Chapter 22

What Is a Test of Significance?

If chance of observing an outcome sampled from a population with an *assumed* parameter is small, then choice of outcome is unlucky or, more likely, choice of population parameter is wrong. Chance in this situation is called *P-value.* Procedure of deciding whether population parameter is correct or not is called *test of significance.* Tests of significance involve following ratio,

$$\text{standard score} = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}$$

where, for population proportion p in particular,

$$\text{observation} = \hat{p}, \qquad \text{mean} = p \qquad \text{and} \qquad \text{standard deviation} = \sqrt{\tfrac{p(1-p)}{n}},$$

and, for population mean μ,

$$\text{observation} = \bar{x}, \qquad \text{mean} = \mu \qquad \text{and} \qquad \text{standard deviation} = \tfrac{s}{\sqrt{n}}.$$

Tests of significance can be approximated by simulation.

Exercise 22.1 (What Is a Test of Significance?)

1. *Test for proportion p: defective batteries.*
   In a battery factory, 8% of all batteries made are assumed to be defective. Technical trouble with production line, however, has raised concern percent defective has increased in past few weeks. Of $n = 600$ batteries chosen at random, $\frac{70}{600}$ths $\left(\frac{70}{600} \approx 0.117\right)$ of them are found to be defective. Does data support concern about defective batteries?

   (a) *Statement.* Choose one.

      i. $H_0 : p = 0.08$ versus $H_a : p < 0.08$
      ii. $H_0 : p \leq 0.08$ versus $H_a : p > 0.08$
      iii. $H_0 : p = 0.08$ versus $H_a : p > 0.08$

(b) *Test.*
Chance $\hat{p} = \frac{70}{600} \approx 0.117$ or more, *if* $p = 0.08$,
is equivalent to chance standard score Z is greater than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.117-0.08}{\sqrt{\frac{0.08(1-0.08)}{600}}} = 3.31$$

which equals **0.0005** / **0.0500** / **4.6500**.
(Using Table B, score 3.3 corresponds to percentile 99.95,
so P-value is $100 - 99.95 = 0.05\%$ or probability 0.0005.)

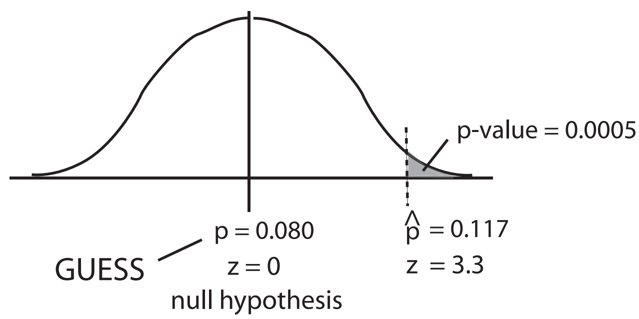

Figure 22.1 (P-value for sample $\hat{p} = 0.117$, if guess $p = 0.08$)

(c) *Conclusion.*
Since P-value = 0.0005 is so small,
do not reject / **reject** null guess: $H_0 : p = 0.08$.
So, sample $\hat{p}$ indicates population proportion p
is less than / **equals** / **is greater than** 0.08: $H_1 : p > 0.08$.

(d) *A comment: null hypothesis and alternative hypothesis.*
In this hypothesis test, we are asked to choose between
(choose one) **one** / **two** / **three** alternatives (or *hypotheses*, or guesses):
a *null* hypothesis of $H_0 : p = 0.08$ and an *alternative* of $H_a : p > 0.08$.
Null hypothesis is a statement of "status quo", of no change; test statistic used to reject it or not. Alternative hypothesis is "challenger" statement.

