Statistics 1040 Fall 2010 Exam 1

Do not open the exam until you are instructed to do so.

Directions: You have 75 minutes to complete the exam, which has 100 total possible points. Be sure to answer every question. Where numeric answers are required, you must show your work to receive credit. The formulas given below and the table on the last page may be useful for some of the questions. Where calculations are required, round to two decimal places.

Student Name: (Solutions)

Under your recitation leader's name, circle your recitation time:

Oleksandr Gromenko	Bryan Stephenson	Anthonie Nichols
10:30	8:30	8:30
11:30	12:30	10:30
12:30	1:30	11:30
		1:30

Possibly useful "formulas" for this exam

- $SD = \sqrt{[\text{ave. of (value}^2)] - (\text{ave. of value})^2}$
- $r = \dfrac{(\text{ave. of x·y}) - (\text{ave. of x})(\text{ave. of y})}{(\text{SD for x})(\text{SD for y})}$
- $slope = r \cdot \dfrac{\text{SD for y}}{\text{SD for x}}$
- rms error $= (\sqrt{1 - r^2}) \cdot (\text{SD for y})$
- calculator average: $\bar{x}$
- calculator SD: σ_n

1. A June 2010 article in Newsweek states that "observational studies going back to 1991 found that ... lower levels of homocysteine (an amino acid in the blood) are associated with a lower risk of [heart attack]." It is known that folic acid and vitamin B12 cause a reduction in homocysteine levels, and for several years many health experts and the press have recommended taking folic acid to reduce the risk of heart attack.

   This article goes on to report on a new double-blind randomized controlled experiment. According to this article, the study found that those who took folic acid and vitamin B12 did not have a lower risk of another heart attack, compared to those who did not take the folic acid and B12. The Newsweek article concludes by saying, "you have to wonder how many more times we - the press as well as supposed experts - will make the mistake of basing health advice on observational studies."

   (a) In one or two brief sentences, explain what it means for an experiment like this new one to be:

      i. (6 points) "controlled"

      key: researchers control assignment to treatment and non-treatment groups

      [+3 if only say there was a control group; +2 if only say control confounding; full points if identify two separate groups]

      ii. (6 points) "randomized"

      key: subjects are assigned at random to receive treatment or not

      [+3 if only say randomly selected; +2 if only say eliminates confounding]

      iii. (6 points) "double-blind"

      key: neither subjects nor evaluators know the subjects' treatment assignment

      [+3 if only specify subjects or evaluators]

   (b) (4 points) The Newsweek article also states that "In the case of homocysteine and heart health, it isn't clear yet why the observational studies - high homocysteine equals higher risk of cardiovascular disease - were misleading. It might be that high levels of homocysteine are a marker for the real culprit, and fixing it leaves that culprit unscathed." What is the statistical term we have used to refer to this "real culprit"?

      confounding factor [causation okay]

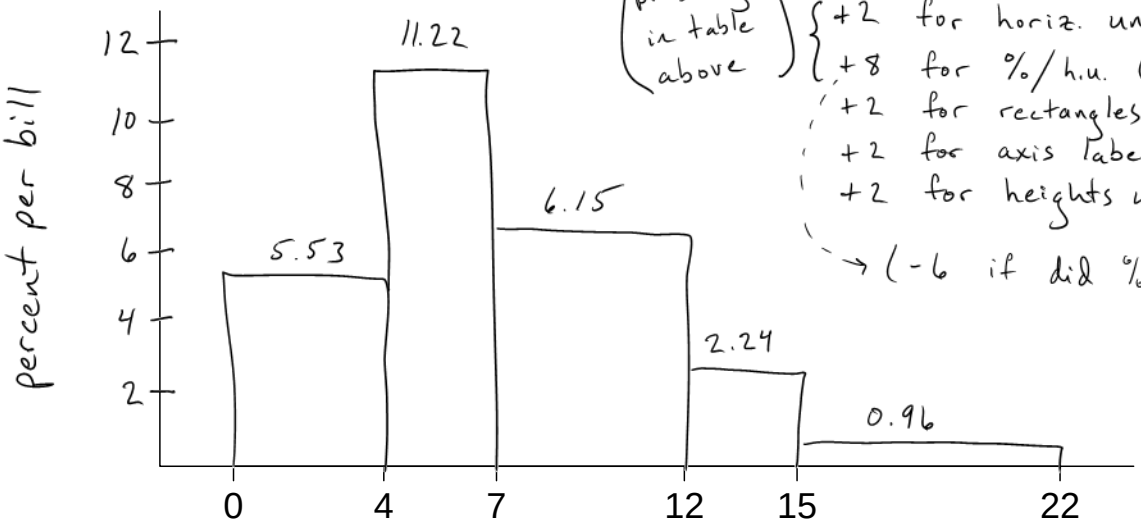
2. In the 2009 session of the Utah Legislature, a total of 706 bills were filed by the 104 members of the legislature. The table below summarizes the numbers of bills filed by each member of the legislature; as with the textbook and classroom examples, intervals include the left but not the right endpoint.

