a&f

EXERCISES

2 1. a. discrete take only certain values on scale
continuous can take any value on scale
b. qualitative: order on scale has no meaning
quantitative: order on scale matters
c. nominal: qualitative (numbers don't mean anything)
ordinal: categories naturally ordered (quantitative)
→ matter because different statistics used.

2. a. quant. b. qual. c. qual. d. quant. e. qual. f. quant. g. qual. h. quant.
3. a. ordinal b. nominal c. interval d. nominal e. nominal f. ordinal
g. interval h. ordinal i. nominal j. interval/ordinal (depends on test) k. nominal ✓
5. a. interval b. ordinal c. nominal ✓
7. ordinal ✓
9. ~~a~~ b c d ~~e~~ f ~~g~~ ✓
11. not probability sample (based on volunteers) ✓
13. a. pick 5 random numbers from list
b. pick 1 random number from first ~~300-120~~ 60 pages and then the same on every 60th. ✓
17. · actual question asked ; sampling error ; maybe different methods; quite accurate ✓
20. a. nonprobability (preselect ♀ at streetcorner); ⊕ easy ; ⊖ non-repr.; need to know population for quota
b. cluster: take random within naturally occurring cluster; same with strata; the random selection within the groups is where it differs.
21. · response rate (≱50%) ; major and slight as one categ. ✓
23. ~~a~~ b c ~~d~~ ✓
27. if know one member of sample → know all ; not so in true random sampling.
→ two consequent entries cannot be in same sample; thus not equal likelihood ←--

2004 Ruedin, D. & Okamoto-Kaminski, K. in Oxford

REFERENCE:
Agresti, A. & Finlay, B. (1997) Statistical Methods for the Social Sciences, Upper Saddle River, Prentice Hall. (3rd edition, ISBN= 0135265266)

3

2

1 c/d
5 b/c → f
7 b/c
9 a/b
11 a/b
13 b/c
15 a/b
17 a/b
19 b
21
27
29
31 a–i
33 a/b
37
39
41
43 a–c
45
49
57
59
61
63
65
69 a/b

1. c.

#	#	#	#	#	#
1	23.6	4	15.1	7+	1.4
2	31.2	5	6.7	8	
3	16.9	6	2.2		

$$\bar{Y} = \frac{\Sigma Y_i}{n} = \frac{1\cdot 23.6 + 2\cdot 31.2 + 3\cdot 16.9 + 4\cdot 15.1 + 5\cdot 6.7 + 6\cdot 2.2 + 8\cdot 1.4}{23.6 + 31.2 + 16.9 + 15.1 + 6.7 + 2.2 + 1.4} =$$

$$\frac{23.6 + 62.4 + 50.7 + 60.4 + 33.5 + 13.2 + 11.2}{97.1} = \frac{255.0}{97.1} \approx \underline{\underline{2.6}}$$

d. $n = 97.1$ → median @ 48.55; $48.55 - 23.6 = 15. > 31.2$ → $\underline{\underline{2}}$

→ mode @ $\underline{\underline{2}}$

5. a.

0	4 6 7 9
1	1 3 3
2	0
3	9
4	4

b. $\bar{Y} = \frac{\Sigma Y_i}{n} = \frac{4+6+7+9+11+13+13+20+39+44}{10} = \frac{166}{10} = 16.6 \sim 17$

c. median $= \frac{11+13}{2} = \underline{\underline{12}}$

d. $s = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{12.6^2 + 10.6^2 + 9.6^2 + 8^2 + 6^2 + 4^2 + 4^2 + 3^2 + 12^2 + 27^2}{9}}$

$= \sqrt{\frac{169 + 121 + 81 + 64 + 36 + 16 + 16 + 9 + 144 + 729}{9}} = \sqrt{\frac{1385}{9}}$

$\approx \sqrt{154} \simeq \underline{\underline{12.5}}$

e.

	0	4 6 7 9
8 7 5	1	1 3 3
4 4 0	2	0
2 1	3	9
0	4	4
5	5	

$\bar{Y}_b = \frac{15+17+18+20+24+24+31+32+40+55}{10} = \underline{\underline{27.6}} \sim 28$

$\text{median}_b = \underline{\underline{24}}$

$S_b = \sqrt{\frac{\Sigma(Y_i - \bar{Y}_b)^2}{n-1}} = \sqrt{\frac{13^2 + 11^2 + 10^2 + 8^2 + 4^2 + 4^2 + 3^2 + 4^2 + 12^2 + 27^2}{9}}$

$= \sqrt{\frac{169+121+100+64+16+16+9+16+144+729}{9}} = \sqrt{\frac{1394}{9}}$

$= \sqrt{155} \simeq \underline{\underline{12.5}}$

$\bar{Y}_b > \bar{Y}_a$: thus stayed longer with same variance...

3

NB: The key is wrong for this question. I have contacted Agresti and confirmed that the key is wrong.

3

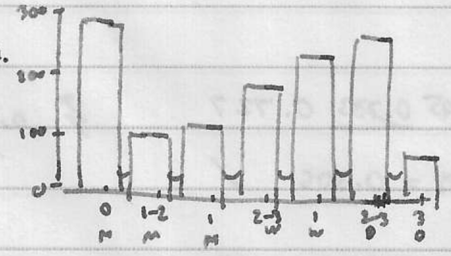
5.f. $median_c$ {15 17 18 20 24 24 31 32 40 40 55} = 24

$\bar{Y}_c \geq \frac{\sum Y_i}{n} \geq \frac{316}{11} \geq 28.7$

$\cancel{18.7}$

~~A-ve~~ = wrong

7. a.



b. median : 2-3 times a month

mode : not at all

c. $\bar{Y} = \frac{\sum(Y_i)}{n} = \frac{0 \cdot 292 + 0.1 \cdot 99 + 1 \cdot 108 + 2.5 \cdot 181 + 4.3 \cdot 233 + 10.8 \cdot 265 + 17 \cdot 72}{292 + 99 + 108 + 181 + 233 + 265 + 72}$ (9.9 + 108 + 452.5 + 1001.9 + 2865 + 1224) $= \frac{5661.3}{1250} = 4.5$ ✓

9. a. $\bar{Y} = \frac{\sum Y_i}{n} = \frac{33 + 16 + 304 + 2 + 11}{5} = \frac{366}{5} = 73.2$ ✓

b. med = 16 ; california as an outlier?

11. a. $\bar{Y}_a = 612.84$

$med_a = 515$ $\bar{Y} >$ med thus skewed to right

b. outlier affects: mean, standard deviation, maximum, range; others not so ✓

13. b. $\bar{Y}_1 = \frac{\sum Y_i}{n} = \cancel{\frac{60+47+77+64+41+44+42+68+51+31+77+60+40+}{19}}$ $\frac{88+84+70+77+77+81}{6}$

$= \cancel{\frac{1199}{19} = 63.1}$ $\frac{477}{6} = 79.5$ $\cancel{79.5}$

$\bar{Y}_2 = \frac{\sum Y_{i_2}}{n} = \frac{38 + 44 + 39 + 39 + 24 + 26 + 31 + 34 + 28 + 32 + 38}{11} = \frac{373}{11} = 33.9$ ✓

c. $med_1 = \cancel{64}$ n=19→10 3.5 $\frac{77+81}{2} = 79$

$med_2 =$ 34 n=11→6 ✓

15. a. $\bar{Y} = \frac{\sum Y_i}{n} = \cancel{\frac{57.9 \cdot 32368 + 8 \cdot 18660}{57.9 + 8}}$ (1874107.2, 149280) $= \cancel{\frac{2023387.2}{65.9}} = \cancel{\$30703}$ (USD)

$\frac{57.9 \cdot 40708 + 8 \cdot 25409}{65.9}$ (2356993.2, 203272) $= \frac{2560265.2}{65.9} = 38850.8$ USD ✓

b. $\bar{Y} = \frac{\sum Y_i}{n} = \frac{57.9 \cdot 40708 + 8 \cdot 25409 + 5.9 \cdot 30291}{71.8}$ (178716.9) $= \frac{2738982.1}{71.8} = 38147.4$ USD ✓

17. a. assume

thus $\bar{Y} >$ med

b. (i) x never negative (iv) probably $\bar{Y} < 1000000$ which cannot be. ✓

(ii) x too narrow

(iii) ✓

3

19. a.