(e) *Another comment: P-value.* In this hypothesis test, P-value is chance observed proportion defective is 0.117 or more, guessing population proportion defective is $\boldsymbol{p = 0.117}$ / $\boldsymbol{p = 0.08}$. In general, *p-value is probability of observing test statistic or more extreme, assuming null hypothesis true.*

(f) *And another comment: test statistic different for different samples.*
If a second *sample* of 600 batteries were taken at *random* from all batteries, observed proportion defective of this second group of 600 batteries would probably be (choose one) **different from** / **same as**
first observed proportion of 600 batteries given above, $\hat{p} = 0.117$, say, $\hat{p} = 0.093$ which may change conclusions of hypothesis test.

2. *Test for proportion p: defective batteries again.*
As before, but of $n = 600$ batteries chosen at random, $\frac{56}{600}$ths $\left(\frac{56}{600} \approx 0.093\right)$ of them are found to be defective. Does data support concern about defective batteries?

   (a) *Statement.* Choose one.

      i. $H_0 : p = 0.08$ versus $H_a : p < 0.08$
      ii. $H_0 : p \leq 0.08$ versus $H_a : p > 0.08$
      iii. $H_0 : p = 0.08$ versus $H_a : p > 0.08$

   (b) *Test.*
   Chance $\hat{p} = \frac{56}{600} \approx 0.093$ or more, *if* $p = 0.08$,
   is equivalent to chance standard score Z is greater than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.093-0.080}{\sqrt{\frac{0.08(1-0.08)}{600}}} \approx 1.2$$

   which equals **0.1151** / **11.51** / **1151**.
   (Using Table B, score 1.2 corresponds to percentile 88.49,
   so P-value is $100 - 88.49 = 11.51\%$ or probability 0.1151.)

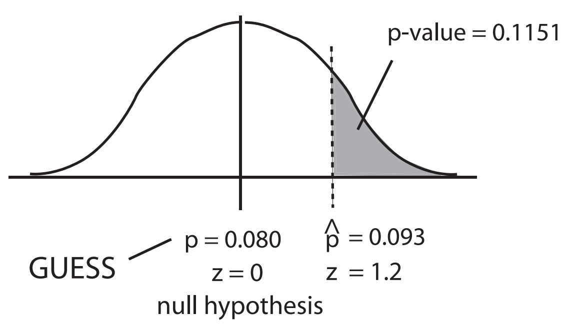

Figure 22.2 (P-value for sample $\hat{p} = 0.093$, if guess $p = 0.08$)

   (c) *Conclusion.*
   Since P-value = 0.1151 is so large,
   do not reject / **reject** null guess: $H_0 : p = 0.08$.
   So, sample $\hat{p}$ indicates population proportion p
   is less than / **equals** / **is greater than** 0.08: $H_0 : p = 0.08$.

   (d) *Comparing hypothesis tests.*
   P-value associated with $\hat{p} = 0.117$ (P-value = 0.0005) is
   (choose one) **smaller** / **larger**
   than P-value associated with $\hat{p} = 0.093$ (P-value = 0.1151).
   Sample defective proportion $\hat{p} = 0.117$ is
   (choose one) **closer to** / **farther away from**,
   than $\hat{p} = 0.093$, to null guess $p = 0.08$.

It makes sense we reject null guess of $p = 0.08$ when observed proportion is $\hat{p} = 0.117$, but not reject null guess when observed proportion is $\hat{p} = 0.093$.

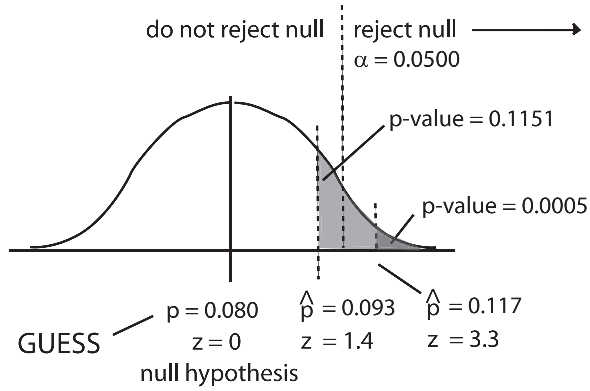

Figure 22.3 (Tend to reject null for smaller P-values)

If p–value is smaller than level of significance, α, reject null. If null is rejected when p–value is small, typically less than $\alpha = 0.05$, test is *significant*. If null is not rejected, test is *not significant*.

