B. Actual Sampling Distribution Creation

If we wanted to create this sampling distribution by hand, that would help us test these hypotheses we would need to do the following.

1. Collect *M* random samples (2) WITH (WITH/WITHOUT)replacement from the population distribution of **female** U.K. smokers.
2. Then calculate the (3) Sample mean of each of these random samples and put them in a list.
3. Collect *M* random samples (2) WITH (WITH/WITHOUT) replacement from the population distribution of **male** U.K. smokers.
4. Then calculate the (3) Sample mean of each of these random samples and put them in a list.
5. Finally, to create the sampling distribution we would subtract the values in both lists.

C. Theoretical Sampling Distribution

HOWEVER, we don't actually need to create this sampling distribution above, because we know the following things about this sampling distribution.

1. The mean of this sampling distribution is approximately (4) $\mu_F - \mu_M$.
2. The standard deviation of this sampling distribution (aka the standard error) is approximately (5) $\sqrt{\frac{\sigma_F^2}{n_F} + \frac{\sigma_M^2}{n_M}}$.
3. Because the following (6) Central Limit Theorem conditions for (1) Sample mean differences below hold,

1. The sample size of females is n = 234 >30 (we could have also tried to check if the sample/population distribution of the number of cigarettes smoked by females on the weekends is normally distributed).
2. The sample of UK females is randomly collected.
3. The sample size n=234 <10% of the UK female smoker population.
4. The sample size of males is n = 187 >30 (we could have also tried to check if the sample/population distribution of the number of cigarettes smoked by males on the weekends is normally distributed).
5. The sample of UK males is randomly collected.
6. The sample size n=187 <10% of the UK males smoker population.
7. We should also check that the males and females were collected independently in this study. (For instance, if the males and females in this sample were married to to eachother, then the male and female respondents would not be indenpendent of each other).

a) _____.

Hypotheses

Ho: Null Hypothesis

Ha: Alternative Hypothesis

The General Definition of a p-value is:

P(observed sample statistic that is at least as suspicious (in favor of the alternative hypothesis) as **the sample statistic we actually observed** *assuming* **null hypothesis is true**)

D. What does a p-value really mean?

Because the sampling distribution of (1) Sample mean differences is (7) normal, we are able to calculate the p-value which represents

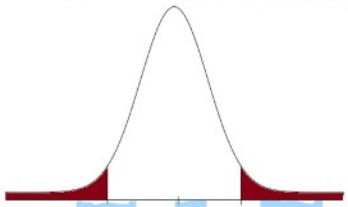
p-value = P((1) Sample mean difference that are at least as suspicious (of the alternative hypothesis) as (8) -3.87 *assuming that* (9) $H_0: \mu_F - \mu_M = 0$

For this problem, a (1) Sample mean difference that is at exactly as suspicious (of the alternative hypothesis as (8) -3.87 is (10) 3.87.

E. Calculating the p-value

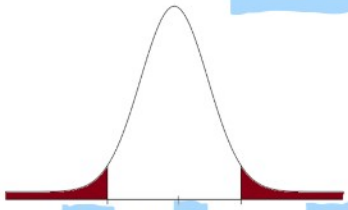
We will assume $H_0: \mu_F - \mu_M = 0$ in this entire box.

Sampling Distribution of (1) Sample mean differences



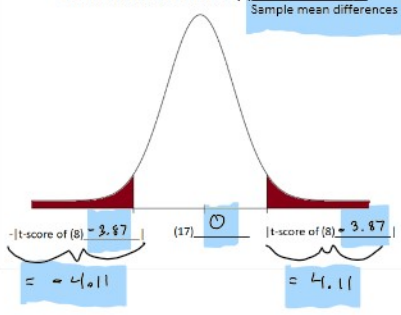
Furthermore, because the sampling distribution of (1) Sample mean differences is (7) normal, the distribution of the z-scores of the (1) Sample mean differences follows the (12) Standard normal distribution, which we know has a mean of (13) 0 and a standard deviation of (14) 1.

Distribution of the Z-scores of (1) Sample mean differences



Finally, because the sampling distribution of (1) Sample mean differences is (7) normal, the distribution of the **t-scores** of the (1) Sample mean differences follows the (15) t-distribution distribution, with (16) $\min(n_F-1, n_M-1)$ degrees of freedom which we know is centered at (17) 0.

Distribution of the t-scores of (1) Sample mean differences



The t-score of (8) = (18) -4.11.

Interpreting the p-value

The p-value is the area represented in red (in all three plots to left). → 0.000059

Therefore, because our p-value that we calculated (19) IS (IS NOT) smaller than $\alpha = 0.01$, it (20) DOES (DOES/DOESN'T) make us suspicious enough of our base assumption, to make us claim that there is sufficient evidence to suggest $H_A: \mu_F - \mu_M \neq 0$.