C9 and 10 Test Tip Sheet!
There's A LOT of info here. Make sure to pay attention to details.

Section 9.1

- **A significance test** assesses the evidence provided by data against a **null hypothesis** H_0 and in favor of an **alternative hypothesis** H_a.
- The hypotheses are usually stated in terms of population parameters. Often, H_0 is a statement of no change or no difference. The alternative hypothesis states what we hope or suspect is true.
- A **one-sided alternative** H_a says that a parameter differs from the null hypothesis value in a specific direction. A **two-sided alternative** H_a says that a parameter differs from the null value in either direction.

- The reasoning of a significance test is as follows. Suppose that the null hypothesis is true. If we repeated our data production many times, would we often get data as inconsistent with H_0, in the direction specified by H_a, as the data we actually have? If the data are unlikely when H_0 is true, they provide evidence against H_0 and in favor of H_a.
- The ***P*-value** of a test is the probability, computed supposing H_0 to be true, that the statistic will take a value at least as extreme as the observed result in the direction specified by H_a.
- Small *P*-values indicate strong evidence against H_0. To calculate a *P*-value, we must know the sampling distribution of the test statistic when H_0 is true.
- If the *P*-value is smaller than a specified value α (called the **significance level**), the data are **statistically significant at level α**. In that case, we can **reject** H_0 and say that we have convincing evidence for H_a. If the *P*-value is greater than or equal to α, we **fail to reject** H_0 and say that we do *not* have convincing evidence for H_a.
- A **Type I error** occurs if we reject H_0 when it is in fact true. In other words, the data give convincing evidence for H_a when the null hypothesis is correct. A **Type II error** occurs if we fail to reject H_0 when H_a is true. In other words, the data don't give convincing evidence for H_a, even though the alternative hypothesis is correct.
- In a fixed level α significance test, the probability of a Type I error is the significance level α.

	H_0 is true	H_a is true
Reject H_0	Type I error α	Correct Decision **Power**
Do not reject H_0	Correct Decision	Type II error β

Note that: Power + β = 1

Section 9.2

- The conditions for performing a significance test of $H_0: p = p_0$ are:
  - **Random:** The data were produced by a well-designed random sample or randomized experiment.
    - **10%:** When sampling without replacement, check that the population is at least 10 times as large as the sample.
  - **Large Counts:** The sample is large enough to satisfy $np_0 \geq 10$ and $n(1 - p_0) \geq 10$ (that is, the expected counts of successes and failures are both at least 10).
- The **one-sample z test for a population proportion** is based on the **test statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

with P-values calculated from the standard Normal distribution.

- Follow the four-step process when you are asked to carry out a significance test:

STATE: What *hypotheses* do you want to test, and at what *significance level*? Define any *parameters* you use.

PLAN: Choose the appropriate inference *method*. Check *conditions*.

DO: If the conditions are met, perform *calculations*.

- Compute the **test statistic**.
- Find the ***P*-value**.

CONCLUDE: Make a *decision* about the hypotheses in the context of the problem.

Section 9.3

- The conditions for performing a significance test of H_0: $\mu = \mu_0$ are:
  - **Random:** The data were produced by a well-designed random sample or randomized experiment.
    - **10%:** When sampling without replacement, check that the population is at least 10 times as large as the sample.
  - **Normal/Large Sample:** The population distribution is Normal *or* the sample size is large ($n \geq 30$). When the sample size is small ($n < 30$), examine a graph of the sample data for any possible departures from Normality in the population. You should be safe using a t distribution as long as there is no strong skewness and no outliers are present.
- The **one-sample t test for a mean** uses the test statistic

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

  with P-values calculated from the t distribution with $n - 1$ degrees of freedom.
- Confidence intervals provide additional information that significance tests do not—namely, a set of plausible values for the parameter μ. A two-sided test of $H_0: \mu = \mu_0$ at significance level α gives the same conclusion as a $100(1 - \alpha)\%$ confidence interval for μ.
- Analyze **paired data** by first taking the difference within each pair to produce a single sample. Then use one-sample t procedures.
- There are three factors that influence the sample size required for a statistical test: significance level, effect size, and the desired power of the test.
- Very small differences can be highly significant (small P-value) when a test is based on a large sample. A statistically significant difference need not be practically important.
- Many tests run at once will probably produce some significant results by chance alone, even if all the null hypotheses are true.

