Estimating Proportions

MGMT 230: Introductory Statistics

Goals of this class meeting

- Learn about estimating proportions from sample data.
- Learn how to construct confidence intervals and conduct hypothesis tests with proportions.

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Proportions

- **Proportion:** Percentage of times some characteristic occurs.
  - Example: percentage of self-identified Democrats who support Hillary Clinton for president.
- Notation:
  - π: population proportion.
  - p: sample proportion.
- Sample proportion:

$$p = \frac{X}{n} = \frac{\text{Number of items in sample having characteristic}}{\text{sample size}}$$

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Sampling Distribution

- Sample size must be sufficiently large.
- Population is *not normally distributed*.
- Central Limit Theorem:
  - Mean of the sampling distribution of p will equal π.
  - Standard deviation of the sampling distribution will be,

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

  - Sampling distribution of p will be normal.

Practical Uses

- The standard deviation requires knowledge of π.
- Use p instead. Use t instead of z.
- Book's suggestion: Use p instead and go ahead and use z as long as $np > 5$ and $n(1-p) > 5$.
- Conservative estimate: suppose $\pi = 0.5$. Use z.

Practical Uses

- The standard deviation requires knowledge of π.
- Use p instead. Use t instead of z.
- Book's suggestion: Use p instead and go ahead and use z as long as $np > 5$ and $n(1-p) > 5$.
- Conservative estimate: suppose $\pi = 0.5$. Use z.

Practical Uses

- The standard deviation requires knowledge of π.
- Use p instead. Use t instead of z.
- Book's suggestion: Use p instead and go ahead and use z as long as $np > 5$ and $n(1-p) > 5$.
- Conservative estimate: suppose $\pi = 0.5$. Use z.

Practical Uses

- The standard deviation requires knowledge of π.
- Use p instead. Use t instead of z.
- Book's suggestion: Use p instead and go ahead and use z as long as $np > 5$ and $n(1-p) > 5$.
- Conservative estimate: suppose $\pi = 0.5$. Use z.

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2, n-1}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Confidence Intervals

- Confidence interval for π:

$$(p - E,\ p + E)$$

- If using z, set $\pi = 0.5$, then,

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

- If using t,

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Example

Suppose in a sample of 50 people, 15 support Hillary Clinton. Construct a 95% confidence interval for the percentage that support Clinton.

- Using Z:

$$E = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} = 1.96\sqrt{\frac{0.25}{50}} = 0.139$$

$$(0.3 - 0.139, 0.3 + 0.139) = (0.161, 0.439)$$

- Using T: $t_{\alpha/2,n-1} = 2.0096$

$$E = t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}} = 2.0096\sqrt{\frac{(0.3)(0.7)}{50}} = 0.130$$

$$(0.3 - 0.130, 0.3 + 0.130) = (0.170, 0.430)$$

Hypothesis testing

- Always use a z when doing a hypothesis test.
- Use null hypothesis value of π.
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Hypothesis testing

- Always use a z when doing a hypothesis test.
- Use null hypothesis value of π.
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Hypothesis testing

- Always use a z when doing a hypothesis test.
- Use null hypothesis value of π.
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject H_0*.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject* H_0.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject* H_0.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject* H_0.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject* H_0.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Example

Suppose you want to estimate what percentage of the democrat population support Barack Obama. You take a sample of 30 democrats and find that 9 support Obama. Test the hypothesis with $\alpha = 0.05$ that Obama has less than 40% of the vote.

- $H_0: \ \pi = 0.4$, $H_a: \ \pi < 0.4$
- Critical Region: $z_{critical} = -1.645$, reject when $z < -1.645$
- Test statistic:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{30}}} = -1.118$$

- Test statistic is not in the critical region, *fail to reject* H_0.
- There is insufficient statistical evidence to claim that Barack Obama will get less than 40% of the vote.

Desired Sample Sizes

- If you want a margin of error no larger than E, for $(1-\alpha)$ confidence level.

$$n = \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2}\right)$$

- Use conservative estimate $\pi = 0.5$.

Desired Sample Sizes

- If you want a margin of error no larger than E, for $(1-\alpha)$ confidence level.

$$n = \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2} \right)$$

- Use conservative estimate $\pi = 0.5$.

Desired Sample Sizes

- If you want a margin of error no larger than E, for $(1-\alpha)$ confidence level.

$$n = \left(\frac{z_{\alpha/2}^{2} \pi (1-\pi)}{E^{2}} \right)$$

- Use conservative estimate $\pi = 0.5$.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z^2_{\alpha/2}\pi(1-\pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5\,(1-0.5)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z_{\alpha/2}^2 \pi (1 - \pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5 \left(1 - 0.5\right)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z_{\alpha/2}^2 \pi (1 - \pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5 \left(1 - 0.5\right)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z^2_{\alpha/2} \pi (1 - \pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5 \left(1 - 0.5\right)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2}\right)$$

$$n = \left(\frac{1.96^2 0.5\,(1-0.5)}{0.02^2}\right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5\,(1-0.5)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Example

- Suppose you want to estimate what percentage of voters support Clinton, within a 2% accuracy with a 95% level of confidence. What sample size must you take.

$$n = \left(\frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2} \right)$$

$$n = \left(\frac{1.96^2 0.5\,(1-0.5)}{0.02^2} \right)$$

$$n = 2401$$

- Must collect at least 2401 observations.
- Seems small compared to the rest of the country / state.
- Better make sure you have a *random sample* of the country.

Homework

- Inferences about proportions: Section 7.5 problems 7.30 through 7.33.
- Confidence intervals: Section 8.3 problems 8.26 through 8.30.
- Hypothesis tests: Section 9.5 problems 9.69 through 9.72.