Unit 4 - Inference for numerical variables

Suggested reading: OpenIntro Statistics, Chapter 5

Suggested exercises:

* *Part 1 - Comparing two means: 5.1, 5.5, 5.7, 5.9, 5.11*

* *Part 2 - Bootstrapping*

* *Part 3 - Inference with the t-distribution: 5.15, 5.19, 5.21, 5.29, 5.31, 5.33, 5.35*

* *Part 4 - Comparing three or more means (ANOVA): 5.39, 5.41, 5.43, 5.45*

LO 1. Describe how bootstrap distributions are constructed, and recognize how they are different from sampling distributions.

LO 2. Construct bootstrap confidence intervals using one of the following methods:

- Percentile method: XX% confidence level is the middle XX% of the bootstrap distribution.

- Standard error method: If the standard error of the bootstrap distribution is known, and the distribution is nearly normal, the bootstrap interval can also be calculated as $\bar{x}_{boot} \pm z^\star SE_{boot}$.

LO 3. Recognize that when the bootstrap distribution is extremely skewed and sparse, the bootstrap confidence interval may not be reliable.

* *Test yourself:*

    1. *How is a bootstrap distribution different from a sampling distribution?*

    2. *If a bootstrap distribution is constructed using 200 simulations, how would we find the 95% bootstrap confidence interval? Hint: Draw a sketch.*

    3. *When is a bootstrap confidence interval not appropriate?*

* *Suggested Reading: Section 5.1 of OpenIntro Statistics*

LO 4. Define observations as paired if each observation in one dataset has a special correspondence or connection with exactly one observation in the other data set.

LO 5. Carry out inference for paired data by first subtracting the paired observations from each other, and then treating the set of differences as a new numerical variable on which to do inference (such as a confidence interval or hypothesis test for the average difference).

∗ *Test yourself:*

1. *20 cardiac patients' blood pressure is measured before taking a medication, and after. For a given patient, are the before and after blood pressure measurements dependent (paired) or independent?*

2. *A random sample of 100 students were obtained and then randomly assigned into two equal sized groups. One group went on a roller coaster while the other in a simulator at an amusement park. Afterwards their blood pressure measurements were taken. Are the measurements dependent (paired) or independent?*

∗ *Suggested Reading: Section 5.2 of OpenIntro Statistics*

LO 6. Calculate the standard error of the difference between means of two independent samples as

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

and use this standard error in hypothesis testing and confidence intervals comparing means of independent groups.

LO 7. Recognize that a good interpretation of a confidence interval for the difference between two parameters includes a comparative statement (mentioning which group has the larger parameter).

LO 8. Recognize that a confidence interval for the difference between two parameters that doesn't include 0 is in agreement with a hypothesis test where the null hypothesis that sets the two parameters equal to each other is rejected.

∗ *Test yourself:*

1. *Describe how the two sample means test is different from the paired means test, both conceptually and in terms of the calculation of the standard error.*

2. *A 95% confidence interval for the difference between the number of calories consumed by mature and juvenile cats ($\mu_{mat} - \mu_{juv}$) is (80 calories, 100 calories). Interpret this interval, and determine if it suggests a significant difference between the two means.*

∗ *Suggested Reading: Section 5.3 and 5.4 of OpenIntro Statistics*

LO 9. Explain why means of small samples ($n < 30$) follow the t-distribution (instead of the normal, z, distribution).

LO 10. Describe how the t-distribution is different from the normal distribution, and what "heavy tail" means in this context.

LO 11. Note that the t-distribution has a single parameter, degrees of freedom, and as the degrees of freedom increases this distribution approaches the normal distribution.

LO 12. Use a t-statistic, with degrees of freedom $df = n-1$ for inference for a population mean using data from a small sample:

$$\text{CI: } \bar{x} \pm t^\star_{df} SE \qquad\qquad \text{HT: } T_{df} = \frac{\bar{x} - \mu}{SE}$$

where $SE = \frac{s}{\sqrt{n}}$.

LO 13. Use a t-statistic, with degrees of freedom $df = min(n_1 - 1, n_2 - 1)$ for inference for difference between means of two population means using data from two small samples, where $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

LO 14. Make note of the pooled standard deviation but use it in very rare circumstances where the standard deviations of the populations being compared are known to be very similar.

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

LO 15. Describe how to obtain a p-value for a t-test and a critical t-score ($t^\star_{df}$) for a confidence interval.

* *Test yourself:*

    1. *What is the $t^\star$ for a 95% confidence interval for a mean, where the sample size is 13.*

    2. *What is the p-value for a hypothesis test where the alternative hypothesis is two-sided, the sample size is 20, and the test statistic, T, is calculated to be 1.75?*

* *Suggested Reading: Section 5.5 of OpenIntro Statistics*

LO 16. Define analysis of variance (ANOVA) as a statistical inference method that is used to determine - by simultaneously considering many groups at once - if the variability in the sample means is so large that it seems unlikely to be from chance alone.

LO 17. Recognize that the null hypothesis in ANOVA sets all means equal to each other, and the alternative hypothesis suggest that at least one mean is different.

    $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

    $H_A :$ At least one mean is different

LO 18. List the conditions necessary for performing ANOVA

    (1) the observations should be independent within and across groups

    (2) the data within each group are nearly normal

    (3) the variability across the groups is about equal

and use graphical diagnostics to check if these conditions are met.

LO 19. Recognize that the test statistic for ANOVA, the F statistic, is calculated as the ratio of the mean square between groups (MSG, variability between groups) and mean square error (MSE, variability within errors). Also recognize that the F statistic has a right skewed distribution

with two different measures of degrees of freedom: one for the numerator ($df_G = k - 1$, where k is the number of groups) and one for the denominator ($df_E = n - k$, where n is the total sample size).

  - Note that you won't be expected to calculate MSG or MSE from the raw data, but you should have a conceptual understanding of how they're calculated and what they measure.

LO 20. Describe why calculation of the p-value for ANOVA is always "one sided".

LO 21. Describe why conducting many t-tests for differences between each pair of means leads to an increased Type 1 Error rate, and we use a corrected significance level (Bonferroni correction, $\alpha^\star = \alpha/K$, where K is the number of comparisons being considered) to combat inflating this error rate.

  - Note that $K = \frac{k(k-1)}{2}$, where k is the number of groups.

LO 22. Describe why it is possible to reject the null hypothesis in ANOVA but not find significant differences between groups when doing pairwise comparisons.

  * *Test yourself:*

  1. *We would like to compare the average income of Americans who live in the Northeast, Midwest, South, and West. What are the appropriate hypotheses?*

  2. *Suppose the sample in the question above has 1000 observations, what are the degrees of freedom associated with the F-statistic?*

  3. *Suppose the null hypothesis is rejected. Describe how we would discover which regions' averages are different from each other. Make sure to discuss how many pairwise comparisons we would need to make, and what the corrected significance level would be.*

  4. *What visualizations are useful for checking each of the conditions required for performing ANOVA?*