Thought: *Never do card tricks for the people you play poker with.*

Thought: *To succeed in politics, it is often necessary to rise above your principles.*

Monday 3/25/02 : OPTIONAL REVIEW, ask questions about homework, course material, sample exam, etc. ***Suggestion - print out Sample Exam 2 and bring it to class with you***.

Help for Exam 2

- Monday, periods 3–8, FLO 104
- Monday, 6:00 pm – 10:00 pm, TUR L005

Tuesday 3/26/02 : EXAM 2–during your regularly scheduled discussion section.

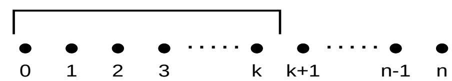
Wednesday 2/13/02 : Pages 334 – 338, 347–351

Chapter 4 : Discrete Random Variables

- Can COUNT the number of distinct values of the variable.
- The Binomial Probability Distribution
  - Some discrete random variables are binomial – NOT ALL
  - Binomial Experiment p. 179 – 5 criteria
  - If $n =$ number of trials, $P(S) = p$: (p. 183)

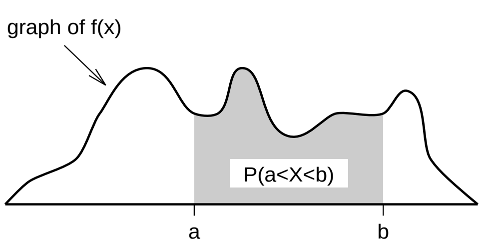
$$p(x) = \binom{n}{x} p^x \, q^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

  - Mean (p. 185) : $\mu = np$.
  - Variance (p. 185) : $\sigma^2 = npq$.
  - Tables : Contain $P(x \le k)$ for $k = 0, 1, \cdots, n$.



Chapter 5 : Continuous Random Variables

- Possible values are all those associated with one or more line intervals.
- Probabilities are areas under "density function".



- If x is a continuous r.v.,

$$P(a \le x \le b) = P(a < x \le b) = P(a \le x < b) = P(a < x < b)$$

- Normal distribution is special case.

- Areas under normal curves between ***z-scores*** of 0 and z, for $z > 0$ in Table IV, p. 809.
- Key to finding correct areas (probabilities) : ***draw pictures***

Chapter 6: If we plan to take a random sample of size n from a **population** with mean μ and standard deviation σ,

- $\bar{x}$ is a ***random variable.***
- Dist. of $\bar{x}$ is called its ***sampling distribution.*** (p. 255)
- $\mu_{\bar{x}} = E(\bar{x}) = \mu$. So $\bar{x}$ is an ***unbiased*** estimator of μ. (p. 266, 261)
- $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ (p. 266), so dist. of $\bar{x}$ is more concentrated around μ for larger sample sizes. $\sigma_{\bar{x}}$ is called the ***standard error*** of $\bar{x}$.(p. 266)
- If the population has a normal dist., then so does $\bar{x}$, i.e., $\bar{x} \sim$ N$(\mu, \sigma/\sqrt{n})$. True for any n.

- **Central Limit Theorem (CLT):** (p. 267) If n is large ($n \geq 30$), then the sampling distribution of $\bar{x}$ is approximately normal, i.e., $\bar{x} \mathrel{\dot\sim} \text{N}(\mu, \sigma/\sqrt{n})$, regardless of the shape of the population distribution.

Chapter 6:

- A ***Parameter*** is a meaningful number associated with a Population. μ, σ^2, etc. (p. 254)
- A ***Statistic*** is a meaningful number associated with a Sample.
- All statistics have ***sampling distributions***

Chapter 7:

- Point Estimator (p. 261)
- Interval Estimator (p. 282)
- Confidence Coefficient (p. 282)
- Confidence Level (p. 282)

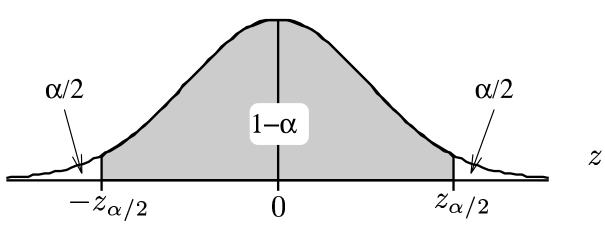
Parameter	Estimator	Standard Error of Est.	Estimated Standard Error of Est.
μ	$\overline{x}$	$\dfrac{\sigma}{\sqrt{n}}$	$\dfrac{s}{\sqrt{n}}$
p	$\hat{p} = \dfrac{x}{n}$	$\sqrt{\dfrac{pq}{n}}$	$\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

