DATA ANALYSIS

Week 11: Special cases (one sample z and t-test)

today's agenda

hypothesis testing with a single mean

general hypothesis testing approach

- we want to explain some variation in the world/population
- we take a sample and obtain a sample statistic based on an underlying model
- we then assume that our model is wrong / unnecessary
- we create a (hypothetical) sampling distribution of the sample statistic
- we ask what are likely and unlikely values of the sample statistic under this assumption
- we assess the likelihood of the observed sample statistic under this sampling distribution
- we make the decision to reject or fail to reject the “null” assumption

our progress so far

data = model + error

thus far

- data = mean + error
- data = X (interval/ratio) + error
- data = X (nominal) + error

after midterm 2

- data = X + Y + error
- data (NOIR) = model (NOIR) + error

our progress so far

data = model + error

correlation t-test
or
regression F-test

thus far

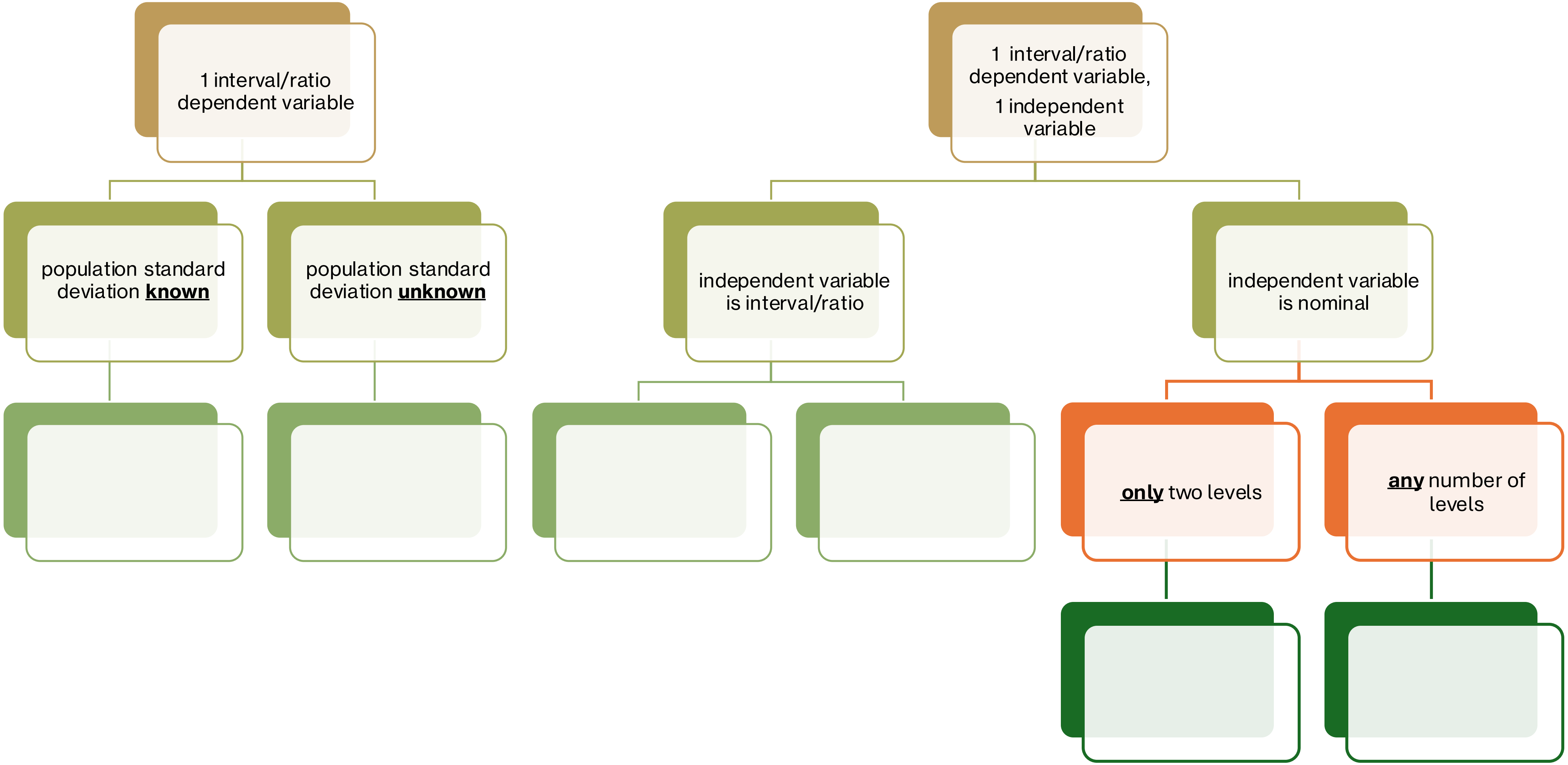
- data = mean + error
- data = X (interval/ratio) + error
- data = X (nominal) + error

