Formula Sheet for W3600.

Ch. 1. Simpson's paradox occurs when some overall conclusion governing a set of objects fails to hold within each collection of subsets of those subjects.

Sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2\}$.

Ch. 2.

1. Sets. $A \cap B = A$ and B, $A \cup B = A$ or B, $A^c =$ not A. Mutually exclusive: $A \cap B = \emptyset$.

2. Probability: a. $0 \leq P(A) \leq 1$. b. $P(S) = 1$ c. If A_1, A_2, ... are mutually exclusive, then $P(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} P(A_i)$.

3. Calculation of probability:

a. $P(\emptyset) = 0$, and $P(A^c) = 1 - P(A)$.

b. $P(A) = \frac{f}{N}$, where $f = \#$ of outcomes belong to A, and $N =$ total number of outcomes in S.

c. The Product Rule. $n_1 * n_2 * \cdots * n_r$

d. The Combination Rule

$$\frac{n*(n-1)*\cdots*(n-r+1)}{1*2*\cdots*r} = \frac{n!}{(n-r)!r!} = \binom{n}{r}.$$

4. Conditional probability: if $P(B) > 0$, $P(A$, given $B) = P(A|B) = \frac{P(A\cap B)}{P(B)}$.

5. Independence. Two events A and B are said to be independent, if $P(A|B) = P(A)$, or $P(A \cap B) = P(A)P(B)$, or $P(B|A) = P(B)$.

6. The law of total probability $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$. In general, $P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j)$, if $\cup_{i=1}^{n} B_i = S$ and B_i, $1 \leq i \leq n$, are mutually exclusive.

7. Bayes' formula:

$$P(B_i|A) = \frac{P(A|B_i)P(B)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}.$$

Ch. 3.

1. Distribution Function $F(b) = P(X \leq b)$.

2. Expectation for discrete random variables: $E(X) = \sum_x xP(X = x) = \mu$. $E(g(X)) = \sum_x g(x)P(X = x)$.

3. Variance. $Var(X) = E((X-\mu)^2) = \sigma^2 = EX^2 - (E(X))^2$, and $SD(X) = \sqrt{Var(X)} = \sigma$.

4. $Var(aX) = a^2Var(X)$, $E(aX + b) = aE(X) + b$, $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$.

5. Bernoulli random variable: $P(X = 1) = p$, $P(X = 0) = 1 - p$. Mean and variance $E(X) = p$, $Var(X) = p(1-p)$.

6. Binomial random variable: $X = \sum_{k=1}^{n} Y_k$, where Y_k's are independent Bernoulli random variables. It is often denoted by $Bino(n, p)$.

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad 0 \leq i \leq n.$$

$E(X) = np$. $Var(X) = np(1-p)$.

7. Geometric random variable: $P(X = k) = (1-p)^{k-1}p = q^{k-1}p$, for $k = 1, 2, ..., \infty$, where $q = 1-p$. Mean: $E(X) = 1/p$. Variance: $Var(X) = \frac{1-p}{p^2}$.

8. Negative binomial random variable: $X = \sum_{i=1}^{r} Y_i$, where Y_i are independent geometric random variables.

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, ..., \infty.$$

Mean and variance $E(X) = r/p$, $Var(X) = \frac{r(1-p)}{p^2}$.

9. Sampling without replacement and hypergeometric random variable: $P(X = x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$, where $\max(0, n+M-N) \leq x \leq \min(n, M)$. Mean and variance $E(X) = nM/N$, $Var(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$.

10. Markov Inequality: $P(X \geq a) \leq E(X^2)/a^2$, if $X \geq 0$. Chebyshev Inequality. Let X be a random variable with mean μ and variance σ^2. Then for any positive k, $P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$.

11. Poisson Distribution. It comes as an approximation for binomial distribution, if $np \to \lambda$ as $n \to \infty$. $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0, 1, 2, ...$ Mean and variance $E(X) = \lambda$, $Var(X) = \lambda$.

Ch. 4

1. Density function $f(x)$: a. $f(x) \geq 0$. b. $\int_{-\infty}^{\infty} f(x)dx = 1$. c. $P(a \leq X \leq b) = \int_a^b f(x)dx$.

2. Distribution function: $F(b) = P(X \leq b) = \int_{-\infty}^{b} f(x)dx$. $\frac{d}{dx}F(x) = f(x)$.

3. Expectation. $E(X) = \int_{-\infty}^{\infty} xf(x)dx$. Variance: $Var(X) = E(X^2) - (E(X))^2$.

In general, for a continuous random variable, $Eg(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$.

4. Uniform Distribution: the density function is $f(x) = \frac{1}{b-a}$, if $a \leq x \leq b$. The (cumulative) distribution function is given by $F(x) = P(X \leq x) = \frac{x-a}{b-a}$, $a \leq x \leq b$.

Mean and Variance $E(X) = \frac{a+b}{2}$, $Var(X) = \frac{1}{12}(b-a)^2$.

