Exam 2 Study Guide

Materials

- Calculator (no graphing calculator)
- Pencil and eraser
- UCLA Student ID
- **<u>ONE</u>** 8.5x11 cheat sheet (both sides). Handwritten or typed font Arial size 12. Only formulas and definitions. Hand drawn graphs or plots are acceptable. No examples or problems. No lecture slides or screenshots!

- The exam will be all multiple choice. It will be on Thursday, June 6 in class (same time and place as lecture).
- The exam covers Chapters 6 through 9.

Chapter 6

- Discrete vs Continuous
- Discrete Probability Distribution Function (PDF) - table and graph
- Continuous Probability Distribution Function
- Normal Model - continuous
  - Make sure you know how to write the shorthand, draw the normal curve, shade the area, and use the z-table.
  - Finding probabilities, z-scores, z-table, inverse normal (measurement from percentile), left area vs right area.
- Binomial Model - discrete
  - Checking conditions.
  - Make sure you know how to write the shorthand, how to graph and how to use the binomial table.
  - How to find the center and spread?

Chapter 7 (Ch 7.1-7.3)

- Survey terminology: population, parameter, census, sample, statistics
- Population size vs Sample size
- Notation for population mean and standard deviation vs sample mean and standard deviation
- Sampling Inference
- Bias - Make sure you know how to identify it!
- Simple Random Sampling (SRS) - precision, accuracy
- Sampling Distributions - If you were to take many samples from a population and you would like to look at a statistic, say the proportion (probability of the event's success) for each sample, then the probability distribution of that statistic (the proportion in this case) is called a sampling distribution. The standard deviation is called the standard error. Know how to find the standard error.
- Sample proportion ($\hat{p}$) vs true population proportion (p)
  - The sample proportion is the proportion (probability of success) for the sample.
  - The population proportion the proportion (probability of success) for the population.
- Central Limit Theorem (CLT) - conditions, normal distribution (shorthand), standard deviation vs standard error

Chapter 7 (Ch 7.4-7.5)

- Estimating the population proportion (p) with confidence intervals
  - standard error
  - margin of error
  - finding critical z-score
  - checking CLT conditions: random and independent, large sample (at least 10 successes/failures), big population
  - setting up and finding confidence intervals
  - interpreting confidence intervals
- Confidence intervals for the difference between two population proportions
  - We use this when we have two populations to compare.
  - checking CLT conditions: random and independent, large samples (at least 10 successes/failures), big populations, independent samples
  - setting up and finding confidence intervals for two proportions
  - interpreting confidence intervals for two proportions

Chapter 8

- Hypothesis testing for one proportion
- What are null and alternative hypotheses?
- One-tailed vs Two-tailed (or one-sided vs two-sided)
- What is a p-value? How do we graphically show a p-value?
- What is a test statistic?
- What is the relationship between a p-value and test statistic?
- What type of mistakes that can be made in hypothesis testing?
- What is the significance level?
- Hypothesis testing for two proportions

Chapter 9

- accuracy (measured by bias) and precision (measured by standard error)
- population mean vs sample mean; population standard deviation vs sample standard deviation
- CLT for means
- Different distributions (population, sample, sampling)
- t-distribution and t-statistic
- degrees of freedom
- Confidence intervals for one mean
- Hypothesis testing for one mean
- Independent vs Dependent/Paired samples
- Confidence intervals for 2 sample means (independent samples)
- Hypothesis testing for 2 sample means (independent samples)
- Confidence intervals for mean of a difference (dependent samples)
- Hypothesis testing for mean of a difference (dependent samples)

One Proportion

- CLT conditions:
  - Random and Independent: The sample is collected randomly from the population, and the observations are independent of each other. The sample can be collected either with or without replacement.
  - Large Sample: The sample size, n, is large enough that the sample expects at least 10 successes (yes's) and 10 failures (no's). $np \geq 10$ and $n(1-p) \geq 10$. If you don't have p then use the sample proportion instead $\hat{p}$.
  - Big Population: If the sample is collected without replacement, then the population size must be at least 10 times bigger than the sample size.
- Formula for Confidence Interval:

$$\hat{p} \pm ME = \hat{p} \pm z^* \times SE(\hat{p}) = \hat{p} \pm z^* \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- General Interpretation: We are ??% confident that the true population proportion lies in the interval (?, ?).

