STAT 571A — Advanced Statistical Regression Analysis

Appendix A NOTES

Review of Probability and Statistics

Fair Warning

- The material presented from Appendix A is meant <u>completely as a review</u> to establish notation and to act as a refresher.
- Students should have learned this Appendix's material previously and be immediately familiar with it. If not, a previous course in statistics or matrix algebra is needed before undertaking STAT 571A.

§A.1: Sums and Products

- Observations: $Y_1, Y_2, \ldots, Y_n$
- Summation: $Y_1+Y_2+\cdots+Y_n = \sum_{i=1}^{n} Y_i$

  Consequences:
  - $\sum_{i=1}^{n} k = nk$
  - $\sum_{i=1}^{n} (Y_i + Z_i) = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i$
  - $\sum_{i=1}^{n} (k + cY_i) = nk + c\sum_{i=1}^{n} Y_i$

Sums and Products (cont'd)

Double sum: $\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}$

$$= \sum_{i=1}^{n}(Y_{i1}+Y_{i2} \cdots +Y_{im})$$

$$= (Y_{11}+Y_{12}+ \cdots +Y_{1m})$$

$$+ \cdots + (Y_{n1}+Y_{n2}+ \cdots +Y_{nm})$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{n} Y_{ij}$$

Product: $(Y_1)(Y_2) \cdots (Y_n) = \prod_{i=1}^{n} Y_i$

§A.2: Probability Rules

- **Events: $A_1, A_2, \ldots, A_n$**
- **Probability of union:**
  $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$
- **Multiplication rule:**
  $P(A_i \cap A_j) = P(A_i)P(A_j|A_i) = P(A_j)P(A_i|A_j)$

$P(A|B)$ is a *conditional probability*

Probability Rules (cont'd)

Complementary event: $\overline{A}_j = \{\text{not } A_j\}$

Complement rule: $P(\overline{A}_j) = 1 - P(A_j)$

So, e.g., $P(\overline{A_i \cup A_j}) = P(\overline{A}_i \cap \overline{A}_j)$

§A.3: Random Variables

- A **random variable** is a numerical outcome of some random process.
- Notation: upper-case Latin letter, say, Y.
- If Y takes on discretely many values it is a **Discrete Random Variable**.
- If Y lies in a continuum of values, it is a **Continuous Random Variable**.

Probability Functions

- The **Probability Function**, $f_Y(y)$, of Y gives the mass or density of probability for Y. For instance, in the discrete case write

  $f_Y(y_s) = P(Y = y_s)$ over $s = 1,\ldots,k$.
- If so, we write $Y \sim f_Y(y)$.
- The tilde (~) is read "is distributed as".

Expectation

- The **Expected Value** of any function of Y, g(Y), is E[g(Y)] $= \sum_{s=1}^{k} g(y_s)\, f_Y(y_s)$ (discrete case)

  $= \int_{-\infty}^{\infty} g(y)\, f_Y(y)dy$ (contin. case)

- The expectation operator, E[·], satisfies
  - E[a] = a (for constant a)
  - E[aY] = aE[Y]
  - E[a + cY] = a + cE[Y]

- The **Population Mean** of Y is μ_Y = E[Y].

Variance

- The **Population Variance** of Y is

$$\sigma^2[Y] = E[(Y - \mu_Y)^2]$$
$$= E[Y^2] - E^2[Y] = E[Y^2] - \mu_Y^2.$$

- Thus,
  - $\sigma^2[c] = 0$ (for constant c)
  - $\sigma^2[cY] = c^2\sigma^2[Y]$
  - $\sigma^2[c + Y] = \sigma^2[Y]$
  - $\sigma^2[c + dY] = d^2\sigma^2[Y]$

- The **Popl'n Standard Deviation** is $\sigma[Y]$.

Mean Squared Error

- Suppose we estimate a parameter ω with a statistic W. The mean of W is E[W] and the variance of W is $\sigma^2\{W\} = E\{(W - E[W])^2\}$.
- We define the **Mean Squared Error** of W as $MSE\{W\} = E\{(W - \omega)^2\}$.
- Notice that this is

$$
\begin{aligned}
MSE\{W\} &= E\{(W - E[W] + E[W] - \omega)^2\} \\
&= E\{[(W - E[W]) + (E[W] - \omega)]^2\} \\
&= E\{(W - E[W])^2\} + E\{(E[W] - \omega)^2\} \\
&\qquad + 2E\{(W - E[W])(E[W] - \omega)\}
\end{aligned}
$$

Mean Squared Error (cont'd)

- But now
  - $E\{(W - E[W])^2\}$ is just $\sigma^2\{W\}$
  - $E[W] - \omega$ has no stochastic features, so $E\{(E[W] - \omega)^2\} = (E[W] - \omega)^2 = \text{Bias}^2\{W\}$
  - And then, $E\{(W - E[W])(E[W] - \omega)\}$

$$
\begin{aligned}
&= (E[W] - \omega)\, E\{W - E[W]\} \\
&= (E[W] - \omega)\, (E\{W\} - E[W]) \\
&= (E[W] - \omega)\, (0) \;\; = 0
\end{aligned}
$$

- So we find $\text{MSE}\{W\} = \sigma^2\{W\} + \text{Bias}^2\{W\}$, i.e., MSE = Variance + Squared Bias.

