Lecture 14

Last time

We talked about the steps of hypothesis testing for population proportions, p.

Hypothesis testing for means

Example: A recent study compared different psychological therapies for teenage girls suffering from anorexia (Brian Everitt, Institute of Psychiatry, London). Each girls' weight was measured before and after a period of therapy. The variable of interest was the weight change, defined as weight at the end of the study minus weight at the beginning of the study. The weight change was positive if the girl gained weight an negative if she lost weight. The therapies were designed to aid weight gain.
In this study, 29 girls received the cognitive behavioral therapy. This form of psychotherapy stresses identifying the thinking that causes the undesirable behavior and replacing it with thoughts designed to help to improve this behavior. See the table below for the data. The weight change for the 29 girls has a sample mean of $\bar{X} = 3.00$ pounds and the standard deviation is $\sigma = 7.32$ pounds.

Girl	**Weight Change**	**Girl**	**Weight Change**	**Girl**	**Weight Change**
1	1.7	11	11.7	21	-1.4
2	0.7	12	6.1	22	-0.8
3	-0.1	13	1.1	23	2.4
4	-0.7	14	-4	24	12.6
5	-3.5	15	20.8	25	1.9
6	14.9	16	-9.3	26	3.9
7	3.5	17	2.1	27	0.1
8	17.1	18	1.4	28	15.4
9	-7.6	19	-0.3	29	-0.7
10	1.6	20	-3.7		

We will use this example to show the five steps to hypothesis testing about a mean.

Steps of hypothesis testing about a mean

Step 1: Assumptions

- The data are obtained using randomization
- Either the population distribution is normal OR the sample size is large enough that the central limit theorem can be used to say that $\bar{X}$ is approximately normal

The anorexia study is a convenience sample causing the inferences to be tentative. They are more convincing if researchers can argue that the girls in the sample representative of the population of girls who suffer from anorexia. The study did employ randomization in assigning girls to one of three therapies, only one of which (cognitive behavioral) is considered in this example.

Step 2: Hypotheses

One-sided alternative hypothesis:

$$H_o : \mu \leq \mu_o \; VS. \; H_a : \mu > \mu_o \tag{1}$$

$$H_o : \mu \geq \mu_o \; VS. \; H_a : \mu < \mu_o \tag{2}$$

Two-sided alternative hypothesis:

$$H_o : \mu = \mu_o \; VS. \; H_a : \mu \neq \mu_o \tag{3}$$

where μ_o denotes a particular value for the population mean.

For example, let μ denote the mean weight change for the population represented by the sample of anorexic girls. If the therapy has no effect then $\mu = 0$. If the therapy has a beneficial effect on weight, as the study expected, then $\mu > 0$. To test that the therapy has no effect against the alternative that it has a beneficial effect, we test $H_o : \mu \leq 0 \; VS. \; H_a : \mu > 0$. In practice, the two-sided alternative $H_a : \mu \neq 0$ is more common, to take an objective approach that can detect either a positive or negative effect on the therapy.

Step 3: Test statistic

The test statistic is the distance between the sample mean $\bar{X}$ and the null hypothesized value μ_o, as measured by the number of standard deviations between them. This test statistic is

$$Z = \frac{\bar{X} - \mu_o}{\frac{\sigma}{\sqrt{n}}} = \frac{sample\ mean - null\ hypothesis\ mean}{standard\ deviation\ of\ the\ sample\ mean} \tag{4}$$

In the anorexia study, the sample mean $\bar{X} = 3.0$ and the population standard deviation is given to be $\sigma = 7.32$. The test statistic for this example is

$$Z = \frac{\bar{X} - \mu_o}{\frac{\sigma}{\sqrt{n}}} = Z = \frac{3-0}{\frac{7.32}{\sqrt{29}}} = \frac{3}{1.36} = 2.21$$

The observed test statistic value is 2.21 meaning that our $\bar{X} = 3.0$ is 2.21 standard deviations away from $\mu_o = 0$.

Step 4: P-value

The **P-value** is the probability that the test statistic takes a value like the observed test statistic or even more extreme, if the null hypothesis is true. It is a single tail or a two-tail proability depending on whether the alternative hypothesis is one-sided or two-sided.

Hypothesis	**P-value**
$H_o : \mu \leq \mu_o \; VS. \; H_a : \mu > \mu_o$	$P(Z > \textit{observed test statistic})$
$H_o : \mu \geq \mu_o \; VS. \; H_a : \mu < \mu_o$	$P(Z < \textit{observed test statistic})$
$H_o : \mu = \mu_o \; VS. \; H_a : \mu \neq \mu_o$	$P(\lvert Z \rvert > \textit{observed test statistic})$

For the anorexia study with $H_a : \mu \neq 0$, the P-value is the two tail probability of a test statistic value farther out in each tail than the observed value of 2.21. This probability is double the single-tail probability.

