Midterm II Review

PH142 Fall 2025

Logistics

- **Date:** Friday, October 31st
- **Time:** 8:10–9:00AM, arrive no later than 8:00AM
- **Location(s):** Wheeler, Stanley, Dwinelle
    - Same room assignments as Midterm 1
- **Material Covered:** Lectures 11–22, Lab 4–7

What to Bring

- **Student ID**
- **Pencil/Pen**
- **Cheat Sheet** (single sided, handwritten, 8.5x11”)
- **Scientific Calculator** (non-graphing)

Probability

Probability

Overview

Probability is the language we use to describe uncertainty. It quantifies the chance of a specific event happening.

Probability Model: An action or process with an uncertain outcome (e.g., flipping a coin, rolling a die, taking a random sample). The model includes:

- **Sample Space (S):** The set of *all possible outcomes* of the experiment. For a coin flip, S = {Heads, Tails}. For a six-sided die, S = {1, 2, 3, 4, 5, 6}.
- **Event:** A specific outcome or a collection of outcomes that we are interested in (e.g., getting heads, rolling an even number).

Probability

Probability Rules

1. **Range of Probability:** Probabilities are numbers between 0 and 1.
2. **Sample Space:** All possible outcomes of a sample space (S) together have a probability of 1.
3. **Disjoint Events:** If two events have a joint probability of 0 (i.e., no overlap in their event spaces) then they are disjoint. If A and B are disjoint: $P(A \text{ AND } B) = 0$.
4. **Complement Rule:** The probability of an event not occurring (the “complement”) is 1 minus the probability of the event occurring.

Probability

Independence of Events

Independent Events

- A and B are independent if knowing whether A occurs <u>does not</u> change the probability of B

Dependent Events

- A and B are dependent if knowing whether A occurs <u>does</u> change the probability of B

Probability

Calculating Probabilities

- **Conditional Probability:** when $P(A) > 0$, the conditional probability of B given A is: $P(B|A) = P(A \cap B) / P(A)$
- **General addition rule:** for any two events A and B: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - When A and B are disjoint, $P(A \cup B) = P(A) + P(B)$
- **General multiplication rule:** $P(A \cap B) = P(A)P(B|A)$
    - When A and B are independent, $P(A \cap B) = P(A) \times P(B)$
- **Bayes' Theorem:** If independent, $P(A|B) = P(A)$. But if they aren't:

  $$P(A|B) = P(B|A)P(A) / [P(B|A)P(A) + P(B|A^c)P(A^c)]$$

Probability

Probability Tip

Look for patterns in language! Some words to look for are “given”, “and”, “or”, “both”, and “neither”.

“Given”	Conditional Probability
“And”	Intersection (A and B)
“Both”	Intersection (A and B)
“Or”	Union (A or B)
“Neither”	Complement

Probability

Screening Tests

Sensitivity: Probability the test is positive given disease present

- Sensitivity = **P(Test+ | Disease+)**

Specificity: Probability the test is negative given disease absent

- Specificity = **P(Test- | Disease-)**

Sensitivity and specificity are properties of the test. They do not change based on prevalence of disease in the population.

Probability

Predictive Values

Positive Predictive Value (PPV): Probability the disease is present given a positive test result

- PPV = **P(Disease+ | Test+)**

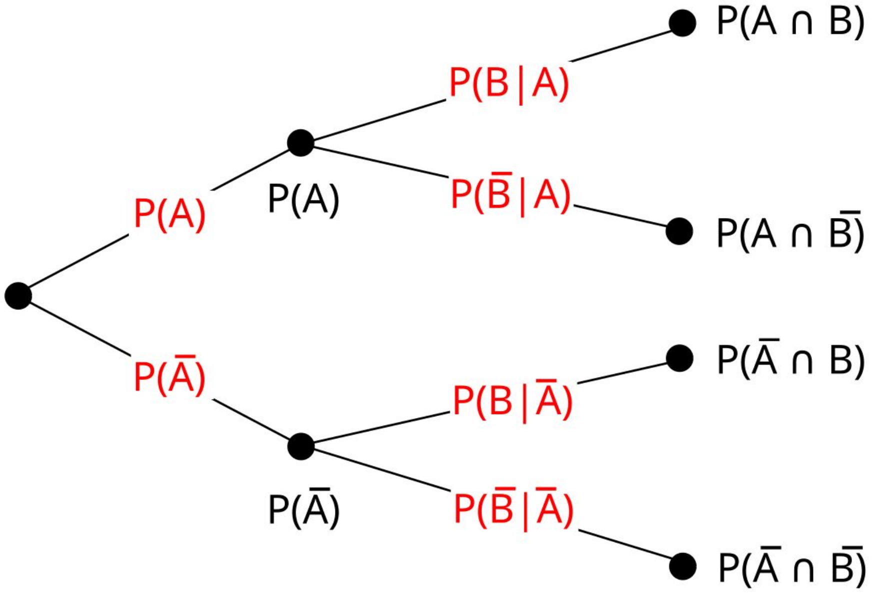
Negative Predictive Value (NPV): Probability the disease is absent given a negative test result

- NPV = **P(Disease- | Test-)**

PPV and NPV change with the prevalence of disease in the population.

Probability

Tree Diagrams



The Normal Distribution

The Normal Distribution

Overview

The Normal distribution is a bell-shaped, symmetric distribution, with notation $\mathbf{X \sim N(\mu, \sigma)}$

- μ = mean
- σ = standard deviation

The normal distribution is continuous. This means the probability of a single exact value is equal to zero.

- $P(X=x) = 0$

The Normal Distribution

Standardization

With normally distributed data, we can use standardization to transform our distribution using the **z-score**.

