Statistics 1040, Section 006, Quiz 1 (20 Points)

September 6, 2002

Your Name: ____________________

from FPP, p. 26, #7, a, b, d)

Question 1: Observational Studies and Experiments (**12 Points**)

According to a study done at Kaiser Permanente in Walnut Creek, California, users of oral contraceptives have a higher rate of cervical cancer than non–users, even after adjusting for age, education, and marital status. Investigators concluded that the pill causes cervical cancer. Answer the following three questions:

(2) 1. Is this a contolled experiment or an observational study? Circle your answer.

7. (a) This is an observational study - the patients chose to be on oral contraceptives.

(5) 2. Why did the investigators adjust for age, education, and marital status?

(b) Age, education, and marital status are variables that are known to have associations with cirvical cancer and pill use. They adjusted for these variables to try to eliminate problems due to these confounding factors.

−1 if "confounding factor" not mentioned

(5) 3. Were the conclusions of the study justified by the data? Answer yes or no, and explain briefly.

(d) The conclusions are not justified by the study. There are several possible confounding factors that were not adjusted for. For example, the number of sexual partners is likely to be higher of pill-users than non-pill users, and it could be that cirvical cancer is a sexually transmitted disease.

−3 if yes & explanation

−2 if no & no explanation

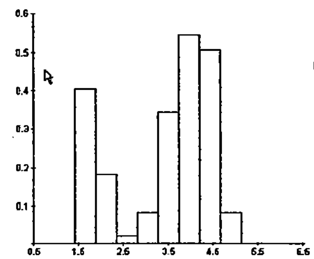
Please turn over!

−1 if no & some explanation (but not speaking of "confounding factors")

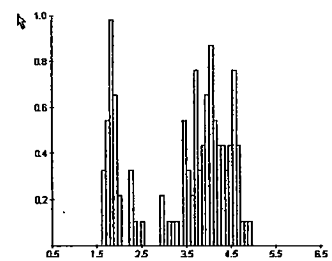
Solutions from Workbook.

Question 2: Histograms (**8 Points**)

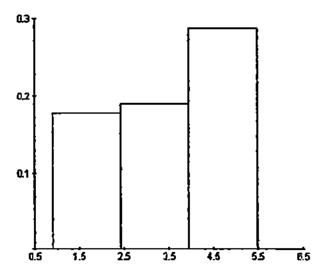
The following three histograms are based on the Old Faithful data set. The observations are the duration (in minutes) for eruptions of the Old Faithful geyser in Yellowstone National Park. There exist two types of eruptions: shorter ones (about 2 minutes) and longer ones (about 4 minutes). Which of the three histograms best describes the underlying data. Shortly explain your answer and indicate why you think the other 2 histograms don't represent the data as well as the one you have selected.



(4) best (matches the description of the data from above, i.e., the 2 types of eruptions are clearly visible)



(2) too spiky (too many classes – not clear that there are 2 types)



(2) too general (too few classes – the 2 types are not visible at all)

Statistics 1040, Section 006, Quiz 2 (20 Points)

September 13, 2002

Your Name: ____________________

Question 1: Measures of Center and Spread (**20 Points**)

The table below, published in USA Today on Friday, May 15, 1998, lists the 15 most widely held stocks and their change year–to–date (YTD). Suppose we hold one share each of AT&T (-6.5), Bell Atlantic (+0.1), Coca–Cola (+16.6), Merck (+11.4), and SBC Comm. (+3.4).

6

1. Find the **average change** (YTD) for the 5 stocks we own. **Show your work!**

MOST WIDELY HELD STOCKS

Stock	Thurs.	Chg.	Pctg. Change Day	YTD
AT&T	$57[illegible]	[illegible]	−1.6	−6.5
Ameritech	$43¼	−¼	−0.6	+7.5
Bell Atlantic	$9[illegible]	[illegible]	+1.2	+0.1
BellSouth	$68 9/16	+1 9/16	+2.3	+21.1
Coca-Cola	$77[illegible]	+½	+0.6	+16.6
Compaq	$31 9/16	−5/16	−1.0	+11.7
Exxon	$73[illegible]	unch	[illegible]	+20.3
GE	$84⅛	−½	−0.6	+14.7
Intel	$84[illegible]	[illegible]	−0.3	+20.4
IBM	$125 13/16	+3 15/16	+3.2	+20.3
Johnson & John	$72[illegible]	[illegible]	−0.3	+9.6
Lucent	$71 5/16	−13/16	−1.1	+78.3
Merck	$118[illegible]	[illegible]	+0.6	+11.4
Microsoft	$88 15/16	+2	+2.3	+37.6
SBC Comm	$37[illegible]	[illegible]	−0.5	+3.4

Stocks included are those held by the largest number of accounts at Merrill Lynch.

$$\text{avg} = \frac{-6.5 + 0.1 + 16.6 + 11.4 + 3.4}{5} = \frac{25}{5} = 5$$

−3 if only final result (correct!)
−2 for incorrect sum
−2 for missing value (in sum)

4

2. Find the **median change** (YTD) for the 5 stocks we own.

sorted list: −6.5 0.1 3.4 11.4 16.6

Center value: 3.4

median = 3.4

−3 for incorrect median

Please turn over!

