A Review

So far, we have

- used the probability distribution of a random variable Y to characterize the population.
- viewed the sample observations (i.e., data) as outcomes of random variables from this population.
- considered an i.i.d. sample $Y_1, Y_2, \ldots, Y_n$ of sample size n, where Y_i denotes the random variable for the ith observation in the sample.
- studied the distribution of the sample mean $\bar{Y}$ and the sample variance S^2.
- learned that $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ assuming $Y_i \sim_{\text{i.i.d}} N(\mu, \sigma^2)$.
- to learn that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ assuming $Y_i \sim_{\text{i.i.d}} N(\mu, \sigma^2)$

We will next turn to *statistical inference*.

Sample Variance

- Lake clarity example: As the sample size n increases, the sample variance gets closer to the population variance.
- Again, consider $Y \sim N(\mu, \sigma^2)$.
- Let $Y_1, Y_2, \ldots, Y_n$ denote an i.i.d. sample from this population $N(\mu, \sigma^2)$.
- The sample variance is defined as

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

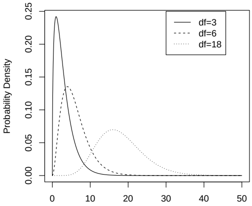
- Note that S^2 is also a random variable.

Standardization of Sample Variance

- Suppose $Y_1, Y_2, \ldots, Y_n$ is an i.i.d. sample from a normal distribution with mean μ and σ^2.
- Let S^2 be the sample variance.
- Define
$$V^2 = \frac{(n-1)S^2}{\sigma^2}$$
- V^2 follows a chi-squared distribution with $n-1$ degrees of freedom.
- $\bar{Y}$ and S^2 are independent.

Chi-Squared Distribution

- Next goal: Make probability statements about S^2.
- A **chi-squared distribution with** m **degrees of freedom** is a model for a random variable denote by $V^2 \sim \chi^2_m$.
- Properties of $V^2 \sim \chi^2_m$:
  - The range of possible values of V^2 is from 0 to $+\infty$.
  - The distribution is right-skewed.
  - $E(V^2) = m$ and $Var(V^2) = 2m$.



- If $Z_1, \ldots, Z_m$ are independent $N(0, 1)$, then $\sum_{j=1}^{m} Z_j^2 \sim \chi^2_m$.

Properties of Sample Variance

- The expected value of the sample variance is

$$E\left(S^2\right) = \sigma^2$$

- Interpretation: Unbiased estimator
- The variance of the sample variance is

$$Var(S^2) = \frac{2\sigma^4}{n-1}.$$

- Interpretation: As n increases, variance decreases and precision to estimate σ^2 increases.

Outline

Two-Sample Studies

Two-sample studies for comparing two populations are more common than one-sample studies. For example, the goals are to:

A. Compare milk yield of cows on two different diets.

B. Compare timber volumes of two species of trees.

C. Compare heart rates of patients before and after a drug treatment.

D. Compare test scores of 7th graders before and after the summer break.

Paired vs. Independent Two Samples

- There are two types of two-sample studies:
  - Two samples are **paired**.
  - Two samples are **independent** or **unpaired**.
- A **paired two-sample study (or experiment)** is a study (or experiment) with two levels of a factor (or treatment) where each observation on one level of the factor (or treatment) is naturally paired with an observation on the other level of the factor (or treatment).
- An **independent two-sample study (or experiment)** is a study (or experiment) with two levels of a factor (or treatment) where there is no relationship between the observations on the two levels of factor (or treatment).

Paired vs. Independent Two Samples

- The choice of a paired-sample study versus an independent two-sample study is an important design issue.
- When to use which study?
- For example, consider
    - Heart rates of 10 patients before and after a drug treatment.
    - Heart rates of 10 patients before the drug treatment and heart rates of another 10 patients after the drug treatment.
- Which study would be better for detecting the drug effect?

- The method we use for data analysis should follow the study design.

