Statistical Inference

Inference for Numerical Variables

Behnam Bahrak
Spring 2020

Conditions for ANOVA

1. **Independence:**
   - ✓ **within groups:** sampled observations must be independent
   - ✓ **between groups:** the groups must be independent of each other (non-paired)

2. **Approximate normality:** distributions should be nearly normal within each group

3. **Equal variance:** groups should have roughly equal variability

(1) Independence

within: sampled observations must be independent of each other

- random sample / assignment
- each n_j less than 10% of respective population
- always important, but sometimes difficult to check

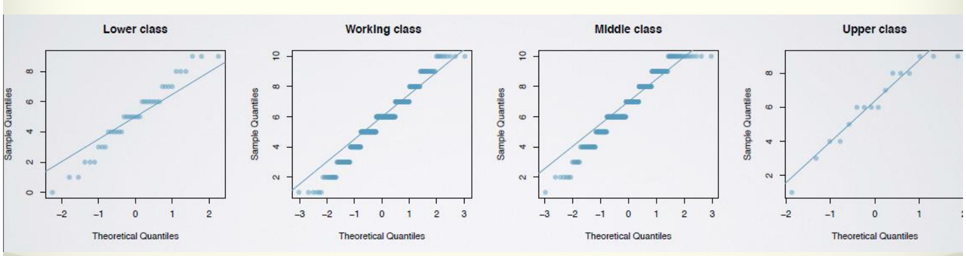
between: groups must be independent of each other

- carefully consider whether the groups may be dependent

→ *repeated measures anova*

(2) Approximately Normal

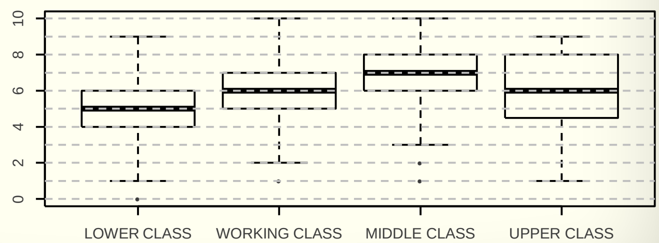
- Distribution of response variable within each group should be approximately normal
- Especially important when sample sizes are small



(3) Constant Variance

- Variability should be consistent across groups: homoscedastic groups
- Especially important when sample sizes differ between groups

	n	sd
lower class	41	2.24
working class	407	1.87
middle class	331	1.89
upper class	16	2.34
overall	**795**	**1.98**



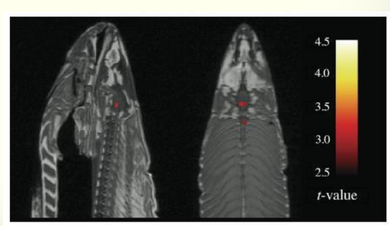
Misconducting Statistical Tests

Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction

Craig M. Bennett[1]*, Abigail A. Baird[2], Michael B. Miller[1] and George L. Wolford[3]

[1]Department of Psychology, University of California at Santa Barbara, Santa Barbara, CA 93106
[2]Department of Psychology, Blodgett Hall, Vassar College, Poughkeepsie, NY 12604
[3]Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hanover, NH 03755

- The salmon was shown the same social perspective taking task that was later administered to a group of human subjects.
- Statistics that were uncorrected for multiple comparisons showed active voxel clusters in the salmon's brain cavity and spinal column.



Which means differ?

- ANOVA method only tells us that at least one pair of means are different, but not which pair of means are different.
- Two sample t tests for differences in each possible pair of groups
- multiple tests → inflated Type 1 error rate
- solution: use modified significance level

Multiple Comparisons

- Testing many pairs of groups is called multiple comparisons
- The Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests
  - Adjust α by the number of comparisons being considered:
  $$\alpha^* = \alpha / K$$
  - K : number of comparisons
  $$K = k(k-1)/2$$

Example

- The social class variables has 4 levels. If $\alpha = 0.05$ for the original ANOVA, what should the modified significance level be for two sample t tests for determining which pairs of groups have significantly different means?

$$k = 4$$

$$K = \frac{4 \times 3}{2} = 6$$

$$a^* = \frac{0.05}{6} = 0.0083$$

Pairwise Comparisons

- Constant variance → re-think standard error and degrees of freedom:
  - use consistent standard error and degrees of freedom for all pairwise comparison tests

- Compare the p-values from each test to the modified significance level α^*

Pairwise Comparisons

- Standard error for multiple pairwise comparisons:

$$SE = \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

- Independent groups test:

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

- Degrees of freedom for multiple pairwise comparisons:

$$df = df_E = n - k$$

$$df = \min(n_1 - 1\,, n_2 - 1)$$

Example

➢ Is there a difference between the average vocabulary scores between middle and lower class Americans?

