Topics for Final Exam

1. Interpret a stemplot. Find Medians and Quartiles.

For **example,** consider the following sample of observations:

11,15,18,20,22,28,39,60,61

We can construct the following Stem and leaf plot to represent this data.

Stem	Leaf
1	1 5 8
2	0 2 8
3	9
4	
5	
6	0 1

Leaf unit=1.0
N=9

If N is even
Median is the avg of (N/2)th observation and [(N/2)+1]th observation
Eg: N=10
Med = Avg of 5^{th} & 6^{th} observation

If N is odd:
N=9
The Median is $[(N+1)/2]^{th}$ observation = 5^{th} observation = 22

n1 = No. of observations before (or after) the median = 4
Q1 = ((n1 +1)/2) = Avg of the 2^{nd} and 3^{rd} observation = (15+18)/2 = 16.5
Q3 = Avg of the 7^{th} and 8^{th} observation = (39+60)/2 = 49.5

2. Calculate means and st. deviations for data in a single column or data in table-form in two columns.

Data in a single column : STAT > EDIT > Enter data in L1 and STAT > CALC > 1 Var-Stats

Data in a table form in 2 columns : STAT > EDIT > Enter values in L1 and frequency in L2 and STAT > CALC > 1 Var-Stats > Enter L1 and L2

3. Enter data into L1 and L2 and then create L3 using a formula (i.e. L3=(L2-L1)/L1*100.

STAT > EDIT > Go to column heading L3 (not the first cell in column L3) and calculate it as ((L2-L1)/L1)*100

4. Find mean and median from a Dotplot.

Enter data values into L1 and the frequencies into L2.

Finding Mean and Median:
STAT > EDIT > Enter values in L1 and frequency in L2 and STAT > CALC > 1 Var-Stats > Enter L1 and L2

5. Use your calculator to create a Boxplot and a 5 number summary of a data set.

STAT> EDIT > Enter data in L1

Boxplot:
2^{nd}> STATPLOT > [1] > Select the box plot (the 4th graph type available in the menu) using the right arrow key and press [ENTER] > Xlist is L1 & Freq default to 1

5 number summary:
STAT > CALC > 1 Var Stats

6. Use the 1.5IQR rule to test for outliers in a data set.

IQR = Q3 – Q1
Lower Fence: Q1 – 1.5*IQR
Upper Fence: Q3 + 1.5*IQR

Any observations < Lower fence or any observations > Upper Fence are OUTLIERS.

7. Know how to interpret the slope of a linear model.

Linear Model: **y=a+bx**
Intercept : *a*
Slope: *b*
Interpretation of Slope: When 'X' increases by 1 unit, 'Y' , on average, increases(if *b* is positive) / decreases(if *b* is negative) by '*b*' units.

8. Create a linear regression model and store it in Y1. Store the predictions and store the residuals. Decide on how useful the model is by looking at the scatter of the data around the regression line and evaluating the R-Sq value and interpreting the residual plot.

Enter the data into the calculator: X in L1 and Y in L2

STAT → CALC → LinReg(a+bx) → XList: L_1 ; YList: L_2 ; Store ReqEq*: Y_1 → Calculate
*to store in Y1, do this: VARS → Y-VARS → FUNCTION → Y1

9. Calculate a predicted y-value using the regression equation.

to store predicted values: STAT? EDIT > L3= → VARS → Y-VARS → FUNCTION → Y1→"Y1(L1)"
X values are stored in L1

L3(column heading for L3) = Y1(L1)

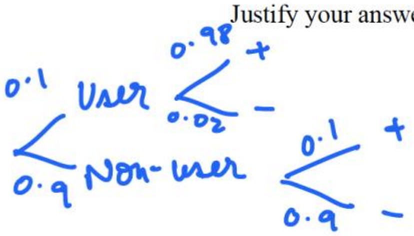
Residuals in L4 = L2 (actual Y values) – L3 (Predicted Y values)

10. Construct a tree diagram for a repetitive process and for a 2 step process and use it to calculate probabilities

Consider the following **example:**
Suppose that a drug test for an illegal drug is such that it is 98% accurate in the case of a user of that drug (e.g. it produces a positive result with probability .98 in the case that the tested individual uses the drug) and 90% accurate in the case of a non-user of the drug (e.g. it is negative with probability .9 in the case the person does not use the drug). Suppose it is known that 10% of the entire population uses this drug. You test someone and the test is positive.

I. Probability of anyone testing positive for using the drug is: **0.188** T / F

Justify your answer:



$$P(+) = P(+ \ \& \ \text{User}) + P(+ \ \& \ \text{Non-user})$$
$$= [0.1 \times 0.98] + [0.9 \times 0.1]$$
$$= 0.098 + 0.09$$
$$= 0.188$$

II. Suppose that a person tests positive for using the drug. Then, the probability that this person is actually a non-user is equal to: **0.232** T / F

Justify your answer:

$$P(\text{Non user} \mid +) = \frac{P(\text{Non-user} \ \& \ +)}{P(+)} = \frac{0.9 \times 0.1}{0.188} = \frac{0.09}{0.188} = 0.478$$

11. Find the mean for a discrete model using a table of the pdf.

Consider the following **example**:

What is the expected value of winnings in the Arizona Pick 3 lottery, where players pay $1 to select a three-digit number, and win $500 if the numbers match in the specified order? The historical probability of winning is 1 in 1000 tickets sold.

