2013 Autumn STAT 22000 Session 02 Midterm

Name (print): ______________________________

THIS IS THE MIDTERM FOR THE 1:30 SESSION (Yibi).

Raise your hand to get the right exam
if you sit in the 10:30 session (Zhou) .

1. Do not sit directly next to another student.
2. Do not turn the page until told to do so.
3. **If a question asks that you do some calculations, you must show your work to receive full credit.**
4. If you do not have enough room for your work in the place provided, use the last blank page. Be sure to mark clearly which problem the material on the back of any page refers to. If you pull the pages apart, sign all the pages.
5. Whenever appropriate, parts of a question will be graded conditionally on how you answered the preceding part(s). For example, even if you get part (a) of a question wrong, you will still get credit for the rest of the question provided your answers to parts (b), (c), etc. are consistent with how you answered part (a).
6. If you are unsure of what a question is asking for, **do not hesitate to ask Yibi for clarification**.
7. This exam has 12 pages.

Question	*Points Available*	*Points Earned*
MLB 2013	6	
Foot length	8	
Hollywood Movies I	16	
Hollywood Movies II	18	
Texting while Driving	6	
Lie Detector	12	
Laptop Warranty	31	
Pets Survey	4	
TOTAL	101	

1. [MLB 2013] [6 points]

The Major League Baseball consisting of 15 teams from each of the two leagues: the American League (AL) and the National League (NL). The main difference between the two leagues is that pitchers take at bats in the National League but not in the American League. The table below lists the total number of homeruns hit during the 2013 season (except playoffs) for each of the 30 teams.

Team	League	Homeruns	Team	League	Homeruns	Team	League	Homeruns
ARI	NL	130	HOU	AL	148	PHI	NL	140
ATL	NL	181	KCR	AL	112	PIT	NL	161
BAL	AL	212	LAA	AL	164	SDP	NL	146
BOS	AL	178	LAD	NL	138	SEA	AL	188
CHC	NL	172	MIA	NL	95	SFG	NL	107
CHW	AL	148	MIL	NL	157	STL	NL	125
CIN	NL	155	MIN	AL	151	TBR	AL	165
CLE	AL	171	NYM	NL	130	TEX	AL	176
COL	NL	159	NYY	AL	144	TOR	AL	185
DET	AL	176	OAK	AL	186	WSN	NL	161

(a) [5 points] Make a back-to-back stemplot of the numbers of homeruns for the two leagues.

(b) [1 point] Does it appear that teams in the American League have more homeruns?

2. [Foot Length] [8 points]

The average foot length for male adults in the U.S. is about 10.5 inches with a standard deviation of 0.5 inch. A histogram for the foot lengths is very close to the normal curve.

(a) [4 points] What percentage of male adults in the U.S. have feet longer than 11.8 inches?

(b) [4 points] If a guy has foot length at the 90th percentile, how long are his feet?

3. [Hollywood Movies I] [16 points]

Opening weekend box office revenue is an important source of income to the movie industry and a crucial preliminary indicator of long-run profitability of a motion picture. The following are some numerical summaries for the Opening-Weekend Revenues (in millions of dollars) for the 136 Hollywood movies in 2011.

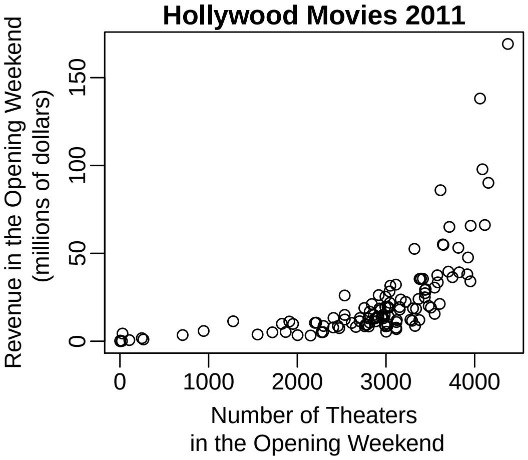
```
> summary(OpeningWeekend)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    7.71   13.10   20.34   25.00  169.20
> sd(OpeningWeekend)
[1] 24.80566
```

(a) [2pts] Is the movie with the highest Opening-Weekend Revenue, $169.20 million, an outlier based on the 1.5× IQR rule? (In fact, that movie was "*Harry Potter and the Deathly Hallows, Part 2*").

(b) [4pts] Sketch a boxplot (not the modified boxplot) of the Opening-Weekend Revenues. Draw the plot to scale. Clearly label the revenue values along the axis for each element of the boxplot.

(c) [2pts] Give one reason why the five-number summary is a better numerical summary for the Opening-Weekend Revenue than the mean and the standard deviation.

One way to increase the opening weekend box office revenue is to increase the number of theaters showing the movie. The scatter plot below shows the revenue of 136 Hollywood movies in 2011 versus the number of theaters showing them in the opening weekend. Their correlation coefficient is $r = 0.589$.



(d) [2pts] Describe the relationship (form, direction, strength) between the two variables based on the scatter plot.