- Both estimators are UNBIASED
- If n is "large", both estimators are approximately NORMALLY distributed.
- How large is "large"?
  - For valid CI for μ : $n \geq 30$.
  - For valid CI for p :

$$n \geq 9 \left(\frac{\text{larger of } (\hat{p},\ \hat{q})}{\text{smaller of } (\hat{p},\ \hat{q})} \right)$$

- $(1-\alpha) \times 100\%$ Confidence Interval for a PARAMETER

$$\underbrace{\text{estimator}}_{\text{formula sheet}} \pm \underbrace{z_{\alpha/2}}_{\text{table}} \underbrace{\left(\text{standard errors}\right)}_{\text{formula sheet}}.$$



- Population mean, μ (P. 283)

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \approx \bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

- Population Proportion, p (P. 300)

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

Finding the sample size to estimate μ.

- Want : Correct to within "B" units with $(1-\alpha)100\%$ confidence.
- $z_{\alpha/2} \times \left(\text{standard error}\right) = B$ and SOLVE
- $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = B$ and solve for n (p. 307)
  - Use ballpark value for σ if you have one. Maybe use $\sigma \approx \dfrac{\text{Range}}{4}$.

Finding the sample size to estimate p.

- Want : Correct to within "B" units with $(1-\alpha)100\%$ confidence.
- $z_{\alpha/2} \times \left(\text{standard error}\right) = B$ and SOLVE
- $z_{\alpha/2}\sqrt{\dfrac{pq}{n}} = B$ and solve for n (p. 333)
  - Use "ballpark" value for p if you have one, if not use $p = q = .5$ to get sample size that will work for ***any*** value of p.

Chapter 8 – Large Sample Hyp. Testing

Parts of a statistical test. (p. 322)

- The hypothesis of ***MAIN INTEREST*** is the ***ALTERNATIVE*** or ***RESEARCH*** hypothesis, – light bulb ex. H_a, $(H_a : \mu > 1325)$. (p. 322) What we are "trying to prove" in an objective, fair manner
- The ***"other"*** hypothesis is called the ***NULL HYPOTHESIS***, – H_o, light bulb ex. $(H_o : \mu = 1325)$ (p. 322)

Errors: p. 325

Decision	**Reality** H_o *true*	H_a *true*
Accept H_o	Correct	Type II error
Reject H_o	Type I error	Correct

- $\alpha = P($ Type I error $)$ (p. 323), ***SIGNIFICANCE LEVEL*** of the test.
- $\beta = P($ Type II error $)$ (p. 325)
- $\alpha \uparrow$, $\beta \downarrow$ and/or $\alpha \downarrow$, $\beta \uparrow$
- $\alpha = P\{$accepting H_a when H_o true $\}$
- $\alpha =$ $P\{$saying what we "want" to say when we should not$\}$
- In our lightbulb example, $\alpha = P\{$ saying $\mu > 1325$ when $\mu = 1325\}$

Thought: *When a man is wrapped up in himself, he makes a pretty small package – (John Ruskin)*

Assignments

Today: P. 334–338, 347–351

Tuesday : Exercises 8.29, 8.33, 8.34, 8.38–41, 8.59, 8.61, 8.67–69

Wednesday: P. 288 – 294 (Sec. 7.2), P. 341 – 345 (Sec. 8.4)

Thursday: Exer. 7.27, 7.30, 7.33, 8.49, 8.50, 8.53, 8.55 – 8.57

Last Time: Large Sample Hypothesis Testing about μ

- $\mu =$ UNKNOWN population mean
- $H_o : \mu = \mu_o$

$$H_a = \left\{ \begin{array}{lll} \mu > \mu_o & I & z > z_\alpha \\ \quad \text{OR} & & \\ \mu < \mu_o & II & z < -z_\alpha \end{array} \right\} \textbf{RR}$$

- Test Statistic

$$z = \frac{\overline{x} - \mu_o}{\sigma/\sqrt{n}} = \frac{\text{estimator}^* - \text{hypothesized value}^{**}}{\text{standard error}^*}$$

$*$ Estimator and Standard Error from Formula Sheet

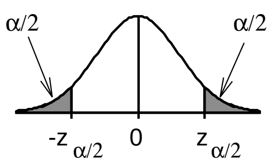
$**$ Hypothesized Value from NULL hypothesis

Ex. : pH of 7 is neutral, over 7 is alkaline, under 7 indicates acidity. Randomly select 30 water specimens from a recreational lake. Can we claim that the mean pH is NOT that of neutral water at the $\alpha = .01$ level?