10

3. Find the **standard deviation of the changes** (YTD) for the 5 stocks we own. **Show your work!**

−5 if only final result (correct!)
−2 each mistake (but not for follow-up mistakes)
−9 if some formula only (no values at all)
−2 for missing value
−2 if (avg − entry)
−4 if summation of values (and not squares)
−4 if avg not subtracted in step 2
−4 if not divided by 5 in step 4
−4 if no $\sqrt{\ }$ in step 5

1. avg = 5

2.
$-6.5 - 5 = -11.5$
$0.1 - 5 = -4.9$
$16.6 - 5 = 11.6$
$11.4 - 5 = 6.4$
$3.4 - 5 = -1.6$

3.
$(-11.5)^2 = 132.25$
$(-4.9)^2 = 24.01$
$(11.6)^2 = 134.56$
$(6.4)^2 = 40.96$
$(-1.6)^2 = 2.56$

4. $$\frac{132.25 + 24.01 + 134.56 + 40.96 + 2.56}{5} = \frac{334.34}{5} = 66.868$$

5. $$SD = \sqrt{66.868} = 8.177 \approx 8.2$$

Formulas:

$$\text{avg} = \frac{\text{sum of all numbers}}{\text{how many numbers}}$$

$$SD = \sqrt{\text{average of [(deviations from avg)}^2]}$$

Statistics 1040, Section 006, Quiz 3 (20 Points)

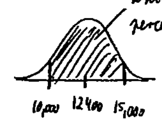
Sepyember 20, 2002

Your Name: ____________________

Question 1: Normal Approximation for Data (**20 Points**)

Car drivers in the United States average 12,400 miles a year, nearly 50 percent more than European drivers (*The Economist*, June 22, 1996). Assume that the number of yearly miles by U.S. drivers approximately follows a normal curve with a standard deviation of 3,200 miles.

(13) 1. Determine the percentage of drivers who travel between 10,000 and 15,000 miles in a year.



① Convert each number into standard units:

$\frac{10{,}000-12{,}400}{3{,}200} = -0.75$ s.u. $\qquad \frac{15{,}000-12{,}400}{3{,}200} = 0.81 \approx 0.80$ s.u.

② Find the area (percentage) related to the numbers from ① (use Table):

area from -0.75 to 0.75: 54.67%

area from -0.80 to 0.80: 57.63%

③ Find the area (percentage) of interest:

area from -0.75 to 0: $\frac{1}{2} \cdot 54.67\% = 27.335\%$

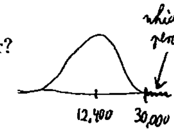
area from 0 to 0.80: $\frac{1}{2} \cdot 57.63\% = 28.815\%$

total area from -0.75 to 0.80: 27.335% + 28.815% = 56.15%

-3 for each incorrect (or missing) s.u.
-3 for each incorrect table value
-3 for each incorrect final result
-2 for each minor (computational) error
+3 if correct graph only (each)

Answer: about 56.15% of all US drivers travel between 10,000 and 15,000 miles a year (in 1996)

(7) 2. And what percentage of drivers travels more than 30,000 miles in a year?



$\frac{30{,}000-12{,}400}{3{,}200} = 5.5$ s.u.

value (5.5) not in Table! – but notice that area from -4.45 to 4.45 is 99.9991%

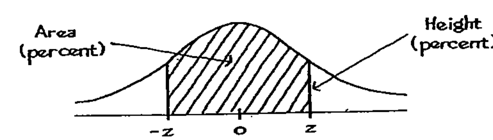
Show your work!

area from -5.5 to 5.5 is even closer to 100%

therefore, area above 5.5 is very close to 0%,

i.e., basically no driver travels more than 30,000 miles a year

Tables



A NORMAL TABLE

z	Area	z	Area	z	Area
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991
				5.5	99.9999... ≈ 100

Statistics 1040, Section 006, Quiz 4 (20 Points)

September 27, 2002

Your Name: ____________________

Question 1: Percentiles and the Normal Curve (**12 Points**)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For all college seniors and graduates who took the exam in the past few years, the mean score for the verbal ability portion of the exam was 497 with a standard deviation of 115. Assuming the scores are bell–shaped, fill in the blanks below. **Show your work!**

(6) 1. A student who received a score of 650 on the verbal ability portion of the GRE exam was at the ___91___ th percentile of the score distribution.

area? 497 650

Transfer 650 in s.u.: $\frac{650-497}{115} = 1.33$ s.u. -2 for incorrect s.u.

