ISQA 8160 Cheat Sheet

Brian Detweiler

Summer, 2016

1 Distributions

Distribution	pmf/pdf	cdf	μ	σ^2
Bernoulli	p(0) = (1 - p), p(1) = p	0, (1- p), p†	p	$p(1-p)$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-t^2} dt$	μ	σ^2
Uniform	$\frac{1}{d-c}$ for $c \leq x \leq d$	$\frac{x-a}{b-a}$ for $x \epsilon [a, b)$	$\frac{c+d}{2}$	$\frac{1}{12}(d-c)^2$
Chi-Squared	$\frac{1}{\frac{\nu}{2}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}} e^{-\frac{x}{2}}$, $x > 0$	$\frac{1}{\Gamma(\frac{\nu}{2})}\gamma(\frac{k}{2}, \frac{x}{2})$	ν	2ν
Student's t	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \frac{2F_1(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}$	0*	$\frac{\nu}{\nu-2}$**
Fisher	$\frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\frac{\nu_1}{2}-1} (1 + \frac{\nu_1}{\nu_2} f)^{\frac{-\nu_1+\nu_2}{2}}$		$\frac{\nu_2}{\nu_2-2}$***	$\frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$****
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$\sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$
Geometric	$(1-p)^{x-1}p$	$1 - (1-p)^x$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
H-geometric	$\frac{\binom{X}{x}\binom{N-X}{n-x}}{\binom{N}{n}}$		$n\frac{x}{N}$	$n\frac{X}{N}\frac{N-K}{N}\frac{N-n}{N-1}$
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$I_X(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$	$\frac{\alpha}{\beta+\alpha}$	$(\frac{\alpha\beta}{\alpha+\beta})^2(\alpha+\beta+1)$
Exponential	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}$	$\frac{1}{\Gamma(\alpha)}\gamma(\alpha, \lambda x)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Multinom	$\frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$		np_i	
Poisson	$\frac{\lambda^x}{x!} e^{-\lambda}$	$e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}$	λ	λ

† 0 for $x < 0$, (1 - p) for $0 < x < 1$, and 1 for $x \geq 1$
* $\nu > 0$, undefined elsewhere
** for $\nu > 2$, ∞ for $1 < \nu \leq 2$
*** $\nu_2 > 2$
**** $\nu_2 > 4$

2 Means and Variances

$$
\begin{aligned}
E[X] &= \sum_{\forall x} x \cdot P(x) \text{ or } \int_{-\infty}^{\infty} xP(x)dx \\
Var(X) &= E[X^2] - \big(E[X]\big)^2 \\
\overline{X} &= \frac{\sum_{i=1}^{n} X_i}{n} \\
S^2 &= E\big[(X-\mu)^2\big] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
E[\overline{X}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu \\
Var[\overline{X}] &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}Var(\sigma^2) = \frac{n\sigma^2}{n} = \frac{\sigma^2}{n} \\
E[S^2] &= \sigma^2 \\
Var(S^2) &= Var\left(\frac{(n-1)S^2}{\sigma^2} \cdot \frac{\sigma^2}{(n-1)}\right) \\
&= \left[\frac{\sigma^2}{(n-1)}\right]^2 Var\left(\frac{(n-1)S^2}{\sigma^2}\right) \\
&= \frac{\sigma^4}{(n-1)^2} \cdot 2\nu \\
&= \frac{2\sigma^4}{(n-1)}
\end{aligned}
$$

3 PDF and CDF Definitions

$$
\begin{aligned}
F_X(x) &= P(X \le x) \\
&= \int_{-\infty}^{x} f(t)dt \\
P(a \le X \le b) &= \int_{a}^{b} f(t)dt \\
F'_X(x) &= f_X(x)
\end{aligned}
$$

4 Moment Generating Functions

$$M_X(t) = E[e^{tX}] = \sum_{x=1}^{\infty} e^{tx} f(x)$$

or

$$= \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

5 Normal Distribution

Z-scores located in *Table III.* The larger the score, the farther from μ.

$$z = \frac{x - \mu}{\sigma}$$

6 t-Distribution

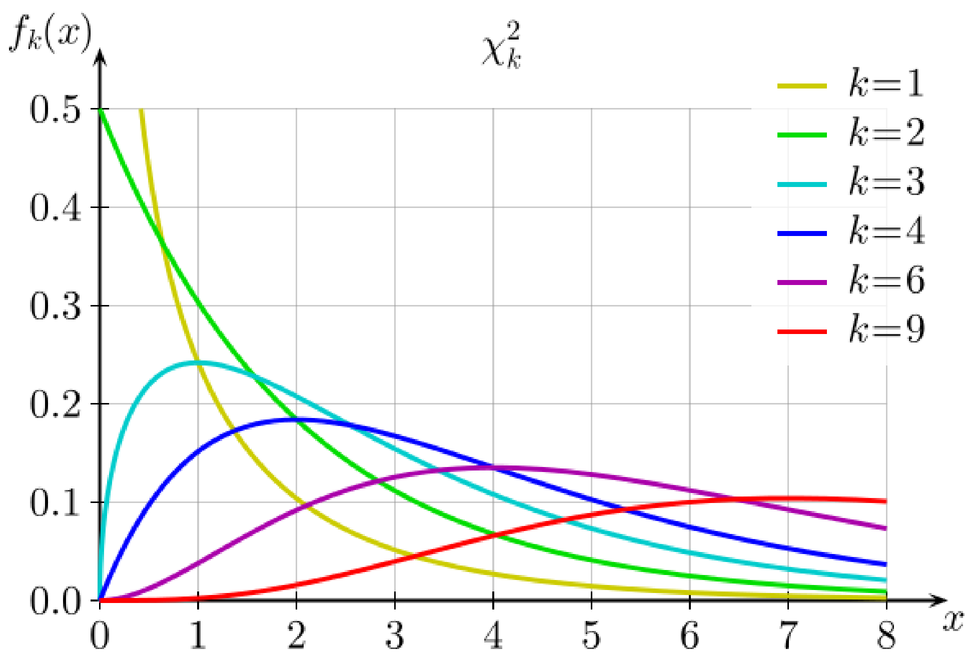
σ is unknown, but we know s.
$(\nu - 1)$ degrees of freedom.
If 1-degree of freedom, then Cauchy Distribution.
If the area of the tails is more than 0.10 (0.05 + 0.05, due to symmetry), then we do not have sufficient evidence to reject the claim (null hypothesis).

