Chapter 6
Introduction to Hypothesis Testing

6.1 **Decision Rule:** Reject H_0 if coin comes up heads every time in n tosses
Retain H_0 if tails comes up at least once in n tosses

α = Pr(Reject H_0 | H_0 is true) = Pr(All heads | The coin is fair) = f(n) = ${}_nC_n(1/2)^n(1-1/2)^0 = 1/2^n$

β = Pr(Retain H_0 | H_0 is false) = Pr(At least one tails | Coin is two-headed) = 0

a. $\alpha = 1/2 = 0.50$, $\beta = 0$

b. $\alpha = 1/16 = 0.0625$, $\beta = 0$

c. $\alpha = 1/33554432 = 2.9\text{x}10^{-8}$, $\beta = 0$

d. $\alpha = 2^{-100} = 7.8\text{x}10^{31}$, $\beta = 0$

6.3 **Decision Rule:**
Reject H_0 if coin comes up heads 12 times
Retain H_0 if tails comes up at least once in the 12 tosses

α = Pr(Reject H_0 | H_0 is true) = Pr(12 heads | The coin is fair) = 2^{-12}

β = Pr(Retain H_0 | H_0 is false) = Pr(At least one tails | Coin is bent) = $1 - (0.75)^{12} = 0.9683$

6.5 **Decision Rule:** Reject H_0 is $K \geq 11$ Retain H_0 if $K < 11$

a. α = Pr(Reject H_0 | H_0 is true) = $\Pr(K \geq 11 | \pi = .5)$ = Pr(K = 11) + Pr(K = 12) =
${}_{12}C_{11}(.5)^{11}(1-.5)^1 + {}_{12}C_{12}(.5)^{12}(1-.5)^0 = 13/(2^{12}) = .0032$

b. β = Pr(Retain H_0 | H_0 is false) = $\Pr(K < 11 | \pi = .80) = 1 - \{\Pr(K = 11) + \Pr(K = 12)\}$ =
$1 - \{{}_{12}C_{11}(.80)^{11}(.20)^1 + {}_{12}C_{12}(.80)^{12}(.20)^0\} = 1 - 0.2749 = 0.7251$

c. β = Pr(Retain H_0 | H_0 is false) = $\Pr(K < 11 | \pi = .95) = 1 - \{\Pr(K = 11) + \Pr(K = 12)\}$ =
$1 - \{{}_{12}C_{11}(.95)^{11}(.05)^1 + {}_{12}C_{12}(.95)^{12}(.05)^0\} = 1 - 0.8816 = 0.1184$.

6.7 H_0: The preferences for Pepsi and Coke are the same. H_a: People prefer Pepsi
Let K be the number in the sample who prefer Pepsi
Decision Rule: Reject H_0 is $K \geq 6$ Retain H_0 if $K < 6$

a. $\alpha = \Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \Pr(K \geq 6 \mid \pi = .5) = 1 - 0.6230 = 0.3770$

b. $\beta = \Pr(\text{Retain } H_0 \mid H_0 \text{ is false}) = \Pr(K \leq 5 \mid \pi = .60) = 0.3669$

c. $\beta = \Pr(\text{Retain } H_0 \mid H_0 \text{ is false}) = \Pr(K < 5 \mid \pi = .70) = 0.1503$

6.9 **This is the six-step procedure to test the hypothesis that the pulse goes down after a physical fitness program, using the sign test.**

Let π be the proportion of times that the pulse goes down.

Step 1: H_0: $\pi = .5$

Step 2: H_a: $\pi > .5$

Step 3: $\alpha = 0.10$

Step 4: Look at the sample of 5 subjects. Assign (+) if pulse goes down, assign (–) is the pulse goes up, and assign (0) if they are tied. We have $k_{(0)} = 1$, $k_{(+)} = 4$, and $k_{(-)} = 0$.
This makes $n = 5 - 1 = 4$.

Test statistic is $k_s = 4$, $n = 4$. Having 4 successes out of 4 is evidence for the alternative hypothesis. So we continue with the next step.

