Module 1 – Section 4

Hypothesis Tests for a Population Proportion

Outline

- Binomial Exact Test for p
- Score Test for p
- Hypothesis Testing Errors and the Rejection Regions of a Hypothesis Test for p
- Power and Sample Size Calculations

Binomial Random Variables

- Random event with 2 outcomes
- Outcomes
  - Success = Category of Interest
  - Failure = Not in Category of Interest
- Probabilities
  - Success = p
  - Failure = $1 - p$

Binomial Random Variables

- Y = number of successes in n independent and identical trials of random event
- Independent – outcome on one trial does not affect outcomes on other trials
- Identical – same probability of success

Binomial Exact Test for p

- Value of p is generally unknown
- Conduct hypothesis test for value of p
- Null Hypothesis
  - $H_0: p = p_0$
- Alternative Hypothesis
  - $H_A: p < p_0$ or $H_A: p > p_0$

Binomial Exact Test for p

- If H_0 is true, Y follows a binomial distribution with number of trials (sample size) n and probability of success p_0.
- Test Statistic
  - Observed value of Y: denoted y

Binomial Exact Test for p

- p-value
  - Obtained from binomial distribution with $p = p_0$
- $H_A: p < p_0$
  - p-value = $P(Y \leq y | p = p_0)$
- $H_A: p > p_0$
  - p-value = $P(Y \geq y | p = p_0)$

Ex. ESP

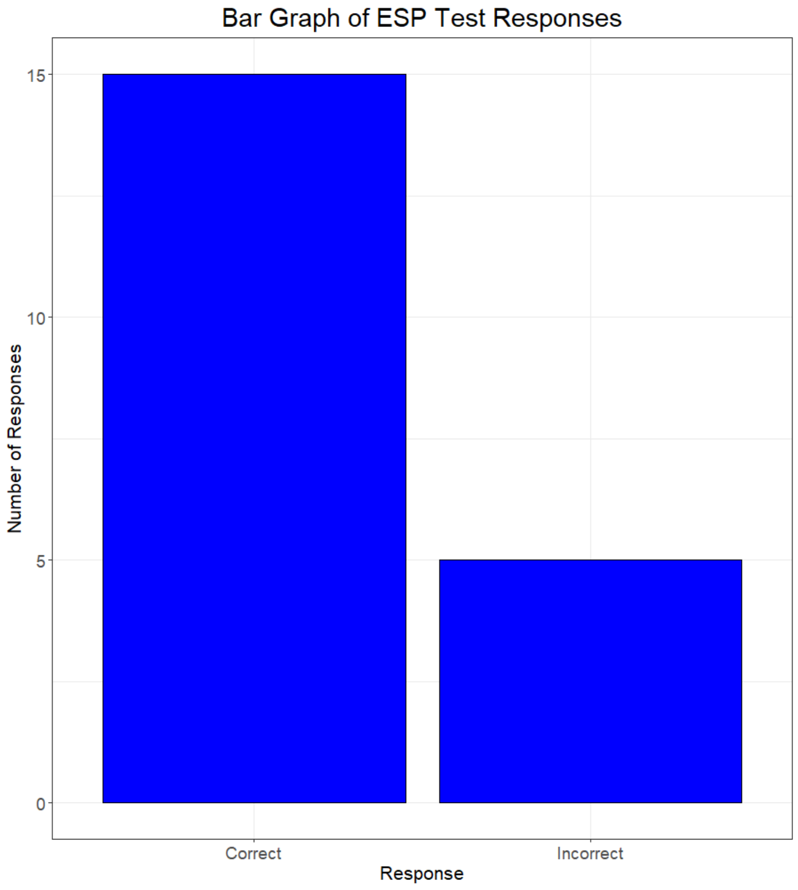
- A person claims to have ESP. You decide to test their claim by having the person guess the result on 20 flips of a coin. The person correctly guesses on 15 out of the 20 flips. Does this person's claim hold?