One Proportion

- Hypothesis Testing:
  - State the null and alternative hypotheses.

    H_0:$p=p_0$ and H_a:$p>p_0$ or $p<p_0$ or $p\neq p_0$
  - Type of test: one-proportion z-test
  - Check CLT conditions
  - Find the test statistic (z-score).

$$z = \frac{\text{observed value - null value}}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

  - Find the p-value.
    - If H_a has a < sign, p-value = Prob($z<z_{obs}$).
    - If H_a has a > sign, p-value = Prob($z>z_{obs}$).
    - If H_a has a ≠ sign, p-value = Prob($z<|z_{obs}|$)= Prob($z<-z_{obs}$)+ Prob($z>z_{obs}$). Make sure you are looking at the tails, you want the probability of both tails.
  - Conclusion: Either reject or fail to reject the null hypothesis and put your conclusion in context.
    - If p-value < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
    - If p-value > α, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible.

Two Proportions

- CLT conditions
  - Random and Independent: Both samples are randomly drawn from their populations and observations are independent of each other.
  - Large Samples: Both sample sizes are large enough that at least 10 successes and 10 failures can be expected in both samples. Here we use the pooled $\hat{p}$.

$$n_1\hat{p} \geq 10 \text{ and } n_1(1-\hat{p}) \geq 10$$
$$n_2\hat{p} \geq 10 \text{ and } n_2(1-\hat{p}) \geq 10$$

  - Big Populations: If the sample are collected without replacement, then both population sizes must be at least 10 times bigger than their samples.
  - Independent Samples: The samples must be independent of each other.
- Formula for Confidence Interval:

$$\left(\hat{p}_1-\hat{p}_2\right) \pm z^* \times SE_{est} = \left(\hat{p}_1-\hat{p}_2\right) \pm z^* \times \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1}+\frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}$$

- General Interpretation: If the CI contains 0, it means the population proportions could be equal. If it doesn't contain 0, then we are confident that the population proportions are not equal, and we should note whether the values in the interval are all positive (the first population proportion is greater than the second) or all negative (the first is less than the second).

Two Proportions

- Hypothesis Testing:
  - State the null and alternative hypotheses.

    H_0: $p_1 = p_2$ and H_a: $p_1 > p_2$ or $p_1 < p_2$ or $p_1 \neq p_2$
  - Type of test: two-proportion z-test
  - Check CLT conditions
  - Find the test statistic (z-score).

$$z = \frac{\text{estimator - null value}}{SE} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\text{where } \hat{p} = \frac{\text{number of successes in sample 1} + \text{number of successes in sample 2}}{n_1 + n_2}$$

  - Find the p-value.
    - If H_a has a < sign, p-value = Prob($z < z_{obs}$).
    - If H_a has a > sign, p-value = Prob($z > z_{obs}$).
    - If H_a has a ≠ sign, p-value = Prob($z < |z_{obs}|$)= Prob($z < -z_{obs}$)+ Prob($z > z_{obs}$). Make sure you are looking at the tails, you want the probability of both tails.
  - Conclusion: Either reject or fail to reject the null hypothesis and put your conclusion in context.
    - If p-value < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
    - If p-value > α, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible.

One Mean

- CLT conditions:
    - Random and Independent: The sample is collected randomly and the trials are independent of each other.
    - Large Sample:
        - If the population distribution is normal, then the sampling will be normal as well, regardless of the sample size.
        - OR if the population distribution is not normal, then we need a large enough sample ($n \geq 25$) to ensure that the sampling distribution will be normal.
    - Big Population: If the sample is collected without replacement, then the population size is at least 10 times the sample size.
- Formula for Confidence Interval:

$$\bar{x} \pm t_{df}^{*} \times SE(\bar{x}) = \bar{x} \pm t_{df}^{*} \times \frac{s}{\sqrt{n}} \qquad df = n - 1$$

Note: Since σ is typically unknown we use s and t* is found using the t-distribution.

- General Interpretation: We are ??% confident that the true population mean of <in context of problem> lies in the interval (?, ?).