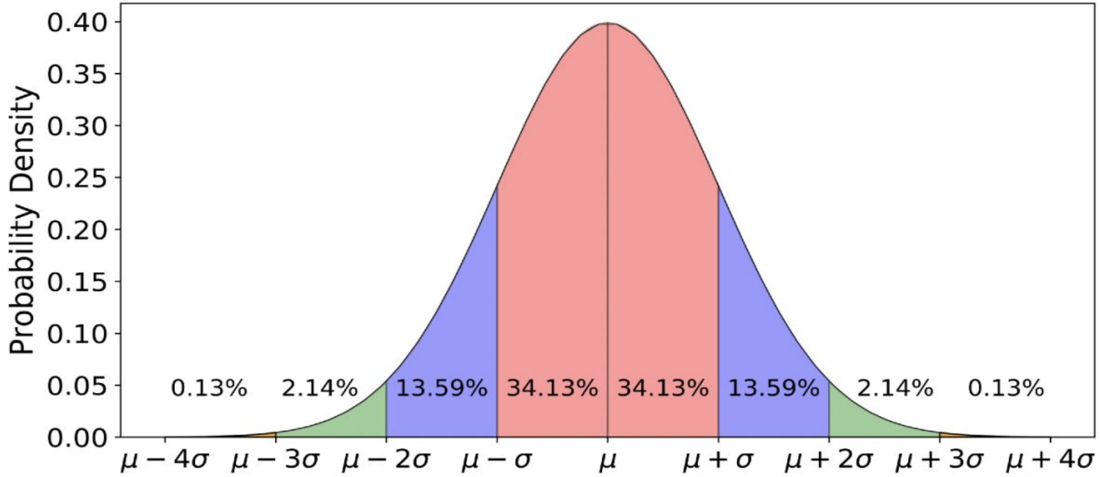
Z-score formula: $$z = \frac{x - \mu}{\sigma}$$

x = observation from a distribution
μ = mean
σ = standard deviation

Interpretation: *z* is the number of standard deviations that *x* is above or below the mean of the data.

The Normal Distribution

Example



≈ 68% of data within one SD of the mean.

≈ 95% of data within two SDs of the mean

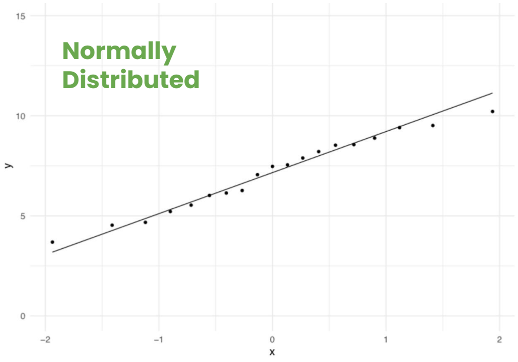
≈ 99.7% of data within three SDs of the mean

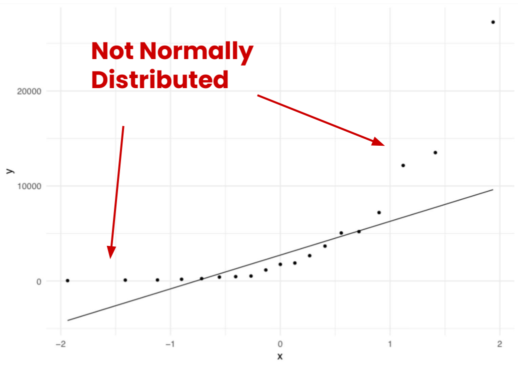
The Normal Distribution

Q-Q Plots

We use **Q-Q Plots** to examine whether a continuous variable follows a Normal distribution.

If the data are normally distributed, the points will fall roughly along the straight line.





The Normal Distribution

Functions

Function	Use	Example
pnorm()	Compute the probability at a specified x or below. Is a cumulative probability. Need to also specify the mean and sd. Remember P(X=2) = 0 . This holds for any value "x".	P(X<=2): pnorm(2, mean = 3, sd = 0.5) P(X<2): pnorm(2, mean = 3, sd = 0.5)
rnorm()	Generate a random variable that is drawn from a normal distribution. We use this to create/simulate data to get a sense of what Normally distributed data looks like	rnorm(10, mean = 3, sd = 0.5) generates 10 random draws from a Normal distribution
qnorm()	Is the opposite of pnorm(). You provide a percentage and it returns the x value such that x percent of the area is at/below that value in the lower tail.	P(X<=x) = 75%, what is x? qnorm(0.75, mean = 3, sd = 0.5) will return the value x.

The Binomial Distribution

The Binomial Distribution

Overview

The Binomial distribution models discrete random variables (only whole numbers), with notation **X ~ Binom(*n*, *p*)**

- *n* = total number of "trials"
- *p* = probability of success

Binomial mean and standard deviation:

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

The Binomial Distribution

Assumptions

1. The number of trials is fixed
2. The trials are independent
3. There are only two outcomes: success or failure
4. The probability of success is the same for each trial

The Binomial Distribution

Normal Approximation

When the data are Binomially distributed with a large sample size *n*, we can use the Normal distribution to perform calculations.

General rule: When **n*p >= 10** and **n*(1-p) >= 10**

The Binomial Distribution

Probability Function

If X has a binomial distribution, we can use the **probability distribution function** to compute (by hand) the probability that X = x for every x between 0 and n.

