Name: ______________________________

Practice Second Midterm Exam

Statistics 1000
Dr. Nancy Pfenning

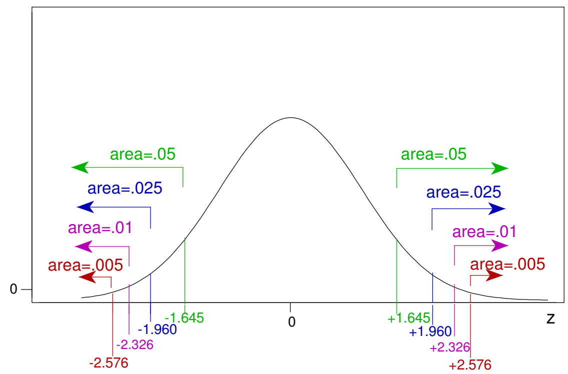
This is a closed book exam worth 150 points. You are allowed to bring a calculator and two two-sided sheets of notes. There are 12 problems, with point values as shown. If you want to receive partial credit for wrong answers, show your work. Don't spend too much time on any one problem.

1. (5 pts.) Circle any of the following that can be expected to have a fairly normal distribution, according to our usual Rule of Thumb: the distribution of sample proportion for random samples of
   (a) size $n = 10$ from a population of students with proportion $p = .5$ living on campus
   (b) size $n = 40$ from a population of students with proportion $p = .5$ living on campus
   (c) size $n = 10$ from a population with proportion of ambidextrous people $p = .03$
   (d) size $n = 40$ from a population with proportion of ambidextrous people $p = .03$

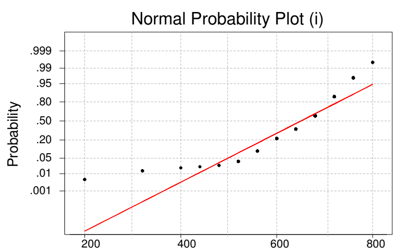
2. (5 pts.) Circle any of the following that can be expected to have a fairly normal distribution, according to our usual guidelines: the distribution of sample mean for random samples of
   (a) size $n = 10$ from a fairly normal population with mean 610, standard deviation 72
   (b) size $n = 40$ from a fairly normal population with mean 610, standard deviation 72
   (c) size $n = 10$ from a moderately skewed population with mean 3.776, standard deviation 6.503
   (d) size $n = 40$ from a moderately skewed population with mean 3.776, standard deviation 6.503

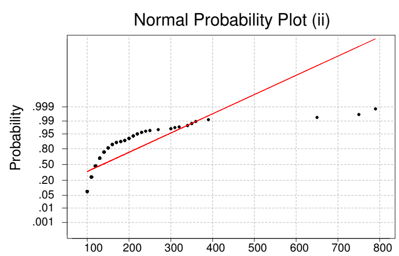
3. (10 pts.) Verbal SAT scores for a certain population are approximately normal with mean 592 and standard deviation 73.

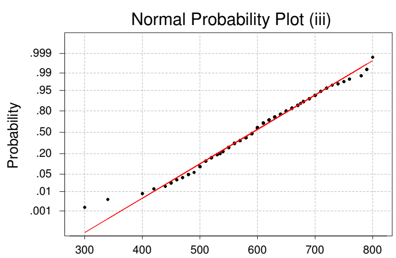
(a) Use the sketch of the tails of the z curve to estimate the probability of a score above 750: between
(i) 0 and .005 (ii) .005 and .01 (iii) .01 and .025 (iv) .025 and .05 (v) .05 and 1

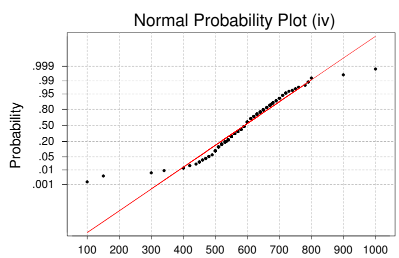


(b) Which one is the normal probability plot for the Verbal SAT scores?
(i) (ii) (iii) (iv)









4. (20 pts.) The proportion of Americans over the age of five who speak another language besides English at home is .20. Suppose we take a random sample of 64 Americans over the age of five.

   (a) We check if the sample is large enough to expect at least 10 who do and don't speak another language in order to justify

      i. claiming that the distribution of sample proportions is centered at .20
      ii. using our formula for standard deviation of sample proportion
      iii. finding probabilities based on the normal distribution

   (b) We check if the population of interest is more than 640 in order to justify

      i. claiming that the distribution of sample proportions is centered at .20
      ii. using our formula for standard deviation of sample proportion
      iii. finding probabilities based on the normal distribution

   (c) Find the mean ______and standard deviation ______of the sample **proportion** $\hat{p}$ who speak another language at home.

   (d) If 23% in a sample of 64 Americans over the age of five speak another language at home, what conclusion can we draw?

      i. We are willing to believe that the sample was coming from a population where 20% speak another language at home.
      ii. We are not willing to believe that the sample was coming from a population where 20% speak another language at home.
      iii. Results are borderline.