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Class	3	236.56	78.855	21.735	<0.0001
Residuals	**791**	2869.80	**3.628**		
Total	794	3106.36			

	n	mean
lower class	41	5.07
middle class	331	6.76

$$H_0: \mu_{\text{middle}} - \mu_{\text{lower}} = 0 \quad , \qquad H_A: \mu_{\text{middle}} - \mu_{\text{lower}} \neq 0$$

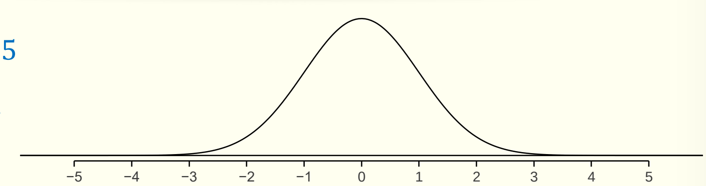
$$T = \frac{(\bar{x}_{\text{middle}} - \bar{x}_{\text{lower}}) - 0}{\sqrt{\frac{MSE}{n_{\text{middle}}} + \frac{MSE}{n_{\text{lower}}}}} = \frac{6.76 - 5.07}{\sqrt{\frac{3.628}{331} + \frac{3.628}{41}}} = \frac{1.69}{0.315} = 5.365$$

$$df = 791$$

Example

$T = 5.365$

$df = 791$




```
R
> 2 * pt(5.365, df = 791, lower.tail = FALSE)
[1] 1.063895e-07
```

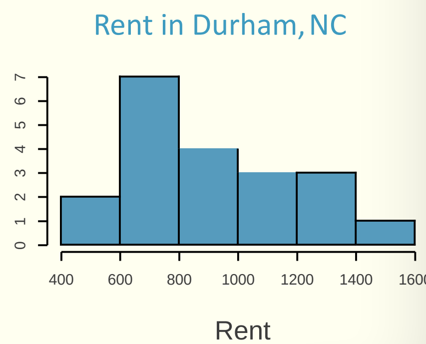
$\alpha^* = 0.0083$

$p - value < \alpha^* \rightarrow$ Reject H_0

Bootstrapping

- Twenty 1+ bedroom apartments were randomly selected in Durham, NC. Is the mean or the median a better measure of typical rent in Durham?

- Can we apply CLT based methods we have learned so far to construct confidence intervals for both?

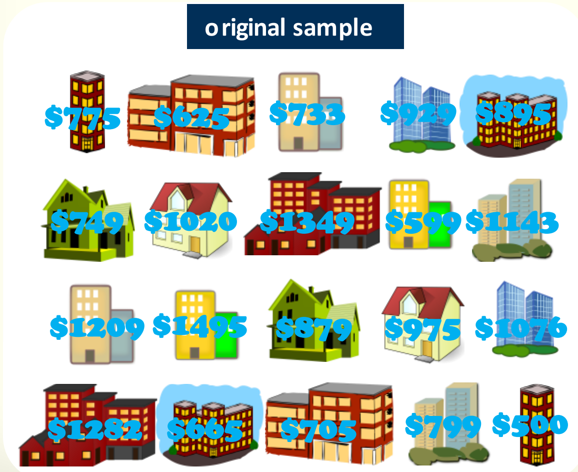


Bootstrapping

- *"Pulling oneself up by one's bootstraps"*: a metaphor for accomplishing an impossible task without any outside help

