Exercise Set 7-1

1) a) The standard deviation of the observations is σ. This comes from the fact that the second parameter of the normal distribution is equal to the distribution's variance (equation 5.22).
b) The standard error of the estimate is $\mathrm{SE}(\hat{\theta}_n) = \sigma/\sqrt{n}$. Recall from exercise set 6-2, problem 1, that $\mathrm{Var}(\hat{\theta}_n) = \sigma^2/n$. The standard error follows immediately from this fact and equation 7.1.
c) Plugging in the numbers in the problem gives $\mathrm{SE}(\hat{\theta}_n) = 1/5 = 0.2$.
d) Use the `norm.samps()` function (from exercise set 6-1, problem 2, and reprinted here)

```
norm.samps <- function(mu = 0, sigma = 1, n = 25, nsamps =
10000){
  samps <- rnorm(n*nsamps, mu, sigma)
  samp.mat <- matrix(samps, nrow = nsamps, ncol = n)
  return(samp.mat)
}
```

to draw a set of samples:

```
> s.mat <- norm.samps(0, 1, 25, 10000)
```

Calculate the median of each sample using `apply()` (or using the `for()` loop given in exercise set 6-1, problem 2):

```
> ests.median <- apply(s.mat, 1, median)
```

Use the `sd()` command to estimate the standard error of the sample median. (Or use `sqrt(var())`):

```
> sd(ests.median)
```

You will get an answer of approximately 0.25, which is larger than the standard error of the sample mean.

2) We seek the probability $\mathrm{P}(\hat{\theta}_n - \omega < \theta < \hat{\theta}_n + \omega)$. θ is fixed, and in this problem, so is ω, so this is really a probability statement about $\hat{\theta}_n$, the only random variable in the expression. The statement

$$\hat{\theta}_n - \omega < \theta < \hat{\theta}_n + \omega$$

is equivalent to

$$\theta - \omega < \hat{\theta}_n < \theta + \omega,$$

so $\mathrm{P}\left(\hat{\theta}_n - \omega < \theta < \hat{\theta}_n + \omega\right) = \mathrm{P}\left(\theta - \omega < \hat{\theta}_n < \theta + \omega\right) = \mathrm{P}\left(\hat{\theta}_n < \theta + \omega\right) - \mathrm{P}(\hat{\theta}_n < \theta - \omega)$. Recalling that normally distributed random variables are continuous, we know that $\mathrm{P}\left(\hat{\theta}_n < \theta +\right.$

$\omega\big) = \mathrm{P}(\hat{\theta}_n \leq \theta + \omega)$, and therefore, if the cumulative distribution function of $\hat{\theta}_n$ is written $F_{\hat{\theta}_n}$, then the probability we seek is

$$F_{\hat{\theta}_n}(\theta + \omega) - F_{\hat{\theta}_n}(\theta - \omega).$$

Because $\hat{\theta}_n$ is normally distributed with expectation θ and standard deviation ω, this is the probability that a normally distributed random variable falls within 1 standard deviation of its mean. You can look up this probability in a table or use `R`'s `pnorm()` function, which evaluates the cumulative distribution function of the normal distribution:

```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
```

Thus, the interval $(\hat{\theta}_n - \omega, \hat{\theta}_n + \omega)$ would contain θ about 68% of the time. There is a subtle but important point here, which will be emphasized in the next section: the probability statement is about the *interval*, which is random, and not about θ, which is fixed.

b) Reasoning along similar lines as in part (a), we evaluate the necessary probability in `R` as

```
> pnorm(2)-pnorm(-2)
[1] 0.9544997
```

Thus, the interval $(\hat{\theta}_n - 2\omega, \hat{\theta}_n + 2\omega)$ will contain θ about 95% of the time.

3) The first question is to evaluate $\mathrm{E}\left(\left[\hat{\theta}_a - \hat{\theta}_b\right]^2\right)$. One way to do it is to expand the squared term and use the linearity of expectation to write

$$\mathrm{E}\left(\left[\hat{\theta}_a - \hat{\theta}_b\right]^2\right) = \mathrm{E}\left({\hat{\theta}_a}^2 - 2\hat{\theta}_a\hat{\theta}_b + {\hat{\theta}_b}^2\right) = \mathrm{E}\left({\hat{\theta}_a}^2\right) - 2\mathrm{E}\left(\hat{\theta}_a\hat{\theta}_b\right) + \mathrm{E}\left({\hat{\theta}_b}^2\right).$$