5. (15 pts.) Birth order X for live births in the United States in 2002 had this probability distribution, with mean 2, standard deviation 1.

Order	1	2	3	4	5	6
Probability	.40	.33	.17	.07	.02	.01

(a) What are the mean ______and standard deviation ______of sample mean birth order for a random sample of 25 births?

(b) Shape of the distribution of sample mean birth order $\bar{X}$ for samples of 25 births is

i. exactly normal
ii. approximately normal
iii. skewed, but not as skewed as the distribution of X
iv. the same as the shape of the distribution of X

(c) The probability of a sample mean less than 1.65 (for a sample of size 25) is (i) less than .025 (ii) more than .025 (iii) can't tell with given information

6. (10 pts.) A Type I Error is rejecting the null hypothesis, even though it is true; a Type II Error is failing to reject the null hypothesis, even though it is false.

(a) Which type of error is more likely to be made if only a very small sample is considered?

i. a Type I Error
ii. a Type II Error
iii. sample size has no effect on conclusions

(b) Merck developed an AIDS vaccine that showed promise in initial tests. Unfortunately, in a subsequent test conducted internationally on a large group of volunteers, those vaccinated were no less likely to become infected than those who were not vaccinated. Apparently, Merck's conclusion in the initial tests was

i. a Type I Error
ii. a Type II Error
iii. both I and II
iv. neither I nor II

7. (10 pts.) Suppose 100 repeated random samples of 40 students are taken from a population of students where .65 have pierced ears.

(a) Suppose each sample is used to test the true null hypothesis that the population proportion is .65 against the two-sided alternative. About how many of these 100 tests should reject at the $\alpha = .05$ level? ______

(b) Suppose each sample is used to up a 95% confidence interval for the population proportion with pierced ears. About how many of the 100 confidence intervals should contain the true population proportion?______

8. (20 pts.) In a study of 3,000 children, 9% fulfilled criteria for ADHD. *Of those 9% with ADHD*, only 48% had been previously diagnosed.

   (a) What is a 95% confidence interval for the proportion of all children *with ADHD* who have been diagnosed? ____________

   (b) Assuming the sample was representative, can we assert that a minority of all children with ADHD have been diagnosed?
   (i) yes, definitely (ii) no, not at all (iii) yes, just barely (iv) no, not quite

   (c) Which of these would produce a narrower interval?

      i. use a lower level of confidence
      ii. study fewer children
      iii. both (i) and (ii)
      iv. neither (i) nor (ii)

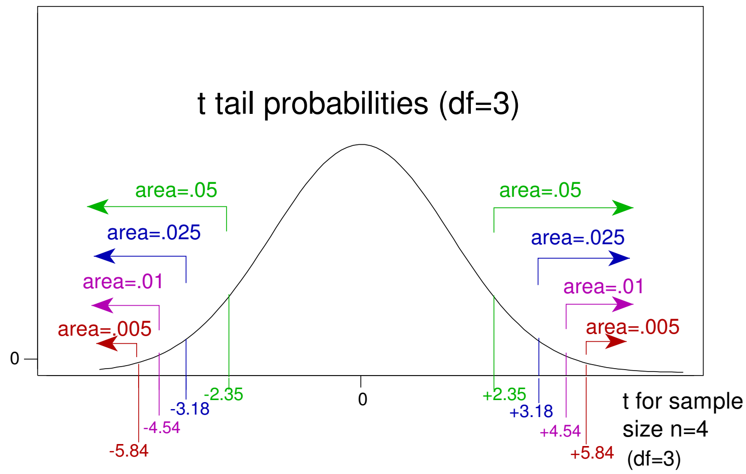
   (d) Which one of these is the correct interpretation of the interval you produced in (a)?

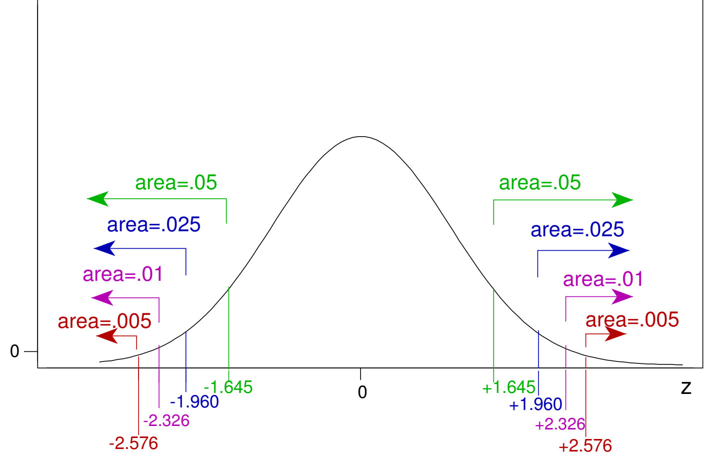
      i. We are 95% sure that population proportion falls in this interval.
      ii. We are 95% sure that sample proportion falls in this interval.
      iii. 95% of the population fall in this interval.
      iv. 95% of the sample fall in this interval.
      v. The population proportion has a 95% probability of falling in this interval.
      vi. The sample proportion has a 95% probability of falling in this interval.

