7. STATISTICAL INFERENCE: FPP Ch 21,23,26,27
7.1 CONFIDENCE INTERVALS

• We select a sample (subset) from the population
We compute a statistic based on the sample, to estimate the population parameter.

• We can use a percentage in a sample to estimate a percentage in a population.

• We can use a sample average to estimate a population mean.

• But there is always chance error.

• We know the percentage or fraction (average of "0" and "1" counts), or a sample average, will, for large samples, have the normal distribution shape.

• We can use the normal distribution to figure the chances that our sample average is within some amount of the true value: that is, that the chance error is smaller than some amount.

• We construct a confidence interval from our sample.

• For example, a 95% confidence interval covers the true value in 95% of repetitions of the sampling process.

• NOTE 1: the chances are in the sampling, not in the true value.

• NOTE 2: sampling from a large population (without replacement) is just like sampling from a box (with replacement).

Blank page for your notes

7.2 CONFIDENCE INTERVAL FOR A PERCENTAGE

- Suppose we want to estimate the percentage of households in a very large city with incomes over $50K.
- We take a sample of households: this is like sampling from a box of "0" and "1", but we do not know the fraction of "1"s.
- We observe the percentage of households is our sample with incomes over $50K. This is our <u>estimate</u> of the population percentage – or the fraction of "1"s in the box.
- The <u>expected value</u> (EV) of our estimate is the true population percentage.
- The SE is $\sqrt{\text{fraction of 1} \times \text{fraction of 0}}/\sqrt{\text{sample size}}$
- But we do not know the fraction of "1"s: use the sample fraction in the SE formula, to get an <u>estimated SE</u>.
- Now we know (sample-value - EV)/SE is like a z-score. We know it is between -2 and +2, with 95% chance.
- So the interval from (observed- 2×SE) to (observed + 2×SE) is a 95% confidence interval for the true population percentage.
- That is, for 95% of samples the confidence interval will include the true value.
- The chances are in the sampling, not in the true value.

Blank page for your notes

7.3 EXAMPLE

• We take a random sample of 1000 households from our city.

We find 400 (40% or fraction 0.4) have incomes over $50K per year.

• We estimate that 40% of households in the city have incomes over $50K per year, but we also want to know how accurate our estimate is likely to be.

• We estimate the SE for this percentage:

$\sqrt{0.4 \times 0.6}/\sqrt{1000} = 0.49/31.62 = 0.015$ **or 1.5%**

• Our 95% confidence interval is from (40-2×1.5) to (40 + 2×1.5), or from 37% to 43%.

• Our 68% confidence interval is from 38.5% to 41.5%

• If a large number of people take samples, and construct a confidence interval in this way, then 95% of the 95% confidence intervals will cover the true value.

• In polls, our 95% confidence interval is often stated as 40% plus-or-minus 3 percentage points.

Blank page for your notes

7.4 CONFIDENCE INTERVAL FOR A POPULATION MEAN

- **Now we want a confidence interval for a population mean – for example, mean household income in the city.**
- **Our estimate is the sample average: for example, for 1000 sampled households, as above, suppose \$48K.**
- **The SE for the sample average is**

  $$(\text{SD of box})/\sqrt{\text{sample size}}$$

- **But we do not know SD of the population (or box). So use SD of the sample, as an estimate – for example \$15K.**
- **Estimated SE is 15,000/$\sqrt{1000}$ = \$470**
- **The 95% confidence interval for the mean household income in the city is from 48,000-2×470 to 48,000 + 2× 470, or \$47,060 to \$48,940.**
- **Note again the randomness is in the sample: 95% of intervals constructed from samples in this way will cover the true value.**
- **We do not know which 95%: we do not know whether our particular interval does or doesn't.**
- **Note we are NOT measuring the spread of household incomes in the city: we are measuring our uncertainty about the MEAN household income in the city.**

Blank page for your notes

7.5 OVERVIEW OF INFERENCE

• Population has some histogram of values, but we do not know it.

• Histogram for a simple random sample of subjects should be "somewhat like" the population histogram.
NOTE: these histograms are NOT bell-shaped.

• So we use the sample values to estimate the population values

Use the sample percent over $50K to estimate population percent of households over $50K.

Use the sample mean to estimate population mean.

• But we need to know the size of the chance error.

• So we need an SE, but we do not know the population SD needed to compute it.

• So we use the sample SD to approximate the population SD.

• Then we can figure the relevant SE.

• Then we can figure z-score, confidence intervals or P-values using this estimated SE.

Blank page for your notes

7.6 WHICH SE ?

• If a question specifies the SD, we can use it to compute an SE.

• We than then use the SE to compute z-scores and chances.

CONFIDENCE INTERVAL

• If we are constructing a confidence interval, we must use the sample SD to compute an estimated SE.

• We then use this estimated SE in computing the confidence interval.

HYPOTHESIS TESTING: (next page)

• If a hypothesis specifies the SD, then use the hypothesized value to compute an SE.