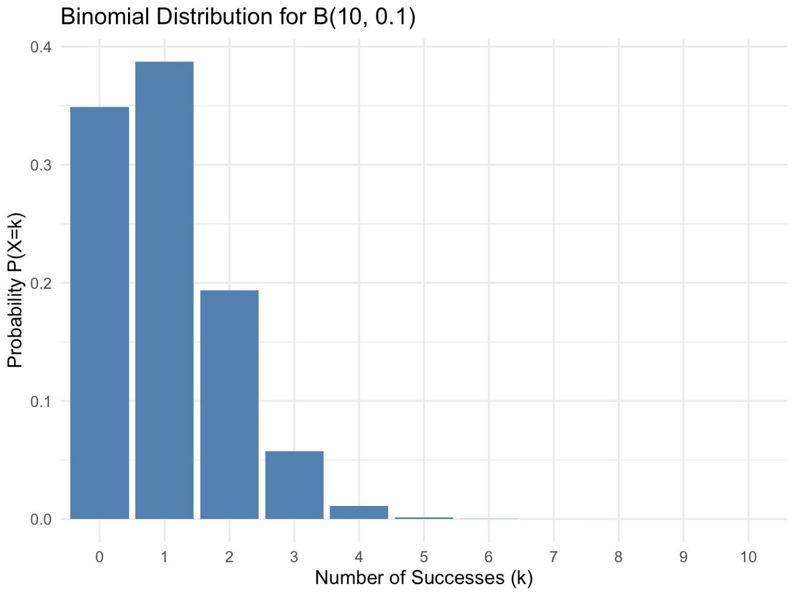
$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

The chance of observing x successes and n-x failures for one specific ordering of successes and failures.

# of ways x successes can be arranged in n observations

The Binomial Distribution

Example



The Binomial Distribution

Functions

Function	Use	Example
pbinom()	To calculate cumulative probabilities	To compute P(X<=2), where X follows a Binomial(10, 0.5) we use pbinom(2, 10, 0.5) For X <2 use pbinom(1, 10, 0.5)
dbinom()	To calculate the probability mass function	To compute the mass of a Binomial(10, 0.5) at 2, we use dbinom(2, 10, 0.5)
rbinom()	To generate random samples from a binomial distribution	To generate 3 random numbers from a Binomial(10, 0.5), we use rbinom(3, 10, 0.5)

The Poisson Distribution

The Poisson Distribution

Overview

A Poisson distribution describes the number of event occurrences in fixed, finite intervals of time or space, with notation **X ~ Pois(λ)**

- λ = the average number of events in an interval

Poisson mean and standard deviation:

$$\mu = \lambda$$

$$\sigma = sqrt(\lambda)$$

The Poisson Distribution

Assumptions

1. Events occur independently
2. The rate at which events occur is constant (the rate cannot be higher in some intervals and lower in other intervals)

The Poisson Distribution

Probability Function

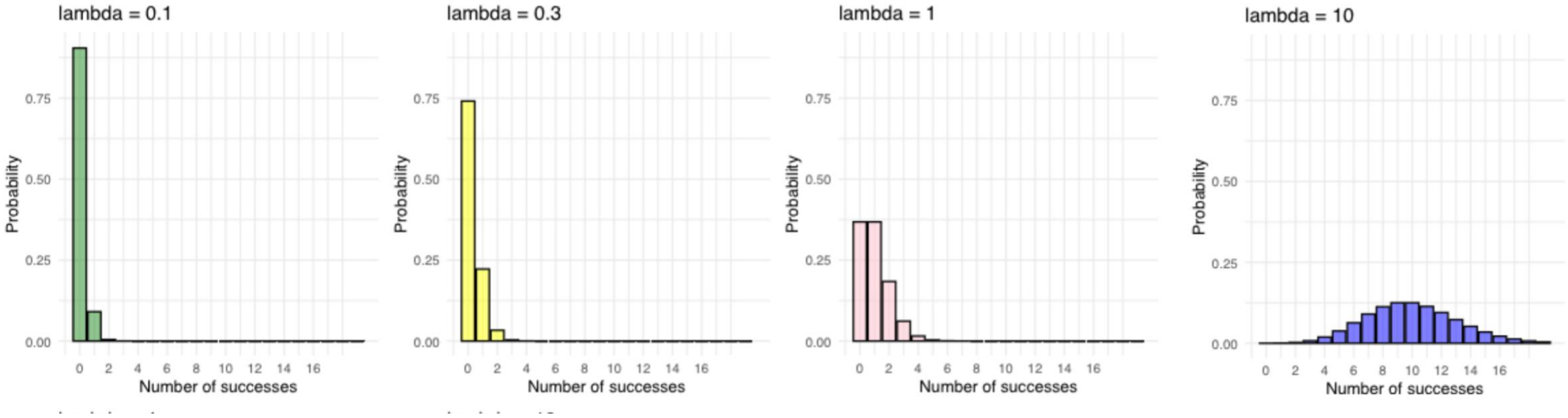
If X has a Poisson distribution, we can use the **probability distribution function** to compute (by hand) the probability that X = k for every k between 0 and n.

