1 On these review notes and the exam

1. You are responsible for the correctness of all of the formulae on this review sheet. (There are undoubtedly ytopgraphical errors :-).

2. You should know, *and understand*, everything in these review notes.

3. You should be able to perform hypothesis tests and calculate confidence intervals using the Z, Student's T, Chi-squared and F tables. You should be able to calculate a P-value using the normal table.

4. In the exam, there will be questions on the implementation and development of the methods. For example, you should know how to construct a T confidence interval, when you construct a T confidence interval, what are the assumptions of the T confidence interval and why (given the assumptions) a T confidence interval yields 95% (for example) coverage.

5. The exam format will be a series of multiple choice and short answer questions. Tedious calculations will be avoided.

6. You can bring a *non-fancy* (you know what I mean) scientific calculator. It must be able to take logs and raise numbers to exponents.

7. You can bring in one sheet of 8.5×11 paper filled, front and back, with formulae and notes.

2 Set theory

1. Notation - $\subset$ means "is a subset of", $\in$ means "is an element of".
2. The **sample space**, Ω, is the space of all possible outcomes of an experiment.
3. An **event**, say $A \subset \Omega$, is subset of Ω.
4. The **union** of two events, $A \cup B$, is the collection of elements that are in A, B or both.
5. The **intersection** of two events, $A \cap B$, is the collection of elements that are in both A and B.
6. The **compliment** of an event, say $\bar{A}$ or A^c, is all of the elements of Ω that are not in A.
7. The **null** or **empty** set is denoted $\emptyset$.
8. Two sets are **disjoint** or **mutually exclusive** if their intersection is empty, $A \cap B = \emptyset$.
9. **DeMorgan's laws** state that $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

3 Probability basics

1. A **probability measure**, say P, is a function on the collection of events to $[0, 1]$ so that:
   a. $P(\Omega) = 1$.
   b. If $A \subset \Omega$ then $P(A) \geq 0$.
   c. If $A_1, \ldots, A_n$ are disjoint then (**finite additivity**) $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
2. $P(\bar{A}) = 1 - P(A)$.
3. The **odds** of an event, A, are $P(A)/(1 - P(A)) = P(A)/P(\bar{A})$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. If $A \subset B$ then $P(A) \leq P(B)$.
6. Two events A and B are **independent** if $P(A \cap B) = P(A)P(B)$. A collection of events, A_i, are **mutually independent** if $P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$.
7. Pairwise independence of a collection of events does not imply mutually independence, though the reverse is true.
8. Given that $P(B) > 0$, the conditional probability of A given that B has occurred is $P(A|B) = P(A \cap B)/P(B)$.
9. Two events A and B are **independent** if $P(A|B) = P(A)$.
10. The **law of total probability** states that if A_i are a collection of *mutually exclusive events* so that $\Omega = \cup_{i=1}^n A_i$, then $P(C) = \sum_{i=1}^n P(C|A_i)P(A_i)$ for any event C.

11. **Baye's rule** states that if A_i are a collection of *mutually exclusive events* so that $\Omega = \cup_{i=1}^n A_i$, then
$$P(A_j|C) = \frac{P(C|A_j)P(A_j)}{\sum_{i=1}^n P(C|A_i)P(A_i)}.$$
for any set C (with positive probability). Notice A and $\bar{A}$ are disjoint and $A \cup A^c = \Omega$ so that we have
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

12. The **sensitivity** of a diagnostic test is defined to be $P(+|D)$ where $+$ $(-)$ is the event of a positive (negative) test result and D is the event that a subject has the disease in question. The **specificity** of a diagnostic test is $P(-|\bar{D})$.

13. Baye's rule yields that
$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)},$$
and
$$P(D^c|-) = \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)}.$$

14. The **likelihood ratio** of a positive test result is $P(+|D)/P(+|\bar{D}) =$ sensitivity$/(1-$specificity$)$. The likelihood ratio of a negative test result is $P(-|\bar{D})/P(-|D) =$ specificity$/(1-$sensitivity$)$.