Joint Probability and Covariance

- The **Joint Probability Function** of U and V is
  $f_{U,V}(u_s, v_t) = P(\ U = u_s \cap V = v_t\ )$ (discrete case)
- The **Covariance** of U and V is
  $$\sigma[U,V] = E\{(U - E[U])(V - E[V])\}$$
  $$= E[UV] - E[U]E[V] = E[UV] - \mu_U \mu_V$$
- The **Correlation** between U and V is
  $$\rho[U,V] = \sigma[U,V]/\{\sigma[U]\sigma[V]\}$$
  $$= \sigma\{\ (U - \mu_U)/\sigma[U]\ ,\ (V - \mu_V)/\sigma[V]\ \}$$
  where $-1 \le \rho[U,V] \le 1$.

Covariance (cont'd)

- Notice that if

  $$\sigma[U,V] = E[UV] - \mu_U\mu_V$$

  then:

  - $\sigma[a_1 + c_1U\ ,\ a_2 + c_2V] = c_1c_2\sigma[U,V]$
  - $\sigma[a_1\ ,\ a_2 + c_2V] = 0$
  - $\sigma[a_1 + U\ ,\ a_2 + V] = \sigma[U,V]$
  - $\sigma[U,U] = \sigma^2[U]$

- Also, if $\sigma[U,V] = 0$, then $\rho[U,V] = 0$.

Independence

- Two random variables U and V are **independent** if their joint prob. function factors:

$$f_{U,V}(u_s,v_t) = f_U(u_s)\, f_V(v_t) \quad \text{for all } u_s, v_t$$

- Then we can show

$$\sigma[U,V] = \rho[U,V] = 0,$$

but not vice versa (except in very special cases).

Sums of Random Variables

- If $Y_i \sim f_{Y_i}(y)$ for $i = 1,\ldots,n$, then
  - $E[\, \sum a_i Y_i \,] = \sum a_i E[Y_i]$
  - $\sigma^2[\, \sum a_i Y_i \,] = \sum_i \sum_j a_i\, a_j\, \sigma[Y_i\,,\, Y_j]$
- So, e.g., if n = 2:
  - $E[a_1 Y_1 + a_2 Y_2] = a_1 E[Y_1] + a_2 E[Y_2]$
  - $\sigma^2[a_1 Y_1 + a_2 Y_2] = a_1^2 \sigma^2[Y_1]$
    $+ a_2^2 \sigma^2[Y_2] + 2 a_1 a_2 \sigma[Y_1, Y_2]$

Sums of Variables (cont'd)