Ex. ESP – Data

Response
Correct
Correct
Correct
⋮
⋮
Incorrect
Incorrect

Ex. ESP – Summary Data

Response	Count	Proportion
Correct	15	0.75
Incorrect	5	0.25
Total	20	1.00



Ex. ESP

- Null Hypothesis
  - Person does not have ESP
  - $H_0: p = 0.5$
- Alternative Hypothesis
  - Person has ESP
  - $H_A: p > 0.5$

Ex. ESP

- Test statistic: $y = 15$
- p-value = $P(Y \geq 15|p = 0.5) = 0.0207$
- Conclusion: We have moderately strong evidence to support the person's claim of having ESP.

Two-Sided Alternative Hypothesis

- What if $H_A: p \neq p_0$?
- For some other two-sided alternative hypotheses, test statistic has symmetric distribution (ex. z or t).
  - Double the p-value from one-sided hypothesis
- Except for $p = 0.5$, binomial distribution is not symmetric.

Two-Sided Alternative Hypothesis

- Methods for finding the p-value
  - Double the one-sided p-value
  - Use equal distance from expected value
  - Method of small p-values

Ex. Balanced Die

- A die is rolled 60 times, and the number of times a 6 appears is 6. Is there evidence the die is unbalanced?

Ex. Balanced Die

- Null Hypothesis
  - Die is balanced
  - $H_0: p = \frac{1}{6}$
- Alternative Hypothesis
  - Die is unbalanced
  - $H_A: p \neq \frac{1}{6}$

Ex. Balanced Die

- Double the one-sided p-value
  - Observed value: $y = 6$
  - Expected value: $E(Y) = np = 60 * \frac{1}{6} = 10$
  - One-sided p-value = $P\left(Y \leq 6 | p = \frac{1}{6}\right) = 0.1081$
  - Overall p-value = $2 * 0.1081 = 0.2162$

Ex. Balanced Die

- Use equal distance from expected value
  - Observed value: $y = 6$
  - Expected value: $E(Y) = np = 60 * \frac{1}{6} = 10$
  - $y = 14$ is equidistant to 10
  - p-value

$$P\left(Y \leq 6 | p = \frac{1}{6}\right) + P\left(Y \geq 14 | p = \frac{1}{6}\right) = 0.2232$$

Ex. Balanced Die

- Method of small p-values
  - $P\left(Y = 6 | p = \frac{1}{6}\right) = 0.05686$
  - Find y such that $P\left(Y = y | p = \frac{1}{6}\right) < 0.05686$
    - $P\left(Y = 13 \middle| p = \frac{1}{6}\right) = 0.0751$ $\quad$ $P\left(Y = 14 \middle| p = \frac{1}{6}\right) = 0.0504$
  - p-value

$$P\left(Y \leq 6 | p = \frac{1}{6}\right) + P\left(Y \geq 14 | p = \frac{1}{6}\right) = 0.2232$$

Ex. Balanced Die

- p-values
  - Double the one-sided p-value = 0.2162
  - Use equal distance from expected value = 0.2232
  - Method of small p-values = 0.2232
- Conclusion: We do not have sufficient evidence to conclude the probability of obtaining a six on this die is different from 1/6.

Binomial Exact Test for p

- Historically we used this test for small sample sizes n only.
  - Calculations for binomial distribution probabilities were difficult and time consuming.
- Today we could use this test for all sample sizes.
- Convention says we still use this test for small sample sizes n only – where np and $n(1-p) \geq 10$ fails.

Sampling Distribution for $\hat{p}$

- Use $\hat{p} = \frac{Y}{n}$ as estimate for p.
- If $np \geq 10$ and $n(1-p) \geq 10$:

$$\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \text{ or } \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Score Test for p

- Value of p is generally unknown
- Conduct hypothesis test for value of p
- Null Hypothesis
  - $H_o: p = p_o$
- Alternative Hypothesis
  - $H_A: p \neq p_o$ or $H_A: p < p_o$ or $H_A: p > p_o$

Score Test for p

- Assume H_0 is true and $np_0 \geq 10$ and $n(1-p_0) \geq 10$
- Define test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- Distribution of test statistic z should be approximately equal to $N(0,1)$.

