Lecture 12

Part 3 Estimation and Hypothesis Test

Estimation

I Have a Problem...

- I ask myself at night, "Am I smarter than the average person?"
- Step 1: Establish a variable of interest, such as IQ.
- Step 2: Compare my IQ to a benchmark (population mean IQ).

I Have a Problem...

- Randomly select a representative sample from the population.
- Compare my IQ to the mean IQ of the sample.
- So we are using a sample statistic (sample mean IQ) to estimate a population parameter (population mean IQ).

Sample Statistics as Estimators

- What did we see last topic?
- On average, the sample mean is approximately equal to the population mean.
  - The expected value of the sample mean was equal to the population mean, i.e., $E(\bar{X}) = \mu$.
- As the sample size increased, the sample mean was much closer to the population mean.
  - The variance of the sample mean decreased as the sample size increased, i.e., $V(\bar{X}) = \frac{\sigma^2}{n}$.

Two Types of Estimators

- **Point estimator:** Draws inferences about a population by using a single value, calculated from a sample, to estimate an unknown population parameter.
- **Interval estimator:** Draws inferences about a population by using an interval or range of values,calculated from a sample, to estimate an unknown population parameter.

Point Estimators

- We have already seen examples of point estimators, e.g., $\bar{X}$ for μ and s^2 for σ^2.
- But there are many different sample statistics that we could use to estimate any particular population parameter.
- For example, we could also use the *sample median* to estimate μ.
- Given a population parameter, how can we choose which sample statistic to use as an estimator?

Bias of a Point Estimator

- Let θ be some population parameter and let $\hat{\theta}$ denote a point estimator of θ.
- The **bias** of a point estimator is defined to be:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- A point estimator is **unbiased** if $B(\hat{\theta}) = 0$, i.e., if $E(\hat{\theta}) = \theta$.

Example of a Biased Point Estimator

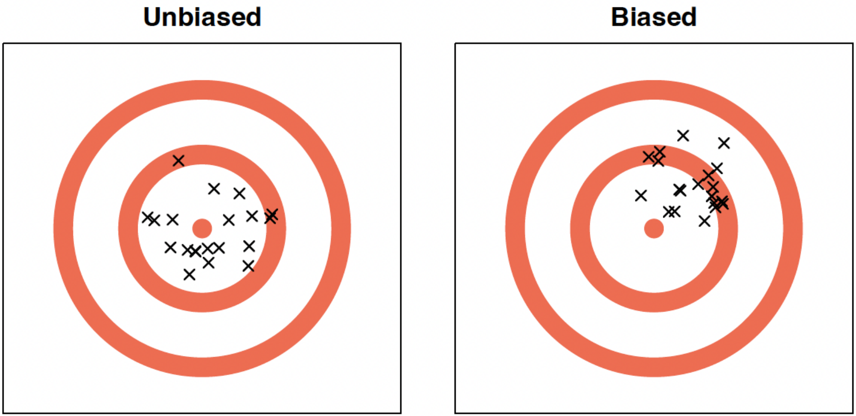
- Consider estimating the population variance σ^2 using the sample variance:

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

- This estimator is biased because:

$$E(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Bias of a Point Estimator



A Real-world Example

- Scenario: Estimating average fish length in a lake
- Biased method: Using a net with 10cm gaps between threads
- This method is biased because:

  Fish smaller than 10cm escape through the gaps

  Estimated average length will be higher than true average
- If true average length is 12cm:

  Biased estimate might be 15cm

  Bias = 15cm - 12cm = 3cm (overestimation)
- Unbiased alternative: Use a net with very small gaps or a variety of gap sizes

Bias of a Point Estimator

- Unbiasedness is a desirable quality of a point estimator.
- We know that $E(\bar{X}) = \mu$, so $\bar{X}$ is an unbiased estimator of μ.
- We have shown that $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator of σ^2.