- <u>If</u> Y_1 and Y_2 are independent, then

$$\sigma[Y_1,Y_2] = 0,$$

so

$$\sigma^2[a_1Y_1 + a_2Y_2] = a_1^2\sigma^2[Y_1] + a_2^2\sigma^2[Y_2].$$

- More generally, <u>if</u> the Y_i's are (mutually) independent

$$\sigma^2[\sum a_iY_i] = \sum a_i^2\sigma^2[Y_i]$$

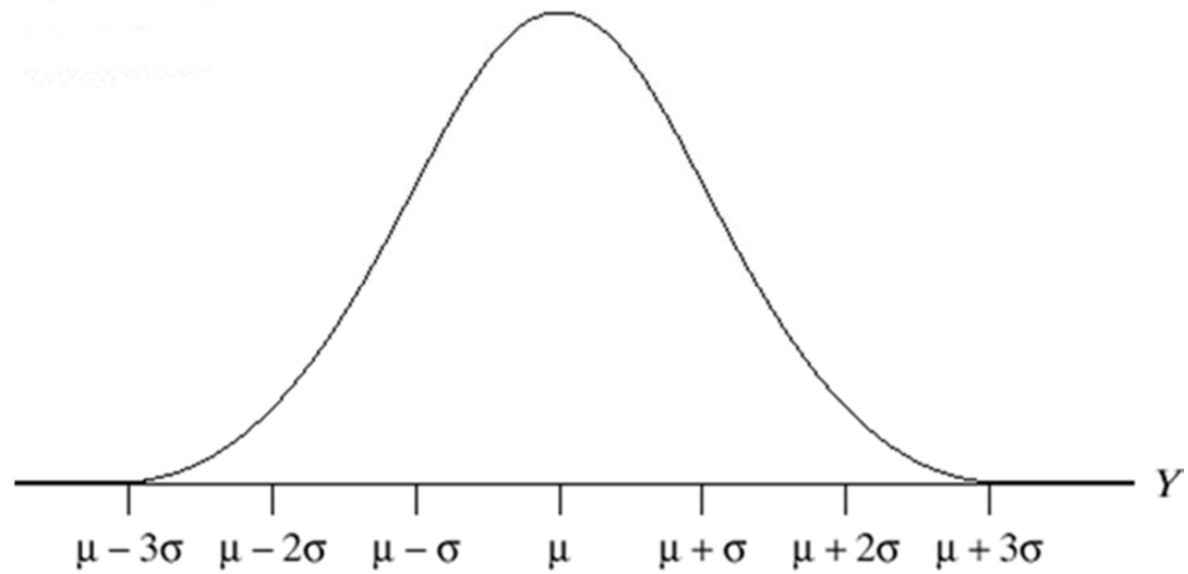
§A.4: Normal Distribution

- The **Normal (Probability) Distribution** has prob. function

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma^2}\right\}$$

where "exp" is the base of the natural logarithm.

- This has a (famous) "bell shape"



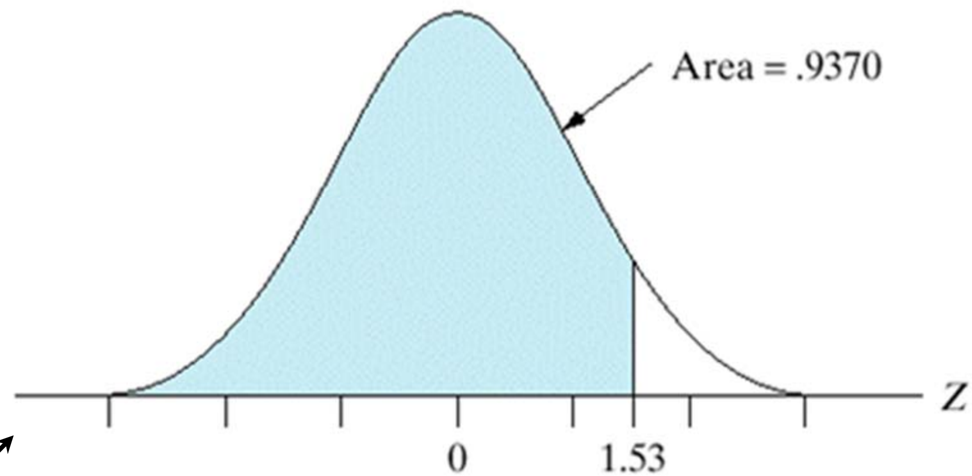
Normal Dist'n (cont'd)

- The normal is also called the "Gaussian" distribution.
- Notation: $Y \sim N(\mu,\sigma^2)$
- Here, $\mu = E[Y]$ and $\sigma^2 = \sigma^2[Y]$
- Can show: $a + cY \sim N(a + c\mu, c^2\sigma^2)$, and in particular, $Z = (Y - \mu)/\sigma \sim N(0,1)$.
- Z then has a **Standard Normal Distribution**

Normal Dist'n (cont'd)