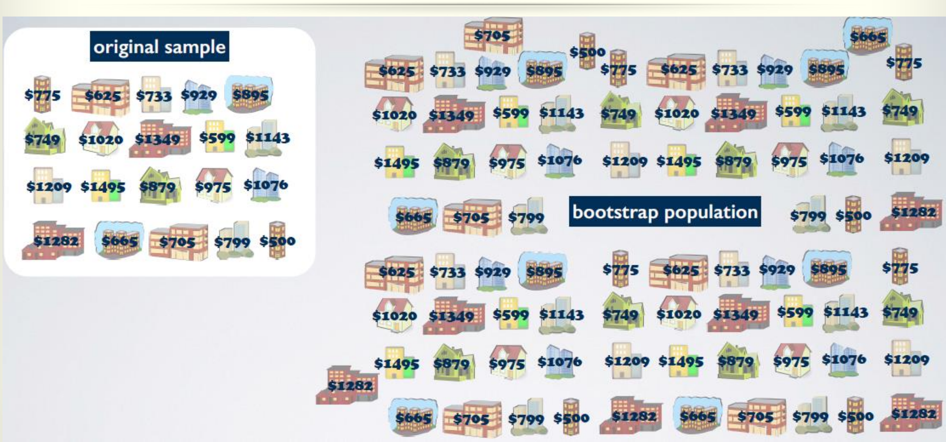
- In this case, the ~~im~~possible task is estimating a population parameter, using data from only the given sample.

Original Sample



sample median
$887

Bootstrap Population



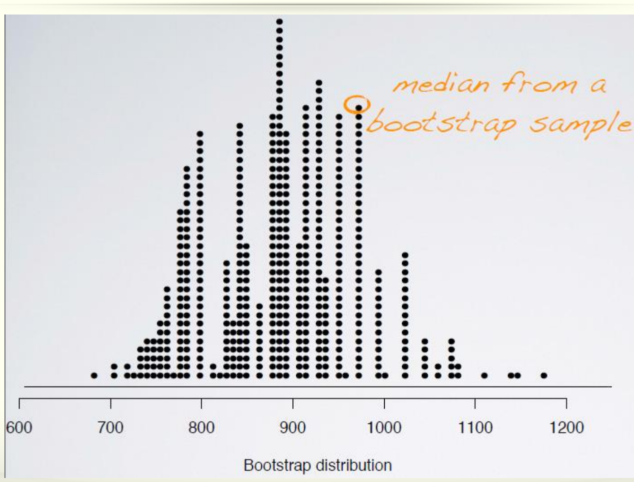
Bootstrapping Scheme

(1) Take a bootstrap sample - a random sample taken **with replacement** from the original sample, of the same size as the original sample

(2) Calculate the bootstrap statistic - a statistic such as mean, median, proportion, etc. computed on the bootstrap samples

(3) Repeat steps (1) and (2) many times to create a bootstrap distribution - a distribution of bootstrap statistics

Bootstrap Distribution



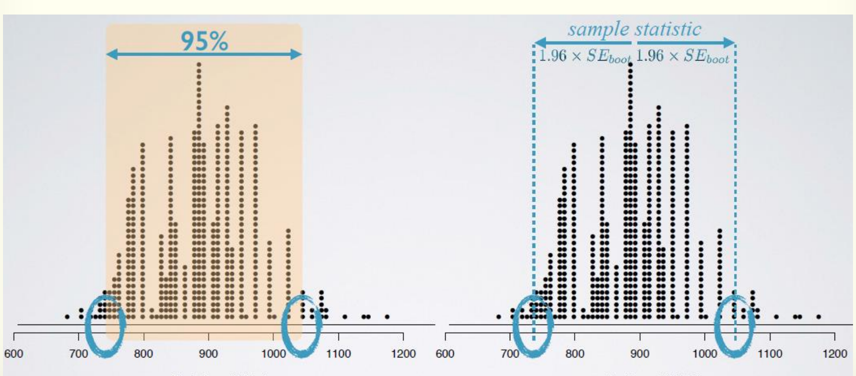
Confidence Interval

1) **Percentile Method:**
   - Simply as the middle 95% of the bootstrap distribution.
   - So the bounds of the interval are the 2.5th and the 97.5th percentiles of the bootstrap distribution.