Outline

- Sample Variance

1. Statistical Inference and Study Design
   - Examples
   - Paired vs. Independent Two Samples

2. Comparison of Two Population Means: Paired T Inference
   - Hypothesis Testing
   - Confidence Intervals
   - Power Analysis

Example: Lake Clarity 1980 vs. 1990

	Wisconsin	
Lake	1980	1990
1	2.11	3.67
2	1.79	1.72
3	2.71	3.46
4	1.89	2.60
5	1.69	2.03
6	1.71	2.10
7	2.01	3.01
8	1.36	1.82
9	2.08	2.64
10	1.10	2.23
11	1.29	1.39
12	2.11	2.08
13	2.47	2.92
14	1.67	1.90
15	1.78	2.44
16	1.68	2.23
17	1.47	2.43
18	1.67	1.91
19	2.31	3.06
20	1.76	2.26
21	1.58	1.48
22	2.55	2.35
sample mean	1.854	2.351
sample variance	0.168	0.354
sample sd	0.410	0.595

Null Hypothesis vs. Alternative Hypothesis

- Y_{1i}: Random variable of Secchi depth of the ith lake in 1990 for $i = 1, \ldots, n$.
- Y_{2i}: Random variable of Secchi depth of the ith lake in 1980 for $i = 1, \ldots, n$.
- $\mu_1 = E(Y_{1i})$: Population mean Secchi depth in 1990.
- $\mu_2 = E(Y_{2i})$: Population mean Secchi depth in 1980.
- Our goal is to test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.
- Under the **null hypothesis** $H_0 : \mu_1 = \mu_2$
- The null hypothesis H_0 is generally the claim initially favored or believed to be true.

- Under the **alternative hypothesis** $H_A : \mu_1 \neq \mu_2$
- The alternative hypothesis H_A is generally the departure from H_0 that one wishes to be able to detect.

Example: Lake Clarity 1980 vs. 1990

	Wisconsin		
Lake	1980	1990	DIFF
1	2.11	3.67	1.56
2	1.79	1.72	-0.07
3	2.71	3.46	0.75
4	1.89	2.60	0.71
5	1.69	2.03	0.34
6	1.71	2.10	0.39
7	2.01	3.01	1.00
8	1.36	1.82	0.46
9	2.08	2.64	0.56
10	1.10	2.23	1.13
11	1.29	1.39	0.10
12	2.11	2.08	-0.03
13	2.47	2.92	0.45
14	1.67	1.90	0.23
15	1.78	2.44	0.66
16	1.68	2.23	0.55
17	1.47	2.43	0.96
18	1.67	1.91	0.24
19	2.31	3.06	0.75
20	1.76	2.26	0.50
21	1.58	1.48	-0.10
22	2.55	2.35	-0.20
sample mean	1.854	2.351	0.497
sample variance	0.168	0.354	0.190
sample sd	0.410	0.595	0.435

Null Hypothesis vs. Alternative Hypothesis

- $D_i = Y_{1i} - Y_{2i}$: Secchi depth difference of the ith lake between 1990 and 1980.
- $\mu_D = E(D_i) = \mu_1 - \mu_2$: Population mean Secchi depth difference between 1990 and 1980.
- Equivalent to testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$, we now consider testing

$$H_0 : \mu_D = 0 \text{ vs. } H_A : \mu_D \neq 0.$$

- A **statistic** is the sample mean Secchi depth difference $\bar{D}$ based on an i.i.d. sample of size $n = 22$ $(D_1, D_2, \ldots, D_{22})$.

Test Statistic

- Assume that the $H_0 : \mu_D = 0$ holds.
- Assume that $D_i \sim_{\text{i.i.d.}} N(0, \sigma_D^2)$.
- What is the distribution of $\bar{D}$?

$$\bar{D} \sim N\left(0, \frac{\sigma_D^2}{n}\right).$$

Why not $\bar{D} \sim N(0, 0.190/22)$?