15. The odds of disease after a positive test are related to the odds of disease before the test by the relation
$$\frac{P(D|+)}{P(D^c|+)} = \frac{P(+|D)}{P(+|D^c)}\frac{P(D)}{P(D^c)}.$$
That is, the posterior odds equal the prior odds times the likelihood ratio. Correspondingly,
$$\frac{P(D^c|-)}{P(D|-)} = \frac{P(-|D^c)}{P(-|D)}\frac{P(D^c)}{P(D)}.$$
This yields a method for evaluating the results of a diagnostic test without knowledge of the disease prevalence.

4 Random variables

1. A **random variable** is a function from Ω to the real numbers. A random variable is a random number that is the result of an experiment governed by a probability distribution.

2. A **Bernoulli** random variable is one that takes the value 1 with probability p and 0 with probability $(1-p)$. That is, $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

3. A **probability mass function** (pmf) is a function that yields the various probabilities associated with a random variable. For example, the probability mass function for a Bernoulli random variable is $f(x) = p^x(1-p)^{1-x}$ for $x = 0, 1$ as this yields p when $x = 1$ and $(1-p)$ when $x = 0$.

4. The **expected value** or (population) **mean** of a discrete random variable, X, with pmf $f(x)$ is
$$\mu = E[X] = \sum_x x f(x).$$
The mean of a Bernoulli variable is then $1f(1) + 0f(0) = p$.

5. The **variance** of any random variable, X, (discrete or continuous) is
$$\sigma^2 = E\left[(X-\mu)^2\right] = E[X^2] - E[X]^2.$$
The latter formula being the most convenient for computation. The variance of a Bernoulli random variable is $p(1-p)$.

6. The (population) **standard deviation**, σ, is the square root of the variance.

7. **Chebyshev's inequality** states that for any random variable $P(|X-\mu| \geq K\sigma) \leq 1/K^2$. This yields a way to interpret standard deviations.

8. A **Binomial** random variable, X, is obtained as the sum of n Bernoulli random variables and has pmf
$$P(X=k) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k (1-p)^{n-k}.$$
Binomial random variables have expected value np and variance $np(1-p)$.

5 Continuous random variables

1. **Continuous** random variables take values on a continuum.

2. The probability that a continuous random variable takes on any specific value is 0.

3. Probabilities associated with continuous random variables are governed by **probability density functions** (pdfs). Areas under probability density functions correspond to probabilities. For example, if f is a pdf corresponding to random variable X, then
$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$
To be a pdf, a function must be positive and integrate to 1. That is, $\int_{-\infty}^{\infty} f(x)dx = 1$

4. If h is a positive function such that $\int_{-\infty}^{\infty} h(x)dx \leq \infty$ then $f(x) = h(x)/\int_{-\infty}^{\infty} h(x)dx$ is a valid density. Therefore, if we only know a density up to a constant of proportionality, then we can figure out the exact density.

5. The expected value, or mean, of a continuous random variable, X, with pdf f, is
$$\mu = E[X] = \int_{-\infty}^{\infty} t f(t)dt.$$

6. The variance is $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$.

7. The **distribution function**, say F, corresponding to a random variable X with pdf, f, is

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt.$$

(Note the common convention that X is used when describing an unobserved random variable while x is for specific values.)

8. The p^{th} **quantile** (for $0 \leq p \leq 1$), say X_p, of a distribution function, say F, is the point so that $F(X_p) = p$. For example, the $.025^{th}$ quantile of the standard normal distribution is -1.96.

6 Properties of expected values and variances

The following properties hold for all expected values (discrete or continuous)

1. Expected values commute across sums: $E[X+Y] = E[X] + E[Y]$.

2. Multiplicative and additive constants can be pulled out of expected values $E[cX] = cE[X]$ and $E[c+X] = c + E[X]$.

3. For independent random variables, X and Y, $E[XY] = E[X]E[Y]$.

4. In general, $E[h(X)] \neq h(E[X])$.

5. Variances commute across sums *for independent variables* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

6. Multiplicative constants are squared when pulled out of variances $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$.

