Review for Exam 4 and final. Math 483 Exam 4 is Friday, April. 25. Final: Thursday, May 8, 10:15-12:15 in Neckers 156. (10 sheets of notes for Exam 4, 20 for the final). Sections 1.3, 3.2, 3.3, 3.9, 4.2, 4.3, 4.9, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, part of 6.3, 6.4, part of 7.2, 7.3, 7.5, 8.1, 8.2, 8.3, 8.6, 8.7, 8.8, 9.3 (p. 450), 9.6, 9.7, 10.1, 10.3, 10.6, and 10.8 will be emphasized on the final. Exam 4 will only cover t-test from ch. 10. **From exams 1, 2, and 3:** See the last page of Exam 2 review and the 3 pages of the Exam 3 review (especially points 1) - 11)).

Know Exam 3 material. Material Since Exam 3: The material on quizzes 8, 9 and 10 will be especially important for exam 4. **Point 10) from exam 3 review** is important for exam 4. Material on quiz 11 will be important for the final.

12) **Common Final Problem:** Normal approximation to the binomial. Let Y count the number of successes in n trials where the probability of a success in p then Y is binomial$(n = 50, p = 0.3)$. Let X be a normal RV with mean $\mu = np$ and SD $\sigma = \sqrt{np(1-p)}$. Then $P(Y \geq 18) = P(X \geq 17.5)$ and $P(Y \leq 18) = P(X \leq 18.5)$. Ideally, this approximation should not be used unless $n > 9p/(1-p)$ and $n > 9(1-p)/p$. See Q8 3?, Q9 ?, HW16 A), B), C).

13) **Common Final Problem:** Find the bias and MSE (as a function of n and c) of an estimator $T = c\sum_{i=1}^n Y_i$ or $(T = b\overline{Y})$ of θ when $Y_1, ..., Y_n$ are iid with $E(Y_1) = \mu = h(\theta)$ and $V(Y_i) = \sigma^2$. Solution: $E(T) = c\sum_{i=1}^n E(Y_i) = nc\mu$, $V(T) = c^2\sum_{i=1}^n V(Y_i) = nc^2\sigma^2$, $B(T) = E(T) - \theta$ and $MSE(T) = V(T) + [B(T)]^2$. (For $T = b\overline{Y}$, use $c = b/n$.) See Q8 4ab?, Q9 ?, HW17 B).

14) **Common Final Problem:** Method of Moments Estimator. Let $Y_1, ..., Y_n$ be iid from a distribution with a given pdf $f(y|\theta)$ or probability function $p(y|\theta)$. Find $E(Y) = \int_{-\infty}^{\infty} y f(y) dy = h(\theta)$, then $\hat{\theta} = h^{-1}(\overline{Y})$. That is, solve $h(\theta) = \overline{Y}$ for θ. See Q10, HW 20, A), B), C), D), E).

15) **Common Final Problem:** Maximum Likelihood Estimator (MLE). Let $g(y|\theta)$ be the probability function or pdf of a random sample $Y_1, ..., Y_n$. Then **the likelihood function** $L(\theta) = \prod_{i=1}^n g(y_i|\theta)$. It is crucial to observe that the likelihood function is a function of θ (and that $y_1, ..., y_n$ act as fixed constants). Let $\hat{\theta}$ be the parameter value at which $L(\theta)$ attains its maximum as a function of θ with $Y_1, ..., Y_n$ held fixed. Then the maximum likelihood estimator (**MLE**) of the parameter θ is $\hat{\theta}$. On exams, to find the MLE, i) find $L(\theta)$ and then find the log likelihood $\log L(\theta)$.
Let $h(y_i|\theta) = f(y_i|\theta)$ or $h(y_i|\theta) = p(y_i|\theta)$ depending on whether Y has a pdf or probability function. Suppose that the constants a_i do not depend on θ. If $h(y_i|\theta) = a_i b_i(\theta) c_i(\theta)$, then

$$L(\theta) = \prod_{i=1}^n a_i b_i(\theta) c_i(\theta) = \prod_{i=1}^n a_i \prod_{i=1}^n b_i(\theta) \prod_{i=1}^n c_i(\theta),$$

and

$$\log L(\theta) = \log(\prod_{i=1}^n a_i) + \sum_{i=1}^n \log(b_i(\theta)) + \sum_{i=1}^n \log(c_i(\theta)).$$

The derivative of the first term with respect to θ is 0.
ii) Find the derivative $\frac{d}{d\theta} \log L(\theta)$, set the derivative equal to zero and solve for θ. The solution is a candidate for the MLE.

iii) **Invariance Principle:** If $\hat{\theta}$ is the MLE of θ, then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.
iv) Show that $\hat{\theta}$ is the MLE by showing that $\hat{\theta}$ is the global maximizer of $\log L(\theta)$. Usually this is done by noting that $\hat{\theta}$ is the unique solution to the equation $\frac{d}{d\theta} \log L(\theta) = 0$ and that the 2nd derivative evaluated at $\hat{\theta}$ is negative: $\frac{d^2}{d\theta^2} \log L(\theta)|_{\hat{\theta}} < 0$.
See Q9, Q10, HW21 A), B), C), D).