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$
$$P(T < -t) + P(T > t)$$

7 χ^2-Distribution

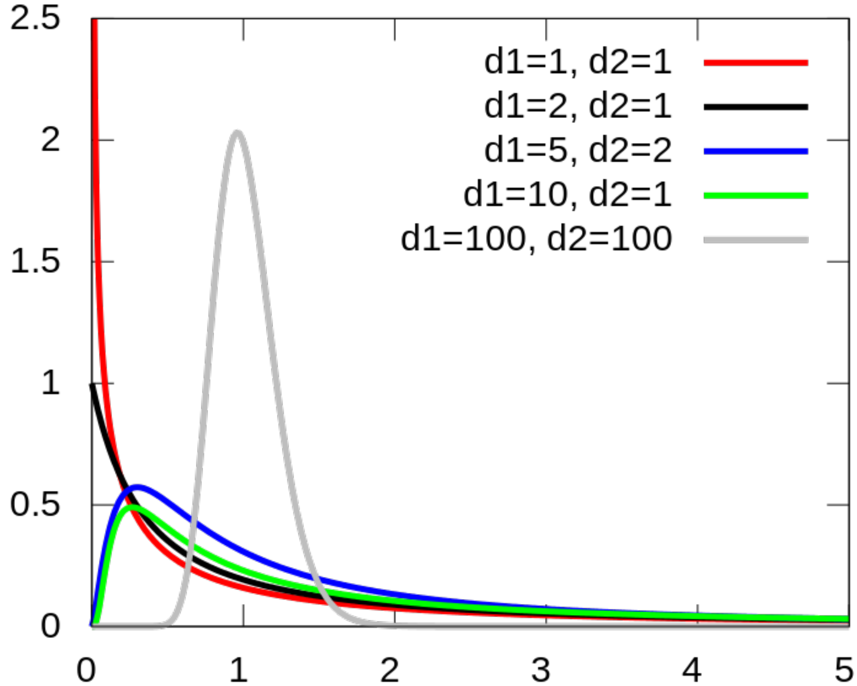
Assumes a normal distribution. Squares a normal.
Interested in σ^2.
ν degrees of freedom.
If $X_1, X_2, \ldots, X_n$ are chi-squared i.i.d. with pdf $N(0,1)$, then $Y = X_1^2 + X_2^2 + \ldots + X_n^2 \sim \chi^2(n)$.
Sum of χ^2's are still χ^2.
If S^2 exceeds a particular value (if $S^2 \geq \alpha$), we reject claim.
P(Rejecting claim when in fact it is true) $= P(\text{error}) = \alpha$.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$



8 f-Distribution

Let $U \sim \chi^2(\nu_1), V \sim \chi^2(\nu_2)$ be independent r.v.s.
Then $F = \frac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}}$ is an f-dist with (ν_1, ν_2) degrees of freedom.
$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 2)$
$P(F \leq f) = P\big(\frac{1}{F} \geq \frac{1}{f}\big)$
If $F(a,b)$, then $\frac{1}{F} = F(b,a)$.



9 Chebyshev's Inequality

$$P\big(|\overline{X}-\mu| < k\frac{\sigma}{\sqrt{n}}\big) \geq 1 - \frac{1}{k^2}$$

10 CLT

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

$$P(\frac{a-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{b-\mu}{\frac{\sigma}{\sqrt{n}}})$$

where a and b are the values we are testing and σ and n are given.

11 Incomplete Beta Function

$$B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$$

$$I_x(a,b) = \frac{B(x;a,b)}{B(a,b)}$$

$$I_0(a,b) = 0$$

$$I_1(a,b) = 1$$

$$I_x(a,1) = x^a$$

$$I_x(1,b) = 1-(1-x)^b$$

$$I_x(a,b) = 1 - I_{1-x}(b,a)$$

$$I_x(a+1,b) = I_x(a,b) - \frac{x^a(1-x)^b}{aB(a,b)}$$

$$I_x(a,b+1) = I_x(a,b) + \frac{x^a(1-x)^b}{bB(a,b)}$$

12 Order Statistics

$$P(Y_i \leq y) = \sum_{k=i}^{n} \binom{n}{k} \big(F(y)\big)^{k-1}\big(1-F(y)\big)^{n-k}$$

$$g_{Y_k}(y) = \frac{n!}{(k-1)!(n-k)!}\big[F(y)\big]^{k-1} \cdot \big[1-F(y)\big]^{n-k} \cdot f(y)$$

$$g_{Y_1}(y) = n \cdot \big[1-F(y)\big]^{n-1} \cdot f(y)$$

$$g_{Y_n}(y) = n\big[F(y)\big]^{n-1} \cdot f(y)$$

13 Maximum Likelihood Estimators

PROCEDURE TO FIND MLE

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm (ln) of $L(\theta)$.
3. When applicable, differentiate $\ln L(\theta)$ with respect to θ, and then equate the derivative to zero.

4. Solve for the parameter θ, and we will obtain $\hat{\theta}$.
5. Check whether it is a maximizer or global maximizer.

Geometric:

$$\theta(1-\theta)^{x-1}\ ,\ \hat{\theta}=\frac{1}{\overline{X}}$$

Normal:

$$\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-\mu)^2}{2\theta}}\ ,\ \hat{\theta}=\overline{X}$$
$$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\theta)^2}{2\sigma^2}}\ ,\ \hat{\theta}=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$
$$\frac{1}{\sqrt{2\pi\theta_2}}e^{-\frac{(x-\theta_1)^2}{2\theta_2^2}}\ ,\ \hat{\theta}_1=\overline{X}\ ,\ \hat{\theta}_2=\frac{n-1}{n}S^2$$

Uniform:

$$\frac{1}{\theta}\ ,\ \hat{\theta}=X_{(n)}$$

Poisson:

$$\frac{\theta^x e^{-\theta}}{x!}\ ,\ \hat{\theta}=\overline{X}$$

14 Natural Log Properties

$$\begin{aligned}
\ln(x\cdot y) &= \ln x+\ln y\\
\ln\frac{x}{y} &= \ln x-\ln y\\
\ln x^y &= y\cdot\ln x\\
\ln x\frac{d}{dx} &= \frac{1}{x}\\
\int\ln x dx &= x\cdot(\ln x-1)+c\\
\ln -x &= \text{ undefined}\\
\ln 0 &= \text{ undefined}\\
\ln 1 &= 0\\
\lim_{x\to\infty}\ln x &= \infty\\
\ln\frac{1}{(\theta\sqrt{2\pi})^n} &= -\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\theta
\end{aligned}$$