7. Additive constants do not change variances: $\mathrm{Var}(c+X) = \mathrm{Var}(X)$.

7 The normal distribution

a. The **Bell curve** or **normal** or **Gaussian** density is the most common density. It is specified by its mean, μ, and variance, σ^2. The density is given by $f(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu)^2/2\sigma^2\}$. We write $X \sim \mathrm{N}(\mu, \sigma^2)$ to denote that X is normally distributed with mean μ and variance σ^2.

b. The **standard normal** density, labeled ϕ, corresponds to a normal density with mean $\mu = 0$ and variance $\sigma^2 = 1$.

$$\phi(z) = (2\pi)^{-1/2}\exp\{-z^2/2\}.$$

The standard normal distribution function is usually labeled Φ.

c. If f is the pdf for a $\mathrm{N}(\mu, \sigma^2)$ random variable, X, then note that $f(x) = \phi\{(x-\mu)/\sigma\}/\sigma$. Correspondingly, if F is the associated distribution function for X, then $F(x) = \Phi\{(x-\mu)/\sigma\}$.

d. If X is normally distributed with mean μ and variance σ^2 then the random variable $Z = (X - \mu)/\sigma$ is standard normally distributed. Taking a random variable subtracting its mean and dividing by its standard deviation is called "standardizing" a random variable.

e. If Z is standard normal then $X = \mu + Z\sigma$ is normal with mean μ and variance σ^2.

f. 68%, 95% and 99% of the mass of any normal distribution lies within 1, 2 and 3 (respectively) standard deviations from the mean.

g. Z_α refers to the α^{th} quantile of the standard normal distribution. $Z_{.90}$, $Z_{.95}$, $Z_{.975}$ and $Z_{.99}$ are 1.28, 1.645, 1.96 and 2.32.

h. Sums and means of normal random variables are normal (regardless of whether or not they are independent). You can use the rules for expectations and variances to figure out μ and σ.

i. The sample standard deviation of iid normal random variables, appropriated normalized, is a Chi-squared random variable (see below).

8 Sample means and variances

Throughout this section let X_i be a collection of iid random variables with mean μ and variance σ^2.

1. We say random variables are **iid** if they are independent and identically distributed.

2. For random variables, X_i, the **sample mean** is $\bar{X} = \sum_{i=1}^n X_i/n$.

3. $E[\bar{X}] = \mu = E[X_i]$ (does not require the independence or constant variance).

4. If the X_i are iid with variance σ^2 then $\mathrm{Var}(\bar{X}) = \mathrm{Var}(X_i)/n = \sigma^2/n$.

5. The **sample variance** is defined to be
$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

6. $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$ is a shortcut formula for the numerator.

7. $\sigma/\sqrt{n}$ is called the **standard error** of $\bar{X}$. The estimated standard error of $\bar{X}$ is $S/\sqrt{n}$. Do not confuse dividing by this $\sqrt{n}$ with dividing by $n-1$ in the calculation of S^2.

8. An estimator is **unbiased** if its expected value equals the parameter it is estimating.

9. $E[S^2] = \sigma^2$, which is why we divide by $n-1$ instead of n. That is, S^2 is unbiased. However, dividing by $n-1$ rather than n does increase the variance of this estimator slightly, $\mathrm{Var}(S^2) \geq \mathrm{Var}((n-1)S^2/n)$.

10. If the X_i are normally distributed with mean μ and variance σ^2, then $\bar{X}$ is normally distributed with mean μ and variance σ^2/n.

11. The **Central Limit Theorem**. If the X_i are iid with mean μ and (finite) variance σ^2 then

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

will limit to a standard normal distribution. The result is true for small sample sizes, if the X_i iid normally distributed.

12. If we replace σ with S; that is,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

then Z still limits to a standard normal. If the X_i are iid normally distributed, then Z follows the Students T distribution for small n.

9 Confidence intervals for a mean using the CLT.

1. Using the CLT, we know that

$$P\left(-Z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq Z_{1-\alpha/2}\right) = 1 - \alpha$$

for large n. Solving the inequalities for μ, we calculated that in repeated sampling, the interval

$$\bar{X} \pm Z_{1-\alpha/2}\frac{S}{\sqrt{n}}$$

will contain μ 100$(1-\alpha)$% of the time.

2. The probability that μ is in an observed confidence interval is either 1 or 0. The correct interpretation is that in repeated sampling, the interval we obtain will contain μ $100(1-\alpha)\%$ of the time. (Assumes that the CLT has kicked in).

