Two-sample Quantitative Data

Ryan Miller

Introduction

- Last week we introduced the idea of *two-sample* data, or scenarios involving a comparison of two distinct groups of cases
  - We focused on comparing categorical outcomes (proportions) using Normal probability models as a basis for statistical inference

Introduction

- Last week we introduced the idea of *two-sample* data, or scenarios involving a comparison of two distinct groups of cases
  - We focused on comparing categorical outcomes (proportions) using Normal probability models as a basis for statistical inference
- This week we'll look at *two-sample quantitative* data
  - More specifically, we'll see how the t-distribution can be used to compare the *means of two groups*

Review of the t-distribution

- William Gosset, a statistician working for Guinness Brewing company in the 1890s, discovered the t-distribution by noticing the Z-test appeared to produce erroneous results when used on small samples of quantitative data

Review of the t-distribution

- William Gosset, a statistician working for Guinness Brewing company in the 1890s, discovered the t-distribution by noticing the Z-test appeared to produce erroneous results when used on small samples of quantitative data
- The problem was in the standard error used in the Z-value, where the sample standard deviation, s, was being used as estimate of the population standard deviation, σ
  - Estimating this extra parameter introduced additional uncertainty beyond what was accounted for by the Normal model

Review of the t-distribution

- William Gosset, a statistician working for Guinness Brewing company in the 1890s, discovered the t-distribution by noticing the Z-test appeared to produce erroneous results when used on small samples of quantitative data
- The problem was in the standard error used in the Z-value, where the sample standard deviation, s, was being used as estimate of the population standard deviation, σ
  - Estimating this extra parameter introduced additional uncertainty beyond what was accounted for by the Normal model
- The t-distribution has thicker tails to properly account for the added uncertainty introduced by estimating the population standard deviation

The t-distribution for Two-sample Data

When comparing the means of two groups, probability theory suggests the following Normal model:

$$\bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

The t-distribution for Two-sample Data

When comparing the means of two groups, probability theory suggests the following Normal model:

$$\bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

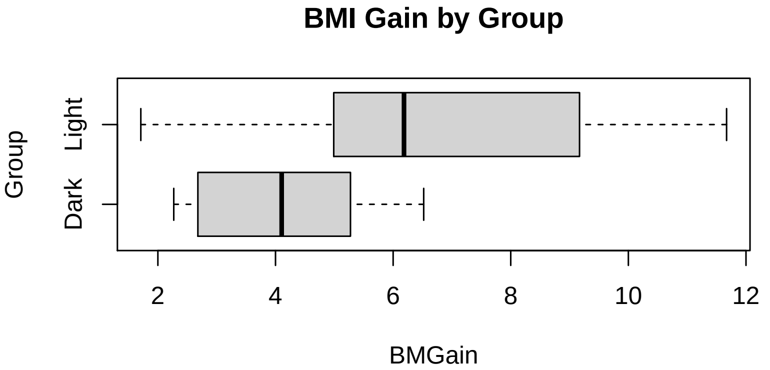
This model serves the basis of the **two-sample t-test**:

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

- Usually we're interested in testing $H_0: \mu_1 = \mu_2$, which implies $\mu_1 - \mu_2 = 0$ under the null hypothesis
- Degrees of freedom are complicated, they're typically calculated using statistical software

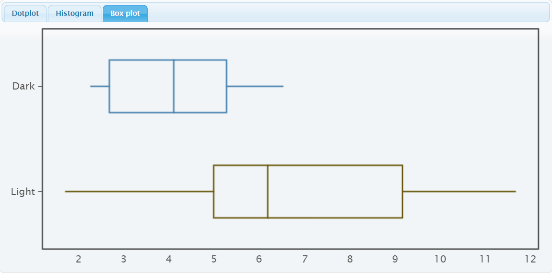
Example - Two-sample t-test

To explore whether artificial light at night contributes to weight gain, researchers randomized 18 young mice to live in lab environments with either complete darkness, or an artificial nightlight, during evening hours:



Example - Two-sample t-test

The data from this experiment are pre-loaded into StatKey at this link



Summary Statistics

Statistics	Light	Dark	Overall
Sample Size	10	8	18
Mean	6.732	4.114	5.568
Standard Deviation	2.966	1.557	2.729
Minimum	1.71	2.27	1.71
Q_1	4.99	2.68	4.00
Median	6.19	4.11	5.16
Q_3	9.17	5.28	6.94
Maximum	11.67	6.52	11.67

Despite the small sample sizes, the data appear Normally distributed, so we can confidently use the t-distribution for hypothesis testing

Example - Two-sample t-test

1) $H_0 : \mu_1 = \mu_2$, or the nightlight has no impact on the average weight gain in each group

Example - Two-sample t-test

1) $H_0 : \mu_1 = \mu_2$, or the nightlight has no impact on the average weight gain in each group
2) We can evaluate how far the observed data are from what we'd expect under this hypothesis using a T-value:

$$T = \frac{(6.732 - 4.114) - 0}{\sqrt{2.966^2/10 + 1.557^2/8}} = 2.41$$

Example - Two-sample t-test

1) $H_0 : \mu_1 = \mu_2$, or the nightlight has no impact on the average weight gain in each group
2) We can evaluate how far the observed data are from what we'd expect under this hypothesis using a T-value:

$$T = \frac{(6.732-4.114)-0}{\sqrt{2.966^2/10+1.557^2/8}} = 2.41$$

3) From software, $df = 14.1$, so the two-sided p-value is 0.03
4) We conclude this experiment provides statistically significant evidence that artificial nightlights lead to weight gain in mice

Confidence Interval Estimation

Similar to other scenarios, the t-distribution can be used as the basis for P% confidence interval estimates for a difference in means:

$$\text{Estimate} \pm \text{Margin of Error} = \bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- t^* is a percentile from a t-distribution with the proper *df* (found using software) that defines the middle P% of the distribution

Example - Confidence Interval Estimation

- To see whether satiety is influenced by visual cues, researchers randomized 54 individuals to sit at tables containing 18oz soup bowls that were identical in appearance
  - Half of these bowls were specially designed to slowly refill as they emptied

Example - Confidence Interval Estimation

- To see whether satiety is influenced by visual cues, researchers randomized 54 individuals to sit at tables containing 18oz soup bowls that were identical in appearance
  - Half of these bowls were specially designed to slowly refill as they emptied
- The researchers tracked how many ounces of soup were consumed by each group (those with the refilling bowls vs. those with the ordinary bowls)
  - How much a difference does refilling have on the amount people consume? (assume $df = 47.5$)

	n	mean	sd
Ordinary	27	8.5	6.1
Refilling	27	14.7	8.4

Example - Confidence Interval Estimation

A 95% CI for the mean difference in ounces of soup consumed in these two groups is calculated below:

$$\begin{aligned} \text{Estimate} \pm \text{Margin of Error} &= (14.7 - 8.5) \pm 2.011\sqrt{\frac{8.4^2}{27} + \frac{6.1^2}{27}} \\ &= (2.2, 10.2) \end{aligned}$$

Thus, we can be 95% confident that people will consume between 2.2 oz and 10.2 oz more soup if the visual cue of their bowl emptying is removed

Relating Confidence Intervals and Hypothesis Tests

- It is important to recognize the complimentary nature of confidence interval estimation and hypothesis testing
  - Our 95% CI of (2.2, 10.2) suggests it is not plausible that members of each group consume the same amount of soup
- Based upon this interval, we can be certain the two-sided p-value when testing $H_0 : \mu_1 = \mu_2$ will be less than 0.05 (the mathematical complement of the interval's confidence level)

Assumptions for using the t-distribution

When using the t-distribution as the basis for statistical inference, it is important that one of the following is met:

1) The data in both groups is approximately Normal (regardless of sample size)
2) The sample sizes in both groups are larger than 30 (regardless of how the data are distributed)

If neither of these is met, the results of our inference (confidence intervals and p-values) may be unreliable

Conclusion

We've now finished discussing methods of statistical inference for four common scenarios

1) A single proportion (one-sample categorical data)
2) Comparing two proportions (two-sample categorical data)
3) A single mean (one-sample/paired quantitative data)
4) Comparing two means (two-sample quantitative data)

While some of the details for these scenarios differ (different standard errors, whether to use the Normal or t-distribution, etc.), the main concepts are the same