Step 5: P-value $= \Pr(K \geq k_s) = \Pr(K \geq 4) = {}_4C_4(.5)^4(1 - 0.5)^0 = 2^{-4} = 1/16 = 0.0625 < \alpha$. So we reject H_0.

Step 6: Conclusion in English. There is sufficient evidence (P-value = 0.0625) to conclude that the at-rest pulse goes down after a three-month physical fitness program.

6.11 **This is the six-step procedure to test the hypothesis that men tend to exaggerate their height, using the sign test.**

Let π be the proportion of times that the reported height is greater than the measured height.

Step 1: H_0: $\pi = .5$

Step 2: H_a: $\pi > .5$

Step 3: $\alpha = 0.10$

Step 4: Look at the sample of 12 subjects. Assign (+) if reported height is greater than measured height, assign (–) if it's the other way around, and assign (0) if they are tied. We have $k_{(0)} = 1$, $k_{(+)} = 10$, and $k_{(-)} = 1$. This makes n = 12 – 1 = 11.

Test statistic is $k_s = 10$, n = 11. Having 10 successes out of 9 is evidence for the alternative hypothesis. So we continue with the next step.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 10) = {}_{11}C_{10}(.5)^{11}(1–.5)^1 + {}_{11}C_{11}(.5)^{11}(1–.5)^0 = 12/2^{11} = 0.0059 < \alpha$. So we reject H_0.

Step 6: Conclusion in English. There is strong evidence (P-value = 0.0059) to conclude that men exaggerate their height.

6.13 This is the six-step procedure to test the hypothesis that consumers prefer the new formulation, using the sign test.

Step 1: H_0: $\pi = .5$

Step 2: H_a: $\pi > .5$

Step 3: $\alpha = .05$

Step 4: There were 5 people who could not tell the difference between the two; so $k_{(0)} = 5$. This makes n = 100 – 5 = 95. There are 57 who prefer the new formulation; so $k_{(+)} = 57$. There are 38 who prefer to old formulation; so $k_{(-)} = 38$. Test statistic is $k_s = 57$ and n = 95. Having 57 successes out of 95 is consistent with alternative hypothesis. So we continue with the next step.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 2) =$ 1 – binomcdf(95,.5,56) = 1 – 0.9679 = 0.0321 < α. So we reject H_0.

Step 6: Conclusion in English. There is sufficient evidence (P-value = 0.0321) to conclude that the consumers prefer the new formulation.

6.15

	Color-blind	Not	Totals
Men	50	950	1000
Women	4	996	1000
Totals	54	1946	2000

Among color-blind subjects, let π be the proportion of those who are men.

Step 1: H_0: Men and women are just as likely to be color-blind: $\pi = 0.50$

Step 2: H_a: A higher percentage of color-blind people are men: $\pi > .5$

Step 3: $\alpha = .01$

Step 4: In the sample of 54 color-blind subjects, let k_s be the sample count of the men. $k_s = 50$

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 50) = 1 - \Pr(K \leq 49) = 1 -$ binomcdf(54,.5,49) = $1.901 \times 10^{-11} < \alpha$. Reject null hypothesis.

Step 6: Conclusion in English. There is very strong evidence (P-value = 1.901×10^{-11}) to conclude that men are more likely than women to be color-blind.

6.17

	Fatigue	No fatigue	Row total
Skipped breakfast	38	62	100
Ate breakfast	45	155	200
Column Total	83	547	300

Among the fatigued people, let π be the proportion of those who skipped breakfast.

Step 1: H_0: People who skip breakfast are not more fatigued. Since those who skipped breakfast comprise 100/300 = 1/3 of the sample, this means that $\pi = 1/3$

Step 2: H_a: People who skip breakfast are more likely to suffer late morning fatigue. $\pi > 1/3$

Step 3: $\alpha = .05$

Step 4: In the sample of 83 who suffer late morning fatigue, 38 skipped breakfast, $k_s = 38$.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 38) = 1 - \Pr(K \leq 37) = 1 -$ binomcdf(83,1/3,37) = $1 - 0.9877 = 0.0123 < \alpha$. Reject null hypothesis.