3. As n increases, the interval gets narrower.

4. As S increases, the interval gets wider.

5. As the **confidence level**, $(1-\alpha)$, increases, the interval gets wider.

6. Fixing the confidence level controls the **accuracy** of the interval. A 95% interval has 95% coverage regardless of the sample size. (Again, assuming that the CLT has kicked in.) Increasing n will improve the precision (width) of the interval.

7. Prior to conducting a study, you can fix the **margin of error** (half width), say δ, of the interval by setting $n = (Z_{1-\alpha/2}\sigma/\delta)^2$. Round up. Requires an estimate of σ.

10 Confidence intervals for a variance and T confidence intervals

1. If X_i are iid normal random variables with mean μ and variance σ^2 then $\frac{(n-1)S^2}{\sigma^2}$ follows what is called a Chi-squared distribution with $n-1$ degrees of freedom.

2. Using the previous item, we know that

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha,$$

   where $\chi^2_{n-1,\alpha}$ denotes the α^{th} quantile of the Chi-squared distribution. Solving these inequalities for σ^2 yields

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$$

   is a $100(1-\alpha)$% confidence interval for σ^2. Recall this assumes that the X_i are iid Gaussian random variables.

3. Chi-squared tests and intervals for variances are not robust to the normality assumption.

4. If Z is standard normal and X is and independent Chi-squared with df degrees of freedom then $\frac{Z}{\sqrt{X/df}}$ follows what is called a Student's T distribution with df degrees of freedom.

5. The Student's T density looks like a normal density with heavier tails (so it looks more squashed down).

6. By the previous item, if the X_i are iid $\mathrm{N}(\mu, \sigma^2)$ then

$$Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

   follows a Student's T distribution with $(n-1)$ degrees of freedom. Therefore if $t_{n-1,\alpha}$ is the α^{th} quantile of the Student's T distribution then

$$\bar{X} \pm t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}$$

   is a $100(1-\alpha)$% confidence interval for μ.

7. The Student's T confidence interval assumes normality of the X_i. However, the T distribution has quite heavy tails and so the interval is conservative and works well in many situations.

8. For large sample sizes, the Student's T and CLT based intervals are nearly the same because the Student's T quantiles become more and more like standard normal quantiles as n increases.

9. For small sample sizes, it is difficult to diagnose normality/lack of normality. Regardless, the robust T interval should be your default option.

11 EDA

1. The p^{th} **empirical quantile** of a data set is that point so that $100p\%$ of the data lies below it. The sample **median** is the $.50^{th}$ quantile. Empirical quantiles estimate population quantiles.

2. A **boxplot** plots a box with a centerline at the sample median and the box edges at the lower and upper quartiles. "Whiskers" extend to the largest data point that is within 1.5 of the IQR (inter quartile range). Side by side boxplots are useful to compare groups.

3. A **quantile-quantile** (qq) plot, plots empirical quantiles versus the theoretical quantiles. For normal random variables with mean μ and variance σ^2, let X_p be the p^{th} quantile. Then, $X_p = \mu + Z_p \sigma$. Therefore plotting the empirical quantiles versus the standard normal quantiles can be used to diagnose non-normality (a **normal qq** plot). Any deviation from a straight line indicates non-normality.

4. **Kernel density estimates**, **histograms** and **stem and leaf** plots show estimates of the density. Each relies on tuning parameters that you should vary. KDEs and histograms should only be used if you have enough data.

12 The bootstrap

1. The (non-parametric) **bootstrap** can be used to calculate **percentile bootstrap confidence intervals**.

2. The **bootstrap principle** is to use the empirical distribution defined by the data to obtain an estimate of the sampling distribution of a statistic. In practice the bootstrap principle is always executed by **resampling** from the observed data.

3. Assume that we have n data points. The bootstrap obtains a confidence interval by sampling m complete data sets by drawing with replacement from the original data. The statistic of interest, say the median, is applied to all m of the resampled data sets, yielding m medians. The percentile confidence interval is obtained by taking the $\alpha/2$ and $1-\alpha/2$ quantiles of the m medians.

