Exam 3 is Wed., Nov. 20. No notes. **You are allowed a TI-30 calculator.** The exam covers ch. 14-15, 17-20. Memorize $\overline{x}$, sd s, zscore, and how to use tables A and C. Quizzes 6, 7, 8, 9 and 10 are relevant with much less emphasis on quizzes 6 and 7.

Ignore power p. 406–408, a df formula on p. 486, an F test, a plus four CI on p. 500–501, and the plus four CI on p. 520.

Types of problems. CHECK THAT THE FORMULAS ON THIS REVIEW SHEET ARE CORRECT. YOU ARE RESPONSIBLE FOR ANY ERRORS.

45) The following problem is **very important for exam 3 and the final.** Recognizing which of the 6 tests or CI's to use. See HW10 B, C, D.
i) one sample z, (Q7, but not very important)
ii) one sample t, HW9 H, I HW10 H, Q8, Q9
iii) matched pairs, HW10 A, G, Q9
iv) two sample t, HW10 E, , HW11 A, B, L, Q9
v) one sample z for proportion, HW11 G, I, M, Q10
vi) two sample z for 2 proportions, HW11 J, K, Q10

I usually do not put z intervals and tests on midterms or finals, so the tests and CI's from chapters 17–20 are **much more important** than the tests and confidence intervals from ch. 14–15.

46) A one sample t-interval and/or test **will be on the exam**. See Q8 1-4, Q9.

47) A forwards calculation for $\hat{p}$ **will be on the exam**.
Step 0) $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$. Step i) draw line with $\hat{p}$ value(s) and p. Step ii) get zscore: z $= \dfrac{value - p}{\sigma_{\hat{p}}}$. Step iii) draw z–curve. Step 4 use table A to get the appropriate probability. Q10, HW11, D.

48) **Know how to choose the sample size** for a desired margin of error for a CI for p: use $n = \left(\dfrac{z*}{m}\right)^2 p^*(1-p^*)$ where p^* is a good guess of p. Use $p^* = 0.5$ if no good guess is given. Round n up. Keywords: to within 0.05 or $\pm.05$ means $m = 0.05$. See HW11 H and Q10.

49) Know the interpretation of the CI as an interval of reasonable values. If 100 95% CI's are generated, about 95 will contain the parameter and about 5 will not. Understand figure 14.2 on p. 356. See Q7 5.

50) Know the interpretation of $\alpha = \delta$ for tests of hypotheses. If $\alpha = \delta = 0.05 =$ P(type I error) and if 100 tests of hypotheses are performed where Ho is true for all 100 of the tests, then about 95 will fail to reject Ho, but about 5 will wrongly reject Ho.

51) Know how to compute $\hat{p} = X/n =$ (count of successes)/(sample size). HW 11 C.

52) You may need to compute $\overline{x}$ and s in order to make a t–interval. HW9 H.

53) The degrees of freedom df tells you what table to use for tests, not what procedure should be used, eg σ unknown and s given make a t–procedure regardless of n.

54) Using output is important. Output makes CI's easy and steps ii) and iii) of hypothesis tests easy. Also check whether the test statistic is given. See HW10 F, H, HW11 L, M, Q8 2?, Q9 4?, Q10.

55) You need to know when the methods can not be used, eg when the CLT does not hold and when the sample sizes for proportions are too small. Can not use the methods if the data is from a voluntary response sample or a sample of convenience. See pages 3–4 of this review. See HW11 G.

56) For the one and two sample z-intervals for proportions and for t–intervals if df > 30, use the third line from the bottom of table C to get the cutoff z* or t* (1.645 for 90%, 1.96 for 95% or 2.576 for 99%). For t–intervals with $df \leq 30$, find the column with the level, eg 95% and the row with the df. Intersect to get t*. Eg if df = 5 then for a 95% CI, t* = 2.571.