• If the hypothesis specifies only a mean, not an SD, then again use the sample SD to compute an estimated SE.

• To test a hypothesis, we compute a z-score, and hence get the chance of observing something as-or-more extreme if the null hypothesis is true (the P-value).

Blank page for your notes

7.7 HYPOTHESES AND SIGNIFICANCE TESTS

• A hypothesis is a statement about a population value, or chance process. For example:

This is a fair coin.
40% of households have incomes over $50K.
Mean household income in this city is $48K.

• Our data (tosses of coin, incomes in sample of households) then tell us whether we can reject this null hypothesis.

• For example, if we see 370 households in our sample of 1000 (that is 37%) can we reject that the percentage in the population is 40%.

• That is, is the difference we see significant or could it have just happened by chance?

• We consider whether our results could happen "by chance" if the null hypothesis is true.

• If the null hypothesis is true, the EV for the percentage of households is 40% and the SE is 1.5% (see 7.3).

• So our z-score would be (37 -40)/1.5 = -2.

• The chance of getting a z-score at least as big as this (in size) is only 5%.

• We say the significance level is 5% (P=0.05).

This means, if the null hypothesis is true the chance of being this far out (or further) is only 5%.

• Small significance levels are evidence against the null hypothesis.

Blank page for your notes

7.8 TESTING A GIVEN VALUE OF A PERCENTAGE

- Question: Is this a fair coin?
- Null hypothesis: this is a fair coin.
- Data: the percentage of heads in N tosses.
- If the null hypothesis is true, the EV is 50% or 0.5 and the SE is $\sqrt{(0.5) \times (0.5)}/\sqrt{N} = 0.5/\sqrt{N})$

	N=50 SE=0.071=7.1%		N=500 SE=0.022=2.2%		N=5000 SE=0.007=0.7%	
obsved	z-score	signif	z-score	signif	z-score	signif
50%	0	100%	0	100%	0	100%
48%	-0.28	78%	-0.9	37%	-2.8	0.5%
54%	+0.56	48.4%	1.82	7%	+5.6	≈0
60%	+1.40	16%	4.54	0.001%	—	—
36%	-1.97	5%	—	—	—	—
30%	-2.81	0.5%	—	—	—	—

- The significance level (P-value) measures the chance of getting a value at least as far from the EV as observed, if the null hypothesis is true.
- If we toss a fair coin 50 times, the chance we get more than 54% heads or less than 46% heads is quite large (48.4%). If we see 54% heads we cannot reject that the coin is fair.
- If we toss a fair coin 500 times, the chance we get more than 54% heads or less than 46% heads is only 7%. If we see 54% heads we might suspect the coin is not fair.
- If we toss a fair coin 5000 times, the chance we

get more than 54% heads or less than 46% heads is practically 0. If we see 54% heads we will reject that the coin is not fair.

- Recall the Law of Averages.

Rest of this page for your notes

7.9 TESTING A GIVEN VALUE OF A MEAN

- **Question: Is the mean height of women 65 inches?**
- **Hypothesis: The mean height of women is 65 inches.**
- **Data: heights of sample of women from the population**
  Avg of these heights; SD of these heights.
- **EV for sample average = population (or box) mean**
  SE for sample average = (SD of box)/$\sqrt{\text{sample size}}$
- **But we do not know the SD of the box, so we estimate it by the SD of the sample.**
- **If hypothesis is true: z-score = (sample avg - 65)/SE.**
- **Example, sample 100 women:**
  sample average= 64.5 inches, sample SD=3 inches.
  Estimated SE for sample average = 3/$\sqrt{100}$ = 0.3"
  z-score = (64.5 - 65)/0.3 = 1.67
  From FPP A-105: between-area = 90%
  Significance level (P-value) = 10%
- **This one, we likely would not reject the null hypothesis.**
- **Suppose same sample avg and SD, but 500 women.**
  Estimated SE = 3/$\sqrt{500}$ = 0.134
  z-score = (64.5-65)/0.134 = 3.72
  From FPP A-105: between-area = 99.98%
  Significance level (P-value) = 0.02 % = 0.0002.
- **CLEARLY, now we reject the hypothesis.**

- **NOTE: 64.5 inches seems close to 65 inches.**
  Most women differ from the mean by more than 0.5"
  But our sample contains both taller and shorter women: the sample average should be very close to the population mean.
- **The SE tells us how close.**

Rest of this page for your notes

7.10 THE SE OF A DIFFERENCE

• Often we are interested in differences:
in height between fathers and sons
in income between men and women retirees
in percentage of polio cases among vaccinated and controls

• We know the SE for each sample average or percentage. What is the SE for the difference?

• It is larger than each SE, because there is chance error in both averages or percentages.

• If is smaller than the sum of the SE's: chance errors average out a bit.