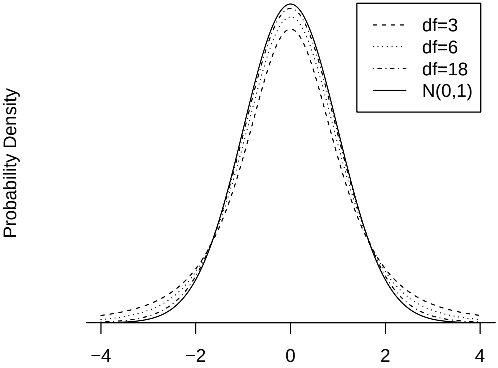
- The **test statistic** is a function of the data that is useful in testing and leads to a value that can be directly interpreted by using an appropriate statistical table.

$$T = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}},$$

T Distribution

- Under the $H_0 : \mu_D = \mu_D^0$,
$$T = \frac{\bar{D} - \mu_D^0}{\frac{S_D}{\sqrt{n}}} \sim T_{n-1}$$
where T_{n-1} is a T **distribution** with $n-1$ degrees of freedom (df).
- The df are the same as the df of the sample variance S_D^2.
- The value
$$t = \frac{\bar{d} - \mu_D^0}{\frac{s_D}{\sqrt{n}}}.$$
is referred to as the ***t*-score**.
- The denominator $s_D/\sqrt{n}$ is the **standard error** of the sample mean $\bar{d}$.

T Distribution



T Distribution

- The possible range of values of T is from $-\infty$ to ∞.
- The distribution is symmetric and centered around 0.
- The distribution of T is similar to that of Z; probabilities are interpreted as areas under a probability density curve.
- However, the T-distribution has less concentration of probability close to 0 and more probability in the tails than Z.
- Also, the distribution of T depends on the sample size n.
- For larger n, the distribution curve has thinner tails, less spread, and more Z-like.

Table B

- Table B gives the critical values a for a set of upper tail probabilities $P(T \geq a)$ for the T distribution with degrees of freedom.
- The values a are inside of Table B and the corresponding probabilities are at the top of Table B (column index).
- For example,

df	...	0.10	0.05	0.025	...
1					
2					
...					
9	...		1.833		...
...					

Table C

- Compute the upper tail probabilities as

$$P(T_9 \geq 1.833) = 0.05$$

- Or, find the the 0.95th quantile a such that

$$P(T_9 \geq a) = 0.05.$$

Thus, $a = 1.833$ or $q_{0.95} = 1.833$.

Example: Lake Clarity 1980 vs. 1990

- From the summary statistics, we have $n = 22$, $\bar{d} = 0.497$, and $s_D = 0.435$.
- The standard error is:
  $s_d/\sqrt{n} = 0.435/\sqrt{22} = 0.0927$
- The observed test statistic is:
  $t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{0.497-0}{0.0927} = 5.357$
- Compute a **p-value** defined as the probability of observing a value as extreme or more extreme than what we observed, *if the H_0 is true*.
  $2 \times P(T_{21} \geq 5.357)$ which is less than 0.002 from Table B.

Interpretation of the p-value

- The p-value can be interpreted as evidence again H_0. The smaller the p-value, the greater the evidence.
- In the classical hypothesis testing, a threshold value α is determined and the p-value is compared against it.
  - If the p-value is less than α, then we **reject** the H_0.
  - If the p-value is greater than α, then we **do not reject** the H_0.
- Lake clarity 1980 vs. 1990 example: Reject H_0 at the 5% level. There is very strong evidence that the mean Secchi depths in 1980 and 1990 are different.

Example: Lake Clarity 1980 vs. 1990

- How about testing

$$H_0 : \mu_1 = \mu_2 + 0.5 \text{ vs. } H_A : \mu_1 > \mu_2 + 0.5$$

- The test statistic is:

$$T = \frac{\bar{D} - 0.5}{S_D/\sqrt{n}} \sim T_{n-1}$$

- The standard error is:
  $s_d/\sqrt{n} = 0.435/\sqrt{22} = 0.0927$

- The observed test statistic is:
  $t = \frac{\bar{d}-0.5}{s_d/\sqrt{n}} = \frac{0.497-0.5}{0.435/\sqrt{22}} = -0.0294$

Point Estimators

- Most probability distributions are indexed by one or more **parameters**. For example, $N(\mu, \sigma^2)$.
- In hypothesis tests, we have used **point estimators** for parameters.
  For example, consider an i.i.d. sample $D_1, D_2, \ldots, D_n \sim_{\text{i.i.d}} N(\mu_D, \sigma_D^2)$. Let

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_n, \quad S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_n - \bar{D})^2.$$

  Then $\bar{D}$ is a point estimator of μ_D and S_D^2 is a point estimator of σ^2.
- We know that $E(\bar{D}) = \mu_D$ and $E(S_D^2) = \sigma_D^2$.
- That is, $\bar{D}$ is an **unbiased estimator** of μ_D and S_D^2 is an unbiased estimator of σ_D^2.

