Stat 4202: Mathematical Statistics II

Lecturer: **Professor Andrew Kerr**
Notes by: Farhan Sadeek

Spring 2025

1 January 6, 2025

STAT 4202 will rely a lot on STAT 4201. So we need to have a pretty good understanding of those concepts.

1.1 Review of Probability Theory

Definition 1
The **Sample Space**, denoted by $\mathcal{S}$, is the set of all outcomes from an experiment.

Definition 2
An **Event**, usually denoted by a capital letter such as A or B, is a subset of the Sample Space.

The probability function

- $P(A) \geq 0$
- $P(\mathcal{S}) = 1$
- For disjoint sets $A_1, A_2, \cdots, A_n$:
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

If an event A is a subset of another event B, then the probability of A is less than or equal to the probability of event B. That is to say, if $A \subseteq B$, then $P(A) \leq P(B)$

The complement of an event A, denoted by A^c, has a probability equal to one minus the probability of the event A. That is,

$$P(A^c) = 1 - P(A)$$

A partition of a sample space $\mathcal{S}$ is an exhaustive, non-overlapping collection of events $A_1, A_2, \cdots, A_n$ that is exhaustive and mutually exclusive:

$$\bigcup_{i=1}^{n} A_i = \mathcal{S}$$

and

$$A_i \cap A_j = \emptyset \quad \forall i \neq j$$

For any partition, we have

$$\sum_{i=1}^{n} P(A_i) = 1$$

Two events A and B are **independent** if the outcome of one doesn't affect the likelihood of the occurrence of the other. For two independent events, we have

$$P(A \cap B) = P(A)P(B)$$

The **conditional probability** of A given B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Lemma 3

Note that if A and B are independent, then

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(B)}{P(B)} \\ &= P(A) \end{aligned}$$

Corollary 4

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$

1.2 Random Variables

Definition 5

A random variable is a function that takes outcomes from the sample space $\mathcal{S}$ to the real numbers $\mathbb{R}$. That is, a random variable is a function $X : \mathcal{S} \to \mathbb{R}$.

We then use a probability mass function (pmf) in the discrete case or a probability density function (pdf) in the continuous case:

pmf:	$f_X(x) = P(X = x)$	when X is discrete
pdf:	$\int_a^b f_X(x)\mathrm{d}x = P(a \leq X \leq b)$	when X is continuous

The cumulative distribution function (cdf) gives the probability of observing a value less than or equal to a given value x:

$$F_X(x) = P(X \leq x)$$

When X is a continuous random variable, the pdf is the derivative of the cdf:

$$f_X(x) = F_X'(x)$$

1.3 Expected Value and Variance

For random variable X, the **expected value** is denoted by $E(X)$ and is given by:

$$E(X) = \begin{cases} \sum_x x f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

The **variance** of a random variable X is denoted by $\mathrm{Var}(X)$ and is given by:

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right]$$

1.4 Covariance

The **covariance** of two random variables X and Y is denoted by:

$$\mathrm{Cov}(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

If two random variables X and Y are independent, then

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

So, we will be using these formulas to estimate the mean and the variance throughout the semester.

2 January 8, 2025

2.1 Statistical Models

In statistics, we often model data $X_1, X_2, \cdots, X_n$ as a random sample from a population. We assume that the data are independent and identically distributed (iid) random variables. The goal is to estimate the parameters of the population distribution.

Definition 6

A **parameter** of a distribution are values that describe a certain characteristic of the given distribution.

Some examples of parameters include:

- The mean height of **all** OSU incoming freshmen.
- The proportion of registered voters that voted for a particular candidate.
- The standard deviation of waiting times for **all** customers shopping at a store during a week.

Fact 7

If $X_1, X_2, \ldots, X_n \overset{iid}{\approx} f_X(x)$ then $\mu = E(X_i)$ is a parameter, which is the mean of the distribution. The variance is also a parameter: $\sigma^2 = E[(X - \mu)^2]$

Example 8

Suppose we are examining the efficacy difference between a newly developed drug and an existing drug. We look at the differences, Δi, from a series of n comparative samples. Note that these will all come from some distribution:

$$\Delta_1, \Delta_2, \ldots, \Delta_n, \overset{iid}{f}(x)$$

Fact 9

Here the independed is a really important to look for we will look thorogh that through the semester.

For a parametric model

$$\{f_g(x)_{\theta \in \mathbb{R}}\}$$

Which is indexed by a vector θ of parameters.