Score Test for p

- p-value
  - Obtained from the standard normal distribution (Z).
- $H_A: p < p_0$
  - p-value = $P(Z < z)$
- $H_A: p > p_0$
  - p-value = $P(Z > z)$
- $H_A: p \neq p_0$
  - p-value = $2 * P(Z > |z|)$

Ex. Ingots

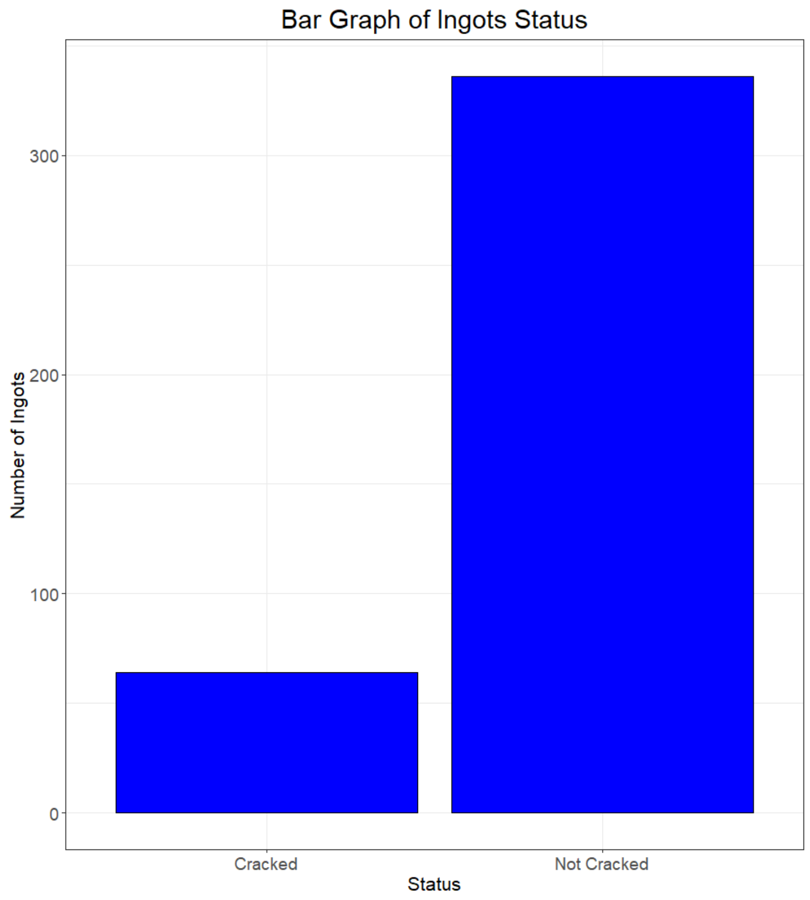
- The cracking rate of ingots used in manufacturing airplanes is 20%. A new process is designed to lower the proportion of cracked ingots. In a sample of 400 ingots, 64 of them were cracked. Did the new process actually lower the proportion of cracked ingots? Use $\alpha = 0.05$

Ex. Ingots – Data

Status
Cracked
Cracked
Cracked
⋮
⋮
Not Cracked
Not Cracked

Ex. Ingots – Summary Data

Status	Count	Proportion
Cracked	64	0.16
Not Cracked	336	0.84
Total	400	1.00



Ex. Ingots

- Null Hypothesis
  - Cracking rate is unchanged
  - $H_0: p = 0.2$
- Alternative Hypothesis
  - Cracking rate has decreased
  - $H_A: p < 0.2$

Ex. Ingots

- Check Assumption
  - $np = 400(0.2) = 80$
  - $n(1-p) = 400(0.8) = 320$
- Test Statistic

$$z = \frac{0.16 - 0.2}{\sqrt{\frac{0.2(0.8)}{400}}} = -2$$

Ex. Ingots

- p-value $= P(Z < -2) = 0.0228$
- Conclusion: We have moderately strong evidence the cracking rate of the ingots has decreased to less than 20%.

Hypothesis Testing Errors

- Two Possible Truths
  - H_0 is true
    - We do not want to reject it
  - H_0 is false
    - We want to reject it
- "The Truth" is unknown.

Hypothesis Testing Errors

	Decision	
"The Truth"	Reject H_0	Do not reject H_0
H_0 is true	Type I error	Correct Decision
H_0 is false	Correct Decision	Type II error

Hypothesis Testing Errors

- Two Errors
  - Type I Error = rejecting H_0 given H_0 is true
  - Type II Error = failing to reject H_0 given H_0 is false
- Since "The Truth" is unknown, we do not know if we committed one of these errors.