(e) *Population, Sample, Statistic, Parameter.* Match columns.

terms	battery example
(a) population	**(A)** all (defective or nondefective) batteries
(b) sample	**(B)** proportion defective, of all batteries, p
(c) statistic	**(C)** 600 (defective or nondefective) batteries
(d) parameter	**(D)** proportion defective, of 600 batteries, $\hat{p}$

terms	(a)	(b)	(c)	(d)
example				

3. *Test for proportion p: conspiring Earthlings.*
It appears 6.5% of Earthlings are conspiring with little green men (LGM) to take over Earth. Human versus Extraterrestrial Legion Pact (HELP) claims more than 6.5% of Earthlings are conspiring with LGM. In a random sample of 100 Earthlings, 7 $\left(\frac{7}{100} = 0.07\right)$ are found to be conspiring with little green men (LGM). Does this data support HELP claim at $\alpha = 0.05$?

(a) *Statement.* Choose one.

i. $H_0 : p = 0.065$ versus $H_a : p < 0.065$
ii. $H_0 : p \leq 0.065$ versus $H_a : p > 0.065$
iii. $H_0 : p = 0.065$ versus $H_a : p > 0.065$

(b) *Test.*
Chance $\hat{p} = \frac{7}{100} = 0.070$ or more, *if* $p = 0.065$, is
is equivalent to chance standard score Z is greater than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.070-0.065}{\sqrt{\frac{0.065(1-0.065)}{100}}} \approx 0.2$$

which equals **0.5793** / **0.4207** / **0.1151**.

(Using Table B, score 0.2 corresponds to percentile 57.93,

so P-value is 100 − 57.93 = 42.07% or probability 0.4207.)

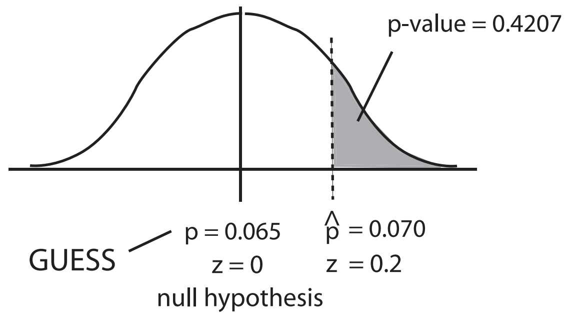

Figure 22.4 (P-value for sample $\hat{p} = 0.070$, if guess $p = 0.065$)

(c) *Conclusion.*
Since P-value $= 0.42 > \alpha = 0.05$,
(circle one) **do not reject** / **reject** null guess: $H_0 : p = 0.065$.
So, sample $\hat{p}$ indicates population proportion p
is less than / **equals** / **is greater than** 0.065: $H_0 : p = 0.065$.

(d) *Another Comment.* P-value, in this case, is chance (choose one)

i. population proportion 0.07 or more, if observed proportion 0.065.
ii. observed proportion 0.07 or more, if observed proportion 0.065.
iii. population proportion 0.07 or more, if population proportion 0.065.
iv. observed proportion 0.07 or more, if population proportion 0.065.

4. *Test for proportion p: overweight in Indiana.*
An investigator wishes to know whether proportion of overweight individuals in Indiana *is less than* national proportion of 70% or not. A random sample of size $n = 600$ results in 390 $\left(\frac{390}{600} = 0.65\right)$ who are overweight. Test at $\alpha = 0.05$.