Example 10

Suppose we wanted to estimate the height and weight of all incoming students at Ohio State. We could take a random sample of n of the incoming students and observe the height (H) and weight (W) of each student, giving the following sample data:

$$(H_1, W_1), (H_2, W_2), \ldots, (H_n, W_n)$$

We can then consider the following model:

$$N(\mu, \Sigma)$$

3 January 8, 2025

We went over the **Recitation Logistics** and **Quiz 1**.

4 January 10, 2025 (In-Person)

We wanted to check how to get estimators. We will do the backwards this week for.

4.1 Unbiased Estimator

Definition 11
An **estimator** $\hat{(}\theta)$

Definition 12
An **unbiased estimator** is an estimator that is equal to the parameter it estimates. That is, if $\hat{\theta}$ is an unbiased estimator of θ, then $E(\hat{\theta}) = \theta$.

4.1.1 Interval Estimation

Definition 13
A **confidence interval** is an interval estimate for a parameter θ that provides a range of values within which the parameter is expected to lie with a certain degree of confidence. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are values of the random variables $\hat{\theta}_1$ and $\hat{\theta}_2$ such that

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$$

for some specified probability $1 - \alpha$, we refer to the interval

$$\hat{\theta}_1 < \theta < \hat{\theta}_2$$

as a $(1 - \alpha)100\%$ confidence interval for θ. The probability $1 - \alpha$ is called the degree of confidence, and the endpoints of the interval are called the lower and upper confidence limits.

Theorem 14
If $\bar{X}$, the mean of a random sample of size n from a normal population with the known variance σ^2, is to be used as an estimator of the mean of the population, the probability is $1 - \alpha$ that the error will be less than $\frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$.

Theorem 15
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with mean μ and variance σ^2. If $\bar{X}$ is the sample mean, then

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

Then the interval

$$\bar{X} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$$

is a $(1 - \alpha)100\%$ confidence interval for the mean of the population.

Theorem 16
If $\bar{X}$ and s are the values of the mean and the standard deviation of a random sample of size n from a normal population, then

$$\bar{X} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

is a $(1-\alpha)100\%$ confidence interval for the mean of the population.

Fact 17
When $n < 30$, the t-distribution should be used instead of the normal distribution to account for the increased variability in the estimate of the standard deviation.

Theorem 18
If x_1 and x_2 are the values of the means of independent random samples of sizes n_1 and n_2 from normal populations with the known variances σ_1^2 and σ_2^2, then

$$(x_1 - x_2) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (x_1 - x_2) + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

is a $(1-\alpha)100\%$ confidence interval for the difference between the two population means.

Theorem 19
If x_1, x_2, s_1, and s_2 are the values of the means and the standard deviations of independent random samples of sizes n_1 and n_2 from normal populations with equal variances, then

$$(x_1 - x_2) - t_{\alpha/2,n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (x_1 - x_2) + t_{\alpha/2,n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

is a $(1-\alpha)100\%$ confidence interval for the difference between the two population means.

5 February 21, 2025

5.1 Hypothesis Testing

Definition 20 (Statistical Hypothesis)
A **Statistical Hypothesis** is an assertion or conjecture about the distribution of one or more random variables. For example, the claim that $\mu_2 \geq \mu_1$

Definition 21 (Null Hypothesis)

The hypothesis that we would like to provide evidence against is called the **null hypothesis** and is denoted by H_0.

Example 22

In the drug example, the null hypothesis is

$$H_0 : \mu_2 \geq \mu_1 \text{ (or } \mu_1 = \mu_2\text{)} \tag{1}$$

The hypothesis $mu_2 > \mu_1$ is called the **alternative hypothesis** and is denoted by H_a or H_1.

Definition 23

The rejection region is also referred as the **critical region**. The size of the critical region is also known as the **Level of Significance** of the test, and the level of significance is denoted by the probability of type I error, or by α.

Example 24

Suppose we wish to test the hypotheses that

$$H_0 : \theta = 0.9 \text{ vs } H_1 : \theta = 0.6$$

For a binomial distribution with $n = 20$ samples, with the random variable X defined as the count of the number of successes. The rejection region for this test is when $X \leq 14$.

- What is the significance level, α, for this test?
- What is the probability of a type II error, β?
- What happens to the values of α and β when we change the rejection region to be $X \leq 15$?
- What happens to the values of α and β when we change the rejection region to be $X \leq 13$?

Example 25

Suppose we take a random sample $X_1, X_2, X_3, \ldots, X_n \sim N(\mu, 1)$ and we wish to test the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_1$$

with $\mu_1 > \mu_0$. Our test procedure is to reject H_0 if $\bar{X} > k$ for some real number k. Find the value of k such that the probability of a type I error is 0.05.