Hypothesis Testing Errors

- Probability of Type I error
  - This is α
  - Control for probability of Type I error
- Probability of Type II error
  - This is called β
  - Want β to be small

Hypothesis Testing Errors

- Inverse Relationship between probabilities of Type I and Type II errors
- All other things being equal,
  - If $\alpha \uparrow$, $\beta \downarrow$
  - If $\alpha \downarrow$, $\beta \uparrow$

Rejection Region

- Rejection Region of Hypothesis Test
  - Values of the test statistic where you will reject H_0
- Determined based on
  - Test Statistic
  - Alternative Hypothesis
  - probability of Type I Error

Binomial Exact Test for p

Alternative Hypothesis	Rejection Region
$H_A: p < p_0$	Find y so that $P(Y \leq y) \leq \alpha$
$H_A: p > p_0$	Find y so that $P(Y \geq y) \leq \alpha$
$H_A: p \neq p_0$	Find y_1 and y_2 so that $P(Y \leq y_1) + P(Y \geq y_2) \leq \alpha$

Ex. ESP

y	$P(Y \geq y \mid p = 0.5)$
16	0.0059
15	0.0207
14	0.0577
13	0.1316

- Rejection Regions
  - $Y \geq 16$ if $\alpha = 0.01$
  - $Y \geq 15$ if $\alpha = 0.05$
  - $Y \geq 14$ if $\alpha = 0.1$

Score Test for p

Alternative Hypothesis	Rejection Region
$H_A: p < p_0$	For test statistic $z < -z_{1-\alpha}$
$H_A: p > p_0$	For test statistic $z > z_{1-\alpha}$
$H_A: p \neq p_0$	For test statistic $z < -z_{1-\frac{\alpha}{2}}$ or $z > z_{1-\frac{\alpha}{2}}$

Ex. Type I Error = 0.05

Alternative Hypothesis	Rejection Region
$H_A\!: p < p_0$	For test statistic $z < -1.645$
$H_A\!: p > p_0$	For test statistic $z > 1.645$
$H_A\!: p \neq p_0$	For test statistic $z < -1.96 \text{ or } z > 1.96$

Observed Type 1 Error Rate and α

- Probability of obtaining a test statistic in the rejection region when the null hypothesis is true.
- For many applications, this probability is extremely close to α.
- Hypothesis tests for population proportions do not always have this property.

Observed Type 1 Error Rate and α

- Binomial Exact Test
  - Will be no more than the stated α level, but will usually be lower.
  - Discrete nature of binomial distribution.
- Score Test
  - Could be more or less than the stated α level.
  - Continuous approximation of a discrete distribution.

Ex. ESP

- For $\alpha = 0.05$, rejection region is $Y \geq 15$.
- Observed Type I Error Rate = Probability of obtaining 15 or more successes when the probability of success is 0.5.

$$P(Y \geq 15 | p = 0.5) = 0.0207$$

Ex. Ingots

- For $\alpha = 0.05$, rejection region is $z < -1.645$.
- Observed Type I Error Rate = Probability of obtaining a test statistic z of -1.645 or smaller when $p = 0.2$

$$
\begin{aligned}
P(z < -1.645 | p = 0.2) &= P\left(\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2\,(0.8)}{400}}} < -1.645\right) \\
= P(\hat{p} < 0.1671) &= P\left(\frac{Y}{400} < 0.1671\right) \\
= P(Y < 66.85) &= P(Y \leq 66) = 0.0433
\end{aligned}
$$

Power of Hypothesis Test

- Power = probability of (correctly) rejecting H_0 when H_0 is false
  - $P(\text{reject } H_0 | H_0 \text{ is false})$
- Power is complement of probability of Type II error.
  - Power = $1 - \beta$.
- Want power to be large

Power of Hypothesis Test

- Helpful in understanding potential research results
  - Data collection not necessary for calculations
- Factors affecting power
  - Test Statistic
  - Type I error rate = α
  - Sample size n
  - Difference between true value of $p = p_a$ and p_0

