Chapter 8

This chapter introduces the concept of *statistical testing*. It's all about the degree to which statistical evidence argues *against* a hypothesis. Here are a couple of tests in action:

a. Republicans have pre-election poll results. The poll is equal-probability, with-replacement, random sample of n = 400 likely voters. The poll finds 171 of them intend to vote Republican. If the *population* vote is actually *tied* what is the probability we'd have seen so few Republican votes *just by chance?*

b. Quality control engineers at a production facility are tasked with setting up an on-going inspection program. At prescribed intervals, items are to be selected from production and inspected. If too many of these are found defective then production will be stopped until remedies have been undertaken. How to decide what "too many" should be, and how frequently to sample?

pg. 287 Null hypothesis H0, test statistic, rejection region, rejecting or failing to reject H0 based upon the sample information.

pg. 288 Type I error (rejecting H0 when it is true) and type II error (failing to reject H0 when it is false). Fixing alpha = P(type I error) and evaluating beta = P(type II error).

pg. 301 t-test (pg. 299 is the special case of z-test).

pg. 308 Sample size needed to achieve specified alpha and beta.

pp. 313, 314, 315

pg. 319 Statistical significance vs practical significance.

pg. 320 Likelihood ratio principle.

Chapter 8 exercises, below, are based on *examples* from the book. They are due Wednesday 11-19-08 at the close of class.

Example 8.1 Use cumulative binomial Table A1. But use the null hypothesis H_0: p = 0.2 versus alternative hypothesis H_a: p > 0.2, a sample size of n = 20, and rejection region {7, 8, ... , 19, 20}. Show your work with detail similar to 8.1.

a. Determine type I error probability α.

b. Type II error probability $\beta(0.25)$ when p = 0.25.

c. Circle your answers to (a) (b), in the plot of

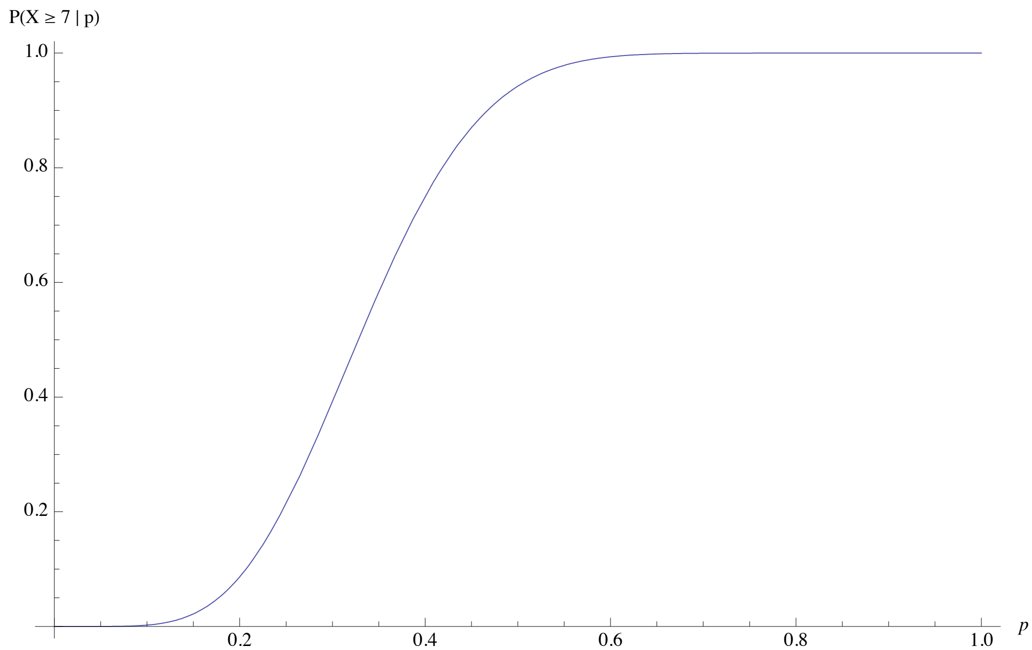
$$P(\text{reject } H_0 \mid p) \text{ for } 0 \le p \le 1$$

just below.

