Inference for A Single Proportion

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of OpenIntro and John McGreedy of Johns Hopkins University

Reminders

- Mini-project 2 is DUE today
- If you have not turned in a final project proposal yet you should ASAP

RECAP: Inference for **means**

Match the following scenarios to the appropriate test (Z-test, T-test, ANOVA):

- Comparing a sample mean to a population mean, N = 50
- Comparing a sample mean to a population mean, N = 20
- Comparing paired means, N = 100
- Comparing paired means, N = 25
- Comparing un-paired means, $N_a = 15$, $N_b = 17$
- Comparing un-paired means, $N_a = 55$, $N_b = 56$
- Comparing more than 2 means, $N_a = 55$, $N_b = 56$, $N_c = 54$, $N_d = 55$

Proportions

- We can make similar inferences for proportions instead of means
- Today we will start with one-sample proportions, compared to a population proportion
- This an analogous to testing whether a one-sample mean differs from a population mean

Motivating Example

The following survey was used to determine if American's have a good understanding of experimental design.

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

(a) All 1000 get the drug
(b) 500 get the drug, 500 don't

Motivating Example

The following survey was used to determine if American's have a good understanding of experimental design.

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

(a) All 1000 get the drug
(b) 500 get the drug, 500 don't

Results from the GSS

Below is the distribution of responses from the 2010 survey:

All 1000 get the drug	99
500 get the drug 500 don't	571
Total	670

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don't”? What are the parameter of interest and the point estimate?

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don't”? What are the parameter of interest and the point estimate?

Parameter of interest: proportion of *all* Americans who have good intuition about experimental design.

p a population proportion

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don't”? What are the parameter of interest and the point estimate?

Parameter of interest: proportion of *all* Americans who have good intuition about experimental design.

p a population proportion

Point estimate: proportion of *sampled* Americans who have good intuition about experimental design.

$\hat{p}$ a sample proportion

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

We can answer this research question using a confidence interval, which we know is always of the form

point estimate ± ME

Inference on a proportion

What percent of all Americans have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

We can answer this research question using a confidence interval, which we know is always of the form

point estimate ± ME

And we also know that *ME = critical value times the SE* of the point estimate.

And:

$$SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$$

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

- But of course this is true only under certain conditions… remember what they are?

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

- But of course this is true only under certain conditions… remember what they are?

independent observations, at least 10 successes and 10 failures

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

- But of course this is true only under certain conditions… remember what they are?

independent observations, at least 10 successes and 10 failures

---

Note: If *p* is unknown (most cases), we use $\hat{p}$ in the calculation of the standard error.

Back to the example ...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design.

Given: $n = 670$, $\hat{p} = 0.85$.

Back to the example ...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design.

Given: $n = 670$, $\hat{p} = 0.85$.

First check conditions.

1. Independence
2. Success-Failures

Back to the example ...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design.

Given: $n = 670$, $\hat{p} = 0.85$.

First check conditions.

1. *Independence*: The sample is random, and 670 < 10% of all Americans, therefore we can assume that one respondent's response is independent of another.

Back to the example ...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Estimate (using a 95% confidence interval) the proportion of all Americans who have good intuition about experimental design.

Given: $n = 670$, $\hat{p} = 0.85$.

First check conditions.

1. *Independence*: The sample is random, and 670 < 10% of all Americans, therefore we can assume that one respondent's response is independent of another.
2. *Success-failure*: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 10.

Practice

We are given

$n = 670$, $\hat{p} = 0.85$

We know:

*CI = point estimate ± critical value * SE,* $SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Practice

We are given
$n = 670$, $\hat{p} = 0.85$
We know:
*CI = point estimate ± critical value * SE,* $SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$ → (0.82, 0.88)

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Practice

We are given

$n = 670$, $\hat{p} = 0.85$

We know:

*CI = point estimate ± critical value * SE,* $SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$ → (0.82, 0.88)

What is the interpretation of this interval?

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Practice

We are given

$n = 670$, $\hat{p} = 0.85$

We know:

*CI = point estimate ± critical value * SE,* $SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85\times0.15}{670}}$ → (0.82, 0.88)

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85\times0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85\times0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571\times99}{670}}$

What is the interpretation of this interval?

We are 95% confident that 82 – 88% of all Americans have good intuition about experimental design.

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^{\star} \times SE$$

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^{\star} \times SE$$

We want: $0.01 \geq ME$

$$0.01 \geq z * \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \text{Use estimate for } \hat{p} \text{ from previous study}$$

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^{\star} \times SE$$

$$0.01 \quad \geq \quad 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \textit{Use estimate for } \hat{p} \textit{ from previous study}$$

solve for n

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^\star \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \textit{Use estimate for } \hat{p} \textit{ from previous study}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$ solve for n

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^{\star} \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \text{Use estimate for } \hat{p} \text{ from previous study}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%?

$$ME = z^{\star} \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \textit{Use estimate for } \hat{p} \textit{ from previous study}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

$$n \geq 4898.04 \rightarrow n \textit{ should be at least } 4{,}899$$

What if there isn't a previous study?

