

Statistics for Economists II

Week 2: Sampling Distributions & Order Statistics

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 From Last Week to This Week
- 2 The Chi-Squared Distribution
- 3 Student's t -Distribution
- 4 The F -Distribution
- 5 Relationships Between the Distributions
- 6 Sampling Distribution of $\bar{X}$ and S^2
- 7 Order Statistics
- 8 Key Takeaways

Where We Stand

Last week we established two powerful results:

- The **Law of Large Numbers**: $\bar{X}_n \xrightarrow{P} \mu$ (sample means converge to population mean)
- The **Central Limit Theorem**: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0, 1)$ (sample means are approximately normal)

This week's aim: The CLT uses σ , but in practice we almost **never know** σ . When we replace σ with its estimate S , the resulting statistic no longer follows a normal distribution. We need three new distributions:

- ① The χ^2 **distribution**: governs the behavior of sample variances
- ② The t -**distribution**: governs means when σ is estimated
- ③ The F -**distribution**: governs ratios of variances

These three distributions are the **building blocks** for confidence intervals (Weeks 7–8) and hypothesis testing (Weeks 9–11). Every t -test and F -test in econometrics rests on this week's material.

The Big Question

Recall from the CLT:

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

Problem: In practice, we almost **never** know σ .

We must *estimate* it with $S = \sqrt{S^2}$, where

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

Today's journey: Normal $\longrightarrow \chi^2 \longrightarrow t \longrightarrow F \longrightarrow$ Order Statistics

The Key Issue

Replacing σ with S introduces **extra randomness**. The resulting ratio is *not* normal anymore—it follows a ***t*-distribution**. Understanding *why* requires knowing the χ^2 distribution first.

Why Learn These Three Distributions?

Every time you run a regression in econometrics, the software reports:

- t -**statistics** next to each coefficient
- An F -**statistic** for the whole model
- Tests for **variance** based on χ^2

The Practical Payoff

- χ^2 : Variance tests, goodness of fit
- t : Coefficient tests, confidence intervals
- F : Joint hypotheses, model comparison

Definition: Chi-Squared Distribution

Definition

Let $Z_1, Z_2, \dots, Z_k$ be independent $N(0, 1)$ random variables. Then

$$Q = Z_1^2 + Z_2^2 + \dots + Z_k^2 \sim \chi^2(k)$$

is said to have a **chi-squared distribution with k degrees of freedom**.

The χ^2 distribution arises from **squaring and summing** standard normals. Think of it as measuring the *total squared deviation* in k independent directions.

- It is always **non-negative** (sum of squares!)
- It is **right-skewed**, especially for small k
- As $k \rightarrow \infty$, it becomes approximately normal (by the CLT itself!)

Where Does $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ Come From?

Step 1: The χ^2 distribution.

If $Z_1, Z_2, \dots, Z_n \stackrel{iid}{\sim} N(0, 1)$, then *by definition*:

$$\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

Where Does $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ Come From?

Step 1: The χ^2 distribution.

If $Z_1, Z_2, \dots, Z_n \stackrel{iid}{\sim} N(0, 1)$, then *by definition*:

$$\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

Step 2: Standardize with the true mean.

Let $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Since $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0, 1)$:

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

This is a sum of n *independent* squared standard normals.

Where Does $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ Come From?

Step 3: Replace μ with $\bar{X}$.

We don't know μ , so we use $\bar{X}$. The sample variance is:

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

Rearranging:

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

Where Does $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ Come From?

Step 3: Replace μ with $\bar{X}$.

We don't know μ , so we use $\bar{X}$. The sample variance is:

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

Rearranging:

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

Step 4: Why $n-1$ instead of n ?

This *looks like* a sum of n squared standard normals, but $\bar{X}$ is computed from the same data. This imposes a constraint:

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

Once you know $n-1$ of the deviations, the last one is determined.