Remember that $\hat{\theta}_a$ and $\hat{\theta}_b$ are independent, which guarantees that $\mathrm{E}\left(\hat{\theta}_a\hat{\theta}_b\right) = \mathrm{E}\left(\hat{\theta}_a\right)\mathrm{E}(\hat{\theta}_b)$ (exercise set 5-2, problem 3c). Because the samples are identically distributed and of equal size, and because the same estimator is applied to each sample, the two estimates are identically distributed. This means that $\mathrm{E}\left(\hat{\theta}_a\right) = \mathrm{E}\left(\hat{\theta}_b\right)$, which implies that $\mathrm{E}\left(\hat{\theta}_a\right)\mathrm{E}\left(\hat{\theta}_b\right) = \mathrm{E}\left(\hat{\theta}_a\right)^2 = \mathrm{E}\left(\hat{\theta}_b\right)^2$, and it also implies that $\mathrm{E}\left({\hat{\theta}_a}^2\right) = \mathrm{E}\left({\hat{\theta}_b}^2\right)$. Combining these insights, we have

$$\mathrm{E}\left(\left[\hat{\theta}_a - \hat{\theta}_b\right]^2\right) = 2\mathrm{E}\left({\hat{\theta}_a}^2\right) - 2\mathrm{E}\left(\hat{\theta}_a\right)^2 = 2\mathrm{Var}\left(\hat{\theta}_a\right) = 2\omega^2.$$

Thus, the expected squared difference between two estimates derived by applying the same estimator to two independent, equally-sizes samples is twice the square of the standard error of the estimate. Similarly, the square root of the expected squared difference is $\sqrt{2}\omega$.

Exercise Set 7-2

1) a) $z_{a/2} \approx 0.674$, which you can verify with the `R` command `qnorm(0.75)`. The following `R` code plots an appropriate picture.

```
a <- 0.5 #parameter: 1 minus the confidence level.

#plot a standard normal distribution
x <- seq(-3.5, 3.5, length.out = 10000)
z.a2 <- qnorm(1 - a/2)
fx <- dnorm(x)
plot(x, fx, type = "l")

#Shade in the appropriate area
x.zs <- seq(-z.a2, z.a2, length.out = 10000)
fx.zs <- dnorm(x.zs)
polygon(c(-z.a2, x.zs, z.a2), c(0, fx.zs, 0), col = "grey",
border = FALSE)
```

The area of the shaded region is $1/2$.

b) By equation 7.3, the confidence interval was constructed as $\left(\hat{\theta} - \omega z_{a/2}, \hat{\theta} + \omega z_{a/2}\right)$. Thus, $\hat{\theta}$ is at the midpoint between the boundaries of the interval—in this case, $\hat{\theta} = 3$.

c) The lower bound of the interval is 2, and $\hat{\theta} = 3$, giving the equation $3 - \omega z_{a/2} = 2$, which implies that $\omega z_{a/2} = 1$. By part a, $z_{a/2} \approx 0.674$, and it follows that the standard error is $\omega = 1/z_{a/2} \approx 1/0.674 \approx 1.48$.

d) For a 95% confidence interval, we need $z_{0.025}$, which is approximately 1.96. The 95% confidence interval is then $\left(\hat{\theta} - \omega z_{a/2}, \hat{\theta} + \omega z_{a/2}\right) \approx (3 - 1.48 * 1.96, 3 + 1.48 * 1.96) \approx (0.09, 5.91)$. This is a larger range than that covered by the 50% confidence interval.

e) $3/1.48 \approx 2.02$. The estimate was about two standard errors away from zero.

f) We need the value of a that solves the equation $\hat{\theta} - \omega z_{a/2} = 3 - 1.48 * z_{a/2} = 0$. That is, we need the value of a that gives $z_{a/2} = 3/1.48 \approx 2.02$. The necessary value of a is thus one minus the probability that a random variable drawn from a normal distribution will fall within 2.02 standard deviations of its expectation. The probability that such a random variable will fall more than 2.02 standard deviations *above* its expectation is $1 - \phi(2.02)$, or in `R`, `1-pnorm(2.02)`. Because the normal distribution is symmetric, the probability that such a random variable will fall more than 2.02 standard deviations *above* its expectation *or* more than 2.02 standard deviations *below* its expectation is therefore double this quantity, or approximately 0.043. This is the necessary value of a.

g) If $\theta = 0$, then the probability of observing $|\hat{\theta}| > 3$ is the probability of observing a normal random variable more than $(\hat{\theta} - \theta)/\omega = 3/1.48 = 2.02$ standard deviations away from its expectation. By the reasoning in the solution to part (f), this probability is 0.043.