Number of Bills Filed	Count	Percent	horizontal units	% / h.u.
0 to 4	23	22.12	4	5.53
4 to 7	35	33.65	3	11.22
7 to 12	32	30.77	5	6.15
12 to 15	7	6.73	3	2.24
15 to 22	7	6.73	7	0.96
(Total: 706)	104	100.00		

(a) (2 points) What is the average number of bills filed per legislator?

$$\frac{706}{104} = 6.79$$

(b) (16 points) Draw a histogram to summarize these data. Be sure to label both axes, and clearly write the heights on your histogram.



Partial credit:
(probably in table above)
+2 for horiz. units
+8 for % / h.u. (correct heights)
+2 for rectangles
+2 for axis labels
+2 for heights written
→ (-6 if did % × h.u.)

(c) (4 points) The Senate President filed 5 bills; in what percentile is he for this distribution?

$4 \times 5.53 + 1 \times 11.22 = 33.34$, or 33rd percentile

[+2 for using base × height to find areas of rectangles in histogram;
+2 for using correct rectangles over 0-4 and 4-5;
-3 for 33.65 or 4-7 or 11.22; -2 for 5.53 + 11.22 = 16.75]

3. (8 points) In the 2009 session of the Utah Legislature, the eight Democrats in the Senate filed a total of 53 bills, as summarized in the table below. Calculate the SD of the number of bills filed by Democrats in the Senate.

Senator Initials	Number of Bills Filed	value^2
P.J.	6	36
K.M.	8	64
B.G.	4	16
G.D.	7	49
S.M.	15	225
R.R.	5	25
L.R.	7	49
K.M.	1	1
ave:	6.63	58.13
	+1	+3

Calculator:
+8 for $\sigma_n = 3.77$
+6 if $\sigma_n = 3.33$ (didn't hit M+ after last data value)
+6 for $\sigma_{n-1} = 4.03$
+2 for any other calculator SD

$SD = \sqrt{58.13 - 6.63^2}$ } +3
$= \sqrt{14.17} \approx 3.76$ } +1

4. The histogram below represents the distribution of the ages of 30 pregnant women (in years) at the time they gave birth, with ages ranging from 12 to 35 years, and with area representing percent. The median is 21 years, the average is 21.8, and the SD is 4.3.

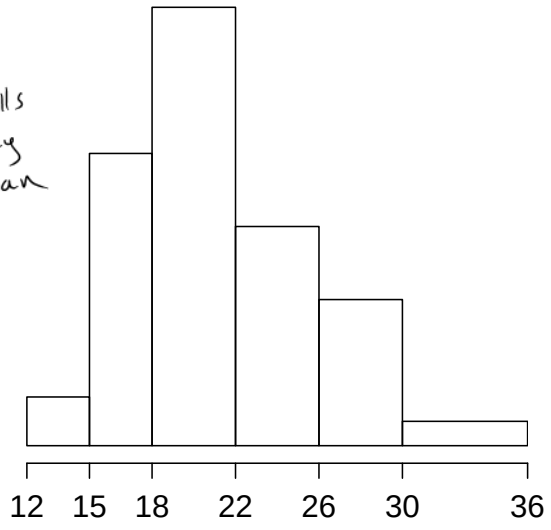
(a) (3 points) This distribution is (circle one):