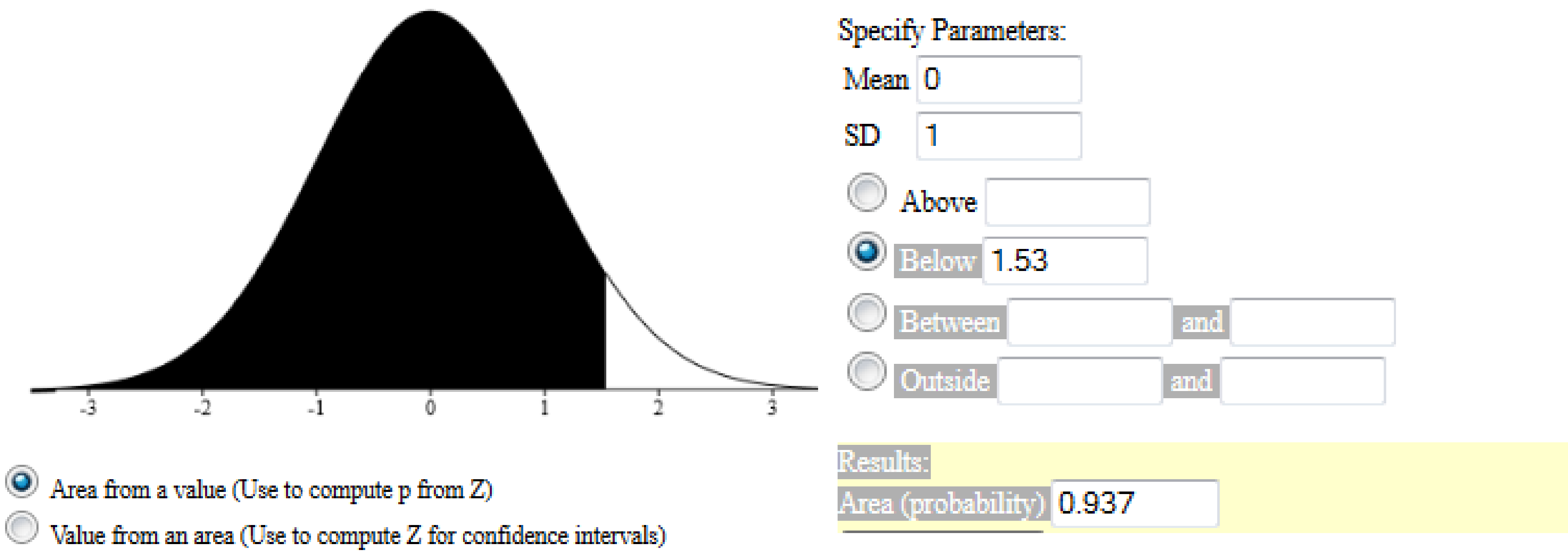
- If $Z \sim N(0,1)$ we call $\Phi(z) = P(Z \le z)$ the <u>cumulative</u> <u>distribution</u> <u>function</u> of Z. See Appendix Table B.1

- $P(Z \le z)$ has interpretation as the area under the normal prob. function. For instance, $P(Z \le 1.53) = 0.937$



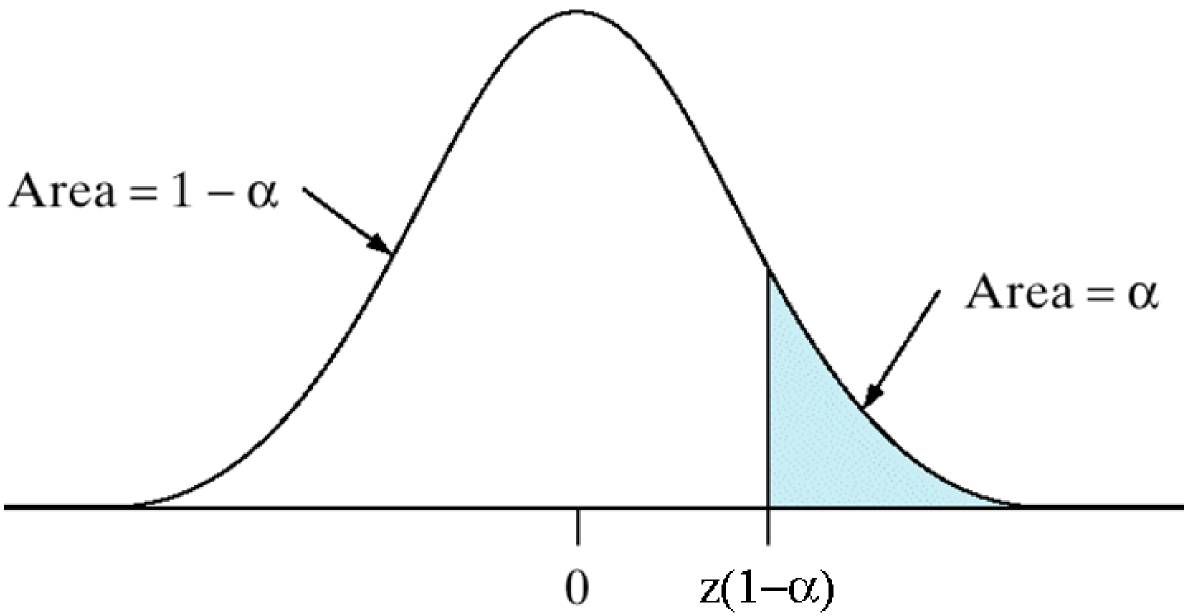
Normal Dist'n (cont'd)

A useful online app for visualizing the std. normal is at http://davidmlane.com/normal.html



Normal Critical Points

- We can reverse the process and ask, what value of $z(1-\alpha)$ gives $P[Z > z(1-\alpha)] = \alpha$:



- This is called the **upper-α critical point** of Z.
- Notice, by symmetry, that $-z(\alpha) = z(1-\alpha)$.