5	7 7 9
6	1
7	
8	4 5 8 8

19. b. $\bar{y} = \frac{\sum y_i}{n} = \frac{0.884 + 0.884 + 0.579 + 0.591 + 0.578 + 0.842 + 0.611 + 0.856}{8}$

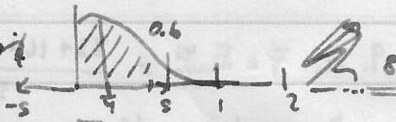
$= \frac{5.825}{8} = 0.728$

med $= \frac{4+5}{2} = \frac{0.611 + 0.842}{2} = 0.733 \quad 0.727$ ⚡ ~~0.726~~ ~~(rounding error)~~

range $= [0.578, 0.884] \rightarrow 0.884 - 0.578 = 0.306$ ✓

21. empirical rule assumes normal distrib., so not appropriate

$\bar{y} = 0.145$, $s = 0.457$ } within 1s : [0 − 0.602] : ~~66%~~
thus within 2s : [0 − 1.05] : 88.8% + 88% = 99.6%



only ⊙ within range → 88.8% ?

27. $\bar{y} = 1921$, $s = 11{,}495$

min = 4
Q1 = 256
med = 530
Q3 = 1105
max = 320,000

→ $\bar{y}$ > med : skewed to right ; v. much so ($\bar{y}$ > Q3) → maybe outliers?

29. $\bar{y} = 7.1$, $s = 6.2$ } skewed to right ; not much left (all values > 0)

31. a. $\bar{y} = 90$, $s = 10$ skewed to left
b. normal
c. skewed to ~~left~~ right
d. skewed to ~~right~~ left
e. skewed to right
f. skewed to left? ✓
g. $\bar{y} = 17$, med = 10 → skewed to right ✓
h. because mode(j−1) → skewed to right
i. U-shaped ✓

33. a. mode = (@51%) not far enough
b. $\bar{y}$ no; not continuous
med = 50% = not far enough ✓

37. 1964: mode = med (2), $\bar{y} = 2.8$ → skewed to right
$s = 2.4$ v. much so (can't go < 0) ✓

39. $\bar{y} = 80$ impossible : −20 (never negative) ✓

41. nearly all : within 3s of $\bar{y}$ | $\bar{y} = 700$ $s = 100$
700 − 300 to 700 + 300 = \$[400 − 1000] ✓

43. a. 68% within s : 120000 ± 40000 → 80000 − 160000
95% within 2s : 120000 ± 80000 → 40000 − 200000
almost all within 3s : 120000 ± 120000 → ≤ 240000

3

5

43. b. $\bar{y} + \frac{1}{2}s$ = 120000 + 20000 = 140 000

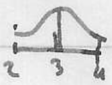
c. skewed to right (ie. more cheaper houses)

~~45.~~ → ~~connectt~~ as outlier ←

49. a. $\bar{y}$ for sample mean ; μ for population mean

b. s for " standard deviation; σ " standard deviation

57. $\bar{y}$ = 145 000, med = 44000 } because assume distrib. to be skewed to right (wealthy people)
↓ $\bar{y}$ > med

59.  $3s = 1 \rightarrow s = \frac{1}{3}$

61. a. b. not
c. bimodal (always wrong; not wrong at all)
d. → not (ordinal)

63. needs weighing

~~65.~~ → just multiply by 1.6

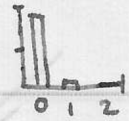
69. a. (i) $\frac{1}{(2s)^2} = \frac{1}{4}$
(ii) $\frac{1}{(3s)^2} = \frac{1}{9}$
(iii) $\frac{1}{(10s)^2} = \frac{1}{100}$ → proportion

b. 2s: 1/4 ~~.25~~ 0.25 ← max possible (any curve)
 $\frac{95}{2}$ ~~= 47.5~~ 0.05 ← because of shape of curve ↓ normal...
~~target given information out shape of curve~~

4: 1b	~~19a-e~~	37
3a/b	21a-d	39a-c
5a-c	23a-c	43a,b,d
7a-d	25a-b	47
9	27a-e	49
11	29a-d	51a-b-c
13 a/b	31 a-b	53a-g
15a-d	33	~~55a-c~~
17	35a-e	~~57~~

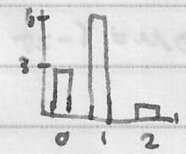
6

4

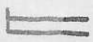
1\. a. 

b. $Y = \frac{\sum x_i}{n} > \frac{0 \cdot .9 + 1 \cdot \cancel{0}.08 + 2 \cdot \cancel{0.02}}{.9 + .08 + .02} = \underline{\underline{.12}}$ (.08, .04) ✓

3\. b. n=10 $\underbrace{2 \cdot \frac{1}{10}}_{p(2)=1/10} + \underbrace{1 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10}}_{p(1) = 6/10} + \underbrace{0 \cdot \frac{1}{10} + 0 \cdot \frac{1}{10} + 0 \cdot \frac{1}{10}}_{p(0)=3/10}$ $\frac{8}{10} = \frac{8}{10} = \underline{0.8}$

a. 

5\. a. [0 .. 9]

b.  p (i) = 0.1

c. 4.5

d. 2.9 (0.7 too close, 9 never, 5 = max distance) ✓

7\. a. ~~50% 34% 28%~~ z = 1 → p = 0.1587

b. "

c. 0.2514 (2 tables)

d. 0.0099 ✓

9\. a. $\mu + z\sigma \geq .01$ → 2.33 (0.0099)

b. 0.025 → 1.96

c. 0.05 → 1.65 : 1.64 (same)

d. 0.1 → 1.29 : 1.28 (closer)

e. 0.25 → 0.68 : 0.67 (closer)

f. 0.5 → 0 ✓

11\. a. 1.28

b. 1.65

c. 2.05

d. 2.33 ✓

13\. b. 25% (Q1) z = .67

a. 75% (Q3) ✓

17\. = 13. (± 0.67) ✓

15\. a. z = 2.1 → p = 0.0179 → 1-p = 98.3%

b. "

c. 1.7% (0.9821)

d. 1-2p = 96.4% (0.9642) ✓

19\. a. $\bar{Y} = 100$ $\sigma = 16$

Y=120 ~~z=1.4σ~~ → ~~0.0808~~

$\bar{Y} + 1.25\sigma$ → z-score (table) = 0.1056

b. 1 - 0.1056 = 0.8944 (less than 80)

c. z = 1.28

→ Y = 100 + 1.28·16 = 120.48

d. Y = 100 - 1.28·16 = 79.52

e. Q1 = μ - zσ = 100 - 0.67·16 = 89.28

Q3 = μ + zσ = 100 + 0.67·16 = 110.72

p = 0.25 → z = 0.67

med = mean (normal distr.)

f. range = 110 - 89 = 21 ✓

21. a. $\mu = 16 \quad \sigma = 5$

$20 = \mu + 0.8\sigma \rightarrow .2119$ (table)

b. $Q1 = \mu - z\sigma = 16 - 0.67 \cdot 5 = 19.35$

$Q3 = \mu + z\sigma = 16 + 0.67 \cdot 5 = 12.65$

$IQR = 6.8$ $(Q3 - Q1)$

c. $Y = 20 \quad \sigma = 5$

$p = 0.05 \rightarrow z = ?\ 1.65$ $\quad Y = \mu + z\sigma \Rightarrow \mu = Y - z\sigma$

$20 = \mu + 1.64 \cdot 5 \rightarrow \mu =$ $\quad$ 11.8 !

d. skewed to right. ✓

23. $\mu = 750 \quad \sigma = 150$

a. $1000 = 750 + 1.67\sigma \rightarrow$ ~~0.0548~~ 0.0475

b. $600 = 750 - \sigma \rightarrow 0.1587$

c. $[500 - 1000] = \mu \pm 1.67\sigma \Rightarrow 1 - 2 \cdot 0.0475 = 1 - 0.095 = 0.905$ $\quad$ 0.904 (calculated! with 0.048)

25. $\mu = 800 \quad \sigma = 500$

a. $z = 1 \rightarrow p = 0.1587 \rightarrow 16\%$ $\quad$ 0.159 (question asks %)

b. not, μ not much bigger than σ

27. a. $p(\text{head}) = 1/2$

$p(\text{tails}) = 1/2$

b. $p(H) = 1/2$, $p(T) = 1/2$ $\left\{ \begin{array}{l} p(HH) = 1/4 \\ p(HT) = 1/2 \\ p(TT) = 1/4 \end{array} \right.$

c. $p(HHH) = 1/8$
$p(HHT) = 3/8$
$p(TTH) = 3/8$
$p(TTT) = 1/8$

d. $p(HHHH) = 1/16$
$p(HHHT) = 4/16$
$p(HHTT) = 6/16$
$p(HTTT) = 4/16$
$p(TTTT) = 1/16$ ✓

e. approaches normal distribution

29. $\mu = 13.6 \quad \sigma = 3$

a. $n = 9 \quad \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} \quad \frac{3}{3} = 1 \quad \bar{Y} = \mu$

b. $n = 36 \quad \frac{3}{6} = \frac{1}{2}$ "

c. $n = 100 \quad \frac{3}{10}$ " ✓

31. $\mu = 250 \quad \sigma = 75$

a. $Y = 300 = \mu + z\sigma \rightarrow z = \frac{300 - \mu}{\sigma} = \frac{50}{75} = \frac{2}{3} \rightarrow p = .2546$ (table)

b. $\bar{Y} = \mu \quad \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{75}{3} = 25$