Properties of the χ^2 Distribution

If $Q \sim \chi^2(k)$, then:

$$\mathbb{E}(Q) = k, \quad \text{Var}(Q) = 2k$$

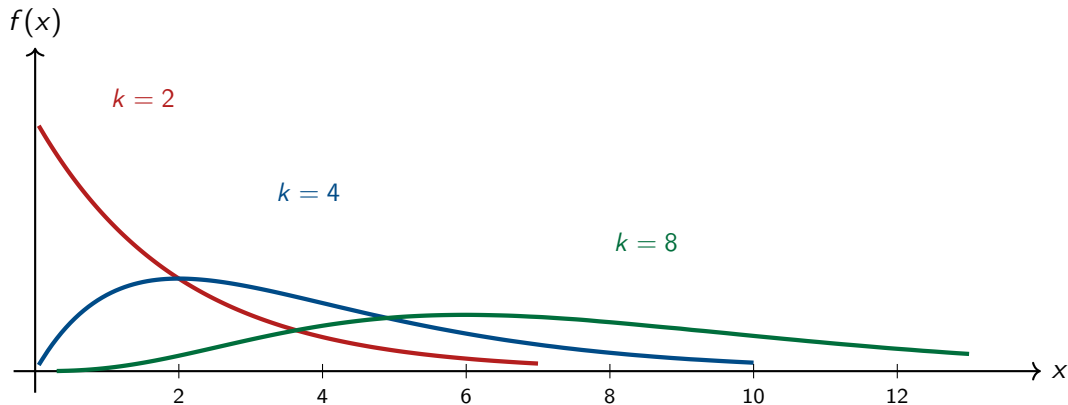
Key properties:

① **Additivity:** If $Q_1 \sim \chi^2(k_1)$ and $Q_2 \sim \chi^2(k_2)$ are independent, then

$$Q_1 + Q_2 \sim \chi^2(k_1 + k_2)$$

The mean equals the degrees of freedom. This makes sense: each Z_i^2 contributes $\mathbb{E}(Z_i^2) = 1$ to the sum. The variance grows at rate $2k$ because squared normals are more variable than normals themselves.

Shapes of the χ^2 Distribution



- $k = 1, 2$: strongly right-skewed

- As k increases: shifts right, more symmetric

The χ^2 and the Sample Variance

Theorem

Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then:

① $\bar{X}$ and S^2 are independent

② $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

- **Why $n-1$?** Estimating μ by $\bar{X}$ imposes one constraint ($\sum(X_i - \bar{X}) = 0$), so we lose one degree of freedom.
- **Why independence?** A special property of the normal distribution. It is what makes the t -test work cleanly.
- **Economics:** This lets us build confidence intervals for variance—useful for measuring risk in financial markets.

Definition: Student's t -Distribution

Definition

Let $Z \sim N(0, 1)$ and $W \sim \chi^2(\nu)$ be **independent**. Then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t(\nu)$$

has a **Student's t -distribution** with ν (“nu”) degrees of freedom.

The t -distribution is what you get when you take a standard normal and **divide by a noisy estimate of its scale**. The extra noise makes the tails **heavier**—extreme values are more likely because the denominator can be small by chance.

Notation Guide

We use ν (Greek “nu”) for degrees of freedom. When testing a sample mean: $\nu = n - 1$ (e.g., $n = 12 \Rightarrow T \sim t(11)$). Some books write k or df —same thing.

Why We Need the t -Distribution

When $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$:

If σ is known

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

Exact. Use z -tables.

If σ is **unknown** (realistic)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$$

Exact. Use t -tables.

This is one of the most important results in applied statistics. Every time you run a regression, software reports t -statistics and p -values based on this distribution. We almost never know σ in economics, so we *always* estimate it.