One Mean

- Hypothesis Testing:
  - State the null and alternative hypotheses.

    $H_0: \mu = \mu_0$ and $H_a: \mu > \mu_0$ or $\mu < \mu_0$ or $\mu \neq \mu_0$
  - Type of test: one-sample t-test
  - Check CLT conditions
  - Find the test statistic (t-statistic).

$$t = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

  - Find the p-value. Find the degrees of freedom (df=n-1) and use the t-distribution table. You will get a range not an exact value based on the table.
  - Conclusion: Either reject or fail to reject the null hypothesis and put your conclusion in context.
    - If p-value < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
    - If p-value > α, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible.

Two Means (Independent)

- CLT conditions:
  - Random Samples and Independence: Both samples are randomly taken from their populations, or subjects are randomly assigned to one of the two groups, and each observation is independent of any other.
  - Independent Samples: The two samples are independent of each other (not paired or dependent).
  - Large Samples: The populations are approximately Normal, or the sample size in each sample is 25 or more.
- Formula for Confidence Interval:

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Note: The multiplier t* is based on an approximate t-distribution. The degrees of freedom (df) is equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

- Interpretation: If the interval doesn't include 0, then we have evidence that the two means are different from each other. If it contains all positive values, then we are confident that the first mean is greater than the second. If it contains all negative values, then we are confident that the first mean is less than the second. If the interval does include 0, then it is plausible that the two means are equal to each other.

Two Means (Independent)

- Hypothesis Testing:
  - State the null and alternative hypotheses.

    $H_0: \mu_1 = \mu_2$ and $H_a: \mu_1 > \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 \neq \mu_2$
  - Type of test: two-sample t-test
  - Check CLT conditions
  - Find the test statistic (t-statistic). $$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
  - Find the p-value. Find the degrees of freedom (smaller one of the two) and use the t-distribution table. You will get a range not an exact value based on the table.
  - Conclusion: Either reject or fail to reject the null hypothesis and put your conclusion in context.
    - If p-value < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
    - If p-value > α, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible.

Two Means (Dependent or Paired)

- With paired samples, we turn two samples into one, by finding the difference in each pair. Then we continue to compute a one-sample confidence interval and hypothesis test.
- CLT conditions:
    - Random and Independent: The sample is collected randomly and the trials are independent of each other.
    - Large Sample:
        - If the population distribution is normal, then the sampling will be normal as well, regardless of the sample size.
        - OR if the population distribution is not normal, then we need a large enough sample ($n \geq 25$) to ensure that the sampling distribution will be normal.
    - Big Population: If the sample is collected without replacement, then the population size is at least 10 times the sample size.
- Formula for Confidence Interval: Use the mean and standard deviation for differences.

$$\bar{x} \pm t_{df}^{*} \times SE(\bar{x}) = \bar{x} \pm t_{df}^{*} \times \frac{s}{\sqrt{n}} \qquad df = n - 1$$

- General Interpretation: We are ??% confident that the mean difference of <in context of problem> lies in the interval (?, ?).

Two Means (Dependent or Paired)

- Hypothesis Testing:
  - State the null and alternative hypotheses.

