Announcements

Make sure to sign in on the google form (I send a list of which section questions are useful for which pset questions afterwards)

Pset 2 due Saturday 9/30

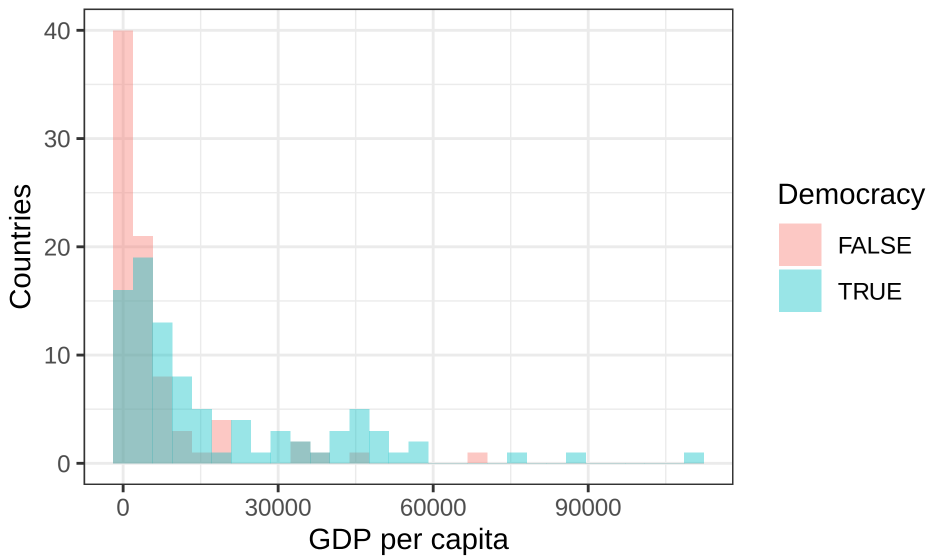
Introductions (again)

- Name
- One question or thought related to lecture last week (ranks, bootstrap, randomization)

Hypothesis testing on real data

These problems will deal with a dataset of country-level statistics from UNdata and Varieties of Democracy.

1. Suppose we want to test for a difference in mean 2010 GDP per capita between democracies and non-democracies. The following plots show the distributions. Which tests would be valid?



2. Perform a formal rank-sum test for the difference in GDP per capita between democracies and non-democracies.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dem_gdps and nondem_gdps
## W = 5443, p-value = 7.754e-08
## alternative hypothesis: true location shift is not equal to 0
```

3. Perform a formal log-transformed t-test for the difference in GDP per capita between democracies and non-democracies. Give a 95% confidence interval for the ratio of medians.

```
##
##  Welch Two Sample t-test
##
## data:  log(dem_gdps) and log(nondem_gdps)
## t = 5.8451, df = 169.64, p-value = 2.533e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.8196649 1.6556519
## sample estimates:
## mean of x mean of y
##  9.015952  7.778294
```

Variance by decomposition

Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$. Let $X + Y = r$.

1. Find the variance of $X|r$ by using the variance of a known distribution (See 3.9.2 in the Stat 110 book for a hint).

2. Find the variance of $X|r$ by using the fact that $\text{Var}(X + Y|r) = 0$ and treating X and Y as the sum of Bernoulli random variables. Verify that the two answers are the same. (Hint: Once you get to the Bernoulli random variables, think about how knowing the sum is r makes p irrelevant.)

Everything everywhere all at once (all two-sample continuous comparisons)

Let $X_1, \dots X_{n_1} \sim \text{Exp}(1/\mu_1)$ and $Y_1, \dots Y_{n_2} \sim \text{Exp}(1/\mu_2)$.

1. Name three tests we've learned so far that would not be applicable for comparing the Xs and Ys.

2. We'll be comparing the Type I and II error for the following tests: a two sample t test, a log-transformed t-test, a rank-based test, and a permutation test. We'll consider two scenarios:

- First, $n_1 = 5$, $n_2 = 15$ with $\mu_1 = \mu_2 = 5$ when calculating the Type I error rate and $\mu_1 = 5$ and $\mu_2 = 3$ when calculating the Type II error rate.
- Second, the same as before but with $n_1 = 20$.

Why should we use $\mu_1 = \mu_2$ when calculating the Type I error rate but $\mu_1 \neq \mu_2$ when calculating the Type II error rate?

3. Compare the results. What has the highest power? Which maintain their nominal false positive rates? Which test is best in which situations?

Test (First scenario)	Type I	Type II
t test	0.074	0.949
log t test	0.061	0.840
Rank test	0.043	0.885
Permutation test	0.053	0.780

Test (Second scenario)	Type I	Type II
t test	0.044	0.696
log t test	0.044	0.796
Rank test	0.047	0.774
Permutation test	0.051	0.743

4. What assumptions do we need for each test and what hypotheses are we testing?

5. The following simulation uses the same set-ups as above to calculate a t-based confidence interval for the difference in means, a t-based confidence interval for the ratio of medians, a percentile bootstrap interval for the difference in means, and a reversed percentile bootstrap interval for the difference in means. Shown below are the coverage probability and interval width for each. Comment on the results.

Interval (First scenario)	Coverage probability (means different)	Interval width (means different)	Coverage probability (means same)	Interval width (means same)
t interval	0.902	10.83	0.935	11.38
Transformed t interval	0.939	8.71	0.942	5.25
Percentile bootstrap	0.821	7.30	0.860	8.33
Rev. Perc. bootstrap	0.824	7.30	0.876	8.33

Interval (Second scenario)	Coverage probability (means different)	Interval width (means different)	Coverage probability (means same)	Interval width (means same)
t interval	0.951	5.40	0.953	6.81
Transformed t interval	0.953	3.80	0.949	2.28
Percentile bootstrap	0.915	4.92	0.915	6.18
Rev. Perc. bootstrap	0.925	4.92	0.932	6.18

6. Based on the results above, which is the best confidence interval to use?

7. Imagine now that we wanted to construct a confidence interval for μ_1 by using a studentized bootstrap interval. If we knew the data were distributed exponentially, what's one small change we could make to the confidence interval for μ_1 so that the interval is equally as wide or narrower while keeping the same confidence level?