Section 10.1

- Choose independent SRSs of size n_1 from Population 1 with proportion of successes p_1 and of size n_2 from Population 2 with proportion of successes p_2. The sampling distribution of $\hat{p}_1 - \hat{p}_2$ has the following properties:
  - **Shape** Approximately Normal if the samples are large enough that n_1p_1, $n_1(1 - p_1)$, n_2p_2, and $n_2(1 - p_2)$ are all at least 10.
  - **Center** The mean is $p_1 - p_2$.
  - **Spread** As long as each sample is no more than 10% of its population, the standard deviation is $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$.
- Confidence intervals and tests to compare the proportions p_1 and p_2 of successes for two populations or treatments are based on the difference $\hat{p}_1 - \hat{p}_2$ between the sample proportions.
- Before estimating or testing a claim about $p_1 - p_2$, check that these conditions are met:
  - **Random:** The data come from two independent random samples or from two groups in a randomized experiment.
    - **10%:** When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples.
  - **Large Counts:** The counts of "successes" and "failures" in each sample or group—$n_1\hat{p}_1$, $n_1(1 - \hat{p}_1)$, $n_2\hat{p}_2$, and $n_2(1 - \hat{p}_2)$—are all at least 10.
- When conditions are met, an approximate $C\%$ confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

  where z^* is the standard Normal critical value with $C\%$ of its area between $-z^*$ and z^*. This is called a **two-sample z interval for $p_1 - p_2$**.
- Significance tests of $H_0: p_1 - p_2 = 0$ use the **pooled (combined) sample proportion** in the standard error formula:

$$\hat{p}_C = \frac{\text{count of successes in both samples combined}}{\text{count of individuals in both samples combined}} = \frac{X_1 + X_2}{n_1 + n_2}$$

When conditions are met, the **two-sample *z* test for $p_1 - p_2$** uses the test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\dfrac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \dfrac{\hat{p}_C(1-\hat{p}_C)}{n_2}}}$$

with *P*-values calculated from the standard Normal distribution.

- Inference about the difference $p_1 - p_2$ in the effectiveness of two treatments in a completely randomized experiment is based on the **randomization distribution** of $\hat{p}_1 - \hat{p}_2$. When conditions are met, our usual inference procedures based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately correct.
- Be sure to follow the four-step process whenever you construct a confidence interval or perform a significance test for comparing two proportions.

Section 10.2

- Choose independent SRSs of size n_1 from Population 1 and size n_2 from Population 2. The sampling distribution of $\bar{x}_1 - \bar{x}_2$ has the following properties:
    - **Shape:** Normal if both population distributions are Normal; approximately Normal otherwise if both samples are large enough ($n_1 \geq 30$ and $n_2 \geq 30$) by the central limit theorem.
    - **Center:** Its mean is $\mu_1 - \mu_2$.
    - **Spread:** As long as each sample is no more than 10% of its population, its standard deviation is $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$.
- Confidence intervals and tests for the difference between the means of two populations or the mean responses to two treatments μ_1 and μ_2 are based on the difference $\bar{x}_1 - \bar{x}_2$ between the sample means.
- Because we almost never know the population standard deviations in practice, we use the **two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

  This statistic has approximately a t distribution. There are two options for using a t distribution to approximate the distribution of the two-sample t statistic:
    - **Option 1 (Technology)** Use the t distribution with degrees of freedom calculated from the data by a somewhat messy formula. The degrees of freedom probably won't be a whole number.
    - **Option 2 (Conservative)** Use the t distribution with degrees of freedom equal to the *smaller* of $n_1 - 1$ and $n_2 - 1$. This method gives wider confidence intervals and larger P-values than Option 1.
- Before estimating or testing a claim about $\mu_1 - \mu_2$, check that these conditions are met:
    - **Random:** The data are produced by independent random samples of size n_1 from Population 1 and of size n_2 from Population 2 or by two groups of size n_1 and n_2 in a randomized experiment.

    - **10%:** When sampling without replacement, check that the two populations are at least 10 times as large as the corresponding samples.
  - **Normal/Large Sample:** Both population distributions (or the true distributions of responses to the two treatments) are Normal or both sample sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$). If either population (treatment) distribution has unknown shape and the corresponding sample size is less than 30, use a graph of the sample data to assess the Normality of the population (treatment) distribution. Do not use two-sample t procedures if the graph shows strong skewness or outliers.

- An approximate $C\%$ confidence interval for $\mu_1 - \mu_2$ is

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

  where t^* is the critical value with $C\%$ of its area between $-t^*$ and t^* for the t distribution with degrees of freedom from either Option 1 (technology) or Option 2 (the smaller of $n_1 - 1$ and $n_2 - 1$). This is called a **two-sample *t* interval for $\mu_1 - \mu_2$**.