Chi-square Distribution

- If Y_i ~ indep. $N(\mu_i,\sigma_i^2)$ for i = 1,…,n, then

  $$\sum a_i Y_i \sim N\left(\sum a_i \mu_i \,,\, \sum a_i^2 \sigma_i^2 \right) \qquad \text{(A.40)}$$

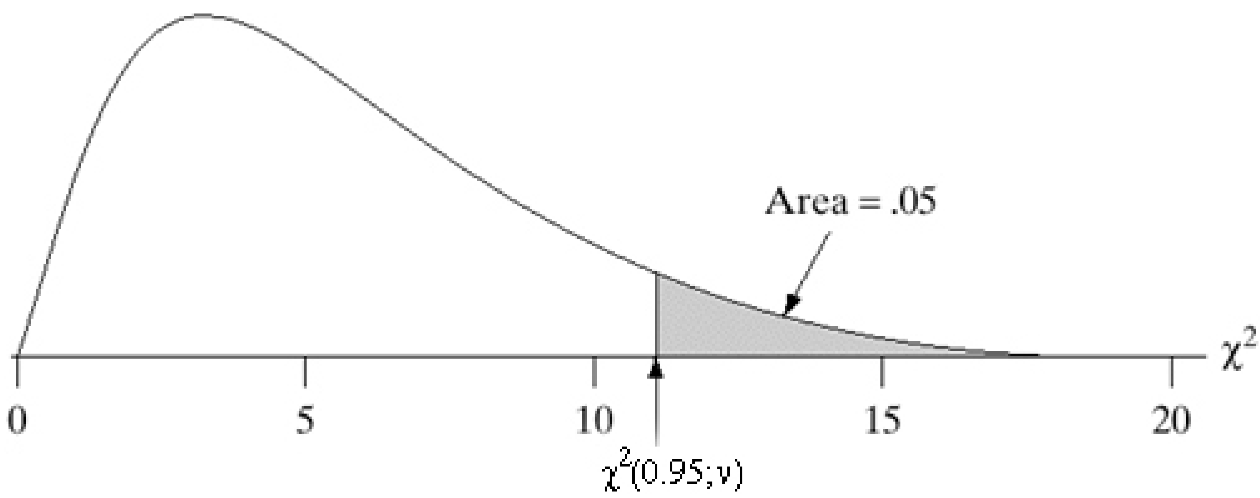
- Now, suppose Z_i ~ indep.N(0,1) i = 1,…,ν. Then $U = \sum_{i=1}^{\nu} Z_i^2$ has a special form:

  $$U \sim \chi^2(\nu)$$

- We say U is distributed as "chi-square" with ν degrees of freedom (d.f.).
- Can show: $E[U] = \nu$ and $\sigma^2[U] = 2\nu$.

Chi-square Critical Points

- The upper-α critical point of $U \sim \chi^2(\nu)$ is $\chi^2(1-\alpha;\nu)$ such that $P[\, U > \chi^2(1-\alpha;\nu)\,] = \alpha$:



- Find these in Appendix Table B.3.

t-distribution

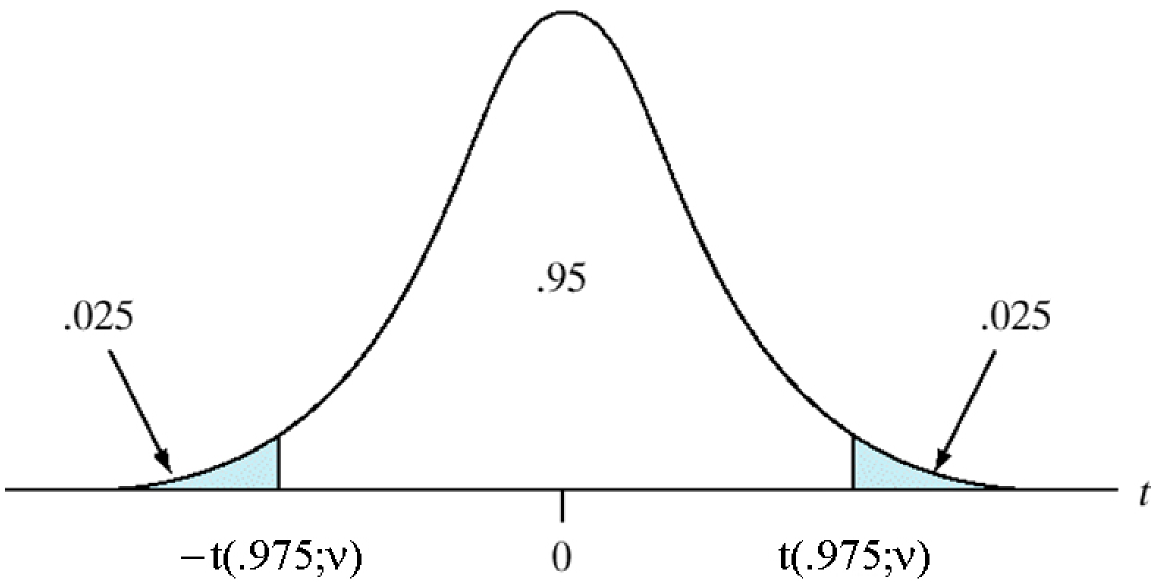
- Now, suppose $Z \sim N(0,1)$ is indep. of $U \sim \chi^2(\nu)$. Let