Area between -1.35 to 1.35 (from table): 82.30% -1 for incorrect table value

Area from 0 to 1.35 : 82.30%/2 = 41.15% -2 for incorrect area

Area below 1.35 : 50% + 41.15% = 91.15%

i.e., about 91% of the area lies below 650

(6) 2. A graduate school program in English will admit only students with GRE verbal ability scores in the top 30%. Therefore, the lowest GRE score they will accept is 40% about ___555 (or 560)___. Find z such that area from -z to z is about 40%: z = 0.50 (38.29%) or 0.55 (41.77%)

30% 30% -z 497 z ← 70th percentile!

Transfer into original units: ~~for~~ z = 0.50: 0.50 · 115 + 497 = 554.5

for z = 0.55: 0.55 · 115 + 497 = 560.25

A score of 555 (or 560) was needed to be accepted. -2 for incorrect area

Question 2: Correlation (8 Points) from FPP, p. 130, #6 -2 for incorrect z

-2 for incorrect original units

(4) 1. If women always married men who were five years older, the correlation between the ages of husbands and wives would be ___(e) exactly 1___. Choose one of the options below, and explain.

men women

6. (a) All the points on the scatter diagram would lie on a line sloping up, so the correlation would be 1.

(4) 2. The correlation between the ages of husbands and wives in the U.S. is ___(d) close to 1___. Choose one option, and explain.

men women

(b) Close to 1; this is like part (a), with some noise thrown into the data.
Comment. In the March 1993 Current Population Survey, the correlation between the ages of the husbands and wives was about 0.95; the husbands were, on average, 2.7 years older than their wives.

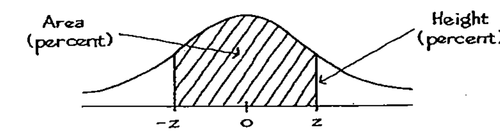
Options: (a) exactly -1 (b) close to -1 (c) close to 0 (d) close to 1
(e) exactly 1

"slightly" wrong answers:
1: (d)
2: (c), (e)

1

0: wrong answer, no explanation
1: wrong answer, some explanation
2: "slightly" wrong answer, no explanation
3: "slightly" wrong answer, some explanation
3: correct answer, no explanation

Tables



A NORMAL TABLE

z	Area	z	Area	z	Area
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

Statistics 1040, Section 006, Quiz 5 (20 Points)

October 11, 2002

Your Name: ________________

Question 1: The Regression Line (**20 Points**)

In one study, the correlation between the educational level of husbands and wives in a certain town was about 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.

X: husband's ed. level
Y: wife's ed. level

1. Find the regression equation for predicting the educational level of a wife from the educational level of her husband. (10 Points)

   $\text{slope} = r \cdot \frac{SD_Y}{SD_X} = 0.50 \cdot \frac{3}{3} = 0.50$

   $\text{intercept}: avg_Y - \text{slope} \cdot avg_X = 12 - 0.50 \cdot 12 = 6$

   regression equation:

   $\boxed{\text{ed. level wife} = 6 + 0.50 \cdot \text{ed. level husband}}$

   or $\boxed{y = 6 + 0.50 \cdot x}$

2. Use the regression equation from part 1. to predict the educational level of a woman whose husband has completed 18 years of schooling. (5 Points)

   $\text{ed. level wife} = 6 + 0.50 \cdot 18 = 6 + 9 = 15$

   (i.e., 15 years of education is the predicted average education level for a woman whose husband has completed (18 years of schooling)

3. Find the r.m.s. error for predicting the wife's educational level from the educational level of her husband. (5 Points)

   $\text{r.m.s. error} = \sqrt{1-r^2} \cdot SD_y = \sqrt{1-0.5^2} \cdot 3 = \sqrt{1-0.25} \cdot 3$

   $= \sqrt{0.75} \cdot 3 = 0.866 \cdot 3 = 2.598 \approx 2.6$

Please turn over!