$300 = \bar{Y} + z\sigma_{\bar{Y}} \rightarrow z = \frac{300 - \bar{Y}}{\sigma_{\bar{Y}}} = \frac{50}{25} = 2 \rightarrow p = 0.0228$ ✓

4

8

33. $\mu = 2.6$ $\sigma = 1.5$

$n = 225$ $\rightarrow \bar{Y} = \mu$; $\sigma_{\bar{Y}} = \frac{1.5}{\sqrt{225}} = \frac{1.5}{15} = 0.1$

$Y = \mu_{\bar{Y}} + z\sigma_{\bar{Y}}$ $= \cancel{2.6+0.1 \cdot 0.1 = 2.61}$

0.4602

$\cancel{2.6} \pm 1 \cdot 0.1$

within 0.1 $\rightarrow \pm 1$ s $\rightarrow z = 1$

$p = 0.1587 \cdot 2$

0.3174

outside ↓

inside = 1 − p = 0.6826 ✓

35. $\mu = 100$ $\sigma = 15$

a. $z = 0$ $p = 0.5$

b. $z = \frac{Y-\mu}{\sigma} = \frac{3}{15} = \frac{1}{5} = 0.2 \rightarrow p = .4207$

c. $= 2(0.5 - .4207) = 2 \cdot 0.0793 = 0.1586$ ✓

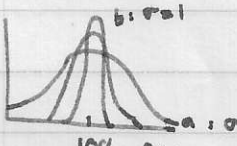
symmetrical ↗ | 50% less than mean ↓ greater than 103

d. $Y = 90 = \mu + z\sigma$ $z = \frac{Y-\mu}{\sigma} = \frac{-10}{15} = \frac{2}{3} = .6\overline{7} \rightarrow p = .2514$ ✓

↳ not surprising

e. normal standard; mean = 0 ; s = 1

37.



$\rightarrow \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{225}} = \frac{15}{15} = 1$ ✓

$\Rightarrow \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{25}} = \frac{15}{5} = 3$

39. $\sigma = 240$ aim: 50

$z = \frac{Y-\mu}{\sigma}$ 1−2p

a. $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{240}{6} = 40$ $z = \frac{50}{\sigma} = \frac{50}{40} = 1.25 \rightarrow p = .1056 \rightarrow$ ~~.8944~~ .7888

b. $= \frac{240}{8} = 30$ $z = \frac{50}{\sigma} = \frac{5}{3} = 1.6\overline{7} \rightarrow p = .0465 \rightarrow$ ~~.9535~~ .9070

c. $= \frac{240}{12} = 20$ $z = \frac{50}{\sigma} = 2.5 \rightarrow p = .0062 \rightarrow$ ~~.9938~~ .9876

43. a. $\bar{Y} = .5$ d. assumedly $\mu = .5$ ✓

b. $\bar{Y} = .6$ ✓

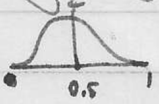
$\sigma = ?$ $z = 1$ $p = 0.158$

$\sigma = \sqrt{\frac{(0-0.5)^2 + (1-0.5)^2}{2}} = \sqrt{\frac{0.25 + 0.25}{2}} = \sqrt{\frac{0.5}{2}} = \sqrt{0.25} = 0.5$

47. a v c v d ✓

49. a ✓

sample means



∴ prob. distr. $\rightarrow s = 1 \rightarrow z = 1$ $p = 0.1587$

51. $\mu_A = 500$ $\mu_B = 450$ $\sigma = 100$

a. A: $400 = \mu \cdot z\sigma$ $z = \frac{-100}{100} = -1 \rightarrow p = .1587$

B: $400 = \mu \cdot z\sigma$ $z = \frac{-50}{100} = -\frac{1}{2} = -.5 \rightarrow p = .3085$

C: A: $z = \frac{-200}{100} = -2 \rightarrow p = .0228$

B: $z = \frac{-150}{100} = -1.5 \rightarrow p = .0668$

b. $\frac{1}{3}$ A $\frac{2}{3}$ B $\Rightarrow$ 667

$\rightarrow \frac{.0668}{.0668 + .0228} = \frac{.0668}{.0896} = 0.750$ B

$\boxed{4}$ 9

53. $N = 50{,}000$ $\mu = 60$ $\sigma = 16$

 $n = 100$ $\bar{Y} = 58.3$ $s = 15$

a. skewed to left $\mu = 60$ $\sigma = 16$

b. similar $\bar{Y} = 58.3$ $s = 15$

c. normally $\bar{Y} = \mu$ $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{10} = 1.6$

d. whatever (a point) $\bar{Y} = \mu$ $\sigma_{\bar{Y}} = \sigma = 16$ $\bar{Y} = 60$ ✓ $\sigma = 16$ ✓ ?? skewed left =

e. v. close to normal $\bar{Y} = \mu$ $\sigma \approx 0$

f. normal $\bar{Y} = \mu$ $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = 1.6$ ✓

g. 1. $Y = \mu + 2\sigma$ $\rightarrow z = \frac{Y-\mu}{\sigma} = \frac{40-60}{16} = \frac{-20}{16} = \frac{-5}{4} = -1.25$ $\rightarrow$ $p = 0.0156$ not unusual

2. but ~~n=100~~ $\cancel{\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{100}} = \frac{16}{10} = 1.6}$ sample mean $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = 1.6$

$z = \frac{-20}{1.6} = -12.5$ times st. dev. from true mean ✓

55. 1. $\sigma^2 = \sum (Y-\mu)^2 p(Y) = (0-.5)^2 .5 + (1-.5)^2 .5 = 2(.25 \cdot \frac{1}{2}) = \underline{.25}$

$0-.5$ $\quad \overset{\shortparallel}{.5}$

$1-.5$

2. $\sigma^2 = \left[\sum Y^2 p(Y)\right] - \mu^2 = \cancel{\ldots - .5^2 = -.25 + .5 = .25}$

$0 \quad \overset{\shortparallel}{.5} \quad \overset{\shortparallel}{.5^2} \qquad 0 + .5 \quad - .25 \quad = \underline{.25}$ ✓

1

(...)

5. 1	15	29	55
3a-c	17	31	57a,b
5a-d	19a,b	33	
7a-b	21a,b	35a,b	
9a-c	23a,b	45a,b	
11a-b	25	49	
13a-b	27a-d	51	

5

10

1. $\bar{Y} = \frac{\sum Y_i}{n} = \frac{18}{6} = 3$

$s = \sqrt{\frac{\sum(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{2^2+1^2+3^2+3^2+3^2+4^2}{5}} = \sqrt{\frac{4+1+9+9+9+16}{5}} = \sqrt{\frac{49}{5}} \approx \sqrt{10} \approx \underline{\underline{3.3}}$ ✓

3. a. $\bar{Y} = 4.171$; $\hat{\sigma}_{\bar{Y}} = 0.0259$

interval: $\bar{Y} \pm z\hat{\sigma}_{\bar{Y}} = \bar{Y} \pm 1.96\,\sigma_{\bar{Y}} = 4.171 \pm 1.96 \cdot 0.0259 = \underline{\underline{4.171 \pm 0.05}} \rightarrow (4.12, 4.22)$ ✓

b. $\bar{Y} \pm z\hat{\sigma}_{\bar{Y}} = 4.171 \pm 2.56 \cdot 0.03 = 4.171 \pm 0.07 \rightarrow (4.10, 4.24)$ ∴ increased. ✓

c. interval ✓

5. a. $\bar{Y} = 1.314$; SE = 0.215 ; SD = 5.42 ✓

b. $SE = \frac{\sigma}{\sqrt{n}} = \frac{SD}{\sqrt{n}} = \frac{5.42}{\sqrt{637}} \approx \frac{5.42}{25.2} \approx \underline{0.215}$ ✓

c. $\bar{Y} \pm z\sigma_{\bar{Y}} = 1.31 \pm 1.96 \cdot 0.215 \approx 1.31 \pm 0.42 \rightarrow (0.89, 1.73)$ ✓

d. $\bar{Y} \pm z\sigma_{\bar{Y}} = 1.31 \pm \cancel{1.96}\, 2.58 \cdot 0.215 \approx 1.31 \pm 0.55 \rightarrow (0.76, 1.86)$ ✓

7. $n = 100$ $\bar{Y} = 5.3$ $\sigma_{\bar{Y}} = 1.0$

a. $\bar{Y} \pm z\underbrace{\sigma_{\bar{Y}}}_{=1.0}$ $5.3 \pm 1 \rightarrow (4.3, 6.3)$

b. double precicion: quadruple sample size.

