Session 8

Comparisons Between Two Groups

Managerial Decisions (revisited)

Are female managers paid less in our company than male managers?

Are sales persons in one zone more productive than others?

Did average household income increase after liberalization in 1991?

Does one of our manufacturing plants have better quality than the other?

Does Angioplasty yield better outcomes than bypass surgery?

Did the child welfare scheme increase the number of school going children?

Learning Objectives

- How to compare means of two populations using paired observations?
- When and how to compare two populations means using independent samples?
- How to test for differences in two population proportions?

Example: Weight reduction programs

- A nutrition expert would like to assess the effect of organized diet programs on the weight of the participants.

- She randomly chooses 36 participants of the Atkins diet program and measures their weight (in kg) just before enrolling in the program and immediately after the completion of the program.

- Based on this evidence, is the Atkins diet program effective in reducing weight?

Before	After	Before	After
130	123	130	127
123	124	139	132
132	134	120	110
150	152	138	140
146	143	141	136
153	143	120	118
137	133	153	154
140	137	126	125
148	152	148	143
158	149	141	135
144	132	137	135
160	155	159	152
146	142	152	148
146	142	140	138
153	155	140	134
137	130	151	147
138	130	141	144
125	124	139	128

ISB

Transforming into a single variable

- Let W be the change in weight of a randomly chosen participant after the diet program
  - Mean: μ_W
  - Standard Deviation: σ_W

- "The diet program is effective" ←→ "Average change in weight is negative"

- H_0: $\mu_W \geq 0$ and H_A: $\mu_W < 0$

- It is natural and also feasible to take before and after measurements on the same subjects → Paired test

Example: A Health Chain

- A health chain can recommend conventional low calorie diet for free or can recommend Atkins diet by paying $200,000 *licensing fee.*

- The firm has determined that it is worth paying the licensing fee if they can gain enough additional members, which is possible if Atkins diet reduces average weight by 2 pounds or more compared to the conventional low calorie diet.

- The firm collects weight loss data from two simple random samples of people, one of whom goes through Atkins diet and the other through the conventional diet for 6 months.

	Number	Mean	Std Dev	Skewness	Kurtosis
Atkins	33	15.42	14.37	-0.052	0.100
Conventional	30	7.00	12.36	0.342	-0.565

Hypotheses: Difference Between Means

- We wish to test hypotheses of the following form (where μ_1 and μ_2 are the means of the two populations and D_0 is the least acceptable difference)

$$H_0 : \mu_1 - \mu_2 \leq D_0$$
$$H_A : \mu_1 - \mu_2 > D_0$$

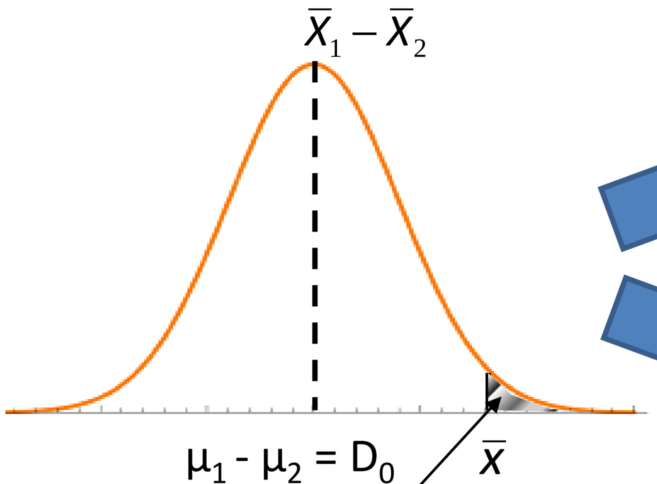
- Similar to earlier approach, we will use $\overline{X}_1 - \overline{X}_2$ to make statements about $\mu_1 - \mu_2$

Sampling Distributions of Means of the Two Samples

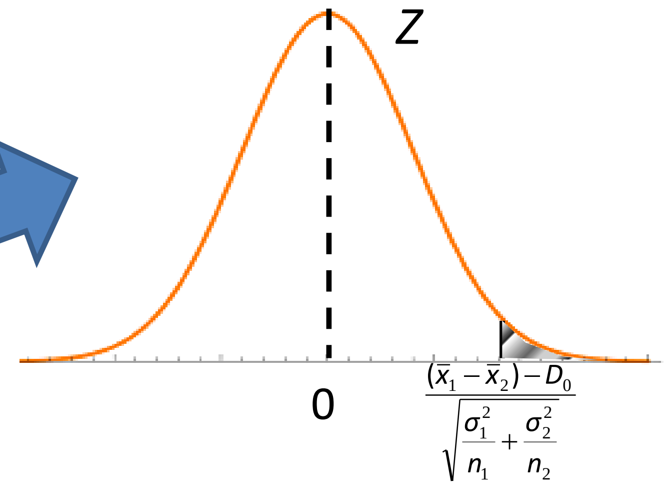
- The two sampling distributions of means are normal provided CLT condition is met separately for each
- Independence
  - Who is in a sample does not influence who else is in that sample
- Size conditions
  - Number of observation in each sample must exceed 10 times the absolute value of Kurtosis and 10 times the squared skewness within that sample