57) Given a test statistic, **know how to find the p–value**. Know that $0 \leq$ p–value ≤ 1. **Making a sketch of the normal or t curve is a useful book keeping technique.**

A) Always use z-table A for z-tests and for t-tests (including matched pairs) if df > 30. If a t-test is used, let $to = z_o$.

i) For a right tail test (Ha >), pval $= 1 - P(z < z_o)$. If $z_o > 3.49$, then pval = 0.0, if $z_0 < -3.49$, then pval = 1.0.

ii) For a left tail test (Ha <), pval $= P(z < z_o)$. If $z_o > 3.49$, then pval = 1.0, if $z_0 < -3.49$, then pval = 0.0.

iii) For two tail (Ha $\neq$), pval $= 2(P(z < -|z_o|))$. If $z_o > 3.49$, then pval = 0.0, if $z_0 < -3.49$, then pval = 0.0.

B) Use table C to approximate p-values for t tests (including matched pairs) if df ≤ 30. Tip: if df > 5 then the pval from table A should be within 0.1 of the pval from table C.

i) For right tail, if to falls between two t* values, then the pval is between 2 "One-sided P" pvalues (eg if df = 5 and to = 3.05, then $0.01 <$ pval < 0.02). If to < 0 or if to $<$ smallest t* value (eg df = 5 and to = 0.555), then pval $> .25$ (the "One-sided P" pvalue furthest to the left). If to $>$ largest t* value (eg df = 5 and to = 17.75), then pval = 0.0 (less than 0.0005, the "One-sided P" pvalue furthest to the right).

ii) For left tail, if to > 0, then pval > 0.25. If to < 0 then compute $|to|$ and use (symmetry and) the rules for the right tail test: that is, if to < 0 and $|to|$ is between two t* values, then the p-value is between two "One-sided P" pvalues (eg if df = 5 and to = -1.57, then $0.05 <$ pval < 0.10). If to < 0 and $|to|$ is bigger than the largest t* value (eg df = 5 and to = -44.67), then pval = 0.0. If to < 0 and $|to|$ is less than the smallest t* value (eg df = 5 and to = -0.17), then pval > 0.25.

iii) For two tail, use the last line of Table C: Two-sided P. If $|to|$ is between two t* values, then the pval is between two "Two-sided P" pvalues. If $|to|$ is bigger than the largest t* value (eg df = 5 and $|to| = 33.79$), then pval = 0.0 (less than 0.001, the "Two-sided P" pvalue furthest to the right). If $|to|$ is smaller than the smallest t* value (eg df = 5 and $|to| = 0.37$), then pval > 0.5 (the "Two-sided P" value furthest to the left).

Confidence intervals: A confidence interval is an interval of reasonable values for the parameter and has the form estimator $\pm$ cutoff SE(estimator) (= estimator $\pm$ margin of error) where SE(estimator) is the standard error of the estimator. The cutoff t^* is obtained from table C. Use third line from the bottom of table C if the cutoff is z^* (or if df > 30). If the cutoff is z^*, then 1.645 is used for the 90% CI, 1.96 for the 95% CI and 2.576 for the 99% CI.

tests of hypotheses: All tests of hypotheses have the same 4 steps:
i) State Ho and Ha.
ii) Obtain the test statistic (possibly from output).
iii) Find the p–value (possibly from output).
iv) If the p–value $\leq \delta$, reject Ho, otherwise fail to reject Ho. Write a nontechnical sentence explaining the decision. (Use $\delta = \alpha = 0.05$ if $\delta = \alpha$ is not given.)

P–values are obtained from table A if they are z-tests or if they are t-tests with df > 30. Use table C to get p–values if the test is a t-test with df < 30. Often p-values and the test statistic are given in computer output.

Use the notes given in class for more details of the following six procedures.

i) z–test and interval for μ: HARDLY EVER ON EXAMS. σ is KNOWN.
The test statistic for Ho: $\mu = \mu_o$ is $z_o = \dfrac{\overline{x} - \mu_o}{\sigma/\sqrt{n}}$, and the CI for μ: $\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

ii) **t test and interval** for μ: ALWAYS ON EXAMS. Sample SD s is given and pop. SD σ is unknown.
The test statistic for Ho: $\mu = \mu_o$ is $t_o = \dfrac{\overline{x} - \mu_o}{s/\sqrt{n}}$

Get p–value from table C if $df = n - 1 \leq 30$ otherwise use table A.