Tips: a) $\exp(a) = e^a$ and $\log(y) = \ln(y) = \log_e(y)$ is the **natural logarithm**.
b) $\log(a^b) = b \log(a)$ and $\log(e^b) = b$.
c) $\log(\prod_{i=1}^n a_i) = \sum_{i=1}^n \log(a_i)$.
d) $\log L(\theta) = \log(\prod_{i=1}^n g(y_i|\theta)) = \sum_{i=1}^n \log(g(y_i|\theta))$.
e) If t is a differentiable function and $t(\theta) \neq 0$, then $\frac{d}{d\theta} \log(|t(\theta)|) = \frac{t'(\theta)}{t(\theta)}$ where $t'(\theta) = \frac{d}{d\theta} t(\theta)$. In particular, $\frac{d}{d\theta} \log(\theta) = 1/\theta$.
f) Anything that does not depend on θ is treated as a constant with respect to θ and hence has derivative 0 with respect to θ.

Suppose $\frac{d}{d\theta} \log L(\theta_o) = 0$. The 2nd derivative test states that if $\frac{d^2}{d\theta^2} \log L(\theta_o) < 0$, then θ_o is a local max.

If $\log L(\theta)$ is strictly concave ($\frac{d^2}{d\theta^2} \log L(\theta) < 0$ for all θ), then any local max of $log L(\theta)$ is a global max.

You should know how to find the MLE for the normal distribution (including when μ or σ^2 is known, memorize the MLE's $\overline{Y}$, $\sum_{i=1}^n (Y_i - \overline{Y})^2/n$, $\sum_{i=1}^n (Y_i - \mu)^2/n$) and for the uniform distribution. Also $\overline{Y}$ is the MLE for several brand name distributions.

Confidence intervals: Confidence intervals are intervals of plausible values for the parameter and have the form estimator $\pm$ cutoff $\sqrt{V(estimator)}$ where V(estimator) is the estimated variance of the estimator. The cutoff $t_{\alpha/2}$ is obtained from the t-table (use the bottom of the t-table if the cutoff is $z_{\alpha/2}$).

TESTS OF HYPOTHESES WILL BE ON QUIZ 10, 11 AND THE FINAL BUT EXAM 4 ONLY COVERS THE t-test. **tests of hypotheses:** All tests of hypotheses have the same 4 steps.
i) State Ho and Ha.
ii) Calculate the test statistic.
iii) Find the p–value.
iv) If the p–value $\leq \alpha$, reject Ho, otherwise fail to reject Ho. Write a nontechnical sentence explaining the decision.

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

i) Z test and interval for μ is HARDLY EVER ON EXAMS since σ known:
test statistic: $z_o = \dfrac{\overline{Y} - \mu_o}{\sigma/\sqrt{n}}$ CI: $\overline{Y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

ii) **one sample t test and interval** for μ is ALWAYS ON EXAM 4 and FINAL. Sample SD s (or sample variance s^2) is given and pop. SD σ is unknown:
test statistic for Ho: $\mu = \mu_o$: $t_o = \dfrac{\overline{Y} - \mu_o}{S/\sqrt{n}}$

Get p–value from t–table if $df = n - 1 \leq 29$, otherwise use the z-table.

CI for μ: $\overline{Y} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$ Get $t_{\alpha/2}$ with $df = n - 1$ if df ≤ 29, otherwise use the bottom row of the t–table (1.645, 1.96 or 2.576). Use if there is one sample of iid data, eg measurements from an experiment. The central limit theorem (CLT) should hold for $\overline{Y}$.

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

test statistic for $\mu_1 = \mu_2$: $t_o = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

Get p-values from z table if $n_1 \geq 30$ and $n_2 \geq 30$.

CI for $\mu_1 - \mu_2$: $(\overline{Y}_1 - \overline{Y}_2) \pm z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$.

Use bottom of t–table to get $z_{\alpha/2}$ (1.645, 1.96 or 2.576) if $n_1 \geq 30$ and $n_2 \geq 30$.