(i.e., on average, the predicted education level is 2.6 years off the observed education level)

Formulas:

$$\text{r.m.s. error} = \sqrt{1-r^2} \times SD_y$$

$$\text{slope} = r \times \frac{SD_y}{SD_x}$$

$$\text{intercept} = avg_y - \text{slope} \times avg_x$$

Grading Criteria:

in 1., 2., & 3.:
- −2 for each calculation error
- −2 for each incorrect value used
- −2 for swapping x and y

in 1.:
- −3 for incorrect formula for slope
- −3 for incorrect formula for intercept
- −2 if no final equation stated
- −1 if only part of the equation stated (eg., 6+0.50x)

in 2.:
- −3 for incorrect formula for prediction
- −1 if correct results, but according to old method

in 3.:
- −3 for incorrect formula for r.m.s. error

Statistics 1040, Section 006, Quiz 6 (20 Points)

October 18, 2002

Your Name: ________________

Question 1: Chance/Probability (**20 Points**)

1. A deck of 52 cards is shuffled and two cards are drawn without replacement.

   (a) (**3 Points**) What is the chance that the first card is a ♡ or a ♢?

   mutually exclusive events
   chance that 1st is ♡ : $\frac{13}{52}$
   chance that 1st is ♢ : $\frac{13}{52}$

   chance that 1st is ♡ or ♢ : $\frac{13}{52} + \frac{13}{52} = \frac{26}{52} = \frac{1}{2} = 50\%$

   Addition rule

   (b) (**4 Points**) What is the chance that the first card is a ♡ and the second card is a ♢ ?

   dependent events
   chance that 1st is ♡: $\frac{13}{52}$
   chance that 2nd is ♢, given that 1st is ♡: $\frac{13}{51}$

   chance that 1st is ♡ and 2nd is ♢:
   $\frac{13}{52} \cdot \frac{13}{51} = \frac{1}{4} \cdot \frac{13}{51} = \frac{13}{204} = 0.0637 = 6.4\%$

   (c) (**4 Points**) What is the chance that both cards are ♡ ?

   dependent events
   chance that 1st is ♡: $\frac{13}{52}$
   chance that 2nd is ♡, given that 1st is ♡: $\frac{12}{51}$

   Multiplication rule
   $\frac{13}{52} \cdot \frac{12}{51} = \frac{1}{4} \cdot \frac{12}{51} = \frac{3}{51} = 0.0588 = 5.9\%$
   chance that both cards are ♡:

   (d) (**4 Points**) What is the chance that neither card is a ♡?

   dependent events
   chance that 1st is not ♡: $\frac{39}{52}$
   chance that 2nd is not ♡, given that 1st is not ♡: $\frac{38}{51}$

   chance that neither card is ♡:
   $\frac{39}{52} \cdot \frac{38}{51} = \frac{3}{4} \cdot \frac{38}{51} = \frac{114}{204} = 0.5588 = 55.9\%$

   Multiplication rule

2. (**5 Points**) There are two options:

   (i) (a) You toss a coin 100 times; on each toss, if it lands heads you win \$1, if it lands tails you lose \$1.

   (ii) (b) You draw 100 times at random with replacement from the box $\boxed{1}\boxed{0}$. On each draw, you are paid (in dollars) the number on the ticket.

   Which option is better? Or are they the same? Explain briefly.

Chapter 13, Review question 11, page 236:
Workbook answer:
"Option (ii) is better. You have the same 50-50 chance of winning one dollar, but for option (ii) you have no way to lose money."

Grading Criteria:

1) general: −1 for calculation error
−1 for incorrect 1st chance
−2 for incorrect 2nd chance (e.g., if not conditional)
−2 for incorrect rule (e.g., addition ↔ multiplication)

2) −3 for answer (a)
−1 if no (or incorrect) explanation

Statistics 1040, Section 006, Quiz 7 (20 Points)

October 25, 2002

Your Name: ____________________

Chapter 17, Review question 4, p. 305

Question 1: Box Models, EV, and SE (**15 Points**)

A large group of people get together. Each one rolls a die 180 times and counts the number of ones. **Show your work!**

1. (**4 Points**) Find the box model.