Exercise Set 7-3

1) There are many possible responses. It's a good idea to divide responses into two: first, are there reasons why, if the hypothesis were true, you would not be convinced that the theory is true? Second, are there reasons why, if the hypothesis were false, you would not be convinced that the theory is false? So first, suppose that I had found that Arizonan women were shorter than other women from the US. There are many reasons not to take such a finding as evidence in favor of the theory that smaller people are attracted to hotter climates. The most important is *confounding*. Arizona has many properties besides hot weather—for example, it is dry, it borders Mexico, its population includes a higher-than-average proportion of members of the Church of Jesus Christ of Latter-Day Saints, it contains most of the Navajo Nation, its population is less dense than average, it is a popular destination for retirees, and it has historically been a major center of copper mining. When one chooses to study Arizonans, one does not get a set of people who live in a place that is hotter than average and otherwise just like everywhere else. One instead gets a set of people who live in a place with a slew of unique or unusual properties. Any difference between Arizonans and other people could in principle—though perhaps not always plausibly—be due to any of these properties rather than to the Arizonan heat. That is confounding.

Secondly, a failure to find a significant difference between the heights of Arizonan women and other women does not kill the theory. For one thing, I only sampled 25 women. As we will see soon, a sample of 25 is not large enough to detect small or moderate differences between population means. The assumption that Arizona is a good test case for the effects of a hot climate is crude. Though the major population centers—Phoenix and Tucson—are hot, the northern part of the state is at high altitude and experiences reasonably cold winters. Sampling from across the state thus dilutes the effects of heat, possibly masking effects that would be apparent if only the hottest parts of the state were sampled. Using height—an imperfect proxy of body size—as the measured variable has the same effect. Confounding can also mask true effects in addition to generating spurious ones. It may be that smaller people are genuinely attracted to hotter climates but that some other feature of Arizona has attracted tall people, masking the effect.

This list of problems is not exhaustive but gives a sense of the many difficulties facing any empirical researcher.

2) a) The standard error is the standard deviation divided by the square root of the sample size, or in this case, 1mm.

b) From equation 7.3, a confidence interval is $\left(\hat{\theta} - \omega z_{a/2}, \hat{\theta} + \omega z_{a/2}\right)$, where $\hat{\theta}$ is the estimator, ω is the standard error, and $z_{a/2} = \phi^{-1}(1 - a/2)$, the $(1 - a/2)$th quantile of the standard normal distribution. For a 95% confidence interval, $z_{a/2}$ can be found using the `R` command `qnorm(0.975)`, and it is equal to 1.96. Thus, because $\omega = 1$, a 95% confidence interval is given by $(\bar{x} - 1.96, \bar{x} + 1.96)$, where $\bar{x}$ is the sample mean.

c) We start by finding the values of $\bar{x}$ that would give two-sided $p = 0.05$. The lower of these values is the same as would give a one-sided p of 0.025. Thus, we want to find the value of $\bar{x}$ that satisfies the equation $\Phi_{100,1}(\bar{x}) = 0.025$, where $\Phi_{100,1}$ is the cumulative distribution

function of the Normal(100,1) distribution, or equivalently, $P(\bar{X} \leq \bar{x} | \mu = 100) = 0.025$. That is, we need $\Phi^{-1}_{100,1}(0.025) = \bar{x}$. In R, we use the command `qnorm(0.025, mean = 100, sd = 1)` to get $\bar{x} = 98.04$, or $\bar{x} = 100 - 1.96$. By symmetry, the two-sided p is 0.05 if $\bar{x} = 100 + 1.96$. Thus, the two-sided p of a test of the null hypothesis that $\mu = 100$ will be less than 0.05 if either $\bar{x} < 98.04$ or $\bar{x} > 101.96$. Notice that, by the result of part (b), these are exactly the cases in which a 95% confidence interval for μ excludes 100.

d) Here is a function that computes a two-sided p for a test of a mean from a normal sample:

```
twotailed.p.normal <- function(x.bar, mu, stand.err){
  abs.diff <- abs(x.bar - mu)
  2 * pnorm(mu - abs.diff, mean = mu, sd = stand.err)
}
```

This function takes a sample mean (`x.bar`), the mean under the null hypothesis (`mu`), and the standard error of the sample mean (`stand.err`).