$\underline{400}$ $\bar{Y} \pm 2\frac{s}{\sqrt{n}}$ $\bar{Y}\sqrt{n} = 2s \rightarrow \sqrt{n} = \frac{2s}{\bar{Y}} \rightarrow n = \left(\frac{2s}{\bar{Y}}\right)^2$ $2s \rightarrow 2s$; $n=1 \rightarrow n = 2^2 = \underline{4}$

9. a. (1) $\bar{Y} = 4.1$ $\overset{SD}{\hat{\sigma}_{\bar{Y}}} = 3.0 \xrightarrow{n=60} \overset{SE}{\sigma_{\bar{Y}}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{7.8} \sim 0.38$

$\bar{Y} \pm z\hat{\sigma}_{\bar{Y}} = 4.1 \pm 1.96 \cdot 0.38 = 4.1 \pm 0.75$ (3.35, 4.85)

(2) wider; z is larger (always...)

b. $SD = 6.0 \rightarrow \sigma_{\bar{Y}} = \frac{6}{7.8} \sim 0.77$

$\bar{Y} \pm z\sigma_{\bar{Y}} = 4.1 \pm 1.96 \cdot 0.77 = 4.1 \pm 1.51$ (2.59, 5.61) ; doubles spread (width)

c. $n = 240$ (4x(a)) $\sigma_{\bar{Y}} = \frac{3}{\sqrt{240}} \approx \frac{3}{15.4} \sim 0.19$

$\bar{Y} \pm z\sigma_{\bar{Y}} = 4.1 \pm 1.96 \cdot 0.19 = 4.1 \pm 0.38$ (3.72, 4.48) ; ½ the spread (width) ✓

11. a. $\bar{Y} = 2.8$ SE = 0.05

$\bar{Y} \pm z\sigma_{\bar{Y}} = 2.8 \pm 1.96 \cdot 0.05 \approx (2.7, 2.9)$ ✓

b. $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$ [n = 1964]

$\sigma = \sigma_{\bar{Y}} \cdot \sqrt{n} \approx 0.05 \cdot 44 = \underline{2.2}$; not normal ($\bar{Y}$, not much more than 1 s away from 0). ✓

5 11

13\. a. $n = 987$

$A = 17$ (yes) $= 0.02$ $\left(\frac{A}{n}\right)$

$B = 970$ (no) $= 0.98$ $\left(\frac{B}{n}\right)$

$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.02 \cdot 0.98}{987}} = \sqrt{\frac{0.0196}{987}} = \sqrt{0.00002}$ ~ 0.005

b. $\hat{\pi} \pm z\hat{\sigma}_{\hat{\pi}} = 0.02 \pm 1.96 \cdot 0.005 = 0.02 \pm 0.01 \rightarrow (0.01, 0.03)$ ✓ < 0.05

so, yes ✓

15\. a. $n = 1958$

$Y = 1425$ $\pi \Rightarrow \frac{1425}{1958} = 73\%$

$73\% \pm 2\%$

$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$ $\sqrt{\frac{0.73 \cdot 0.27}{n}} = \sqrt{\frac{0.1971}{1958}} = \sqrt{0.0001} = 0.01$ ✓

$\pi \pm z \cdot \hat{\sigma}_{\hat{\pi}} = 0.73 \pm 1.96 \cdot 0.01 = .73 \pm .02$ $\rightarrow = 2\%$

(assumed 95%)

17\. $\hat{\pi} = 25.5\%$ $n = 42000$ $z = 2.58$ $\rightarrow \hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = \sqrt{\frac{0.26 \cdot 0.74}{42000}} = \sqrt{\frac{.1924}{42000}} = \sqrt{0.000005}$ ~ 0.00225

$\pi \pm z\hat{\sigma}_{\hat{\pi}} = 0.25 \pm 2.58 = 0.25 \pm 0.006$ $(\rightarrow 0.249, 0.261)$ ✓

19\. a. $n = 1400$

$D = 742$ $\pi_D = \frac{742}{1400} = 0.53$

$R = 658$ $\pi_R = \frac{658}{1400} = 0.47$

$\sigma_{\pi_D} = \sqrt{\frac{\pi_D(1-\pi_D)}{n}} = \sqrt{\frac{0.53 \cdot 0.47}{1400}} = \sqrt{\frac{0.2491}{1400}} = \sqrt{0.0002} \approx 0.015$

$\hat{\pi} \pm z\hat{\sigma}_{\hat{\pi}} = 0.53 \pm 1.96 \cdot 0.015 = 0.53 \pm 0.03$ $\rightarrow$ ~~Len~~ $(0.50, 0.56)$ $\rightarrow$ just enough.

b. $\hat{\pi} \pm z\sigma_{\hat{\pi}} = 0.53 \pm 2.58 \cdot 0.015 = 0.53 \pm 0.04$ $(0.49, 0.57)$

$\hookrightarrow$ not always sure enough that $D > R$. ✓
(ie. $R > D \neq$ unplausible)

21\. a. $n = 577006$

$B = 412878$

$\pi_B = \frac{B}{n} = \frac{412878}{577006} = 0.72$

b. $\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.72 \cdot 0.28}{577006}} = \sqrt{\frac{0.2016}{577006}} = \sqrt{0.0000003} = 0.0005$

$\hat{\pi} \pm z\hat{\sigma}_{\hat{\pi}} = 0.72 \pm 2.58 \cdot 0.0005 = 0.72 \pm 0.001$ $(0.719, 0.721)$ > 0.5 $\therefore$ yes, majority

c. no measurement error, no non-response ✓

5

23. a. $n = 400$

$J = 160$

$S = 240$

$\pi_j = \frac{160}{400} = 0.4$

$\sigma_\pi = \sqrt{\frac{0.4 \cdot 0.6}{400}} = \sqrt{\frac{0.24}{400}} = \sqrt{0.0006} \approx 0.025$

$\hat{\pi} \pm z\sigma_\pi = 0.4 \pm 2.58 \cdot 0.025 = 0.4 \pm 0.06$ $(0.34, 0.46)$; will lose since < 0.5

b. $n = 40$

$J = 16$

$S = 24$

$\pi_j = 0.4$ $\sigma_\pi = \sqrt{\frac{0.24}{40}} = \sqrt{0.006} \approx 0.08$

$\hat{\pi} \pm z\sigma_\pi = 0.4 \pm 2.58 \cdot 0.08 = 0.4 \pm 0.21$ $(0.19, 0.61)$; can't say

25. $n = 30$

$O = 3$

$\cancel{\pi_O = \frac{3}{30} = \frac{1}{10} = 0.1}$ $\boxed{\pi(1-\pi) \hat{=} 0.25}$ (conservative estd. if no data)

$n = \pi(1-\pi)\left(\frac{z}{B}\right)^2$ $B = 0.07$, $z = 1.96$

$= \frac{\cancel{0.1 \cdot 0.9}}{0.25}\left(\frac{1.96}{0.07}\right)^2 = \frac{\cancel{0.09}}{0.25} \cdot 28^2 = \frac{\cancel{0.09 \cdot 784}}{0.25} = \cancel{70.6}$ 196

$0.25\left(\frac{z}{B}\right)^2 \Rightarrow n = 196$ ✓

27. a. $B = 0.1$ $z = 1.96$ $n = \pi(1-\pi)\left(\frac{z}{B}\right)^2 = 0.25\left(\frac{1.96}{0.1}\right)^2 = 96$

b. $B = 0.05$ $z = 1.96$ $= 0.25\left(\frac{1.96}{0.05}\right)^2 = 384$

c. $B = 0.05$ $z = 2.58$ $= 0.25\left(\frac{2.58}{0.05}\right)^2 = 666$

d. $B = 0.01$ $z = 2.58$ $= 0.25\left(\frac{2.58}{0.01}\right)^2 = 16641$ ✓

29. $\bar{Y} = 1400$ $s = 1000$ (σ)

$B = 100$ $z = 1.64$ $90\% \to 0.05$

$n = \pi(1-\pi)\left(\frac{z}{B}\right)^2 = \cancel{0.25}\ 1000^2 \left(\frac{1.64}{100}\right)^2 = 269$ ✓