$$T = \frac{Z}{\sqrt{U/\nu}}$$

- Then we say T is distributed as "Student's t" with ν degrees of freedom: $T \sim t(\nu)$.
- Can show: $E[T] = 0$ and $\sigma^2[T] = \nu/(\nu - 2)$.

t Critical Points

- The upper-α critical point of $T \sim t(\nu)$ is $t(1-\alpha;\nu)$ such that $P[\ T > t(1-\alpha;\nu)\ ] = \alpha$:



- By symmetry, $-t(\alpha;\nu) = t(1-\alpha;\nu)$; e.g., $-t(.975;\nu) = t(.025;\nu)$
- Find these in Appendix Table B.2.

F-distribution

- Now, suppose U_i ~ indep. $\chi^2(\nu_i)$, i = 1,2. Let

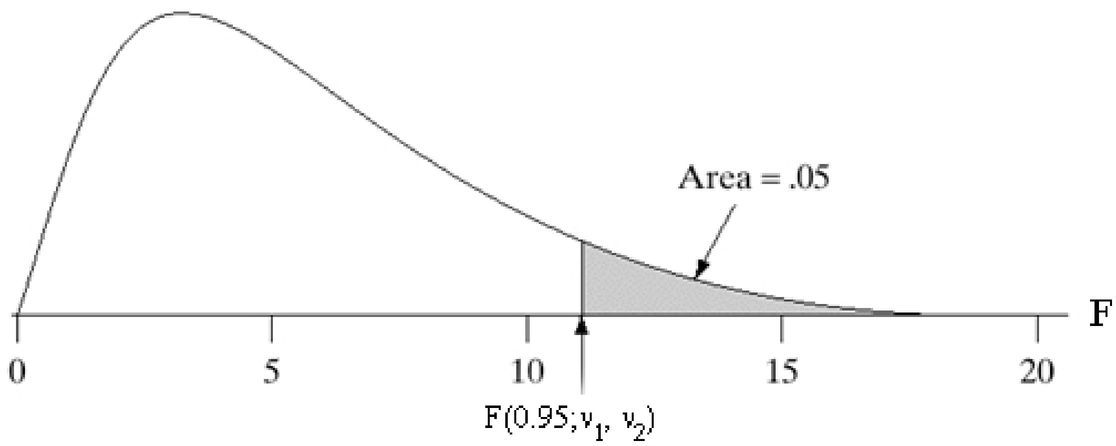
$$F = \frac{U_1/\nu_1}{U_2/\nu_2}$$

- Then we say F is distributed as, well, 'F' with ν_1 numerator d.f. and ν_2 denominator d.f. (order *is* important): $F \sim F(\nu_1,\nu_2)$.
- Can show (sorta' obvious): $1/F \sim F(\nu_2,\nu_1)$

F Critical Points

- The upper-α critical point of $F \sim F(\nu_1,\nu_2)$ is $F(1-\alpha;\ \nu_1,\nu_2)$ such that

$$P[\ F > F(1-\alpha;\ \nu_1,\nu_2)\ ] = \alpha:$$



- Find (some of) these in Appendix Table B.4.

Central Limit Theorem

- The **Central Limit Theorem** states that if $Y_i \sim \text{indep.}(\mu,\sigma^2)$ for $i = 1,\ldots,n$, then

$$\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu}{\sigma/\sqrt{n}} \mathrel{\dot{\sim}} N(0,1)$$

where the $\cdot$ over the $\sim$ reads "is approximately distributed as."

- The approximation improves as $n \to \infty$.

§A.5: Statistical Estimation

- Suppose some parameter of a prob. function $f_Y(y)$, say, θ, is <u>unknown</u>.
- A **statistical estimator** of θ is generically denoted by $\hat{\theta}$
- $\hat{\theta}$ is **unbiased** for θ if $E[\hat{\theta}] = \theta$

Statistical Estimation (cont'd)

- To find an estimator of θ we can employ the **Method of Least Squares** (LS).
- Given $Y_i \sim$ indep. $f_{Y_i}(y)$ for $i = 1,\ldots,n$, with $E[Y_i] = \theta$. The LS estimator of θ minimizes the objective quantity

$$Q = \sum (Y_i - \theta)^2$$

- We can model θ as a function of other parameters to expand the setting.