Step 5: Conclusion

The conclusion of a hypothesis test reports the P-value and interprets what it says about the question that motivated the test. Sometimes this includes a decision about the validity of H_o. We reject the null hypothesis when the P-value is less than or equal to the pre-selected significance level.
In our example, the small P-value of 0.04 provides considerable evidence against the null hypothesis that the therapy has no effect. If we had pre-selected a significance level of 0.05, this would be enough evidence to reject $H_o : \mu = 0$ in favor of $H_a : \mu \neq 0$.
The alternative hypothesis states that the population mean weight change μ is not equal to zero. The positive value for the sample mean, $\bar{X} = 3.0$ suggests that $\mu > 0$. The cognitive behavioral therapy seems to be beneficial. The effect may be small in practical terms, however. The 95% confidence interval predicts that μ falls between 0.2 and 5.8 pounds.

Results of two-sided test and results of confidence intervals agree

For the anorexia study, we got a P-value of 0.036 for testing $H_o : \mu = 0 \; VS \; H_a : \mu \neq 0$ for the mean weight change with the cognitive behavioral therapy. With the 0.05 significance level, we would reject H_o. We also know that the 95%confidence interval is (0.2, 5.8) pounds. The confidence interval shows just how different from 0 the population mean weight change is likely to be. It is estimated to fall between 0.2 and 5.8 pounds. We infer that the population mean weight change μ is positive because all the numbers in this interval are greater than 0, but the effect of the therapy could be as small as 0.2.

Both the hypothesis test and the confidence interval suggested that μ differs from 0. Conclusions about means using a two-sided test are the same as conclusions using confidence intervals. If a two-sided test says you can reject the hypothesis that $\mu = 0$, then 0 is not in the corresponding confidence interval.

Comparing two groups: hypothesis testing for two means

Example:
A 30-month study evaluated the degree of addiction that teenagers form to nicotine once they begin experimenting with smoking (J. DiFranza, *Archives of Pediatric and Adolescent Medicine*, 156, 2002). Random numbers were used to sample 679 seventh-grade students in two Massachusetts cities. of them, the 332 students who had ever used tobacco by the start of the study were the subjects evaluated. The response variable was constructed using a questionnaire developed for the study, called the Hooked on Nicotine Checklist (HONC). This is a list of ten questions such as "Have you ever tried to quit but couldn't", "Do you ever have strong cravings to smoke?" and "Is it hard to keep from smoking in places where you are not supposed to, like school?". The HONC score is the total number of questions to which a student answered yes during the study. Each student's HONC score falls between 0 and 10. The higher the score, the more hooked on nicotine a student is judged to be.
The study considered explanatory variables such as gender, that might be associated with the HONC score. Below shows the descriptive statistics for the data.

Group	**Sample Size**	**Mean of HONC score**	**Standard Deviation σ_i**
Males	150	2.8	3.6
Females	182	1.6	2.9

We want to be able to compare the sample HONC scores for females and males.

Standard deviation for comparing two means

If we denote males with a subscript 1 and females with a subscript 2, how well does the difference $(\bar{X}_1 - \bar{X}_2)$ between two sample means estimate the difference between the two population means $(\mu_1 - \mu_2)$? This is described by the standard deviation of the sampling distribution of $(\bar{X}_1 - \bar{X}_2)$. We can find the standard deviation of $(\bar{X}_1 - \bar{X}_2)$ by combining the two standard deviations with the following formula

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{(standard\ deviation\ of \bar{X})^2 + (standard\ deviation\ of \bar{X})^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (5)$$

Example:
Another explanatory variable in the teenage smoking study was whether a subject was still a smoker when the study ended. The study had 75 smokers and 257 ex-smokers at the end of the study. The HONC means describing nicotine addiction were $\bar{X}_1 = 5.9 (\sigma_1 = 3.3)$ for the smokers and $\bar{X}_2 = 1.0 (\sigma_2 = 2.3)$ for the ex-smokers.
First compare smokers and ex-smokers on their mean HONC scores.
Since $(\bar{X}_1 - \bar{X}_2) = 5.9 - 1.0 = 4.9$, on the average smokers answered yes to nearly five more questions than ex-smokers did on the ten-question HONC scale. That's a large sample difference.
What is the standard deviation of the difference in the sample mean of HONC scores? Interpret.
Applying the formula for the standard deviation of $(\bar{X}_1 - \bar{X}_2)$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{3.3^2}{75} + \frac{2.3^2}{257}} = 0.41$$

This describes the spread of the sampling distribution of $(\bar{X}_1 - \bar{X}_2)$.