... use $\hat{p} = 0.5$

why?

What if there isn't a previous study?

... use $\hat{p} = 0.5$

why?

- if you don't know any better, 50-50 is a good guess

What if there isn't a previous study?

... use $\hat{p}$ = 0.5

why?

- if you don't know any better, 50-50 is a good guess
- $\hat{p}$ = 0.5 gives the most conservative estimate -- highest possible sample size

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

What are the hypotheses?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What test should we use?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.

Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What test should we use?

- Z-test, N > 50

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What test should we use?

- Z-test, N > 50

Do we meet the conditions to use it?

Hypothesis Test for proportions

Success-failure condition:

- CI: At least 10 *observed* successes and failures
- HT: At least 10 *expected* successes and failures, calculated using the null value

Independence:

- Observations must be independent of each other

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What test should we use?

- Z-test, N > 50

Do we meet the conditions to use it?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What test should we use?

- Z-test, N > 50

Do we meet the conditions to use it? (At least 10 *expected* successes and failures, calculated using the null value)

- 0.8 * 670 = 536, 0.2 * 670 = 134 → Success-failure condition $\boxed{\checkmark}$
- Independent observations $\boxed{\checkmark}$

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.

Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score) $Z = \frac{obs - null}{SE}$
- Find p-value for that test statistic

CI vs. HT for proportions

Standard error:

- CI: calculate using observed sample proportion:

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

- HT: calculate using the null value:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.

Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$$Z = \frac{obs \; - null}{SE} \qquad SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.

Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score) $Z = \frac{obs\ - null}{SE}$ $\quad SE = \sqrt{\frac{p_0(1-p_0)}{n}}$
- Find p-value for that test statistic

$$Z = \frac{obs - null}{SE} = \frac{0.85 - .80}{SE}$$

$$= \frac{0.85 - .80}{0.0155} = 3.22$$

$$SE = \sqrt{\frac{p_0(1 \ - \ p_0)}{n}} = \sqrt{\frac{0.8\ (1 \ - 0.8)}{670}} = 0.0155$$

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$$Z = \frac{obs \; - null}{SE} \qquad SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$Z = \frac{obs - null}{SE} = \frac{0.85 - .80}{SE} \qquad SE = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.8\,(1 - 0.8)}{670}} = 0.0155$$

$$= \frac{0.85 - .80}{0.0155} = 3.22$$

What does this z-score tell us?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$$Z = \frac{obs \; - null}{SE} \qquad SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$Z = \frac{obs \; - null}{SE} = \frac{0.85 - .80}{SE} \qquad SE = \sqrt{\frac{p_0(1 \; - \; p_0)}{n}} = \sqrt{\frac{0.8\,(1 \; - 0.8)}{670}} = 0.0155$$

$$= \frac{0.85 - .80}{0.0155} = 3.22$$

What does this z-score tell us?

Our observation is over 3 SE’s away from the null

Hypothesis Testing

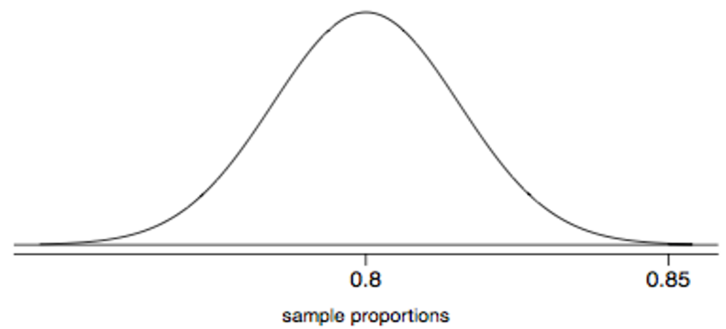
The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$Z = 3.22$ Our observation is over 3 SE's away from the null



Hypothesis Testing

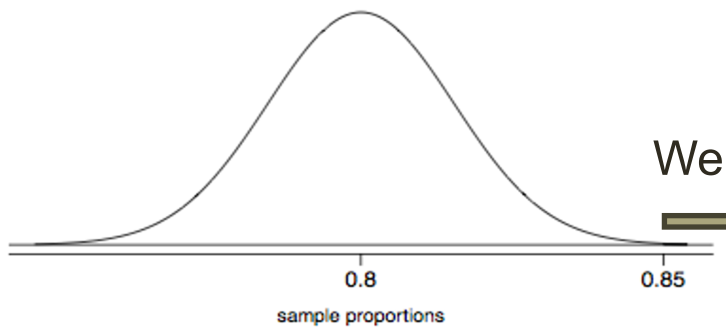
The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$Z = 3.22$ Our observation is over 3 SE's away from the null



