Choosing Sample Size. Power of a Test.

A. Skripnikov[1]

[1]New College of Florida

IDC 5205

How big of a sample we need?

Question: Ever wondered how sample sizes are determined for exit polling? E.g. is $1,000$ people enough? 100? 745?

Criteria: **Precision** of the results.

Question: What's meant by precision when talking about statistical inference? (**Hint**: Thinks of "precise"/"imprecise" confidence interval.)

Margin of Error.

Question: What's the formula for margin of error in case of confidence intervals for:

- population proportion?

- population mean?

Both depend on

1. **Standard error**, which, in its turn, depends on **sample size**.
2. Confidence level $(1-\alpha)$

Sample Size for Proportion.

Hence, to determine

"How large of a sample size n we need to estimate a population proportion?"

one should specify

1. Desired **margin of error.**
2. Desired **confidence level**.

Example. We'd like to predict the proportion of votes for a candidate in the political election with

1. margin of error $m = 0.04$,
2. confidence level of 95% (hence, significance level $\alpha = 0.05$).

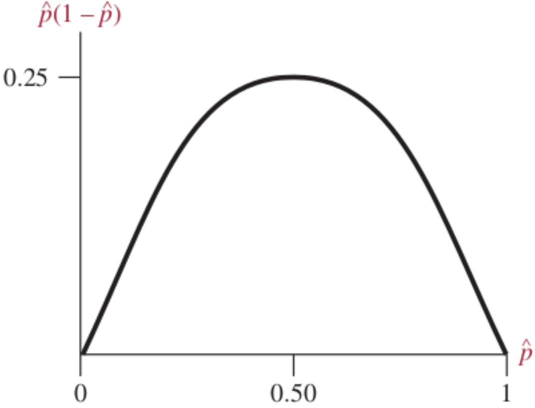
Question: What is the sample size needed?
Derived in class, using the formula for 95% **CI for proportion**:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sample Size for Proportion.

Derivation.

Selecting a Sample Size Without Guessing a Value for $\hat{p}$.



Sample Size for Proportion.

SUMMARY: Sample Size for Estimating a Population Proportion

The random sample size n for which a confidence interval for a population proportion p has margin of error m (such as $m = 0.04$) is

$$n = \frac{\hat{p}(1 - \hat{p})\, z^2}{m^2}.$$

The z-score is based on the confidence level, such as $z = 1.96$ for 95% confidence. You either guess the value you'd get for the sample proportion $\hat{p}$ based on other information or take the safe approach of setting $\hat{p} = 0.5$.

Sample Size for Proportion.

Example (cont'd). The election is expected to be close, hence margin of 0.04 isn't sufficient.

Question: What sample size would we need for

- Margin of error 0.02, at confidence level of 95%.

- Margin of error 0.01, at confidence level of 99%.

Sample Size for Mean.

Confidence Interval for Population Mean:

$$\bar{\mathbf{x}} \pm t_{n-1,[1-\alpha/2]} \frac{s}{\sqrt{n}}$$

Task: Given margin of error m and confidence level $1 - \alpha$ (significance level of α), derive the formula for sample size.

Sample Size for Mean.

Question: What's the most glaring issue with the formula below?

$$n = \frac{t^2_{n-1,[1-\alpha/2]}\ s^2}{m^2}$$

For simplicity, presuming that we'll need $n \geq 30$, we may write **confidence interval** as

$$\bar{\mathbf{x}} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

Q: What does the formula for sample size become now?

Sample Size for Mean.

SUMMARY: Sample Size for Estimating a Population Mean

The random sample size n for which a confidence interval for a population mean μ has margin of error approximately equal to m is

$$n = \frac{\sigma^2 z^2}{m^2}.$$

The z-score is based on the confidence level, such as $z = 1.96$ for 95% confidence. To use this formula, you need to guess the value for the population standard deviation σ.

Sample Size for Mean.

Question: How do we estimate σ?

Answer: **Very crudely**, by making assumptions about:

- data range,
- data distribution.

Task. How many **randomly sampled** Data Scientists should we ask about their salary to get a

1. $5k\$$ margin of error at 95% confidence level,
2. $10k\$$ margin of error at 99% confidence level.

Sample Size for Mean.

Task (cont'd).

1. $5k\$$ margin of error at 95% confidence level,

2. $10k\$$ margin of error at 99% confidence level.

Other Factors Affecting Choice of the Sample Size.

Four factors playing a role in determining study's sample size:

1. **Precision** (m, margin of error).
2. **Confidence level** ($1 - \alpha$).
3. **Variability** of data.

   Question: If data variability is high, do we need a small or large sample? Why?

4. **Cost**.

   Larger samples are more time consuming to collect. They may be more expensive than a study can afford.

   When planning a study, typical question to ask yourself is:

   "Should we go ahead with the smaller sample that we can afford, even though the margin of error will be greater than we would like?"

P(Type II error), Power of a Test.

Another important aspect of a statistical study is its **power**, defined as:

$$Power = 1 - P(\text{Type II error})$$

In significance testing, there're two types of errors we can commit:

- Type I error (**Explain.**)

- Type II error (**Explain.**)

P(Type II error), Power of a Test.