15 Gamma Function

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^{infty} e^{-x}x^{\alpha-1}dx\\
\Gamma(\alpha+1) &= \alpha\Gamma(\alpha)\\
\Gamma(n+1) &= n!\\
\Gamma(\frac{1}{2}) &= \sqrt{\pi}\\
\Gamma(\frac{3}{2}) &= \frac{1}{2}\Gamma(\frac{1}{2})=\frac{\sqrt{\pi}}{2}
\end{aligned}$$

16 Beta Function

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt \text{ for } x > 0, y > 0$$

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \text{ for } x, y \epsilon \mathbf{I}, x > 0, y > 0$$

17 Method of Moments

$$m_k' = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$$m_k = \mu_k \text{ NOTE: May need to find the kth moment}$$

$$m_1 = \mu_1$$

$$\overline{x} = \mu$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2$$

After solving parameter, plug it back into the distribution $f(x;\theta)$).

18 Unbiased Estimators

For a normal population $N(\mu, \sigma^2)$,

$$\begin{aligned}
\hat{\mu} &= \overline{X} \\
\hat{\sigma^2} &= \frac{n-1}{n} S^2 \\
E[\hat{\mu}] &= E[\overline{X}] = \mu \text{ (this is unbiased)}\checkmark \\
E[\hat{\sigma^2}] &= E\left[\frac{n-1}{n} S^2\right] \\
&= E\left[\frac{(n-1)S^2}{\sigma^2} \cdot \frac{\sigma^2}{n}\right] \text{ (first term is } \chi^2(n-1) \text{ and second is constant)} \\
&= \frac{\sigma^2}{n} E\left[\frac{(n-1)S^2}{\sigma^2}\right] \\
&= \frac{\sigma^2}{n}(n-1) \neq \sigma^2 \text{ biased } \boxtimes \\
E[S^2] &= \sigma^2 \text{ (this is unbiased)}\checkmark
\end{aligned}$$

18.1 Asymptotically Unbiased Estimators

Unbiased for large samples. $\lim_{n \to \infty} \frac{n}{n+1} \theta = \theta$

18.2 Consistent Estimators

If $Var(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ.

18.3 Fisher Information

$$I(\theta) = -E\left[\frac{\partial}{\partial \theta^2} \ln f(x; \theta)\right]$$

18.4 Cramér-Rao Lower Bound

Let $\hat{\theta}$ be an unbiased estimator of θ. Then,

$$Var(\hat{\theta}) \geq \frac{1}{n} \cdot \frac{1}{I(\theta)}$$

is the minimum possible value of variance (Cramér-Rao lower bound).

18.5 Minimum Variance Unbiased Estimator (MVUE)

If $Var(\hat{\theta}) = \frac{1}{n} \frac{1}{I(\theta)}$ then $\hat{\theta}$ is the MVUE.

18.6 Efficient Estimators

Smaller variance is more efficient. The ratio of the C-R bound, $\frac{C-R}{Var(\hat{\theta})} \epsilon [0, 1]$ is called the efficiency of $\hat{\theta}$. If $\frac{C-R}{Var(\hat{\theta})} = \frac{1}{2} \iff 2 \cdot C - R = Var(\hat{\theta})$, we we need twice as many observations to do as well an estimation as can be done with the MVUE.

The ratio $\frac{Var(\hat{\theta}_1}{Var(\hat{\theta}_2}$ is the efficiency of $\hat{\theta}_2$ relative to $\hat{\theta}_1$.

18.7 Sufficient Estimators

Let $f(x_1, x_2, \ldots, x_n; \theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)$ be the joint pdf of $(x_1, \ldots, x_n)$.

The statistic $\hat{\theta}$ is a **sufficient estimator** of θ iff $f(x_1, x_2, \ldots, x_n; \theta)$ can be written as $f(x_1, x_2, \ldots, x_n; \theta) = \phi(\hat{\theta}, \theta) \cdot h(x_1, x_2, \ldots, x_n)$ where ϕ depends only on $\hat{\theta}, \theta$ and h doesn't depend on θ.

19 Estimation of Means

19.1 σ^2 is known

Start with $N(\mu, \sigma^2)$. $\overline{X}$ is a "nice" estimate of μ. By normalizing it, we can say $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

If $1-\alpha$ is the area of the "body" of our distribution, then $\frac{\alpha}{2}$ are the tails.

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\left(\overline{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \text{ is a } (1-\alpha)100\% \text{ CI for } \mu$$

$$Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \text{ is called the "max error"}$$

$$Z_{\frac{0.10}{2}} = Z_{0.05} = 1.645 \text{ is } 90\% \text{ Confidence}$$
$$Z_{\frac{0.05}{2}} = Z_{0.025} = 1.96 \text{ is } 95\% \text{ Confidence}$$
$$Z_{\frac{0.01}{2}} = Z_{0.005} = 2.575 \text{ is } 99\% \text{ Confidence}$$

Given the error, find the sample size needed

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$
$$\Rightarrow E\sqrt{n} = Z_{\frac{\alpha}{2}}\sigma$$
$$\Rightarrow (\sqrt{n})^2 = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$$
$$\Rightarrow n \geq \left\lceil \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 \right\rceil$$

19.2 σ^2 is unknown

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim T(n-1) \text{ for n small}$$
$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim N(0,1) \text{ for n large}$$

$$\left(\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right)$$

20 Estimation of Differences between Means

20.1 σ_1^2, σ_2^2 Are Known

With n_1, n_2 large, $\overline{X}_1, \overline{X}_2$ are point estimates for μ_1, μ_2.

$$\overline{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$
$$\overline{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$
$$(\overline{X}_1 - \overline{X}_2) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$
$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$\left((\overline{X}_1 - \overline{X}_2) - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\overline{X}_1 - \overline{X}_2) + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \text{ is a } (1-\alpha)100\% \text{ CI for } (\mu_1 - \mu_2)$$

$$\left((\overline{X}_1 - \overline{X}_2) - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\overline{X}_1 - \overline{X}_2) + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \text{ is a } (1-\alpha)100\% \text{ CI for } (\mu_1 - \mu_2)$$