§A.6: Inference

- **Normal sampling:** $Y_i \sim \text{i.i.d.} N(\mu,\sigma^2)$ for $i = 1,\ldots,n$, where "i.i.d." stands for "independent and identically distributed."
- **The sample mean is** $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$
- **The sample variance is** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2\right\}$$

- **The sample std. deviation is** $S = \sqrt{S^2}$

t-Statistic

- The **standard error (of the mean)** is

$$s(\bar{Y}) = S/\sqrt{n}$$

- This is used in the t-statistic

$$t = (\bar{Y} - \mu)/s(\bar{Y}) = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

- Here, $t \sim t(n-1)$.

Interval Estimates

- An **Interval Estimate** for μ is based on the t-statistic, and its reference t-dist'n:

$$\bar{Y} - t\left(1 - \tfrac{\alpha}{2};\ n-1\right)\frac{S}{\sqrt{n}} < \mu < \bar{Y} + t\left(1 - \tfrac{\alpha}{2};\ n-1\right)\frac{S}{\sqrt{n}}$$

or simply

$$\bar{Y} \ \pm\ t\left(1 - \tfrac{\alpha}{2};\ n-1\right)\frac{S}{\sqrt{n}}$$

- This is called a **1–α Confidence Interval** for μ. (Note: confidence is **not** probability!)

Statistical Inferences

- Confidence intervals are forms of statistical inference, where a statement about a population parameter is constructed using probability arguments.
- Another form of such inference is called hypothesis testing, where hypotheses about an unknown parameter are tested $\longrightarrow$

Hypothesis Test for μ

Null hypoth.	Altern. hypoth.	Rejection Region
$H_o: \mu = \mu_o$	$H_a: \mu \neq \mu_o$	$\lvert t^* \rvert > t\left(1 - \frac{\alpha}{2}; n-1\right)$
$H_o: \mu = \mu_o$	$H_a: \mu < \mu_o$	$t^* < -t(1-\alpha; n-1)$
$H_o: \mu = \mu_o$	$H_a: \mu > \mu_o$	$t^* > t(1-\alpha; n-1)$

where the **test statistic** is $t^* = \dfrac{\overline{Y} - \mu_o}{S/\sqrt{n}}$

P-values

- The P-value from an hypoth. test is the probability of recovering a test statistic as extreme or more extreme than t* under H_o.

- "More extreme" is defined in the direction of H_a:

$H_o: \mu = \mu_o \quad H_a: \mu \neq \mu_o \qquad P = 2P[t(n-1) > |t^*|]$

$H_o: \mu = \mu_o \quad H_a: \mu < \mu_o \qquad P = P[t(n-1) < t^*]$

$H_o: \mu = \mu_o \quad H_a: \mu > \mu_o \qquad P = P[t(n-1) > t^*]$

Significance and Error Rates

- The quantity α here is the **significance level** of the test ($0 < \alpha < 1$).
  Can relate this to the P-value: always reject H_o in favor of H_a when $P < \alpha$.
- Interpretation is based on **error rates**:
  - α = P[reject H_o | H_o true] = P[false positive error]
  - β = P[accept H_o | H_o false] = P[false neg. error]
- We say $1–\beta$ is the **power** of the test (see §2.3).

Error Rates

- Older terminology for a false positive error is a **Type I error**,

  while that for a false negative error is a **Type II error**.

- Can think of it this way:

		True Situation	
		H_0 *true*	H_0 *false*
Our Decision	Do not reject H_0	Correct	Type II error
	Reject H_0	Type I error	Correct

Default is Two-Sided

- In any hypothesis testing scenario, the decision to chose a one-sided vs. a two-sided alternative hypothesis MUST be made *prior* to sampling the data.
- If the subject-matter cannot guide this decision then use a two-sided alternative hypothesis, by default.

Tautology

- There is a tautology between confidence intervals and hypothesis tests: they are two forms of the same inference!
  - For the "two-sided" case with H_o: $\mu = \mu_o$ vs. H_a: $\mu \neq \mu_o$, we reject H_o at signif. level α if and only if μ_o is not contained in the $1 - \alpha$ confidence interval

$$\bar{Y} \pm t\left(1-\tfrac{\alpha}{2};\, n-1\right)\frac{S}{\sqrt{n}}$$

Tautology (cont'd)