In[150]:=
```
power[p_] :=
 Apply[Plus,
  Table[ (20! / (x! (20 - x)!)) p^x (1 - p)^(20 - x), {x, 7, 20}]]
```

In[151]:=
```
Plot[power[p], {p, 0, 1}, AxesLabel → {p, "P(X ≥ 7 | p)"}]
```

Out[152]=

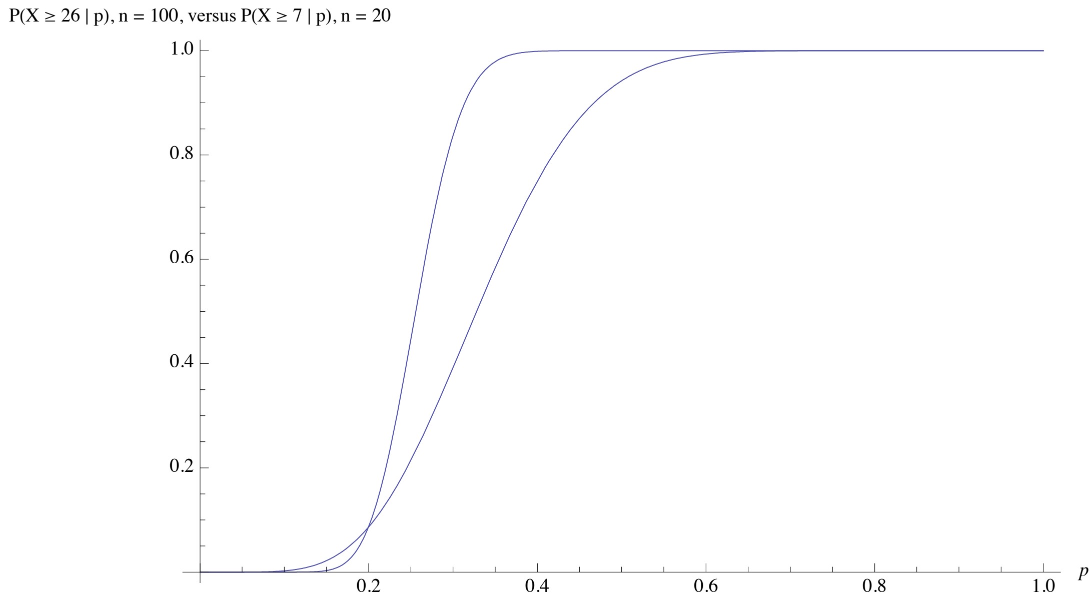


d. Below, plot (c) has been overlaid with a plot for the test of the same hypotheses, but using n = 100, rejection region {26, 27, ... , 99, 100}.

In[153]:=
```
power100[p_] :=
 Apply[Plus,
  Table[ (100! / (x! (100 - x)!)) p^x (1 - p)^(100 - x), {x, 26, 100}]]
```

In[154]:=
```
Show[Plot[power100[p], {p, 0, 1},
  AxesLabel → {p, "P(X ≥ 26 | p), n = 100, versus P(X ≥ 7 | p), n = 20"}],
 Plot[power[p], {p, 0, 1}, AxesLabel → {p, "P(X ≥ 7 | p)"}]]
```

Out[154]=



d1. Note that both tests have around the same α (their curves appear to cross at p = 0.2).

d2. Note that n = 100 test uses a lot more information. Its plot should be around $\sqrt{100/20} = 2.23$ times steeper at its steepest point. Does it seem to be so?

d3. Plot an "ideal" curve for the n = ∞ for these hypotheses. What, ideally, are

P(reject | p = 0.199999999999999999999999999999),

P(reject | p = 0.200000000000000000000000000001)?

Example 8.2 When σ is known we can for every n > 0 exploit the fact that for N(μ, σ^2) samples (in control)

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

has z-distribution. Tests and CI exploit the above.

Consider the null hypothesis H_0: $\mu = 70$ versus alternative hypothesis H_a: $\mu < 70$. Assume $\sigma = 10$, a sample size of n = 20, and rejection region defined by cutoff c = -1.5. That is, reject H_0 if

$$\text{test statistic } \frac{\overline{\overline{X}} - 70}{\frac{10}{\sqrt{20}}} < -1.5$$

a. Determine type I error probability α.

b. Determine type II error probability β when $\mu = 67$. Use:

$$P\left(\frac{\overline{\overline{X}} - 70}{\frac{10}{\sqrt{20}}} < -1.5 \,\Big|\, \mu\right) = P\left(\frac{\overline{\overline{X}} - \mu + \mu - 70}{\frac{10}{\sqrt{20}}} < -1.5 \,\Big|\, \mu\right)$$

$$= P\left(\frac{\overline{\overline{X}} - \mu}{\frac{10}{\sqrt{20}}} < \frac{70 - \mu}{\frac{10}{\sqrt{20}}} - 1.5 \,\Big|\, \mu\right) = P\left(Z < \frac{70 - \mu}{\frac{10}{\sqrt{20}}} - 1.5\right)$$

$$= \Phi\left(\frac{70 - \mu}{\frac{10}{\sqrt{20}}} - 1.5\right)$$

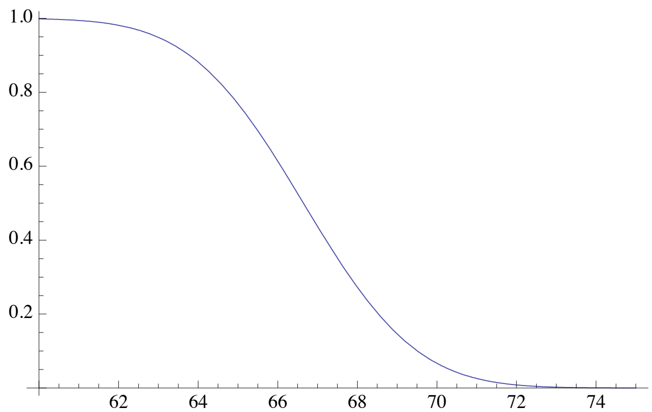
at $\mu = 67$. The beauty of knowing $\sigma = 10$ is that it makes finding β perfectly simple in the case of normal observations. It is another benefit of being under statistical control.