+1 i. skewed left
+3 (ii.) skewed right
+0 iii. symmetric

key: long tail pulls average away from median

(b) (3 points) If we added a 50-year old delivering mother to this data set, the SD would (circle one):

+3 (i.) increase
+0 ii. decrease
+0 iii. stay the same

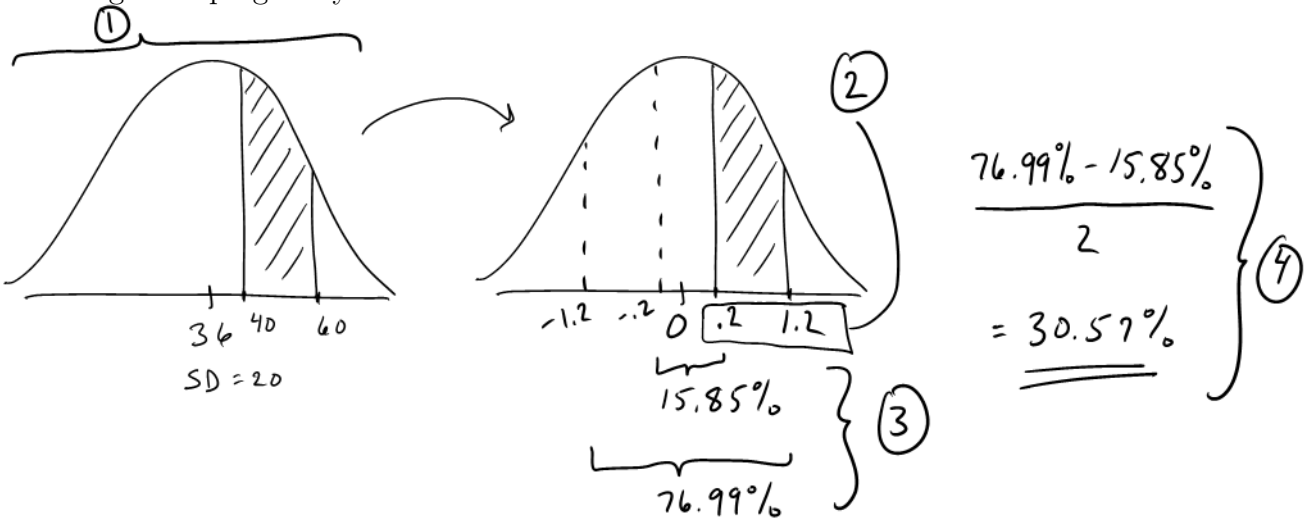


5. (3 points) A team of marketing researchers has the question, "Do men's preference for salty snacks change as the men age?" The type of study that would be most appropriate to address this question is (circle one):

cross-sectional (+0) longitudinal (+3)