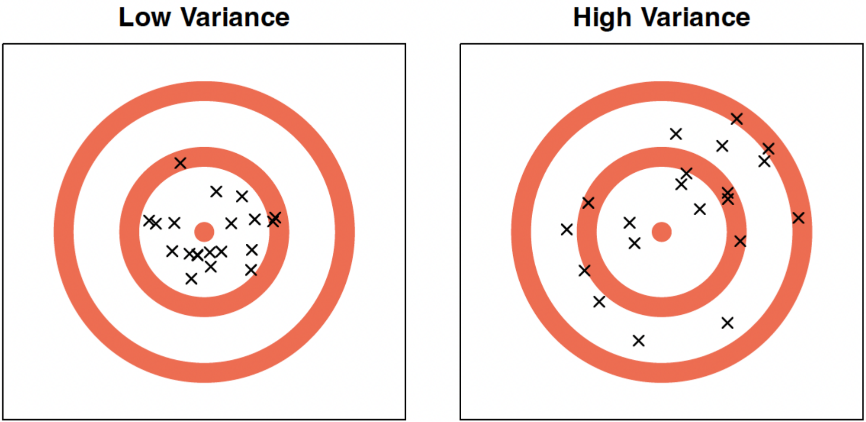
Variance of a Point Estimator

- If $\hat{\theta}$ is a point estimator of θ, the **variance** of $\hat{\theta}$ is:

$$
\begin{aligned}
V(\hat{\theta}) &= E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right) \\
&= E(\hat{\theta}^2) - \left(E(\hat{\theta})\right)^2
\end{aligned}
$$

- We would like our estimators to have low variance.

Variance of a Point Estimator



Low vs High Variance Estimators

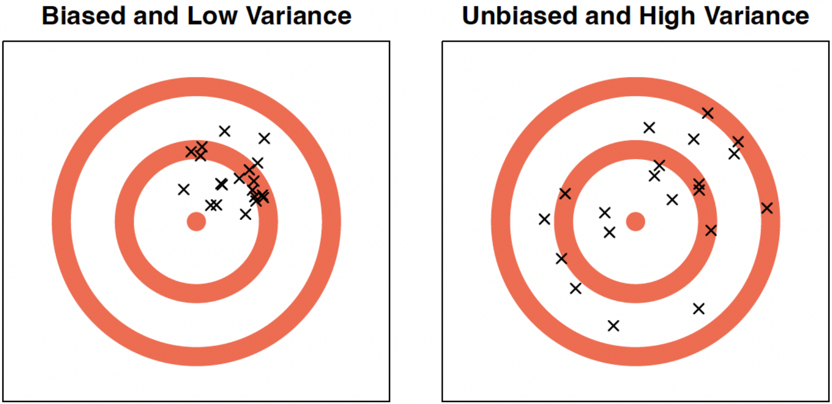
- If you want to estimate average daily sales in a coffee shop
- Low Variance Estimator: Average sales over 30 days

  Example results: \$985, \$1010, \$995 (multiple samples)
  Low spread around true value
- High Variance Estimator: Sales from a single randomly chosen day

  Example results: \$1200, \$800, \$1100 (multiple samples)
  High spread around true value
- Implications:

  Low variance: More reliable estimates
  High variance: Less reliable, more prone to extreme values

Bias and Variance



Trade-off

- There is often a trade-off between minimizing bias and minimizing variance. Why the trade-off?
  - Reducing bias often increases complexity of the estimator
  - More complex estimators tend to have higher variance

Mean Squared Error of a Point Estimator

- If $\hat{\theta}$ is a point estimator of θ, the **mean squared error** of $\hat{\theta}$ is defined to be:

$$MSE(\hat{\theta}) = E\left(\left(\hat{\theta} - \theta\right)^2\right)$$
$$= V(\hat{\theta}) + \left(B(\hat{\theta})\right)^2$$

- MSE can be useful for comparing point estimators.
- Mean Squared Error (MSE) $=$ Variance $+$ Bias^2