5. Gamma function $\Gamma(t)$ is defined as $\Gamma(t) = \int_0^{\infty} e^{-y}y^{t-1}dy$. In particular, $\Gamma(t) = (t-1)\Gamma(t-1)$. If n is a positive integer, $\Gamma(n) = (n-1)!$ and $\Gamma(1/2) = \sqrt{\pi}$.

6. Exponential distribution. Density: $f(x) = \frac{1}{\theta}e^{-x/\theta}$, $x \geq 0$. Mean and variance: $E(X) = \theta$, $Var(X) = \theta^2$. (Cumulative) distribution function $P(X \leq t) = 1 - e^{-t/\theta}$, $t \geq 0$. The exponential distribution has the "memoryless property": $P(X \geq t+t_0 | X \geq t_0) = P(X \geq t)$

7. Gamma distribution. Density $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$, $x \geq 0$. Mean and variance: $E(X) = \alpha\beta$, $Var(X) = \alpha\beta^2$. In particular, if $\alpha = 1$, then it becomes an exponential random variable with mean β.

If X_1, ..., X_n represent independent gamma random variables with parameters α and β. Then $Y = \sum_{i=1}^{n} X_i$ is still a gamma random variable with parameters $n\alpha$ and β.

8. Standard normal random variable $Z = N(0,1)$. Density $f(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$, $-\infty < z < \infty$. $\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$. If $\frac{X-\mu}{\sigma} = Z$, then X is called a normal random variable with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$.

9. χ^2 distribution with the d.f. ν is a special case of the gamma distribution with $\alpha = v/2$ and $\beta = 2$. Also $\chi^2 = \sum_{i=1}^n Z_i^2$, where $Z_1, ..., Z_n$ are independent $N(0, 1)$ random variables.

Ch. 5

1. Bivariate probability distribution and bivariate density.

a. Discrete case. Joint probability mass function: $p(x_1, x_2) = P(X_1 = x_1,\ X_2 = x_2)$. Marginal probability functions: $p_1(x_1) = P(X_1 \le x_1) = \sum_{x_2} p(x_1, x_2)$ and $p_2(x_2) = P(X_2 \le x_2) = \sum_{x_1} p(x_1, x_2)$.

b. Continuous case. Joint density $P((X, Y) \in C) = \int\int_{(x,y)\in C} f(x, y)dxdy$. Marginal densities: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$, and $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$.

2. Conditional densities. $f_{X|Y}(x|y) = f(x, y)/f_Y(y)$, if $f_Y(y) > 0$, is a density function of x for every fixed y. Similarly, $f_{Y|X}(y|x) = f(x, y)/f_X(x)$, if $f_X(x) > 0$, is a density function of y for every fixed x.

3. Independence.

a. Discrete case. $P(X = x, Y = y) = P(X = x)P(Y = y)$.

b. Continuous case. $f(x, y) = f_X(x)f_Y(y)$, $f_{X|Y}(x|y) = f_X(x)$, $f_{Y|X}(y|x) = f_Y(y)$.

4. Expectation of functions of random variables

a. Discrete case $E[g(X_1, X_2)] = \sum_{x_1}\sum_{x_2} g(x_1, x_2)p(x_1, x_2)$.

b. Continuous case $E[g(X_1, X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)f(x_1, x_2)dx_1dx_2$.

c. Independence implies $E(g(X_1)h(X_2)) = E(g(X_1))E(h(X_2))$.

5. Covariance. $Cov(X_1, X_2) = E(X_1X_2) - \mu_1\mu_2$, where $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$. If X_1 and X_2 are independent then $Cov(X_1, X_2) = 0$. However, the converse is not true.

6. Correlation coefficient between X_1 and X_2, $\rho = \frac{Cov(X_1,X_2)}{\sqrt{Var(X_1)Var(X_2)}}, 1 \le \rho \le 1$. For two independent random variables $\rho = 0$. But the converse is not true.

7. Let X_1, ..., X_m and Y_1,..., Y_n be random variables. Define $U_1 = \sum_{i=1}^n a_iY_i$ and $U_2 = \sum_{j=1}^m b_jX_j$ for constants $a_1, ..., a_n$ and b_1, ..., b_n. Then $E(U_1) = \sum_{i=1}^n a_iE(Y_i)$ and $E(U_2) = \sum_{j=1}^m b_jE(X_j)$.

$$Var(U_1) = \sum_{i=1}^n a_i^2 Var(Y_i) + 2\sum\sum_{i<j} a_ia_jCov(Y_i, Y_j),$$

where the double sum is over all pairs (i, j) with $i < j$. $Cov(U_1, U_2) = \sum_{i=1}^n\sum_{j=1}^m a_ib_jCov(Y_j, X_j)$.

If X_1, ..., X_m are independent, then $Var(\sum_{j=1}^m b_jX_j) = \sum_{j=1}^m b_j^2 Var(X_j)$

8. $E(X) = E(E(X|Y))$.

9. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then $E(\bar{X}) = \mu$, $Var(\bar{X}) = \sigma^2/n$, and $Cov(X_i - \bar{X},\ \bar{X}) = 0$. Define $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}[(\sum_{i=1}^n X_i^2) - n(\bar{X})^2]$.