ANOVA F-test

after midterm 2

- data = X + Y + error
- data (NOIR) = model (NOIR) + error

hypothesis testing flowchart



sampling with means

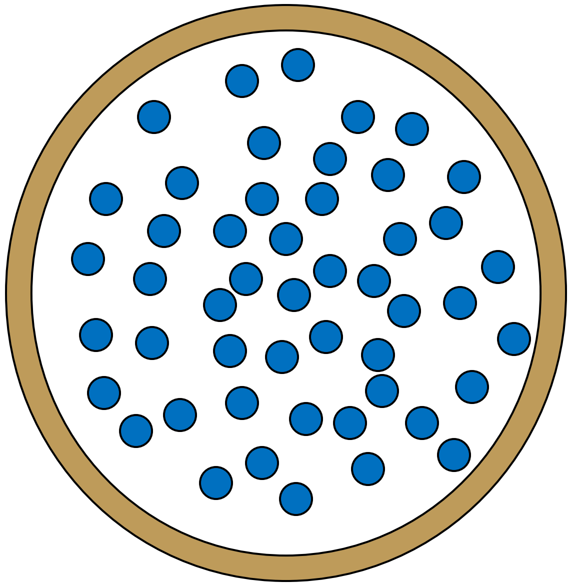
- the average mother sea turtle lays $\mu = 80$, $\sigma = 6$ eggs per mating season. We work for an endangered species foundation, and are testing the effectiveness of a new hormone (X15) on turtle fertility. We predict that turtles treated with the hormone will produce *different* nest sizes from the average turtle (no direction). We collect a sample of n=30 turtles from the above population and treat them with the hormone. We then count the number of eggs in their nest and get a mean of 84.
- is the fertility hormone effective?

framing the problem

- population characteristics (usual turtles): $\mu = 80, \sigma = 6$
- sample characteristics (our 30 turtles): $M = 84$
- two possible explanations for the difference in sample mean (M) and population mean (μ)
  - sampling error (H_0: null hypothesis)
  - true effect of hormone (H_1:alternative hypothesis)
- we assume the null hypothesis is true and set out to reject this assumption with our evidence

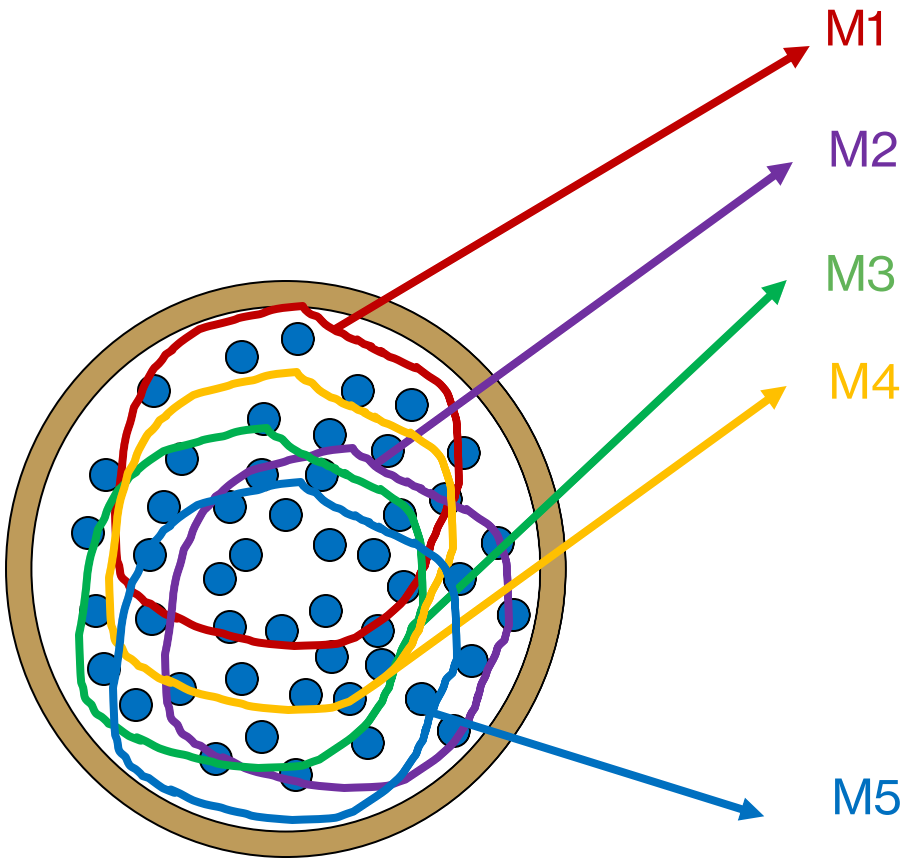
sampling with means

- we know some things about the population
- $\mu = 80, \sigma = 6$
- we could take random samples from this original population and examine the distribution of means
- sampling distribution of means



sampling with means

- we know some things about the population
- $\mu = 80, \sigma = 6$
- we could take random samples from this original population and examine the distribution of means
- sampling distribution of means
- once we know the form of this sampling distribution (t / F / normal), we can assess the probability of obtaining a sample as or more extreme as ours



central limit theorem (CLT)