- To test $H_0: \mu_1 - \mu_2 =$ hypothesized value, use a **two-sample *t* test for $\mu_1 - \mu_2$**. The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

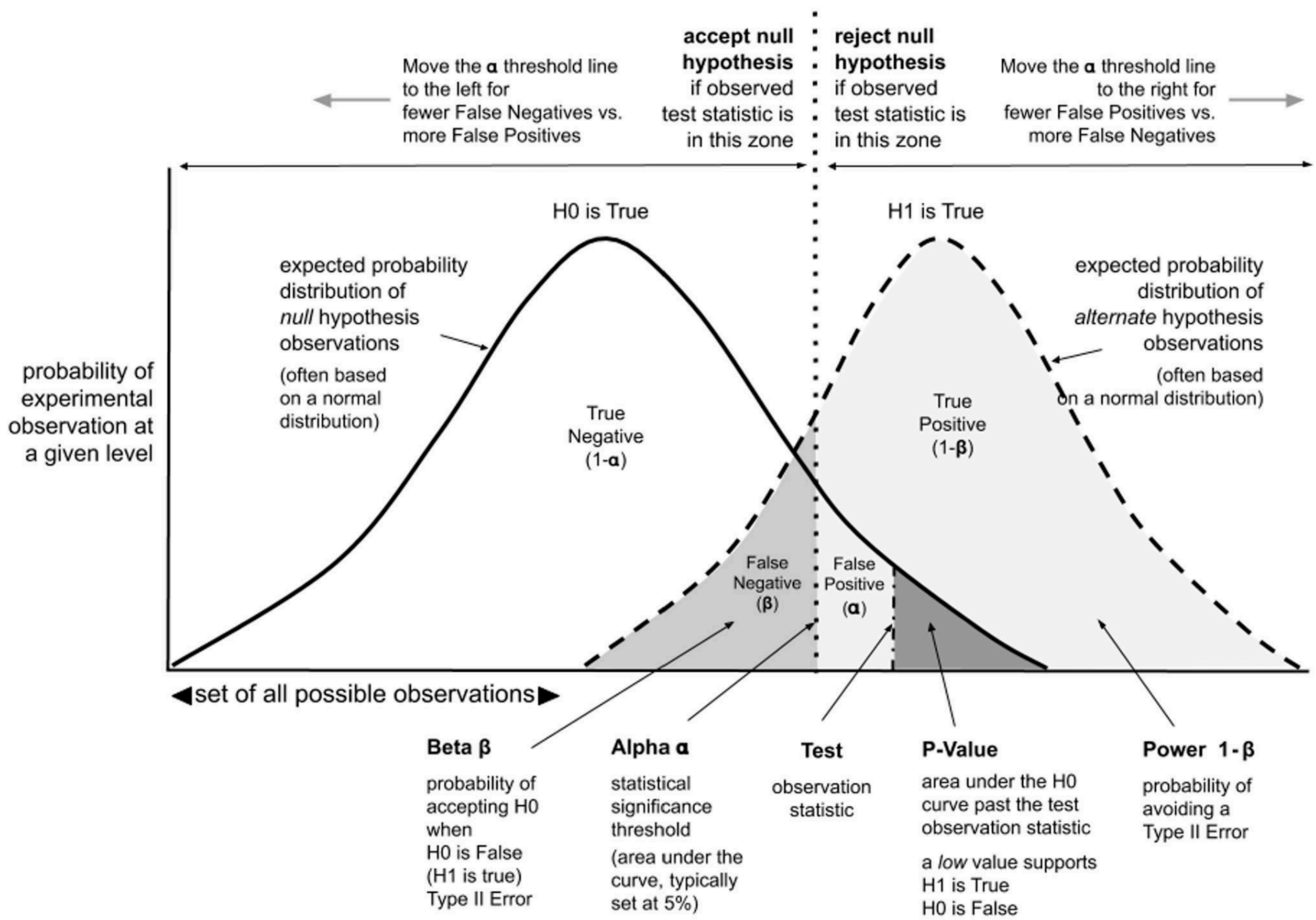
  P-values are calculated using the t distribution with degrees of freedom from either Option 1 (technology) or Option 2 (the smaller of $n_1 - 1$ and $n_2 - 1$).

- Inference about the difference $\mu_1 - \mu_2$ in the effectiveness of two treatments in a completely randomized experiment is based on the **randomization distribution** of $\bar{x}_1 - \bar{x}_2$. When the conditions are met, our usual inference procedures based on the sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be approximately correct.
- Don't use two-sample t procedures to compare means for paired data.
- Be sure to follow the four-step process whenever you construct a confidence interval or perform a significance test for comparing two means.

Note on confidence intervals:
When interpreting a confidence interval for a difference of means, proportions, etc., we need two statements:

- **Interpret:** We are ___% confident that the interval from A to B captured… etc.

- **Conclude:** Since the interval is ___________ (entirely positive, entirely negative, or includes zero), we (have or do not have) statistically significant evidence that…. etc..