To simulate the means of 10,000 samples of size four, we have two options. We can either simulate the samples (here, storing them in a matrix) and take their means:

```
sim.mat <- matrix(rnorm(40000, mean = 100, sd = 2), ncol = 4,
     nrow = 10000)
sim.means <- rowMeans(sim.mat)
```

Or we can simulate the means directly, remembering that the means are normally distributed with expectation 100 and variance 1:

```
sim.means <- rnorm(10000, mean = 100, sd = 1)
```

To get the distribution of ps, we could use a `for()` loop:

```
ps <- numeric(10000)
for(i in 1:10000){
  ps[i] <- twotailed.p.normal(sim.means[i], 100, 1)
}
```

Or, even better, we could use `sapply()`, a version of `apply()` that takes vectors as input:

```
ps <- sapply(sim.means, FUN = twotailed.p.normal, mu = 100,
     stand.err = 1)
```

We can plot the distribution of the p values with `hist(ps)`. It is approximately uniform. To find the proportion of p values less than 0.05, use `mean(ps < 0.05)`. It should be approximately 0.05. Similarly, the proportion of p values less than 0.10 should be about 0.10. This is a good result—it means that when the null hypothesis is true, the test works approximately as advertised.

e) To simulate normal samples of size 4 from a Normal(101,4) distribution and test the null hypothesis that $\mu = 100$, use the following R code:

```
sim.mat <- matrix(rnorm(40000, mean = 101, sd = 2), ncol = 4,
     nrow = 10000)
sim.means <- rowMeans(sim.mat)
ps <- sapply(sim.means, FUN = twotailed.p.normal, mu = 100,
     stand.err = 1)
```

The distribution of p values is no longer uniform—it has a concentration of low p values, representing samples that would be unlikely to be drawn if μ were in fact 100. I find that about 17% of the p values are less than 0.05 and that about 26% are less than 0.10.

f) To simulate normal samples of size 4 from a Normal(102,4) distribution and test the null hypothesis that $\mu = 100$, use the following R code:

```
sim.mat <- matrix(rnorm(40000, mean = 102, sd = 2), ncol = 4,
     nrow = 10000)
sim.means <- rowMeans(sim.mat)
ps <- sapply(sim.means, FUN = twotailed.p.normal, mu = 100,
     stand.err = 1)
```

Again, the distribution of p values shows a concentration of low values, even more pronounced than in part (e). I find that about 51% of the p values are less than 0.05 and that about 64% are less than 0.10.

g) To simulate normal samples of size 16 from a Normal(101,4) distribution and test the null hypothesis that $\mu = 100$, use the following R code:

```
sim.mat <- matrix(rnorm(160000, mean = 101, sd = 2), ncol = 16,
     nrow = 10000)
sim.means <- rowMeans(sim.mat)
ps <- sapply(sim.means, FUN = twotailed.p.normal, mu = 100,
     stand.err = 1/2)
```

Again, the distribution of p values shows a concentration of low values, even more pronounced than in part (e). I find that about 51% of the p values are less than 0.05 and that about 64% are less than 0.10. Notice that these are the same values as in part (f): doubling the difference between the true mean and the mean under the null hypothesis had the same effect on the distribution of p values as quadrupling the sample size. There is a good reason for this—both changes have the effect of doubling the number of standard errors separating the true parameter from the value postulated by the null hypothesis.

Exercise Set 7-4

1) a) The proportion of the time that the null hypothesis is true is

$$\tau = \frac{t_n + f_p}{t_n + f_p + f_n + t_p} = t_n + f_p.$$

The simplification comes from the fact that the denominator $t_n + f_p + f_n + t_p = 1$.

b) The proportion of the time the null hypothesis is false is $\varphi = f_n + t_p$.

c) The power of the test is

$$\pi = \mathrm{P}(R|H_0^C) = \frac{t_p}{t_p + f_n}.$$

By part (b), this quantity is also equal to t_p/φ.

d) In the notation of the table, the false discovery rate is

$$\mathrm{P}(H_0|R) = \frac{f_p}{t_p + f_p}.$$

One way to express the false discovery rate in terms of τ, φ, γ, and π is to use Bayes' theorem. By Bayes' Theorem,

$$\mathrm{P}(H_0|R) = \mathrm{P}(R|H_o)\frac{\mathrm{P}(H_0)}{\mathrm{P}(R)}.$$

$\mathrm{P}(R|H_o) = \gamma$ by the definition in equation 7.4, and $\mathrm{P}(H_0) = t_n + f_p$ by part a. To get the remaining term, use

$$\mathrm{P}(R) = \mathrm{P}(R|H_0)\mathrm{P}(H_0) + \mathrm{P}(R|H_0^C)\mathrm{P}(H_0^C) = \gamma\tau + \pi\varphi.$$