(a) *Statement.* Choose one.

i. $H_0 : p = 0.7$ versus $H_a : p > 0.7$
ii. $H_0 : p = 0.7$ versus $H_a : p < 0.7$
iii. $H_0 : p = 0.7$ versus $H_a : p \neq 0.7$

(b) *Test.*
Chance $\hat{p} = \frac{390}{600} = 0.65$ or *less*, *if* $p = 0.7$, is
is equivalent to chance standard score Z is *less* than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.65-0.70}{\sqrt{\frac{0.7(1-0.7)}{600}}} \approx -2.7$$

which equals **0.35** / **0.0035** / **0.000035**.

(Using Table B, score -2.7 corresponds to percentile 0.35, so P-value is 0.35% or probability 0.0035.)

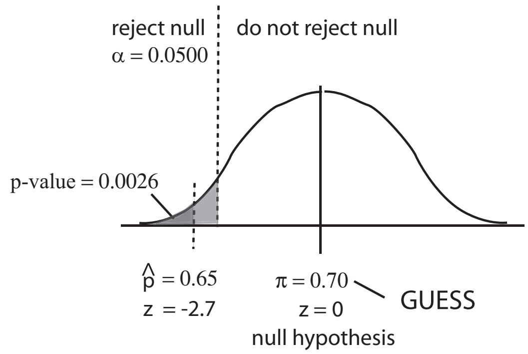

Figure 22.5 (P-value for sample $\hat{p} = 0.65$, if guess $p = 0.70$)

(c) *Conclusion.*
Since P-value $= 0.0035 < \alpha = 0.0500$,
(circle one) **do not reject** / **reject** null guess: $H_0 : p = 0.70$.
In other words, sample $\hat{p}$ indicates population proportion p
(circle one) **less than** / **equals** / **does not equal** 0.70: $H_a : p < 0.70$.

5. *Simulation of test for proportion p: overweight in Indiana.*
An investigator wishes to know whether proportion of overweight individuals in Indiana *is less than* national proportion of 70% or not. A random sample of size $n = 600$ results in 360 $\left(\frac{360}{600} = 0.6\right)$ who are overweight. Simulate P-value, chance proportion overweight at most 0.6, $\hat{p} \leq 0.6$, *if* $p = 0.7$.

(a) *Statement.* Choose one.

i. $H_0 : p = 0.7$ versus $H_a : p > 0.7$
ii. $H_0 : p = 0.7$ versus $H_a : p < 0.7$
iii. $H_0 : p = 0.7$ versus $H_a : p \neq 0.7$

(b) *Simulation of Test, 10 repetitions.*
scan ten digits at a time; start line 121
if 0-6, overweight; if 7-9, not overweight;
calculate proportion of ten which are overweight

71487 09984 $\frac{4}{10} = 0.4$	29077 14863 $\frac{6}{10} = 0.6$	61683 47052 $\frac{8}{10} = 0.8$	62224 51025 $\frac{10}{10} = 1.0$	13873 81598 $\frac{5}{10} = 0.5$
95052 90908 $\frac{6}{10} = 0.6$	73592 75186 $\frac{6}{10} = 0.6$	87136 87136 $\frac{6}{10} = 0.6$	54580 81507 $\frac{7}{10} = 0.7$	27102 56027 $\frac{8}{10} = 0.8$

Since 6 of 10 simulations have $\hat{p} \leq 0.6$, P-value $\approx \frac{6}{10} =$ **0.5** / **0.6** / **0.7**.

(c) *Conclusion.*
Since P-value $= 0.60 > \alpha = 0.05$,
(circle one) **do not reject** / **reject** null guess: $H_0 : p = 0.70$.
In other words, sample $\hat{p}$ indicates population proportion p
(circle one) **less than** / **equals** / **does not equal** 0.70: $H_0 : p = 0.70$.

6. *Testing μ, left–sided: weight of coffee*
Label on a large can of Hilltop Coffee states average weight of coffee contained in all cans it produces is 3 pounds of coffee. A coffee drinker association claims average weight is less than 3 pounds of coffee, $\mu < 3$. Suppose a random sample of 30 cans has an average weight of $\bar{x} = 2.95$ pounds and standard deviation of $s = 0.18$. Does data support coffee drinker association's claim at $\alpha = 0.05$?