- the central limit theorem states that for **any** population with mean μ and standard deviation σ , the distribution of sample means for sample size *n* will have:
  - a mean of $\mu_M = \mu$ = expected value of M
  - a standard deviation of $\sigma_M = \frac{\sigma}{\sqrt{n}}$ = standard error of the mean or M
  - will approach a normal distribution as n approaches infinity
  - distribution of sample means will be normally distributed **even if the population was not normally distributed (if n is large enough!)**
  - typically n (number of scores in a sample) around 30 yields a reasonably normal distribution
- CLT only applies to the distribution of sample means, i.e., if our sample statistic is NOT a mean, and our hypotheses are NOT about means, then we must use a different sampling distribution for the null hypothesis

typical sampling distributions

sample statistic	sampling distribution	standard error
means	normal (central limit theorem)	$\sigma_M = \frac{\sigma}{\sqrt{n}}$
slopes / correlation	Student's *t* distribution	$SE_r = s_r = \sqrt{\frac{1 - r^2}{n - 2}}$ $SE_{model} = \sqrt{\frac{SS_{error}}{n - 2}}$
ratio of squared errors	F-distribution	$F_{observed} = \frac{MS_{model}}{MS_{error}}$

step 1: stating the hypotheses

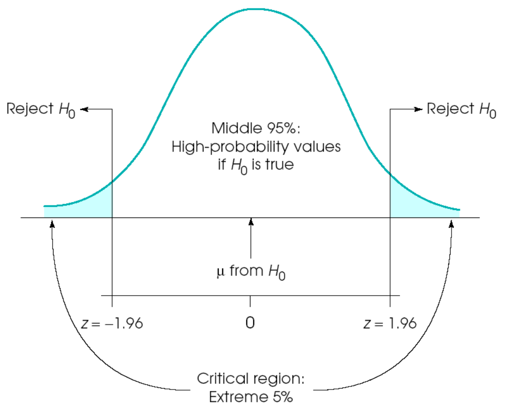
- null hypothesis: X15 does not have an effect on nest size, i.e., it stays the same and any variation observed in samples is simply due to sampling error

$$H_o: \mu = 80$$

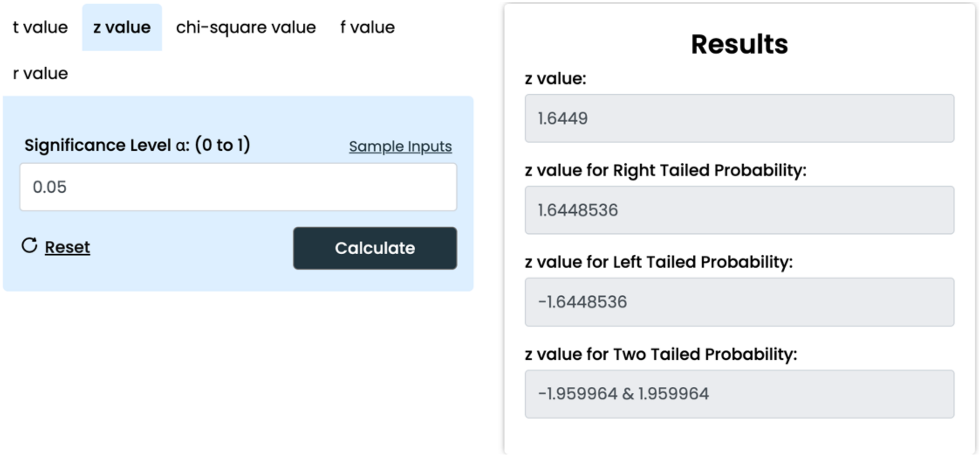
- alternative hypothesis: X15 has an effect on nest size