$\underline{| \boxed{1} \quad 5 \times \boxed{0} |}$

number of draws: 180

die → box

$\boxed{1} \rightsquigarrow \boxed{1}$

$\boxed{2} \ldots \boxed{6} \rightsquigarrow \boxed{0}$

−2 for minor error

−3 for major error (e.g., box with $\boxed{1} \ldots \boxed{6}$)

−1 if number of draws not stated

2. (**5 Points**) Find the expected value for the number of ones.

$$\text{box average} = \frac{1 + 5\cdot 0}{6} = \frac{1}{6}$$

$$EV_{sum} = 180 \cdot \frac{1}{6} = 30$$

in 2. & 3.: −1 each calculation error

−2 for each minor error

−3 for each major error

3. (**6 Points**) Find the standard error.

$$\text{box } SD = \sqrt{\frac{\left(1-\frac{1}{6}\right)^2 + 5\cdot\left(0-\frac{1}{6}\right)^2}{6}}$$

$$= \sqrt{\frac{\left(\frac{5}{6}\right)^2 + 5\cdot\left(-\frac{1}{6}\right)^2}{6}}$$

$$= \sqrt{\frac{\frac{25}{36} + \frac{5\cdot 1}{36}}{6}}$$

$$= \sqrt{\frac{30}{216}} = \sqrt{0.1388} \approx 0.3727$$

$$SE_{sum} = \sqrt{180} \cdot 0.3727 = 13.416 \cdot 0.3727 = 5.00$$

Please turn over!

Question 2: Law of Averages (**5 Points**)

The meaning of "The probability of a Head is 1/2" in tossing a coin is best expressed by saying:

1. The coin has only two sides, so the chance of each is 1/2.
2. The coin will come up Heads exactly half the time: 50 Heads in 100 tosses, 500 Heads in 1000 tosses, and so on.
3. The odds against a Head are 2 to 1.
4. The fraction of tosses that come up Head will get ever closer to 1/2 as more tosses are made,

Explain your answer!

1. correct, but not the best choice
2. false – this is very unlikely to happen
3. false – actually the odds against a Head are 1 to 1
4. correct, and best choice: another way how to phrase the Law of Averages

−1 for correct answer, no explanation

−2 for 1. with explanation

−3 for 1., no explanation

−4 for 2. or 3. with explanation

−5 for 2. or 3., no explanation

Formulas:

$$\text{box average} = \frac{\text{sum of all numbers in box}}{\text{how many numbers in box}}$$

$$\text{box SD} = \sqrt{\text{average of [(deviations from box average)}^2]}$$

$$EV_{sum} = \text{number of draws} \times \text{box average}$$

$$SE_{sum} = \sqrt{\text{number of draws}} \times \text{box SD}$$

Statistics 1040, Section 006, Quiz 8 (20 Points)

November 1, 2002

Your Name: ____________________

Question 1: EV, SE, and Normal Curve (**20 Points**)

According to the U.S. Census Bureau, 68% of Utah residents are 18 years of age or older. Suppose that 200 Utah residents have been randomly chosen to participate in a survey.

1. (**4 Points**) Find the box model.

$\underline{|\ 68 \times \boxed{1} \quad 32 \times \boxed{0}\ |}$

number of draws = 200

1 = 18 years of age or older
0 = below 18 years
-1 if slightly incorrect number of $\boxed{0}$/$\boxed{1}$'s in box
-2 if box given as $|\boxed{0}\ \boxed{1}|$ etc.
-3 if box something else than $\boxed{0}$/$\boxed{1}$'s
-1 if number of draws missing or incorrect

2. (**8 Points**) Find the expected number of Utah residents in this sample of 200 who are 18 years of age or older. What is the corresponding SE?

box average $= \frac{68}{100} = 0.68 = 68\%$

box SD $= \sqrt{\frac{68}{100} \cdot \frac{32}{100}} = \sqrt{0.68 \cdot 0.32} = \sqrt{0.2176} = 0.466$

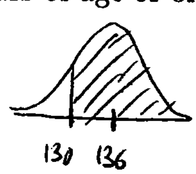
$EV_{sum} = 200 \cdot 0.68 = \underline{\underline{136}}$

$SE_{sum} = \sqrt{200} \cdot 0.466 = 14.14 \cdot 0.466 = \underline{\underline{6.59}}$

-1 for each calculation error
-1 for each minor mistake
-2 for each major mistake (or step missing)

3. (**8 Points**) Using the normal curve, find the chance that **at least** 130 of the Utah residents in the sample are 18 years of age or older.

s.u.: $\frac{130-136}{6.59} = -\frac{6}{6.59} = -0.91$



area between -0.90 and 0.90: 63.19%

area above -0.90: $50\% + \frac{63.19\%}{2} = 81.595\%$

$\approx \underline{\underline{81.6\%}}$

There is a chance of about 81.6% that at least 130 of the Utah residents in the sample are 18 years of age or older.

-1 for each calculation error
-2 for incorrect curve parameters, i.e., anything else than EV and SE
-2 for incorrect s.u.
-2 for incorrect table value
-2 for incorrect area under the curve

Please turn over!