20.2 σ_1^2, σ_2^2 Are Unknown

Assume $\sigma_1^2 = \sigma_2^2$. The following "pooled estimate", is an unbiased estimate of σ^2

$$\begin{aligned} S_p^2 &= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \\ &= \frac{n_1-1}{n_1+n_2-2}S_1^2 + \frac{n_2-1}{n_1+n_2-2}S_2^2 \end{aligned}$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T(n_1+n_2-2)$$

The $(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$ is

$$\left((\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}, n_1+n_2-2} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\overline{X}_1 - \overline{X}_2) + t_{\frac{\alpha}{2}, n_1+n_2-2} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

21 Estimation of Proportions

21.1 Normal Approximation to the Binomial

$\hat{\theta} = \frac{X}{n}$ is a *nice* estimator.

$$E[\hat{\theta}] = E\left[\frac{X}{n}\right] = \frac{n\theta}{n} = \theta \checkmark$$
$$Var(\hat{\theta}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}n\theta(1-\theta) = \frac{\theta(1-\theta)}{n} \rightarrow_{n\rightarrow\infty} 0 \checkmark$$

$$\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}} \sim N(0,1) \text{ if } n\theta \geq 5, n(1-\theta) \geq 5$$

$$\Rightarrow \frac{\frac{X}{n}-n\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \sim N(0,1)$$

$$\left(\hat{\theta} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}, \hat{\theta} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\right), \text{ With } \hat{\theta} = \frac{X}{n}$$

21.2 Differences in Normal Approximations to the Binomial

$\hat{\theta}_1 = \frac{X_1}{n_1}, \hat{\theta}_2 = \frac{X_2}{n_2}$ are *nice* estimators, so we construct $\hat{\theta}_1 - \hat{\theta}_2$.

$E[\hat{\theta}_1 - \hat{\theta}_2] = E\left[\frac{X_1}{n_1}\right] - E\left[\frac{X_2}{n_2}\right] = \frac{1}{n_1}n_1\theta_1 - \frac{1}{n_1}n_1\theta_1 = \theta_1 - \theta_2$.

$Var(\hat{\theta}_1 - \hat{\theta}_2) = Var(\hat{\theta}_1) + Var(\hat{\theta}_2) = \frac{\hat{\theta}_1(1-\hat{\theta}_1}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2}{n_2}$.

We use the statistic constructed by normalizing,

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2}{n_2}}}$$

So the $(1-\alpha)100\%$ CI for $\theta_1 - \theta_2$ is given by

$$\left((\hat{\theta}_1 - \hat{\theta}_2) - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}, (\hat{\theta}_1 - \hat{\theta}_2) + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}\right), \text{ With } \hat{\theta}_1 = \frac{X_1}{n_1}, \hat{\theta}_2 = \frac{X_2}{n_2}$$

22 Estimation of Variances

Starting with a normal population $N(\mu, \sigma^2)$,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right)$$
$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},(n-1)}}, \frac{(n-1)S^2}{\chi^2_{\frac{1-\alpha}{2},(n-1)}},\right)$$

NOTE: take the square root to get Standard Deviation

22.1 Proportions of Variances

Recall,

$$F \sim F(n_1-1, n_2-1)$$
$$\frac{1}{F} \sim F(n_2-1, n_1-1)$$

$$\left(\frac{\frac{S_1^2}{S_2^2}}{f_{\frac{\alpha}{2},n_1-1,n_2-1}}, \left(\frac{S_1^2}{S_2^2}\right) f_{\frac{\alpha}{2},n_2-1,n_1-1}\right)$$

23 Hypothesis Testing

1.) Start with a normal population, $N(\mu, \sigma^2)$.
Want to test validity of a claim on value of μ.

2.) Null Hypothesis
$H_0 : \mu = \mu_0$, where $\mu_0 =$ some value.

Alternative Hypothesis
$H_1 : \mu \neq \mu_0$ (**Two Sided / Simple Hypothesis**), or $\mu < \mu_0$, or $\mu > \mu_0$ (**One Sided / Composite Hypotheses**).

3.) Test statistic

$$\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

4.) Critical region of size α. $P\left(\left|\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right| < Z_{\frac{\alpha}{2}}\right) = \alpha$

H_0	H_1	C
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right\rvert \geq z_{\frac{\alpha}{2}}$
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -z_\alpha$
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha$

$$C = \left\{ \overline{x} \middle| \left|\frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right| \geq Z_{\frac{\alpha}{2}} \right\}$$

5.) Variations

- Table valid for non-normal populations with large samples
- σ unknown; replace $\sigma \to s$. For large samples use the normal statistic above. For small samples, use the T-distribution.

$$\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim T(n-1)$$

23.1 P-Values

Proceed as normal with other hypothesis testing, but then when we get our actual Z-value, find the corresponding probability from the tables to get the critical region.

For instance, if we are measuring against an α of 0.05, then the Z-value we are measuring against is 1.96 (two-tailed). If our test statistic produces 2.60, then finding the Z-value from *Table III* gives us 0.4953. Since it is a two-tailed test, we multiply by two and subtract the whole thing from 1. This is our P-value (the area in the critical region).

23.2 Hypothesis Tests Concerning Means

H_0	H_1	C
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \right\rvert \geq z_{\frac{\alpha}{2}}$
$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -z_\alpha$
$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha$

Table is valid for $N(\mu, \sigma^2)$, where σ is known and non-normal populations w/ large samples.

If σ is unknown, replace it by s.

If $n \geq 30$, use Normal.

If $n < 30$ use T-distribution and assume normality.

23.3 Hypothesis Tests Concerning Differences of Means

H_0	H_1	C
$\mu_0 - \mu_1 = \delta$	$\mu_0 - \mu_1 \neq \delta$	$\left\lvert \frac{\overline{x}_1-\overline{x}_2-\delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \right\rvert \geq Z_{\frac{\alpha}{2}}$
$\mu_0 - \mu_1 = \delta$	$\mu_0 - \mu_1 < \delta$	$\frac{\overline{x}_1-\overline{x}_2-\delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} < -Z_\alpha$
$\mu_0 - \mu_1 = \delta$	$\mu_0 - \mu_1 > \delta$	$\frac{\overline{x}_1-\overline{x}_2-\delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} > Z_\alpha$

Consider two independent normal populations, $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$.

Assume σ_1^2, σ_2^2 are known.

Sample sizes are n_1, n_2.

23.4 Hypothesis Tests Concerning One Variance

1.) Assume a Normal population $N(\mu, \sigma^2)$.