$$P(X = k) = \frac{e^{-\mu}\mu^k}{k!}$$

μ (or λ) = mean # of occurrences per interval

The Poisson Distribution

Example



The Poisson Distribution

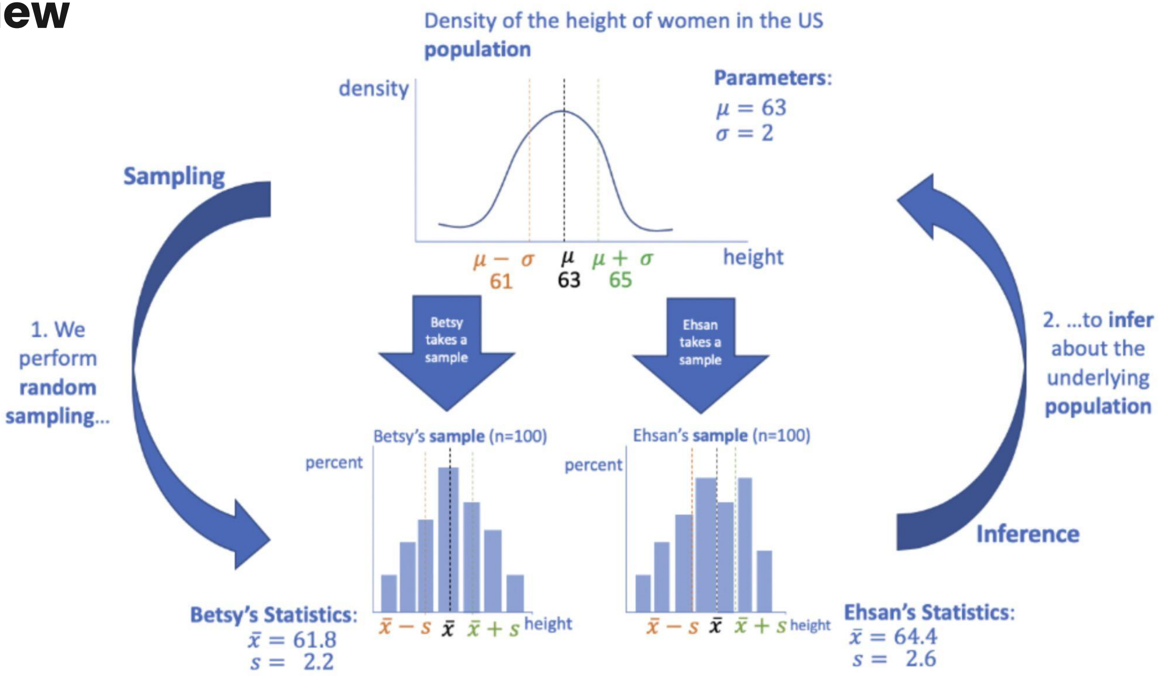
Functions

Function	Use	Example
ppois()	Compute the probability at a specified x or below (lower.tail=T by default). Is a cumulative probability. Need to also specify lambda, the mean of the Poisson distribution	P(X<=2): ppois(2, lambda = 0.5) P(X<5): ppois(4, lambda = 0.5) P(X>5) = 1-P(X<=5) = 1 – ppois(5, lambda = 0.5)
dpois()	Compute an exact probability at a specified value for x, where x is a discrete value between 0 and infinity. Need to also specify lambda, the mean of the Poisson distribution	P(X=2): dpois(2, lambda = 0.5) P(X=1) + P(X=2) (can add them together)
rpois()	Generate a random variable that is drawn from a Poisson distribution. We use this to create/simulate data to get a sense of what Poisson data looks like	rpois(10, lambda = 0.5) generates 10 random draws from a Poisson distribution

Sampling Distributions

Sampling Distributions

Overview



Sampling Distributions

Definitions

We almost never have data for an entire population. However, we can use smaller samples to make inferences about that population

- **Parameter:** A number that describes a characteristic of the population. It's a fixed, single value, but it's usually unknown.
  - Example: The true mean height (μ) of all US women; the true proportion (p) of all California voters who support a certain policy.
- **Statistic:** A number that is computed from a sample. We use statistics to estimate unknown parameters.
  - Example: The mean height $(\bar{x})$ from a sample of 100 women; the proportion $(\hat{p})$ from a poll of 1500 voters.

Sampling Distributions

Definitions

A **sampling distribution** is the distribution of estimates that we have for our parameter. In our case, we either make distributions of estimates for μ or population proportion p.

Sampling Distributions

Definitions

	Population Parameter	Sample Statistic	Mean of Sampling Distribution	Standard Error (SD of Sampling Distribution)
Mean	μ	$\bar{x}$	μ	σ/sqrt(n)
Proportion	p	$\hat{p}$	p	sqrt(p(1-p)/n)

Sampling Distributions

Central Limit Theorem

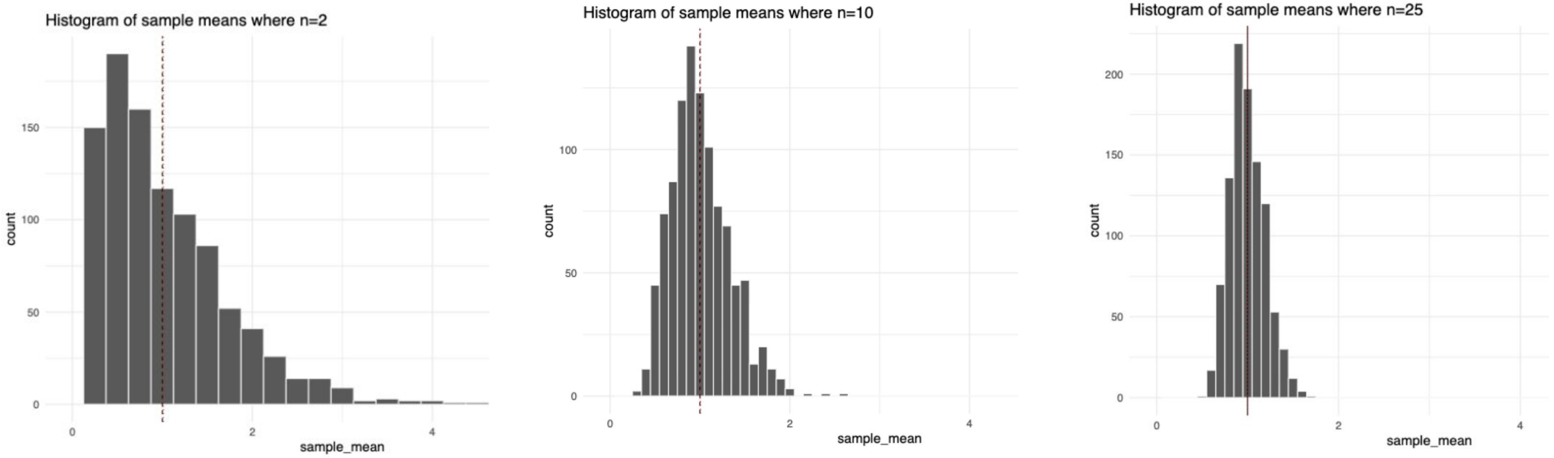
Definition: When n is large, the sampling distribution of the mean $\bar{x}$ is approximately Normal. The shape of the original population distribution does not matter!