Steps for hypothesis testing about two means

Step 1: Assumptions

- We have quantitative response variables for two groups - The data are independent and obtained using randomization
- For each group: either the population distribution is normal OR the sample size is large enough that the central limit theorem can be used to say that $\bar{X}_i$ is approximately normal
In the HONC study our two groups are smokers and ex-smokers with the response variable being HONC scores.
Because the students were chosen randomly we can assume they are independent.
Our sample sizes are $n_1 = 75$ smokers and $n_2 = 257$ ex-smokers so these are large enough samples to assume both $\bar{X}_1$ and $\bar{X}_2$ have an approximately normal distribution.

Step 2: Hypotheses

One-sided alternative hypothesis:

$$H_o : \mu_1 \leq \mu_2 \ (\rightarrow \mu_1 - \mu_2 \leq 0) \ \ VS. \ \ H_a : \mu_1 > \mu_2 \ (\rightarrow \mu_1 - \mu_2 > 0) \tag{6}$$

$$H_o : \mu_1 \geq \mu_2 \ (\rightarrow \mu_1 - \mu_2 \geq 0) \ \ VS. \ \ H_a : \mu_1 < \mu_2 \ (\rightarrow \mu_1 - \mu_2 < 0) \tag{7}$$

Two-sided alternative hypothesis:

$$H_o : \mu_1 = \mu_2 \ (\rightarrow \mu_1 - \mu_2 = 0) \ \ VS. \ \ H_a : \mu_1 \neq \mu_2 \ (\rightarrow \mu_1 - \mu_2 \neq 0) \tag{8}$$

For example, let $(\mu_1 - \mu_2)$ denote the mean HONC score difference between smokers and ex-smokers for the population represented by the sample of teenagers. If there is no difference between the HONC scores of the two groups then $H_o : \mu_1 - \mu_2 = 0$. We will test this idea against the possibility that there is a difference in HONC scores between the two groups using a two-sided alternative $H_o : \mu_1 - \mu_2 \neq 0$.

Step 3: Test statistic

The test statistic is the distance between the sample mean difference $(\bar{X}_1 - \bar{X}_2)$ and the null hypothesized value $(\mu_1 - \mu_2)$, as measured by the number of standard deviations between them. This test statistic is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{sample\ mean\ difference - null\ hypothesis\ mean\ difference}{standard\ deviation\ of\ the\ sample\ mean\ difference} \tag{9}$$

In the HONC study, the sample mean difference is $(\bar{X}_1 - \bar{X}_2) = 5.9 - 1.0 = 4.9$ and the we found the standard deviation of the sample mean difference to be $\sqrt{\frac{3.3^2}{75} + \frac{2.3^2}{257}} = 0.41$. The test statistic for this example is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = Z = \frac{(4.9) - 0}{\sqrt{\frac{3.3^2}{75} + \frac{2.3^2}{257}}} = \frac{4.9}{0.41} = 11.95$$

The observed test statistic value is 11.95 meaning that our $(\bar{X}_1 - \bar{X}_2) = 4.9$ is 11.95 standard deviations away from $\mu_1 - \mu_2 = 0$.

Step 4: P-value

The **P-value** is the probability that the test statistic takes a value like the observed test statistic or even more extreme, if the null hypothesis is true. It is a single tail or a two-tail proability depending on whether the alternative hypothesis is one-sided or two-sided.

Hypothesis	**P-value**
$H_o : \mu_1 - \mu_2 \leq 0 \;\; VS. \;\; H_a : \mu_1 - \mu_2 > 0$	$P(Z > \textit{observed test statistic})$
$H_o : \mu_1 - \mu_2 \geq 0 \;\; VS. \;\; H_a : \mu_1 - \mu_2 < 0$	$P(Z < \textit{observed test statistic})$
$H_o : \mu_1 - \mu_2 = 0 \;\; VS. \;\; H_a : \mu_1 - \mu_2 \neq 0$	$P(\lvert Z \rvert > \textit{observed test statistic})$

For the HONC study with $H_a : \mu_1 - \mu_2 \neq 0$, the P-value is the two tail probability of a test statistic value farther out in each tail than the observed value of 11.95. This probability is double the single-tail probability.

Step 5: Conclusion

In our example, our P-value is so small that it definitely provides evidence against the null hypothesis that there is not a difference between groups. As seen from the large difference in sample means (4.9), smokers and ex-smokers HONC scores do differ on average.