$s = \sqrt{\pi(1-\pi)} \to s^2 = \pi(1-\pi)$

$\cancel{\frac{s^2}{\pi} = 1 - \pi}$

$\cancel{\pi = 1 - \frac{s^2}{n} = 1 - \frac{1000^2}{n}}$

30. $n = \sigma^2\left(\frac{z}{B}\right)^2 = 200^2 \cdot \left(\frac{1.96}{25}\right)^2 = \underline{246}$

$\bar{Y} \pm z \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{200}{\sqrt{246}} = \underline{26}$

31. $\cancel{B = 25}$

$\cancel{P = 0.95}$

$\cancel{s = 400}$

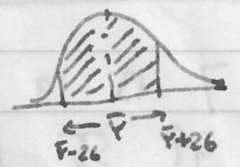
$\cancel{n = s^2\left(\frac{1.96}{25}\right)^2 = 983\ 983}$ $\cancel{160000 \cdot 0.00615}$

$\cancel{\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{400}{31} \approx 13}$

$\cancel{\frac{400}{246} \leftarrow ?}$

$\cancel{Y \pm z\sigma_\pi \to Y \pm 1.96 \cdot 13 \to Y \pm 25}$ 26

$\cancel{Y \pm 50}$



33. $n = 20$

$p = .95$

$M = \text{med}\{10, 17, 4, 15, 14, 0, 5, 2, 8, 10, 0, 3, 4, 100, 3, 0, 25, 12, 9, 36\}$

mid $= \frac{21}{2} = \frac{10+11}{2}$ $\to \frac{8+9}{2} = 8{,}500$

$s = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{n-1}}$

$M \pm z \frac{1.25 s}{\sqrt{n}}$

6th value $= 3000$

$\to$ 6th value $= 14000$ ✓

$M \pm \sqrt{n} \to \frac{(n+1)}{2} \pm \sqrt{n} = 10.5 \pm 4.5 \to (6, 15)$ ✓

$\sqrt{n} = 4.5$

5 35. a. $z = 1.64$ $\frac{n+1}{2} \pm 0.5 z \sqrt{n} = \frac{55}{2} \pm 0.5 \cdot 1.64 \sqrt{54} = 27.5 \pm 1.64 \cdot 7.8$
$= 27.5 \pm 6$ $(21.5^{th}, 33.5^{th})$
lower = 11, higher = 19

b. $z = 2.58$ $\frac{n+1}{2} \pm 0.5 \cdot 2.58 \sqrt{54} = \pm 9.5 = 27.5 \pm 9.5$
$= (18^{th}, 37^{th})$
values $\Rightarrow$ $(10, 19)$ ✓

45. a. z (+)
b. n smaller $\rightarrow \sigma_{\bar{Y}}$ (+) ✓

49. a (cf. 45.a :-)

51. $n = 1467$
$p = .95$
$(6.8, 8.01)$

b, e ✓
μ!

55. $n = 50$
$p = .95$
$(4.0, 5.6) \rightarrow \bar{Y} = 4.8$ ✓
$\bar{Y} \pm z\sigma_{\bar{Y}} =$ ~~...~~ $= 17.3$?
$\bar{Y} \pm z \frac{s}{\sqrt{n}}$
$5.6 = \bar{Y} + z \frac{s}{\sqrt{n}}$ $5.6 - \bar{Y} = +z \frac{s}{\sqrt{n}} \rightarrow \frac{5.6 - \bar{Y}}{z} \cdot \sqrt{n} = s = \frac{5.6 - 4.8}{1.96} \cdot \sqrt{50} = 2.9$ ✓ (7.07)

~~57.~~ basically from Q 30

31. $n = \sigma^2 \left(\frac{z}{B}\right)^2 = 246$
$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{400}{\sqrt{246}} = \frac{400}{15.7} \approx 25$
$Y \pm z \cdot \sigma_{\bar{Y}} = 1.96 \cdot 25 \approx 50$

49 / 51 / 53 / 61
6: 1 7 13 19 25 31 37 35 / 37 / 39 / 41 / 43 / 49 / 53
43 49 55 ~~67~~ ~~[illegible]~~

7: 1 7 13 19 25 31 ~~37~~
~~43~~ 45 ~~49~~ 55 ~~61~~
~~[illegible]~~
51

0.001996

6

14

1\. a. $z = 1.04$

$\rightarrow p' = .1492$

P value $= 2 \cdot .1492 = .2984$ cannot reject Ho

b. $z = 2.50$

$\rightarrow p' = 0.0062$

P value $= 2 \cdot 0.0062 = 0.0124$ → stronger evidence against Ho. ✓

7\. a. Ho : $\mu = 0$ Ha : $\mu \neq 0$ ✓

b. $z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} = \frac{-0.052 - 0}{\frac{1.253}{\sqrt{996}}} = 1.307 \rightarrow p' = 0.0951$

$\sqrt{996} = 31.5$, $\sigma = 0.0397$

$P = 2 \cdot p' = \underline{0.1902}$ ✓

↓

Ho : $\mu = 0$ not unlikely

c. equal spacing: $-2 \quad -1 \quad 0 \quad 1 \quad 2$ (spacing x, $2x$)

8\. $n = 26$

~~$\frac{10}{0.22} = 37 \ldots n$~~

$c = 7 \quad \pi_c = \frac{7}{26} = 0.27$

a. Ho : $\pi_0 = 0.5$

$s = \sqrt{\pi(1-\pi)} = \sqrt{0.5 \cdot 0.5} = 0.05$

$\sigma = \frac{s}{\sqrt{n}} = \frac{0.05}{5.1} = 0.0098$

$z = \frac{\hat{\pi}_c - \pi_0}{\sigma_\pi} = \frac{0.27 - 0.5}{s/\sqrt{n}} = \frac{-0.23}{0.0098} = -2.35$ ($\sqrt{n} = 5.1$)

$\rightarrow p' = .0094$

$\rightarrow p = 2 \cdot p' = 0.019$ ✓ ~?

↓ can reject Ho

b. $n = 31$

$c = 1$

Ho : $\pi_0 = 0.5$

$\sigma_\pi = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.25}{31}} = 0.0898$

$z = \frac{\pi_c - \pi_0}{\sigma_\pi} = \frac{1 - 0.5}{\sigma_\pi} = \frac{1 - 0.5}{0.0898} = 5.57$

$\rightarrow p' = ~~0.28~~$ ~~$\rightarrow P = 2 \cdot p' = 0.56$~~

$0.00000028 \rightarrow P = 0.0000005$

↓

reject Ho very strongly.

$\sqrt{\frac{0.5 \cdot 0.5}{25}} = \sqrt{\frac{0.25}{25}} = \sqrt{0.01} = 0.1$ ✓

19\. $n = 25$

$\pi = 18 = 0.72$ → $\sigma_\pi = \sqrt{\frac{\pi(1-\pi)}{n}}$

a. Ho : $\pi_0 = 0.5$

$z = \frac{\pi - \pi_0}{\sigma_\pi} = \frac{0.72 - 0.5}{0.1} = \frac{0.22}{0.1} = 2.2 \rightarrow p' = .0139 \rightarrow P = 0.0278$

↓ can reject Ho

b. $\alpha = 0.05$

$P_a = 0.014 < 0.05$ → reject Ho → yes ✓

c. TII ~~type I~~ – ~~we can reject it (it is~~

TI (we think we can reject it, but really we cannot → this mistake)

d. $\hat{\pi} \pm 2\sigma_\pi$ $0.75 \pm 1.96 \cdot 0.0433 = 0.75 \pm 0.08 = (0.67, 0.83)$ → yes, include majority. ✓

$\pi = \frac{75}{100} = 0.75$ 95%

$\sigma = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.75 \cdot 0.25}{100}} = \sqrt{\frac{.1875}{100}} = \sqrt{0.001875} = 0.0433$

15

25. $n=20$ a. assumptions: - random sample
- quantitative
- pop = norm distrib

$\bar{Y}=4.0$

$s=4$

$\bar{Y} \pm z\,\hat{\sigma}_{\bar{Y}} = \bar{Y} \pm \cancel{1.96}^{2.09} \frac{4}{\sqrt{20}} = 4 \pm \cancel{1.96}^{2.09} \cdot 0.894 = 4 \pm \cancel{1.75}^{1.87} = (2.13, 5.87)$

$\sqrt{20} \approx 4.\cancel{47}$

2.093

↑

$df = 20-1 = 19$

t-table $t_{0.025}$

b. skewed to right; min = 0 ⇒ only 1s away from $\bar{Y}$...
but OK w/ robustness of t-stats.