(a) *Statement.* Choose one.

i. $H_0 : \mu = 3$ versus $H_1 : \mu < 3$
ii. $H_0 : \mu < 3$ versus $H_1 : \mu > 3$
iii. $H_0 : \mu = 3$ versus $H_1 : \mu \neq 3$

(b) *Test.*
Chance observed $\bar{x} = 2.95$ or *less*, *if* $\mu = 3$, is
is equivalent to chance standard score Z is *less* than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{2.95-3}{\frac{0.18}{\sqrt{30}}} \approx -1.5$$

which equals **0.07** / **0.0007** / **0.000007**.
(Using Table B, score -1.5 corresponds to percentile 6.68, so P-value is 6.68% or probability 0.0668.)

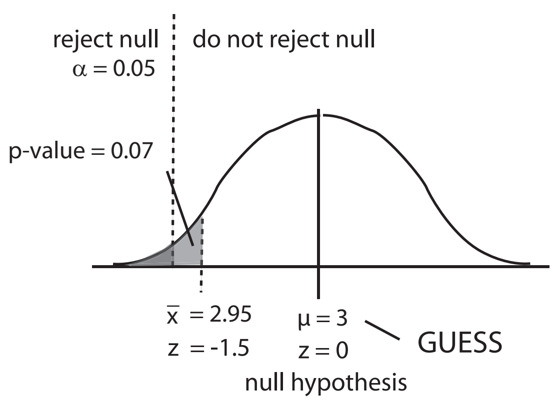

Figure 22.6 (Left-sided test of μ: coffee can weights)

(c) *Conclusion.*
Since p–value $= 0.07 > \alpha = 0.05$,
(circle one) **do not reject** / **reject** the null guess: $H_0 : \mu = 3$.
In other words, sample $\bar{x}$ indicates population average weight μ
(circle one) **is less than** / **equals** / **is greater than** 3: $H_0 : \mu = 3$.

7. *Testing μ, two–sided: golf ball flight.*
Superflight golf balls are rated to fly a distance of 280 yards. An inspection process tries to determine if the flight of these golf balls *differs* from 280 yards. Suppose, in a random sample of 36 golf balls, average distance is $\bar{x} = 278.5$ with a standard deviation of $s = 12$. Does this data support inspection "claim" $\mu \neq 280$ at $\alpha = 0.05$?

   (a) *Statement.* Choose one.
      i. $H_0 : \mu = 280$ versus $H_1 : \mu \neq 280$
      ii. $H_0 : \mu < 280$ versus $H_1 : \mu > 280$
      iii. $H_0 : \mu = 280$ versus $H_1 : \mu < 280$

   (b) *Test.*
   Since two-sided, chance equals 2× chance standard score Z is less than

$$\text{standard score} = \frac{\text{observation–mean}}{\text{standard deviation}} = \frac{278.5-280}{\frac{12}{\sqrt{36}}} \approx -0.8$$

   which equals **0.07** / **0.2119** / **0.4238**.
   (Using Table B, score -0.8 corresponds to percentile 21.19,
   so, since *two*-sided, P-value is $2 \times 21.19\% = 42.38\%$ or probability 0.4238.)

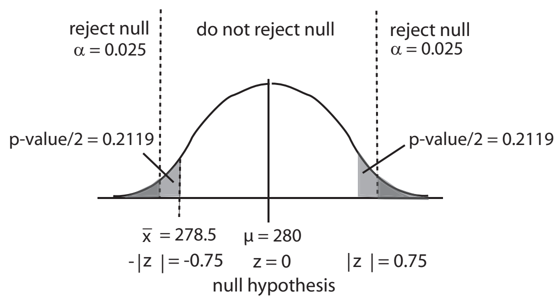

Figure 22.7 (Two–sided test and golf ball flight)

   (c) *Conclusion.*
   Since p–value $= 0.4238 > \alpha = 0.0500$,
   (circle one) **do not reject** / **reject** the null guess: $H_0 : \mu = 280$.
   In other words, sample $\bar{x}$ indicates population average length μ
   (circle one) **equals** / **does not equal** 280: $H_0 : \mu = 280$.