We need to find the proportion of more extreme observations

Hypothesi

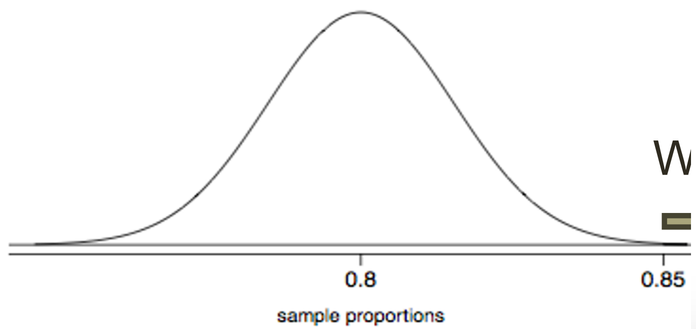
The GSS found that
question on experim
Do these data provi
Americans have a g

What are the steps

- Calculate
- Find p-val

$Z = 3.22$ Our ob

W



Positive Z

Z	Second decimal place of 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
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

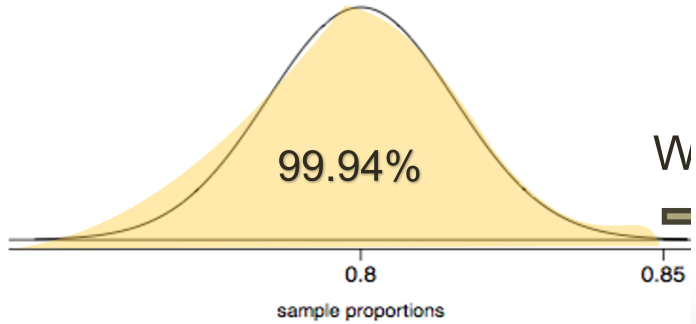
*For $Z \geq 3.50$, the probability is greater than or equal to 0.9998.

Hypothesi

The GSS found that
question on experim
Do these data provi
Americans have a g

What are the steps
- Calculate
- Find p-val

$Z = 3.22$ Our ob



W

Positive Z

Z	Second decimal place of 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
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

*For $Z \geq 3.50$, the probability is greater than or equal to 0.9998.

Hypothesis Testing

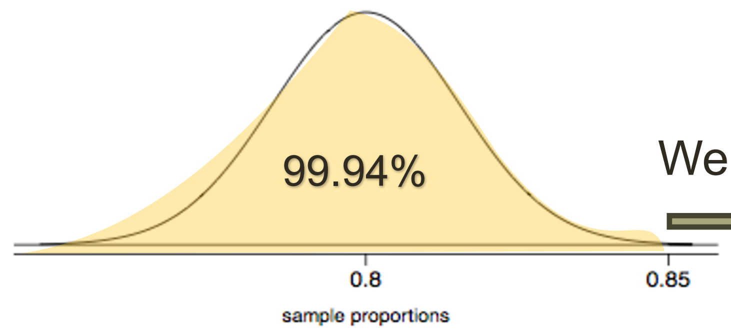
The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$\text{Z} = 3.22$ Our observation is over 3 SE's away from the null



We need to find the proportion of more extreme observations

$$p = 1 - 0.9994 = 0.0006$$

Hypothesis Testing

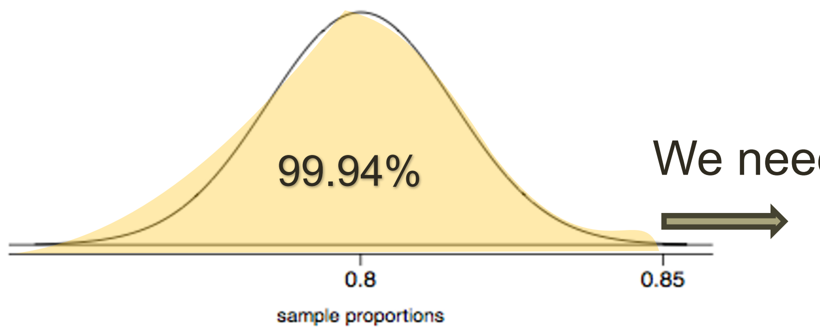
The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

What are the steps for a Z-test?