4. Make sure you do enough resamples so that your confidence interval has stabilized.

5. Bootstrap intervals are interpreted the same as frequentist intervals.

6. To guarantee coverage, the bootstrap interval requires large sample sizes.

7. There are improvements to the percentile method that are not covered in this class.

13 The log-normal distribution

1. We use "$\log$" to represent the natural logarithm (base e).

2. A random variable X is log-normal with parameters μ and σ^2 if $Y = \log X$ is normal with mean μ and variance σ^2.

3. μ is $E[Y] = E[\log X]$. Because the mean and median are the same for the normal distribution, μ is also the median for $\log X$. Notice that $\exp\{E[\log X]\} = e^\mu \neq E[X]$. However, because μ is the median for $\log X$

$$.5 = P(\log X \leq \mu) = P(Y \leq e^\mu).$$

   Therefore e^μ is also the median on the original data scale.

4. Assuming log-normality, exponentiating a Student's T confidence interval for μ (using the logged data) yields a confidence for the median on the original data scale.

14 Hypothesis testing for a single mean

1. The null, or status quo, hypothesis is labeled H_0, the alternative H_a or H_1 or H_2 ...

2. A **type I error** occurs when we falsely reject the null hypothesis. The probability of a type I error is usually labeled α.

3. A **type II error** occurs when we falsely fail to reject the null hypothesis. A type II error is usually labeled β.

4. A **Power** is the probability that we correctly reject the null hypothesis, $1 - \beta$.

5. The Z test for $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ or $H_2 : \mu \neq \mu_0$ or $H_3 : \mu > \mu_0$ constructs a test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ and rejects the null hypothesis when

   H_1 $TS \leq -Z_{1-\alpha}$

   H_2 $|TS| \geq Z_{1-\alpha/2}$

   H_3 $TS \geq Z_{1-\alpha}$

   respectively.

6. The Z test requires the assumptions of the CLT and for n to be large enough for it to apply.

7. If n is small, then a Student's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df.

8. Tests define confidence intervals by considering the collection of values of μ_0 for which you fail to reject a two sided test. This yields exactly the T and Z confidence intervals respectively.

9. Conversely, confidence intervals define tests by the rule where one rejects H_0 if μ_0 is *not in* the confidence interval.

10. A **P-value** is the probability of getting evidence as extreme or more extreme than we actually got under the null hypothesis. For H_3 above, the P-value is calculated as $P(Z \geq TS_{obs} | \mu = \mu_0)$ where TS_{obs} is the observed value of our test statistic. To get the P-value for H_2, calculate a one sided P-value and double it.

11. The P-value is equal to the **attained significance level**. That is, the smallest α value for which we would have rejected the null hypothesis. Therefore, rejecting the null hypothesis if a P-value is less than α is the same as performing the rejection region test.

12. The power of a Z test for H_3 is given by the formula (know how this is obtained)

$$P(TS > Z_{1-\alpha} | \mu = \mu_1) = P\left(Z \geq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + Z_{1-\alpha}\right).$$

Notice that power required a value for μ_1, the value under the null hypothesis. Correspondingly for H_1 we have

$$P\left(Z \leq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}\right).$$

For H_2, the power is approximately the appropriate one sided power using $\alpha/2$.

13. Some facts about power.

a. Power goes up as α goes down.

b. Power of a one sided test is greater than the power of the associated two sided test.

c. Power goes up as μ_1 gets further away from μ_0.

d. Power goes up as n goes up.

14. The prior formula can be used to calculate the sample size. For example, using the power formula for H_1, setting $Z_{1-\beta} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}$ yields

$$n = \frac{(Z_{1-\beta} + Z_{1-\alpha})^2 \sigma^2}{(\mu_0 - \mu_1)^2},$$

which gives the sample size to have power $= 1 - \beta$. This formula applies for H_3 also. For the two sided test, H_2, replace α by $\alpha/2$.

15. Determinants of sample size.

a. n gets larger as α gets smaller.

b. n gets larger as the power you want gets larger.

c. n gets lager the closer μ_1 is to μ_0.

15 Binomial confidence intervals and tests

1. Binomial distributions are used to model proportions. If $X \sim \mathrm{Binomial}(n,p)$ then $\hat{p} = X/n$ is a sample proportion.

2. $\hat{p}$ has the following properties.

   a. It is a sample mean of Bernoulli random variables.

   b. It has expected value p.

   c. It has variance $p(1-p)/n$. Note that the largest value that $p(1-p)$ can take is $1/4$ at $p = 1/2$.

   d. $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$ follows a standard normal distribution for large n by the CLT. The convergence to normality is fastest when $p = .5$.