c. Circle your answers to (a) (b) in the plot, just below, of $P(\text{reject } H_0 \mid \mu) = \Phi\left(\frac{70 - \mu}{\frac{10}{\sqrt{20}}} - 1.5\right)$ vs μ.

In[155]:=
```
Φ[z_] :=
  Integrate[Exp[-r^2 / 2] / Sqrt[2 Pi], {r, -∞, z}]
```

In[156]:= $\text{Plot}\left[\Phi\left[\frac{70-\mu}{\frac{10}{\sqrt{20}}}-1.5\right], \{\mu, 60, 75\}\right]$

Out[156]=

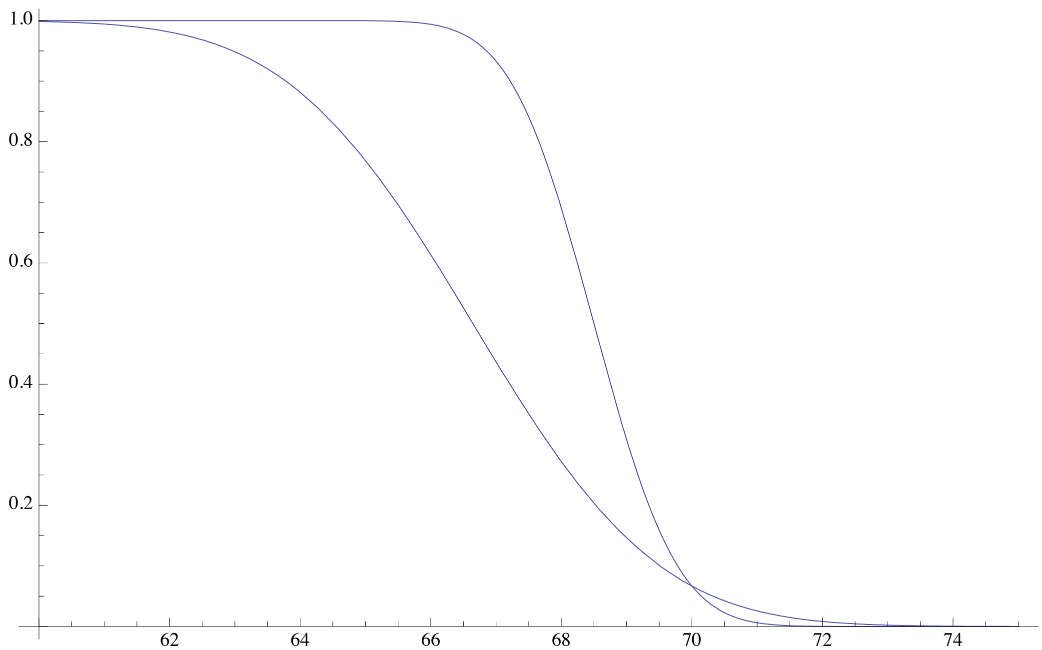


d. Below, plot (c) has been overlaid with a plot for the test of the same hypotheses using n = 100, with test statistic and rejection region defined by

$$\text{test statistic } \frac{\overline{X}-70}{\frac{10}{\sqrt{100}}} < -1.5$$

In[157]:= $\textbf{Show}\left[\textbf{Plot}\left[\Phi\left[\frac{70-\mu}{\frac{10}{\sqrt{20}}}-1.5\right],\{\mu,60,75\}\right],\textbf{Plot}\left[\Phi\left[\frac{70-\mu}{\frac{10}{\sqrt{100}}}-1.5\right],\{\mu,60,75\}\right]\right]$

Out[157]=



d1. Note that both tests have the same α (their curves cross at $\mu = 70$).

d2. Note that this other test uses a lot more information, 100 samples vs 20. Its plot should be around $\sqrt{100/20} = 2.23$ times steeper at its steepest point. Does it seem to be so?

d3. Plot an "ideal" curve for the n = ∞ for these hypotheses. What, ideally, are

P(reject | μ = 70.0000000000000000000000000001),

P(reject | μ = 69.0000000000000000000000000001)?

d4. Calculate β at $\mu = 67$ for the n = 100 test and compare with its value in the plot.