10. The law of large numbers: For any $\epsilon > 0$, $P(|\bar{X} - \mu| > \epsilon) \to 0$, as $n \to \infty$. The result can be proved by using the Markov inequality.

11. The central limit theorem: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0, 1)$ and $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \approx N(0, 1)$.

12. Normal approximation for binomial. Use np and $np(1 - p)$ as mean and variance for the normal approximation, with possible continuity corrections (by adding 0.5).

Chapters 6 and 7

1. Unbiased estimator: $E(\hat{\theta}) = \theta$.

2. Mean square error: $E[(\hat{\theta} - \theta)^2]$.

3. The conditional expectation is the best predictor in terms of minimizing the mean square error. More precisely, for any deterministic function g, $E[(g(X)-Y)^2] \geq E[(E(Y|X) - Y)^2]$.

4. One-sample confidence intervals: $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$(for the large sample case), $\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ (for the small sample case, d.f.=$n-1$), $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, $(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}})$.

5. $(n-1)S^2/\sigma^2 = \chi^2_{n-1}$, if X_1, X_2, ... are normally distributed.

Ch. 8

1. The definitions of H_0, H_a, p-value, test statistic, rejection region, and the power of a test.

2. One-sample test statistics: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ (for the large sample case), $T_{n-1} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ (for the small sample case, d.f.=$n-1$), $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$, $\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$.

Ch. 9

1. Two-sample confidence intervals: $(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (for the large sample case), $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$(for the small sample case with $\sigma_1 = \sigma_2$, d.f.=n_1+n_2-2), $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$, $(\hat{p}_1-\hat{p}_2) \pm z_{\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$, $(\frac{s_2^2}{s_1^2}\frac{1}{F_{\alpha/2}(n_2-1,n_1-1)}, \frac{s_2^2}{s_1^2}F_{\alpha/2}(n_1-1, n_2-1))$

2. Two-sample test statistics: $Z = (\bar{X}_1 - \bar{X}_2 - D_0)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (for the large sample case), $T_{n_1+n_2-2} = (\bar{X}_1 - \bar{X}_2 - D_0)/\{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\}$(for the small sample case with $\sigma_1 = \sigma_2$, d.f.=n_1+n_2-2), $Z = (\hat{p}_1 - \hat{p}_2 - D_0)/\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$, $F_{n_1-1,n_2-1} = \frac{s_1^2}{s_2^2}$.

Ch. 12

1. The simple linear regression model: $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon = N(0, \sigma^2)$.

2. Notation: $SS_{xy} = \sum(X_i - \bar{X})(Y_i - \bar{Y})$, $SS_{xx} = \sum(X_i - \bar{X})^2$, $SS_{yy} = \sum(Y_i - \bar{Y})^2$.

3. $\hat{\beta}_1 = SS_{xy}/SS_{xx}$, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$, $\hat{\sigma}^2 = \frac{1}{n-2}SSE = \frac{1}{n-2}\sum_{i=1}^n e_i^2 = \frac{1}{n-2}(SS_{yy} - \hat{\beta}_1 SS_{xy})$, where $e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$.

4. $SD(\hat{\beta}_1) = \sigma/\sqrt{SS_{xx}} \approx \hat{\sigma}/\sqrt{SS_{xx}}$, and $SD(\hat{\beta}_0) = \sigma\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{SS_{xx}}} \approx \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{SS_{xx}}}$.

5. C.I's. $\hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \cdot SD(\hat{\beta}_1)$, $\hat{\beta}_0 \pm t_{\alpha/2,\, n-2} \cdot SD(\hat{\beta}_0)$. Test statistics: $T_{n-2} = \frac{\hat{\beta}_1 - \beta_1^{(H_0)}}{SD(\hat{\beta}_1)}$, $T_{n-2} = \frac{\hat{\beta}_0 - \beta_0^{(H_0)}}{SD(\hat{\beta}_0)}$.

6. $r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SS_{yy}} = \frac{SS_{xy}^2}{SS_{xx}SS_{yy}}$. Taking $\sqrt{}$, r is also called the sample correlation.

7. Prediction: the estimators will be the same $\hat{\beta}_0 + \hat{\beta}_1 X_0$. For the prediction of the mean response, the SD is about $\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{SS_{xx}}}$, while for the individual response,

it is $\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{SS_{xx}}}$. The c.i. is given by estimator$\pm t_{\alpha/2,n-2}$SD(estimator).

Ch. 14.

1. χ^2 tests. goodness of fit test: $\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$, d.f. $=k-1$, for the multinomial case.

2. for $(r\times c)$ table: $\sum\frac{(O_i-E_i)^2}{E_i}$, d.f. $=(r-1)(c-1)$, where E_i is the product of the corresponding row sum and column sum divided by the total sum.

Assumption: each cells must be large enough.

3. Test of independence and test of homogeneity for the $(r\times c)$ table.

4. A general formula: d.f.$=n-m$, where n is the number of free cells, and m is the number of parameters needed to be estimated. From this we can show that the d.f.'s for test of homogeneity and independence for $(r\times c)$ tables are both equal to $(r-1)(c-1)$.

4. If there are only two cells, the χ^2 test is the same as the z test for binomial p.