Properties of the t -Distribution

If $T \sim t(\nu)$:

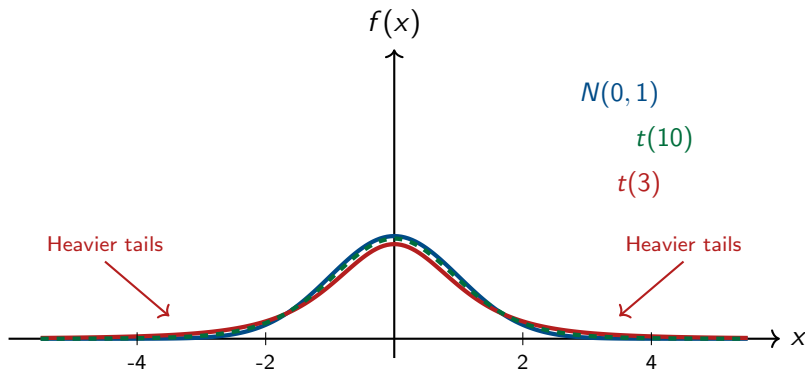
- $\mathbb{E}(T) = 0$ for $\nu > 1$ (symmetric about zero)
- $\text{Var}(T) = \frac{\nu}{\nu - 2}$ for $\nu > 2$ (greater than 1, so wider than normal!)
- As $\nu \rightarrow \infty$: $t(\nu) \rightarrow N(0, 1)$

Practical Rule

For $\nu \geq 30$, the t -distribution is nearly indistinguishable from the standard normal. This is why many textbooks say “use z when $n \geq 30$.” But it is better practice to always use t when σ is estimated.

The variance $\nu/(\nu - 2) > 1$ captures the **extra uncertainty** from estimating σ . With few observations (ν small), S^2 is a noisy estimate, so the tails are heavy. With many observations (ν large), $S \approx \sigma$ and the t looks like a z .

Comparing t and Normal Distributions



The t distribution has the **same bell shape** but **heavier tails**. As df increases, $t \rightarrow N(0, 1)$.

Definition: F -Distribution

Definition

Let $U \sim \chi^2(d_1)$ and $V \sim \chi^2(d_2)$ be **independent**. Then

$$F = \frac{U/d_1}{V/d_2} \sim F(d_1, d_2)$$

has an **F -distribution** with d_1 and d_2 degrees of freedom.

The F -distribution is a **ratio of two variance estimates**. It answers the question: “Are these two sources of variability comparable, or is one significantly larger than the other?”

- Always **non-negative** (ratio of non-negative quantities)
- Right-skewed, especially when d_1 or d_2 is small
- Named after R. A. Fisher, a founder of modern statistics

Properties of the F -Distribution

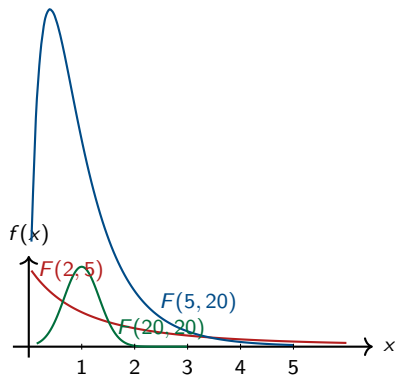
If $F \sim F(d_1, d_2)$:

- $\mathbb{E}(F) = \frac{d_2}{d_2 - 2}$ for $d_2 > 2$ (close to 1 when d_2 is large)
- If $T \sim t(\nu)$, then $T^2 \sim F(1, \nu)$ (connection to t -test!)
- If $F \sim F(d_1, d_2)$, then $1/F \sim F(d_2, d_1)$

Where F Appears in Economics

- ➊ **Comparing variances:** Are returns of stock A more volatile than stock B?
- ➋ **ANOVA:** Do wages differ significantly across three industries?
- ➌ **Regression:** The overall F -test checks whether *any* of your explanatory variables matter
- ➍ **Chow test:** Did a structural break occur (e.g., policy change)?