(e) [2pts] Why is the correlation coefficient r NOT appropriate for measuring the strength of association between the two variables in the scatterplot?

(f) [2pts] If John, a film producer, wants to predict the Opening-Weekend Revenue of a movie shown in 3000 theaters in the opening weekend using simple linear regression, his prediction will ____________________ the actual Opening-Weekend Revenue. The blank should be

(i) tend to underestimate
(ii) tend to overestimate
(iii) about equally likely to be above or below

Circle one and explain briefly.

(g) [2pts] Which of the following words best describes the shape of the histogram of the "number of theaters during the opening weekend"?

(i) right-skewed (ii) symmetric (iii) left-skewed
(iv) bimodal (v) uncertain, not enough information

Circle one. No explanation is required.

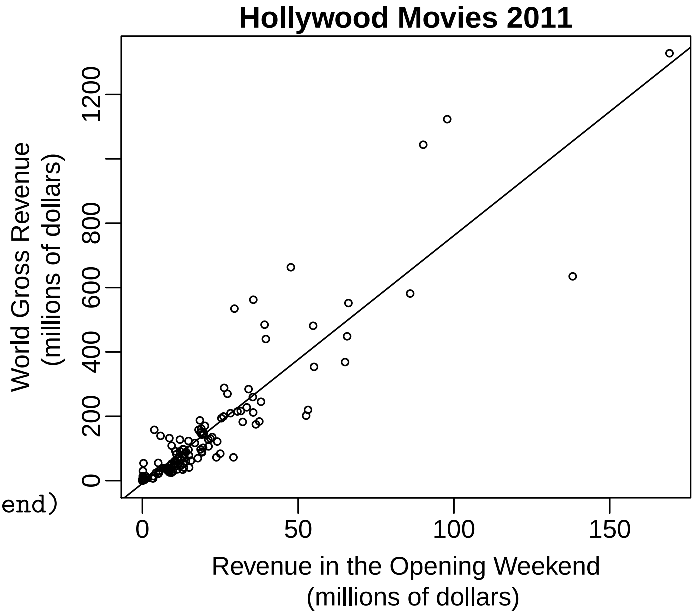
4. [Hollywood Movies II] [18 points]

Opening weekend box office revenue is an important source of income to the movie industry and a crucial preliminary indicator of long-run profitability of a motion picture. Here are some R outputs as well as a scatter plot for the Opening-Weekend Revenue and World Gross Revenue for the 136 Hollywood movies in 2011.

```
> mean(OpeningWeekend)
[1] 20.50153
> sd(OpeningWeekend)
[1] 24.96158
> mean(WorldGross)
[1] 149.217
> sd(WorldGross)
[1] 212.7556
> cor(OpeningWeekend, WorldGross)
[1] 0.9035918
> lm(WorldGross ~ OpeningWeekend)

Call:
lm(formula = WorldGross ~ OpeningWeekend)

Coefficients:
   (Intercept)  OpeningWeekend
        -8.678           7.702
```



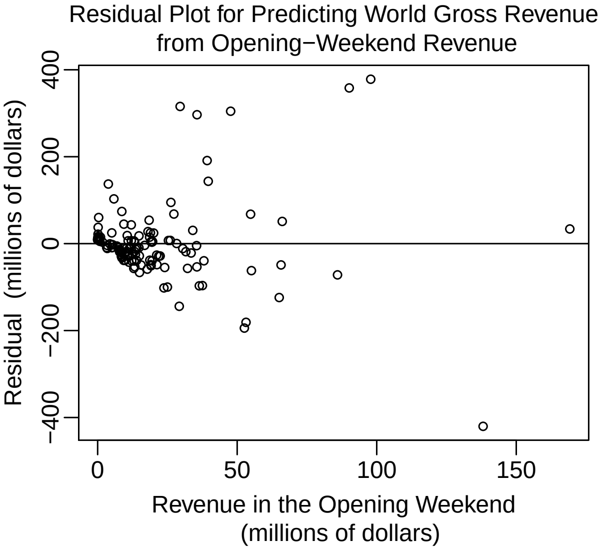
(a) [2 points] Write down the equation of the regression line for predicting the World Gross Revenue of a movie from its Opening-Weekend Revenue.

(b) [3 points] The movie “*Harry Potter and the Deathly Hallows, Part 2* had an Opening-Weekend Revenue of $169.19 millions and a World Gross Revenue of $1328.111 millions. Using the regression line, what is its predicted World Gross Revenue? What is the residual (prediction error)?

(c) [1 point] On the scatter plot, circle the dot has the largest NEGATIVE residual.

(d) [2 points] What fraction of the variance of the World Gross Revenue was explained by the least square regression on the Opening-Weekend Revenue?

(e)



[6 points] The plot on the right is the residual plot for predicting the World Gross Revenue from the Opening-Weekend Revenues using simple linear regression. For each of the four statements below, determine whether it is TRUE or FALSE based on facts about residuals and the residual plot on the right.

