Inferences about two means

Math 130: Introductory Statistics

Goals of this section

- Learn about how to make hypothesis tests about the difference between two samples.
- Learn how to treat independent samples and dependent samples.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic = $\bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Hypothesis tests

- Hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example from last section: hypothesis testing about μ:
  - sample statistic $= \bar{x}$.
  - standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Therefore:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- If σ is known, use z-statistic.
- If σ is not known, use s instead, and use t-statistic.

Confidence intervals

- Confidence intervals are all the same:

$$(\text{sample statistic} - E,\ \text{sample statistic} + E)$$

$$E = (\text{critical t or z})\,(\text{standard deviation of the sampling distribution})$$

- Examples with confidence intervals about μ:

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Confidence intervals

- Confidence intervals are all the same:

$$(\text{sample statistic} - E,\ \text{sample statistic} + E)$$

$$E = (\text{critical t or z})\,(\text{standard deviation of the sampling distribution})$$

- Examples with confidence intervals about μ:

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Inferences about two means

- Suppose we take two samples and expect the two means to be different.
- We'll want to make hypothesis tests and confidence intervals about $\bar{x}_1 - \bar{x}_2$.
- We'll need to know the standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

Inferences about two means

- Suppose we take two samples and expect the two means to be different.
- We'll want to make hypothesis tests and confidence intervals about $\bar{x}_1 - \bar{x}_2$.
- We'll need to know the standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

Inferences about two means

- Suppose we take two samples and expect the two means to be different.
- We'll want to make hypothesis tests and confidence intervals about $\bar{x}_1 - \bar{x}_2$.
- We'll need to know the standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

Independent and dependent samples

- Independent samples: when observations of one sample are not at all related to observations of another sample.
- May assume independent samples may have different variances, or they may be the same.
- Dependent samples (or paired samples): when the observations in the two samples are related.
  - For example, the same participants are in each sample, before and after some kind of treatment.

Independent and dependent samples

- Independent samples: when observations of one sample are not at all related to observations of another sample.
- May assume independent samples may have different variances, or they may be the same.
- Dependent samples (or paired samples): when the observations in the two samples are related.
    - For example, the same participants are in each sample, before and after some kind of treatment.

Independent and dependent samples

- Independent samples: when observations of one sample are not at all related to observations of another sample.
- May assume independent samples may have different variances, or they may be the same.
- Dependent samples (or paired samples): when the observations in the two samples are related.
    - For example, the same participants are in each sample, before and after some kind of treatment.

Independent and dependent samples

- Independent samples: when observations of one sample are not at all related to observations of another sample.
- May assume independent samples may have different variances, or they may be the same.
- Dependent samples (or paired samples): when the observations in the two samples are related.
  - For example, the same participants are in each sample, before and after some kind of treatment.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Independent samples

- Assume two samples are independent.
- Both of the sample sizes must be large ($n_1 > 30$ and $n_2 > 30$) or..
- Assume both populations are normally distributed.
- Allow the standard deviations of each sample may be different.
- Standard deviation of sampling distribution of $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- n_1 and σ_1 are the sample size drawn from population 1, and the population 1 standard deviation.
- n_2 and σ_2 are the sample size drawn from population 1, and the population 1 standard deviation.

Hypothesis test and confidence intervals

- Suppose you want to test if the mean of sample one is greater than the mean of sample two.
- $H_0 : \mu_1 - \mu_2 = 0$
- $H_a : \mu_1 - \mu_2 > 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$.
- Confidence interval for $\bar{x}_1 - \bar{x}_2$:

$$[(\bar{x}_1 - \bar{x}_2) - E, (\bar{x}_1 - \bar{x}_2) + E]$$

$$E = t_{\alpha/2,df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Hypothesis test and confidence intervals

- Suppose you want to test if the mean of sample one is greater than the mean of sample two.
- $H_0 : \mu_1 - \mu_2 = 0$
- $H_a : \mu_1 - \mu_2 > 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$.
- Confidence interval for $\bar{x}_1 - \bar{x}_2$:

$$[(\bar{x}_1 - \bar{x}_2) - E, (\bar{x}_1 - \bar{x}_2) + E]$$

$$E = t_{\alpha/2,df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Hypothesis test and confidence intervals

- Suppose you want to test if the mean of sample one is greater than the mean of sample two.
- $H_0 : \mu_1 - \mu_2 = 0$
- $H_a : \mu_1 - \mu_2 > 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$.
- Confidence interval for $\bar{x}_1 - \bar{x}_2$:

$$[(\bar{x}_1 - \bar{x}_2) - E, (\bar{x}_1 - \bar{x}_2) + E]$$

$$E = t_{\alpha/2,df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Hypothesis test and confidence intervals

- Suppose you want to test if the mean of sample one is greater than the mean of sample two.
- $H_0 : \mu_1 - \mu_2 = 0$
- $H_a : \mu_1 - \mu_2 > 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$.
- Confidence interval for $\bar{x}_1 - \bar{x}_2$:

$$[(\bar{x}_1 - \bar{x}_2) - E, (\bar{x}_1 - \bar{x}_2) + E]$$

$$E = t_{\alpha/2, df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Hypothesis test and confidence intervals

- Suppose you want to test if the mean of sample one is greater than the mean of sample two.
- $H_0 : \mu_1 - \mu_2 = 0$
- $H_a : \mu_1 - \mu_2 > 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Degrees of freedom is the smaller of $n_1 - 1$ and $n_2 - 1$.
- Confidence interval for $\bar{x}_1 - \bar{x}_2$:

$$[(\bar{x}_1 - \bar{x}_2) - E, (\bar{x}_1 - \bar{x}_2) + E]$$

$$E = t_{\alpha/2,df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Equal variances

- What if you could assume the true population variance is the same for both samples.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Use a pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Standard deviation of the sampling distribution:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- Degrees of freedom:

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Inferences about paired samples

- Each sample has same observations taken at different points in time.
  - Example: Weight recorded for a group of individuals before and after an exercise program.
- Each sample therefore has the same sample size, call it n.
- Sample statistic: *for each observation*, compute the difference $x_{1,i} - x_{2,i}$
- Take the average of the differences, call it $\bar{d}$.
- Take the standard deviation of the differences, call it s_d.
- Standard deviation of the sampling distribution of $\bar{d}$:

$$\frac{s_d}{\sqrt{n}}$$

Homework

- Section 9-3, problems 9-18 odds.