9	5
8	9
7	17
6	8
5	7
4	6

$\bar{x} = 71$

$M = 74$

ST 305: Exam 3

By handing in this completed exam, I state that I have neither given nor received assistance from another person during the exam period. I have used no resources other than the exam itself and the basic mathematical functions of a calculator (ie, no notes, electronic communication, notes stored in calculator memory, etc.) Using your calculator for values from probability distributions like the normal or t is OK; however, if you are doing a calculation from a normal or t distribution show your work all the way to the point of calculating z/t-values. I have not copied from another person's paper. I understand that the penalty if I am found guilty of any such cheating will include failure of the course and a report to the NCSU Office of Student Conduct. **I understand that I must show all work/calculations, even if they seem trivial, to get credit for my answers.**

For all relevant questions, use a significance level of $\alpha = .05$ unless instructed otherwise.

Name: KEY

ID#: ______________________

$$\bar{x} = \frac{1}{n}\sum x_i$$

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

$$Z = \frac{X - \mu}{\sigma}$$

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n-1}$$

$$b_1 = r\frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$\text{residual} = y - \hat{y}$$

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

$$P(A^C) = 1 - P(A)$$

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

$$\mu_X = \sum x_i p_i$$

$$\mu_{a+bX} = a + b\mu_X$$

$$\mu_{X+Y} = \mu_X + \mu_Y$$

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

$$\sigma_{a+bX}^2 = b^2\sigma_X^2$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

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

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

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

$$\mu_X = np$$

$$\sigma_X = \sqrt{np(1-p)}$$

$$\hat{p} = X / n$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$P(X = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \sigma/\sqrt{n}$$

$$m = z^* \sigma/\sqrt{n}$$

$$\bar{x} \pm m$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

Chapter 7 Stuff

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}, \text{ df} = n - 1$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \text{ df} = \min(n_1, n_2) - 1$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \text{ df} = n - 1$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \text{ df} = \min(n_1, n_2) - 1$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \text{ df} = n_1 + n_2 - 2$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ df} = n_1 + n_2 - 2$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Definitions. (5 points each)

Carefully define each of the following terms.

1. P-value:

Probability, computed assuming H_0 is true, of getting data as extreme – or more extreme – than the data observed

2. Standard error:

Estimated std dev of a statistic

3. Central Limit Theorem

If $\bar{X}$ is computed for a SRS of size n from a pop'n w/ mean μ & std dev σ,

$\bar{X} \dot{\sim} N(\mu, \sigma/\sqrt{n})$ for large n

What is wrong? (3 points each)

If a statement is incorrect, edit it to make it correct by replacing the underlined word with another word or short phrase. If the statement is correct, simply state that it is correct.

4. Increasing the sample size ~~increases~~ the confidence level of a confidence interval.
   does not change

5. We reject the null hypothesis when the p-value is smaller than the selected significance level.
   correct

6. The width of a confidence interval is narrower when the population standard deviation is ~~large~~. small

7. The mean of a t distribution is ~~larger~~ when the degrees of freedom (df) is larger.
   zero (unchanged)

8. The ~~proportion~~ of successes in a SRS often has a binomial distribution.
   count

Multiple Choice. (3 points each)

9. If X has a t distribution with 10 df and Z has a standard normal distribution,

B
a. Z has a smaller mean than X
b. Z has a smaller standard deviation than X
c. $P(Z > k)$ is larger than $P(X > k)$, if $k > 0$

10. The Law of Large Numbers tells us that

A
a. the sample mean converges to the population mean as the sample size gets large
b. the sample mean converges to a normal distribution as the sample size gets large
c. both (a) and (b) are true

11. The Central Limit Theorem does not apply if

C
a. the population being studied has a very skewed distribution
b. the population standard deviation is unknown
c. the sample from a population is not a SRS

12. If the p-value for testing $H_0: \mu = 20$ is 0.03

A
a. the result is significant at the $\alpha = .05$ level
b. the result is significant at the $\alpha = .01$ level
c. we need to know if the alternative hypothesis is 1-sided or 2-sided before we can tell if the result is significant

13. If we are testing $H_0: \mu = 20$

A
a. we do not use the data in determining the form of our alternative hypothesis
b. we use a ">" alternative hypothesis if the sample mean is greater than 20
c. we use a "≠" alternative hypothesis unless the data suggests otherwise

14. Using only a single sentence or phrase for your answer, why do we almost always use hypothesis test and confidence interval procedures based on the *t* distribution instead of those based on the *z* distribution? (5 points)

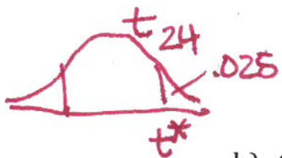
t procedures account for the typical need to estimate the unknown value of σ

15. A study was carried out to compare the effectiveness of face-to-face courses and online courses at NCSU. A SRS of 36 face-to-face courses and 25 online courses was chosen. The average course GPA (on a scale of 0 to 4.0) was computed for each class. Graphical methods suggest that GPAs are approximately normal for both types of courses. Here is a summary of the results:

	$\bar{X}$	s	
Online	2.9	0.5	$n=25$
Face-to-face	3.1	0.4	$n=36$

a) Find a 95% CI for the average GPA in online NCSU courses. (5 points)

$$\bar{x} \pm t^* \, s/\sqrt{n} \qquad 2.9 \pm 2.064 \frac{0.5}{\sqrt{25}}$$



$t^* = 2.064$

$$\boxed{\begin{array}{c} 2.9 \pm .21 \\ (2.69, 3.11) \end{array}}$$

b) Compute a 95% CI for the difference in average GPA between online and face-to-face courses at NCSU. (5 points)

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (3.1 - 2.9) \pm 2.064\,(.12)$$

0.12

$0.2 \pm .25$

$(-.05, .45)$

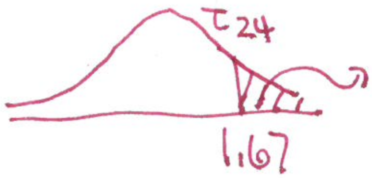
c) Is the mean GPA of online courses lower than that of face-to-face courses at NCSU? Carry out a statistical analysis to justify your answer. (5 points)

F-F O/L

$H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 > 0$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.2}{0.12} = 1.67$$



$.05 < p < .10$ ⇒ Fail to reject Ho

There is not strong evidence that the mean GPA for face-to-face courses is higher.