As n increases…

- The variability of $\bar{x}$ decreases
- The distribution of $\bar{x}$ becomes more Normal

Sampling Distributions

Central Limit Theorem Example



Inference

Inference

Confidence Intervals

A **confidence interval (CI)** gives us a range of plausible values for the true population parameter.

A confidence interval provides a range of values that likely contain μ. The CI reflects our estimate with some uncertainty/variability, and is given by:

CI = Point estimate ± Margin of error

Inference

Confidence Intervals

CI Formula: $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

$\bar{x}$ = sample mean

z^* = critical value

$\frac{\sigma}{\sqrt{n}}$ = standard error of the mean

Inference

Confidence Intervals

Common Critical Values (z*):

- 90% CI: z* = 1.645
- **95% CI: z* = 1.960**
- 99% CI: z* = 2.576

Inference

Confidence Intervals

Interpretation: "If we were to take 100 different random samples from this population and construct a 95% CI for each, we would expect about 95 of those intervals to contain the true population mean, μ."

Note: Do not use the textbook's shorthand that "we are 95% confident that μ is contained in the CI"

Inference

Hypothesis Testing Steps

Hypothesis testing is the process of using statistics to say how likely something is to be true under certain assumptions

Step 1: Specify the parameter of interest that addresses the research questions

Step 2: Specify the null hypothesis (H_0)

Step 3: Specify the alternative hypothesis (H_A)

Step 4: Set the significance level (α)

Step 5: Calculate the test statistic and p-value based upon the null distribution

Step 6: Write a conclusion

Inference

Hypothesis Testing

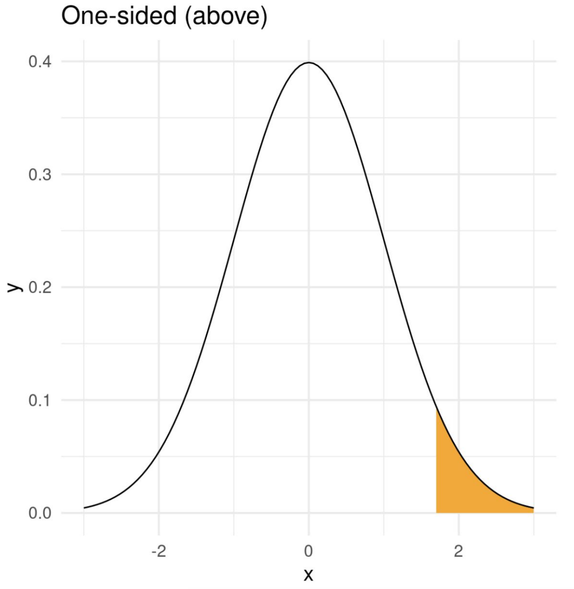
Null Hypothesis (H_0): A statement of no effect or difference between variables.

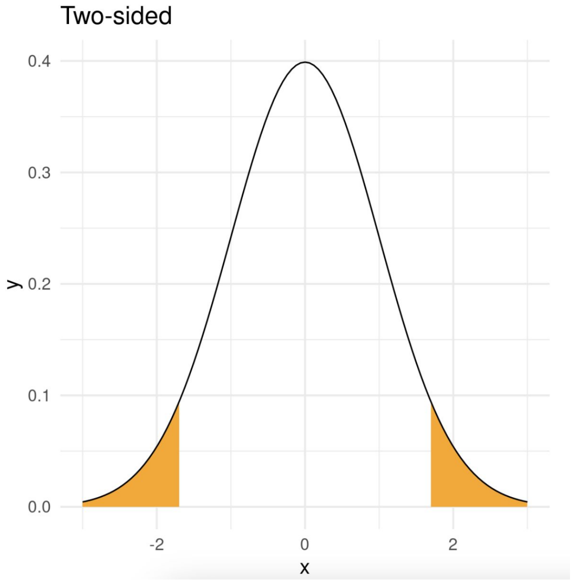
Alternative Hypothesis (H_A): A statement claiming some effect or difference between variables.

- One-sided (H_A): Checks for effect in one direction (ex. higher OR lower)
- Two-sided (H_A): Checks for effect difference in any direction

Inference

Hypothesis Testing





Inference

Z-Test Statistic

In a hypothesis test, we can use the **Z-Test Statistic** to determine if the sample mean is significantly different from the population mean.

Z-Test Statistic: $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$\bar{x}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size

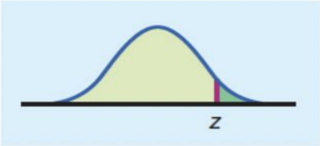
Inference

Finding the p-value

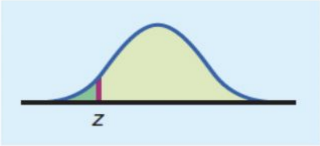
After finding our z-test statistic, we can find the p-value under the Standard Normal curve by marking the observed value of z.