$$H_1: \mu \neq 80$$

step 2: set criteria for decisio

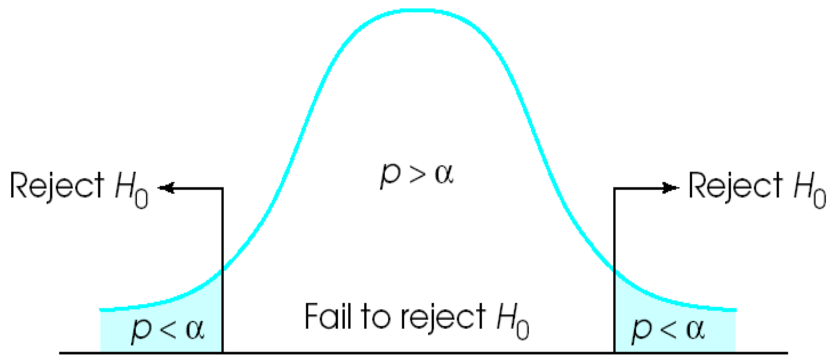


- examine the distribution of sample means
- the central limit theorem states that for **any** population with mean μ and standard deviation σ , the distribution of sample means for sample size *n* will have:
  - a mean of $\mu_M = \mu$ = expected value of M = 80
  - a standard deviation of $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{30}}$ = 1.095
- represent the problem graphically
- criteria: α-level = .05 (extreme 5%)
- find $z_{critical}$ = ±1.96, $p_{critical}$ = .05 (calculator)
- for a p-value smaller than .05, the obtained sample mean is unlikely to have come from this expected sampling distribution for the null hypothesis



step 3 and 4: collect data and decide!

- we collect the data and evaluate the probability of obtaining the data as extreme as this, under the null hypothesis
  - P (data | null)
  - compute z-score of sample mean under the sampling distribution
  - $z_{observed} = \frac{M - \mu}{\sigma_M} = \frac{84 - 80}{1.095} = 3.65$
  - remember that $z_{critical} = \pm 1.96$
  - look up the probability, $p_{observed} < .001$
  - this sample is very rare if the null hypothesis was true
- conclusion: we reject the null hypothesis that the hormone does not produce a difference
- reporting: X15 has a significant effect on sea turtle fertility (*M* = 84, *z* = 3.65, p < .001)



W11 Activity 1

- complete on Canvas

W11 Activity 1a

- If the final exam scores for the population have a standard deviation of $\sigma = 12$, does the sample provide enough evidence to conclude that the new online course is significantly different from the traditional class? Use a two-tailed test with $\alpha = .05$.

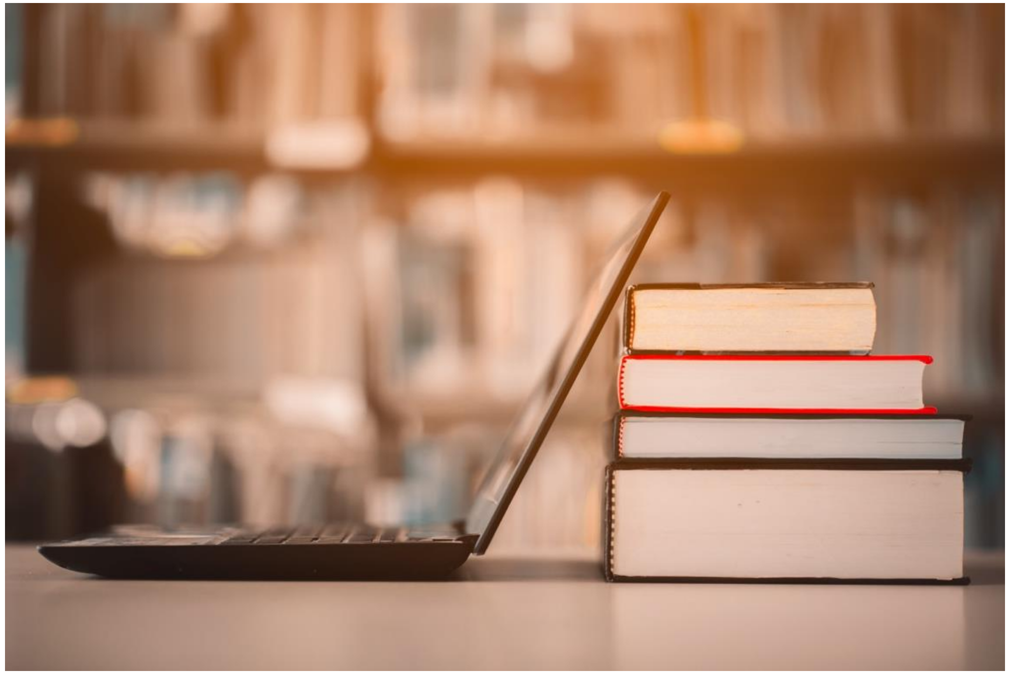
step 1: stating the hypotheses

- null hypothesis: the new online course does not produce any change in scores compared to the traditional class and any variation observed in samples is simply due to sampling error