2.) We want to test
$H_0 : \sigma^2 = \sigma_0^2$
$H_1 : \sigma^2 \neq \sigma_0^2, \sigma^2 < \sigma_0^2, \sigma^2 > \sigma_0^2$

3.) Get our test statistic

$$\frac{(n-1)s^2}{\sigma^{2^*}} \sim \chi^2(n-1)$$

* Here, the σ^2 will be replaced by σ_0^2.

H_0	H_1	C
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\frac{\alpha}{2},n-1}$
	OR	$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\frac{\alpha}{2},n-1}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\alpha,n-1}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\alpha,n-1}$

23.5 Hypothesis Tests Concerning Two Variances

1.) Assume independent Normal populations $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ with sample sizes n_1, n_2.

2.) We want to test
$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2, \sigma_1^2 < \sigma_2^2, \sigma_1^2 > \sigma_2^2$

3.) Get our test statistic

$$\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 1)$$

* Here, the σ^2 will be replaced by σ_0^2.

H_0	H_1	C
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\frac{s_1^2}{s_2^2} > f_{\frac{\alpha}{2},n_1-1,n_2-1}$
	OR	$\frac{s_2^2}{s_1^2} > f_{\frac{\alpha}{2},n_2-1,n_1-1}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$\frac{s_1^2}{s_2^2} > \frac{1}{f_{\alpha,n_2-1,n_1-1}}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$\frac{s_1^2}{s_2^2} < f_{\alpha,n_1-1,n_2-1}$

23.6 Hypothesis Tests Concerning Proportions

1.) If $n > 30$, use normal approximation to the binomial.

$$\frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{\frac{X}{n} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \sim N(0,1)$$

H_0	H_1	C
$\theta = \theta_0$	$\theta \neq \theta_0$	$\left\lvert \frac{\frac{X}{n} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \right\rvert > Z_{\frac{\alpha}{2}}$
$\theta = \theta_0$	$\theta > \theta_0$	$\frac{\frac{X}{n} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} > Z_{\alpha}$
$\theta = \theta_0$	$\theta < \theta_0$	$\frac{\frac{X}{n} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} < -Z_{\alpha}$

2.) If $n < 30$, we have to get the P-value using the binomial pdf.

$$2P(X \leq x) = 2\left[\sum_{i=0}^{x} \binom{n}{i} \theta^i (1-\theta)^{n-i}\right]$$

23.7 Hypothesis Tests Concerning Differences among K-Proportions

	< 30k	≥ 30k	Total
A	f_{11} / e_{11}	f_{12} / e_{12}	n_1
B	f_{21} / e_{21}	f_{22} / e_{22}	n_2
	$f_{\cdot 1}$	$f_{\cdot 2}$	f

Where

$$
\begin{aligned}
f_{i1} &= x_i \\
f_{i2} &= n_i - x_i \\
e_{i1} &= n_i \hat{\theta} \\
e_{i2} &= n_i (1 - \hat{\theta})
\end{aligned}
$$

Case 1. Known value

1.) Hypothesis

$$
\begin{aligned}
&H_0 : \theta_1 = \theta_2 = \ldots = \theta_0 \text{ known value} \\
&H_1 : \exists i \in \{1, 2, \ldots, k\}, \theta_i \neq \theta_0
\end{aligned}
$$

2.) Test statistic **chi-squared**

$$\sum_{i=1}^{k} \left(\frac{x_i - n_i \theta_i}{\sqrt{n_i \theta_i (1 - \theta_i)}} \right)^2 \sim \chi^2(k)$$

3.) Critical region is determined by

$$\chi^2 > \chi^2_{\alpha,k}$$

Case 2. No known value

1.) Hypothesis

$$
\begin{aligned}
&H_0 : \theta_1 = \theta_2 = \ldots = \theta_k \text{ no known value} \\
&H_1 : \text{not all are equal}
\end{aligned}
$$

2.) Pooled Estimate $\hat{\theta}$.

$$\hat{\theta} = \frac{x_1 + x_2 + \ldots + x_k}{n_1 + n_2 + \ldots + n_k}$$

3.) Statistic

$$\sum_{i=1}^{k}\left(\frac{x_i - n_i\hat{\theta}}{\sqrt{n_i\hat{\theta}(1-\hat{\theta})}}\right)^2 \sim \chi^2(k-1)$$

4.) Critical region given by

$$\chi^2 > \chi^2_{\alpha,k-1}$$
$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} > \chi^2_{\alpha,\text{k - 1}}$$

23.8 Contingency Tables / rXc Tables (Multinomial)

We want θ_{ij}, which is the probability of the jth outcome for the ith population.

$$\begin{aligned}
i &: r \text{ (row)}\\
j &: c \text{ (column)}\\
H_0 &: \theta_{1j} = \theta_{2j} = \ldots = \theta_{rj}, j = 1, 2, \ldots, c\\
H_1 &: \text{ not all equal}\\
\theta_{ij} &= \text{ probability of falling in cell (i, j)}\\
\hat{\theta}_{i\cdot} &= \frac{f_{i\cdot}}{f}\\
\hat{\theta}_{\cdot j} &= \frac{f_{\cdot j}}{f}\\
\hat{\theta}_{ij} &= \hat{\theta}_{i\cdot}\cdot\hat{\theta}_{\cdot j}\\
&= \frac{f_{i\cdot}}{f}\cdot\frac{f_{\cdot j}}{f}\\
&= \frac{(f_{i\cdot})\cdot(f_{\cdot j})}{f^2}\\
\chi^2 &: \sum\frac{(F-E)^2}{E} > \chi^2_{\alpha,(r-1)(c-1)}
\end{aligned}$$

	Poor	Fair	Good	Total
A	f_{11} / e_{11}	f_{12} / e_{12}	f_{12} / e_{12}	$f_{1\cdot}$
B	f_{21} / e_{21}	f_{22} / e_{22}	f_{12} / e_{12}	$f_{2\cdot}$
C	f_{21} / e_{21}	f_{22} / e_{22}	f_{12} / e_{12}	$f_{3\cdot}$
	$f_{\cdot 1}$	$f_{\cdot 2}$	$f_{\cdot 3}$	f

Make the following table:

Rays	(F) Occur.	$P(X = x)$	E	$\frac{(F-E)^2}{E}$
0	19	0.09071795329	20.86512926	0.1667234888
1	54	0.2177230879	50.07631022	0.08608989148
2	58	0.2612677055	60.09157226	0.5809386751
3	23	0.2090141644	48.07325781	40.40608318
4	6	0.1254084986	28.84395468	7.963738403
5	6	0.06019607934	13.84509825	32.3240659
6	6	0.02407843174	5.538039299	3.594971472
7+	6	0.01159407926	2.66663823	0.6667022138
	230	1	230	85.78931322