- The direction of your hypothesis will determine the direction of your calculation
- Use pnorm() functions to calculate the p-value

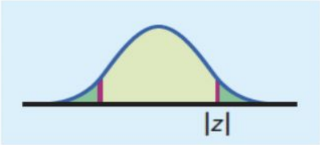
$H_a\colon \mu > \mu_0$ is $P(Z \geq z)$



$H_a\colon \mu < \mu_0$ is $P(Z \leq z)$



$H_a\colon \mu \neq \mu_0$ is $2P(Z \geq |z|)$



Power, Type I, and Type II Error

Power, Type I, Type II Error

Definitions

Type I Error: the chance of making a wrong decision when the null hypothesis is true, equal to significance level (α)

Type II Error: the chance of making a wrong decision when the alternative hypothesis is true, known as (β)

Power: the chance of making the right decision when the alternative is true, known as ($1-\beta$)

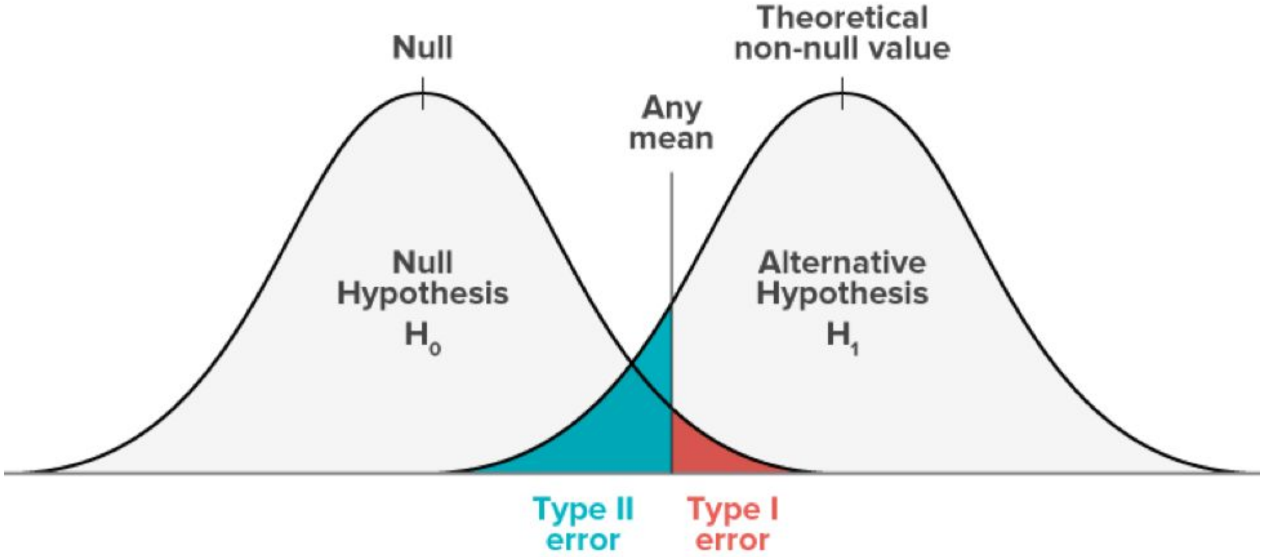
Power, Type I, Type II Error

Error Table

	Reality: H_0 is True	Reality: H_A is True
Fail to Reject H_0	Correct Decision	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision (Power)

Power, Type I, Type II Error

Visualization



Power, Type I, Type II Error

Power Definition

Power is the probability of detecting an effect if one truly exists.
We want studies to have high power (reduce Type II error)!

Methods to increase power:

- Increase the sample size (n)
- Increase the effect size
- Increase the significance level (α)

Practice Problems

Probability Example

1. A study of hearing impairment in dogs examined over 5000 dalmatians for both hearing impairment and iris color. Being impaired was defined as deafness in either one or both ears. Dogs with one or both irises blue (a trait due to low iris pigmentation) were labeled blue. The study found that 28% of the dalmatians were hearing impaired, 11% were blue eyed, and 5% were hearing impaired and blue eyed.

1A. Use probability notation to write the known probabilities from the prompt.
1B. What is the probability of dalmatians being either blue eyed or hearing impaired?
1C. What is the probability of a dalmatian neither being impaired nor a blue-eyed dog?
1D. Are blue eyes and impaired hearing independent or dependent? Why?

Probability Example (Key)

1. A study of hearing impairment in dogs examined over 5000 dalmatians for both hearing impairment and iris color. Being impaired was defined as deafness in either one or both ears. Dogs with one or both irises blue (a trait due to low iris pigmentation) were labeled blue. The study found that 28% of the dalmatians were hearing impaired, 11% were blue eyed, and 5% were hearing impaired and blue eyed.

1A. $P(I) = 0.28$, $P(B) = 0.11$, $P(B \cap I) = 0.05$

1B. $P(B \cup I) = P(B) + P(I) - P(B \cap I) = 0.11 + 0.28 - 0.05 = 0.34$

1C. $P(B^c \cap I^c) = P((B \cup I)^c) = 1 - P(B \cup I) = 1 - 0.34 = 0.66$

1D. $P(B) * P(I) = 0.28 \times 0.11 = 0.0308 \neq P(B \cap I) = 0.05$

So the events B and I are dependent.

Diagnostic Testing Example

2. Suppose that iHealth designed a COVID test that ALWAYS gives a positive result regardless of the disease status. (Hypothetically of course… this test would be useless in real life.) In reality, 2% of the population who take the test actually have COVID.

2A. Are the results of the test and the disease status independent events?
2B. Consider the three metrics for this test: sensitivity, specificity, and positive predictive value. Which of these values would be guaranteed to be 1? Which of these would be guaranteed to be 0? Calculate the remaining metric.
2C. How would the values of sensitivity and specificity from question 2 change if the COVID test ALWAYS gave a negative result?

Diagnostic Testing Example (Key 2A)

2. Suppose that iHealth designed a COVID test that ALWAYS gives a positive result regardless of the disease status. (Hypothetically of course… this test would be useless in real life.) In reality, 2% of the population who take the test actually have COVID.

2A. Yes, they are independent since knowing the result of the test does NOT change the probability of having the disease. (Or equally: knowing the disease status of the individual does not change the probability of getting a positive/negative test.)