Statistics 1040, Section 6, Quiz 9 (20 Points)

November 15, 2002

Your Name: ____________________

Chapter 20, Review question 3 (e), p. 371

Question 1: EV$_{\%}$, SE$_{\%}$, and Normal Curve (**20 Points**)

(−2) each calculation error

A group of 50,000 tax forms has an average gross income of \$37,000, with an SD of \$20,000. Furthermore, 20% of the forms have a gross income over \$50,000. A group of 900 forms is chosen at random for audit. Estimate the chance that between 19% and 21% of the forms chosen for audit have gross income over \$50,000. **Show your work!**

1: income over \$ 50,000
0: income under \$ 50,000

Note: avg = 37,000 & SD = 20,000 are not used to answer this question!

box: $\boxed{20 \times \boxed{1} \quad 80 \times \boxed{0}}$ or $\boxed{10{,}000 \times \boxed{1} \quad 40{,}000 \times \boxed{0}}$ (4)

number of draws: 900 number of draws: 900 (−1) if # draws missing

$$\text{box avg} = \frac{20}{100} = \frac{10{,}000}{50{,}000} = 0.2 \quad (2)$$

$$\text{box SD} = \sqrt{\frac{20}{100} \cdot \frac{80}{100}} = \sqrt{\frac{10{,}000}{50{,}000} \cdot \frac{40{,}000}{50{,}000}} = \sqrt{0.2 \cdot 0.8} = \sqrt{0.16} = 0.4$$

$$EV\% = 20\% \quad (1)$$

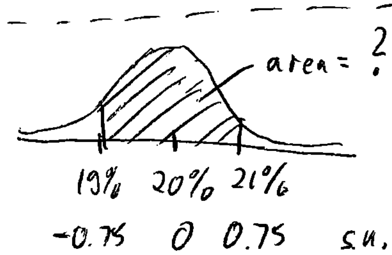
[$EV_{sum} = 900 \cdot 0.2 = 180$ not needed]

$$SE_{sum} = \sqrt{900} \cdot 0.4 = 30 \cdot 0.4 = 12 \quad (2)$$

$$SE\% = \frac{12}{900} \cdot 100\% = 1.33\% \quad (2)$$

$$\text{s.u.:} \quad \frac{19\% - 20\%}{1.33\%} = -0.75 \quad (2)$$

$$\frac{21\% - 20\%}{1.33\%} = 0.75 \quad (2)$$



area between −0.75 to 0.75: 54.67% ≈ 55% (4)

Please turn over!

Statistics 1040, Section 006, Quiz 10 (20 Points)

November 22, 2002

Your Name: ____________________

Chapter 23, Review question 3, p. 426

Question 1: Confidence Intervals (**20 Points**)

A real estate office wants to make a survey in a certain town, which has 50,000 households, to determine how far the head of household has to commute to work. A simple random sample of 1,000 households is chosen, the occupants are interviewed, and it is found that on average, the heads of the sample households commuted 8.7 miles to work; the SD of the distances was 9.0 miles. (All distances are one–way; if someone isn't working, or is working at home, the commute distance is defined to be 0.)

Fill the blanks in the statements below and **show your work!**

1. (**10 Points**) The average commute distance of all 50,000 heads of households in the town is estimated as 8.7 miles (1), and this estimate is likely to be off by 0.3 miles (1) or so.

box: unknown

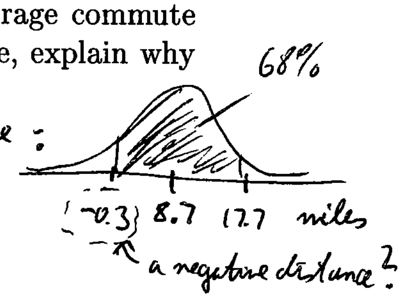
box avg = sample avg = 8.7 (2)

box SD = sample SD = 9.0 (2) bootstrap

$SE_{sum} = \sqrt{1000} \cdot 9.0 = 31.6 \cdot 9.0 = 284.4$ (2)

$SE_{avg} = \frac{284.4}{1,000} = 0.2844 \approx 0.3$ (2)

2. (**10 Points**) If possible, find a 95%–confidence interval for the average commute distance of all heads of households in the town. If this isn't possible, explain why not.



Note that the data do not follow the normal curve, otherwise:

(4) for yes/possible, any calculation based on avg

But:

The average of 1,000 draws will follow the normal curve.

So we can obtain confidence intervals:

$8.7 \pm 2 \cdot 0.3 = 8.7 \pm 0.6$

(2) (2) (2) $= 8.1$ to 9.3 miles

Please turn over!

