FINAL EXAM REVIEW PROBLEMS

These questions are adapted from material provided courtesy of Debaleena Sain.

1. The following data represents the average yearly rainfall (in inches) in Riverside:

Year	2013	2014	2015	2016	2017	2018
Rainfall	3.4	5.3	7.5	8.8	8.5	7.2

   (a) Find the mean yearly rainfall over these 6 years.

   (b) Find the standard deviation of average yearly rainfall for these 6 years.

   (c) Find the median and range for the rainfall data.

2. You decide to protect your jewel collection by putting them in a drawer with a bunch of fake jewels. You have 3 real jewels and 13 fake ones. Your cat breaks into the drawer and loses 5 of your jewels. What is the probability that he did not lose any of the real jewels?

3. UC Riverside wanted to know about people's favorite animals in order to develop better stress relief fairs. A survey was conducted among 50 randomly selected UCR undergrads. The survey found that 27% of students prefer dogs (D), 31% prefer cats (C), and 22% love both.

   (a) What is the population?

   (b) What is the experimental unit?

   (c) What is the variable? Is it qualitative or quantitative?

   (d) What is the probability that a randomly selected student likes either animal?

   (e) Are the preference of cats and the preference of dogs independent?

   (f) Are the preference of cats and the preference of dogs disjoint?

4. Suppose a lizard expert claims that 50% of bearded dragons (a type of lizards) are female. I want do a study to test this claim and I want to be accurate within 2% at the 95% level of confidence. How many lizards will I need to sneak past my husband in order to conduct this study in my house?

5. The average monthly income in California is well-approximated by a normal distribution with mean \$6500 and standard deviation \$2100.

   (a) Find the proportion of people who have monthly incomes between \$5000 and \$10000. Your final answer should be written in terms of left-tail probabilities for Z.

   (b) Find the 90th percentile of the monthly income of California households. Your final answer should be written in terms of left-tail probabilities for Z.

6. A researcher believes that daily Vitamin C consumption improves immunity. In a self-reported study of 1337 people who took Vitamin C regularly for six months, 891 reported that their health conditions improved.

   (a) Use the confidence interval approach to test this claim at the 5% level of significance.

   (b) Use the test statistic approach to test this claim at the 1% level of significance.

7. The average weight of a male labrador retriever is believed to be well-approximated by a normal distribution with mean 71.5 lbs and standard deviation 7.5 lbs. A veterinarian is skeptical of this claim and takes a random sample of 15 labs that come into her clinic. The 15 labs have an average weight of 80 lbs with standard deviation 5.1. Use the test statistic approach at the 5% level to test if the vet's data suggests that our original belief is incorrect.

No. of Cigarettes	8	6	5	3	10
CD4-T Cell Count (per mm^3)	202.4	427.1	219.3	529.19	111.7

8. Some HIV/AIDS research suggests that cigarette smoking has an impact on T-cell counts. The table above shows the T-cell counts for 5 HIV+ individuals and the average number of cigarettes they smoke per day.

   (a) Which is the response and which is the predictor variable?

   (b) The regression line for this data is given by $\hat{y} = 633.1889 - 52.5392x$. Find the residuals for all 5 HIV+ individuals.

   (c) Predict the T cell count for an HIV+ person who smokes on average 2 cigarettes per day. Do you have any concerns with this prediction?

   (d) Find the correlation coefficient for this data. Based on this correlation, how would you describe the relationship between the two variables?

Table 1: Depression Data

Depression Score x	0	1	2	3
Proportion of People $P(x)$	0.614	0.196	0.128	0.062

9. Table 1 shows the distribution of US people who were diagnosed with depression (measured on the Hamilton scale), where 0 indicates no depression and 3 indicates very severe depression.

   (a) Find the mean and standard deviation of the depression score.

   (b) What proportion of people are above the expected depression score?

   (c) Consider two categories: (1) people with no depression and (2) people with some level of depression. A random sample of 10 patients were diagnosed in a clinic. Let Y be the random variable which counts number of people with some level of depression. What is the distribution of Y?