Binomial Exact Test

$$P(\text{reject } H_0 | H_0 \text{ is false}) = P(\text{reject } H_0 | p = p_a)$$

Alternative Hypothesis	Power
$H_A: p < p_0$	$P(Y \leq y \mid p = p_a)$
$H_A: p > p_0$	$P(Y \geq y \mid p = p_a)$
$H_A: p \neq p_0$	$P(Y \leq y_1 \mid p = p_a) + P(Y \geq y_2 \mid p = p_a)$

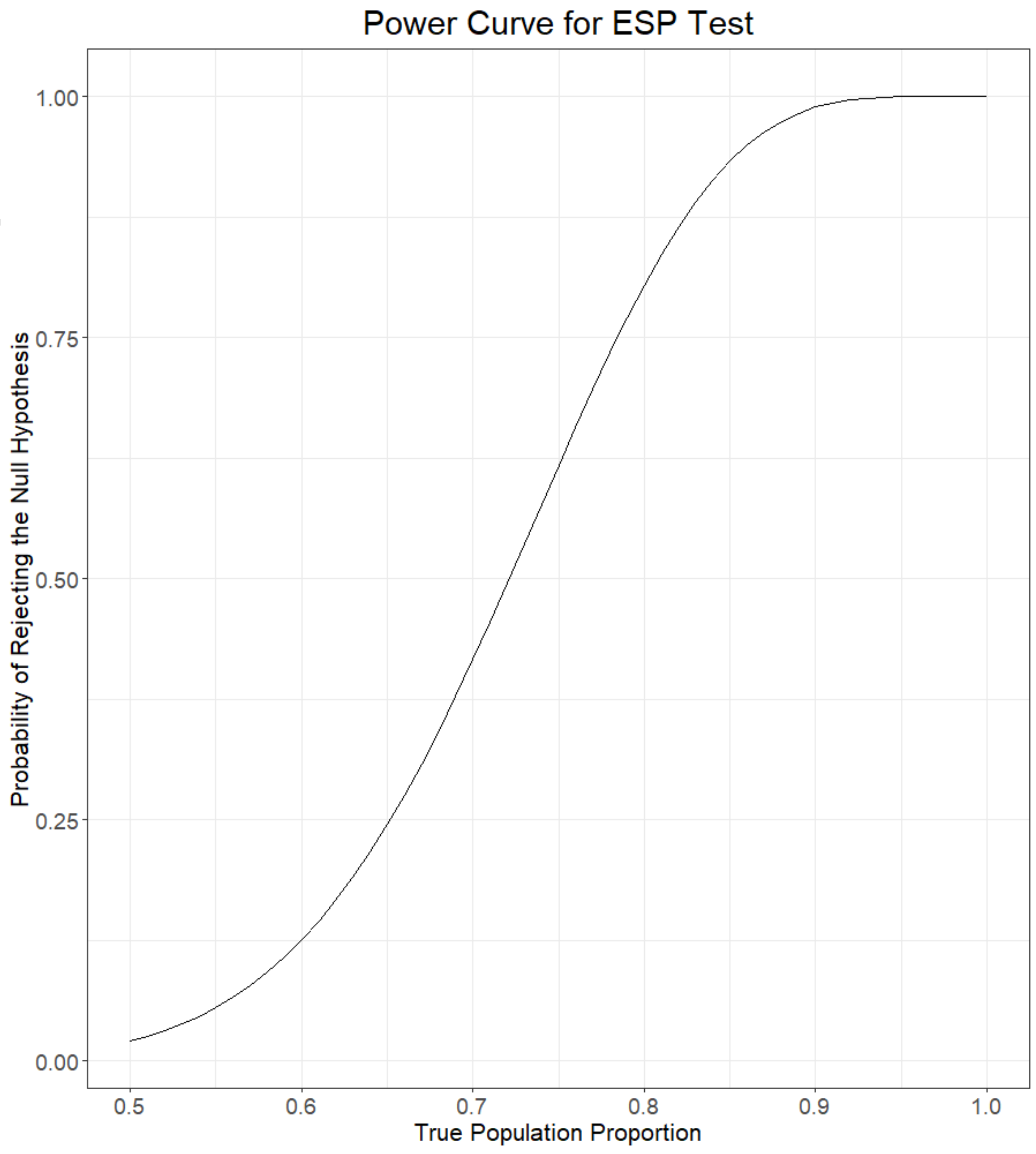
The values of y, y_1, y_2 come from rejection regions.