−9 for no

−9 calculation for %

Statistics 1040, Section 006, Quiz 11 (20 Points)

December 2, 2002

Your Name: ____________________

Chapter 26, Review question 8, p. 499

Question 1: Tests of Significance(**20 Points**)

Bookstores like education, because national data show that 71% of college graduates have read a book in the past year, compared to 54% of the general population age 18 and over. The data also show the nationwide average educational level to be 13 years of schooling completed, with an SD of about 3 years, for persons age 18 and over.
A bookstore is doing a market survey in a certain county, and takes a sample of 1,000 people age 18 and over. They find the average educational level to be 14 years, and the SD is 5 years. Can the difference in average educational level between the sample and the nation be explained by chance variation? If not, what other explanations can you give? Please **follow the steps below** in answering these questions.

1. (5 points) State the null and the alternative hypothesis for this problem, in words and in terms of the box model.

Null: County's avg education level matches nationwide level, i.e, box avg = 13 years

Alternative: County's avg education level is different from nation, i.e., box avg $\neq$ 13 years

2. (5 points) Calculate the appropriate test statistic.

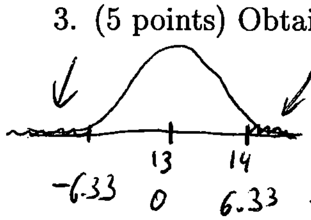
observed avg = 14
expected avg = 13
sample SD = 5
(use this!)

$SE_{sum} = \sqrt{1000} \cdot 5 = 158.1$

$SE_{avg} = \frac{158.1}{1{,}000} = 0.158$

$\rightsquigarrow z = \frac{\text{obs}-\text{exp}}{SE} = \frac{14-13}{0.158} = 6.33$

3. (5 points) Obtain the P–value (use the normal table on the back).



P-value is area on both sides, but this is about 0%

4. (5 points) State conclusions in terms of rejecting the null hypothesis and in your own words.

reject null hypothesis, result is highly statistically significant (P-value ≈ 0% < 1%); the avg education level for this county is different from the national average, more specifically it is above the national average; this could be a rich suburban county

Please turn over!

Grading Criteria:

1) Wrong null and alternative hypotheses, e.g.:

- swapped null and alternative — -3
- "14" in hypothesis instead of "13" — -3

Hypotheses stated in words only or in numbers only — -1 each

2) incorrect z, e.g. $\frac{\text{exp} - \text{obs}}{SE}$ or $\frac{\text{obs} - \text{exp}}{SD}$ — -2 [correct: $\frac{\text{obs} - \text{exp}}{SE}$]

incorrect SE_{avg} — -2

calculation error — -1 each

3) incorrect area — -2

incorrect table value — -2

calculation error — -1

4) reject null if p-value > 5% — -4

(or do not reject null if p-value < 5% — -4)

no explanation (e.g., reject, but no conclusion) — -2

correctly rejecting, but explanation mixed up — -2

if not speaking of rejecting / not rejecting — -2

if not speaking of (highly) statistically significant — -1

Statistics 1040, Section 006, Quiz 12 (20+ Points)

Due on or before December 11, 2002

Your Name: ____________________

This is a take–home quiz. You should work on it on your own and bring it to me on or before the final examination day. Please work on this quiz independently, getting as little help as possible from your friends, books, and notes.

Question 1:

(20 Points) A thermostat used in an electrical device is to be checked for the accuracy of its design setting of 200 degrees Fahrenheit. Ten thermostats were tested to determine their actual setting, resulting in the following data:

202.2	203.4	200.4	202.5	206.3	198.0	203.7	200.8	201.3	199.0

Is the mean setting of these thermometers different from 200 degrees Fahrenheit? State the null and the alternative hypothesis, calculate test statistic (after finding the average and SD of the sample), obtain the P–value, and clearly state your conclusions. Assume that the thermometer settings follow the normal curve.

$$\text{avg} = \frac{202.2 + \dots + 199.0}{10} = 201.76 \quad (2)$$

-15 ... -19

if no t-test

$$SD = \sqrt{\frac{(202.2 - 201.76)^2 + \dots + (199.0 - 201.76)^2}{10}} = 2.29 \quad (2)$$

1, null: mean setting of thermometers doesn't differ, i.e., avg = 200 F (1) (1)

alternative: mean setting of thermometers differs, i.e., avg ≠ 200 F (1) (1)

2, sample size < 30
SD unknown
data follows normal curve

$\Rightarrow$ t-test (two-sided!)