$$H_o: \mu = 71$$

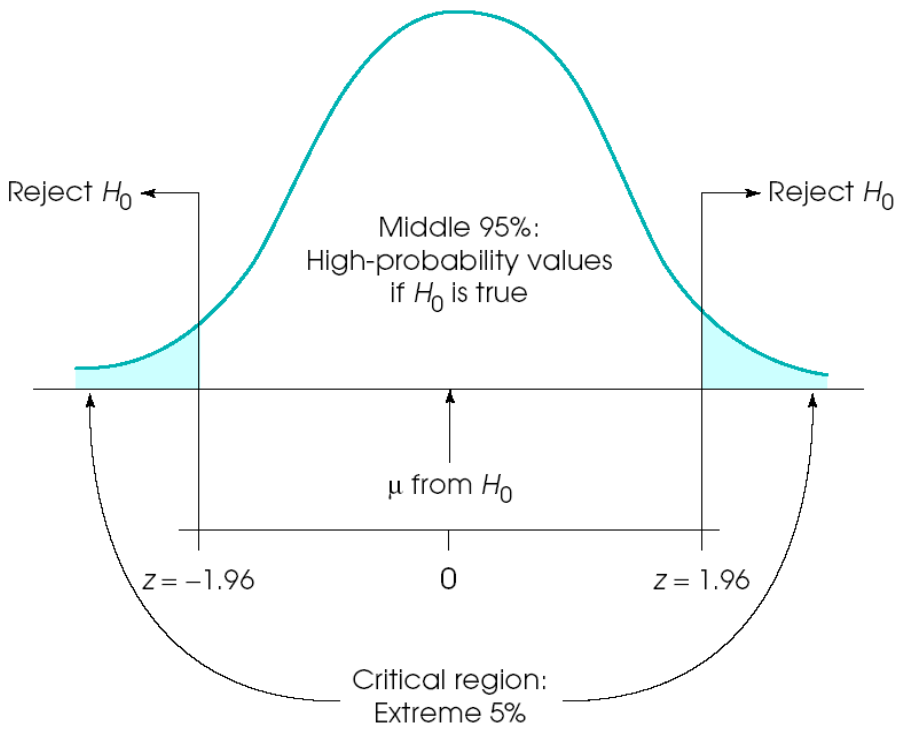
- alternative hypothesis: the new online course is significantly different from the traditional class

$$H_1: \mu \neq 71$$



step 2: set criteria for decision

- examine the distribution of sample means
- n = 36
- $\mu_M = \mu$ = expected value of M = 71
- $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2$
- criteria:
  - α-level = .05
  - $z_{critical} = \pm 1.96$ (extreme 5%)



step 3 and 4: collect data and decide!

- compute z-score of this mean under the sampling distribution
  - $z_{observed} = \frac{M - \mu}{\sigma_M} = \frac{76 - 71}{2} = 2.5$
  - $p_{observed} = .012$
  - remember that $z_{critical}$ = ±1.96 and alpha = .05
- this sample is very rare if the null hypothesis was true
- conclusion: we reject the null hypothesis
- reporting: online course has a significant effect on scores (z = 2.5, p = .012)

Z=2.5

The two-tailed P value equals 0.0124

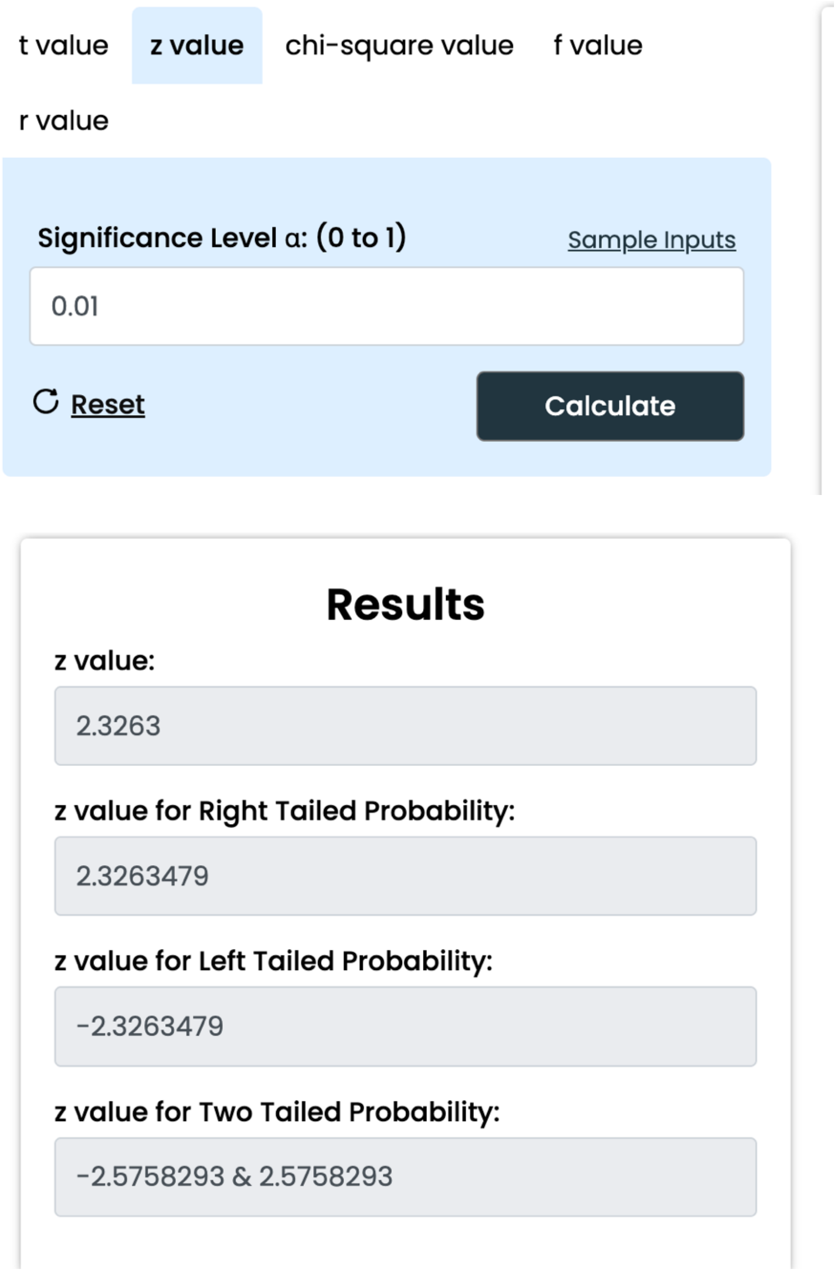
By conventional criteria, this difference is considered to be statistically significant.