The CI for μ is $\overline{x} \pm t^* \dfrac{s}{\sqrt{n}}$

Get t^* from table C with $df = n - 1$ if $df \leq 30$. Otherwise use $t^* = z^*$ line near the bottom of table C. The central limit theorem (CLT) should hold for $\overline{x}$. The data should be from a SRS or measurements from an experiment.

iii) **matched pairs t test and interval** for μ: The n pairs (x_i, y_i) are independent, but x_i and y_i are dependent, eg x and y are **two measurements on the same person or thing (taken at the same time or "before and after"), or on twins, or on litter mates**. Often 2 "treatments" are randomly assigned to the same individual or thing. Suppose x has mean μ_1 and y has mean μ_2. The matched pairs procedures are simply the one sample t procedures applied to the differences $d_i = x_i - y_i$. Let $\mu_d = \mu_1 - \mu_2$. Let $\overline{x}_d$ be the sample mean and let s_d be the sample standard deviation of the differences d_i. The subscript d is often not used.
The test statistic for Ho: $\mu_d = 0$ is $t_o = \dfrac{\overline{x}_d - 0}{s_d/\sqrt{n}}$.

Get p–value from table C if $df = n - 1 \leq 30$ otherwise use table A.

The CI for $\mu_d = \mu_1 - \mu_2$ is $\overline{x}_d \pm t^* \dfrac{s_d}{\sqrt{n}}$.

Get t^* from table C with $df = n - 1$ if $df \leq 30$. Otherwise use $t^* = z^*$ line near the bottom of table C. The central limit theorem (CLT) should hold for $\overline{x}_d$. The data should be from a SRS of pairs or pairs of measurements from an experiment.

iv) **2 sample t test and interval** for $\mu_1 - \mu_2$:

The test statistic for Ho: $\mu_1 = \mu_2$ is $$t_o = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

The CI for $\mu_1 - \mu_2$ is $(\overline{x}_1 - \overline{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Use df from output or use df = smaller of $n_1 - 1$ and $n_2 - 1$. Get p–values and cutoff t^* from table C if df ≤ 30 otherwise get the p-value from table A and cutoff $t^* = z^*$ near the bottom of table C. The data should be from two independent SRS's or from measurements from an experiment on two groups (where the individuals were randomly assigned to each group). If $n_1 = n_2 \equiv n$ then the procedure can be used for $n \geq 5$ if both populations have the **same shape**. Otherwise, the CLT should hold for both $\overline{x}_1$ and $\overline{x}_2$.

v) **z test and interval for** p:

The test statistic for Ho: $p = p_o$ is $z_o = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$.

The CI for p is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Sample size $n \approx \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*)$, round up. The value p^* is a good guess for p. If no good guess is available, use $p^* = 0.5$.

Get the p-value from table A and cutoff z^* near the bottom of table C.

The data should be a proportion obtained from a SRS or an experiment. The population size should be at least ten times the sample size. For a test, need both $np_o \geq 10$ and $n(1 - p_o) \geq 10$. For a CI, need both the number of successes $X = n\hat{p} \geq 15$ and the number of failures $n - X = n(1 - \hat{p}) \geq 15$.

vi) **z test and interval for** $p_1 - p_2$**:**

The test statistic for Ho:$p_1 = p_2$ is: $z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ where $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$. Here X_i is the count of successes in sample i, $i = 1, 2$.

The CI for $p_1 - p_2$ is: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.

Get the p-value from table A and cutoff z^* near the bottom of table C. The data should be proportions obtained from two ind. SRS's (or a randomized controlled experiment). The population size should be at least ten times the sample size. For a test, need both the number of successes $X_i = n_i\hat{p}_i \geq 5$ and the number of failures $n_i - X_i = n_i(1 - \hat{p}_i) \geq 5$. For a CI, need both $X_i = n_i\hat{p}_i \geq 10$ and $n_i - X_i = n_i(1 - \hat{p}_i) \geq 10$ for $i = 1, 2$.