- Calculate test statistic (z-score)
- Find p-value for that test statistic

$Z = 3.22$ Our observation is over 3 SE's away from the null



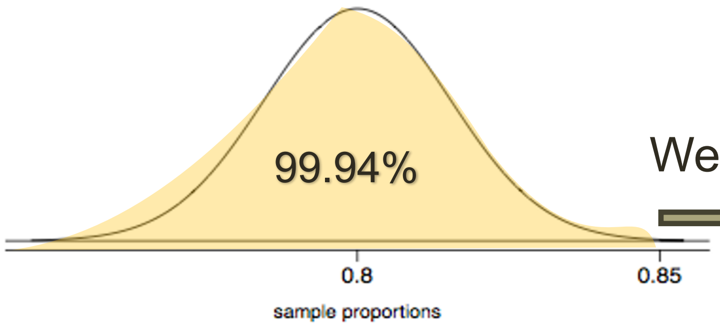
We need to find the proportion of more extreme observations

$p = 1 - 0.9994 = 0.0006$, if $\alpha = 0.05$ what is the result of this test?

Hypothesis Testing

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly.
Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$



We need to find the proportion of more extreme observations

$p = 1 - 0.9994 = 0.0006$, if $\alpha = 0.05$ what is the result of this test?

Since the p-value is low, we **reject *H_0***. The data provide convincing evidence that more than 80% of Americans have a good intuition on experimental design.

Practice

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey is ±3%. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween." At 95% confidence level, is this news piece's statement justified?

(a) Yes
(b) No
(c) Can’t tell

Practice

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey is ±3%. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween." At 95% confidence level, is this news piece's statement justified?

(a) Yes
(b) No → (8%, 14%) includes 8%, 9%, 10%
(c) Can’t tell

Practice

6.6 Elderly drivers. The Marist Poll published a report stating that 66% of adults nationally think licensed drivers should be required to retake their road test once they reach 65 years of age. It was also reported that interviews were conducted on 1,018 American adults, and that the margin of error was 3% using a 95% confidence level.[9]

(a) Verify the margin of error reported by The Marist Poll.

(b) Based on a 95% confidence interval, does the poll provide convincing evidence that *more than* 70% of the population think that licensed drivers should be required to retake their road test once they turn 65?

6.12 Is college worth it? Part I. Among a simple random sample of 331 American adults who do not have a four-year college degree and are not currently enrolled in school, 48% said they decided not to go to college because they could not afford school.[15]

(a) A newspaper article states that only a minority of the Americans who decide not to go to college do so because they cannot afford it and uses the point estimate from this survey as evidence. Conduct a hypothesis test to determine if these data provide strong evidence supporting this statement.

(b) Would you expect a confidence interval for the proportion of American adults who decide not to go to college because they cannot afford it to include 0.5? Explain.

Practice

6.6

(a) With a random sample from $< 10\%$ of the population, independence is satisfied. The success-failure condition is also satisfied. Hence, the margin of error can be calculated as follows:

$$ME = z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.66 \times 0.34}{1018}} = 0.029 \approx 3\%$$

(b) A 95% confidence interval for the proportion of adults who think that licensed drivers should be required to re-take their road test once they reach 65 years of age can be calculated as

$$0.66 \pm 0.03 = (0.63, 0.69).$$

No, our confidence interval is below 70%, it suggests less than 70% of Americans think drivers should be required to re-take the road test.

Practice

6.12

(a) The hypotheses are as follows:

$H_0 : p = 0.5$ (50% of Americas who decide not to go to college because they cannot afford it do so because they cannot afford it)

$H_A : p < 0.5$ (Less than 50% of Americas who decide not to go to college because they cannot afford it do so because they cannot afford it)

Before calculating the test statistic we should check that the conditions are satisfied.

1. Independence: The sample is representative and we can safely assume that $331 < 10\%$ of all American adults who decide not to go to college, therefore whether or not one person in the sample decided not to go to college because they can't afford it is independent of another.
2. Success-failure: $331 \times 0.5 = 165.5 > 10$ and $331 \times 0.5 = 165.5 > 10$.

Since the observations are independent and the success-failure condition is met, $\hat{p}$ is expected to be approximately normal. The test statistic can be calculated as follows:

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

$$= \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{331}}} = \frac{-0.02}{0.0275} = -0.73$$

$$p - value = P(\hat{p} < 0.48 | p = 0.5) = P(Z < -0.73) = 0.2327$$

Since the p-value is large, we fail to reject H_0. The data do not provide strong evidence that less than half of American adults who decide not to go to college make this decision because they cannot afford college.

(b) Yes, since we failed to reject $H_0 : p = 0.5$.

Recap - inference for one proportion

Population parameter: p, point estimate: $\hat{p}$

Recap - inference for one proportion

Population parameter: p, point estimate: $\hat{p}$

Conditions

- independence
  - random sample and 10% condition
- at least 10 successes and failures
  - if not → randomization

Recap - inference for one proportion

Population parameter: p, point estimate: $\hat{p}$

Conditions

- independence
  - random sample and 10% condition
- at least 10 successes and failures
  - if not → randomization

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$

- for CI: use $\hat{p}$
- for HT: use p_0