Interval Estimators

- Now we turn to **interval estimators** to give a reasonable interval for parameters.
  - For μ: $[a_1, a_2]$ for some constants a_1, a_2 based on data
  - For σ^2: $[b_1, b_2]$ for some constants b_1, b_2 based on data
- The assumptions are the same as in hypothesis testing, but we do not need a null hypothesis about the parameters (e.g. $\mu_D = \mu_1 - \mu_2 = 0$).

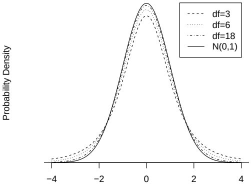
Confidence Interval for μ_D

- Suppose $D_1, D_2, \ldots, D_n$ is an i.i.d. sample from $N(\mu_D, \sigma_D^2)$ and σ_D^2 is unknown.
- Note that

$$\frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim T_{n-1}.$$

- Let $t_{n-1,\alpha/2}$ denote the t critical value such that

$$P(-t_{n-1,\alpha/2} \leq T_{n-1} \leq t_{n-1,\alpha/2}) = 1 - \alpha.$$



Confidence Interval for $\mu_D = \mu_1 - \mu_2$

- Then we have

$$1 - \alpha = P\left(\mu_D \in \left[\bar{D} - t_{n-1,\alpha/2}\frac{S_D}{\sqrt{n}},\ \bar{D} + t_{n-1,\alpha/2}\frac{S_D}{\sqrt{n}}\right]\right)$$

- A $(1-\alpha)$ CI for μ_D is

$$\mu_D \in \left[\bar{d} - t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}},\ \bar{d} + t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}\right]$$

  or

$$\bar{d} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}$$

- The half width of this CI is $t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}$.
- The width of this CI is $2 \times t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}$.

Confidence Intervals for $\mu_D = \mu_1 - \mu_2$

- A $(1-\alpha)$ CI for μ_D is

$$\mu_D \in \left[\bar{d} - t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}},\ \bar{d} + t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}\right]$$

- In the lake clarity 1980 vs. 1990 example, a 95% CI for μ_D is

$$0.497 - 2.080 \times \frac{0.435}{\sqrt{22}} \leq \mu_D \leq 0.497 + 2.080 \times \frac{0.435}{\sqrt{22}}$$

which is [0.30, 0.69] or 0.497 ± 0.195.

Remarks

- By convention, CIs are two-sided. But one-sided confidence bounds are possible.
- It is not true that $P(0.30 \leq \mu_D \leq 0.69) = 0.95$. Why not?

  because once a sample is observed, there is nothing random.
- The 0.95 probability concerns with the repeated random sampling. It is interpreted as, 95% of the time, the (random) CIs calculated in this way contains (fixed) μ_D.
- For a single case, it is interpreted as ? having 95% confidence that μ_D is between 0.30 m and 0.69 m.

- The interval $[0.30, 0.69]$ (or 0.497 ± 0.195) can be thought of as a plausible range of μ_D.
- What are the assumptions made when we perform a paired T test or construct a corresponding confidence

Ingredients of Hypothesis Testing

- i.i.d. sample: $Y_1, Y_2, \ldots, Y_n \sim N(\mu, \sigma^2)$
- Parameter of interest: μ or σ^2
- H_0 versus H_A
- Testing $H_0 : \mu = \mu_0$
  - Normal population σ^2 known:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

  - Normal population σ^2 unknown:

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim T_{n-1}, \quad \text{where } S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2},$$

- Testing for σ^2: $H_0 : \sigma^2 = \sigma_0^2$

$$V^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

- p-value and conclusion

Example: $H_0 : \mu = 60$ vs. $H_A : \mu \neq 60$

- Consider testing
$$H_0 : \mu = 60 \text{ vs. } H_A : \mu \neq 60.$$
- Suppose that an i.i.d. sample of size $n = 16$ is possible and that the population distribution is $N(\mu, 36)$.
- Under H_0, what is the distribution of the sample mean $\bar{Y}$?
$\bar{Y} \sim N(\mu_0 = 60, \frac{\sigma^2}{n} = \frac{36}{16} = (1.5)^2)$
- Suppose $\alpha = 0.05$. That is, we reject H_0 if the p-value is less than or equal to 0.05.
- What values of $\bar{Y}$ would lead to a rejection of H_0?
- Can we answer this question before even having the data?