Diagnostic Testing Example (Key 2B)

2. Suppose that iHealth designed a COVID test that ALWAYS gives a positive result regardless of the disease status. (Hypothetically of course... this test would be useless in real life.) In reality, 2% of the population who take the test actually have COVID.

2B. In this case, since we have independence between the disease status and test results, we know that P(A|B) = P(A). [By the rules of independence]. Thus

a. Sensitivity = P(Test = + |Disease = TRUE) = P(Test = +) = 1
b. Specificity = P(Test = - |Disease = FALSE) = P(Test = -) = 0
c. Positive Predictive Value = P(Disease = TRUE | Test = +) = P(Disease = TRUE) = 0.02

Diagnostic Testing Example (Key 2C)

2. Suppose that iHealth designed a COVID test that ALWAYS gives a positive result regardless of the disease status. (Hypothetically of course… this test would be useless in real life.) In reality, 2% of the population who take the test actually have COVID.

2C. The values of the specificity and sensitivity would switch. (This is an example why the sensitivity and specificity of a test are both important values to look at. Sometimes healthcare scientists define a test based on some cutoff value of a body metric, like an individual's hormone levels, and picking the right cutoff value for the test involves balancing the test's sensitivity and specificity.)

Tree Diagram Example

3. A study of the presence of cutaneous malignant melanoma at a single body location among the Italian population found that 15% of skin cancers are located on the head and neck area, another 41% on the trunk and the remaining 44% on the limbs. 44% of individuals with skin cancer on the head are men, as are 63% of those with skin cancer on the trunk but only 20% of those with skin cancer on the limbs. What percent of all individuals with skin cancer are women?

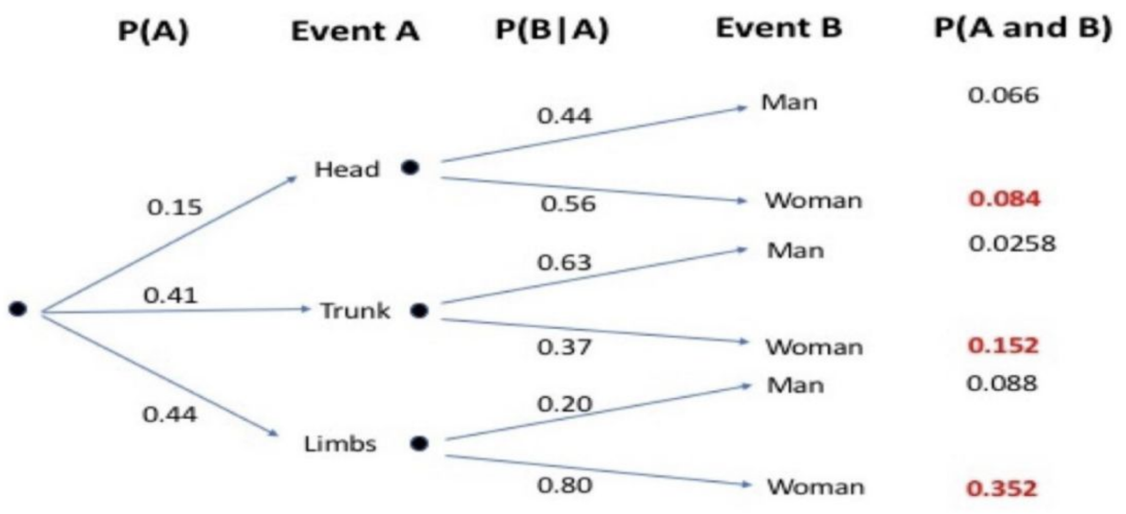
Tip: First, write out your probabilities. Then start filling in the tree diagram.

Tree Diagram Example (Key)

Probabilities:

P(head) = 0.15
P(trunk) = 0.41
P(limbs) = 0.44

P(man|head) = 0.44
P(man | trunk) = 0.63
P(man | limbs) = 0.20



2x2 Table Example

4. Based on the 2x2 table, answer the following questions:

	Has Disease (D+)	No Disease (D-)	Total
Is Exposed (E+)	100	50	150
Not Exposed (E-)	50	300	350
Total	150	350	500

4A. What is the **risk** that someone has the disease?
4B. What are the **odds** that someone has the disease?

2x2 Table Example (Key)

4. Based on the 2x2 table, answer the following questions:

	Has Disease (D+)	No Disease (D-)	Total
Is Exposed (E+)	100	50	150
Not Exposed (E-)	50	300	350
Total	150	350	500

4A. P(Has disease) = P(D+) = 150/500=0.3=30%. The risk of having the disease in this population is 30%.

4B. p/(1-p) = (150/500)/(350/500) = 150/350 = 3/7

Normal Distribution Questions

5. Answer the following questions about the Normal distribution:

5A. What is the difference between Normal and Standard Normal?

5B. What is a z-score? How do you interpret a z-score?

5C. What is the 68-95-99.7 rule for the normal distribution?

Normal Distribution Questions (Key)

5. Answer the following questions about the Normal distribution:

5A. Normal distributions are defined by two parameters: mean (mu) and standard deviation (sigma).A special case is mean=0 and sd=1: called "standard" normal distribution.

5B. For any observation following a known normal distribution, we can "standardize" it, which is the z-score.

E.g. X~N(10,10), then Z=(X-10)/10 ~ N(0,1), z-score following the standard normal distribution

5C. 68-95-99.7 rule states that: under a normal distribution, approximately 68% of the data falls within mu±sigma, 95% of the data falls within mu±2*sigma, 99.7% of the data falls within mu±3*sigma

Binomial Distribution Example

6. The probability that a patient recovers from a widespread flu is 90%. There are now 30 people in a hospital who have contracted the flu independently. What is the probability that 25 patients in that hospital will recover from it?