W11 Activity 1b

- If the final exam scores for the population have a standard deviation of $\sigma = 12$, does the sample provide enough evidence to conclude that the new online course is significantly different from the traditional class? Use a two-tailed test with $\alpha = .01$.

what changes?

- step 1: stating the hypotheses (same!)
- step 2: set criteria for decision
  - n = 36
  - $\mu_M = \mu$ = expected value of M = 71
  - $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2$
  - $z_{critical}$ based on alpha = .01 = $\pm$ 2.58
  - $p_{critical}$ = .01
- step 3: collect data (same!)
  - $z_{observed} = \frac{M - \mu}{\sigma_M} = \frac{76 - 71}{2} = 2.5$
  - $p_{observed} = .012$
- step 4: decide
  - $z_{observed} < z_{critical}$ and $p_{observed} > p_{critical}$
  - cannot reject the null hypothesis!



W11 Activity 1c

- If the population standard deviation is $\sigma = 18$, is the sample sufficient to demonstrate a significant difference? Again, use a two-tailed test with $\alpha = .05$.

what changes?

- step 1: stating the hypotheses (same!)
- step 2: set criteria for decision
  - n = 36
  - $\mu_M = \mu$ = expected value of M = 71
  - $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{36}} = 3$
  - $z_{critical} = \pm 1.96$
- step 3: collect data
  - $z_{observed} = \frac{M - \mu}{\sigma_M} = \frac{76 - 71}{3} = 1.67$
  - $p_{observed} = .0475 + .0475 = .095$
- step 4: decide
  - $z_{observed} < z_{critical}$
  - cannot reject the null hypothesis!

Z=1.67

The two-tailed P value equals 0.0949

W11 Activity 1d

- If the population standard deviation is $\sigma = 18$, is the sample sufficient to demonstrate that online courses **increases** the score compared to traditional class? Use a one-tailed test with $\alpha = .05$.

what changes?

- **step 1: stating the hypotheses**
  - H_1: $\mu > 71$ and H_0: $\mu \leq 71$
- **step 2: set criteria for decision**
  - n = 36, $\mu_M = \mu$ = expected value of M = 71
  - $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{36}} = 3$
  - $z_{critical}$ = +1.65 (only one side)
- **step 3: collect data**
  - $z_{observed} = \frac{M - \mu}{\sigma_M} = \frac{76 - 71}{3} = 1.67$
  - $p_{observed} = \frac{.0949}{2} = .0475$
- **step 4: decide**
  - $z_{observed} > z_{critical}$
  - we can reject the null hypothesis if we use a one-tailed test!

Results

z value:

1.6449

z value for Right Tailed Probability:

1.6448536

z value for Left Tailed Probability:

-1.6448536

z value for Two Tailed Probability:

-1.959964 & 1.959964

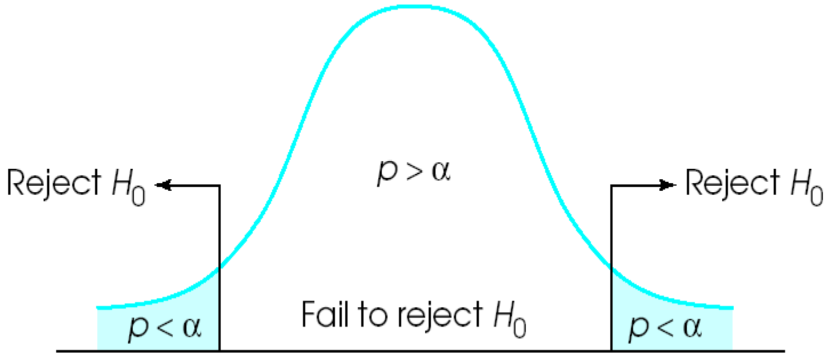
Z=1.67

The two-tailed P value equals 0.0949

summary

- making the critical region region smaller (by decreasing the α-level) makes the z-test more conservative, i.e., $z_{critical}$ will be higher, making it harder to reject the null hypothesis
- higher standard errors (due to high population variance σ or low sample size n) will impact the sample z-score ($z_{observed}$), i.e., where the sample statistic lies relative to the distribution