9. (5 pts.) A 2004 poll of 1,573 likely voters before the presidential election found 0.49 intended to vote for George Bush. Report the margin of error to 3 decimal places, assuming a 95% confidence level. ______

10. (5 pts.) Suppose that for a random sample of Pitt students, we record the values of the variable Credits (how many credits taken that semester) and Bookmoney (how much money they spent on textbooks that semester). Denote the standard deviations as σ_C for Credits, σ_B for Bookmoney, and σ_{C+B} for the sum, Credits plus Bookmoney.

Which one of the following is correct?
(i) $\sigma_{C+B} = \sigma_C + \sigma_B$ (ii) $\sigma^2_{C+B} = \sigma^2_C + \sigma^2_B$ (iii) $\sigma^2_{C+B} > \sigma^2_C + \sigma^2_B$ (iv) $\sigma^2_{C+B} < \sigma^2_C + \sigma^2_B$




area=.05
area=.025
area=.01
area=.005
area=.05
area=.025
area=.01
area=.005
0
-1.645
-1.960
-2.326
-2.576
0
+1.645
+1.960
+2.326
+2.576
z

11. (25 pts.) For this problem, we will assume that the null hypothesis should be rejected as long as the p-value is less than .05. Suppose mean and standard deviation for amount spent on textbooks in a semester by a sample of Pitt undergraduates are used to test whether mean amount spent on textbooks in a semester for **all Pitt undergraduates** exceeds \$600. Refer to the sketch of the t distribution for 3 degrees of freedom, or the standard z distribution (depending on the circumstances) and choose the correct conclusion for each of the following circumstances.

   (a) A **representative** sample of **4** Pitt undergrads produces a standardized sample mean of $t = +2.0$; the data set has a fairly **normal** appearance.

      i. Population mean for all Pitt undergrads exceeds \$600.

      ii. Continue to believe population mean for all Pitt undergrads is \$600 (not more).

      iii. Neither z nor t inference procedures are appropriate in this situation.

   (b) A **representative** sample of **4** Pitt undergrads produces a standardized sample mean of $t = +2.0$; there is a high **outlier** in the data set.

      i. Population mean for all Pitt undergrads exceeds \$600.

      ii. Continue to believe population mean for all Pitt undergrads is \$600 (not more).

      iii. Neither z nor t inference procedures are appropriate in this situation.

   (c) A sample of **400** Pitt **Biology majors** produces a standardized sample mean of $t = +2.0$; the data set has a **normal** appearance.

      i. Population mean for all Pitt undergrads exceeds \$600.

      ii. Continue to believe population mean for all Pitt undergrads is \$600 (not more).

      iii. Neither z nor t inference procedures are appropriate in this situation.

   (d) A **representative** sample of **400** Pitt undergrads produces a standardized sample mean of $t = +2.0$; there is a high **outlier** in the data set.

      i. Population mean for all Pitt undergrads exceeds \$600.

      ii. Continue to believe population mean for all Pitt undergrads is \$600 (not more).

      iii. Neither z nor t inference procedures are appropriate in this situation.

   (e) Which one of these sample means would provide the most evidence against the null hypothesis $H_0 : \mu = 600$ **in favor of** the **two-sided** alternative $H_a : \mu \neq 600$? (i) $\bar{x} = 500$ (ii) $\bar{x} = 550$ (iii) $\bar{x} = 650$

12. (20 pts.) Average wait times for consumers visiting retail outlets, banks, and restaurants were compared for a sample of 18 east coast cities and a sample of 7 west coast cities.

```
location     N      Mean     StDev   SE Mean
east        18     272.7      72.5        17
west         7     231.6      27.8        11
Difference = mu (e) - mu (w)
Estimate for difference:  41.1
95% lower bound for difference: 6.6
T-Test of difference = 0 (vs >): T-Value = 2.05  P-Value = 0.026  DF = 22
```

(a) The design was (i) paired (ii) two-sample (iii) several sample

(b) Does the output above provide evidence that wait times were higher in the east? Use $\alpha = .05$ as your cut-off, and circle the best answer below.

i. Yes, definitely.
ii. Yes, but just barely.
iii. No, not quite.
iv. No, not at all.

(c) The p-value shown is for testing against a one-sided alternative. What would the p-value be if we had tested against a two-sided alternative? ______

(d) Considering the size of the p-value for the two-sided alternative, would a 95% confidence interval for difference between population means contain zero? (No calculations necessary.)

i. Yes, definitely.
ii. No, not even close.
iii. Borderline: either just barely or not quite.
iv. There is not enough information to decide.

13. (Extra Credit 5 pts.) Let $\hat{p}$ be the proportion of even numbers appearing when a die is rolled 2 times. Sketch the probability histogram for the distribution of $\hat{p}$.