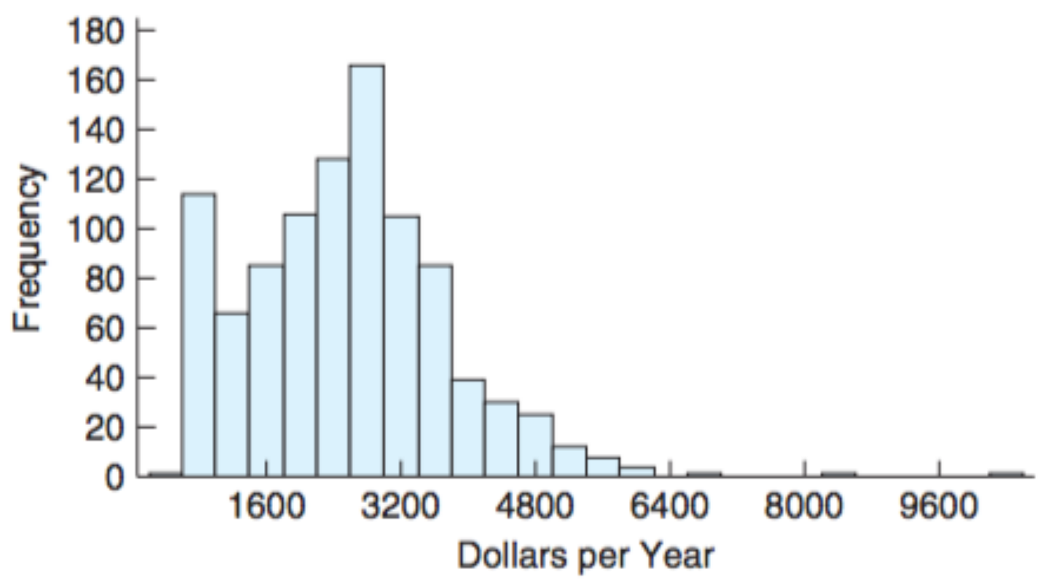
    H_0: $\mu_{\text{difference}}=0$ and H_a: $\mu_{\text{difference}}>0$ or $\mu_{\text{difference}}<0$ or $\mu_{\text{difference}}\neq 0$
  - Type of test: paired t-test
  - Check CLT conditions
  - Find the test statistic (t-statistic).

$$t = \frac{\bar{x}_{difference} - 0}{\frac{s_{difference}}{\sqrt{n}}}$$

  - Find the p-value. Find the degrees of freedom and use the t-distribution table. You will get a range not an exact value based on the table.
  - Conclusion: Either reject or fail to reject the null hypothesis and put your conclusion in context.
    - If p-value < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
    - If p-value > α, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible.

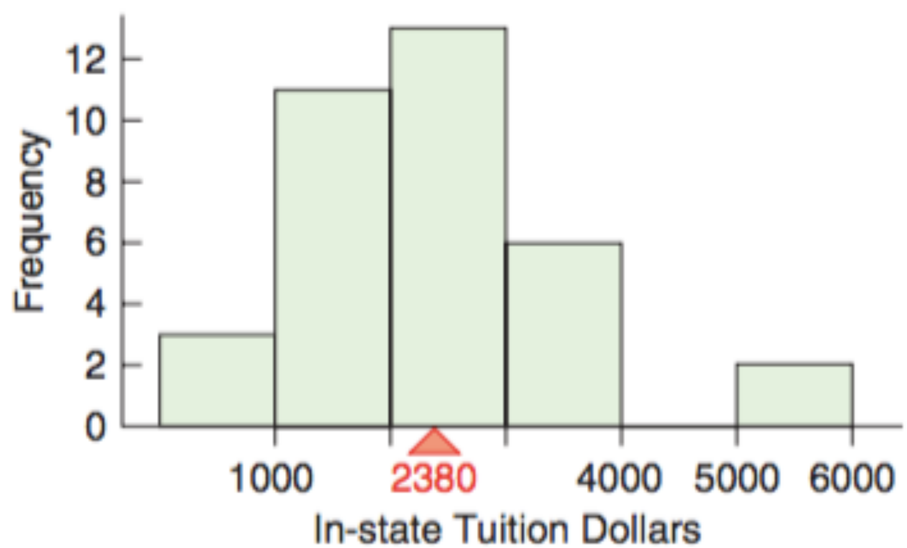
Review of Distributions

- The population distribution is the distribution of values from the population.
- The following histogram shows the distribution of in-state tuition and fees for all two-year colleges in the United States for the 2008-2009 academic year.
- It is an example of a population distribution because it shows the distribution of all two-year colleges.



Review of Distributions

- The distribution of the sample is the distribution of the values from the sample. If the sample size is large, and if the sample is random, then the sample will be representative of the population, and the distribution of the sample will look similar to the population distribution.
- We take a random sample of 30 colleges. The following histogram shows the distribution of this sample.



Review of Distributions

- The sampling distribution is more abstract. If we take a random sample of data and find the sample mean, and then repeat this many, many times, we will get an idea of what the sampling distribution looks like.
- We generate 200 samples of sample size 30 (histogram a) and generate another 200 samples of sample size 60 (histogram b).
- The more observations in our sample, the better an approximation the Normal distribution provides.