$$z = \frac{M - \mu}{\sigma_M} \text{ and } \sigma_M = \frac{\sigma}{\sqrt{n}}$$



Cohen's d: effect size for means

- hypothesis testing depends on sample sizes - higher sample sizes typically lead to statistically significant effects, even if those effects may be small
- effect sizes are a way to quantify the magnitude of the effect independent of sample size
- Cohen's d: effect size for means
- Cohen's d = $\frac{mean\ difference}{standard\ deviation} = \frac{M - \mu}{\sigma}$

Magnitude of *d*	Evaluation of Effect Size
$d = 0.2$	Small effect (mean difference around 0.2 standard deviation)
$d = 0.5$	Medium effect (mean difference around 0.5 standard deviation)
$d = 0.8$	Large effect (mean difference around 0.8 standard deviation)

one sample: z to t-distribution

- inferential statistics = from samples to populations
- but…for a one-sample z-test, we need to know the mean (μ) and standard deviation (σ) of the population! this information is usually not known!
- when σ is unknown and we have to rely on sample standard deviation (s) as an estimator, we cannot use the normal distribution as our sampling distribution

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \text{ vs. } s_M = \frac{s}{\sqrt{n}}$$

- we instead use the student's *t* distribution
- $estimated\ Cohen's\ d = \frac{M-\mu}{s}$

z-test vs. one-sample t-test

z-tests

- **when**: population mean and standard deviation are known
- **want to compare**: sample mean to population mean

one sample t-test

- **when**: population standard deviation is unknown
- **want to compare**: sample mean to population mean

W11 Activity 2

- complete on Canvas

W11 Activity 2a

- research examining the effects of preschool childcare has found that children who spent time in day care, especially high-quality day care, perform better on math and language tests than children who stay home with their mothers (Broberg, Wessels, Lamb, & Hwang, 1997). In a typical study, a researcher obtains a sample of $n = 10$ children who attended day care before starting school. The children are given a standardized math test for which the population mean is $\mu = 50$. The scores for the sample are as follows: 53, 57, 61, 49, 52, 56, 58, 62, 51, 56.
- Is this sample sufficient to conclude that the children with a history of preschool day care are significantly different from the general population? Use a two-tailed test with $\alpha = .01$.

framing the problem

- population
  - mean μ = 50
  - **standard deviation (σ) unknown, we cannot use a z-test**
- sample
  - n = 10, sample scores: X = 53, 57, 61, 49, 52, 56, 58, 62, 51, 56
  - sample mean: $M = \frac{\sum X}{n}$ = 55.5
  - sample standard deviation: $s = \sqrt{\frac{(X-M)^2}{n-1}} = 4.249183$
- we need to use a t-test!
  - degrees of freedom: $df = n - 1 = 9$
  - $s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{(4.249183)^2}{10}} = 1.34371$

sheets solution

hypothesis testing

- step 1: stating the hypotheses
  - H_0: μ = 50; H_1: $\mu \neq 50$
- step 2: setting decision criteria
  - two-tailed test (t-value calculator)
  - $t_{critical}(9) = \pm 3.2498$ for $\alpha = .01$
- step 3: collect data
  - $t_{observed} = \frac{M - \mu}{s_M} = \frac{55.5 - 50}{1.34371} = 4.09$
  - $p_{observed} = 0.0027$ (obtained from p-value calculator for t-score)
- step 4: decide!
  - $p_{observed} < .05\ and t_{observed} > t_{critical}(9)$
  - reject H_0 and conclude that children with a history of preschool day care are significantly different from the general population, *t* (9) = 4.09, *p* = .003.

sheets solution

W11 Activity 2b: Cohen's *d*

- $estimated\ d = \frac{M-\mu}{s}$
- population mean μ = 50
- sample mean: $M = \frac{\sum X}{n}$ = 55.5
- sample standard deviation: $s = \sqrt{\frac{(X-M)^2}{n-1}} = 4.249183$
- d = $\frac{M-\mu}{s} = \frac{55.5-50}{4.249} = 1.29$
- Math achievement scores for children with a history of preschool day care are significantly different from the general population, t (9) = 4.09, p = .003, d = 1.29.