Example. $n = 116$ volunteers provide their

- horoscope info,
- personality chart (according to California Personality Index).

An astrologer is presented with:

- person's horoscope info,
- personality charts of *this person* + *two others* (randomly selected).

Astrologer has to guess which personality chart (out of three) matches this horoscope info.

P(Type II error), Power of a Test.

Task. Proceed to formulate a hypothesis test addressing the question:

"Are astrologers' predictions any different from a random guess?"

Specify what is meant by Type I/II errors in that particular context.

P(Type II error), Power of a Test.

Task (cont'd). If conducting hypothesis test at **significance level** α, then

- P(Type I error) $=$

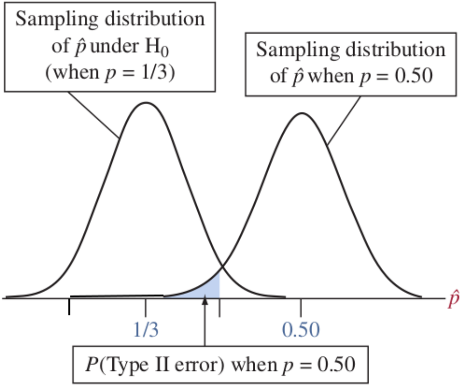
- P(Type II error) $=$

Note: In order to calculate P(Type II error), we need to **fix an alternative value**. E.g.

"Say, H_0 is false and p **is actually** $= 0.50$ **(fixed alternative value)**. What is P(Type II error) of our hypothesis test then?"

That's the only way of having a **well-defined sampling distribution of the statistic under H_a being true** (denoted as "$\hat{p} \mid H_a$")

P(Type II error), Power of a Test.

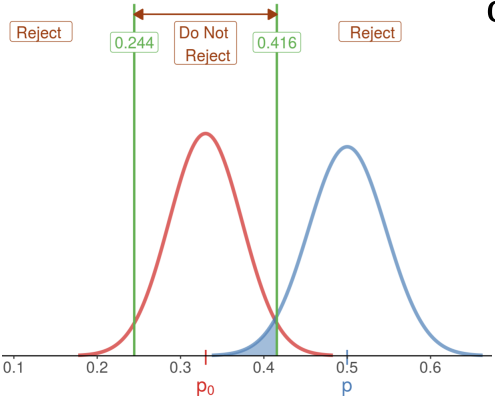


▲ **Figure 9.15 Sampling Distributions of $\hat{p}$ When H_0: $p = 1/3$ Is True and When $p = 0.50$. Question** Why does the shaded area under the left tail of the curve for the case $p = 0.50$ represent P(Type II error) for that value of p?

For better illustration, see `https://istats.shinyapps.io/power/`.

P(Type II error), Power of a Test.

Setting $n = 116$ **in** `https://istats.shinyapps.io/power/`:



Questions:

- What's the red curve? Blue curve?
- What are the green vertical lines?
- What's the blue **shaded** region?
- How is P(Type II error) calculated here?

What happens to P(Type II error) when...

Using the app `https://istats.shinyapps.io/power/`, proceed to check what happens to P(Type II error) whenever:

- Distance between true value p and hypothesized value p_0 increases,

- Sample size increases,

- Significance level α increases.

Practical Effect Size to Be Detected.

Before conducting the study, researchers should find P(Type II error) for **practical effect sizes they want to be able to detect**. E.g.

"Do we want to correctly detect if astrologists' predictive accuracy is..."

- at 0.40 (barely above $1/3$)?

  or

- at 0.50 (considerably above $1/3$)?

Power of a Test.

Power of a significance test is defined as follows

$$Power = 1 - P(\text{Type II error})$$

Indicates your test's ability to **correctly reject** H_0 if it is **actually false**.

Did You Know?

Before granting research support, many agencies (such as the National Institutes of Health) expect research scientists to show that for the planned study, reasonable power (usually, at least 0.80) exists at values of the parameter that are considered practically significant. ◀

P(Type II error), Power of a Test: Example.

Example (you can find a similar one in the book). New medical treatment has been developed to treat a particular condition. Old medical treatment for this condition has success rate of just 10%.

Statement about the proposed trial:

"At a 0.05 significance level, the proposed trial will have 0.82 power if the true success rate of alternative treatment is 25%."

Tasks:

1. Formulate the hypotheses.
2. Interpret what's meant by "0.82 power".

P(Type II error), Power of a Test: Example.

Example (you can find a similar one in the book). New medical treatment has been developed to treat a particular condition. Old medical treatment for this condition has success rate of just 10%.

Statement about the proposed trial:

"At a 0.05 significance level, the proposed trial will have 0.82 power if the true success rate of alternative treatment is 25%."

Tasks:

- In context, what is a Type II error for the proposed trial? What are its consequences?

- If $p = 0.25$, what is the probability of committing a Type II error?

P(Type II error), Power of a Test: Example.

Example (you can find a similar one in the book). New medical treatment has been developed to treat a particular condition. Old medical treatment for this condition has success rate of just 10%.

Statement about the proposed trial:

"At a 0.05 significance level, the proposed trial will have 0.82 power if the true success rate of alternative treatment is 25%."

Tasks:

- What can be done to lower the probability of a Type II error, i.e., to increase the power of the test?