Calculating the probability of type-I error

H_0: $\mu_1 - \mu_2 \leq D_0$



σ known



σ unknown

Two cases depending on
$\sigma_1 = \sigma_2$ or $\sigma_1 \neq \sigma_2$

Probability that I see a sample of $\bar{x}$ or greater when the null hypothesis is true (p-value)

With population standard deviations unknown

If we believe $\sigma_1 = \sigma_2$	If we believe $\sigma_1 \neq \sigma_2$

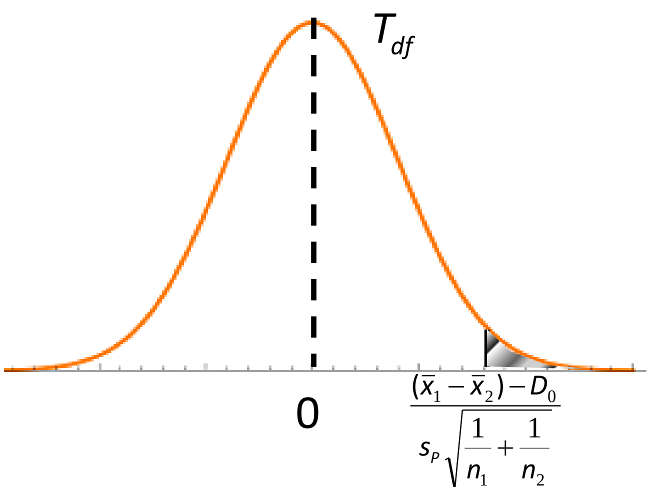
If we believe $\sigma_1 = \sigma_2$

- Calculate the "pooled" sample standard deviation

$$s_P = \sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2}}$$

and degrees of freedom

$$df = n_1 + n_2 - 2$$



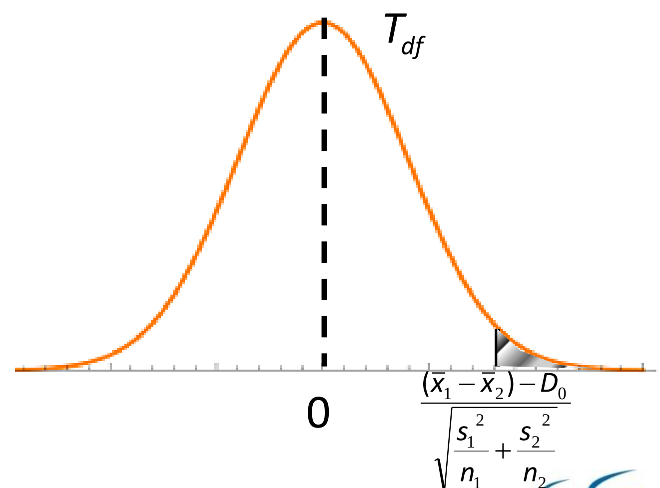
If we believe $\sigma_1 \neq \sigma_2$

- Calculate the standard error

$$se(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and degrees of freedom

$$df = \left\lfloor \frac{\left(s_1^{\,2}/n_1 + s_2^{\,2}/n_2\right)^2}{\left(s_1^{\,2}/n_1\right)^2/(n_1 - 1) + \left(s_2^{\,2}/n_2\right)^2/(n_2 - 1)} \right\rfloor$$



ISB

Example: Health Chain

- Hypotheses
  - $H_o: \mu_A - \mu_C \leq 2$
  - $H_A: \mu_A - \mu_C > 2$

- Recall

	Number	Mean	Std Dev	Skewness	Kurtosis
Atkins	33	15.42	14.37	-0.052	0.100
Conventional	30	7.00	12.36	0.342	-0.565

$T^* = \dfrac{(15.42 - 7.00) - 2}{3.369} \approx 1.91$ and df = 60.82 → P-value = 0.0308 < 0.05.

Can We Attribute the Difference to Diets?

- Could there be other systematic differences between the two groups?
  - Atkins is adopted by younger individuals, who are more motivated to lose weight
  - Atkins is adopted by individuals from higher socio-economic strata who have easier access to healthy food options

- These factors can be "controlled" for using additional variables in a regression model

- An alternative way is to randomly assign individuals to one or the other diet program and then compare the difference → Field Experiment

Example: Proportion of dieters who lose weight

- Suppose an alternate metric to measure the performance of the diet program is proportion of participants who have lost more than 10 pounds

	Number	Successful	Proportion
Atkins	33	20	0.606
Conventional	30	15	0.50

- Hypotheses:
  - H_o: $\pi_A - \pi_c = 0$
  - H_A: $\pi_A - \pi_c \neq 0$

- Similar to previous calculation, we can calculate $z = \dfrac{(p_A - p_C)}{\sqrt{\bar{p}(1-\bar{p})(1/n_A + 1/n_C)}}$ and proceed with hypothesis testing accordingly

- Here $\bar{p}$ is the pooled sample proportion given by $\dfrac{p_A n_A + p_C n_C}{n_A + n_C}$

Summary of Session VIII

- The best way to compare the means of two distributions is using paired observations, if it is feasible
- The average difference of paired sample observations follows normal distribution according to Central Limit Theorem
- When paired observations are not possible, we use independent samples and formulate hypothesis on the difference of two means
- It is important to ensure that subjects are randomly assigned to the two samples to avoid any confounding errors
- Similar approach can be used to test the difference in proportions between two populations

ISB