31. $H_0: Y = 0$ H_a ~~Y≠0~~ $Y > 0$

3, 7, 3, 3 → $\bar{Y}=4$, $n=4$

$t = \frac{\bar{Y}-\mu_0}{s/\sqrt{n}} = \frac{4-0}{2/2} = \frac{4}{1} = \cancel{2.25}\ 4$ $df = n-1 = 3$

$\hat{\sigma} = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n-1}} = \sqrt{\frac{1+9+1+1}{3}} = \sqrt{\frac{12}{3}} = \sqrt{4} \sim 2$

↓

$0.025 > P \geq 0.01$

so, reject $H_0: Y=0$
→ $H_a: Y>0$ rel. good.

37. $\pi = 0.5$

a. $2(0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5) = \cancel{0.008}\ 0.016 \cdot 2 = \underline{0.031}$

b. p(win all) = 0.016

$P_{normal} = 0.5$ (H_0)

? what does Q ask?

43. a. $P_s = 0.8$ $P_f = 0.2$

p(~~su~~) ~~$0.8^4 = 0.4096$~~   p(all fail) $= 0.2^4 = 0.0016$ ✓

$H_0: \pi = 0.8$

$H_a: \pi < 0.8$

↓

b. = P-Value $\pi < \cancel{0.2}\ 0.8$ ~~expected~~

~~49.~~ 55. 15

	reject H_0	do not reject H_0
H_0 true	T I	✓
H_0 false	✓	T II

T I test positive while really HIV−

T II test negative while really HIV+

	Test ⊖	Test ⊕
H_0 false HIV ⊕	T II	✓
H_0 true HIV ⊖	✓	T I

6th only

7

16

1. 1982: $n = 350$, $\bar{Y} = 4.1$, $s = 3.3$; $n = 1965$, $\bar{Y} = 2.8$, $s = 2$

$$\hat{\sigma}_\Delta = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{3.3^2}{350} + \frac{2^2}{1965}} = \sqrt{\cancel{0.0104}\ 0.0331} \sim \cancel{0.3}\ 0.18$$

(10.89, 4; $\cancel{0.0084}$ 0.0311, 0.0020)

a. $(\bar{Y}_2 - \bar{Y}_1) \pm z\hat{\sigma}_\Delta = (2.8 - 4.1) \pm 1.96 \cdot 0.18 = 1.3 \pm 0.36$ $\cancel{(0.94, 1.66)}$ $(0.94, 1.66)$

95% → $z = 1.96$

↳ doesn't include 0, so difference ≠ likely to be 0.

b. $H_0: \mu_1 = \mu_2$ $H_a: \mu_1 \neq \mu_2$

c. $$z_{\mu_2 - \mu_1} = \frac{(\bar{Y}_2 - \bar{Y}_1) - 0}{\hat{\sigma}_\Delta} = \frac{2.8 - 4.1}{0.18} = \frac{1.3}{0.18} = 7.2$$ $p < 0.00001$ → v. unlikely H_0

d. no, skewed to right

7.a. B: $n = 152$, $\bar{Y} = 4.7$, $s = 0.7$; a: 140, 3.1, 1.3

$$\sigma_{\bar{Y}_2 - \bar{Y}_1} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.7^2}{152} + \frac{1.3^2}{140}} = \sqrt{\frac{0.49}{152} + \frac{1.69}{140}} = \sqrt{0.01527}$$

(0.0032, 0.0121) $\sim \underline{0.125}$

$(\bar{Y}_2 - \bar{Y}_1) \pm z\,\sigma_{\bar{Y}_2 - \bar{Y}_1} = (3.1 + 4.7) \pm 1.96 \cdot 0.125 = +1.6 \pm 0.25$

95% → $z = 1.96$ (1.35, 1.85)

b. T_1 ♂: $n = 50$, $\bar{Y} = 2.9$, $s = 1.4$; T_2 ♀: 90, 3.2, 1.2

$$\sigma = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1.4^2}{50} + \frac{1.2^2}{90}} = \sqrt{0.0552} \sim 0.24$$

(0.0392, 0.016)

$$z = \frac{(\bar{Y}_2 - \bar{Y}_1) - 0}{\hat{\sigma}_{\bar{Y}_2 - \bar{Y}_1}} = \frac{0.3}{0.24} = 1.25$$

$p' = 0.1056$ $P = 0.21$ $H_0: \mu_1 = \mu_2$ plausible (21%)

c. cannot reject H_0 since $\alpha = 0.05 < 0.21$

T2 ? type II / I

13. 1982: $F = 152 \Rightarrow \hat{\pi}_1$, $N = 165$, $n = 317$; 1994: $2215 \Rightarrow \hat{\pi}_2$, 580, $n = 2795$

$\hat{\pi}_1 = \frac{152}{317} = 0.48$

$\hat{\pi}_2 = \frac{2215}{2795} = 0.79$

$\cancel{(\bar{Y}_2 - \bar{Y}_1) \pm z\sigma}$

$$\sigma = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}} = \sqrt{\frac{0.48 \cdot 0.52}{317} + \frac{0.71 \cdot 0.21}{2795}} = \sqrt{0.00078 + 0.000053} = \sqrt{0.00083} = \underline{0.029}$$

$(\hat{\pi}_2 - \hat{\pi}_1) \pm z\sigma = 0.31 \pm 0.06$

95% → $z = 1.96$ (0.25, 0.37)

thus π_2 between 0.25 and 0.37 larger than π_1.

7 (16) 17

19. S N

$n_1 = 13$ $n_2 = 17$

$\bar{Y}_1 = 2$ $\bar{Y}_2 = 4.8$

$s_1 = 2.1$ $s_2 = 3.2$

$\sigma_\Delta = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \approx 1$

0.3392 0.6023

$df = n_1 + n_2 - 2 = 28$

a. $(\bar{Y}_2 - \bar{Y}_1) \pm z\,\sigma_\Delta$ $\quad -2.8 \pm 2$ $\quad$ (−4.8, 0.8) ✓

95% → $t = 2.08$

$$\cancel{s^2 = \frac{n_1 - 1 \cdot s_1^2 + n_2 - 1 \cdot s_2^2}{n_1 + n_2 - 2} = \frac{53 + 164}{28} = \frac{217}{28} = 7.74 \to s = 2.78}$$

b. $H_0: Y_1 = Y_2$

$t = \frac{(\bar{Y}_2 - \bar{Y}_1) - 0}{\sigma} = -2.8$ $\quad df = n_1 + n_2 - 2 = 28$ $\quad p = 0.01 + .005$ ✓

25. W B

$n = 832$ 209

$\pi_{6000} = 0.53$ 0.66

$\sigma = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$

$= \sqrt{0.000299 + 0.000730}$

a. $(\pi_2 - \pi_1) \pm z\sigma$

$= \sqrt{0.00103} = 0.01$ ✓

95% → $z = 1.96$

$(0.66 - 0.53) \pm 1.96 \cdot 0.01$

$+0.13 \pm 0.02$ $\quad$ (0.11, 0.15) $\quad$ $\pi_2 > \pi_1$ by (0.11 to 0.15).

b = $z = \frac{(\pi_2 - \pi_1) - 0}{\sigma} = \frac{0.13}{0.01} = 13$ ✓

AF: $\sigma = 0.033 \to z = 3.9$ $\quad \to p < 0.0001$ ✓

~~31.~~

45. p(smoker) = 0.00130 $\quad$ a. $0.00130 - 0.00012 = \underline{0.00118}$

p(non) = 0.00012 $\quad$ Statistically similar! ✓ in absolute risk

b. $\frac{0.00130}{0.00012} = \underline{10.83}$ $\quad$ 10x difference in relative risk. ✓

55. a ✓ $\quad$ 51. a. $(F_1 F_2 F_3)\ (F_1 F_2 M_1)\ (F_1 F_2 M_2) \ldots (M_1 M_2 M_3)$.

b ✗ $\quad$ b. $\underbrace{p(MMF) \vee p(MMM)}_{\text{ie. 10 cases } (\frac{n}{2})} \le \frac{1}{3}$ ✓

c ✓

d ✓ $\quad$ c. $p(MMM) = \frac{1}{20} = 0.05$ ✓

8) \ ~~33, 35, ... 41~~ ~~49, 55, 61~~ Sol

1, 7, 13, 19, 25, 31, ~~43~~, 47, ~~51~~, 57

a) 43, 49, 55

~~55 \ 19 ... 31, 32, 41, 51~~

1 7 13 17 33 39 45 53

exam: ... b ?

18

8

1. a.

	♀	♂
Y	0.85	0.85
N	0.15	0.15

(0.15 , 0.85)

b. v. likely (no different results for ♀/♂). ✓

7.