Binomial Distribution Example (Key)

6. The probability that a patient recovers from a widespread flu is 90%. There are now 30 people in a hospital who have contracted the flu independently. What is the probability that 25 patients in that hospital will recover from it?

Method 1: Calculate "by hand" in R

```
n_trials <- 30
k_success <- 25
probability <- 0.9

choose(n_trials,k_success)*(probability)^k_success*(1-probability)^(n_trials-k_success)

= 0.10
```

Binomial Distribution Example (Key)

6. The probability that a patient recovers from a widespread flu is 90%. There are now 30 people in a hospital who have contracted the flu independently. What is the probability that 25 patients in that hospital will recover from it?

Method 2: Calculate using dbinom()

```
n_trials <- 30
k_success <- 25
probability <- 0.9

dbinom(x=k_success, size=n_trials, prob=probability) → 0.10
```

Poisson Distribution Example

7. Suppose there is a disease that occurs at a rate of 10 per million people for a period of one year. Let X be the number of cases in 1 million people each year.

7A. What is the probability that we see more than 12 cases of the disease in a city within one year?

7B. How about exactly 12?

Poisson Distribution Example (Key)

7. Suppose there is a disease that occurs at a rate of 10 per million people for a period of one year. Let X be the number of cases in 1 million people each year.

7A. What is the probability that we see more than 12 cases of the disease in a city within one year?

ppois(q = 12, lambda = 10, lower.tail = F) → 0.21

7B. How about exactly 12?

dpois(x = 12, lambda = 10) → 0.09

Inference Example

8. An environmentalist group collects a liter of water from each of 45 random locations along a stream. The mean amount of dissolved oxygen per liter is 4.62 mg. Suppose that we know that dissolved oxygen varies among locations according to a Normal distribution with $\sigma = 0.92$ mg. Is there strong evidence that the stream has a mean oxygen content of less than 5 mg per liter? Perform a hypothesis test by completing the following steps:

8A. State your null and alternative hypotheses (hint: is it one-sided or two-sided?)
8B. Calculate the z test-statistic
8C. Find the p-value. Express this using probability notation. What code would you use to find this p-value?
8D. Interpret your p-value in the context of this problem

Inference Example (Key)

8. An environmentalist group collects a liter of water from each of 45 random locations along a stream. The mean amount of dissolved oxygen per liter is 4.62 mg. Suppose that we know that dissolved oxygen varies among locations according to a Normal distribution with σ = 0.92 mg. Is there strong evidence that the stream has a mean oxygen content of less than 5 mg per liter? Perform a hypothesis test by completing the following steps:

8A. $H_0 : \mu = 5mg/L \quad H_a : \mu < 5mg/L$

8B. $z = \frac{4.62 - 5}{.92/\sqrt{45}} = -2.77078$

8C. pnorm(-2.77078, mean = 0, sd = 1) → p-value of 0.0028

8D. We found a p-value of p = 0.0028. This is less than our cut-off value of α = .05. So we have evidence to reject the null hypothesis that the mean amount of dissolved oxygen is 5 mg/L and have evidence to support the alternative hypothesis that the mean amount of dissolved oxygen is less than 5 mg/L.

Type 1 & 2 Error Example

9. You take a sample of 100 people and find that their mean systolic blood pressure is 126 mm Hg. You want to test if the population's average systolic blood pressure is different from the healthy level of 120 mm Hg. Suppose you know the standard deviation of blood pressures in the population is 10 mm Hg.

9A. State the null and alternative hypotheses
9B. Find the confidence interval for this estimate
9C. Suppose after a census you learn that the population's true average systolic blood pressure is 120 mm Hg CI: [119, 121]. What kind of error occurred?

Type 1 & 2 Error Example (Key)

9. You take a sample of 100 people and find that their mean systolic blood pressure is 126 mm Hg. You want to test if the population's average systolic blood pressure is different from the healthy level of 120 mm Hg. Suppose you know the standard deviation of blood pressures in the population is 10 mm Hg.

9A. H0 : μ = 120 mmHg, HA : $\mu \neq$ 120 mmHg

9B. Confidence Interval:

126 - 1.96*(10/sqrt(100)) = 124.04 (lower bound)
126 + 1.96*(10/sqrt(100)) = 127.96 (upper bound)

9C. Reject H0 that mean systolic blood pressure is 120 mm Hg. This is a type I error. We rejected H0 when H0 was actually true.

Power Example

10. You obtain a sample of 30 diabetic individuals from hospitals around Northern California and find that their mean blood sugar level is 207 mg/dL. We know the standard deviation of the population's blood sugar level is 15 mg/dL. We hypothesize that the population of diabetic individuals from hospitals in Northern California have a mean blood sugar level greater than 200 mg/dL.

If we consider an $\alpha = 0.05$ level, how confident can we be that the test will reject H0 when the true underlying mean equals $\mu = 210$ mg/dL?

Power Example (Key)

10. If we consider an $\alpha = 0.05$ level, how confident can we be that the test will reject H0 when the true underlying mean equals $\mu = 210$ mg/dL?

We know the population $\sigma = 15$ mg/dL
n = 30 diabetic individuals from hospitals in Northern California
H0 : $\mu = 200$ mg/dL, Ha : $\mu > 200$ mg/dL
Set $\alpha = 0.05$

Step 1: calculate the minimum z-value required to reject H0

qnorm(p = 0.05, mean = 200, sd = 15/sqrt(30), lower.tail = F)
= 204.5046. So for any sample mean with this value or higher when n = 30, you will reject H0 in favor of HA.

Step 2: calculate the percentile associated with this minimum z-value required to reject H0

pnorm(q = 204.5046, mean = 210, sd = 15/sqrt(30), lower.tail = F)
= 0.9776058 Thus, you have a 98% chance of obtaining evidence in favor of HA when $\mu = 210$ if n = 30.

Good Luck!

- The PH142 Teaching Team