   (d) Find the mean and variance of Y, the number of people out of 10 with depression.

   (e) What is the probability that more than 8 patients out of 10 will have some level of depression?

Variance

$$s^2 = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2$$

Probability

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

Random Variables

$$\begin{aligned} E(X) &= x_1 \times P(X = x_1) + \cdots + x_k P(X = x_k) \\ &= \sum_{i=1}^{k} x_i P(X = x_i) \end{aligned}$$

$$\begin{aligned} Var(X) &= (x_1 - \mu)^2 \times P(X = x_1) + \cdots + (x_k - \mu)^2 \times P(X = x_k) \\ &= \sum_{j=1}^{k} (x_j - \mu)^2 P(X = x_j) \end{aligned}$$

$$E(aX + bY) = aE(X) + bE(y)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$$

Distributions

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}, \qquad E(X) = np, \qquad Var(X) = np(1-p)$$

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad E(X) = \lambda, \qquad Var(X) = \lambda, \qquad e \approx 2.718282$$

$$X \sim N(\mu, \sigma), \qquad z = \frac{x - \mu}{\sigma}, \qquad E(X) = \mu, \qquad Var(X) = \sigma^2$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \qquad E(\bar{X}) = \mu, \qquad Var(\bar{X}) = \sigma^2/n$$

Regression

$$e_i = y_i - \hat{y}$$

$$R = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Confidence Intervals:

$$\textit{point estimate} \pm (\textit{critical value}) \times (\textit{standard error})$$

$$n \geq \left(\textit{critical value} \times \frac{sd}{MoE}\right)^2$$

Case	Test Statistic	Confidence Interval
p, $np \geq 10$ and $n(1-p) \geq 10$	$z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	$\hat{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$
$\bar{x}$, $n \geq 30$	$z = \frac{\bar{x}-\mu_o}{s/\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
$\bar{x}$, $n < 30$, $x \sim N(\mu, \sigma)$, σ known	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$
$\bar{x}$, $n < 30$, $x \sim N(\mu, \sigma)$, σ unknown	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2,(n-1)}\left(\frac{s}{\sqrt{n}}\right)$

Critical Values for z

$(1-\alpha)100\%$	90%	95%	98%	99%
$z_{\alpha/2}$	1.645	1.96	2.33	2.575

Critical Values for t: $t_{\alpha/2,(n-1)}$

	$(1-\alpha)100\%$			
$(n-1)$	90%	95%	98%	99%
1	6.3137	12.706	31.821	63.657
2	2.9200	4.3026	6.9646	9.9248
3	2.3534	3.1824	4.5407	5.8409
4	2.1319	2.7765	3.7470	4.6041
5	2.0151	2.5706	3.3649	4.0321
6	1.9432	2.4469	3.1427	3.7074
7	1.8946	2.3646	2.9979	3.4995
8	1.8596	2.3060	2.8965	3.3554
9	1.8331	2.2622	2.8214	3.2498
10	1.8125	2.2281	2.7638	3.1693
11	1.7959	2.2010	2.7181	3.1058
12	1.7823	2.1788	2.6810	3.0545
13	1.7709	2.1604	2.6503	3.0123
14	1.7613	2.1448	2.6245	2.9768
15	1.7530	2.1315	2.6025	2.9467
16	1.7459	2.1199	2.5835	2.9208
17	1.7396	2.1098	2.5669	2.8982
18	1.7341	2.1009	2.5524	2.8784
19	1.7291	2.0930	2.5395	2.8609
20	1.7247	2.0860	2.5280	2.8453
21	1.7207	2.0796	2.5177	2.8314
22	1.7171	2.0739	2.5083	2.8188
23	1.7139	2.0687	2.4999	2.8073
24	1.7109	2.0639	2.4922	2.7969
25	1.7081	2.0595	2.4851	2.7874
26	1.7056	2.0555	2.4786	2.7787
27	1.7033	2.0518	2.4727	2.7707
28	1.7011	2.0484	2.4671	2.7633
29	1.6991	2.0452	2.4620	2.7564
30	1.6973	2.0423	2.4573	2.7500