Step 6: Conclusion in English. There is sufficient evidence (P-value = 0.0123) to conclude that people who skip breakfast are more likely to suffer late morning fatigue.

6.19

	Headache	No Headache	Totals
Drug Treatment	9	111	120
Placebo	3	117	120
Totals	12	228	240

Among subjects with headaches, let π be the probability the subject is found in the drug treatment group.

Step 1: H_0: The drug group has headaches with probability equal to the proportion of subjects in this group: $\pi = 120/2140 = .5$

Step 2: H_a: Those taking drug have headaches with probability higher than .5: $\pi > .5$

Step 3: $\alpha = .05$

Step 4: In the sample of 12 subjects complaining of headaches, let k_s be the sample count of in the drug treatment group. $k_s = 9$

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 9) = 1 - \Pr(K \leq 8) = 1 -$ binomcdf(12,.5,8) = .0730 > α. So we retain H_0.

Step 6: Conclusion in English. There is insufficient evidence (P-value = .0730) to conclude that the drug treatment increases incidence of headaches.

6.21

	Fatal	Other	Totals
Vaccine	0	200745	200745
Placebo	4	201225	201229
Totals	4	11	401964

Step 1: H_0: Vaccine is not effective in preventing polio deaths.

Step 2: H_a: Vaccine is effective in preventing polio deaths.

Step 3: No α given; let's choose $\alpha = .05$ to show significance.

Step 4: Among the 4 subjects in the fatal group, let $k_s = 0$ be the number observed in the vaccine group. Then $N = 401964$, $N_1 = 200745$, and $n = 4$.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 4) = \dfrac{\binom{200745}{0}\binom{201229}{4}}{\binom{401964}{4}} =$

$$\frac{201229 \cdot 201228 \cdot 201227 \cdot 201226}{401964 \cdot 401963 \cdot 401962 \cdot 401961} = 0.0628$$ $> \alpha$. Retain the null hypothesis.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0628) to conclude that the vaccine prevents polio deaths. Clearly because of the small number of deaths in the study, the test is weak. That is, it is difficult to disprove the null. This is like trying to prove that a coin is not fair based on four tosses. Even if all four tosses come up heads, there is insufficient evidence to conclude that the coin is not fair.

6.23

	Loan officer	Teller	Totals
Women	1	9	10
Men	3	1	4
Totals	4	10	14

a. Binomial exact test.

Step 1: H_0: A fair number of women were hired for the loan officer positions.

Step 2: H_a: Women were less likely to be hired as loan officers. That is, there was sex discrimination in the hiring.

Step 3: $\alpha = .05$

Step 4: Among the 4 loan officer positions, let $k_s = 1$ be the number of women hired.

Step 5: P-value = $\Pr(K \le k_s) = \Pr(K \le 1) = {}_4C_0(.5)^0(1-.5)^4 + {}_4C_1(.5)^1(1-.5)^3 = 0.0733 > \alpha$. Retain the null hypothesis.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0733) to conclude that the company practiced sex discrimination in hiring. Since the binomial exact test failed to show the that the evidence was significant and because n = 4 is less than 5% of the total number of employees hired N = 14, Fisher's exact test is recommended.

b. Fisher's exact test. Steps 1 – 3 are the same as above.

Step 4: Among the 4 loan officer positions, let $k_s = 1$ be the number of women hired. Here N = 14, $N_1 = 1$, and n = 4.

Step 5: P-value = $\Pr(K \le k_s) = \dfrac{\binom{10}{0}\binom{4}{4}}{\binom{14}{4}} + \dfrac{\binom{10}{1}\binom{4}{3}}{\binom{14}{4}} = 0.0410 < \alpha$. Reject the null hypothesis.

Step 6: Conclusion in English: There is sufficient evidence (P-value = .0410) to conclude that the bank discriminated against women in hiring. Since Fisher's exact test has more power than the binomial exact, Fisher's showed significant discrimination, whereas the binomial exact test failed to do so.