Mean Squared Error: Understanding the Bias-Variance Trade-off

- MSE can be decomposed into two components:

$$\text{MSE} = \text{Bias}^2 + \text{Variance}$$

- **Bias**: How far off our predictions are from the true value.
  High bias: Underfitting (oversimplified model)
  Low bias: Captures underlying pattern well
- **Variance**: How much our predictions vary for different samples.
  High variance: Overfitting (too sensitive to noise in data)
  Low variance: More stable predictions across samples

Mean Squared Error: Understanding the Bias-Variance Trade-off

- **The Trade-off**:

  As model complexity increases: Bias $\downarrow$, Variance $\uparrow$
  As model complexity decreases: Bias $\uparrow$, Variance $\downarrow$
- Goal: Find the sweet spot that minimizes total error (MSE)

Bias-Variance Trade-off: House Price Prediction

- Predicting house prices based on square footage
- Data(our sample): 100 houses with their prices and square footage
- We use two ways to estimate
  - Linear: Price $=$ a $\times$ (Square Footage) $+$ b
  - Quadratic: Price $=$ a $\times$ (Square Footage)^2 $+$ b $\times$ (Square Footage) $+$ c

Bias-Variance Trade-off: House Price Prediction

- **Results:**

Model	Bias	Variance	MSE
Linear	High	Low	10,000
Quadratic	Low	High	8,000

- **Interpretation:**
  - Linear: Underfits (high bias), but consistent (low variance)
  - Quadratic: Balances bias and variance, lowest overall error
- The quadratic model finds the "sweet spot" in the bias-variance trade-off for this dataset.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\text{Price}_i - \hat{\text{Price}}_i)^2, \quad n = 100 \text{ here}$$

Why We Need MSE: For Consistency!

- An estimator is said to be *consistent* if it approaches (i.e., gets closer to) the population parameter as the sample size increases.
- In other words...
- We can use the mean squared error to measure closeness.
- Let $\hat{\theta}$ be a point estimator of θ. If $MSE(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a **consistent** estimator of θ.
- Now you should be able to tell me what is a point estimator: Mean, Median, Mode, Variance?

Consistency: Examples

- Is $\bar{X}$ a consistent estimator of μ? Yes!
    - We know that $E(\bar{X}) = \mu$ and $V(\bar{X}) = \frac{\sigma^2}{n}$.
    $\therefore MSE(\bar{X}) = \frac{\sigma^2}{n} + 0^2 \to 0$ as $n \to \infty$

Recall the Sample Proportion We Learned

- $X = \sum_{i=1}^{n} X_i$, where

$$X_i = \begin{cases} 1 & \text{person } i \text{ prefers Coke over Pepsi} \\ 0 & \text{otherwise} \end{cases}$$

- So $\hat{p}$ is just **the sample mean** of a sample of independent Bernoulli random variables:

$$\hat{p} = \frac{X}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Consistency: Examples

- Is the sample proportion, $\hat{p} = \frac{X}{n}$, a consistent estimator of p? Yes!
  - Previously we showed that $E(\hat{p}) = p$ and $V(\hat{p}) = \frac{p(1-p)}{n}$.
    $\therefore MSE(\hat{p}) = \frac{p(1-p)}{n} + 0^2 \to 0$ as $n \to \infty$

Is sample variance a consistent estimator for population variance?

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

- We have proved unbiasedness of s^2.
- And we just proved, as $n \to \infty$:
  - $\bar{X} \to \mu$ (sample mean converges to population mean)

  Recall that

$$\text{MSE} = \text{Bias}^2 + \text{Variance}$$

- For this question, we need to calculate the variance of sample variance!

