Workshop Statistics: Discovery with Data, Second Edition

Topic 19: Confidence Intervals I: Proportions

Activity 19-7: Magazine Advertisements

(a) pages
(b) *Sports Illustrated*: □=54/116 = .4655
.4655 ± 1.96sqrt(.4655(1-.4655)/116) = (.375, .556)
Soap Opera Digest: □=28/130 = .2154
.2154 ± 1.96sqrt(.2154(1-.2154)/130)=(.145, .286)
(c) We are 95% confident that the above interval contains the proportion of pages with advertisements for all of the magazine's pages.
(d) If we were take many random samples from the magazine's population of pages, and construct a confidence interval for each sample, then 95% of the intervals would succeed in capturing the population proportion.
(e) Yes
(f) The sample proportion is the center of the interval.
(g) Answers will vary.

Activity 19-8: Phone Book Gender (*cont.*)

(a) number of first names: 36+36+77+14 = 163
(b) number of female names: 36+14 = 50; 50/163=.307
(c) .307 ± 1.645sqrt(.307(1-.307)/163) = (.247, .366)
(d) Often couples will only list the male names giving us an underestimate of the proportion of women in San Luis Obispo County.
(e) Answers will vary.

Activity 19-9: Closeness of Baseball Games

(a) observational units: games; variable of interest: margin of victory
(b) quantitative, no
(c) categorical, yes
(d) □=38/190 = .20; statistic since these 190 games are a sample of all games played.
(e) parameter
(f) 90%: .20 ± 1.645sqrt(.20(1-.20)/190) = (.152, .248)
95%: .20 ± 1.96sqrt(.20(1-.20)/190) = (.143, .257)
99%: .20 ± 2.576sqrt(.20(1-.20)/190) = (.125, .275)
(g) I'm 95% confident that the proportion of baseball games decided by one run is

between .143 and .257. I could make this interval slightly larger or smaller by changing the confidence level.
(h) No, it's more of a convenience sample, but there might not be bias involved with those days and the margin of victory (especially not being close to the beginning or end of the season).
(i) Yes, then July, 26-August, 8 definitely behave differently than the rest of the season with regard to game time temperature.

Activity 19-10: Home Field Advantage

(a) □will be the center of the interval = .521
(b) .521(190) = 99 games
(c) 99% confidence: .521 $\pm$ 2.576sqrt(.521(1-.521)/190) = (.428, .614)
(d) This indicates that .5 is a plausible value for θ which corresponds to no home field advantage (the home team and the visiting team are equally likely to win a game).

Activity 19-11: Random Babies (*cont.*)

(a) Answers will vary from class to class for the simulation. These are intended to be sample answers.
Here are the simulation results report in the Topic 14 Inclass Solutions.

# matches	0	1	2	3	4	total
count	35	37	22	0	6	100
proportion	.35	.37	.22	.00	.06	1.0

Proportion of repetitions with no mother getting the correct baby: .35
(b).35 $\pm$ 1.96sqrt(.35(1-.35)/100) = (.257, .443)
(c) We are 95% confident that the actual proportion of mothers getting the correct baby when they are returned at random is between .257 and .443.
(d) θ=.375
(e) Yes, .375 is contained in this interval.
(f) 95%
(g) .35 $\pm$ 1.282sqrt(.35(.65)/100) = (.289, .411). This interval also succeeds, but it's much more conceivable that this interval would miss the parameter (20% of the time instead of 5% of the time).

Activity 19-12: Television Magic

(a) population = all *Nick-at-Nite* viewers, parameter = proportion of *Nick-at-Nite* viewers who vote for Samantha (or could choose Jeannie).
(b) No, the sample consisted of people who voluntarily called in to respond to the poll.

(c) □=810,000/(810000+614000) = .5688
97% confidence interval: .5688 ± 2.17sqrt(.5688(1-.5688)/1,424,000) = (.579, .5697)
(d) Because the sample size is so large.