First column is the measurement
Second is the frequency
Third is based on the assumed distribution's PDF
Fourth is the expected value obtained by taking the probability * total frequency
Last is summed to get the χ^2 statistic in the bottom right cell

Finding degrees of freedom

$s =$ total number of cells
$t =$ number of estimated parameters

$$\begin{aligned} s - t - 1 &= r \cdot c - (r + c - 2) - 1 \\ &= r \cdot c - r - c + 2 - 1 \\ &= r \cdot c - r - c + 1 \\ &= r(c - 1) - (c - 1) \\ &= (r - 1)(c - 1) \end{aligned}$$

23.9 Goodness of Fit

# Sold	# Days	$P(X = x)$	E	$\frac{(F-E)^2}{E}$
0	1	0.001	0.3	$\frac{(1-0.3)^2}{0.3}$
1	16	0.027	8.1	$\frac{(16-8.1)^2}{8.1}$
2	55	0.243	72.9	$\frac{(55-72.9)^2}{72.9}$
3	228	0.729	218.7	$\frac{(228-218.7)^2}{218.7}$
	300	1	300	$\chi^2 = 14.1289$

Goal: Determine if a dataset may be looked upon as a random sample having a given distribution.

Need to estimate θ.

Find the expected frequencies of E for the given values $0, 1, 2, \ldots$.

Find $\chi^2 = \sum \frac{(F-E)^2}{E} > \chi^2_{\alpha,s-t-1}$.

Find $\hat{\theta}$ [1]

[1] Note: If PDF of population is given and we don't have to estimate a parameter, $t = 0$.

24 Nonparametric Statistics

24.1 3.1) Binomial Test

1.) Want to test validity of a claim on value of p.

2.) $n \leq 20$

3.) Null Hypothesis
$H_0 : p = p^*$, where $p^* =$ some value.

Alternative Hypothesis
$H_1 : p \neq p^*$ (**Two Sided / Simple Hypothesis**), or $p < p^*$, or $p > p^*$ (**One Sided / Composite Hypotheses**).

4.) Critical region of size α.
Left-tailed tests: $P(y \leq t) = \alpha$
Right-tailed tests: $P(y \leq t) = 1 - \alpha$
Double-tailed tests: $P(y \leq t_1) = \alpha_1$ or $P(y \leq t_2) = \alpha_2$

H_0	H_1	Critical Region	P-value
$p = p^*$	$p \neq p^*$	$c = \{T : T \leq t_1 \text{ or } T > t_2\}$	$min\{2 \cdot P(y \leq T), 2 \cdot P(y \geq T)\}$
$p \geq p^*$	$p < p^*$	$c = \{T : T \leq t\}$	$P(y \leq T)$
$p \leq p^*$	$p > p^*$	$c = \{T : T \geq t\}$	$P(y \geq T)$

24.2 3.1) Binomial Test (Normal Approximation)

1.) Want to test validity of a claim on value of p.

2.) $n > 20$

3.) Null Hypothesis
$H_0 : p = p^*$, where $p^* =$ some value.

Alternative Hypothesis
$H_1 : p \neq p^*$ (**Two Sided / Simple Hypothesis**), or $p < p^*$, or $p > p^*$ (**One Sided / Composite Hypotheses**).

4.) Critical region of size α.
Left-tailed tests: $P(y \leq t) = \alpha$
Right-tailed tests: $P(y \leq t) = 1 - \alpha$
Double-tailed tests: $P(y \leq t_1) = \alpha_1$ or $P(y \leq t_2) = \alpha_2$

H_0	H_1	Critical Region	P-Value
$p = p^*$	$p \neq p^*$	$c = \{T : T \leq np^* + z_{\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}$ or $T > np^* + z_{1-\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}\}$	$min\left\{2P\left(z \leq \frac{T-np^*+0.5}{\sqrt{np^*(1-p^*)}}\right), 2\left(1 - P\left(z \leq \frac{T-np^*+0.5}{\sqrt{np^*(1-p^*)}}\right)\right)\right\}$
$p \geq p^*$	$p < p^*$	$c = \{T : T \leq np^* + z_{\alpha}\sqrt{np^*(1-p^*)}$	$P\left(z \leq \frac{T-np^*+0.5}{\sqrt{np^*(1-p^*)}}\right)$
$p \leq p^*$	$p > p^*$	$c = \{T : T > np^* + z_{1-\alpha}\sqrt{np^*(1-p^*)}$	$1 - P\left(z \leq \frac{T-np^*+0.5}{\sqrt{np^*(1-p^*)}}\right)$

24.3 3.2) Quantile Test

1.) Want to test validity of a claim on value of x^*, which is the $*^{th}$ population quantile.

2.) $n \leq 20$

3.) $T_1 = \#$ of $x_i \leq x^*$
$T_2 = \#$ of $x_i < x^*$

H_0	H_1	Critical Region	P-Value
The p^* quantile of $X = x^*$ $P(X \leq x^*) = p^*$ $p = p^*$	$p \neq p^*$	$c = \{T_1 \leq t_1$ or $T_2 > t_2\}$ find t_1, t_2 s.t. $P(y \leq t_1) = \frac{\alpha}{2}$ $P(y \leq t_2) = 1 - \frac{\alpha}{2}$	$min\{2P(y \leq T_1), 2P(y \geq T_2)\}$
$p \geq p^*$	$p < p^*$	$c = \{T_1 : T_1 \leq t_1\}$ find t_1 s.t. $P(y \leq t_1) = \alpha$	$P(y \leq T_1)$
$p \leq p^*$	$p > p^*$	$c = \{T_2 : T_2 > t_2\}$ find t_2 s.t. $P(y \leq t_2) = 1 - \alpha$	$P(y \geq T_2)$

24.4 3.2) Quantile Test (Normal Approximation)

1.) Want to test validity of a claim on value of x^*, which is the $*^{th}$ population quantile.