Ex. ESP

- Let $\alpha = 0.05$
- Rejection region is $Y \geq 15$
- Power
  - $P(Y \geq 15|p = p_a = 0.6) = 0.1256$
  - $P(Y \geq 15|p = p_a = 0.75) = 0.6172$
  - $P(Y \geq 15|p = p_a = 0.9) = 0.9887$

Power Curve

- Value of p_a versus power for all values of p_a in the alternative hypothesis.
- The greater the distance between p_a and 0.5, the larger the power of the hypothesis test.



Score Test for p

- Power when $H_a: p < p_0$

$$P\left(Z < \frac{p_0 - p_a - z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}}{\sqrt{\frac{p_a(1-p_a)}{n}}}\right)$$

Score Test for p

- Power when $H_a\!: p > p_0$

$$P\left(Z > \frac{p_0 - p_a + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}}{\sqrt{\frac{p_a(1-p_a)}{n}}}\right)$$

Score Test for p

- Power when $H_a: p \neq p_0$

$$2 * P\left(Z > \frac{|p_0 - p_a| + z_{1-\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}}{\sqrt{\frac{p_a(1-p_a)}{n}}}\right)$$

Ex. Ingots

- Let $\alpha = 0.05$
- Rejection region: $Z < -1.645$
- Two values of p_a
  - $p_a = 0.14$
  - $p_a = 0.18$
- $n = 400$

Ex. Ingots

- Power when $p_a = 0.14$

$$\text{Power} \quad = \quad P\left(Z < \frac{0.2 - 0.14 - 1.645\sqrt{\frac{0.2\,(0.8)}{400}}}{\sqrt{\frac{0.14\,(0.86)}{400}}}\right)$$