How to Organize a Statistical Problem: A Four-Step Process		
	Confidence intervals (CIs)	**Significance tests**
STATE:	What *parameter* do you want to estimate, and at what *confidence level*?	What *hypotheses* do you want to test, and at what *significance level*? Define any *parameters* you use.
PLAN:	Choose the appropriate inference *method*. Check *conditions*.	Choose the appropriate inference *method*. Check *conditions*.
DO:	If the conditions are met, perform *calculations*.	If the conditions are met, perform *calculations*. • Compute the **test statistic**. • Find the ***P*-value**.
CONCLUDE:	*Interpret* your interval in the context of the problem.	Make a *decision* about the hypotheses in the context of the problem.

CI: statistic ± (critical value)·(standard deviation of statistic)

$$\text{Standardized test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

Inference about	Number of samples (groups)	Interval or test	Name of procedure (TI Calculator function) Formula	Conditions
Proportions	1	Interval	One-sample z interval for p (1-PropZInt) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	**Random** Data from a random sample or randomized experiment ○ **10%**: $n \leq 0.10N$ if sampling without replacement **Large Counts** At least 10 successes and failures; that is, $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$
		Test	One-sample z test for p (1-PropZTest) $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	**Random** Data from a random sample or randomized experiment ○ **10%**: $n \leq 0.10N$ if sampling without replacement **Large Counts** $np_0 \geq 10$ and $n(1-p_0) \geq 10$
	2	Interval	Two-sample z interval for $p_1 - p_2$ (2-PropZInt) $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	**Random** Data from independent random samples or randomized experiment ○ **10%**: $n_1 \leq 0.10N_1$ and $n_2 \leq 0.10N_2$ if sampling without replacement **Large Counts** At least 10 successes and failures in both samples/groups; that is, $n_1\hat{p}_1 \geq 10,\ n_1(1-\hat{p}_1) \geq 10,$ $n_2\hat{p}_2 \geq 10,\ n_2(1-\hat{p}_2) \geq 10$
		Test	Two-sample z test for $p_1 - p_2$ (2-PropZTest) $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$ where $\hat{p}_c = \frac{\text{total successes}}{\text{total sample size}} = \frac{X_1 + X_2}{n_1 + n_2}$	**Random** Data from independent random samples or randomized experiment ○ **10%**: $n_1 \leq 0.10N_1$ and $n_2 \leq 0.10N_2$ if sampling without replacement **Large Counts** At least 10 successes and failures in both samples/groups; that is, $n_1\hat{p}_1 \geq 10,\ n_1(1-\hat{p}_1) \geq 10,$ $n_2\hat{p}_2 \geq 10,\ n_2(1-\hat{p}_2) \geq 10$

Inference about	Number of samples (groups)	Estimate or test	Name of procedure (TI Calculator function) Formula	Conditions
Means	1 (or paired data)	Interval	One-sample t interval for μ (TInterval) $\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ with df $= n - 1$	**Random** Data from a random sample or randomized experiment ○ **10%:** $n \leq 0.10N$ if sampling without replacement **Normal/Large Sample** Population distribution Normal or large sample ($n \geq 30$); no strong skewness or outliers if $n < 30$ and population distribution has unknown shape
		Test	One-sample t test for μ (T-Test) $t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$ with df $= n - 1$	**Random** Data from a random sample or randomized experiment ○ **10%:** $n \leq 0.10N$ if sampling without replacement **Normal/Large Sample** Population distribution Normal or large sample ($n \geq 30$); no strong skewness or outliers if $n < 30$ and population distribution has unknown shape
	2	Interval	Two-sample t interval for $\mu_1 - \mu_2$ (2-SampTInt) $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ df from technology or $\min(n_1 - 1, n_2 - 1)$	**Random** Data from independent random samples or randomized experiment ○ **10%:** $n_1 \leq 0.10N_1$ and $n_2 \leq 0.10N_2$ if sampling without replacement **Normal/Large Sample** Population distributions Normal or large samples ($n_1 \geq 30$ and $n_2 \geq 30$); no strong skewness or outliers if sample size < 30 and population distribution has unknown shape
		Test	Two-sample t test for $\mu_1 - \mu_2$ (2-SampTTest) $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ df from technology or $\min(n_1 - 1, n_2 - 1)$	**Random** Data from independent random samples or randomized experiment ○ **10%:** $n_1 \leq 0.10N_1$ and $n_2 \leq 0.10N_2$ if sampling without replacement **Normal/Large Sample** Population distributions Normal or large samples ($n_1 \geq 30$ and $n_2 \geq 30$); no strong skewness or outliers if sample size < 30 and population distribution has unknown shape