- Similarly, for the "one-sided" case with H_o: $\mu = \mu_o$ vs. H_a: $\mu > \mu_o$, reject H_o at signif. level α if and only if μ_o exceeds the (one-sided) $1 - \alpha$ confidence bound

$$\bar{Y} + t(1-\alpha;\, n-1)S/\sqrt{n}$$

- For H_o: $\mu = \mu_o$ vs. H_a: $\mu < \mu_o$, reject H_o at signif. level α if and only if μ_o lies below the (one-sided) $1 - \alpha$ confidence bound

$$\bar{Y} - t(1-\alpha;\, n-1)S/\sqrt{n}$$

§A.7: Two-Sample Inference

- $Y_i \sim$ i.i.d.$N(\mu_1,\sigma^2)$ for $i = 1,\ldots,n_1$, indep. of $U_j \sim$ i.i.d.$N(\mu_2,\sigma^2)$ for $j = 1,\ldots,n_2$.
- Find sample means $\bar{Y}$ and $\bar{U}$, and pooled sample variance

$$S^2_{pool} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- The pooled variance estimates the common σ^2.

Two-Sample Inference (cont'd)

Then, find the test statistic

$$T_{12} = (\bar{Y} - \bar{U})/s\{\bar{Y} - \bar{U}\}$$

where $s\{\bar{Y} - \bar{U}\} = S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ is the std. error of $\bar{Y} - \bar{U}$.

A 1–α conf. int. for the difference $\mu_1 - \mu_2$ is then

$$(\bar{Y} - \bar{U}) \pm t(1 - \tfrac{\alpha}{2};\ n_1+n_2-2)s\{\bar{Y} - \bar{U}\}$$

Hypothesis Test for $\mu_1 - \mu_2$

Null hypoth.	Altern. hypoth.	Rejection Region
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 \neq \mu_2$	$\lvert t^* \rvert > t\left(1 - \frac{\alpha}{2}; df\right)$
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 < \mu_2$	$t^* < -t(1-\alpha; df)$
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 > \mu_2$	$t^* > t(1-\alpha; df)$

where $df = n_1 + n_2 - 2$ and the **test statistic** is

$$t^* = (\bar{Y} - \bar{U})/s\{\bar{Y} - \bar{U}\}$$

P-values for $\mu_1 - \mu_2$

Null hypoth.	Altern. hypoth.	P-value
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 \neq \mu_2$	$P = 2P[t(df) > \lvert t^* \rvert]$
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 < \mu_2$	$P = P[t(df) < t^*]$
$H_o: \mu_1 = \mu_2$	$H_a: \mu_1 > \mu_2$	$P = P[t(df) > t^*]$

where $df = n_1 + n_2 - 2$ and the test statistic is

$$t^* = (\bar{Y} - \bar{U})/s\{\bar{Y} - \bar{U}\}$$

Unequal Variances

- If $Y_i \sim$ i.i.d.$N(\mu_1,\sigma_1^2)$ for $i = 1,\ldots,n_1$, indep. of $U_j \sim$ i.i.d.$N(\mu_2,\sigma_2^2)$ for $j = 1,\ldots,n_2$, the variances are heterogeneous. Do NOT use the pooled variance estimator.
- Instead, apply the "Welch-Satterthwaite correction" which uses the individual samples variances and adjusts the t-dist'n d.f. (See your intro. stat. textbook.)

§A.8: Inferences on σ^2

- Let $Y_i \sim \text{i.i.d.}N(\mu,\sigma^2)$ for $i = 1,\ldots,n$.
- Estimate σ^2 with the sample variance S^2.
- In fact, $E[S^2] = \sigma^2$ (unbiased!)
- Also, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. So, a $1-\alpha$ conf. int. for σ^2 is

$$\frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2};\, n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2};\, n-1)}$$

- (But, it's not optimal...)

Hypothesis Tests for σ^2

Null hypoth.	Altern. hypoth.	Rejection Region
$H_o: \sigma = \sigma_o$	$H_a: \sigma \neq \sigma_o$	$X^{2*} > \chi^2(1-\frac{\alpha}{2}; n-1)$ or $X^{2*} < \chi^2(\frac{\alpha}{2}; n-1)$
$H_o: \sigma = \sigma_o$	$H_a: \sigma < \sigma_o$	$X^{2*} < \chi^2(\alpha; n-1)$
$H_o: \sigma = \sigma_o$	$H_a: \sigma > \sigma_o$	$X^{2*} > \chi^2(1-\alpha; n-1)$

where the **test statistic** is $X^{2*} = \dfrac{(n-1)S^2}{\sigma_o^2}$

§A.9: Two Variances

- We can also extend inferences on variances to the two-sample case, to find a confidence interval on the ratio σ_1^2/σ_2^2 or to test hypotheses such as H_o: $\sigma_1^2 = \sigma_2^2$.
- The reference dist'n becomes $F(n_1–1, n_2–1)$. See Appendix A.9 for details.