Notes

Week 3

Module 2 Week 3B

Hypothesis Testing I

Hypothesis testing with one sample

- E.g.:
    - The average temperature in Maryland in July is 88 degrees.
    - A new Honda Accord gets 38 miles per gallon.
    - 64% of all college students graduate in four years.
    - A new lawyer makes an average of $100,000 per year.
    - 70% of all U.S. households have internet access
- These types of claims can be assessed statistically using a *sample of data* to determine if there is sufficient evidence to support them by doing a *Hypothesis Test*
- First, we will cover hypothesis tests about a *single mean or a single proportion.*
- There is a straightforward approach to hypothesis testing, however, remember that we are using a *random sample of data*, and therefore, we can make errors.
- The approach:
    1. Set up *two conflicting hypotheses* (*Always* about the *Population Parameter* of interest, *not* the *Sample Statistic*!)
        - **Null hypothesis (H_0)**: statement of no difference or no effect
            - H_0 will always have $=$ or $\leq$ or $\geq$ (NK: it is most accurate for the Null to only have an "$=$" sign)
        - **Alternative hypothesis (H_a)**: contradicts H_a
            - H_a will always have $\neq$ or $>$ or $<$
            - **Two-Tailed Test**: H_a has $\neq$
            - **Two-Tailed Test**: H_a has $>$ or $<$
    2. Collect a random sample of data
    3. Determine the correct distribution to use for the test
        - Tests about *population means* and *proportions* use a *test statistic* that follows a *normal distribution* or a *Student's t-distribution*.
            - Recall:
                - For one sample means:

- If population SD is *known*: $\bar{X} \sim N(\text{mean} = \mu, \text{SD} = \frac{\sigma}{\sqrt{n}})$
- If population SD is *unknown*: $\bar{X} \sim t_{df}$

- For one sample estimated proportion p': $p' \sim N(\text{mean} = p, \text{SD} = \sqrt{\frac{pq}{n}})$, where $p = \mathbb{P}(\text{"success"})$

4. Analyze/assess the data to determine which hypothesis is supported
   - Determine which hypothesis is supported with the sample
5. Reject or fail to reject hypothesis
   - Think of the statement in H_0 as an assumption
   - We collect a sample of data to determine if it supports the assumption.
   - If the *sample* has properties that are very unlikely given the assumption, there is a good chance that H_0 is incorrect, and therefore, it is rejected.
   - The *sampling distribution* used for a hypothesis test is the sampling distribution *under the assumption that H_0 is true*.
   - We will use a *p-value* to make our decision (reject or fail to reject H_0).
   - A **p-value** is *the probability, under the assumption that H_0 is true,that another random sample would produce results as extreme or more extreme as the results obtained in the given sample.* (I.e. The $\mathbb{P}$ that the results would as adverse or more adverse to the assumption in the *Null Hypothesis*)
   - **Rule: reject Ho if p-value** $< \alpha$
     - Ideally, want to re-perform the test with new random samples.
   - E.g. p-value $= 0.12$
     - The probability of getting results as adverse to H_0 as the ones obtained in the sample used is 0.12.
   - A *large* p-value supports H_0 (I.e. it is likely we would get this result again from another random sample if H_0 is true.)
   - A *small* p-value does not support H_0 (I.e. it is NOT likely we would get this result again from another random sample if H_0 is true.)

- *Assumptions for Hypothesis Tests*:
  - For one sample mean (i.e. *t-test* assumptions):
    - Random sample assumed
    - Variable of interest X is approximately Normally Distributed
      - But: if n is large enough, *t-test* works even if X is not Normally Distributed
    - s used to approximate σ
  - For one sample proportion:
    - Random sample assumed
    - Variable of interest has a *Binomial Distribution*

        - $np > 5$ AND $nq > 5$ to use normal approximation
- Errors can be made when conducting a hypothesis test:
    - **Type I error**: H_0 is rejected when it is true
        - $\mathbb{P}(\text{type I error}) = \text{level of significance} = \alpha$
            - The standard α is 0.05 but is arbitrary and context-specific
    - **Type II error**: H_0 is not rejected when it is false
        - $\mathbb{P}(\text{type II error}) = \beta$
        - **Power** $= \mathbb{P}(\text{"reject the } H_0 \text{ when the } H_0 \text{ is false"}) = 1 – \beta$
            - Increasing sample size can increase *Power*.
    - The two types of errors are *inversely related*; more of one is less of the other and vice-versa.(
    - Which error is more important is context-specific; e.g. if life-and-death of whether $H_0 : \text{drug has no effect}$ and $H_a : \text{drug saves lives}$, *Type I error* can be more acceptable than *Type II error*.
    - If sample size is *large*, it is easier to reject H_0, so a *smaller* $\text{level of significance} = \alpha$ should be specified and vice-versa

- E.g. The average annual salary of a new lawyer is $100,000$
    - H_0: $\mu = 100,000$
    - H_a: $\mu \neq 100,000$
    - Suppose we collect a random sample of 36 new lawyers and calculate a sample mean of $105,000
      and a sample standard deviation of 12,000.
    - This is a test about a mean and the population *standard deviation* is *unknown*, therefore, the appropriate distribution is the *t-distribution with 35 degrees of freedom*.
    - If H_0 is true (i.e. $\mu = 100,000$), how many standard deviations is our sample result (i.e. $\bar{X} = 105,000$) from the population mean?
    - (Because test about a *Mean* AND have to estimate the *SD*) Test-statistic (t-stat) $= \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{105000-100000}{12000/\sqrt{36}} = \frac{5000}{2000} = 2.5$
    - P-value $= 0.017$ <-area to the right of 2.5 ($= 0.086$) + area to the left of -2.5 ($= 0.086$), because we have a *Two-Tailed Test*. Interpretation: it is very *unlikely* we would observe this difference if the *Null Hypothesis* were true =>
    - **Rule: reject Ho if p-value** $< \alpha$
    - $0.017 < 0.05$ => reject H_0 (but would *not* reject if $\alpha = 0.01$)
- E.g. 70% of U.S. households have internet access
    - H_0: $p = 0.70$
    - H_a: $p \neq 0.70$
    - Suppose we collect a random sample of 81 households and calculate a sample proportion of 0.66.
    - This is a test about a proportion, therefore, the appropriate distribution is the *normal distribution*.

- If H_0 is true (i.e. $p = 0.70$), how many standard deviations is our sample result (i.e. $p\prime = 0.66$)
  from the population mean?
- Test-statistic (z-stat) $= \frac{p' - p}{\sqrt{pq/n}} = \frac{0.66 - 0.70}{\sqrt{0.70 * 0.30/81}} = \frac{-0.04}{0.051} = -0.784$
- p-value $= 0.43$ = $\text{Area to the left of -}0.784 + \text{Area to the right of } 0.784$
- **Rule: reject H_0 if p-value $< \alpha$**
- $0.43 > 0.05$ => fail to reject H_0