Activity 19-13: Cat Households (*cont.*)

(a) .273 ± 2.576sqrt(.273(1-.273)/80000) = (0.269, 0.277)
(b) We are 99% confident that the proportion of all American households that own a pet cat is between .269 and .277
(c) If we were to repeatedly take samples from the population of American households and calculate a confidence interval for each sample, we'd expect 99% of the intervals to capture the population proportion.
(d) Would need to know if the sample was selected at random.

Activity 19-14: Charitable Contributions

(a) parameter = proportion of all American households that claim to make a financial contribution to charity.
(b) *n*=250, .685 ± 2.576sqrt(.685(1-.685)/250) = (.609, .761)
(c) *n*=500, .685 ± 2.576sqrt(.685(1-.685)/500) =(.631, .742)
(d) *n*=1000, .685 ± 2.576sqrt(.685(1-.685)/1000) =(.647, .723)
(e) *n*=2000, .685 ± 2.576sqrt(.685(1-.685)/2000) =(.658, .712)
(f) Half-widths: .076, .054, .038, .027
The half-width decreases as increase the sample size. In fact, each half-width is the previous divided by 1.41.
(g) No, it cuts it by a factor of sqrt(2) instead of by 2.
(h) *n*=2719: .685 ± 2.576sqrt(.685(1-.685)/2719) = (.662, .708)
We are 95% confident that the proportion of all American households that claim to make a financial contribution to charity is between .662 and .708.

Activity 19-15: Marriage Ages (*cont.*)

(a) 22 marriages
(b) 16 of the 22 brides are younger (.727)
(c) .727±1.645sqrt(.727(1-.727)/22) = (.571, .883)
(d) .727±1.96sqrt(.727(1-.727)/22) = (.541, .913)
(e) .727±2.576sqrt(.727(1-.727)/22) = (.483, .972)
(f) Only the last interval contains .5. The other intervals fall above .5.
(g) We are 95% confident that the proportion of marriages with the bride younger is above .5, though we are not 99% confident of this. Thus, we have some evidence that more than half of the marriages in this county have an older groom.

Activity 19-16: Newspaper Reading

(a) .430 $\pm$ 1.645 sqrt(.430(1-.43)/1870) = (.412, .449)
(b) No
(c) Yes, we are 95% confident that the proportion of all adult Americans who read a newspaper daily is below .5 since our confidence interval is from .412 to .449.

Activity 19-17: Emotional Support (*cont.*)

(a) Hite margin of error = 1.96sqrt(.96(.04)/4500) = .006, interval (.954, .966)
ABC News/*Washington Post* margin of error = 1.96sqrt(.44(.56)/767) = .035, interval (.404, .474)
(b) These confidence intervals are not at all similar and do not overlap.
(c) The Hite poll has a smaller margin of error
(d) We believe the ABC News/*Washington Post* poll more since they took a random sample and thus the results better reflect the population of all American women.

Activity 19-18: Veterans' Marital Problems

(a) A 95% confidence interval is .370 $\pm$ 1.96sqrt(.370(1-.370)/2101) = (.349, .391)
(b) No since Vietnam veterans are not a good sample of all middle-aged American men. It's plausible that they have a higher divorce rate than the rest of the population.

Activity 19-19: E-Mail Usage (*cont.*)

Answers will vary.

Activity 19-20: Critical Values

(a) 85% confidence corresponds to 92.5% below. Looking this up in Table II, we find z*=1.44
(b) 97.5% corresponds to 98.75% below, so z*=2.24
(c) 51.6% corresponds to 75.8% below, so z*=.70

Activity 19-21: Incorrect Conclusions (*cont.*)

(a) Andrew since his interval is not centered at .4.
(b) Andrew: .62; Becky: .695
(c) Andrew: .062, Becky: .084
(d) We do not know the confidence level used by each researcher.
(e) Andrew: z*=1.28 so 80%; Becky: z*=2.58 so 99%
(f) Andrew: half-width= .051, □=.584
Becky: half-width= .026, □=.576

(g) The interval with the smaller half-width (Becky) goes with the larger sample size.
(h) Both intervals have the same "chance," 90%, of capturing the population proportion since they have the same confidence level.