• In fact,
SE of difference $= \sqrt{(\textbf{first SE })^2 \ + \ (\textbf{second SE})^2}$

• Example:
Sample 100 men aged 50-65,
avg height = 70 inches, SD = 4 inches
Sample 200 men aged 20-35,
avg height = 72 inches, SD = 4.25 inches
First SE $= 4/\sqrt{100} = \mathbf{0.4}$ **inches.**
Second SE $= 4.25/\sqrt{200} = \mathbf{0.3}$ **inches.**
SE of difference $= \sqrt{(0.3)^2 + (0.4)^2} = \mathbf{0.5}$.

• Observed difference = 2 inches = 4 SE.
VERY HIGHLY SIGNIFICANT

• We reject the hypothesis the means are equal.

Blank page for your notes

7.11 TESTING EQUALITY OF MEANS

• In fact, we can get a z-score, and a P-value for testing whether two population means are equal.

Example: Is there a difference in mean incomes of male and female retirees?

Null hypothesis: there is no difference.

• Again, we do not know the SD of the box (or population values), so we use the sample SD to estimate it.

• Example: in units of \$1000:
500 male retirees, mean income 50, SD = 8
400 female retirees, mean income = 49, SD = 6

• For men: SE = 8 $/\sqrt{500}$ = 0.36 (or \$360)
For women: SE = $6/\sqrt{400}$ = 0.30 (or \$300)
SE of difference = $\sqrt{(0.36)^2 + (0.30)^2}$ = 0.469

• z-score = ((50-49) -0)/0.469 = 2.13
Significance level (or P-value) is 0.035 or 3.5%

• We reject the null hypothesis of no difference, since the P-value is less than 5%, but only just.

• A 95% confidence for the difference is
\$1000 $\pm 2\times$ \$ 469 or \$62 to \$1938.

• Note, we are not testing which gender has higher income, only whether there is a difference. (In FPP: this is a two-sided test. We will do ONLY two-sided tests.)

Blank page for your notes

7.12 THE HOMEOPATHY STUDY 9
TESTING MEAN DIFFERENCES: LAB 1 RESULTS

• • • Of Treatment vs Control:

• For the Treatment: mean = 7.94, SD = 3.28
For the Controls mean = 7.13, SD = 3.55

• SE (Trt) = 3.28/$\sqrt{20}$ = 0.73, SE(Cnt) = 0.79,
SE (diff) = $\sqrt{0.73^2 + 0.79^2}$ = 1.08.
Z = (7.94 -7.13)/1.08 = 0.75: NO significant difference.

• • • Over time: First half vs Last half:
• For the first 20, mean = 5.24, SD = 1.91
For the last 20, mean=9.84, SD = 2.98

• SE (First) = 0.43, SE(second) = 0.67, SE(diff) = 0.80
Z = (9.84 -5.24)/0.80 = 5.75: HIGHLY SIGNIFICANT

• • • Of Treatment vs Control: correcting for time:

• Recall the analysts adjusted for time by fitting a curve (5.11). We can test for treatment effects, corrected for time, by testing the RESIDUALS from the fitted curve.

• For the time-adj Treatment: mean = -0.50, SD = 2.56.
For the time-adj Controls: mean = 0.08, SD = 1.08.

• SE(adj-Trt) = 0.57, SE(adj-Cnt) = 0.24,
SE(diff) = 0.62. Z = (0.57 - 0.24)/0.62 = 0.53
STILL NOT SIGNIFICANT: The time-effect is real, but it apparently did not mask a small treatment effect.

Blank page for your notes

7.13 TESTING EQUALITY OF PROPORTIONS

- **The Boston School of Public Health birth defects and childhood leukaemia study, Woburn MA, 1976-1982.**
Women drinking well water: 414. Birth defects 16.
Women not drinking this water: 228. Birth defects 3.

- **Does toxic waste in the well cause birth defects?**
Null hypothesis: the rates of birth defects are equal.

- **For 414 women drinking water from well:**
Fraction = 16/414 = 0.039 = 3.9%.
SE = $\sqrt{0.039 \times 0.961}/\sqrt{414} \;=\; 0.0095$

- **For 228 women not drinking this water:**
Fraction = 3/228 = 0.013 = 1.3%.
SE = $\sqrt{0.013 \times 0.987}/\sqrt{228} \;=\; 0.0075$

- **Observed difference in percentage = 3.9- 1.3 = 2.6%**
SE of difference = $\sqrt{(0.0095)^2 + (0.0075)^2}$ **= 0.0121 = 1.2%**

- **95% confidence interval for difference is**
(2.6 ± $2\times$**1.2) or 0.2% to 5%, which does not contain 0.**

- **z-score = (2.6 -0) /1.2 = 2.17.**
Significance (or P-value) = (100 -97)% = 3 % =0.03.
We reject the null hypothesis.

- **This is firm evidence of a difference.**
It does NOT prove that toxic waste is cause.
But this was careful study, with many other factors controlled for.

Blank page for your notes