8. *Comparing Type I error (α) and Type II error (β).*

chosen ↓ actual →	null H_0 true	alternative H_1 true
choose null H_0	correct decision	type II error, β
choose alternative H_1	type I error, α	correct decision: power $= 1 - \beta$

   (a) *Definitions.* Type I error is accidently rejecting the **null** / **alternative**, and type II error is accidently rejecting the **null** / **alternative**.

(b) *Defective batteries.*
Since test is $H_0 : p = 0.08$ versus $H_a : p > 0.08$,
type I error is finding true percent defective batteries is

i. greater than 8% when, in fact, it is only 8%
ii. equals 8% when, in fact, it is greater than 8%

whereas type II error is finding true percent defective batteries is

i. greater than 8% when, in fact, it is only 8%.
ii. equals 8% when, in fact, it is greater than 8%.

worst error, in this case, is **type I / type II**

(c) *Conspiring Earthlings.*
Since test is $H_0 : p = 0.065$ versus $H_a : p > 0.065$,
type I error is finding true percent conspiring Earthling is

i. greater than 6.5% when, in fact, it is only 6.5%
ii. equals 6.5% when, in fact, it is greater than 6.5%

whereas type II error is finding true percent conspiring Earthling is

i. greater than 6.5% when, in fact, it is only 6.5%.
ii. equals 6.5% when, in fact, it is greater than 6.5%.

worst error, in this case, is **type I / type II**

(d) *Weight of coffee.*
Since test is $H_0 : \mu = 3$ versus $H_a : \mu < 3$,
type I error is finding true average coffee can weight is

i. less than 3 pounds when, in fact, it is 3 pounds
ii. equals 3 pounds when, in fact, it is less than 3 pounds

whereas type II error is finding true average coffee can weight is

i. less than 3 pounds when, in fact, it is 3 pounds.
ii. equals 3 pounds when, in fact, it is less than 3 pounds.

worst error, in this case, is **type I / type II**

(e) *Property.* So, as chance of type I error, α, decreases, chance of type II error, β, **decreases / increases** because if chance of mistakenly rejecting null decreases, this necessarily means chance of mistakenly rejecting alternative increases.

Chapter 23

Use and Abuse of Statistical Inference

Tests of inference and confidence intervals should be used carefully. *Statistical* significance may not necessarily mean *practical* significance. Sample size influences statistical inference: a significant result may be missed if sample size is too small; statistical, but not practical, significance may occur if sample size is too large. Determine statistical inference from one (and only one) SRS because statistical significance will occur by chance alone in one of many SRSs taken.

Exercise 23.1 (Use and Abuse of Statistical Inference)

1. *Statistical inference versus practical significance.*

   (a) *Defective batteries.*
   Test statistic, $\hat{p} = \frac{70}{600} \approx 0.12$, of whether proportion of defective batteries was 0.08 or larger than this was of *statistical* significance because chance of $\hat{p} \approx 0.12$ or more, *if* $p = 0.08$, was small (P-value = 0.0005). Is $\hat{p} \approx 0.12$ of *practical* significance?

      i. No, because difference $0.12 - 0.08 = 0.04$, 4%, is small.
      ii. Yes, because difference $0.12 - 0.08 = 0.04$, 4%, is large.

   This is a discussion-type question: practical significance is subjective, based on whether the actual difference affects things like program goals or safety requirements: 4% defective of 1000 lithium car batteries, 40, may be of serious practical concern if these 40 defective batteries lead to 40 injured or dead, whereas 4% defective of 1000 flashlight batteries may not be of any practical safety concern.