- H_o and H_a?
  H_a : H_o :

How?

$$H_o : \mu = \mu_o$$
$$H_a : \mu \neq \mu_o$$
$$TS : z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
$$RR : z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2}$$



"Two Tailed test" (p. 331)

Back to pH example:

- $n = 30$ $\bar{x} = 7.3$ $s = 0.2.$
- $\alpha = .01$ level test:

  RR :

$$z = \underline{\qquad\qquad} =$$

- Reject H_o at the $.01$ level of significance. In terms of this application: "There   enough evidence at the $.01$ level to indicate that the mean pH reading is not 7."

If the mean pH is NOT 7, what is it?

Construct a 99% confidence interval for μ.

$$1-\alpha = .99, \ \alpha = \qquad , \ \frac{\alpha}{2} = \qquad , \ z_{.005} =$$

$$\bar{x} \pm \qquad \frac{s}{\sqrt{n}}$$

$$7.3 \pm .094 :$$

Do you think that $\mu = 7$?

- — the value "7" is   the 99% confidence interval
- Agrees with two-tailed test!!

Hypothesis Testing

- α is chosen BEFORE the test is performed
- Smaller α
  - to reject H_o.
  - Provides   CONFIDENCE in our decision to reject H_o in favor of H_a (IF we DO SO).

What is the SMALLEST value of α for which H_o could be rejected in favor of H_a?

- The ***p-value*** or ***observed significance level*** (P. 335)

Recall the lightbulb example

$$H_o : \mu = 1325$$
$$H_a : \mu > 1325$$
$$z = 1.95$$

α	rejection region
.05	$z > 1.645$
.04	$z >$
.03	$z >$
.02	$z > 2.050$
.01	$z > 2.330$

$H_o : \mu = 1325$

$H_a : \mu > 1325$

p-value $=$ Probability of a z-value

the one observed

- Larger z-values are indicative that H_a is true.
- ***p-value*** = $P(z > \quad)$

  $=$

- $\alpha \geq$ p-value $\Rightarrow$ REJECT H_o.
- $\alpha <$ p-value $\Rightarrow$ CANNOT reject H_o.
- In our case, p-value = .0256
  - $\alpha = .1 \Rightarrow$ H_o.
  - $\alpha = .05 \Rightarrow$ H_o.
  - $\alpha = .03 \Rightarrow$ REJECT H_o.
  - $\alpha = .02 \Rightarrow$ H_o.
  - $\alpha = .01 \Rightarrow$ H_o.

Instead of "imposing" YOUR CHOICE of α on a person who might be interested in your conclusions, the ***p-value*** allows him/her to assess the "rareness" of the observed event.

Ex. : #8.24, P. 333

$H_o : \mu = 10$

$H_a : \mu < 10$

$z = -2.33$

p-value $=$ Probability of a z-value

the one observed

- Smaller z-values are more indicative that H_a is true.
- ***p-value*** $=$

- H_o for any α that is $\geq .0099$.
- H_o for any α that is $< .0099$.

See page 234 of notes:

```
Z-Test

Test of mu = 10.000 vs mu < 10.000
The assumed sigma = 2.10

Variable     N     Mean    StDev   SE Mean       Z         P
JntsInsp    48    9.292    2.103     0.304   -2.33    0.0099
```


TWO - Tailed Test

- Find the area in whichever "tail" the z-value is in and ***DOUBLE IT***.

EX. : Have done a two-tailed test:

$$H_o : \mu = 17$$
$$H_a : \mu \neq 17$$
$$z = 1.75$$

- $p-$value = .
- $\alpha = .05$?
  - claim that $\mu \neq 17$ with $\alpha = .05$.