The 2 sample procedures are (usually) used if no information about σ_1 and σ_2 is given. There should be 2 samples of iid data, the CLT should hold for both $\overline{Y}_1$ and $\overline{Y}_2$. Since $z_{\alpha/2}$ is used, want $n_i \geq 30$ for $i = 1, 2$ to make df > 29

iv) pooled 2 sample t test and interval for $\mu_1 - \mu_2$: test statistic for Ho $\mu_1 = \mu_2$:
$t_o = \frac{\overline{Y}_1 - \overline{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ where $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

Get p–value from t–table if $df = n_1 + n_2 - 2 \leq 29$, otherwise use the z-table.

CI: $(\overline{Y}_1 - \overline{Y}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Get $t_{\alpha/2}$ with $df = n_1 + n_2 - 2$ if df ≤ 29, otherwise use the bottom row of the t–table (1.645, 1.96 or 2.576). **Use if told** that $\sigma_1 = \sigma_2$. There should be 2 samples of iid data, the CLT should hold for both $\overline{Y}_1$ and $\overline{Y}_2$.

v) **One sample Z test and interval for** p: let $\hat{p}$ = number of "successes"$/n$.

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

Get p-value from z-table.

CI for p: $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Get $z_{\alpha/2}$ from bottom line of t–table (1.645, 1.96 or 2.576).

16) Sample size $n \approx \left(\frac{z_{\alpha/2}}{B}\right)^2 p^*(1 - p^*)$, round up. The value p^* is a good guess for p. If no good guess is available, use $p^* = 0.5$. See Q9.

vi) **Two sample Z test and interval for** $p_1 - p_2$**:**

test statistic for Ho $p_1 = p_2$: $z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ where $\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$.

Get p-value from z-table.

CI for $p_1 - p_2$: $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Get $z_{\alpha/2}$ from bottom line of t–table (1.645, 1.96 or 2.576).

Always use z-table to get the p-value for a z-test (and the bottom of the t-table for z-intervals).

17) COMMON FINAL PROBLEM: Make a t–interval. See Q8 5?.

18) Some of the intervals will be on exam 4 and the FINAL. See Q9.

19) The t-test for μ, the z test for p or $p_1 - p_2$ and/or the two sample t-test will be on the FINAL.

See Q8 5, Q9, Q10, Q11, HW 17 D), E), HW 18 A), C), D), E), F), HW19 A) and HW22 B), C), D), E), F), G).

20) 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 for z-tests and for t-tests if df > 29. If a t-test is used, let $to = z_o$. i) For a right tail test (Ha $>$), pval $= P(z > z_o)$. If $z_o > 2.99$, then pval = 0.0, if $z_0 < -2.99$, then pval = 1.0.
ii) For a left tail test (Ha $<$), pval $= P(z < z_o) = 1 - P(z > z_o)$. If $z_o > 2.99$, then pval = 1.0, if $z_0 < -2.99$, then pval = 0.0.
iii) For two tail (Ha $\neq$), pval $= 2(P(z > |z_o|))$. If $z_o > 2.99$, then pval = 0.0, if $z_0 < -2.99$, then pval = 0.0.

B) Use t-table to approximate p-values for t tests if df ≤ 29. Tip: if df > 5 then the pval from z-table should be within 0.1 of the pval from t-table.

i) For right tail, if to falls between two t* values, then the pval is between 2 upper tail values (eg if df = 5 and to = 3.05, then 0.01 $<$ pval $<$ 0.025). If to < 0 or if to $<$ smallest t* value (eg df = 5 and to = 0.555), then pval $> .10$. If to $>$ largest t* value (eg df = 5 and to = 17.75), then pval = 0.0.

ii) For left tail, if to > 0, then pval > 0.10. 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 upper tail values (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 > 0.10.

iii) For two tail, if $|to|$ is between two t* values, then the pvalue is between two upper tail values each multiplied by 2 (eg if df = 5 and to = 2.68 or to = -2.68 then 2(0.01)=0.02 $<$ pval $<$ 2(.025)=0.05. If $|to|$ is bigger than the largest t* value (eg df = 5 and $|to| = 33.79$), then pval = 0.0. If $|to|$ is smaller than the smallest t* value (eg df = 5 and $|to| = 0.37$), then pval > 0.10.

C) The left tail t-table is useful. If to falls between two t* values, then the pval is between 2 pvalues: left tail for left tail, right tail for right tail and two tail for two tail. If to $>$ largest value, than pval = 0 for two tail and right tail, but pval = 1 for left tail. If to $<$ smallest value, then pval = 0 for two tail and left tail, but pval =1 for right tail.