3. The **Wald test** for $H_0 : p = p_0$ versus one of $H_1 : p < p_0$, $H_2 : p = p_0$, and $H_3 : p > p_0$ uses the test statistic
$$TS = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}}$$
which is compared to standard normal quantiles.

4. The **Wald confidence interval** for a binomial proportion is
$$\hat{p} \pm Z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}.$$
The Wald interval is the interval obtained by inverting the Wald test (and vice versa).

5. The **Score test** for a binomial proportion is
$$ts = \frac{\hat{p} - p}{\sqrt{p_0(1-p_0)/n}}.$$
The score test has better finite sample performance than the Wald test.

6. The **Score interval** is obtained by inverting the score test (and vice versa)
$$\hat{p}\left(\frac{n}{n+Z^2_{1-\alpha/2}}\right) + \frac{1}{2}\left(\frac{Z^2_{1-\alpha/2}}{n+Z^2_{1-\alpha/2}}\right)$$
$$\pm Z_{1-\alpha/2}\sqrt{\frac{1}{n+Z^2_{1-\alpha/2}}\left[\hat{p}(1-\hat{p})\left(\frac{n}{n+Z^2_{1-\alpha/2}}\right) + \frac{1}{4}\left(\frac{Z^2_{1-\alpha/2}}{n+Z^2_{1-\alpha/2}}\right)\right]}.$$

7. An approximate score interval for $\alpha = .05$ can be obtained by taking $\tilde{p} = \frac{X+2}{n+4}$ and calculating the Wald interval using $\tilde{p}$ instead of $\hat{p}$.

8. An exact binomial test for H_3 can be performed by calculating the exact P-value

$$P(X \geq x_{obs} | p = p_0) = \sum_{k=x_{obs}}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p_0^k (1 - p_0)^{n-k}.$$

where x_{obs} is the observed success count. For H_1 the corresponding exact P-value is

$$P(X \leq x_{obs} | p = p_0) = \sum_{k=0}^{x_{obs}} \left(\begin{array}{c} n \\ k \end{array} \right) p_0^k (1 - p_0)^{n-k}.$$

These confidence intervals are **exact**, which means that the actual type one error rate is *no larger than* α. (The actual type one error rate is generally smaller than α.) Therefore these tests are **conservative**. For H_2, calculate the appropriate one sided P-value and double it.

9. Occasionally, someone will try to convince you to obtain an exact Type I error rate using supplemental randomization. Ignore them.

10. Inverting the exact test, choosing those value of p_0 for which we fail to reject H_0, yields an exact confidence interval. This interval has to be calculated numerically. The coverage of the exact binomial interval is no lower than $100(1 - \alpha)\%$.

16 The likelihood for a binomial parameter p

1. The **likelihood** for a parameter is the density *viewed as a function of the parameter*.

2. The binomial likelihood for observed data x is $p^x(1 - p)^{n-x}$. It is standard to drop constants in the parameter from the likelihood (such as the n choose x part).

3. The **principle of maximum likelihood** states that a good estimate of the parameter is the one that makes the data that was actually observed most probable. That is, the principle of maximum likelihood says that a good estimate of the parameter is the one that maximizes the likelihood.

   a. The maximum likelihood estimate for p is $\hat{p} = X/n$.

   b. The maximum likelihood estimate for μ for iid $\mathrm{N}(\mu, \sigma^2)$ data is $\bar{X}$. The maximum likelihood estimate for σ^2 is $(n - 1)S^2/n$ (the biased sample variance).

4. The **law of the likelihood** states that **likelihood ratios** represent the relative evidence comparing one hypothesized value of the parameter to another.

5. Likelihoods are usually plotted so that the maximum value (the value at the ML estimate) is 1. Where reference lines at 1/8 and 1/32 intersect the likelihood depict **likelihood intervals**. Points lying within the 1/8 reference line, for example, are such that no other parameter value is more than 8 times better supported given the data.

17 Group comparisons

1. For group comparisons, make sure to differentiate whether or not the observations are paired (or matched) versus independent.

2. For paired comparisons for continuous data, one strategy is to calculate the **differences** and use the methods for testing and performing hypotheses regarding a single mean. The resulting tests and confidence intervals are called **paired Student's** T tests and intervals respectively.