Rejection Region

- Note $z_{0.025} = 1.96$ (from Table A). We would reject H_0 if the observed $\bar{y}$ is more than 1.96 standard deviation (of the sample mean!) away from the hypothesized mean $\mu = 60$.

- We would reject H_0 if the observed $\bar{y}$ is less than 57.06 or more than 62.94. [Why?]

$$
\begin{aligned}
0.05 &= P(Z \leq -1.96) + P(Z \geq 1.96) \\
&= P\left(\frac{\bar{Y} - 60}{1.5} \leq -1.96\right) + P\left(\frac{\bar{Y} - 60}{1.5} \geq 1.96\right) \\
&= P(\bar{Y} \leq 60 - 1.96 \times 1.5) + P(\bar{Y} \geq 60 + 1.96 \times 1.5) \\
&= P(\bar{Y} \leq 57.06) + P(\bar{Y} \geq 62.94)
\end{aligned}
$$

Type I and II Error

- For a given rejection region, the rule is to reject H_0 if the observed $\bar{y}$ falls in the rejection region and do not reject H_0 otherwise.
- Because $\bar{Y}$ is random, it is possible to make two types of errors.
  - A **Type I error** occurs if we reject H_0 when H_0 is true.
  - A **Type II error** occurs if we accept H_0 when H_0 is false.
- In our example, what is the probability of Type I error?

$$P(\text{Reject } H_0 | H_0) = P(\bar{Y} \leq 57.06 \text{ or } \bar{Y} \geq 62.94 | \mu = 60)$$

Power

- For a given rejection rule and for any given value of μ, the **power** is the probability of rejecting H_0 given the value of μ.
- In the example above, what is the power for $\mu = 60$?

$$P(\text{Reject } H_0|\mu = 60) = P(\bar{Y} \leq 57.06 \text{ or } \bar{Y} \geq 62.94|\mu = 60)$$

- What is the power for $\mu = 62$?

$$P(\text{Reject } H_0|\mu = 62) = P(\bar{Y} \leq 57.06 \text{ or } \bar{Y} \geq 62.94|\mu = 62)$$

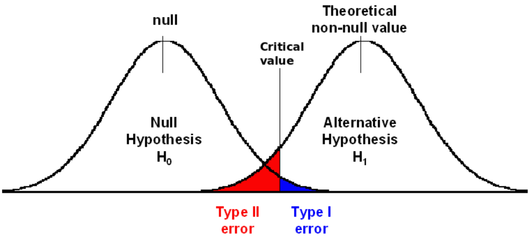
- What is the power for $\mu = 64$?

$$P(\text{Reject } H_0|\mu = 64) = P(\bar{Y} \leq 57.06 \text{ or } \bar{Y} \geq 62.94|\mu = 64)$$

Types of Error and Statistical Power

There are four possible outcomes that could be reached as a result of the null hypothesis being either true or false and the decision being either “fail to reject" or “reject”.

Reality	Our Decision	
	H_0	H_a
H_0	$\surd$	Type I Error
H_a	Type II Error	$\surd$



Types of Error and Statistical Power

Reality	Our Decision H_0	Our Decision H_a
H_0	$\surd$ (Prob $= 1 - \alpha$)	Type I Error (Prob $= \alpha$)
H_a	Type II Error (Prob $= \beta$)	$\surd$ (Prob $= 1 - \beta$)

- The **significance level** α of any fixed level test is the probability of a Type I error.
- The **power** of a fixed level test against a particular alternative is $1 - \beta$ for that alternative.
- In practice, we first choose an α and consider only tests with probability of Type I error no greater than α. Then we select one that makes the probability of Type II error as small as possible (i.e. the most powerful possible test).