2) **Standard error method:**
   - We calculate the interval as the sample statistic plus or minus t^* times the standard error of the bootstrap distribution
   - The critical T-score (t^*) will have $(n-1)$ as its degrees of freedom, where n is the number of bootstrap samples.

Confidence Interval

(1) percentile method (2) standard error method



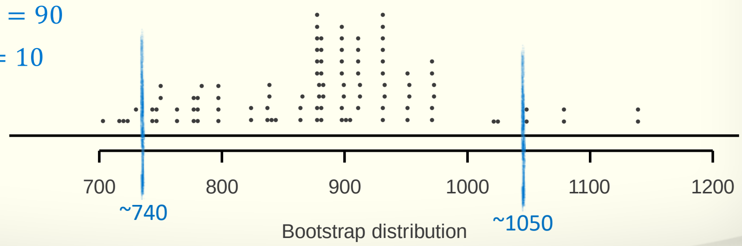
Example

➢ The dot plot below shows the distribution of medians of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval for the median rent based on this bootstrap distribution using the percentile method.

$100 \times 0.90 = 90$

$100 - 90 = 10$

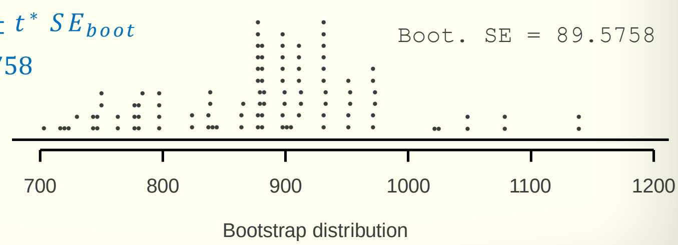
$10/2 = 5$



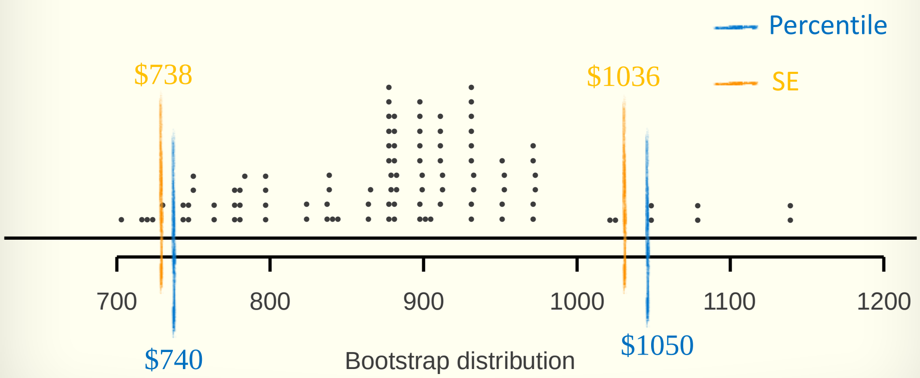
Example

➢ The dot plot below shows the distribution of medians of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval for the median rent based on this bootstrap distribution using the standard error method.

$$sample\ median\ \pm t^* \ SE_{boot}$$
$$= 887 \pm 1.66 \times 89.5758$$
$$\approx (738, 1036)$$



Percentile vs. SE Methods



Bootstrapping Limitations

- Bootstrap intervals do not have as rigid conditions as on sample size and skew essentials limit theorem based methods.
- If the bootstrap distribution is extremely skewed or sparse, the bootstrap interval might be unreliable
- A representative sample is still required — if the sample is biased, the estimates resulting from this sample will also be biased.

Bootstrap vs. Sampling Distribution

- Sampling distribution created using sampling (with replacement) from the **population**
- Bootstrap distribution created using sampling (with replacement) from the **sample**
- Both are distributions of sample statistics

When to use bootstrapping?

- When **to** use bootstrapping?
  - When the theoretical distribution of a statistic of interest is complicated or unknown.
  - When the sample size is insufficient for straightforward statistical inference.
  - When sample size calculations have to be performed.
- When **not to** use bootstrapping?
  - when the underlying population lacks a finite variance (e.g. a power law or Cauchy distribution), then the bootstrap distribution will not converge to the same limit as the sample statistic.
  - Unless one is reasonably sure that the underlying distribution is not heavy tailed, one should hesitate to use the naive bootstrap.