Shape of the F -Distribution



As both d_1 and d_2 grow, F concentrates around ≈ 1 and becomes more symmetric.

Practical Note

You do **not** need to memorize t or F -tables. In exams, tables will be **provided**. Focus on *when* and *why* these tests are used.

Summary of the Three Distributions

	$\chi^2(k)$	$t(\nu)$	$F(d_1, d_2)$
Definition	$\sum Z_i^2$	$Z/\sqrt{V/\nu}$	$(U/d_1)/(V/d_2)$
Support	$[0, \infty)$	$(-\infty, \infty)$	$[0, \infty)$
Mean	k	0	$d_2/(d_2 - 2)$
Variance	$2k$	$\nu/(\nu - 2)$	complicated
Shape	Right-skewed	Symmetric, heavy tails	Right-skewed
Limit	$\rightarrow N(k, 2k)$	$\rightarrow N(0, 1)$	$\rightarrow 1$
Used for	Variance tests	Mean tests (σ unknown)	Comparing variances

Memorize the “**Used for**” row. In your future econometrics courses, you will see t -statistics on every regression coefficient and F -statistics for joint hypothesis tests.

Sampling Distribution of the Sample Mean

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$.

Exact Result

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{exactly, not just approximately})$$

Standardized Versions

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

$$\sigma \text{ unknown: } T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

For normal populations, $\bar{X}$ is **exactly** normal. The t -statistic is the workhorse of applied economics.

Sampling Distribution of the Sample Variance

Theorem

Let $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

From this: $\mathbb{E}(S^2) = \sigma^2$ (unbiased!) and $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$.

- S^2 is **unbiased** for σ^2 —the $n-1$ denominator ensures this
- $\text{Var}(S^2) \rightarrow 0$ as $n \rightarrow \infty$, so S^2 is also **consistent**
- The χ^2 link lets us build **confidence intervals for variance**—crucial for risk measurement

Quality Control Application

Scenario: A factory producing electronic components claims that the variance of component lifetimes is $\sigma^2 = 100$ (hours²). A quality inspector samples $n = 20$ components and finds $s^2 = 165$.

Question: Is the actual variance larger than claimed?

Quality Control Application

Scenario: A factory producing electronic components claims that the variance of component lifetimes is $\sigma^2 = 100$ (hours²). A quality inspector samples $n = 20$ components and finds $s^2 = 165$.

Question: Is the actual variance larger than claimed?

Test statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19 \times 165}{100} = 31.35$$

Under H_0 : $\chi^2 \sim \chi^2(19)$. Critical value: $\chi_{0.05,19}^2 = 30.14$.

Since $31.35 > 30.14$, we **reject** at 5% level.

The factory's quality claim is questionable. The observed variance is significantly larger than stated. This has direct economic consequences: warranty costs, return rates, and customer satisfaction all depend on product variability.

Order Statistics: Sorting the Sample

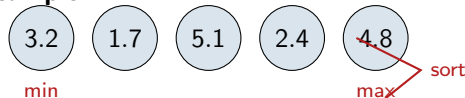
Definition

Given a random sample $X_1, X_2, \dots, X_n$, the **order statistics** are the sample values arranged in ascending order:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

where $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i X_i$.

Original sample:



Order statistics:



Order Statistics: Detailed Setup

Definition. Let $X_1, X_2, \dots, X_n$ be an i.i.d. random sample from a distribution with CDF $F(x)$ and PDF $f(x)$.

Sort them from smallest to largest:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

$X_{(1)} = \min$, $X_{(n)} = \max$, $X_{(k)}$ = the k -th smallest value.

Order Statistics: Detailed Setup

Definition. Let $X_1, X_2, \dots, X_n$ be an i.i.d. random sample from a distribution with CDF $F(x)$ and PDF $f(x)$.

Sort them from smallest to largest:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

$X_{(1)} = \min$, $X_{(n)} = \max$, $X_{(k)}$ = the k -th smallest value.