10	20	[30	60
30	40	[30	100
[10	[20	[10	[40
50	80	70	[200

✓

13. (12: χ^2)

a. ✓

b. ✓ $\frac{2409}{37792} - \frac{3865}{30902} = 0.06$ ✓

c. $\frac{2409}{3865} \Big/ \frac{35383}{27037} = 0.623 / 1.308 = 0.47$

SO: wearing seatbelt makes you 0.47 as likely to get injured compared to not wearing one (ie. 2 times less likely) ✓

expected values

	inj	no inj	c
belt	$\boxed{2409 \; (x=3452)}$	35383 / 34340	37792
no belt	3865 / fe=2882	27037 / 28080	30902
	6274	62420	68694
	A		B

$x = \left(\frac{6274}{68694}\right) \cdot 37792 = 3452 = fe$

$z = \frac{f_o - f_e}{\sqrt{fe(1-\pi_r)(1-\pi_c)}}$

$1 - A/B \qquad 1 - C/B$

$\frac{2409 - 3452}{\sqrt{3452(1-0.091)(1-0.55)}} = 27.7$ ✓

$\pi_r = A/B \qquad \pi_c = C/B$

19. ~~odds = $\frac{p(win)}{p(lose)}$ ⟺ p(win) = odds · p(lose)~~ p(L) = 1 − p(w) →

$p(win_{sea}) = \frac{1.1}{1.1+1} = 0.52$ ← win

$p(win_{Auc}) = \frac{0.3}{0.3+1} = 0.23$

$0.52 + 0.23 \overset{!}{\neq} 1$ ✓

25.

4	2	1
2	2	2
1	2	4

a. no, f_e in all cells <5.

b. P = 0.48 → no evidence of assoc. ✓
(from answer)

31. a. $\hat{\gamma} = .3 = \frac{C' - D'}{C' + D'}$ $\quad C' + D' = 1$, $D' = 1 - C$ $\quad = \frac{C - (1-C)}{C + 1 - C} = \frac{2C - 1}{1} = .3$

~~$c' = .15$, $c' = 1.15$~~

$2c' = 1.3$

$c' = \underline{0.65} \rightarrow D' = \underline{0.35}$ ✓

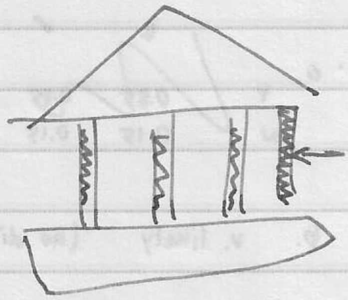
b. $\hat{\gamma}_A = .3$ pos. assoc

$\hat{\gamma}_B = -.7$ neg. assoc. ; stronger. ✓

$|\hat{\gamma}_B| > |\hat{\gamma}_A|$

[8] ~~43.~~

49. X : order matters ✓



55. a. a b / c d (a ↘ d) ; a b / c c (b ↙ c)

$C = a \cdot d$ $D = b \cdot c$

$$Q = \hat{\gamma} = \frac{C-D}{C+D} \quad \frac{(ad-bc)}{(ad+bc)}$$

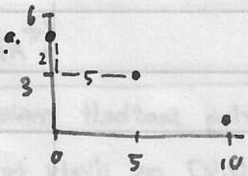
b. if $\underset{c}{a} = 0$: $\frac{-bc}{bc} = -1$

if $\underset{\text{etc.}}{b} = 0$: $\frac{ad}{ad} = 1$ ∎

[9] 1. a. ~~explanatory:~~ highschool ~~college~~ GPA c. years of edu.

b. mothers edu d. annual income ✓

7a.



$\hat{Y} = a + bx$

$= 5 - 0.4x$

$\bar{x} = \frac{15}{3} = 5$

$\bar{Y} = \frac{9}{3} = 3$

$s_y = \sqrt{\frac{4+0+4}{2}} = 2$

$s_x = \sqrt{\frac{25+25}{2}} = 5$

b. $r = b\left(\frac{s_x}{s_y}\right)$

$= -0.4\left(\frac{5}{2}\right)$

$= -2.5 \cdot 0.4$

$= \underline{\underline{-1}}$ ✓

$s_x = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n-1}}$

$s_y = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

A+F: goes through all points: -1 ✓

~~15.~~ 17. Y = birth rate
X = ec. activity

a. $\hat{Y} = a + bx + \varepsilon$

$= 36 - 0.28x$ ✓

b. $r^2 = -0.2973$

$r = \sqrt{r^2} = -\sqrt{0.2973} = -0.55$ ✓

c. Nigeria : $\hat{Y} = 36 - 0.28(51) = 21.72$

$Y_N = 43.3$

residual $= (43.3 - 21.7) = \underline{21.6}$: actually much higher than predicted using regression...

20

9

33. Y = income

a. ~~not at all~~ divide by 1.5

b. not at all (association remains)

39. a. not: lin. reg. assumes equal distance (on avg.) for all values (residuals not more or less at any avg. place).

b. not: "heteroskedasticity"

c.  not linear; b=0

d.  not linear increase ✓

10: ~~41, 43,~~ ~~1~~ ~~3~~ ~~13~~ ~~25~~ ~~27~~ ~~29~~ 5, 7, 15, 21, 31, 37, ~~43~~ 39.

11 ~~5?~~ ~~55~~ ~~25~~ ~~27~~ ~~29~~ ~~31~~ ~~33~~ ~~41~~ ~~45~~

1, 7, 13, 19, 23, ~~33~~, 35, 41, 47, 53

10

5. a. $x_1 \rightarrow Y$, $x_2 \rightarrow Y$ ✓

b. sex $\rightarrow x_1 \rightarrow Y$, sex $\rightarrow x_2 \rightarrow Y$

7. a.

	WC	BC
Dem	265	735
Rep.	735	265

association: yes: BC tend to vote Dem.

b. income for everything ✓

c. more info $x_1 \rightarrow Y$ (inc → vote); $x_1 \rightarrow Y$, x_2 job ✓

d. no, association ≠ change between tables ✓

e. $x_2 \rightarrow x_1 \rightarrow Y$
gen. job → class inc. → vote

f. job x_2, x inc. → Y vote; inc. → job, → vote ??

g. (e) theory makes it more plausible ✓

15. a. x = gender
Y = ? salary
control = ? academic rank ✓

b. $\text{Inc}_{♂} > \text{inc}_{♀}$ ✓ || A+F: $\bar{Y}_♀ < \bar{Y}_♂$ by 9500 $ ✓
(49.6 − 40.1) = 9.5

c. (b) still true for each rank; gap bigger for professors ✓

d. gender (x) → rank (control) → salary (Y)
↑ not reflected in table... ✓

e. # ♂ prof >> # ♀ prof
↓
$\bar{Y}_{♂}$ more skewed

10

21. a. expl. resp. control

S → C. age: → 1.1 → 2.4 → 4.3 ✓

b. interaction: yes, levels of c (thus S → c) vary over ① (increase). ✓

31. 1980 $\bar{Y} = 890$ 1984 $\bar{Y} = 897$

a. $0.72 \cdot 924 + .28 \cdot 805 = 665 + 225 = \underline{890}$ ✓

$.695 \cdot 932 + .305 \cdot 817 = 648 + 249 = \underline{897}$ ✓

b. % whites declined, so their (whites) contrib. relatively less ✓

37. a. % whites $\xrightarrow{\ominus}$ crime (a)

sparent $\xrightarrow{\oplus}$ crime (b)

% whites $\xrightarrow{\ominus}$ sparent (c)

b. W → S → cr → weaker; some effect (much) due to (c). ✓ ✓

11

1. a. ⊕ ✓

b. ⊖ ✓

c. ~~$\hat{Y} = 40.3 - 0.8x_1 + 0.6x_2$~~

$\hat{Y} = -11.526 + 2.6x_1$ ✓

d. $\hat{Y} = 40.3 - 0.8x_1 + 0.6x_2$ ✓

e. high correlation (43%, 68%) with both vars.

f. $x_2 = 0 \rightarrow \hat{Y} = 40.3 - 0.8x_1$

$x_2 = 50 \rightarrow \hat{Y} = 40.3 - 0.8x_1 + 30 = 70.3 - 0.8x_1$

$x_2 = 100 \rightarrow \hat{Y} = 100.3 - 0.8x_1$ ✓

7. a. $\hat{Y} = -1198.5 + \underset{b_1}{18.3}x_1 + \underset{b_2}{7.7}x_2 + 89.4x_3$ ✓

b. $R^2 = .722$ → v. sign. (72% less error cf. $\bar{Y}$) ✓

c. $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ $H_a: \neq 0$

$F = 39.9$ (from table) $df_1 = k = 3$ $p < 0.05$

$df_2 = n - (k+1) = 50 - 4 = 46$ $p < 0.01$

$\underline{p < 0.001}$ → v. unlikely that H_0 = true → at least one var w/ effect. ✓

d. $t = \frac{b_i}{\hat{\sigma}_{b_i}} = \frac{18.3}{6.1} = 3$ $df = df_2 = 46$

→ $p \leq 0.005 \cdot 2$ → $p < 0.01$ → b_1 is v. significant (controlling for other) ✓

$t_{0.025}$? 1.96 A+F take 2.01 in t-table, why?

e. $b_i \pm t_{0.025}\hat{\sigma}_{b_i} = 18.3 \pm 2 \cdot 6.1 = 18.3 \pm 12 = (6, 30)$. $df = 46$ $t_{0.025}$ → ?