____ The prediction for World Gross Revenue is more accurate for movies with larger Opening-Weekend Revenues.

____ The average of the residuals is about 0, though may not be exactly 0.

____ The standard deviation of residuals is smaller than the standard deviation for the World Gross Revenues.

____ On the scatter plot (not the residual plot), the residual of a point is the shortest (signed) distance from that point to the regression line.

(f) [4 points] Write down the equation of the regression line for predicting the Opening-Weekend Revenue from the World Gross Revenues. (Note this line is different from the one in part (a).)

5. [Texting while Driving] [6 points]
To investigate whether or not sending text messages while driving impacts driving ability, we have 100 participants (50 men and 50 women) drive an obstacle course under two conditions:

(1) No texting while driving, and
(2) Sending five text messages while driving.

We measure the accuracy the subjects drove the obstacle course from a scale of 1 to 10 (1 = poor and 10 = excellent).

(a) [3 points] What is the advantage of using a block design over the completely randomized design for this study? What should we block on?

(b) [2 points] Briefly describe how to assign the 100 participants to the two groups using a randomized block design?

(c) [1 point] Is this study blinded? Explain in one sentence.

6. [Lie Detector] [12 points]

To reduce theft among employees, a company requires all of its employees to take a lie-detector test. The test asks if the employee has ever stolen from the company.

The test has been proven to correctly identify guilty employees 90% of the time and innocent employees pass the test at a rate of 96%.

All of the employees have to complete the test. Suppose that 5% of the employees are actually guilty (they steal from the company).

(a) [3 points] What percentage of employees will fail the test?

(b) [5 points] What percentage of those employees fail the test will actually be innocent? Why it is not fair to fire all employees who fail the test?

(c) [4 points] If those who fail the lie-detector test are given an independent second test, what percentage of those employees who fail both tests will actually be innocent?

7. [Laptop Warranty] [31 points]
Suppose a small company buys five laptops of the same model. Because the company is too small to hire IT technicians, whenever a laptop breaks down, the company simply replaces it with a new one, which costs $1,000. From past experience, a laptop will be broken in 3 years with probability 0.2. Suppose the breakdown of one laptop is independent of the breakdown of the rest.

(a) [3 points] What is the probability that at least one of the 5 laptops break down in 3 years?

(b) [3 points] What is the probability that EXACTLY 2 of the 5 laptops break down in 3 years?

(c) [3 points] Suppose the manufacturer sells a three-year warranty, which promises to fix any problem incurred or replace it with a new one. The warranty charges $99 per laptop. Should the company purchase the warranty?

For simplicity, suppose only two parts in a laptop can cause its breakdown — the screen and the motherboard, which happen in 3 years with probability 0.1 and 0.1, and cost \$100 and \$300 to fix, respectively. Suppose the laptop can breakdown at most once in 3 years. Further more, suppose the breakdown of parts in one laptop is independent of the breakdown of others.

(d) [2 points] Based on the assumptions above, are the breakdown of the motherboard and the breakdown of the screen independent? Explain briefly.

(e) [4 points] Let X be the random variable for the potential cost of the manufacturer to cover one laptop. What are the possible values that X may take? Write down the probability distribution of X.

(f) [4 points] Find the mean and the variance of X.

(g) [4 points] Suppose the manufacturer sold 1000 warranties (to cover 1000 laptops). What are the expected value (i.e., the mean) and the standard deviation of the total cost to cover these 1000 laptops?

(h) [4 points] What is the probability that total cost for cover these 1000 laptops is less than \$45,000 [1]?

[1] Recall the manufacturer will have a revenue of $\$99 \times 1000 = 99{,}000$ by selling 1000 warranties. If the total cost is less than \$45,000, the manufacturer will have a net profit of \$99,000-\$45,000=\$44,000.

(i) [4 points] What is the probability that among these 1000 covered laptops, more than 220 of them will breakdown in 3 years?

8. [Pets Survey] [4 points]
The state of Illinois wants to know about the fertility of owned dogs in Illinois. They select a simple random sample of 50 households from each county in the state and ask how many dogs they have, and are they spayed or neutered.

(a) [2 points] What type of sampling method was used to collect the data?

(i) systematic county sampling (ii) stratified sampling
(iii) multistage cluster sampling (iv) simple random sampling

Circle one. No explanation is required.

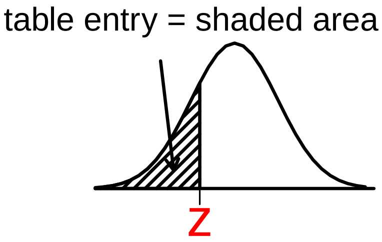
(b) [2 points] Fill in the blank. The average number of dogs per household for the Cook County in Illinois is a ____________________.

(i) population (ii) sample (iii) parameter (iv) statistics

Circle one. No explanation is required.

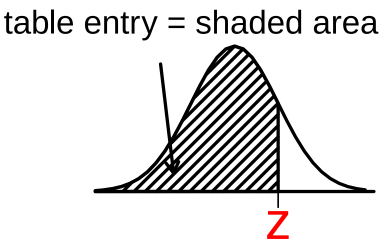
. . . . END OF EXAM . . . .

. . . . YOUR SCRATCH PAPER . . . .



Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998