Goal: Derive the CDF and PDF of $X_{(k)}$.

CDF of $X_{(k)}$: Binomial Counting

Question: What is $P(X_{(k)} \leq x)$?

CDF of $X_{(k)}$: Binomial Counting

Question: What is $P(X_{(k)} \leq x)$?

The k -th smallest value is $\leq x \iff$ *at least* k observations fall at or below x .

CDF of $X_{(k)}$: Binomial Counting

Question: What is $P(X_{(k)} \leq x)$?

The k -th smallest value is $\leq x \iff$ *at least* k observations fall at or below x .

Each observation independently falls below x with probability $F(x)$, so the number of observations $\leq x$ follows a $\text{Binomial}(n, F(x))$ distribution.

CDF of $X_{(k)}$: Binomial Counting

Question: What is $P(X_{(k)} \leq x)$?

The k -th smallest value is $\leq x \iff$ *at least* k observations fall at or below x .

Each observation independently falls below x with probability $F(x)$, so the number of observations $\leq x$ follows a $\text{Binomial}(n, F(x))$ distribution.

Therefore:

$$F_{X_{(k)}}(x) = P(X_{(k)} \leq x) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}$$

PDF of $X_{(k)}$: The Intuition

For $X_{(k)}$ to land in a tiny interval $[x, x + dx]$, three things must happen simultaneously:

PDF of $X_{(k)}$: The Intuition

For $X_{(k)}$ to land in a tiny interval $[x, x + dx]$, three things must happen simultaneously:

1. $k - 1$ obs. fall below x

Probability: $[F(x)]^{k-1}$

Choose which ones: $\binom{n}{k-1}$

ways

PDF of $X_{(k)}$: The Intuition

For $X_{(k)}$ to land in a tiny interval $[x, x + dx]$, three things must happen simultaneously:

1. $k - 1$ obs. fall below x

Probability: $[F(x)]^{k-1}$

Choose which ones: $\binom{n}{k-1}$
ways

2. 1 obs. lands in

$[x, x + dx]$

Probability: $f(x) dx$

$(n - k + 1)$ candidates remain

PDF of $X_{(k)}$: The Intuition

For $X_{(k)}$ to land in a tiny interval $[x, x + dx]$, three things must happen simultaneously:

1. $k - 1$ obs. fall below x

Probability: $[F(x)]^{k-1}$

Choose which ones: $\binom{n}{k-1}$
ways

2. 1 obs. lands in

$[x, x + dx]$

Probability: $f(x) dx$

$(n - k + 1)$ candidates remain

3. $n - k$ obs. fall above x

Probability: $[1 - F(x)]^{n-k}$

PDF of $X_{(k)}$: The Intuition

For $X_{(k)}$ to land in a tiny interval $[x, x + dx]$, three things must happen simultaneously:

- | | | |
|---|---|---|
| 1. $k - 1$ obs. fall below x | 2. 1 obs. lands in $[x, x + dx]$ | 3. $n - k$ obs. fall above x |
| Probability: $[F(x)]^{k-1}$ | | Probability: $[1 - F(x)]^{n-k}$ |
| Choose which ones: $\binom{n}{k-1}$ ways | Probability: $f(x) dx$
$(n - k + 1)$ candidates remain | |

Collecting the combinatorial terms:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

Minimum ($k = 1$):

$$f_{X_{(1)}}(x) = n [1 - F(x)]^{n-1} f(x)$$

One obs. lands at x , all $n - 1$ others must be above it.

Minimum ($k = 1$):

$$f_{X_{(1)}}(x) = n [1 - F(x)]^{n-1} f(x)$$

One obs. lands at x , all $n - 1$ others must be above it.

Maximum ($k = n$):

$$f_{X_{(n)}}(x) = n [F(x)]^{n-1} f(x)$$

One obs. lands at x , all $n - 1$ others must be below it.

Why Do Order Statistics Matter?