6.25

	Endometriosis	No endometriosis	Totals
Low-cal	1	59	60
Well-fed	6	54	60
Totals	7	113	120

Step 1: H_0: A low-calorie diet does not reduce the risk of endometriosis in rhesus monkeys.

Step 2: H_a: A low-calorie diet reduces the risk of endometriosis in rhesus monkeys.

Step 3: $\alpha = .05$

Step 4: Among the 7 cases of endometriosis, let $k_s = 1$ be the number in the low-calorie group.

Step 5: P-value = $\Pr(K \le k_s) = \Pr(K \le 1) = \dfrac{\binom{60}{0}\binom{60}{7}}{\binom{120}{7}} + \dfrac{\binom{60}{1}\binom{60}{6}}{\binom{120}{7}} = 0.0570 > \alpha$.

Retain the null.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0570) to conclude that a low-calorie diet reduced the risk of endometriosis in rhesus monkeys.

Remark: The binomial exact test produces a larger P-value = binomcdf(7,.5,1) = 0.0625. This is the probability of getting 1 or fewer heads in 7 tosses of a fair coin.

6.27 This is the six-step procedure to test the hypothesis more than half of families in Wilmette have income of $50,000 or more.

Let π be the proportion of families in Wilmette with income of at least $50,000 (in 1989).

Step 1: H_0: The median family income is $50,000; that is, $\pi = .5$

Step 2: H_a: The median family income is greater than $50,000; that is, $\pi > .5$

Step 3: $\alpha = .05$

Step 4: Look at the sample of 20 families. For each family, assign (+) if the income is above $50,000, assign (–) if below $50,000, and assign (0) if it is exactly $50,000. We have $k_{(0)} = 0$, $k_{(+)} = 12$, and $k_{(-)} = 8$. This makes $n = 20 - 0 = 20$.

Test statistic is $k_s = 12$, $n = 20$. Having as many as 12 successes out of 20, is consistent with the alternative hypothesis. So we continue with the next step.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 12) = 1 -$ binomcdf(20,.5,11) $= 1 - 0.7483 = 0.2517 > \alpha$. Retain the null.

Step 6: Conclusion in English. There is insufficient evidence (P-value =0.2517) to conclude that the median income in Wilmette is more than $50,000.

6.29 This is the six-step procedure to test the hypothesis more than half of white non-Hispanic girls have BMI above the 50th percentile of the CDC growth chart.

Let π be the proportion of white non-Hispanic girls who are above the CDC 50th percentile.

Step 1: H_0: $\pi = .5$

Step 2: H_a: $\pi > .5$

Step 3: $\alpha = .01$

Step 4: Look at the sample of 34 white non-Hispanic girls. Assign (+) if BMI percentile on the CDC growth chart is above 50, assign (–) if it's below 50, and assign (0) if it is exactly 50. We have $k_{(0)} = 0$, $k_{(+)} = 20$, and $k_{(-)} = 14$. . This makes $n = 34 - 0 = 34$.

Test statistic is $k_s = 20$, $n = 34$. Having as many as 20 successes out of 34 is consistent with the alternative hypothesis. So we continue with the next step.
Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 20) = 1 -$ binomcdf(34,.5,19) = 1 – 0.8042 = 0.1958 > α.
Retain the null.

Step 6: Conclusion in English. There is insufficient evidence (P-value =0.1958) to conclude that more than half of the white non-Hispanic girls exceed the 50^{th} percentile on the CDC growth chart.

6.31 **Let** π be the failure rate of the computer chips.

Step 1: H_0: Computer chips fail at the historic 8% rate: $\pi = 0.08$

Step 2: H_a: Computer chips fail at a higher rate: $\pi > 0.08$

Step 3: $\alpha = .05$

Step 4: In the sample of 100 computer chips, let k_s be the number that failed: $k_s = 15$

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 15) =$ 1 – binomcdf(100,0.08,14) = 1 – 0.9867 = 0.0133 < α.
Reject null hypothesis.

Step 6: Conclusion in English. There is sufficient evidence (P-value = 0.0133) to conclude that the computer chips are failing at a rate higher than usual.