Applying these identities gives

$$\mathrm{P}(H_0|R) = \frac{\gamma\tau}{\gamma\tau + \pi\varphi}.$$

You can verify that this expression is equivalent to $\mathrm{P}(H_0|R) = f_p/(f_p + t_p)$ by replacing τ, φ, γ, and π with their definitions in terms of t_n, f_p, f_n, and t_p. Because $\varphi + \tau = 1$, it is also possible to remove either φ or τ from the expression. For example, we could remove φ by writing

$$\mathrm{P}(H_0|R) = \frac{\gamma\tau}{\gamma\tau + \pi(1-\tau)}.$$

Because τ, γ, and π are all between 0 and 1, the false discovery rate decreases as the power of the test, π, increases. This is one excellent reason for valuing tests with high power to reject false null hypotheses.

e) In the notation of the table, the negative predictive value is

$$\mathrm{P}(H_0|R^C) = \frac{t_n}{t_n + f_n}.$$

Using reasoning parallel to that used in part (d), the negative predictive value is equal to

$$\mathrm{P}(H_0|R^C) = \frac{(1-\gamma)\tau}{(1-\gamma)\tau + (1-\pi)\varphi} = \frac{(1-\gamma)\tau}{(1-\gamma)\tau + (1-\pi)(1-\tau)}.$$

Because τ, γ, and π are all between 0 and 1, the negative predictive value increases as the power of the test, π, increases.

2) a) If the research group adopted the proposed procedure, then they would falsely reject the null hypothesis about 11% of the time, which is more than twice the nominal level of each of their tests. This is true even though each of the individual tests rejects the null hypothesis at the correct rate.

b) Other things equal, increasing the number of measurements increases the probability that at least one of the tests leads to an incorrect rejection of the null hypothesis, also called the type I error rate. Increasing the degree of correlation between the measurements toward 1 tends to decrease the probability that at least one of the tests leads to a rejection of the null hypothesis.

One way to control the familywise error rate is with Bonferroni correction, in which each p value is compared to the value γ/k, where γ is the desired familywise error rate and k is the number of hypothesis tests being conducted.

3) The proposed procedure leads to an incorrect rejection of the null hypothesis about 11-12% of the time, which grows worse with more repeated testing.

```
ps <- serial.testing.sim()

sigs <- ps < .05
colMeans(sigs)
mean(rowMeans(sigs) > 0)
```

Exercise Set 7-5

1) The following block of code provides one way to produce the necessary plot, assuming that the `ps.1sz()` function has been defined:

```
n <- 25
```

```
d <- seq(-2, 2, length.out = 101)
pow <- numeric(length(d))
for(i in 1:length(d)){
  pow[i] <- ps.1sz(d[i], n)
}
plot(d, pow, ylim = c(0,1), type = "l", ylab = "Power")
```

2) a) When d and n are large enough, power is near 1, and the estimated effect size from studies that rejected the null hypothesis is approximately correct. But for smaller d and n, the bias produced by the winner's curse can be substantial. For example, when $d = 0.25$ and $n = 25$, the mean estimated effect size in studies that reject the null hypothesis is approximately twice the true value. Here are some possible parameter choices:

```
wc.1sz( .3, 50, .05)
wc.1sz( .5, 50, .05)
wc.1sz( .1, 50, .05)
wc.1sz( .3, 25, .05)
wc.1sz( .3, 50, .01)
```

b) The following code is one way to produce the requested plots:

```
true.d <- 0.3
ns <- seq(5, 200, by = 5)
pows <- numeric(length(ns))
est.ds <- numeric(length(ns))

#Save power and estimated effect sizes.
for(i in 1:length(ns)){
  wc <- wc.1sz(true.d, ns[i])
  est.ds[i] <- wc[2]
  pows[i] <- wc[3]
}

#First Plot: Cursed effect size estimate as a function of sample
#size.
plot(ns, est.ds, type = "l", lty = 2, lwd = 2, ylim = c(0,
max(est.ds)), ylab = "d", xlab = "n")
lines(ns, rep(true.d, length(ns)), lwd = 2)
legend("topright", lwd = c(2,2), lty = c(2,1), legend =
c("Cursed", "True"))

#Second Plot: Size of the winner's curse effect as a function of
#power.
curse.size <- est.ds - true.d
plot(pows, curse.size, type = "l", lwd = 2, xlab = "Power", ylab
= "Size of Winner's Curse Effect")
```