   (b) *Conspiring Earthlings.*
   Test statistic, $\hat{p} = \frac{7}{100} \approx 0.070$, of whether proportion of conspiring Earthlings was 0.065 or larger than this was of *in*significance because chance of $\hat{p} \approx 0.070$ or more, *if* $p = 0.065$, was large (P-value = 0.42). Is $\hat{p} \approx 0.070$ of *practical* significance?

i. No, difference $0.070 - 0.065 = 0.005$, 0.5%, is small.
ii. Yes, difference $0.070 - 0.065 = 0.005$, 0.5%, is large.

Is 0.5% of 7 billion Earthlings, 35 million, where each extra conspirator leads to 2 other imprisoned Earthlings, a lot of extra conspiring Earthlings?

(c) *Overweight in Indiana.*
Test statistic, $\hat{p} = \frac{390}{600} \approx 0.65$, of whether proportion of overweight was 0.70 or less than this was of statistical significance because chance of $\hat{p} \approx 0.65$ or less, *if* $p = 0.70$, was small (P-value = 0.0035). Is $\hat{p} \approx 0.65$ of *practical* significance?

i. No, difference $0.65 - 0.70 = -0.05$, -5%, is small.
ii. Yes, difference $0.65 - 0.70 = -0.05$, -5%, is large.

What if it costs Indiana companies 30 sick days per extra overweight individual? What if it costs only an extra 2 sick days per overweight individual?

(d) *Coffee can weight.*
Test statistic, $\bar{x} \approx 2.95$, of whether average coffee can weight was 3 pounds or less than this was of statistical *in*significance because chance of $\bar{x} \approx 2.95$ or less, *if* $\mu = 3$, was large (P-value = 0.07). Is $\bar{x} \approx 2.95$ of *practical* significance?

i. No, difference $3 - 2.95 = 0.05$ pounds is small.
ii. Yes, difference $3 - 2.95 = 0.05$ pounds is large.

(e) *Golf ball flight.*
Test statistic, $\bar{x} \approx 278.5$, of whether average golf ball flight was 280 yards or different from this was of statistical *in*significance because twice chance of $\bar{x} \approx 278.5$ or less, *if* $\mu = 280$, was large (P-value = 0.42). Is $\bar{x} \approx 278.5$ of *practical* significance?

i. No, difference $278.5 - 280 = -1.5$ yards is small.
ii. Yes, difference $278.5 - 280 = -1.5$ yards is large.

2. *Sample size and statistical significance.*

(a) *Defective batteries.*
If $n = 600$, $p = 0.08$ and $\hat{p} = \frac{70}{600} \approx 0.117$,
chance standard score Z is greater than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.117-0.08}{\sqrt{\frac{0.08(1-0.08)}{600}}} = 3.31$$

equals **0.0005 / 0.0500 / 4.6500**.

(Using Table B, score 3.3 corresponds to percentile 99.95,

so P-value is $100 - 99.95 = 0.05\%$ or probability 0.0005.)

If $n = 60$, $p = 0.08$ and $\hat{p} = \frac{7}{60} \approx 0.117$,
chance standard score Z is greater than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.117-0.08}{\sqrt{\frac{0.08(1-0.08)}{60}}} = 1.1$$

equals **13.57** / **0.1357** / **0.001357**.

(Using Table B, score 1.1 corresponds to percentile 86.43,

so P-value is $100 - 86.43 = 13.57\%$ or probability 0.1357.)

P-value when $n = 600$ is **smaller** / **larger** than P-value when $n = 60$.
Significant test at $n = 600$ becomes non-significant at $n = 60$ if $\alpha = 0.05$.

(b) *Overweight in Indiana.*
If $n = 600$, $p = 0.70$ and $\hat{p} = \frac{390}{600} \approx 0.65$,
chance standard score Z is *less* than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.65-0.70}{\sqrt{\frac{0.7(1-0.7)}{600}}} \approx -2.8$$

equals **0.0014** / **0.0026** / **0.0056**.