6.33 Among the panel of 350, let π be the probability chosen juror is a woman.

Step 1: H_0: There was no bias against women by judge: $\pi = 102/350$

Step 2: H_a: There was bias against women: $\pi < 102/350$

Step 3: $\alpha = .01$

Step 4: In the jury of 100, let k_s be the number of women: $k_s = 9$

Step 5: P-value = $\Pr(K \leq k_s) = \Pr(K \leq 9) =$ binomcdf(100,102/350,9) = $9.08\text{x}10^{-7} < \alpha$.
Reject null hypothesis.

Step 6: Conclusion in English. There is very strong evidence (P-value = $9.08\ \text{x}10^{-7}$) to conclude sex bias in the choosing of jurors.

6.35

	Success	Failure	Totals
New Drug	22	4	42
Standard Treatment	14	7	21
Totals	62	11	63

Step 1: H_0: New drug is not better than standard treatment.

Step 2: H_a: New drug is better than standard treatment.

Step 3: $\alpha = .01$

Step 4: Among the 48 subjects in the failure group, let $k_s = 4$ be the number observed in the new drug group. Then $N = 63$, $N_1 = 42$, and $n = 11$.

Step 5: P-value = $\Pr(K \le k_s) = \Pr(K \le 4) = \sum_{k=0}^{4} \frac{\binom{42}{k}\binom{21}{11-k}}{\binom{63}{11}} = 0.0254 > \alpha$. Retain the null.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0254) to conclude that new drug treatment is better than the standard treatment.

6.37

	Flu	No flu	Totals
Shot	3	12	15
No shot	18	15	33
Totals	21	27	48

Among those in the flu population, let π be the proportion of those who are in the shot group.

Step 1: H_0: The flu shot does not reduce the risk of the getting the flu: $\pi = 15/48$.

Step 2: H_a: The flu shot reduces the risk of getting the flu: $\pi < 15/48$.

Step 3: $\alpha = .05$

Step 4: In the sample of 21 students who got the flu, let k_s be the sample count of those who received the shot: $k_s = 3$.

Step 5: P-value = $\Pr(K \le k_s) = \Pr(K \le 3)$ = binomcdf(21,15/48,3) = $0.0684 > \alpha$. Retain null.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0684) to conclude that the flu shot reduces the risk of getting the flu.

Note: The binomial exact test is not recommended because $20n = 20(21) > 48 = N$.

6.39

	Success	Failure	Totals
Acupuncture	7	21	28
Placebo	4	23	27
Totals	11	44	55

Among those in the success population, let π be the proportion of those who are in the acupuncture group.

Step 1: H_0: Acupuncture does not increase success rate: $\pi = 0.50$

Step 2: H_a: Acupuncture increases success rate: $\pi > 0.50$

Step 3: $\alpha = .05$

Step 4: In the sample of 11 subjects in the success group, let k_s be the sample count of those who received acupuncture: $k_s = 7$.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 7) = 1 - \text{binomcdf}(11,0.50,6) = 0.2744 > \alpha$. Retain null hypothesis.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.2274) to conclude that acupuncture drug treatment is better than the placebo.

6.41

	Success	Failure	Totals
Acupuncture	7	6	13
Placebo	4	13	17
Totals	11	19	30

Step 1: H_0: Acupuncture is same as placebo.

Step 2: H_a: Acupuncture is better than placebo.

Step 3: $\alpha = .05$

Step 4: Among the n = 11 subjects in the success group, let $k_s = 7$ be the number observed in the acupuncture group. Then $N = 30$, $N_1 = 13$, and $n = 11$.

Step 5: P-value = $\Pr(K \geq k_s) = \Pr(K \geq 7) = 1 - \Pr(K \leq 6) = 1 - \sum_{k=0}^{6} \frac{\binom{13}{k}\binom{17}{11-k}}{\binom{39}{11}} = 0.0927 > \alpha$.

Retain null hypothesis.

Step 6: Conclusion in English. There is insufficient evidence (P-value = 0.0927) to conclude that acupuncture drug treatment is better than the placebo.