11

7. f. because much of poverty explained by single parent → highly correlated.
x_1 x_3

multicollinearity

22

13. a. $\hat{y} = -498.7 + 32.6 x_1 + 9.11 x_2$ (pov. / metro)

$\hookrightarrow \hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$ ✓

→ from answer $= 158.9 - 14.72 x_1 - 1.29 x_2 + 0.76 x_1 x_2$

b. also increase, interaction is also positive (ie. both same) ✓

19. y = vote Dem $\hat{y} = 20 + .3x_1 + .05x_2 + .005 x_1 x_2$
x_1 = reg Dem
x_2 = reg + vote

yes, if x_1 larger, x_2 also increases. ✓

23. y = murders $\hat{y} = 30 - .02x_1 - .1x_2 - 1.2x_3 + .8x_4$
x_1 = police
x_2 = sentence
x_3 = income
x_4 = unemploy.

$\bar{y} = 15$
$\bar{x}_1 = 100$ $\bar{x}_3 = 13$ $s_y = 8$ $s_{x_2} = 10$
$\bar{x}_2 = 15$ $\bar{x}_4 = 7.8$ $s_{x_1} = 30$ $s_{x_3} = 2$
$s_{x_4} = 2$

a. no. ✓ different units of measurement

b. $b_1^* = b_1 \left(\frac{s_{x_1}}{s_y}\right) = -0.02\left(\frac{30}{8}\right) = -0.075$ $b_3^* = b_3\left(\frac{s_{x_3}}{s_y}\right) = -1.2\left(\frac{2}{8}\right) = -0.3$

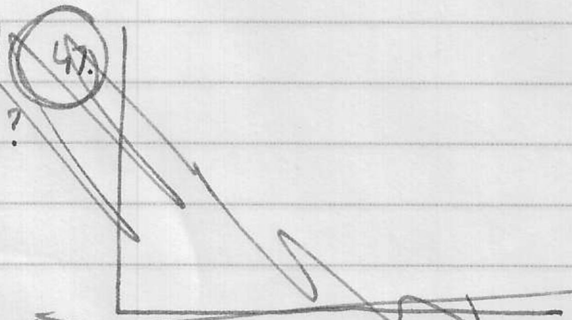
$b_2^* = b_2\left(\frac{s_{x_2}}{s_y}\right) = -.1\left(\frac{10}{8}\right) = -0.125$ $b_4^* = b_4\left(\frac{s_{x_4}}{s_y}\right) = +.8\left(\frac{2}{8}\right) = +0.2$ ✓

d.

35. → b $r^2_{yx_2} > R^2$ ✓
partial cannot explain more than everything
r^2 R^2

41. a ✓
~~b~~ partial effect (P)
~~c~~ wrong sign

47.



?
¿do

~~53.~~ 12 | 12 \ 15 ~~2~~ 31 – 39 → 1, 7, 13, 19, 2?, 41, 47
13 | 13 \ 13 – 21 → 1, 7, 11, 23, 25
14 \ 35 – 47 → 1, 7, 13, 19, 25, 31, 49, 57
D \ 23 – 27
\ 31 – 35 → 1, 7, 13, 19, 29, 37,

12

23

use dummies...

ANOVA

14 + 15 → Mon

1. $(84-81)^2 + 2\cdot(83-81)^2 + (74-81)^2 = 66$ ✓

$\quad 9 \qquad 2\cdot 4 \; + \; 49$

b. WWS = 682

df = 12 : $(n-1)\cdot 4 = 12$, $n = 3$

WMS = $\frac{WWS}{df} = \frac{682}{12} = \cancel{55.1}\ \underline{51.8}$

7. b. $(\bar{Y}_2 - \bar{Y}_1) \pm t_{.005}\,\hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

df = df.

$n_1 + n_2 - 2$

$470 + 357 = 825$

$(4.2 - 4.1) \pm 2.64\cdot\sqrt{1.76}\cdot\sqrt{\frac{1}{470}+\frac{1}{357}} =$

$(-.4\ ,\ .1)$

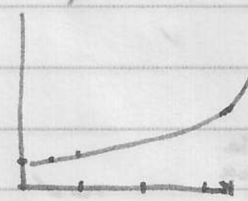
13. $F = \frac{EMS}{MSE} = \frac{900}{5}$

14

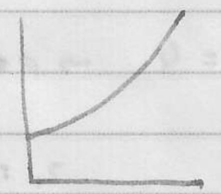
~~1.~~ ~~7.~~ ~~13.~~ 19. a. x_1, x_2 are highly correlated → 'multi collinearity...' ✓

~~25~~. 31.

a.



b.



4% growth/yr. ✓

"Zinseszins"

49. n = 100,000

a. $(n + 4.2\%)^{10}$

$1.042^{10}\cdot 100{,}000 = 1.509\cdot 100{,}000 = \underline{150895}$

b. ⇒ 50.8% increase.

57. · heteroskedasticity → c

· multicollinearity → e

· forward selection → g

· interaction → j

· exp. model → a

· stepwise selection → i

· studentized residual → k

· GLM → h ✓

1, 7, 13, 19, 29, 37

15

1. a. $\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -1 + .2X$

$\hat{\pi} = .5 \Rightarrow \log\left(\frac{.5}{.5}\right) = -1 + 2X$

$\rightarrow Y = -2 + 2X$

$\hat{\beta} \cdot \hat{\pi}(1-\hat{\pi}) = .02 \cdot .5 \cdot .5 = .005$

where from?

b. (i) $\hat{\pi} = .5 \rightarrow 1 = -1 + 2X \rightarrow 2 = 2x \rightarrow \underline{x = 1}$

(ii) above (i)

(i) $\frac{a}{b} = \frac{1}{.02} = 50$ (ii) above (i) ✓ "

c. $\hat{\pi} = \frac{e^{a+bx}}{1+e^{a+bx}}$

(I) $\Rightarrow x = 10$ $= \frac{e^{-1+.2\cdot10}}{1+e^{-1+.2\cdot10}} = \frac{e^{1}}{1+e^{1}} = .31$

(II) $\Rightarrow x = 100$ $= \frac{e^{-1+.2\cdot100}}{1+e^{-1+.02\cdot100}} = \frac{e^{19}}{1+e^{19}} \approx .73$

d. $\frac{\pi}{1-\pi} = e^{a+bx}$ $\rightarrow$ increase of x by 1 unit, odds multiply by e^{b}. ✓

e. $se = 0.05$ $z = \frac{b}{se} = \frac{0.2}{0.05} = 4 \rightarrow p \leq 0.0001$ ✓

7. a. $\text{logit}(\pi) = a + b_1 z_1 + b_2 z_2$

z_1 : 1 for white, 0 for black

z_2 : 1 for AZT, 1 for no AZT ~✓

b. $\pi = \frac{e^{a+b_1z_1+b_2z_2}}{1+e^{a+b_1z_1+b_2z_2}} = \frac{e^{a}}{1+e^{a}}$ ~✓

[?]

13. 19. a. $-0.0414 \rightarrow e^{-.0414} = .96$ ~ $\rightarrow$ thus odds above level: $\frac{1}{.96} = .04 \rightarrow 4\%$.

b. $H_0 : b_{AC} = 0$

test statistic = Wald = 3.5 df=1 (def.) $\rightarrow \chi^2$ distrib. $\rightarrow_{\text{table}}$ $p < \underline{0.1}$ ✓

printout: $p = 0.06$ ✓

c. because assumptions ignored... ✓

~~d.~~

29. a. having done edu makes you 4x more likely to use condoms than those who hasn't. ✓

b. $\text{logit}(\hat{\pi}) = a + b_1 G_1 + b_2 G_2 + b_3 S_1 + b_4 L_1$ dummies. ~✓

c. $\log(1.23) = .21$ $\frac{.21 + 2.56}{2} = \frac{2.77}{2} = \underline{1.385}$ ✓

$\log(1288) = 2.56$

~~37.~~