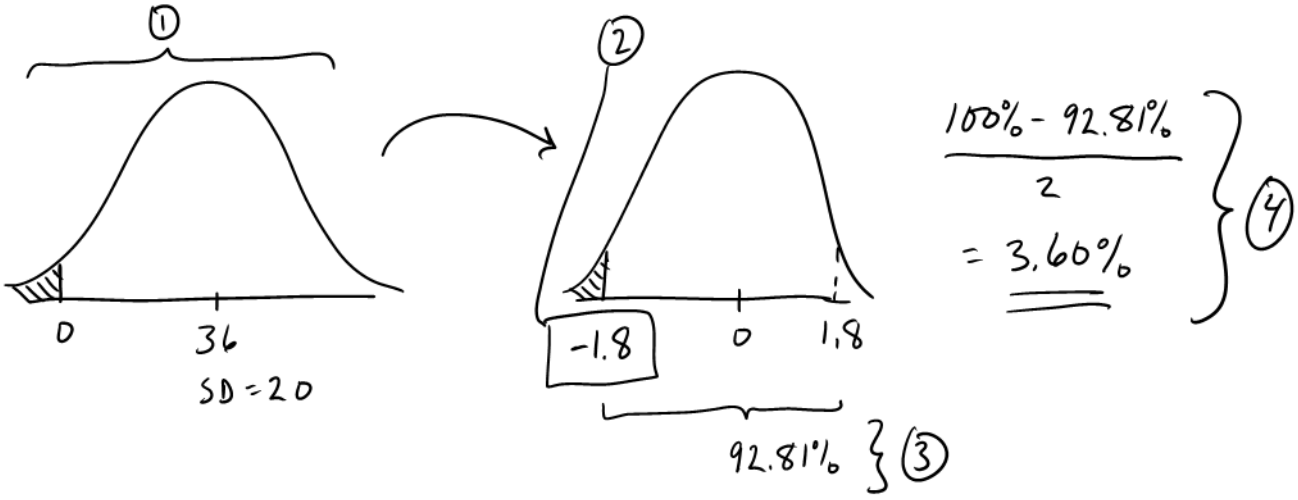
6. In a study of over 20,000 pregnant women participating in the WIC (Women, Infants, and Children) health and nutrition program, researchers found that the distribution of weight gain during pregnancy was close to the normal curve, with an average gain of 36 pounds and SD 20 pounds.

   (a) (9 points) About what percentage of women gained between 40 and 60 pounds during their pregnancy?



Partial credit: ① +2 for picture (including area of interest)
② +2 for conversion to standard units
③ +2 for areas from normal table
④ +2 for correct usage of areas from normal table

   (b) (9 points) About what percentage of women lost weight during their pregnancy?



Partial credit: same as in (a) above.

[Full points if used area to left of -1 instead of 0; then $z = -1.85$, middle area = 93.57%, left tail = 3.22%]

7. In a study of pregnant women participating in the WIC (Women, Infants, and Children) health and nutrition program, there were close to 14,000 non-Latino women. Researchers were interested in the relationship between these women's BMI (body mass index) at the beginning of pregnancy and their weight gain (in pounds) during pregnancy. The scatter plot of BMI versus weight gain was roughly football-shaped, and the correlation coefficient was $r = -0.22$. (Note that $\sqrt{1-r^2} \approx 0.98$.)

y weight gain	x BMI
ave = 36.70	ave = 25.78
SD = 19.37	SD = 6.78

← If switched x & y, same point allocation as below, but lose 6 points on (a): $y = -.08x + 28.72$

(a) (12 points) Find the equation of the line to predict weight gain from BMI.

+2 { $y = mx + b$

$m = r \cdot \frac{SD_y}{SD_x} = (-0.22) \cdot \frac{19.37}{6.78} = -0.63$

(+2) (+2)

$(\text{ave. } y) = m \cdot (\text{ave. } x) + b \rightarrow 36.70 = -0.63(25.78) + b \rightarrow b = 52.94$ } +2

(+2) { $y = -0.63x + 52.94$

(b) (2 points) Predict the weight gain during pregnancy for a non-Latino woman whose BMI at the beginning of pregnancy is 24.5.

[Full points if used equation (a) correctly] $-0.63(24.5) + 52.94 = 37.51$

(c) (7 points) Would you be surprised if the woman referred to in part (b) gained 10 pounds during pregnancy? Circle one and explain briefly using statistical justification:

Yes (No) } +1

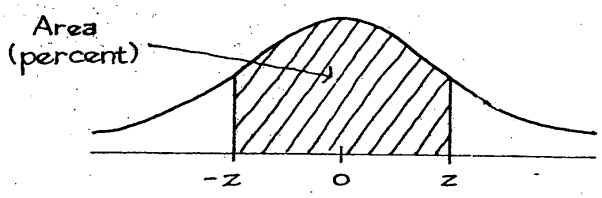
+2 { rms error $= \sqrt{1-r^2}\, SD_y = 0.98 \times 19.37 = 18.98$

+2 { $\frac{10 - 37.51}{18.98} = -1.45$

+2 { 10 pounds is 1.45 (fewer than 2) rms errors away from the predicted 37.51 pounds

(-2 if SD instead of rms error)

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