X=winnings

X	$500	$0
P(X=x)	1/1000	999/1000

Enter the winnings $500 and $0 in L1

and the probabilities 1/1000 and 999/1000 in L2

We can calculate E(X) = E(winnings) using the calculator by calculating

L3 = L1*L2

And E(X) = Sum of Values in L3 (STAT > CALC > 1 Var-Stats for L3 > Use the sum value)

E(X) = $0.5

We see that expected winnings are $0.5 while we pay $1 to buy the ticket. So, in the long run we would lose $0.5.

12. Calculate the mean and st dev for a binomial model and determine whether an outcome is unusual.

If X is a Binomial (n,p) random variable, then
E(X)=n*p
Std dev(X) = \sqrt (n*p*(1-p))

If X is a Binomial (n,p) random variable, then
Binomial PDF = P(X=k) = 2^{nd} > Vars > Binompdf (n,p,k)
Binomial CDF = P(X≤k) = 2^{nd} > Vars > Binomcdf (n,p,k)

Consider the following **Example**:
Suppose that 40% of students in a small class expect to receive a scholarship for their academic achievements.

a. Find the probability that exactly 2 students out of 5 will expect to receive a scholarship for their academic achievements.

   p=0.4
   n=5
   X: # of students out of 5 who will receive a scholarship
   P(X=2) = Binompdf (5,0.4,2) = 0.3456

b. Find the probability that the number of students in a group of 5 who expect to receive a scholarship for their academic achievements is more than 2 but less than 5.

   n=5 ; p=0.4

   P(2 < X < 5) = P(3 ≤X ≤ 4) = P(X≤ 4) – P(X ≤ 2)
   = Binomcdf (5,0.4,4) - Binomcdf (5,0.4,2)
   = 0.3072

13. Use normalcdf to calculate a probability about X

If X is a Normal random variable with mean \mu and std dev \sigma, then
P(L < X < U) = 2^{nd} > Vars > Normalcdf (L,U,\mu,\sigma)

Consider the following **Example**:
Suppose X is a normally distributed random variable with **mean 150** and **std dev 20**. Find P(X<120).

P(X<120) =P(-E99 < X < 120) = Normalcdf (-E99,120,150,20) = 0.0668

14. Use the invNorm to find cut-off points for specific areas under the normal curve.

If X is a Normal random variable with mean \mu and std dev \sigma
If you know P(X<b) = k; where you know 'k', but 'b' is unknown, then
b =2^{nd} > Vars> invnorm (k, \mu,\sigma)

Consider the following **Example**:
Suppose X is a normally distributed random variable with **mean 150** and **std dev 20**. Find the first quartile.

Q1 is that point such that P(X< Q1)=0.25
Therefore, Q1 = invnorm (0.25, 150,20) = 136.6

Invnorm (area to the left/right/center , mean, std dev , Option:left/right/center)

Q1 is a point such that P(X>Q1)=0.75
Q1 = invnorm (0.75, 150,20, Right) = 136.6

15. Calculate a probability about $\bar{x}$ using the normalcdf and the correct standard deviation value: $\sigma/\sqrt{n}$.

The sample mean, $\bar{x}$ is a Normal random variable with mean = \mu and std dev= $\sigma/\sqrt{n}$

P($\bar{x}$ <k) = Normalcdf (-E99, k ,\mu, $\sigma/\sqrt{n}$)

16. CI for μ using the Tinterval and interpret it.

STAT > Tests > Tinterval > Input: Stats (x_bar, S_x, n,c-level)

17. CI for p using 1PropZint and interpret it.

STAT > Tests > 1PropZint (x, n, c-level)

18. Know what "significance" means. How do you make a decision when using p-values and levels of significance – Reject or Fail to Reject Ho.

p-val > \alpha ; Decision is to **fail to reject the null**
p-val < \alpha ; Decision is to **reject the null** in favour of the alternative

19. HT for p and interpret it.

STAT > Test > 1PropZTest (p_0, x, n, alternative)

20. HT for μ and interpret it.

STAT > Test > TTest> Input: Stats (\mu_0, x_bar, S_x, n, alternative)

21. Paired t-test

Create a column of differences and enter into the calculator in L1
STAT > Test > TTest> Input: Data(\mu_0, List=L1, Freq=1, alternative)

22. CI for p1 – p2

STAT > Test > 2PropZInt (x1, n1, x2, n2, c-level)

23. CI for mu1 – mu2 (Independent Samples)

STAT > Test > 2SampleTInt > Input: Stats(x1_bar, S_x1, n1, x2_bar, S_x2, n2, c-level, pooled=No)

24. Z Scores

Z = (X – mean) / Std dev

Consider the sample: 1,2,3,4,5
Mean=5
Std dev = 1.6
Z score for 5 =(5-3)/1.6 = 1.25