Order statistics are the foundation of **non-parametric statistics**—methods that work without assuming a specific distribution. They are critical in economics for measuring **inequality** (percentiles, Gini coefficient), **extreme risk** (Value-at-Risk), and **auction theory** (highest/lowest bids).

Important derived quantities:

- **Sample range:** $R = X_{(n)} - X_{(1)}$ —simplest measure of spread
- **Sample median:** $X_{((n+1)/2)}$ if n is odd—robust measure of center
- **Quantiles:** $X_{(\lceil np \rceil)}$ estimates the p -th population quantile
- **Minimum** $X_{(1)}$: lowest bid in an auction, worst-case scenario
- **Maximum** $X_{(n)}$: highest bid, best-case scenario

Next: We derive the distributions of $X_{(n)}$ (max) and $X_{(1)}$ (min).

PDF of the Maximum and Minimum

Maximum: The event $\{X_{(n)} \leq y\}$ happens iff *all* $X_i \leq y$:

$$F_{X_{(n)}}(y) = [F(y)]^n \quad \Rightarrow \quad \boxed{f_{X_{(n)}}(y) = n[F(y)]^{n-1}f(y)}$$

Minimum: The event $\{X_{(1)} > y\}$ happens iff *all* $X_i > y$:

$$F_{X_{(1)}}(y) = 1 - [1 - F(y)]^n \quad \Rightarrow \quad \boxed{f_{X_{(1)}}(y) = n[1 - F(y)]^{n-1}f(y)}$$

Why does the max formula make sense? For $X_{(n)}$ to be exactly y : one observation lands at y ($f(y)$), all other $n - 1$ must be $\leq y$ ($[F(y)]^{n-1}$), and any of the n could be the one at y (factor n).

Beautiful symmetry: The max uses $F(y)$ (prob. of being $\leq y$), the min uses $1 - F(y)$ (prob. of being $> y$).

Distribution of a Single Order Statistic

Theorem

If $X_1, \dots, X_n$ are i.i.d. with CDF F and PDF f , the PDF of $X_{(k)}$ is:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

Special cases:

$$f_{X_{(1)}}(x) = n [1 - F(x)]^{n-1} f(x) \quad (\text{minimum})$$

$$f_{X_{(n)}}(x) = n [F(x)]^{n-1} f(x) \quad (\text{maximum})$$

The formula has a beautiful **combinatorial** interpretation: for $X_{(k)}$ to be near x , we need exactly $k - 1$ observations below x , one observation at x , and $n - k$ observations above x . The factorials count the arrangements.

Example in class

Key Takeaways from Week 2

- 1 The χ^2 distribution arises from **summing squared standard normals** and governs the behavior of S^2
- 2 Student's t is used when σ is **unknown** (almost always); heavier tails give more conservative inference
- 3 The F -distribution is a **ratio of variances**—central to comparing groups and joint hypothesis tests
- 4 All three distributions **originate from the normal**; normality of the data is the key assumption
- 5 **Order statistics** give us tools for non-parametric methods, extreme-value analysis, and auction theory
- 6 As n grows, $t \rightarrow z$ and $\chi^2 \rightarrow \text{normal}$ —**large-sample and small-sample methods converge**

Coming Up Next

Weeks 3–6: Point Estimation. How do we find the “best” estimator? Properties: unbiasedness, efficiency, sufficiency, consistency.

- Wackerly, Mendenhall & Scheaffer, Ch. 7: *Sampling Distributions and the Central Limit Theorem*
- Hogg, McKean & Craig, Ch. 5: *Some Special Distributions* (Sections on χ^2 , t , F)
- Hogg, Tanis & Zimmerman, Ch. 5–6: *Distributions of Functions of Random Variables*
- Casella & Berger, Ch. 5: *Properties of a Random Sample* (Order statistics)
- Devore & Berk, Ch. 6: *Statistics and Sampling Distributions*