3. For independent groups of iid variables, say X_i and Y_i, *with a constant variance* σ^2 across groups
$$Z = \frac{\bar X - \bar Y - (\mu_x - \mu_y)}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$
limits to a standard normal random variable as both n_x and n_y get large. Here
$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$
is the **pooled estimate** of the variance. Obviously, $\bar X$, S_x, n_x are the sample mean, sample standard deviation and sample size for the X_i and $\bar Y$, S_y and n_y are defined analogously.

4. If the X_i and Y_i happen to be normal, then Z follows the Student's T distribution with $n_x + n_y - 2$ degrees of freedom.

5. The test statistic $TS = \frac{\bar X - \bar Y}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$ can be used to test the hypothesis that $H_0 : \mu_x = \mu_y$ versus the alternatives $H_1 : \mu_x < \mu_y$, $H_2 : \mu_x \neq \mu_y$ and $H_3 : \mu_x > \mu_y$. The test statistic should be compared to Student's T quantiles with $n_x + n_y - 2$ df.

6. $\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$ follows what is called the F distribution with $n_x - 1$ **numerator degrees of freedom** and $n_y - 1$ denominator degrees of freedom.

7. To test the hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ versus th hypotheses $H_1 : \sigma_x^2 < \sigma_y^2$, $H_2 : \sigma_x^2 \neq \sigma_y^2$ and $H_3 : \sigma_x^2 > \sigma_y^2$ compare the statistic $TS = S_1^2/S_2^2$ to the F distribution. We reject H_0 if:

   H_1 if $TS < F_{n_x-1,n_y-1,\alpha}$,

   H_2 if $TS < F_{n_x-1,n_y-1,\alpha/2}$ or $TS > F_{n_x-1,n_y-1,1-\alpha/2}$,

   H_3 if $TS > F_{n_x-1,n_y-1,1-\alpha}$.

8. The F distribution satisfies the property that $F_{n_x-1,n_y-1,\alpha} = F_{n_y-1,n_x-1,1-\alpha}$. So that, it turns out, that our results are consistent whether we put S_x^2 on the top or bottom.

9. Using the fact that
$$1 - \alpha = P\left(F_{n_x-1,n_y-1,\alpha/2} \leq \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \leq F_{n_x-1,n_y-1,1-\alpha/2}\right)$$

we can calculate a confidence interval for $\frac{\sigma_y^2}{\sigma_x^2}$ as $\left[F_{n_x-1,n_y-1,\alpha}\frac{S_x^2}{S_y^2}, F_{n_x-1,n_y-1,1-\alpha/2}\frac{S_x^2}{S_y^2}\right]$. Of course, the confidence interval for $\frac{\sigma_x^2}{\sigma_y^2}$ is $\left[F_{n_y-1,n_x-1,\alpha}\frac{S_y^2}{S_x^2}, F_{n_y-1,n_x-1,1-\alpha/2}\frac{S_y^2}{S_x^2}\right]$.

10. F tests are not robust to the normality assumption.

11. The statistic
$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$
follows a standard normal distribution for large n_x and n_y. It follows an approximate Students T distribution if the X_i and Y_i are normally distributed. The degrees of freedom are given below.

12. For testing $H_0 : \mu_x = \mu_y$ in the event where there is evidence to suggest that $\sigma_x \neq \sigma_y$, the test statistic $TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}}$ follows an approximate Student's T distribution under the null hypothesis when X_i and Y_i are normally distributed. The degrees of freedom are approximated with
$$\frac{(S_x^2/n_x + S_y^2/n_y)^2}{(S_x^2/n_x)^2/(n_x - 1) + (S_y^2/n_y)^2/(n_y - 1)}.$$

13. The power for a Z test of $H_0 : \mu_x = \mu_y$ versus $H_3 : \mu_x > \mu_y$ is given by
$$P\left(Z \geq Z_{1-\alpha} - \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}\right)$$
while for $H_1 : \mu_x < \mu_y$ it is
$$P\left(Z \leq -Z_{1-\alpha} - \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}\right).$$

14. Sample size calculation assuming $n_x = n_y = n$
$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2(\sigma_x^2 + \sigma_y^2)}{(\mu_x - \mu_y)^2}.$$