Example 26

Continuation from the previous example

If $\mu_0 = 10$ and $\mu_1 = 11$, determine the minimal sample size so that $\beta \leq 0.06$ using the test with

$$k = \mu_0 + \frac{1.645}{\sqrt{n}}$$

Definition 27

The **Power of a Test**, given $1 - \beta$, is the probability that H_0 is rejected given that H_0 is false. In our example, $1 - \beta$ is the power of the test, $\theta = \theta_1$ for

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1$$

6 February 28, 2025

6.1 Tests of Significance

A statistical test, which specifies a simple hypothesis, the size of the critical region α, and a composite alternative hypothesis is called a Test of Significance. For such tests, α is referred to as the level of significance.

Example 28

Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ where σ^2 is known. A two-tailed test for

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0$$

is

Example 29

A one-tailed test for

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu < \mu_0$$

is to reject H_0 if

6.1.1 Four Steps to Hypothesis Testing

1. Formulate H_0 and H_a and specify α.
2. Specify the test statistic and define the critical region of size α.
3. Determine the value of the (observed) test statistic from the data.
4. Check whether the value of the test statistic falls in the rejection region and accordingly, reject or fail to reject H_0.

Example 30
Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ where σ^2 is known. Consider the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0$$

1. Formulate H_0 and H_a and specify α.
2. Specify the test statistic and define the critical region of size α.
3. Determine the value of the (observed) test statistic from the data.
4. Check whether the value of the test statistic falls in the rejection region and accordingly, reject or fail to reject H_0.

6.2 P-Values

It is oftentimes more informative to compute the so-called p-value of a test and compare it to α to decide whether to reject H_0 or not.

Example 31
Consider again the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0$$

Test Statistic: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ Rejection Region: $|Z| \geq z_{\alpha/2}$ p-value $= \alpha^*$ so that

In general, we have

$$\text{p-value} = \begin{cases} P(Z \geq z^* \mid H_0) & \text{if } H_a : \mu > \mu_0 \\ P(Z \leq z^* \mid H_0) & \text{if } H_a : \mu < \mu_0 \\ 2P(Z \geq |z^*| \mid H_0) & \text{if } H_a : \mu \neq \mu_0 \end{cases}$$

where Z is the test statistic and z^* is the observed test statistic.

Corresponding to an observed value of a test statistic, the p-value is the lowest level of significance at which the null hypothesis could have been rejected.

By definition of a p-value, we can show that if a p-value $\leq \alpha$, then we would reject H_0 at the level of significance α.

6.2.1 Alternate Testing Procedure

Based on this, we can modify steps 2-4 to be

1. Formulate H_0 and H_a and specify α.
2. Specify the test statistic.
3. Determine the value of the (observed) test statistic and the corresponding p-value from the data.
4. Check if p-value $\leq \alpha$ and accordingly, reject or fail to reject H_0.

6.3 Tests Concerning Means

Suppose we consider the null hypothesis $H_0 : \mu = \mu_0$ and assume that the population variance, σ^2, is known. Let $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ be the test statistic. Given a level of significance α, the rejection region is

$$Z \geq z_\alpha \quad \text{if } H_a : \mu > \mu_0$$

$$Z \leq -z_\alpha \quad \text{if } H_a : \mu < \mu_0$$

$$|Z| \geq z_{\alpha/2} \quad \text{if } H_a : \mu \neq \mu_0$$

The p-value is

$$P(Z \geq z \mid H_a : \mu > \mu_0)$$

$$P(Z \leq -z \mid H_a : \mu < \mu_0)$$

$$2P(Z \geq z \mid H_a : \mu \neq \mu_0)$$

where $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$.

Example 32

Given the following summary statistics

$$\sigma = 0.16, \quad \bar{X} = 8.091, \quad n = 25, \quad \alpha = 0.01$$

Test the hypotheses

$$H_0 : \mu = 8 \quad \text{vs} \quad H_a : \mu \neq 8$$

6.3.1 Another Application

When $n \geq 30$, we can replace σ by s if σ is unknown. In this case, we have

$$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0,1) \text{ approximately}$$

Example 33

Given the following summary statistics

$$\bar{X} = 21819, \quad s = 1295, \quad n = 100, \quad \alpha = 0.05$$

Test the hypotheses

$$H_0 : \mu = 22000 \quad \text{vs} \quad H_a : \mu < 22000$$