$$SD^+ = 2.29 \cdot \sqrt{\frac{10}{9}} = 2.41 \quad (1)$$

$$SE_{sum} = \sqrt{10} \cdot 2.41 = 7.62 \quad (1)$$

$$SE_{avg} = \frac{7.62}{10} = 0.76 \quad (1)$$

$$t = \frac{201.76 - 200}{0.76} = 2.32 \quad (1)$$

3, $df = 10 - 1 = 9$ (1)

t-statistic is 2.32: between 2.26 and 2.82 (1)

2.26 → 2.5%, 2.82 → 1%

two-sided! P-value is between 5% and 2% (2)

4, reject null; (2)

result is statistically significant (1)
(P-value between 2% and 5%),

there is some evidence that the mean setting of thermometers differs from 200 F (1)

The following questions are extra–credit questions. You may obtain a maximum of 20 extra–points if you complete both questions.

(*) Note that not every participant in an experiment produces a valid outcome.

Question 2:

(10 Points) In an experiment to study the dependence of hypertension on smoking habits, the following data were taken on 180 (*) individuals:

obs ¦ exp	Nonsmokers	Moderate Smokers	Heavy Smokers	
Hypertension	21 24	36 35	30 29	87
No hypertension	21 18	26 27	21 22	68
	42	62	51	155 (*)

Is the presence or absence of hypertension independent of smoking habits? Conduct an appropriate statistical test to answer this question.

-7 ... -9 if no χ^2-test for independence

1, null: hypertension and smoking are independent, i.e, boxes are the same (1)

alternative: hypertension and smoking are dependent, i.e, at least 1 box differs (1)

2) χ^2-test for independence

expected: $\frac{42 \cdot 87}{155} = 23.6 \approx 24 \quad \frac{62 \cdot 87}{155} = 34.8 \approx 35 \quad \frac{51 \cdot 87}{155} = 28.6 \approx 29$

18 27 22 (2)

Question 3: (see next page)

(10 Points) A study was made to estimate the difference in salaries of college professors in private and state colleges of North Carolina. A random sample of 100 professors in private colleges showed an average 9–month salary of \$32,000 with a standard deviation of \$1300. A random sample of 200 professors in state colleges showed an average salary of \$32,900 with a standard deviation of \$1400. Is there any statistical evidence that professors in state colleges have **higher average salaries** than professors in private colleges? Conduct an appropriate statistical test to answer this question.

$$\chi^2 = \frac{(21-24)^2}{24} + \frac{(36-35)^2}{35} + \frac{(30-29)^2}{29} + \frac{(21-18)^2}{18} + \frac{(26-27)^2}{27} + \frac{(21-22)^2}{22} = 1.02$$ (2)

3, $df = (2-1)\cdot(3-1) = 2$ (1)

χ^2-statistic is 1.02: between 0.71 and 1.39

p-value is between 70% and 50% (1)

4, do not reject null. (1)

based on this particular study, (1) hypertension and smoking are independent, i.e, this study does not provide enough evidence that smoking is associated with hypertension (although another study might show...)

Question 3

box A: private

avg_A: 32,000

SD_A: 1,300

sample size$_A$: 100

box B: state

avg_B: 32,900

SD_B: 1,400

sample size$_B$: 200

−7...−9 if no 2-sample z-test

1, null: avg salaries are the same, i.e; $avg_B - avg_A = 0$ 1

alternative: avg salaries in state colleges are higher, i.e, $avg_B - avg_A > 0$ 1

2, **2-sample z-test:**

obs difference: $32{,}900 - 32{,}000 = 900$

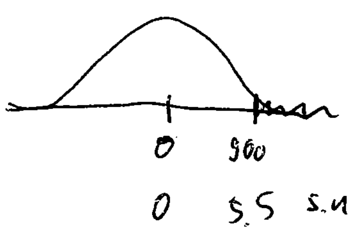
exp. difference: 0

$$\left.\begin{aligned} SE_{sum_A} &= \sqrt{100}\cdot 1{,}300 = 13{,}000 \\ SE_{avg_A} &= \frac{13{,}000}{100} = 130 \end{aligned}\right\} 1 \qquad \left.\begin{aligned} SE_{sum_B} &= \sqrt{200}\cdot 1{,}400 = 19{,}799 \\ SE_{avg_B} &= \frac{19{,}799}{200} = 99 \end{aligned}\right\} 1$$

$$SE_{diff} = \sqrt{130^2 + 99^2} = 163.4 \quad 1$$

$$z = \frac{900-0}{163.4} = 5.5 \quad 1$$

3,



P-value ≈ 0% 1

4, reject null; 1

result is highly statistically significant (P-value < 1%); 1

there is considerable evidence that professors have higher average salaries in state colleges than in private colleges (in North Carolina) 1