2.) $n > 20$

3.) $T_1 = \#$ of $x_i \leq x^*$
$T_2 = \#$ of $x_i < x^*$

H_0	H_1	Critical Region	P-Value
The p^* quantile of $X = x^*$ $P(X \leq x^*) = p^*$ $p = p^*$	$p \neq p^*$	$c = \{T_1 \leq t_1$ or $T_2 > t_2\}$ find t_1, t_2 s.t. $t_1 = np^* + z_{\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}$ and $t_2 = np^* + z_{1-\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}$	$min\{2P(y \leq T_1), 2P(y \geq T_2)\}$
$p \geq p^*$	$p < p^*$	$c = \{T_1 : T_1 \leq t_1\}$ find t_1 s.t. $t_1 = np^* + z_{\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}$	$P(y \leq T_1)$ $= P\left[z \leq \frac{T_1 - np^* + 0.50}{\sqrt{np^*(1-p^*)}}\right]$
$p \leq p^*$	$p > p^*$	$c = \{T_2 : T_2 > t_2\}$ find t_2 s.t. $t_2 = np^* + z_{1-\frac{\alpha}{2}}\sqrt{np^*(1-p^*)}$	$P(y \geq T_1)$ $= 1 - P\left[z \leq \frac{T_2 - np^* + 0.50}{\sqrt{np^*(1-p^*)}}\right]$

24.5 3.3) Tolerance Limits

Method A. To find n when q is known.

How large n should be with $1-\alpha\%$ confidence that greater than or equal to $q\%$ of the population will be from $x^{(1)}$ (lowest) and $x^{(n)}$ (highest) or,

$$X^{(r)} \leq \text{ at least } q\% \text{ of the population } \leq x^{(n+1-m)}$$
$$n \cong \frac{1}{4}\chi^2_{1-\alpha,2(r+m)}\frac{1+q}{1-q}+\frac{1}{2}(r+m-1)$$

Note: If either $r=0$ or $m=0$, it will be a one-sided Tolerance Limit.

Method B. To find q when n is known.

Given we know n, what proportion q of the population (at least) are within a sample range with $1-\alpha\%$ confidence.

$$\text{from } x^{(r)} \text{ and } x^{(n+1-m)}$$

$$q = \frac{4n-2(r+m-1)-\chi^2_{1-\alpha,2(r+m)}}{4n-2(r+m-1)+\chi_{1-\alpha,2(r+m)}}$$

Note: If either $r=0$ or $m=0$, it will be a one-sided Tolerance Limit.

24.6 3.4) The Sign Test ($n \leq 20$)

1.) $n \leq 20$

H_0	H_1	Critical Region	P-Value
$P(+)=P(-)$	$P(+)\neq P(-)$	$c=\{T : T \leq t \text{ or } T \geq n-t\}$	$min\{2P(y\leq T_1), 2P(y\geq T_2)\}$
$P(+)\geq P(-)$	$P(+)<P(-)$ or $P\geq 0.50$ and $P<0.50$	$c=\{T : T\leq t\}$ where $T=\#$ of '+' signs and $t: P(y\leq t)=\alpha$	$P(y\leq T)$
$P(+)\leq P(-)$	$P(+)>P(-)$ or $P\geq 0.50$ and $P<0.50$	$c=\{T : T\geq n-t\}$ where $n=$ excluding $\#$ of ties and $t: P(y\leq n-t)=1-\alpha$	$P(y\leq T)$

24.7 3.4) The Sign Test ($n > 20$)

1.) $n > 20$

H_0	H_1	Critical Region	P-Value
$P(+)=P(-)$	$P(+)\neq P(-)$	$c=\{T : T\leq t \text{ or } T\geq n-t\}$ where $t=\frac{1}{2}(n+z_{\frac{\alpha}{2}}\sqrt{n})$	$min\{2P(y\leq T), 2P(y\geq T)\}$
$P(+)\geq P(-)$	$P(+)<P(-)$	$c=\{T : T\leq t\}$ where $t=\frac{1}{2}(n+z_\alpha\sqrt{n})$	$P(y\leq T)=P\big(z\leq\frac{2T-n+1}{\sqrt{n}}\big)$
$P(+)\leq P(-)$	$P(+)>P(-)$	$c=\{T : T\geq n-t\}$ where $t=\frac{1}{2}(n+z_\alpha\sqrt{n})$	$P(y\geq T)=1-P\big(z\leq\frac{2T-n+1}{\sqrt{n}}\big)$

24.8 3.5) McNemar Test for Significance of Changes

		y_i 0	1
X	0	a	b
	1	c	d

1.) $n = b + c \leq 20$
2.) $T_2 = b$

H_0	H_1	Critical Region	P-Value
$P(x_i = 0) = P(y_i = 0)$	$P(x_i = 0) \neq P(y_i = 0)$	$c = \{T_2 : T_2 \leq t_1 \text{ or } T_2 \geq n - t\}$	$min\{2P(y \leq T_2), 2P(y \geq T_2)\}$

3.) $n = b + c > 20$
4.) $T_1 = \frac{(b-c)^2}{b+c}$

H_0	H_1	Critical Region	P-Value
$P(x_i = 0) = P(y_i = 0)$	$P(x_i = 0) \neq P(y_i = 0)$	$c = \{T_1 : T_1 > \chi^2_{1-\alpha,1}\}$	$min\{2P(z < -\sqrt{T_1}), 2P(z > \sqrt{T_1})\}$

24.9 3.5) Cox and Stuart Test for Trend (like regression)

1.) Split data in half, and pair elements.
1a.) n = Even number of elements split evenly then pair:

$[1, 2, 3, 4, 5, 6, 7, 8] \rightarrow [1, 2, 3, 4], [5, 6, 7, 8] \rightarrow [(1, 5), (2, 6), (3, 7), (4, 8)]$

1b.) n = Odd number of elements, drop the median, then split and pair: $x_0.50 = \frac{n+1}{2}$

$[5, 6, 7, 8, 9, 10, 11, 12, 13] \rightarrow \frac{9+1}{2} = \text{ 5th element} = 9 \rightarrow [5, 6, 7, 8], \cancel{9}, [10, 11, 12, 13] \rightarrow [(5, 10), (6, 11), (7, 12), (8, 13)]$

2.) $T = \#$ of '+'s
3.) $n =$ total excluding $\#$ of ties

H_0	H_1	Critical Region	P-Value
$\beta_1 = 0$	$\beta_1 \neq 0$	$c = \{T : T \leq t \text{ or } T \geq n - t\}$ where $P(y \leq t) = \frac{\alpha}{2}$ $P(y \leq n - t) = 1 - \frac{\alpha}{2}$	$min\{2P(y \leq T), 2P(y \geq T)\}$
$\beta_1 \geq 0$	$\beta_1 < 0$	$c = \{T : T \leq t\}$ where $P(y \leq t) = \alpha$	$P(y \leq T)$
$\beta_1 \leq 0$	$\beta_1 > 0$	$c = \{T : T \geq n - t\}$ where $P(y \leq n - t) = 1 - \alpha$	$P(y \geq T)$