$$= \quad P(Z < 1.56)$$

$$= \quad 0.9406$$

Ex. Ingots

- Power when $p_a = 0.18$

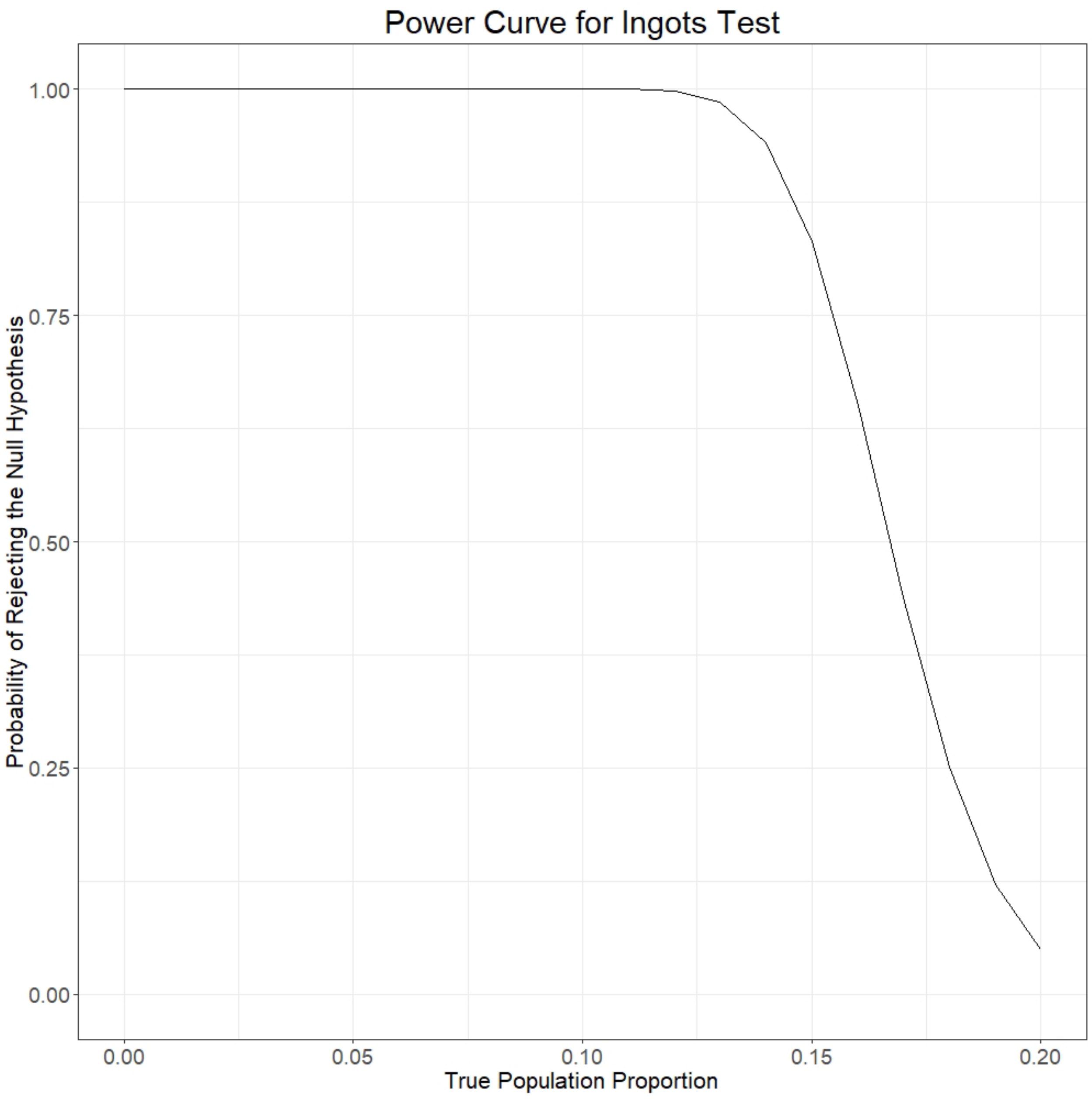
$$\text{Power} = P\left(Z < \frac{0.2 - 0.18 - 1.645\sqrt{\frac{0.2\,(0.8)}{400}}}{\sqrt{\frac{0.18\,(0.82)}{400}}} \right)$$

$$= P(Z < -0.67)$$

$$= 0.2514$$

Power Curve

- Value of p_a versus power for all values of p_a in the alternative hypothesis.
- The greater the distance between p_a and 0.2, the larger the power of the hypothesis test.



Sample Size Calculations

- Before conducting study, determine sample size based on
  - Type I error rate = α
  - Value of p_a
  - Desired Power $(1 - \beta)$ for given value of p_a

Sample Size Calculations

- Calculated for Score Test Only
  - Binomial Exact Test is used for inference for small sample sizes.
  - Assume planning is for large sample sizes

Sample Size Calculations

- $H_A: p < p_0$

$$n \geq \frac{\left[z_{1-\beta}\sqrt{p_a(1-p_a)} + z_{1-\alpha}\sqrt{p_0(1-p_0)}\right]^2}{(p_0 - p_a)^2}$$

Sample Size Calculations

- $H_A\colon p > p_0$

$$n \geq \frac{\left[z_{1-\beta}\sqrt{p_a(1-p_a)} + z_{1-\alpha}\sqrt{p_0(1-p_0)}\right]^2}{(p_0 - p_a)^2}$$

Sample Size Calculations

- $H_A: p \neq p_0$

$$n \geq \frac{\left[z_{(1-\beta)/2}\sqrt{p_a(1-p_a)} + z_{1-\alpha/2}\sqrt{p_0(1-p_0)}\right]^2}{(p_0 - p_a)^2}$$

Ex. Ingots

- Let $\alpha = 0.05$
  - $z_{1-\alpha} = 1.645$
- $p_a = 0.17$
- Power $= 1 - \beta = 0.8$
  - $z_{1-\beta} = 0.84$

Ex. Ingots

$$n \geq \frac{\left[0.84\sqrt{0.17\,(0.83)} + 1.645\sqrt{0.2\,(0.8)}\right]^2}{(0.2 - 0.17)^2}$$

$$= 1053.071$$

- Sample Size: $n = 1054$