16. The daily profit, P, for a small business is equal to its daily income, I, minus its daily expenses, E. Based on past results, the business owner believes that the daily income varies with mean $1,000 and standard deviation $200, while daily expenses vary with mean $500 and standard deviation $100. Daily income and expenses appear to be independent of one another, but their distributions do not appear to be normally distributed.

a) Find the mean and standard deviation of daily profit? (5 points)

$\boxed{P = I - E}$

$\mu_P = \mu_{I-E}$
$= \mu_I - \mu_E$
$= 1000 - 500$
$\boxed{= 500}$

$\sigma_P^2 = \sigma_{I-E}^2 = \sigma_I^2 + \sigma_E^2$
$= 200^2 + 100^2$
$= 50{,}000$

$\Rightarrow \sigma_P = \sqrt{\sigma_P^2} = \boxed{223.6}$

b) If the business is open 300 days each year, find the mean and standard deviation of the annual profit? (5 points)

Ann. Profit $= P_1 + P_2 + \cdots + P_{300} = A$

$\mu_A = \mu_{P_1+P_2+\cdots+P_{300}} = \mu_{P_1} + \mu_{P_2} + \cdots + \mu_{300} = 500 + 500 + \cdots + 500$
$= 300(500) = \boxed{150{,}000}$

$\sigma_A^2 = \sigma_{P_1+P_2+\cdots+P_{300}}^2 = \sigma_{P_1}^2 + \sigma_{P_2}^2 + \cdots + \sigma_{P_{300}}^2 = 50{,}000(300) = 15{,}000{,}000$

$\Rightarrow \boxed{\sigma_A = 3{,}873}$

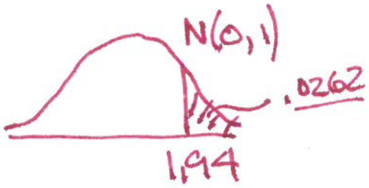
c) If the business will be open 300 days next year, what is the probability that the average daily profit for the entire year will be greater than $525 (5 points)

~~Avg daily profit = $\bar{A}$ (sample mean of the 300 A values)~~

~~By CLT, $\bar{A} \dot\sim N(150{,}00$~~

Avg daily profit = $\bar{P}$ (sample mean of the 300 P's

By CLT, $\bar{P} \dot\sim N(500, \underset{\uparrow \; 12.9}{223.6/\sqrt{300}})$



$Z = \frac{525 - 500}{12.9} = 1.94$

$P(Z > 1.94) = \boxed{.0262}$

17. Suppose that the business in problem 16 makes a (positive) profit on 75% of all days (it doesn't, but let's use this number for simplicity). An auditor takes the records from a SRS of 50 days and counts the number showing a profit. Assume that the daily profits are independent of each other.

a) Find the mean and standard deviation of the number of those 50 days that had a profit. (5 points)

Let $X =$ # of days w/ profit

$X \sim Bin(50, .75)$

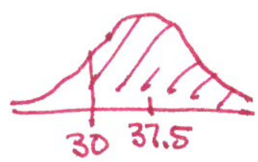
$\mu_X = np = 50(.75) = \boxed{37.5}$

$\sigma_X = \sqrt{np(1-p)} = \sqrt{50(.75)(.25)} = \boxed{3.06}$

b) Find the probability that there was a profit on at least 60% of those 50 days. (5 points)

$P(X \geq 30)$

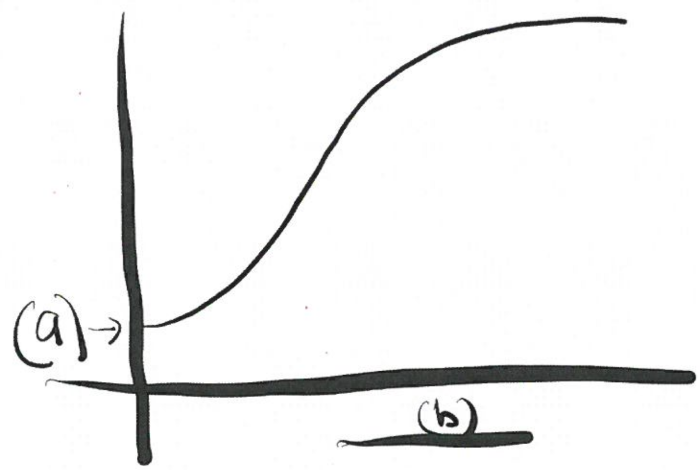
$X \dot{\sim} N(37.5, 3.06)$



$Z = \dfrac{30 - 37.5}{3.06} = -2.45$

$P(Z > -2.45) = \boxed{.9929}$

18. Here is the sketch of the power curve for a test of $H_O: \mu = 20$ using $\alpha = .05$.



a) What is the missing numerical value for (a)? (3 points)

$\alpha = 0.5$

b) What is the missing label (b) for the X-axis? (2 points)

μ

c) What was the alternative hypothesis for this test? How can you tell? (5 points)

$H_A: \mu > 20$

Power increases for large values of μ