24.10 4.1) 2x2 Contingency Table

1.) Random Sample (rows), Random Results (columns)

	Class 1	Class 2	
Population 1.	O_{11}	O_{12}	n_1
Population 2.	O_{21}	O_{22}	n_2
	c_1	c_2	N

2.) Test Statistic

$$T_1 = \frac{\sqrt{N}(O_{11}O_{12} - O_{21}O_{22})}{\sqrt{n_1 n_2 c_1 c_2}}$$

H_0	H_1	Critical Region	P-Value
$p_1 = p_2$	$p_1 \neq p_2$	$c = \{T_1 : T_1 < z_{\frac{\alpha}{2}}$ or $T_1 > z_{1-\frac{\alpha}{2}}$	$min\{2P(z_{\frac{\alpha}{2}} < T_1), 2P(z_{1-\frac{\alpha}{2}} > T_1)\}$
$p_1 \geq p_2$	$p_1 < p_2$	$c = \{T : T < z_\alpha\}$	$P(z_\alpha < T_1)$
$p_1 \leq p_2$	$p_1 > p_2$	$c = \{T_1 : T_1 > z_{1-\alpha}\}$	$P(z_{1-\alpha} > T_1)$

24.11 4.1) Fisher's Exact Test

1.) Fixed Sample (rows), Fixed Results (columns)

	Class 1	Class 2	
r_1	x	$r - x$	r
r_2	$c - x$	$N - r - c + x$	$N - r$
	c	$N - c$	N

2.) $n \leq 20$
3.) Test Statistic

$$P(T_2 = x) = \begin{cases} P(T_2 \leq x) = \frac{\binom{r}{x}\binom{N-r}{c-x}}{\binom{N}{c}} & x = 0, 1, \ldots, min\{r_1, c\}) \\ 0 & \text{for all other values of x} \end{cases}$$

H_0	H_1	Critical Region	P-Value
$p_1 = p_2$	$p_1 \neq p_2$		$min\{2P(T_2 \leq x), 2P(T_2 \geq x)\}$
$p_1 \geq p_2$	$p_1 < p_2$		$P(T_2 \leq x)$
$p_1 \leq p_2$	$p_1 > p_2$		$P(T_2 \geq x)$

2.) $n > 20$
3.) Test Statistic

$$T_3 = \frac{x - \frac{rc}{N}}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}}$$

$$\textit{(P-value) } T_3 = \frac{x - \frac{rc}{N} \pm 0.5}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}}$$

H_0	H_1	Critical Region	P-Value
$p_1 = p_2$	$p_1 \neq p_2$	$c = \{T_3 : T_3 \leq z_{\frac{\alpha}{2}}$ or $T_3 > z_{1-\frac{\alpha}{2}}\}$	$min\{2P(z_{T_3 \leq \frac{\alpha}{2}}, 2P(z_{\frac{\alpha}{2}} \geq T_3)\}$
$p_1 \geq p_2$	$p_1 < p_2$	$c = \{T_3 : T_3 \leq z_\alpha\}$	$P(z_\alpha \leq T_3)$
$p_1 \leq p_2$	$p_1 > p_2$	$c = \{T_3 : T_3 \geq z_\alpha\}$	$P(z_{1-\alpha} \geq T_3)$

24.12 4.1) Mantel-Haenszel Test

	Class 1	Class 2	
Row 1	x_i	$r_i - x_i$	r_i
Row 2	$c_i - x_i$	$N_i - r_i - c_i + x_i$	$N_i - r_i$
	c_i	$N_i - c_i$	N_i

1.) Fixed Sample (rows), Fixed Results (columns)
2.) Test Statistic

$$T_4 = \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^2 (N_i - 1)}}}$$

H_0	H_1	Critical Region	P-Value
$p_{1i} = p_{2i}$	$p_{1i} > p_{2i}$ for some i or $p_{1i} < p_{2i}$ for some i (but not both)	$c = \{T_4 : T_4 < z_\alpha\}$ or $c = \{T_4 : T_4 > z_{1-\alpha}\}$	$min\{2P(z < (T_4 + 0.5)), 2P(z > (T_4 - 0.5))\}$
$p_{1i} \geq p_{2i}$	$p_{1i} < p_{2i}$	$c = \{T_4 : T_4 < z_\alpha\}$	$P(z_\alpha < (T_4 + 0.5))$
$p_{i1} \leq p_{i2}$	$p_{i1} > p_{i2}$	$c = \{T_4 : T_4 > z_\alpha\}$	$P(z_{1-\alpha} > (T_4 - 0.5))$

3.) Random Sample (rows), Random Results (columns)
4.) Test Statistic

$$T_5 = \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^3}}}$$

H_0	H_1	Critical Region	P-Value
$p_{1i} = p_{2i}$	$p_{1i} > p_{2i}$ for some i or $p_{1i} < p_{2i}$ for some i (but not both)	$c = \{T_5 : T_5 < z_\alpha\}$ or $c = \{T_5 : T_5 > z_{1-\alpha}\}$	$min\{2P(z_\alpha < T_5), 2P(z_{1-\alpha} > T_5)\}$
$p_{1i} \geq p_{2i}$	$p_{1i} < p_{2i}$	$c = \{T_5 : T_5 < z_\alpha\}$	$P(z_\alpha < T_5)$
$p_{i1} \leq p_{i2}$	$p_{i1} > p_{i2}$	$c = \{T_5 : T_5 > z_\alpha\}$	$P(z_{1-\alpha} > T_5)$

24.13 4.2) Chi-squared Test for Differences in Probabilities, rxc

	Class 1	Class 2	...	Class c	Totals
Population 1	O_{11}	O_{12}	...	O_{1c}	n_1
Population 2	O_{21}	O_{22}	...	O_{2c}	n_2
...	...	...	...	...	...
Population r	O_{r1}	O_{r2}	...	O_{rc}	n_{rc}
Totals	C_1	C_2	...	C_C	N

1.) Fixed Sample (rows), Fixed Results (columns)
2.) Test Statistic

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ where } E_{ij} = \frac{n_i C_j}{N}$$