Variance of the Sample Variance

The variance of S^2, by the shortcut formula $V(X) = \mathbb{E}(X^2) - \mathbb{E}(X^2)$, is:

$$V(S^2) = \mathbb{E}[(S^2)^2] - (\mathbb{E}[S^2])^2 = \mathbb{E}[(S^2)^2] - \sigma^4$$

After expansion and using properties of expectations:

$$\mathsf{Var}(S^2) = \frac{\mathbb{E}(X_i^4)}{n} - \frac{\sigma^4(n-3)}{n(n-1)} \to 0, \text{ as } n \to \infty$$

Relative Efficiency

- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two *unbiased* point estimators of θ. The **relative efficiency** of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is defined to be:

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

- The unbiased estimator with the smaller variance is said to be **relatively more efficient**.

Relative Efficiency

- For example, for a normal distribution, it can be shown that the sample median has expected value equal to μ and variance equal to $\frac{1.5708 \times \sigma^2}{n}$.

  $\therefore \text{eff}(\bar{X}, \text{Med}) = \frac{V(\text{Med})}{V(\bar{X})} = \frac{\frac{1.5708 \times \sigma^2}{n}}{\frac{\sigma^2}{n}} = 1.5708$
- Since $\text{eff}(\bar{X}, \text{Med}) > 1$, i.e., $V(\bar{X}) < V(\text{Med})$, $\bar{X}$ is relatively more efficient than the sample median for estimating μ in a normal distribution.

Interval Estimators

- Why use interval estimators?
  - Point estimators will almost always be wrong.
  - Difficult to tell how close a point estimator is to the parameter.
  - Point estimators do not reflect the effects of larger sample sizes.
- How do we construct an interval estimator?
  - Recall that sampling distributions gave us the distribution of an estimator (sample statistic).
  - We will use *probabilities* derived from the sampling distribution of the estimator to construct an interval estimator.

Estimating μ (σ^2 Known)

- To construct an interval estimator for μ based on $\bar{X}$ (when the population variance is known) we will use its sampling distribution via the Central Limit Theorem.
- Specifically, we know that the sample mean follows a normal distribution for large sample sizes.
- We will use the z-tables to determine the associated probabilities.

Central Limit Theorem

- From the CLT, we know that for large n:

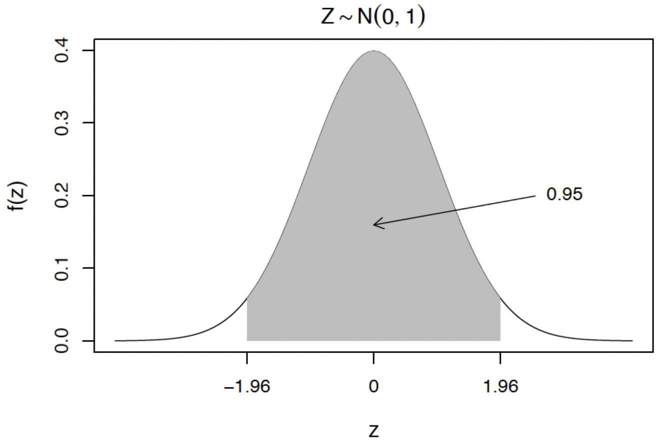
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- And if we standardize, we get:

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

Standard Normal Distribution

$$P(-1.96 < Z < 1.96) = 0.95$$



Standard Normal Distribution

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

Interval Estimator for μ

$$P(-1.96 < Z < 1.96) = 0.95$$

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right) = 0.95$$

$$P\left(-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(-1.96\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < 1.96\frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 0.95$$

$$P\left(1.96\frac{\sigma}{\sqrt{n}} + \bar{X} > \mu > -1.96\frac{\sigma}{\sqrt{n}} + \bar{X}\right) = 0.95$$

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

95% Confidence Interval for μ

- So our interval estimator for μ when σ^2 is known is given by:

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

- This is called a 95% **confidence interval** for μ.
- What this means: In repeated sampling, 95% of the intervals created in this way would contain μ and 5% would not.

95% Confidence Interval for μ

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- What made this a 95% confidence interval?

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$