W11 Activity 2c and 3d

3 Multiple Choice 1 point

Let's say I re-ran the study another 100 times. The distribution of sample means I will obtain from each of those studies will be:

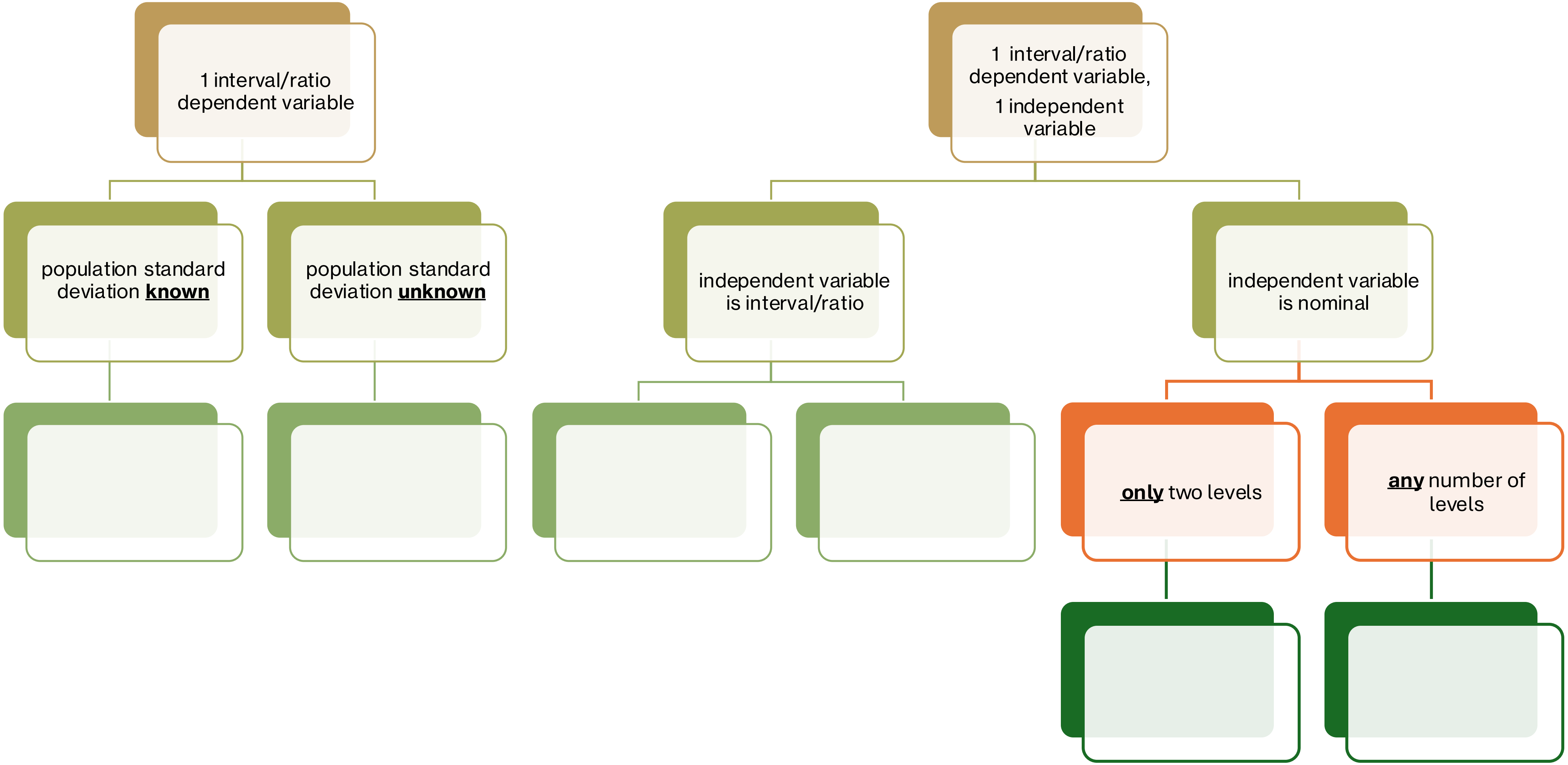
- Normal
- t-distributed
- F-distributed
- Cannot be determined

4 Multiple Choice 1 point

Let's say I re-ran the study another 100 times. If there was NO true effect of daycare in the population, the distribution of sample means I will obtain from each of those studies will have the mean:

- 50
- 55.5
- 10
- Impossible to know

hypothesis testing flowchart



next time

Here are the to-do's for this week:

- Submit Week 11 Quiz
- Submit Problem Set 5
- Complete Practice Midterm 2 (Conceptual)
- Submit any lingering questions here!
- Extra credit opportunities:
  - Submit Exra Credit Questions
  - Submit Optional Meme Submission

Before Tuesday

- [optional but encouraged] Read Chapters 7, 8, and 9 from the Gravetter & Wallnau (2017) textbook.

Before Thursday

- [optional but encouraged] Read Chapter 10 from the Gravetter & Wallnau (2017) textbook.
- Watch: Hypothesis Testing (Independent Sample t-test).
  - Practice Data
  - Solution Sheet

After Thursday

- See Apply section.