Ex. : #8.68, p. 352 In a "Pepsi Challenge", 100 Diet Coke drinkers were given unmarked cups of both Diet Coke and Diet Pepsi. 56 indicated that they preferred the taste of Diet Pepsi. Is there sufficient evidence to indicate that a majority of the Diet Coke drinkers will select Diet Pepsi in a blind taste test?

- $p =$ true proportion of Diet Coke drinkers who would select Diet Pepsi in a blind taste test.

$$H_a : \quad (1)$$
$$H_o : \quad (2)$$

- How???

Large Sample Tests About p (Section 8.5)

Interested in a POPULATION that contains an UNKNOWN but FIXED PROPORTION of items with a particular attribute $(\text{"}S\text{"})$.

- Recall the BINOMIAL EXPERIMENT.
  - $*$ $p =$ the proportion of Diet Coke drinkers who select Diet Pepsi in a blind taste test.
  - $*$ $p =$ the proportion of batteries that fail before guarantee expires.
- GOAL : Test hypotheses about p based on a "large" number of trials
  - $*$ $n =$ # of trials ; $x =$ # of "S" in the n trials
- Estimate for p

$$\hat{p} = \frac{x}{n} = \frac{\#\text{"}S\text{" in the sample}}{\text{sample size}}.$$

- If n is "large"
  - $*$ $\hat{p}$ has an distribution
  - $*$ $\mu_{\hat{p}} = p \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$
  - $*$ That is
    $$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$
    has an approximate distribution.

Consider testing

$$H_o : p = p_o \quad \text{a fixed particular value of } p$$

versus

$$H_a = \begin{cases} p > p_o & I \\ \text{OR} \\ p < p_o & II \\ \text{OR} \\ p \neq p_o & III \end{cases}$$

- if $H_o : p = p_o$ is the null hypothesis, TEST STATISTIC
  $$z = \frac{\text{estimator}^* - \text{hypothesized value}^{**}}{\text{standard error}^*}$$
  - $*$ Estimator and Standard Error from Formula Sheet
  - $**$ Hypothesized Value from NULL hypothesis
  - $\star$ $z = \dfrac{\widehat{p} - p_o}{\sqrt{\dfrac{p_o q_o}{n}}}$
- If H_o is true $(p = p_o)$, z has a STANDARD NORMAL distribution
- Rejection Regions (RR):

$$H_a = \left\{ \begin{array}{lll} p > p_o & I & z > z_\alpha \\ \text{OR} & & \\ p < p_o & II & z < -z_\alpha \\ \text{OR} & & \\ p \neq p_o & III & z < -z_{\alpha/2} \text{ or } \\ & & z > z_{\alpha/2} \end{array} \right\} \textbf{RR}$$

Ex. : #8.68, p. 352 In a "Pepsi Challenge", 100 Diet Coke drinkers were given unmarked cups of both Diet Coke and Diet Pepsi. 56 indicated that they preferred the taste of Diet Pepsi. Is there sufficient evidence to indicate that a majority of the Diet Coke drinkers will select Diet Pepsi in a blind taste test?

- $p =$ true proportion of all voters who think health care reform is the leading priority

$$H_a : p > .50 \quad (3)$$
$$H_o : p = .50 \quad (4)$$

- $\alpha = .05$ level test, RR :

- Assumptions : the 100 individuals participating in the the Pepsi Challenge are a SAMPLE of all Diet Coke drinkers. Note: n is "large"

- Data : $n = 100 \qquad \hat{p} = \underline{\qquad} =$

$$z = \frac{\qquad}{\qquad} = \frac{\qquad}{.05} =$$

- Conclusion :

  reject H_o in favor of H_a ***AT THE*** $\alpha = .05$ ***LEVEL***!!

- In terms of this problem:
  " claim that there is sufficient evidence at the level of significance" (or with 95% confidence) to indicate that the majority of Diet Coke drinkers will select Diet Pepsi in a blind taste test.

- $p-$ value?

- $p-$value =

Minitab?

- Stat→Basic Statistics→1 Proportion
- Click radio button "Summarized Data", type in Number of trials, Number of Successes
- Click Options, Select Alternative, Type in Null Value
- Click Box "Use test and interval based on normal distribution", OK, OK

```
Test and Confidence Interval for One Proportion

Test of p = 0.5 vs p > 0.5

Sample    X    N   Sample p          90% CI          Z-Value    P-Value
1        56  100   0.560000   (0.462710, 0.657290)     1.20      0.115
```