(Using Table B, score -2.8 corresponds to percentile 0.26, so P-value is 0.26% or probability 0.0026.)

If $n = 60$, $p = 0.70$ and $\hat{p} = \frac{39}{60} \approx 0.65$,
chance standard score Z is *less* than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{0.65-0.70}{\sqrt{\frac{0.7(1-0.7)}{60}}} \approx -1.5$$

equals **6.68** / **0.0567** / **0.0668**.

(Using Table B, score -1.5 corresponds to percentile 6.68, so P-value is 6.68% or probability 0.0668.)

P-value when $n = 600$ is **smaller** / **larger** than P-value when $n = 60$.
Significant test at $n = 600$ becomes non-significant at $n = 60$ if $\alpha = 0.05$.

(c) *Coffee can weight.*
If $n = 300$, chance observed $\bar{x} = 2.95$ or less, if $\mu = 3$ and $s = 0.18$,
chance standard score Z is less than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{2.95-3}{\frac{0.18}{\sqrt{300}}} \approx -4.8$$

equals **0.00** / **0.07** / **0.09**.

(Using Table B, score -4.8 is off table, so P-value is 0% or probability 0.00.)

If $n = 30$, chance observed $\bar{x} = 2.95$ or less, if $\mu = 3$ and $s = 0.18$,
chance standard score Z is less than

$$\text{standard score} = \frac{\text{observation}-\text{mean}}{\text{standard deviation}} = \frac{2.95-3}{\frac{0.18}{\sqrt{30}}} \approx -1.5$$

equals **0.07** / **0.0007** / **0.000007**.

(Using Table B, score -1.5 corresponds to percentile 6.68, so P-value is 6.68% or probability 0.0668.)

P-value when $n = 300$ is **smaller** / **larger** than P-value when $n = 30$.
Significant test at $n = 300$ becomes *in*significant at $n = 30$ if $\alpha = 0.05$.

(d) *Coffee can weight, 95% confidence interval.*
If $n = 300$, $\bar{x} = 2.95$, $s = 0.18$, 95% CI of coffee can weight μ

$$\bar{x} \pm 2\left(\frac{s}{\sqrt{n}}\right) = 2.95 \pm 2\left(\frac{0.18}{\sqrt{300}}\right) =$$

2.950 ± 0.021 / **2.950 ± 0.066** / **2.950 ± 0.263**
If $n = 30$, $\bar{x} = 2.95$, $s = 0.18$, 95% CI of coffee can weight μ

$$\bar{x} \pm 2\left(\frac{s}{\sqrt{n}}\right) = 2.95 \pm 2\left(\frac{0.18}{\sqrt{30}}\right) =$$

2.950 ± 0.021 / **2.950 ± 0.066** / **2.950 ± 0.263**
95% CI when $n = 300$ is **shorter** / **longer** than 95% CI when $n = 30$.

3. *Perform statistical inference procedure only once.*

(a) *Defective batteries.*
If $p = 0.08$ (really) is true, we would expect most observed $\hat{p}$ to be close to $p = 0.08$. If a 95% CI is actually calculated many times with different SRSs, 95% of CIs include $p = 0.08$ and so 5% exclude $p = 0.08$. If test of significance with $\alpha = 0.05$ is calculated many times with different SRSs, what percentage are significant? **2.5%** / **5%** / **7.5%**

(b) *Coffee can weight.*
If $\mu = 3$ pounds (really) is true, we would expect most observed $\bar{x}$ to be close to $\mu = 3$ pounds. If a 95% CI is actually calculated many times with different SRSs, 95% of CIs include $\mu = 3$ pounds and so **2.5%** / **5%** / **7.5%** exclude $\mu = 3$ pounds. If test of significance with $\alpha = 0.05$ is calculated many times with different SRSs, what percentage are significant? **2.5%** / **5%** / **7.5%**